Localized ensemble-based inversion of ERT data: synthetic test and application to mining exploration.

A. Vinciguerra[1], G. Marquis[1], J. Girard[1]

[1] University of Strasbourg - CNRS - ITES

Summary

Electrical Resistivity Tomography (ERT) is a geophysical method widely used in mining exploration and characterized by non-linear and ill-posed inverse problem. Even though deterministic inversion represents the standard approach in ERT inversion, Bayesian algorithms allow to tackle the non-linearity of the problem returning a posterior probability density function (pdf) as solution. The Ensemble Smoother Multiple Data Assimilation (ESMDA) is an ensemble-based Bayesian algorithm that solves the inverse problem approximating the sensitivity through cross-covariance and covariance matrices. Due to finite size of the ensemble, the covariance matrices can be contaminated by spurious correlations that affect the sampling of the pdf. We propose a distance-based localization approach to mitigate these effects, smoothing out or removing correlations far away from the data position. After encouraging synthetic test, we apply the proposed strategy to field dataset acquired in the Athabasca basin. The results suggest that the localized ESMDA allows to obtain smooth models while preserving resolution in the relevant portion of the model. In addition, sampling of the pdf helps result interpretation by evaluating the uncertainty associated with each model cell. This approach is particularly valuable in mining exploration, where decision about the drilling need to be taken.

Localized ensemble-based inversion of ERT data: synthetic test and application to mining exploration.

Introduction

Electrical Resistivity Tomography (ERT) is widely used in hydrogeology, civil engineering and mining exploration. It is characterized by a non-linear and ill-posed inverse problem, which brings about non-uniqueness of the solution and local minima. The common inversion approach involves the application of deterministic algorithms such as Gauss-Newton and its variations, which return a single resistivity model as solution. In some contexts, it is valuable to obtain information about the uncertainty of the solution to better interpret the resistivity model. The Bayesian inversion algorithms return the posterior probability density function (*pdf*) as solution of the problem, which can be analyzed in terms of its mean and standard deviation. These algorithms, for instance Markov Chain Monte Carlo, are usually computationally expensive involving tens of thousands iterations to converge and sample the posterior *pdf*. In this work, we focus on ensemble-based approach, applied in meteorology, ocean sciences and reservoir history matching. Instead of sampling the *pdf* through Markov chains, the algorithm approximates the distribution by means of an ensemble of models updated iteratively. In particular, we apply the adaptive version of the Ensemble Smoother Multiple Data Assimilation (ESMDA) (Emerick, 2016), based on the Kalman filter, which was already tested and applied in near surface context by Vinciguerra et al. (2024). The drawback of ESMDA, and in general of the ensemble-based approaches, is mainly that even the updates can be considered as a small Gauss Newton corrections (Reynolds et al., 2006), the sensitivity is not computed mathematically (through adjoint state) but approximated by means of cross-covariance and covariance matrices of the entire ensemble of models. The consequence is that the employment of a too small ensemble may cause uncertainty underestimation and spurious correlations that affect the *pdf* sampling. On the other hand, the use of a large ensemble may lead to a heavy computational burden.
This work aims to investigate the effect of localization in ESMDA inversion maintaining the ensemble dimension as small as possible. The localization regularizes updates to model parameters in function, for instance, of the distance between each cell of the model and observed data (Chen and Oliver, 2017). This strategy is widely applied in other contexts representing the state-of-the-art approach (Chen and Oliver, 2017), but as far as we are aware, it has never been applied to exploration geophysics. There exists various strategies to build the localization function, in this work we employ a distance based approach. First, we test the algorithm on a synthetic inversion simulating three highly conductive bodies within a resistive half-space, then we apply the algorithm to a field dataset acquired in Athabasca basin (Canada) for Uranium exploration purposes.

Methods

Adaptive ESMDA
ESMDA (Emerick and Reynolds, 2013) is an iterative algorithm that employs an ensemble of models to assess posterior *pdf* associated with the model space. In particular, from Reynolds et al. (2006), it arises that the algorithm performs multiple small Gauss-Newton-like corrections to approximate the posterior distribution. At each iteration and for each ensemble member, the update of resistivity model $\mathbf{m}$ is expressed by:

$$\mathbf{m}_k^{i+1} = \mathbf{m}_k^i + \mathbf{C}_{md}(\mathbf{C}_{dd} + \alpha^i \mathbf{C}_d)^{-1}(\mathbf{d}_{kpred} - \mathbf{d}^*_{obs}) \tag{1}$$

where the index i express the i-th iteration, the index k the k-th ensemble member, $\mathbf{C}_{md}$ denotes the cross-covariance between the ensemble of resistivity models and the predicted apparent resistivities, $\mathbf{C}_{dd}$ is the covariance of the ensemble of predicted data, $\mathbf{C}_d$ represents the covariance of the observed apparent resistivity, α^i is the damping factor and $\mathbf{d}_{pred}$ and $\mathbf{d}^*_{obs}$ are the predicted and perturbed observed apparent resistivity, respectively. The $\mathbf{C}_{md}$ matrix has dimension $n \times m$, the matrix $\mathbf{C}_{dd}$ $m \times m$ and $\mathbf{C}_d$ is a diagonal matrix of dimension $m \times m$, where n is the number of model parameters and m the number of data points. In this work we adopt the adaptive ESMDA, where α^i changes according to the value of normalized objective function, damping the model updates at early iterations (Vinciguerra et al., 2024). Introducing the Kalman filter, the 1 becomes:

$$\mathbf{m}_k^{i+1} = \mathbf{m}_k^i + \mathbf{K}(\mathbf{d}_{kpred} - \mathbf{d}^*_{obs}) \tag{2}$$

where it emerges the pivotal role of the Kalman filter, which multiply the difference between observed and predicted data.
The ensemble dimension represents a hyper parameter during the preliminary phase that is usually chosen proportional to the number of parameters. Indeed, a too small ensemble may produce ensemble collapse, uncertainties underestimation and spurious correlations in the matrices $\mathbf{C}_{md}$ and $\mathbf{C}_{dd}$ contaminating the Kalman filter. At the same time, since each ensemble member is associated to a forward calculation, reducing the ensemble dimension guarantees a lower computational cost.

Localization
Spurious correlations may greatly impact model updating (2) especially where we expect low or no sensitivity (Chen and Oliver, 2017). The algorithm approximates the sensitivity information through cross-covariance and covariance matrices computed from the ensemble of models. Consequently, without analytic computation of sensitivity, the resulting posterior *pdf* may be characterized by an unreliable and noisy mean model (Chen and Oliver, 2017).
In other contexts, such as in reservoir history matching, there exists a vast literature about reducing the impact of spurious correlations involving the concept of localization. As the term suggests, it means to "localize" the updates, i.e removing or smoothing-out correlations far away from the data position. In this work we focus on distance-based localization, which allows to relate the localization range to the expected investigation depth of the survey. The localization involves an element-wise product between a correlation type function and Kalman gain at each iteration following the relation:

$$\mathbf{m}_k^{i+1} = \mathbf{m}_k^i + \mathbf{C} \circ \mathbf{K}(\mathbf{d}_{pred} - \mathbf{d}^*_{obs}) \tag{3}$$

where $\circ$ indicates the element-wise product between the correlation function $\mathbf{C}$ and Kalman gain $\mathbf{K}$. The matrix $\mathbf{C}$ has dimension $n \times m$ where each i-th column contains a 2D function that expresses the correlation factor associated with each cell. The appropriate choice of the function may depend on different factors such as the range (critical distance) and the rate at which the spatial correlation diminishes with distance. Since the matrix $\mathbf{C}$ enters into the update of (3), we can consider the localization as regularization to the model update as a function of the distance between cells of the model and data.
Here, after testing different functions, we decided to employ a simple exponential correlation defined as:

$$\mathbf{C}_{i,:} = exp(-(\overline{\mathbf{d}}/r)^k) \tag{4}$$

where $\mathbf{C}_{i,:}$ indicates the i-th column of the matrix, $\overline{\mathbf{d}}$ is the Euclidean distance between each cell of the model and the i-th data point, r is the range which expresses the distance where spurious correlations are smoothed out and k is the order. Since in ERT each quadrupole (of known coordinates) measures an apparent resistivity datum, we assign a correlation factor between the model cells and data points depending on the distance (4). Even though the theory behind the localization is straightforward, the effects on inversion results seem valuable as we show in the next paragraphs.

Results

Synthetic test
Since we aim to apply the method to field ERT survey, we first create a synthetic model that includes the expected macro resistivity domains of the Athabasca basin. In particular, we build a model containing three main conductive bodies that simulate graphitic zones within two horizontal high resistivity domains. The shallow one represents the Athabasca sandstones, whereas the second is the underlying basement (label A and B in Figure 1). They are characterized by high resistivity values, about $3 \cdot 10^3 \Omega \cdot m$ and $5 \cdot 10^3 \Omega \cdot m$, respectively. We simulate an ERT survey with pole-dipole array, which is suitable for sub-vertical resistivity contrasts. The voltage dipole length is 100 m and it is moved along the profile increasing the distance between pole and dipole. We use pyGIMLi (Rücker et al., 2017) as forward operator, which solves the Poisson's differential equation through finite-element approach.
In order to make the test realistic, we contaminate the data with Gaussian noise having standard deviation of 2% of the apparent resistivity values. Since we search for conductive bodies, we start with a prior multivariate distribution having mean value of $10^3 \Omega \cdot m$. After preliminary tests, we set the ensemble dimension equal to 2000, the order k equal to 2, and the r parameter in (4) equal to 1500 meters. The choice has been made interpreting the Depth of Investigation (DOI) index.

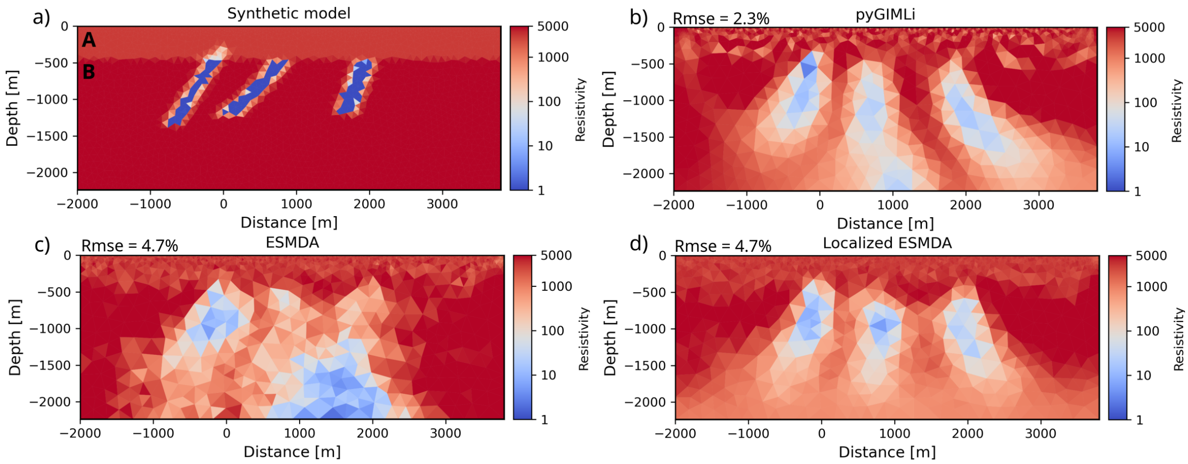

Figure 1 *a) Synthetic model composed by three conductive bodies within a layered host. b) Inversion result with pyGIMLi. c) Mean model we obtain with standard ESMDA. d) Mean model we obtain with Localized ESMDA algorithm.*

The inversion results show that pyGIMLi recovers the conductive bodies top in terms of position, but as expected due to the low resolution of ERT, not in terms of dip. Moreover, the first 500 meters of the model are contaminated by artifacts (Figure 1a). The standard ESMDA without localization returns a low root mean square error (rmse equal to 4.7%), whereas the three bodies are not correctly reconstructed. The localized ESMDA (Equation 3) is able to better reconstruct the conductive bodies where we expect high sensitivity, until the depth of 1500 m. The conductive bodies seem to be more focused even though, as pyGIMLi, the low resistivity anomalies do not reveal the true dip. The mean model is characterized by fewer artifacts in the first 500 m maintaining low rmse. These encouraging results made us apply the algorithm to a field dataset.

Inversion of field dataset

The data we invert in this work have been acquired in the Athabasca basin, an important Uranium play, characterized by extended graphitic fault systems. The array configuration is the same we have used for the synthetic test, pole-dipole with dipole dimension of 100 meters and distance between pole and dipole that increases from 100 to 1650 meters.

We set the same parameters as for the synthetic test, both in terms of prior distribution (mean resistivity equal to 1000 $\Omega \cdot m$), ensemble dimension of 2000 models and localization range (r in 4) equal to 1500 m. The computational time is about 4 hours with a laptop equipped with Intel Core i7-1165G7 @ 2.80GHz. Without localization, the ESMDA inversion returns a mean model able to predict the observed data with a low *rms* error of 3.7% but heavily scattered (Figure 2a). This is probably due to the spurious correlations that contaminate the covariance matrices (1) and then the Kalman filter. The use of a covariance-like localization function is able to smooth-out the updates where we expect low sensitivity, focusing more on the first 1500 meters of depth. The mean model shows that the first 500 meters are characterized by high resistivity values, higher than 1000 $\Omega \cdot m$ to more than 5000 $\Omega \cdot m$ without an evident stratification (Figure 2b). Starting from the depth of 500 m, we observe sub-vertical low resistivity anomalies that seem to be connected to a large anomaly until 1500 meters of depth. We interpret them as graphitic conductors that we know from geological information. To further validate the mean model we employ the pyGIMLi software inversion (Figure 2c). The inverted model shows anomalies that are totally compatible with the mean model in Figure 2b but with a noiser near-surface model. The benefit of Bayesian inversion is to use the posterior *pdf* to better interpret the result evaluating the cells in terms of probability distribution. Figure 2d represents four marginal posterior *pdf*s associated with four cells indicated by colored dots in Figure 2b. As each marginal distributions represent the resistivity values of each cell for all the ensemble members, it can be interpreted in terms of uncertainty (standard deviation). Combined to other geophysical or geological information, the standard deviation may help to interpret relevant resistivity anomalies.

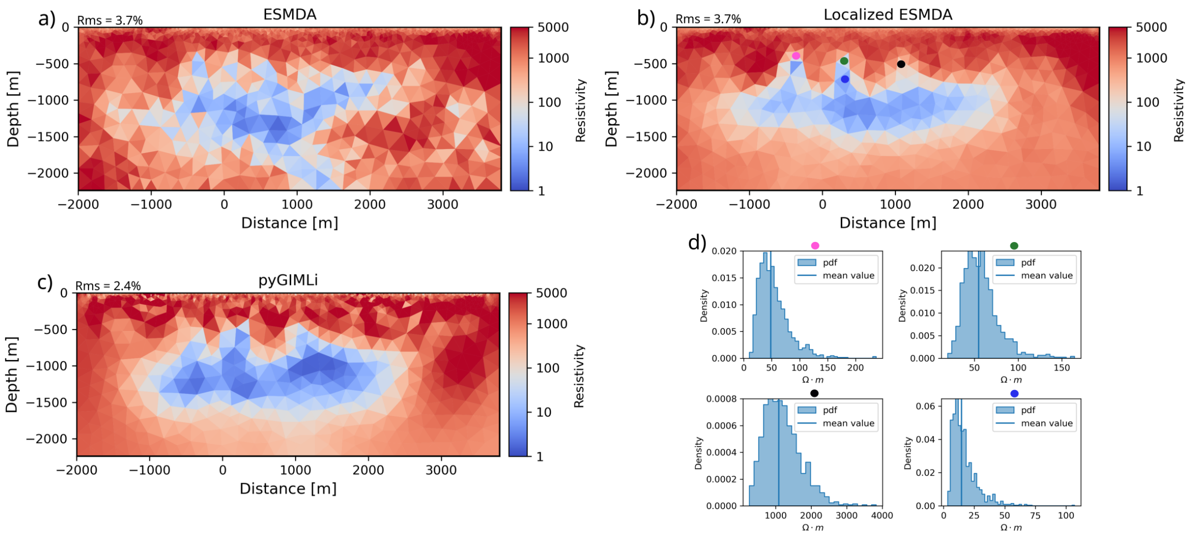

Figure 2 *a) Mean model obtained with adaptive ESMDA inversion without localization. b) Mean model obtained with Localized adaptive ESMDA. c) Model obtained with pyGIMLi. d) Marginal posterior probability density functions associated with four cells of the mean model.*

Conclusions

This research focuses on apply the localization to enhance the result the adaptive ESMDA in context of mineral exploration where we expect large contrasts of resistivity. The synthetic test suggests that the distance-based localization is able not only to regularize the model updates beyond the critical distance but also to focus on relevant portions of the model. The application of the algorithm to field data acquired in the Athabasca basin confirms the benefit of the localization, which helps to reconstruct the large contrast between conductive bodies and shallow high resistivity half-space. The results suggest that localization of the Kalman filter guarantees smoother model in depth without losing resolution in the relevant portion of the model. In addition, the marginal posterior distribution of each cell can help the interpretation of the resistivity anomalies, especially in mining context where decision about drilling has to be taken.

Acknowledgments

We thank Elodie Williard from Orano Mining and Grant Harrison from Orano Canada for their valuable suggestions. We also thank Cameco for providing and allowing us to present the field data we have used use in this work.

References

Chen, Y. and Oliver, D.S. [2017] Localization and regularization for iterative ensemble smoothers. *Computational Geosciences*, **21**(1), 13–30.

Emerick, A.A. [2016] Analysis of the performance of ensemble-based assimilation of production and seismic data. *Journal of Petroleum Science and Engineering*, **139**, 219–239.

Emerick, A.A. and Reynolds, A.C. [2013] Ensemble smoother with multiple data assimilation. *Computers & Geosciences*, **55**, 3–15.

Reynolds, A., Zafari, M. and Li, G. [2006] Iterative Forms of the Ensemble Kalman Filter. European Association of Geoscientists and Engineers.

Rücker, C., Günther, T. and Wagner, F.M. [2017] pyGIMLi: An open-source library for modelling and inversion in geophysics. *Computers and Geosciences*, **109**, 106–123.

Vinciguerra, A., Aleardi, M., Madsen, L.M., Bording, T.S., Christiansen, A.V. and Stucchi, E. [2024] Stochastic inversion of time-lapse electrical resistivity tomography data by means of an adaptive ensemble-based approach. *Geophysical Prospecting*, **72**(1), 268–284.