

Th B 01

## Data Driven Initial Model Estimation with Application to Marine CSEM Data

R. Mittet\* (EMGS)

### Summary

---

Initial model estimation can be fully data driven. Normally only one inversion run is required to achieve a good estimate for a smooth initial model with this approach. The scheme saves the operator in charge of the processing from repeated tests to find a proper initial model. It is very common in a production setting that several initial models are tested by the operator for each survey dataset. Thus there is a search not only for the final model but also for the initial model. These initial-model tests have a fairly high cost since they are full-scale inversions. A data driven start model generation approach has the potential to improve marine controlled-source electromagnetic imaging, especially in the deeper sections. A real data test from Brazil confirms the potential, the approach, giving a more focused inversion that has a better match with a priori information regarding reservoir placement, both laterally and in depth.

## Introduction

Marine controlled-source electromagnetic data (MCSEM) data are as a rule “imaged” by full waveform inversion (FWI). The end result of the inversion is a resistivity cube. Seismic imaging techniques like depth migration and reverse time migration are not effective since MCSEM data are dominated by re-fractions and wave-guide contributions (Mittet, 2015). Sometimes FWI is portrayed as a fully automatic procedure. It is not so in a production setting. The large scale 3D nature of this inverse problem precludes global optimization schemes and some sort of local optimization scheme is forced. The local optimization schemes can roughly be divided into gradient based schemes and Gauss-Newton (GN) schemes where also a Hessian approximation is used predict the update. Both types of scheme are usually designed to accept only updates that reduce the error functional. In this way local schemes become dependent on the initial model. For MCSEM our experience is that GN schemes are more robust than gradient based schemes with respect to the choice of initial model, but GN schemes are far from independent of the initial model. A consequence is that it is very common in a production setting that several initial models are tested by the operator for each survey dataset. Thus there is a search not only for the final model but also for the initial model. Unless proper tools are available it may be that the search for the initial model requires more human and computer resources than the search for the final model.

These initial-model tests have a fairly high cost since they are full scale inversions. They tend to use cluster resources on a large scale. Our general experience is that a half space as initial model is too far from the true model to give optimal final models. On the other hand, it may be dangerous to put too much structure into the start model. If that structure information is false, then the inverse scheme may struggle to remove it to a sufficient degree. A smooth start model is advantageous for MCSEM inversion as is often the case also for seismic inversion. For production work we see that a start model with a gradient in resistivity from the seabed and down performs much better than a half-space model, but it may be possible to perform even better if the search for the initial model is more directly data driven.

The scheme discussed below is a data-driven parametric inversion for a smooth initial model. The parametrized function space ensures the smoothness. The scheme uses a subset of all available receivers in a survey. Typical inline and crossline receiver separations are 5 km, however all towlines are used.

## Theory

Our objective is to construct a smooth resistivity model with a limited number of optimization parameters. The seawater part of the model, including bathymetry variations is kept fixed during all iterations. The parametrized model will have 3D variations. The lateral variations will be described by bivariate polynomials up to second order in the two horizontal coordinates  $x$  and  $y$ . These polynomials will be multiplied by functions that either cover the total model volume or by spatially bandlimited depth dependent functions. The key function used to describe the smooth depth dependence is the Fermi distribution,  $\mathcal{F}$ . The Fermi distribution can be expressed as,

$$\mathcal{F}(z, z_{\mu}^{(1)}, z_w^{(1)}) = \left( 1 + e^{\frac{(z - z_{\mu}^{(1)})}{\beta z_w^{(1)}}} \right)^{-1}, \quad (1)$$

where  $z$  is the depth coordinate. The parameter  $z_{\mu}^{(1)}$  is usually called the Fermi energy or the chemical potential, here it will be called the transition depth. The general behavior of the Fermi distribution is that  $\mathcal{F} = 1$  for  $z < z_{\mu}^{(1)}$  and  $\mathcal{F} = 0$  for  $z > z_{\mu}^{(1)}$ . A value of  $\beta = 0.144$  lets  $z_w^{(1)}$  roughly represent the width of the transition zone. The Fermi distribution is used at the top of the formation and immediately below the seabed. Layers in the middle of the formation are represented by the bandlimited boxcar distribution. This function can be expressed as the sum of two Fermi distributions,

$$\mathcal{B}(z, z_{\mu}^{(i)}, z_w^{(i)}, z_{\mu}^{(i+1)}, z_w^{(i+1)}) = \mathcal{F}(z, z_{\mu}^{(i+1)}, z_w^{(i+1)}) - \mathcal{F}(z, z_{\mu}^{(i)}, z_w^{(i)}), \quad (2)$$

where  $z_{\mu}^{(i)}$  and  $z_w^{(i)}$  are the transition depth and transition width at the top of the layer and  $z_{\mu}^{(i+1)}$  and  $z_w^{(i+1)}$  are the transition depth and transition width at the bottom of the layer. A bandlimited Heaviside



for the predicted fields at iteration  $n$  we write  $F_i^{Prd}(\mathbf{x}_r, \mathbf{x}_s, \omega | \mathbf{p})$ . The uncertainty in the observed data is  $\delta F_i(\mathbf{x}_r, \mathbf{x}_s, \omega)$ . The sum over  $[\mathbf{x}_r, \mathbf{x}_s, \omega], i, F]$  is then the sum over all observations. A smooth model that fit the observed data can be estimated by minimizing the error functional  $\varepsilon(\mathbf{p})$ ,

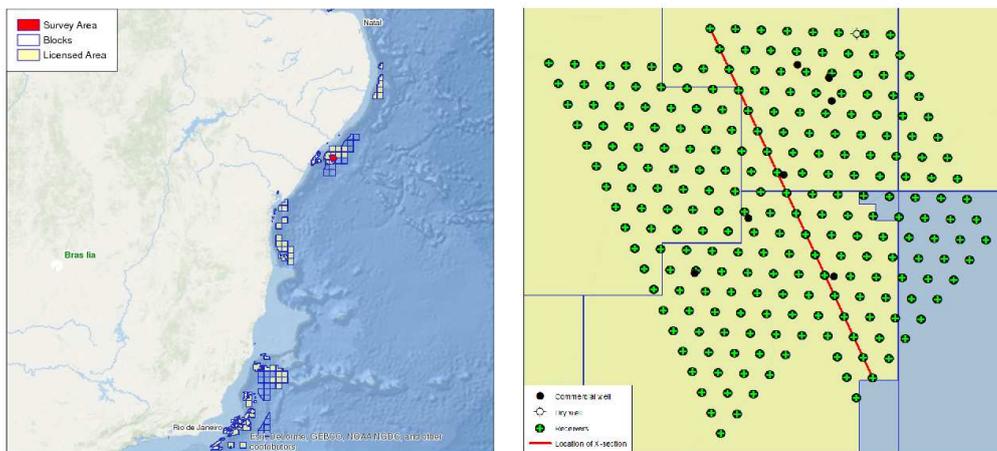
$$\begin{aligned}
 \varepsilon(\mathbf{p}) &= \sum_{\mathbf{x}_r, \mathbf{x}_s, \omega, i, F} \eta^2(\mathbf{x}_r, \mathbf{x}_s, \omega, i, F), \\
 \eta(\mathbf{x}_r, \mathbf{x}_s, \omega, i, F) &= \frac{\sqrt{\Delta F_i^*(\mathbf{x}_r, \mathbf{x}_s, \omega | \mathbf{p}) \Delta F_i(\mathbf{x}_r, \mathbf{x}_s, \omega | \mathbf{p})}}{\delta F_i(\mathbf{x}_r, \mathbf{x}_s, \omega)}, \\
 \Delta F_i(\mathbf{x}_r, \mathbf{x}_s, \omega | \mathbf{p}) &= F_i^{Obs}(\mathbf{x}_r, \mathbf{x}_s, \omega) - F_i^{Prd}(\mathbf{x}_r, \mathbf{x}_s, \omega | \mathbf{p})
 \end{aligned} \tag{8}$$

The above optimization problem is solved with the Levenberg-Marquardt algorithm. The Jacobian elements are estimated by a finite-difference technique. For the examples to follow,  $N_L = 5$  and first order polynomials were used such that 56 parameters must be estimated for a vertical plus a horizontal resistivity model.

## Results

The approach has been tested on a real multi-client data set acquired in 2012 offshore Brazil in the Sergipe-Alagoas Basin, as indicated in Figure 1. The survey consists of 232 receivers in a staggered 2x2 km grid and 15 towlines with 2 km spacing, covering an area of 800 km<sup>2</sup>. This area includes the major Muriu and Moita Bonita discoveries - there is thus ample well control in the area, making it well suited for testing the approach in an actual exploration setting. Both Muriu and Moita Bonita are turbidite sandstone discoveries, situated in the Early Cretaceous section approximately 2500-3000 m below the mudline.

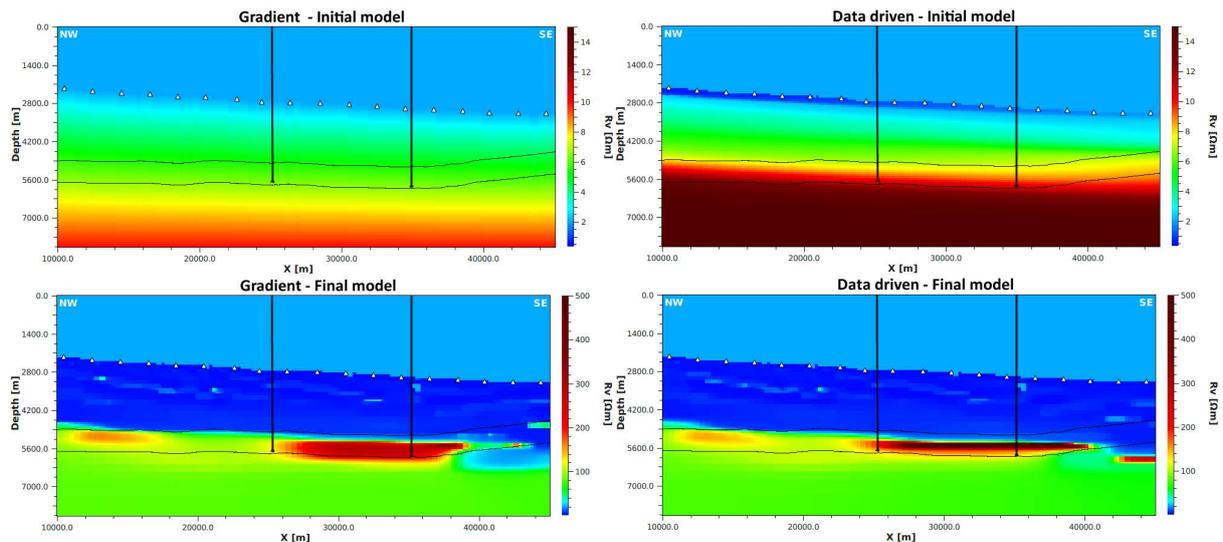
The cross-sections in Figure 2, marked with the red line in Figure 1, covers two wells (3-BRSA-1244-SES to the north and 3-BRSA-1296-SES to the south) which encountered light oil and reservoirs with good permeability and porosity conditions. Two initial models were tested in this exercise. One where



**Figure 1** Survey overview to the left and survey layout to the right.

we used a gradient model derived from 2.5D Gauss-Newton inversions. Several towlines were used and the results were interpolated. Only inline data contributed with this procedure. The second initial model was the result of the parametrized inversion scheme described in the previous section. The input to the data driven start model estimation was a decimated version of the complete data set, using 32 receivers spaced 6 km apart, but including all 15 towlines, such that azimuth data contributed to the optimization. Two Gauss-Newton inversion runs were then performed, including all the receivers and all the towlines. Inversion and regularization parameters were identical - the only difference between the two runs was the start model.

As can be seen in Figure 2, the two start models are not radically different, though the data driven model includes a shallow, low resistive layer not captured by the gradient model as well as being more resistive in the deeper parts around and below the discovery interval. The deepest horizon shown in these figures is mapped from legacy 2D seismic data acquired by Veritas and represents a feature near the Base Campanian, while the upper horizon is the same surface shifted up 800 m in order to represent the prospective interval. It is important to note that these horizons have not been used in the inversion process, and only serve as a visual guide when interpreting the results. The initial RMS data misfit from the gradient model was 6.04 compared to 4.63 for the data driven model. Both inversion runs converged



**Figure 2** Initial and final models. Left is for gradient start model and right is for data driven start model estimation. Initial model at the top. Final models at the bottom.

towards the same RMS data misfit of 1.0, and using a similar amount of iterations. In the shallow section above the prospective level, both inversions seem to recover similar resistive features. However, when inspecting the prospective level in light of what we know from the seismic data and the wells in the area, we observe that the final model from the data driven approach is more focused and has been able to place the resistive anomalies more appropriately in depth. We also see that the lateral distribution differs slightly, especially at the northern 3-BRSA-1244-SES well, where the final model starting from the data driven initial model estimate has a more distinct anomaly than the final model starting from the gradient model.

We also observe discrepancies in the south eastern part of the cross-section, where the gradient approach has recovered a weak, shallow resistor (2000 m BML) above a deeper conductive section, while the data driven approach has recovered a considerably stronger resistive anomaly and placed it deeper, within the conductive section. However, due to the lack of data points at the edge of the survey, this discrepancy is well within the expected uncertainty at the edge of the survey and would need further testing before doing interpretation at that location.

## Conclusions

We see that a data driven start model generation approach has the potential to improve MCSEM imaging, especially in the deeper sections. A real data test from Brazil confirms the potential, the approach, giving a more focused inversion that has a better match with a priori information regarding reservoir placement, both laterally and in depth.

## References

- [1] Mittet, R., 2015, Seismic wave propagation concepts applied to the interpretation of marine controlled-source electromagnetics: *Geophysics*, **80**,E63–E81.