Efficient computation of approximate low-rank Hessian for 3D CSEM inversion

Manuel Amaya*, Department of Mathematical Sciences, NTNU, Jan Petter Morten and Linus Boman, EMGS ASA

SUMMARY

Use of controlled-source electromagnetics for increasingly challenging exploration applications has led to the requirement for more powerful 3D inversion approaches. For 3D cases, application of Gauss-Newton algorithms is limited by the computational cost required to compute the Hessian matrix and solve for the model update. We consider a low-rank approximation to the Hessian matrix, which has the potential to reduce the numerical complexity drastically. The scheme is based on computing Green functions for phase-encoded groups of sources instead of incorporating sources individually. We describe the implementation and demonstrate the feasibility of the approach by numerical examples. We also give a theoretical analysis of the errors introduced by the approximation and how to mitigate them.

INTRODUCTION

The application of 3D controlled-source electromagnetics (CSEM) to image and characterize targets in increasingly challenging environments has motivated the development of more powerful inversion methods. The target response often represents a small perturbation of the measured signal response from complex background resistivity variations. When the geological understanding is limited, we require the 3D CSEM inversion to reconstruct from the data not only the potential hydrocarbon reservoir target, but also an accurate representation of the background resistivity variation and the structural framework.

The Gauss–Newton optimization algorithm is known to work well for inversion of CSEM data when assumptions of lower spatial dimensionality can be applied (Abubakar et al., 2006; Mittet et al., 2007). When a 3D model description is required, and when the input from state-of-the-art 3D acquisition is to be used, the numerical complexity of the Gauss–Newton algorithm can be very large (Abubakar et al., 2009; Sasaki, 2011). The large size of the Jacobian and Hessian matrices, as well as the number of 3D forward simulations can be a severe limitation. This has been addressed by several authors considering schemes to reduce the numerical cost by e.g. model reparameterization (Lin et al., 2013), and input data decimation (Schwarzbach and Haber, 2011). Gradient-based approaches to 3D CSEM inversion, like conjugate-gradient and quasi-Newton are less computationally demanding, and are now commonly used (Mackie et al., 2007; Støren et al., 2008). However, these approaches are most accurate when a good background model has been built. The construction of the background model can be a demanding task if the geology is complex and if little other geophysical data is available.

In this paper we present a Hessian approximation based on the superposition of phase-encoded sources. This approach leads

to a low-rank representation of the Hessian matrix, and alleviates the computational cost of constructing and storing this matrix as well as the solution of the Gauss–Newton equation. We show by numerical examples how the approximation is able to capture important features of the Hessian, at a numerical cost that is up to two orders of magnitude smaller than the exact calculation. The low-rank approximation was introduced by Amaya et al. (2014) and is here expanded to include a more detailed consideration of the grouping of sources and the effect the grouping has on accuracy and on forward modeling.

GAUSS–NEWTON OPTIMIZATION AND HESSIAN APPROXIMATION

The inversion of CSEM data is formulated as an optimization problem

$$\sigma = \arg \min_{\sigma \in \mathcal{M}} \varepsilon(\sigma), \tag{1}$$

where σ is a 3D conductivity model in the set $\mathcal{M}$ of models compatible with a priori information, and ε is the cost function. This cost function includes both regularization terms and data misfit terms that are dependent on the observed data,

$$\varepsilon_{\text{Data}}(\sigma) = \sum_{F,i,f,\mathbf{r}_{\text{rx}},\mathbf{r}_{\text{tx}}} \left| W_i^F(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}},f) \Delta F_i(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}},f;\sigma) \right|^2. \tag{2}$$

Here $\Delta F = F^{\text{Obs}} - F^{\text{Synth}}$ represents the difference between observed and synthetic fields ($F = E$ for electric and $F = H$ for magnetic), W is a datum weight (typically inverse measurement uncertainty), i are the spatial components (x,y) of the field recordings, f are the frequencies, $\mathbf{r}_{\text{rx}}$ is a receiver position, and $\mathbf{r}_{\text{tx}}$ is a source position. The shorthand notation $\kappa = (F,i,f,\mathbf{r}_{\text{rx}},\mathbf{r}_{\text{tx}})$ will uniquely label a measurement.

The non-linear optimization problem is solved by iteratively updating the conductivity model. The Gauss–Newton equation for model updates $\Delta\sigma$ is $\mathbf{H}\Delta\sigma = -\mathbf{g}$ where

$$\mathbf{H} = \mathbf{J}^\dagger \mathbf{J} + \text{c.c.} \tag{3}$$

is the Hessian matrix constructed from the Jacobian matrix $\mathbf{J}$, and

$$\mathbf{g} = \sum_\kappa W_\kappa \Delta F_\kappa^* (\mathbf{J})_\kappa + \text{c.c.} \tag{4}$$

is the model parameter gradient of the cost function. The c.c. denotes complex conjugated term. The Jacobian is a complex $N \times M$ matrix where N is the number of data samples, and M is the number of model parameters. The Jacobian can be constructed from Green functions,

$$(\mathbf{J})_{\kappa,\mathbf{r}} = W_i^F(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}},f) \sum_m G_{i,m}^{F,J}(\mathbf{r}_{\text{rx}}|\mathbf{r},f)$$

$$\times \sum_n G_{m,n}^{E,J}(\mathbf{r}|\mathbf{r}_{\text{tx}},f) J_n(\mathbf{r}_{\text{tx}},f), \tag{5}$$

where $\mathbf{r}$ is the position in the model, $G_{m,n}^{F,J}$ denotes the Green function for field F, component m, from a unit electric current source in direction n, and J_n is a component of the source $(m, n = x, y, z)$. It is straightforward to generalize the expression in (5) to the anisotropic and discrete case. From this expression, we see that explicit construction of the Jacobian requires the Green function associated with every source position $\mathbf{r}_{\text{tx}}$ to be calculated.

The gradient in (4) can be computed efficiently and without explicitly constructing the Jacobian by the use of adjoint state modeling. In this case, the factors $W_\kappa \Delta F_\kappa^*$ are used as the source strength in simultaenous source simulations for each receiver component after the synthetic data has been calculated. An implementation of such approach is described in Støren et al. (2008).

The Hessian matrix is a real $M \times M$ matrix, with rank given by the number of rows in the Jacobian (Grayver et al., 2013), i.e.

$$\text{rank}(\mathbf{H}) = 2\, N_F\, N_f\, N_i\, N_{\text{rx}}\, N_{\text{tx}}. \tag{6}$$

Here, and throughout, the notation N_a denotes the number of unique elements of index a. For a state-of-the-art 3D CSEM survey and with a realistic model representation, the numerical complexity involved with the construction of $\mathbf{H}$ and the solution for the model update can be very large. The number of forward solutions required can be of order 10^4, and the dense linear system for $\Delta\sigma$ can be of size $10^6 \times 10^6$. In this paper, we consider a low-rank approximation where sources in (5) are combined after encoding with a random phase factor, i.e. we construct $\sum_{\kappa \in s} e^{i\phi_\kappa}(\mathbf{J})_\kappa$ for a group of sources s associated with a receiver component, and where ϕ_κ are uniformly distributed random numbers in the interval $[0, 2\pi)$. The number of source groups N_s and the grouping scheme will be discussed below. Following this approach, the factors of $\mathbf{J}$ associated with the sources in a source group and a specific receiver component are calculated from a single simultaneous-source (super-shot) forward solution of the Maxwell equations. We denote the output of such simulation,

$$\tilde{G}_{i,m,\mathbf{r}_{\text{rx}},s}^{F}(\mathbf{r}, f) =$$
$$\sum_{n, \mathbf{r}_{\text{tx}} \in s} W_i^F(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}}, f)\, J_n(\mathbf{r}_{\text{tx}}, f)\, e^{i\phi_{F,i,\mathbf{r}_{\text{tx}},f}}\, G_{m,n}^{E,J}(\mathbf{r}|\mathbf{r}_{\text{tx}}, f). \tag{7}$$

The approximate Hessian matrix $\tilde{\mathbf{H}}$ following from the Jacobian constructed in this approach becomes,

$$\tilde{\mathbf{H}}(\mathbf{r}, \mathbf{r}') = \sum_{F,i,f,\mathbf{r}_{\text{rx}},s} \left[\sum_m G_{i,m}^{F,J}(\mathbf{r}_{\text{rx}}|\mathbf{r}, f)\, \tilde{G}_{i,m,\mathbf{r}_{\text{rx}},s}^{F}(\mathbf{r}, f) \right]$$
$$\times \left[\sum_n G_{i,n}^{F,J}(\mathbf{r}_{\text{rx}}|\mathbf{r}', f)\, \tilde{G}_{i,n,\mathbf{r}_{\text{rx}},s}^{F}(\mathbf{r}', f) \right]^* + \text{c.c.} \tag{8}$$

The rank of the approximation is given by the number of terms in the outer sum,

$$\text{rank}(\tilde{\mathbf{H}}) = 2\, N_F\, N_f\, N_i\, N_{\text{rx}}\, N_s. \tag{9}$$

Note that both the rank and the storage requirement to construct $\tilde{\mathbf{H}}$ scale by N_s instead of N_{tx} as for $\mathbf{H}$. The reduction in numerical complexity from the approximation is described below, but it is realistic that the ratio N_{tx}/N_s can be of order $10 - 100$ meaning a dramatic decrease in complexity.

SOURCE GROUPING STRATEGIES

Several strategies can be used in order to form the group of sources, as introduced in (5). In this section we will explore three configurations, shown schematically in Figure 1 below.

(a) Four groups of sources gathering closest ones in a single group simulation

(b) Four groups of sources gathering randomly the different sources in a single group simulation

(c) Four groups of sources gathering distant sources in a single group simulation
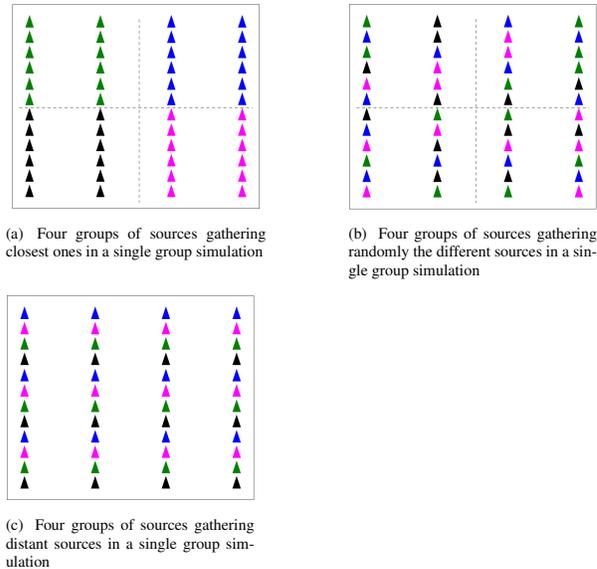
Figure 1: Examples of different strategies for source grouping. The triangles represent source positions, and the points with the same color are grouped.

In the example shown in Figure 1 the sources are arranged in four groups. In Figure 1(a) the sources are grouped such that each group covers a particular area. In Figure 1(b) the sources are grouped randomly, and in 1(c) the groups are selected by maximizing the distance between the sources. Numerical studies described below have shown that using groups based on a largest distance, as in Figure 1(c), gives the best result. This is in agreement with the qualitative argument in the next section which predicts that approximation errors decay with increasing separation between simultaneous sources in (7).

Once the groups of sources have been established, one forward modeling per group of sources and receiver component is performed, using $W_i^F(\mathbf{r}_{\text{rx}}|\mathbf{r}_{\text{tx}}, f)\, e^{i\phi_{F,i,\mathbf{r}_{\text{tx}},f}}$ as the strength for each source in a group. Linearity of the Maxwell equations implies that this is equivalent to the summation of individual terms in (7). The random phase factors $e^{i\phi_{F,i,\mathbf{r}_{\text{tx}},f}}$ are sampled independently for ϕ at each source position.

The number of forward simulations needed for building the Hessian matrix in a standard Gauss-Newton implementation, N_{Sim}, and with the low-rank approach, $\tilde{N}_{\text{Sim}}$, are

$$N_{\text{Sim}} = N_{\text{rx}} \cdot N_c \cdot N_F + N_d \cdot N_{\text{tx}} \tag{10a}$$
$$\tilde{N}_{\text{Sim}} = N_{\text{rx}} \cdot N_c \cdot N_F + N_d \cdot N_s \cdot N_{\text{rx}}. \tag{10b}$$

With realistic values for a modern 3D CSEM survey, N_{Sim} is dominated by the number of sources N_{tx}. The ratio $\tilde{N}_{\text{Sim}}/N_{\text{Sim}}$

then displays a decrease in the number of forward modeling jobs whenever $N_s \cdot N_{\rm rx}/N_{\rm tx} < 1$.

It is not only the reduction in number of forward computations which will alleviate the computational complexity. For the Jacobian, or equivalently the Green functions required to construct it, the volume of data in standard Gauss–Newton inversion scales linearly with $N_{\rm tx}$. Using the approximateion described here, this scaling is instead given by N_s. In the same way, the low-rank approximation allows a Hessian representation using considerably less memory when N_s is small by storing the quantities in square brackets in (8). In addition to the savings on number of forward solutions and memory requirements, we expect that solving the Gauss–Newton equation can be done very efficiently by exploiting the low-rank property of the Hessian matrix in (8).

QUALITATIVE ANALYSIS OF THE APPROXIMATION ACCURACY

Consider now the errors introduced by the approximation of the Hessian in (8). Due to the summation over source positions in $\tilde{G}$, the approximation will include terms involving two different source positions. Such terms are illustrated schematically in Figure 2 where (a) shows contributions from one source position, as in a standard Hessian, and (b) shows the additional terms introduced through the low-rank approximation.
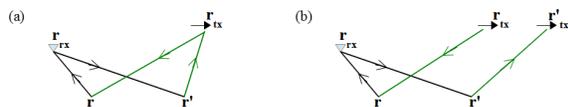
Figure 2: (a) Diagram representing terms included in the Hessian $\mathbf{H}$. Arrows correspond to Green functions, with reverse directions indicating complex conjugation. Source strength factors $W_i^F(\mathbf{r}_{\rm rx}|\mathbf{r}_{\rm tx},f)J_n(\mathbf{r}_{\rm tx},f)e^{i\phi_{F,i,\mathbf{r}_{\rm tx},f}}$ are associated with source positions. (b) Diagram representing the additional cross-talk terms introduced into $\tilde{\mathbf{H}}$, where two separate source positions give a contribution.

We refer to these errors as "cross-talk" and denote their contribution η, such that $\tilde{\mathbf{H}} = \mathbf{H} + \eta$. The source-diagonal terms, involving only a single $\mathbf{r}_{\rm tx}$, are the terms that make up the exact Gauss–Newton Hessian $\mathbf{H}$. The random phase-factors $e^{i\phi_\kappa}$ introduced in (7) will cancel in the source-diagonal terms since they enter as an absolute value. However, for the cross-talk terms, where two different source points are involved, the phase factors remain and act to suppress the cross-talk in the outer summation in (8). This is similar to applications of phase encoding in seismic modeling, see e.g. Bansal et al. (2013). The number of source-diagonal terms contributing to $\mathbf{H}$ is proportional to $N_{\rm tx}$. The number of cross-talk terms contributing to η will scale with the number of sources as $N_{\rm tx}^2$ (assuming $N_s = 1$). However, the magnitude $|\eta|$ should scale by the square root of the number of terms. This is due to the cancellations from random phases $e^{i\phi_\kappa}$ of the cross-talk terms and by analogy to a Gaussian random walk. We therefore expect lin-
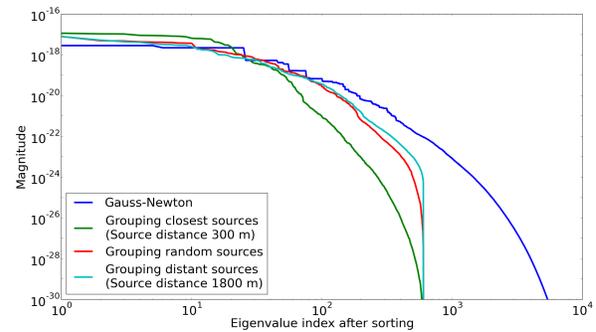
Figure 3: Eigenvalue distribution from various source grouping strategies, with $N_s = 12$ and rank $= 600$.

ear scaling $|\eta| \sim N_{\rm tx}$. Further, the number of significant terms contributing to η should be even less due to the exponential decay of the Green functions. This means that for cross-talk terms where $|\mathbf{r}_{\rm tx} - \mathbf{r}'_{\rm tx}|$ is large the magnitude of the contribution to $|\eta|$ is very small. Thus the magnitude $|\eta|$ scales by the number of sources in a source group as $(N_{\rm tx})^\alpha$ with $\alpha < 1$. In summary, the asymptotic behavior of the approximation is feasible since $\lim_{N_{\rm tx}\to\infty}|\eta|/|\mathbf{H}| = 0$.

We can reduce the error of the approximation by increasing the number of source groups N_s. In fact, using the maximum $N_s = N_{\rm tx}$ makes $\tilde{\mathbf{H}}$ identical to $\mathbf{H}$, but in this case there is no reduction in computational cost. We can optimize the approximation by constructing the source groups with maximum separation between the spatial locations of sources In this case, each cross-talk contribution shown in Figure 2 (b) will be smaller compared to a source-diagonal contribution in Figure 2 (a) by the decay of the Green functions over distance $|\mathbf{r}_{\rm tx} - \mathbf{r}'_{\rm tx}|$. This supports the numerical results where source groups based on the maximum distance between sources were found to give the highest accuracy, as is discussed in the results section. The distance between sources is thus a tuning parameter for the accuracy of the approximation that determines the number of source groups N_s and the strength of the cross-talk noise.

RESULTS

In this section we will show numerical results for the low-rank Hessian approximation. Figure 3 shows the eigenvalue distributions for the three methods of grouping sources illustrated in Figure 1, and also the distribution for the Hessian of standard Gauss–Newton. In this example the groups are chosen such that the rank of the approximate Hessians is 600. It is seen that source grouping based on maximum distance has an eigenvalue distribution that is closer to that of the standard Gauss–Newton Hessian, than the other two methods of grouping. This is consistent with the discussion above that suggests that the grouping based on maximum distance is the most accurate.

The following results show model updates computed for an

example CSEM survey, using both the Gauss–Newton update and the approximation scheme described above. In the example, the inversion parameterization is a regular grid with cell size 200 m × 200 m × 100 m and the total number of cells is 28275. The survey layout is detailed in Figure 4 with a total of 10500 data samples.

(a) $z = 1500$ m

(b) $y = 3000$ m
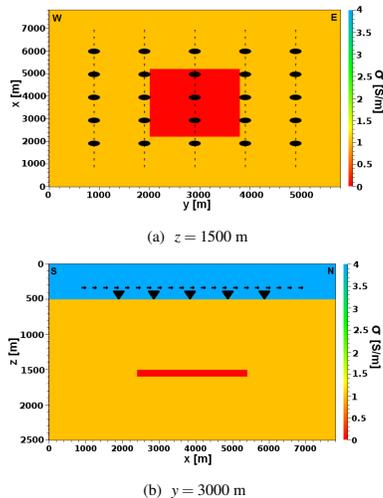
Figure 4: True model and survey layout for the example CSEM survey. The water conductivity is 4 S/m, and the water depth is 500 m. A resistor is located at 1.5 km depth, with dimensions 3 km × 2 km × 0.1 km, and conductivity 0.02 S/m. The formation conductivity is 1 S/m. There are 5 towlines and 25 receivers recording E_x and E_y at 0.25 and 1.0 Hz. The source distance is 300 m along towlines.

(a) $\tilde{\mathbf{H}}\Delta\sigma = -\mathbf{g}$

(b) $\mathbf{H}\Delta\sigma = -\mathbf{g}$

(c) $\Delta\sigma = -\mathbf{g}$
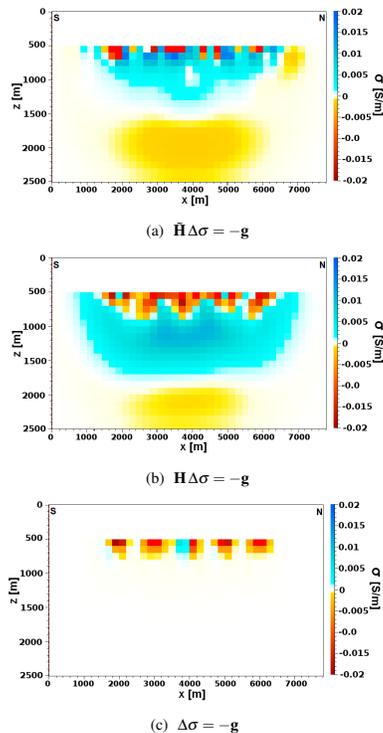
Figure 5: Model updates $\Delta\sigma$ at $y = 3000$ m.

The Gauss–Newton equation was solved using a conjugate gradient method with a small stabilizer and an initial model with the correct background conductivity. Figure 5 shows the solution for three cases using (a) the approximate Hessian matrix $\tilde{\mathbf{H}}$, (b) the exact Gauss–Newton Hessian matrix $\mathbf{H}$, and (c) steepest descent ($\mathbf{H} \rightarrow \mathbf{1}$). The approximate Hessian was obtained using one source group per receiver channel ($N_s = 1$). Comparing (a) with (b) and (c), we see that the solution to the Gauss–Newton equation with the approximate Hessian matrix is qualitatively more similar to a solution with the exact Hessian matrix than a steepest descent solution. In particular, much of the sensitivity information in $\mathbf{H}$ remains in $\tilde{\mathbf{H}}$ as seen at depth. Table 1 shows key characteristics for the computational cost of inversion for the example survey as well as for a larger, more realistically sized 3D CSEM survey. As is shown in the table, the number of forward simulations is reduced with a factor 2.6, however in a larger survey the reduction in computational cost can be much larger while keeping the error at the same level.

Case	Numerical cost	GN	Approx. $\tilde{\mathbf{H}}$	Ratio
a)	N_{Sim}	260	100	2.6
a)	$\mathbf{J}$ size	4.4 GB	43.1 MB	105
b)	N_{Sim}	11800	1000	11.8
b)	$\mathbf{J}$ size	127.4 TB	68.7 GB	1900

Table 1: Computational cost comparison for exact Gauss–Newton and the approximation scheme, for a) the simple example case shown in Figures 4 and 5, as well as b) large-scale realistic survey.

The data for the realistic survey in Table 1 were obtained from a survey area of 30 km×20 km×4 km, with 10 towlines at 2 km line separation and a source distance of 100 m along the towlines. In total we obtain 5700 source positions recorded at 100 receiver sites, measuring the horizontal components of electric and magnetic fields at 4 frequencies. The same discretization as in the smaller example is assumed, for two anisotropy components. The survey has a total of 14.59 million data samples. Using three source groups ($N_s = 3$) we keep the simultaneous source separation at 300 m such as in the smaller example survey discussed above. The size of the Hessian will be 2.6 TB, but the Jacobian representation in the approximation scheme offers a 40-fold reduction in size. The approximation error, η, could be reduced by increasing the number of source groups, but the computational cost would increase.

CONCLUSIONS

We have described a low-rank approximation to the Hessian for Gauss–Newton 3D inversion of CSEM data. The scheme is based on superposition of phase-encoded sources, and we have demonstrated the potential to significantly reduce both the number of forward simulations and memory requirements for inversion.

ACKNOWLEDGMENTS

http://dx.doi.org/10.1190/segam2014-0493.1

EDITED REFERENCES
Note: This reference list is a copy-edited version of the reference list submitted by the author. Reference lists for the 2014 SEG Technical Program Expanded Abstracts have been copy edited so that references provided with the online metadata for each paper will achieve a high degree of linking to cited sources that appear on the Web.

REFERENCES

Abubakar, A., T. Habashy, V. Druskin, D. Alumbaugh, A. Zerelli, and L. Knizhnerman, 2006, Two-and-half dimensional forward and inverse modeling for marine CSEM: 76th Annual International Meeting, SEG, Expanded Abstracts, **150**, 750.

Abubakar, A., T. M. Habashy, M. Li, and J. Liu, 2009, Inversion algorithms for large-scale geophysical electromagnetic measurements: Inverse Problems, **25**, no. 12, 123012, http://dx.doi.org/10.1088/0266-5611/25/12/123012.

Amaya, M., J. P. Morten, and L. Boman, 2014, A low-rank approximation to the Hessian for 3D CSEM Gauss-Newton inversion: Presented at the 76th Annual International Conference and Exhibition, EAGE.

Bansal, R., J. Krebs, P. Routh, S. Lee, J. Anderson, A. Baumstein, A. Mullur, S. Lazaratos, I. Chikichev, and D. McAdow, 2013, Simultaneous-source full-wavefield inversion: The Leading Edge, **32**, 1100–1108, http://dx.doi.org/10.1190/tle32091100.1.

Grayver, A. V., R. Streich, and O. Ritter, 2013, Three-dimensional parallel distributed inversion of CSEM data using a direct forward solver: Geophysical Journal International, **193**, no. 3, 1432–1446, http://dx.doi.org/10.1093/gji/ggt055.

Lin, Y., M. Li, A. Abubakar, and T. Habashy, 2013, A wavelet-based model compression method for three-dimensional electromagnetic data inversion: Presented at the 83rd Annual International Meeting, SEG.

Mackie, R., M. Watts, and W. Rodi, 2007, Joint 3D inversion of marine CSEM and MT data: 77th Annual International Meeting, SEG, Expanded Abstracts, **116**, 574.

Mittet, R., H. Maulana, K. Brauti, and T. A. Wicklund, 2007, CMP inversion of marine CSEM data: Presented at the EGM International Workshop.

Sasaki, Y., 2011, Gauss-newton-based 3D joint inversion of marine CSEM and MT data: 73rd Annual International Conference and Exhibition, EAGE, C027.

Schwarzbach, C., and E. Haber, 2011, 2011, Finite element based inversion for electromagnetic problems using stochastic optimization: 81st Annual International Meeting, SEG, Expanded Abstracts, **110**, 567.

Støren, T., J. Zach, and F. Maaø, 2008, Gradient calculations for 3D inversion of CSEM data using a fast finite-difference time-domain modelling code: 70th Annual International Conference and Exhibition, EAGE, Extended Abstracts, P194.