PROCEEDINGS, TOUGH Symposium 2012
Lawrence Berkeley National Laboratory, Berkeley, California, September 17-19, 2012

GEOPHYSICAL DATA IMPROVE STABILITY AND CONVERGENCE OF HYDROLOGICAL PROPERTY ESTIMATION: A SYNTHETIC CO_2 INJECTION STUDY

J. Doetsch, M.B. Kowalsky, S. Finsterle, C. Doughty, J.B. Ajo-Franklin, and T.M. Daley

Lawrence Berkeley National Laboratory
One Cyclotron Road
Berkeley, CA, 94720, USA
e-mail: jdoetsch@lbl.gov

ABSTRACT

Monitoring and modeling migration of injected CO_2 in the subsurface is critical for assessing the risk of leakage from geologic carbon sequestration sites, but it is also very challenging. Integrating complementary hydrological and geophysical monitoring data in a coupled hydrogeophysical inversion can help to address this challenge. We consider a synthetic CO_2 injection study to analyze the effect of adding cross-borehole electrical resistance and seismic data to inversions of pressure and gas-composition data. The geophysical data are found to significantly improve convergence and stability of the inversions. Parameterizing mean aquifer permeability and differences from this mean value are found to be superior to inverting for the permeability of each layer directly. These results will be the starting basis for analysis of the actual field data.

INTRODUCTION

Geologic sequestration of carbon dioxide (CO_2) is a promising approach for offsetting anthropogenic carbon emissions, and deep saline aquifers have been identified as potential target formations. Several pilot studies are currently under way to demonstrate the feasibility of long-term underground CO_2 storage, to assess risks associated with it, and to improve the overall understanding of CO_2 migration in the subsurface. A key component in these efforts is the integrated analysis of complementary hydrological and geophysical monitoring data. Whereas traditional hydrological measurements are useful for determining the properties immediately surrounding boreholes, geophysical cross-borehole measurements are sensitive to subsurface properties over larger regions, but can have lower spatial or temporal resolution and can be difficult to interpret quantitatively. We combine the advantages of hydrological and geophysical data sets in a fully coupled hydrogeophysical inversion, using iTOUGH2 (Finsterle, 2004).

The fully coupled hydrogeophysical inversion (Kowalsky et al., 2005) calibrates a hydrological flow and transport model simultaneously to the hydrological and geophysical data. Performing this model calibration and setting up an inverse problem involves many steps and decisions (Carrera et al., 2005), some of which are implicit and therefore easily overlooked. The main steps in developing a model can be described as follows (Finsterle and Kowalsky, 2011): (1) data selection, (2) development of a conceptual hydrological model, (3) parameter definition or parameterization, (4) choice of an objective function and optimization algorithm and (5) *a posteriori* assessment of the appropriateness of the choices in steps 1–4, and uncertainty analysis for estimated parameters and model predictions. Each of these points invites extended analysis.

Here, we concentrate on points (1) and (3), analyzing the effect and value of geophysical data for the inversion process and touching upon some aspects of model parameterization. We focus on an application involving the hydrogeophysical monitoring of CO_2 injection in a deep saline aquifer. We present a synthetic study and show some preliminary results after providing a brief overview of a field experiment and corresponding data that motivate the study.

DESCRIPTION OF THE EXPERIMENT

Field site and injection experiment

The synthetic study considered here is based on the general field layout, aquifer properties, and data availability from a large-scale injection experiment at the SECARB-Cranfield CO_2 injection pilot site (Hovorka et al., 2009). The experiment involved the injection of approximately 1 M tonne of CO_2 into a saline aquifer at a depth of ~3200 m. The injection interval is in a segment of the Lower Tuscaloosa Formation referred to as the Tuscaloosa D/E sand; this unit ranges from 15 to 25 m in thickness across the field and consists of relatively permeable fluvial sandstones and conglomerates. Injection began on December 1, 2009, at a rate of 3 kg/s, which was subsequently increased to 7 kg/s.

Borehole logs suggest that the Tuscaloosa unit is separated into two permeable layers of approximately 16 m (top) and 7.5 m (bottom) thickness, with the separating aquitard being laterally continuous. Based on measurements from sidewall cores, the permeability of the aquitard is about 50 times lower than that of the overlying and underlying aquifer layers.

As shown in Figure 1, the aquifer is accessed by three wells, one used as an injection well (F1) and two as monitoring wells (F2 and F3), allowing for the collection of hydrological and geophysical monitoring data. Data availability for our synthetic example is mostly based on that of the field experiment.

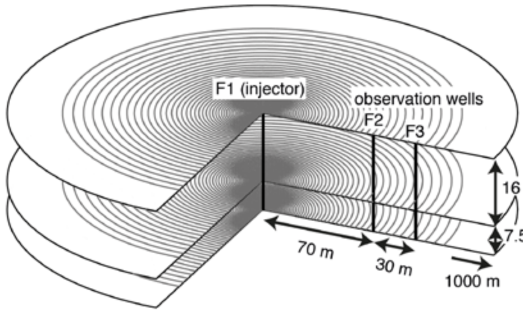

Figure 1. Schematic of the two-layer radial model used for the synthetic example. The model consists of two unconnected layers that are accessed by three wells.

Hydrological data

A variety of hydrological characterization and monitoring data were collected before and during the injection experiment (e.g., core and geophysical logging data, pressure fall-off tests). In this study, we focus mostly on (1) pressure in the injection well (F1), which was measured at a high temporal sampling rate, and (2) U-tube sampling data that provide information on the time-varying gas composition of the fluid in the monitoring wells, using on-site mass spectroscopy (Freifeld et al., 2011). The main gas-phase components that were present include CO_2, CH_4, and the injected tracers SF_6 and Kr.

Geophysical data

Cross-borehole seismic and electrical resistivity tomography (ERT) data were acquired between monitoring wells F2 and F3.

A continuous active-source seismic monitoring system (CASSM, Daley et al., 2011) was installed for the Cranfield experiment, but stopped operating before the start of the CO_2 injection. Therefore, seismic data were only acquired once before the start of the CO_2 injection and once after ~300 days of injection (Ajo-Franklin et al., 2012). For the purpose of this synthetic study, which is intended to examine our parameter estimation methodology, we assume that seismic data were acquired daily as intended.

The time-lapse ERT monitoring began before the CO_2 injection started and continued for ~300 days (Yang et al., 2012). Because directly including thousands of electrical resistance measurements in a coupled hydrological-geophysical inverse modeling procedure (see Kowalsky et al., 2005) is computationally very expensive, we begin by using a reduced form of the data that represents an average response of the aquifer to the CO_2 injection. We calculate an average value of electrical resistivity in the reservoir as a function of time, which can be obtained through traditional tomographic inversion of the ERT data. The output of this pre-processing step is then used as input in the inversion.

Specifically, the reduced form of the ERT data that we incorporate in the inversion procedure is the average change in the reservoir electrical conductivity (EC) relative to the pre-injection value. We assume that the CO_2 and CH_4 gas phase and the host rock are infinitely resistive, and no gas phase exists at time zero. Under these assumptions, Archie's law (Archie, 1942) can be expressed as

$$EC(t) \big/ EC(0) = S_w^n ,$$

with the brine saturation S_w and Archie's saturation exponent n. Formulating the measurement as the change in electrical conductivity makes it unnecessary to know the electrical conductivity of the brine, the porosity, and the cementation exponent, otherwise required for applying Archie's law. Furthermore, the ERT processing and inversion techniques handling the changes in EC rather than the absolute value of EC are more reliable (e.g., Daily et al., 1992).

APPROACH

Coupled hydrological-geophysical model

The hydrological model used to simulate the CO_2 injection and the corresponding synthetic data consists of a two-layer radial model (Figure 1), with the layers unconnected from each other to represent the high permeability layers of the saline aquifer, as described above. The injection well and monitoring wells are connected to both layers to allow for injection to occur into both layers and for hydrological measurements, which represent an average response over both layers, to be simulated. Each layer consists of 99 cells in the radial direction with a maximum radius of 1000 m. The outermost cells in each layer are given a constant-pressure boundary condition.

Hydrological properties are modeled as homogeneous within each layer. However, to reproduce the double arrivals of CO_2 and CH_4 that are observed in the field data, it was necessary to use different values of absolute permeability for each layer (1.5×10^{-13} m^2 for the top, and 2.2×10^{-13} m^2 for the bottom). The simulations were performed using TOUGH2 EOS7C (Pruess et al., 1999; Oldenburg et al., 2004), and the resulting synthetic data appear consistent with the field data (see Figure 2). Note that our intention was to obtain synthetic data for examining a coupled hydrological-geophysical inverse modeling approach. Rather than aiming to develop a more complex and computationally intensive model that reproduces the field data exactly, for this study we chose to use a simplified representation of the system that reproduces the general characteristics observed in the field data.

In the coupled hydrological-geophysical model, the ERT data, that is, the change in average EC as defined above, is simulated daily as a function of the time-varying properties simulated in TOUGH2 for a given set of hydrological input parameters. The change in EC depends on the average gas saturation in the aquifer layers (as well as on Archie's parameter n, which we assume to be known and equal to 2) and, therefore, becomes a function of the hydrological input parameters (e.g., the permeability).

Similarly, seismic travel times are calculated in the hydrological-geophysical model and are a function of the time-varying CO_2 saturation, among other things (Daley et al., 2011). We simulate seismic data for 11 sources in F2 and 11 receivers in F3, yielding a full data set of 121 travel times at each time step. The number of sources and receivers is varied to look at the optimal setup (see below). The conversion from CO_2 saturation to seismic velocities uses the patchy saturation model described by Daley et al. (2011).

Gaussian noise was added to the synthetic hydrological and geophysical data, based on the error level observed in the field data. Table 1 gives an overview of the available data and the assumed errors. The actual field data are compared with the simulated data (with and without noise) in Figure 2. Observe the double arrival in the CO_2, CH_4, and tracer mass fractions; this feature is explained only by the two layers of differing permeability, which is consistent with the choice of the conceptual model used in this study (i.e., the radial model that assumes two unconnected layers).

The simulation time for a single forward run on a 2.7 GHz single core CPU machine is ~2 minutes, which enables relatively fast inversions, even when testing many parameter sets.

Inverse modeling approach

We use the inverse modeling capabilities of iTOUGH2 (Finsterle, 2004) for evaluating the objective function for many parameter sets and studying the inversion performance for different combinations of the synthetic data set. Most analysis is based on inversions for the absolute permeability (or log permeability) of the two model layers.

Table 1. Overview of the available data and assumed measurement errors

Data type	*Time [days]*	*Interval [days]*	*Number of data*	*Error (std. dev.)*
F1 Press.	0 – 300	10	31	2.00E+04 Pa
F2 CO_2	11.5 – 30	0.25 / 0.5 / 1	46	0.04
F3 CO_2	14.5 – 30	0.25 / 0.5 / 1	18	0.04
F2 Tracer	11.5 – 30	0.25 / 0.5 / 1	46	2.00E-06
F3 Tracer	14.5 – 30	0.25 / 0.5 / 1	18	2.00E-06
ERT	0 – 300	1	300	0.01
Seismic	0 - 300	2 / 4 / 8 / 16	55	8.00E-06 ms

THE EFFECT OF PARAMETERIZATION CHOICES ON INVERSION STABILITY

The process of deciding which parameters to estimate and how to formulate the inverse problem is critical for successful inverse modeling. Inverting for too many parameters simultaneously (over-parameterization) can lead to a non-unique solution due to the strong correlation between model parameters and a lack of information independently related to each parameter, leading to large uncertainty in the parameter estimates. Over-parameterization can be caused either by attempting to estimate a property with too much geometrical detail (e.g., defining unknown permeability values in too many pixels or spatial regions) or by estimating too many different properties that are not related spatially (without sufficiently sensitive data to support their estimation). Over-parameterization due to high spatial resolution of parameters is regularly encountered in geophysical inversions and is partly overcome through regularization; smoothing and damping techniques stabilize the inversion and reduce ambiguity for large-parameter fields by preventing too much fine-scale detail (Constable et al., 1987). It is potentially more difficult to reduce non-uniqueness between parameters that are not geometrically related.

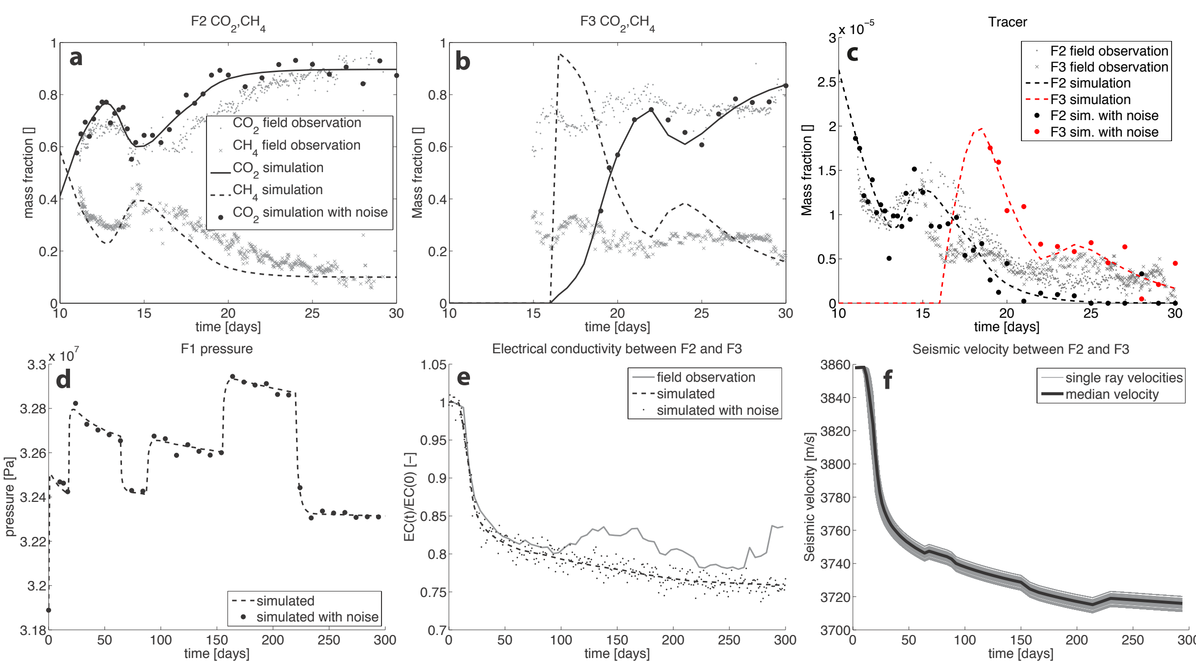

Figure 2. Overview of observed (gray), simulated (lines) and noisy simulated (black and red dots) hydrological and geophysical data. The noisy simulated data is used as input for the inversions.

For the 2-layer radial model example, properties are defined as homogeneous within each layer, but the permeability in each layer is unknown. An obvious choice for estimating the permeability values might be to invert for the permeability (or log permeability) of each layer directly. However, the parameters are highly correlated (>80%) in this example, meaning that changes in the permeability of either layer similarly affect the simulated data, making a unique determination of either parameter difficult. Figure 3a illustrates the objective function with respect to variation of the log permeability of each layer, highlighting the high correlation between the two parameters, the presence of a local minimum, and the poorly defined global minimum for this parameterization.

An alternative way to solve the same problem is to invert for the mean of the log permeability of the two layers, and also for the difference in log permeability between the two layers. This transformation corresponds to a rotation of 45° in parameter space, as is shown in Figure 3b. The transformation—although seemingly equivalent to the previous formulation—reduces

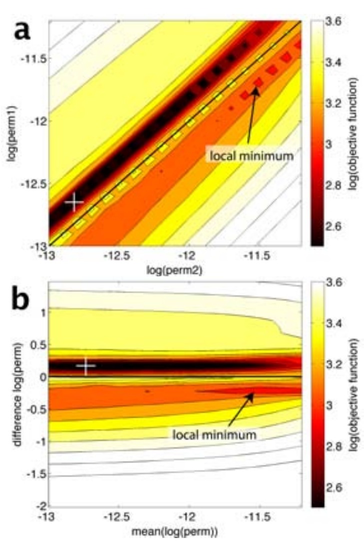

Figure 3. Contour plots of the objective as a function of (a) log permeability in the two layers, and (b) mean log permeability and difference in log permeability between the two layers. The minimum of the objective function is marked by a white '+'.

the parameter correlation and stabilizes the inversion. The importance of proper parameterization, as shown in this simple two-parameter problem, is amplified when attempting a multi-parameter inversion. Finding independent parameters should be viewed as a priority when defining the parameterization for an inverse problem. It is similarly important to determine what types of data should be collected to maximize parameter independence.

THE EFFECT OF DIFFERENT DATA TYPES ON INVERSION STABILITY

For successful inversion, the availability of data with sufficient informational content to estimate the parameters of interest is key. Parameters can only be resolved if the data are sensitive to them and, ideally, the correlations between parameters are low. Another factor influencing the performance of inverse modeling is the shape of the objective function. Clearly, the presence of local minima can make finding the global minimum difficult, especially when gradient-based optimization algorithms are used.

In general, large-scale measurements tend to create smoother objective functions, compared to point measurements. We examine how the inclusion of geophysical measurements in a hydrological-geophysical inversion affects the shape of the objective function and overall inversion performance.

Figure 4 shows the objective function for the different data sets as a function of the difference in log permeability between the two model layers. The objective functions for the CO_2 and tracer mass fractions contain local minima and will likely hinder an optimization from global convergence unless the initial guess is close to the global minimum. The objective functions for the geophysical data, on the other hand, gradually decrease towards a single minimum and thus may facilitate easier convergence.

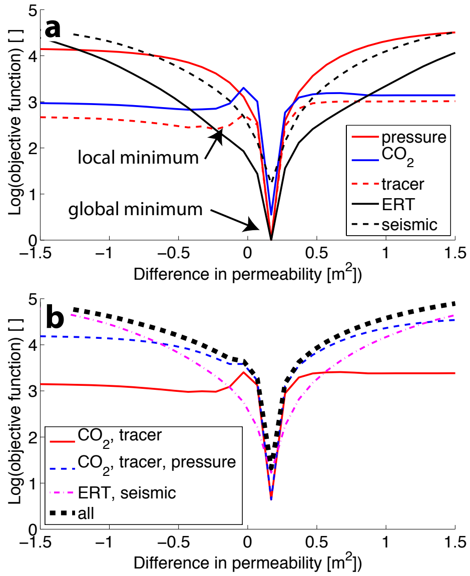

Figure 4. (a) Contributions of each data set to the objective function and (b) objective functions for subsets of the data and for the complete data set containing all hydrological and geophysical data.

When combining hydrological and geophysical data, the geophysical data stabilize the inversion, resulting in better convergence to the true solution. This effect is illustrated in Figure 5, which shows the convergence behavior for inversions performed using the same data subsets shown in Figure 4b with nine different sets of initial guesses for the parameters.

For the cases in which the geophysical data are included (Figure 5c and d), the convex shape of the objective function helps all inversions, each starting from a different initial guess, converge to the correct value. On the contrary, few of the inversions using hydrological data alone converge to the global minimum (Figure 5a and b). Note how the objective function for the combined data sets (Figure 5d) reflects the shape of the contributions from the pressure and geophysical data (Figure 5b and c, respectively), nicely illustrating the link between the choice of data included in a study and the resulting objective function character. It should also be noted that changing the weights and error assumptions of the data sets considered in this study could change the objective function shape and inversion performance.

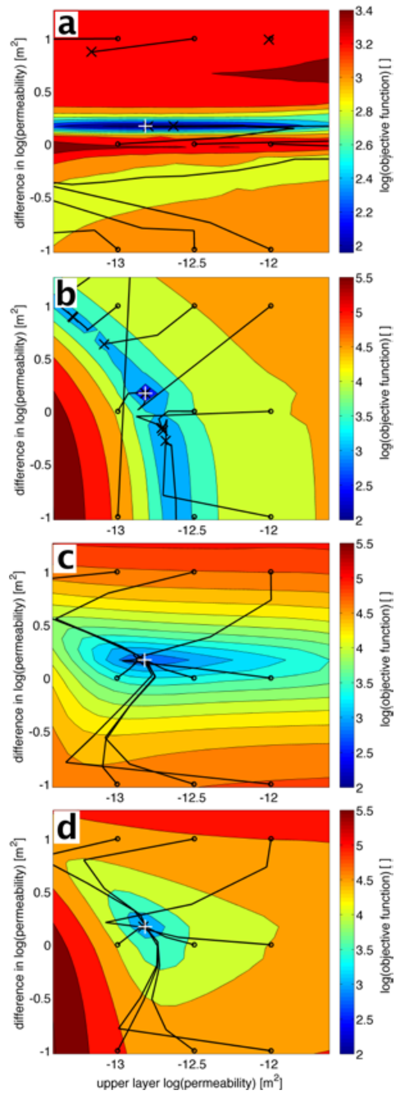

Figure 5. Contour plots of the objective function and example inversion paths (lines) for nine sets of initial values (black dots) using (a) CO_2 and tracer data, (b) CO_2, tracer and pressure data, (c) ERT and seismic data and (d) all available data. Each inversion results is marked with a black "x" and the true value with a white "+".

Due to the high cost and effort of obtaining geophysical data for such a field experiment, optimizing data collection is a worthwhile consideration. For example, the continuous seismic data collection using the CASSM system (Daley et al., 2011) is very promising, but the number of sources is restricted and should be minimized. Analysis of the objective function for the 121 source and receiver combinations shows that it is possible to constrain the global minimum with a single seismic source if the source position is carefully chosen. Figure 6 shows the contributions of each ray to the objective function (gray lines), as well as two zero-offset (horizontal) rays through each layer. If the source position is chosen close to the layer boundary, the difference between the two layers can be resolved equally well as with the full data set.

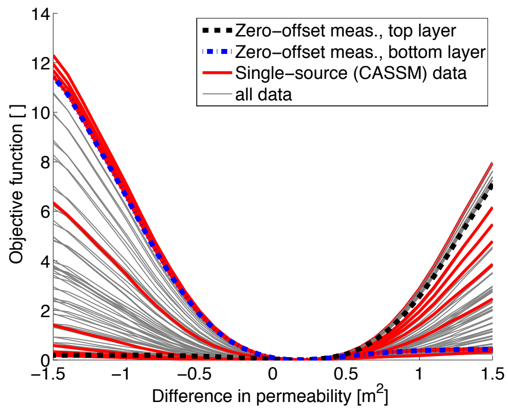

Figure 6. Objective function for different parts of the seismic data. Single-source data can define the global minimum if the source position is well chosen.

DISCUSSION

Parameterization and data selection are critical components for setting up a successful inversion, as they affect the shape of the objective function and the likelihood of obtaining accurate parameter estimates. In order to illustrate the objective function and show the inversion paths (Figures 4 and 5), the above analysis concentrates on two-parameter inversions. When inverting for more than two parameters, the objective function gets more complex and the inversion problem less unique. Choosing a good parameterization and including data with reasonable error assumptions and weighting thus becomes even more critical.

Even in the two-parameter inversion case, our analysis shows how important the geophysical data can be for a successful inversion: only when the ERT and seismic data are included can the inversion converge to the global minimum of the objective function, except when the initial values are very close to the true parameter values.

The results of the particular example discussed here may not be directly applicable to other problems, but certain characteristics are generally valid. For example, geophysical data usually have a larger support volume than measurements in wells, and their inclusion is expected to result in objective functions that are more favorable for inverse modeling.

The integration of geophysical and hydrological data in an inversion can be made difficult by uncertainty or spatial variation in the petrophysical relationships that are needed to translate geophysical parameters into hydrological properties and state variables. Uncertainties in the petrophysical models translate into uncertainties of the geophysical data, and nonlinearity of the petrophysical relationships adds further complications. However, in some cases uncertainty in petrophysical parameters can be accounted for by including their estimation in the inversion process (Kowalsky et al., 2005).

The number of petrophysical parameters and their uncertainty can also be reduced when inverting time-lapse data. For example, in our ERT formulation, dependence on porosity and the cementation exponent was removed, and only the relatively well-known saturation exponent needs to be determined.

CONCLUSIONS

In this study, we analyzed the effect of parameterization and geophysical data on the stability and convergence of coupled hydrogeophysical inversions. The analysis is based on a synthetic study that mimics a CO_2 injection experiment at Cranfield, MS.

We find that:

- Geophysical data (ERT and seismic) greatly stabilize the inversion and improve convergence.
- Parameterizing mean and difference of permeability in different layers rather than the permeability of each layer improves convergence.

- Time-lapse formulation (e.g., for ERT) can reduce the number of uncertain petrophysical parameters.

Although the analysis is specific to this case study, we believe these three points to be generally valid. The presented results will be the starting basis for analysis and inversion of the field data from the CO_2 pilot site at Cranfield, MS.

ACKNOWLEDGEMENTS

The authors would like to thank Susan Hovorka for her leadership and assistance (Texas Bureau of Economic Geology) and Xianjin Yang and Charles R. Carrigan (Lawrence Livermore National Laboratory) for their contributions. This work was supported by the National Risk Assessment Partnership (NRAP) of the US DOE under contract DE-AC02-05CH11231.

REFERENCES

Ajo-Franklin, J.B., J. Peterson, J. Doetsch, and T.M.Daley, High-resolution characterization of a CO_2 plume using crosswell seismic tomography: Cranfield, MS, *Int. J. Greenhouse Gas Control*, under review.

Archie, G. E., The electrical resistivity log as an aid in determining some reservoir characteristics, *Transactions of the American Institute of Mining, Metallurgical and Petroleum Engineers*, *146*, 54-62, 1942.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten, Inverse problem in hydrogeology, *Hydrogeology Journal*, *13*(1), 206-222, 2005.

Constable, S. C., R. L. Parker, and C. G. Constable, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, *52*(3), 289-300, 1987.

Daily, W., A. Ramirez, D. Labrecque, and J. Nitao, Electrical resistivity tomography of vadose water movement, *Water Resour. Res.*, *28*(5), 1429-1442, 1992.

Daley, T. M., J. B. Ajo-Franklin, and C. Doughty, Constraining the reservoir model of an injected CO_2 plume with crosswell CASSM at the Frio-II brine pilot, *Int. J. Greenh. Gas Control*, *5*(4), 1022-1030, 2011.

Finsterle, S., Multiphase Inverse Modeling: Review and iTOUGH2 applications, *Vadose Zone J.*, *3*(3), 747-762, 2004.

Finsterle, S., and M. B. Kowalsky, A truncated Levenberg–Marquardt algorithm for the calibration of highly parameterized nonlinear models, *Computers & Geosciences*, *37*(6), 731-738, 2011.

Freifeld, B., C. Doughty, P. Cook, K. Romanak, J. Lu, and C. Yang, Flow and Geochemical Sampling at SECARB's Phase-III Cranfield Site – Initial Results, *Ann. Conf. Carbon Capture, Utilization & Sequestration, Pittsburgh*, 2011.

Hovorka, S. D., T. A. Meckel, R. H. Treviño et al., *SECARB Phase III – Early Test, Cranfield, MS*, GCCC Digital Publication #09-04, Southern States Energy Board Stakeholders' Meeting, 2009.

Kowalsky, M. B., S. Finsterle, J. Peterson, S. Hubbard, Y. Rubin, E. Majer, A. Ward, and G. Gee, Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data, *Water Resour. Res.*, *41*(11), W11425, 2005.

Oldenburg, C. M., G. J. Moridis, N. Spycher, and K. Pruess, *EOS7C Version 1.0: TOUGH2 Module for Carbon Dioxide or Nitrogen in Natural Gas (Methane) Reservoirs*, Report LBNL-56589, Lawrence Berkeley National Laboratory, Berkeley, Calif., 2004.

Pruess, K., C. Oldenburg, and G. Moridis, *TOUGH2 User's Guide, Version 2.0*, Report LBNL-43134, Lawrence Berkeley National Laboratory, Berkeley, Calif., 1999.

Yang, X., C. Carrigan, and B. Labrecque, ERT monitoring at the Cranfield CO_2 injection site, *Geophysics*, under review.