LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-TR-420989

Development of a Stochastic Inversion Tool To Optimize Agreement Between The Observed And Predicted Seismic Response To CO2 Injection/Migration in the Weyburn-Midale Project

A. L. Ramirez, Y. Hao, D. White, S. Carle, K. Dyer, X. Yang, W. Mcnab, W. Foxall, J. Johnson

December 4, 2009

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Theme 3B: Storage Monitoring Methods
Task 3b.11: Forward-Tuned Stochastic Modeling

Development of a Stochastic Inversion Tool To Optimize Agreement Between The Observed And Predicted Seismic Response To CO_2 Injection/Migration in the Weyburn-Midale Project

Abelardo L. Ramirez[1], Yue Hao[1], Donald J. White[2], Steven F. Carle[1], Kathy Dyer[1], Xianjin Yang[1], Walter Mcnab[1], William Foxall[1], James Johnson[3]

1-Lawrence Livermore National Laboratory
2-Geological Survey of Canada
3-Schlumberger-Doll Research Center

LLNL-TR-420989

Abstract

During Phase 1 of the Weyburn Project (2000-2004), 4D reflection seismic data were used to map CO_2 migration within the Midale reservoir, while an extensive fluid sampling program documented the geochemical evolution triggered by CO_2-brine-oil-mineral interactions. The aim of this task (3b.11) is to exploit these existing seismic and geochemical data sets, augmented by CO_2/H_2O injection and HC/H_2O production data toward optimizing the reservoir model and thereby improving site characterization and dependent predictions of long-term CO_2 storage in the Weyburn-Midale reservoir.

Our initial project activities have concentrated on developing a stochastic inversion method that will identify reservoir models that optimize agreement between the observed and predicted seismic response. This report describes the technical approach we have followed, the data that supports it, and associated implementation activities. The report fulfills deliverable D1 in the project's statement of work. Future deliverables will describe the development of the stochastic inversion tool that uses geochemical data to optimize the reservoir model.

This work has been performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Introduction

When completed, our completed stochastic inversion tool will explicitly integrate reactive transport modeling, facies-based geostatistical methods, and a novel stochastic inversion technique to optimize agreement between observed and predicted storage performance. Such optimization will be accomplished through stepwise refinement of: 1) the reservoir model—principally its permeability magnitude, anisotropy, and heterogeneity—and 2) geochemical parameters—primarily key mineral volume fractions and kinetic data. We anticipate that these refinements will facilitate significantly improved history matching and forward modeling of CO_2 storage. We are unaware of any previous attempts to explicitly integrate seismic and geochemical data, which represent key components of CO_2 monitoring for Weyburn and future commercial storage projects.

We believe that the stochastic inversion tool under development will optimize agreement between predicted and observed storage performance. We expect that the tool will identify optimal: a) permeability distributions in the reservoir, b) mineral volume fraction and c) kinetic data. Our tool uses the Markov Chain Monte Carlo (MCMC) methodology (e.g., Aines, et al., 2002; Ramirez et al., 2005). Model optimization will be carried out in two steps. First, we will use seismic data to refine the spatial distribution of reservoir permeability, which effectively defines the spatial framework of CO_2 migration; then, we will use fluid chemistry data to refine mineral volume fractions and kinetic data, which effectively control CO_2-triggered geochemical changes in the reservoir.

The seismic measurements detect the CO_2 distribution that is controlled by the reservoir permeability. They also define the spatial distribution of physical trapping mechanisms that will keep the CO_2 in the reservoir over long time periods. The fluid chemistry data tracks the compositional evolution controlled by water-CO_2-mineral reactions. The tool will also use CO_2/H_2O injection and HC/H_2O production data to constrain the fluid flow simulations. Core logs, geophysical logs and knowledge of the depositional environment will be used to constrain random reservoir model realizations.

The following models are embedded within our stochastic inversion algorithm: 1) a geostatistical model that proposes geologic realizations of the reservoir that are randomly perturbed about some initial reservoir model, 2) a flow simulator that predicts CO_2 migration and CO_2/H20 production, 3) a seismic forward model that predicts zero-offset seismograms using pressure and saturation information produced by the CO_2 flow simulator, 4) an extended reactive transport model that calculates aqueous chemical reactions as a function of space and time, and predicts dissolved chemical species.

This document describes the methods required to carry out the first inversion step: identify reservoir models that optimize agreement between the observed and predicted seismic response; we will refer to this the permeability step. We will discuss in detail items 1 – 3 in the previous paragraph. Future deliverables will describe methods required to carry out the second inversion step: identify geochemical parameters (mineral volume fraction, kinetic parameters) that optimize agreement between and observed aqueous chemistry.

Approach description

MCMC Introduction

Our MCMC approach is a derivative of the Metropolis algorithm (Metropolis et. al., 1953) as described by Mosegaard and Tarantola (1995). It uses a Markov chain process to control the sampling of the model space X. Within this framework, the solution to an inverse problem is an estimate of the posterior probability distribution defined over model space X. Then, for any potential solution $x_0 \in X$, the method will provide an estimate of the probability and confidence that state x_0 (i.e., proposed model) is the true state of the underlying system.

There are two major components to the approach:

1) A base representation specifying the rules that the proposed reservoir models of the system must obey. These rules are based upon a priori knowledge such as information from geophysical logs and core measurements.

2) A Markov Chain Monte Carlo (MCMC) simulation algorithm that generates samples according to the unknown posterior distribution. Our Bayesian approach is driven by an importance sampling algorithm (described in a subsequent section). It uses a randomized decision rule to accept or reject the proposed models according to their consistency with the observed data.

The main advantage of this approach is that it automatically identifies alternative models that are consistent with all available data, and objectively ranks them according to their posterior probabilities and associated confidences. In most geophysical applications, the inverse problem is substantially under-constrained and ill-posed. Thus, the search for a solution that is unique and possesses a high degree of confidence is generally impossible. Hence, it is wiser to consider approaches that are capable of generating alternative models and ranking them.

The MCMC approach is similar to classical deterministic inversion with the random model generator replacing the deterministic updating scheme (usually based on a gradient search). In both cases, an initial model is chosen and responses are calculated with a forward solver. The calculated responses are compared to observed data. Finally, an updated model is chosen and the process repeats. The two approaches differ in how the updated model is chosen and the final result of the process. Specifically, MCMC produces a probability distribution defined over X; while, deterministic methods produce a single or a collection of states from X that best explain the data.

The inverse problem under consideration may be described as follows. Let D denote the data space, and $\underline{x}$ denotes the model, then suppose that there exists a mapping G such that:

$$\underline{d} = G(\underline{x}) \qquad (1)$$

The goal is to find an $\underline{x}_0$ that corresponds to the set of observations, $\underline{d}_0$. An important point here is that X, the range of possible solutions, is limited by a priori knowledge. In our application, this prior knowledge consists of data such as: approximate locations of lithology boundaries, lithology types, knowledge of the depositional system, and correlations between porosity, lithology and permeability.

The MCMC sampling process can be viewed as consisting of two separate components: prior knowledge and measurements $\underline{d}$. We will describe in detail the first component, the generation of model realizations that are consistent with the available prior knowledge in the upcoming section titled " Propose Reservoir Model Realizations". We now discuss the second component: a decision process that either accepts or rejects these model realizations according to their consistency with the measurements $\underline{d}$.

Specifically, for each reservoir model realization, the seismic forward simulator is used to predict a measurable quantity such as amplitude of a reflector; the seismic simulator will be described in detail in upcoming sections. These predictions are then compared to corresponding measurements to determine the likelihood $L(\underline{x})$ that the given state $\underline{x} \in X$ produced the observed data. An accept/reject decision based upon this likelihood is used to modify the prior sampling process. The result is a new Markov chain, **R**, which samples the posterior distribution, $P(\underline{x})$. These samples provide the basis for estimating the posterior distribution and any subsequent inference concerning the true unknown state of the system.

All MCMC approaches are based on Bayesian inference. Formally, Bayes rule relates the prior and posterior distributions as follows:

$$P(\underline{x}|\underline{d}) = c\, L(\underline{x})\rho(\underline{x}) \qquad (2)$$

where $P(\underline{x}|\underline{d})$ is the posterior distribution, and $L(\underline{x})$ is the likelihood function, a measure of the degree of fit between the data predicted assuming model x and the observed data. Based on the available prior data, the prior $\rho(x)$ can have a distribution that is uniform, Gaussian, or unknown. We do not need to know $\rho(x)$ explicitly because the method only requires the ability to produce samples from the distribution.

Permeability Step Description

We now describe how we have adapted the MCMC approach to carry out the first inversion step: identify reservoir models that optimize agreement between the observed and predicted seismic response. Figure 1 schematically shows a flow diagram indicating the methods required to identify reservoir models that produce CO_2 plumes that provide optimal agreement with the seismic data (permeability step). We explicitly account for uncertainties associated with both limited sensitivity and measurement error by using the Metropolis-Hastings technique, which is the core of our MCMC approach (Moosegrad and Tarantola, 1995).

We assume that CO_2 plume evolution is well mapped by 4D seismic reflection technique. We also assume that the flow/reactive transport simulator (NUFT, Nitao, 1998 a,b) produces sufficiently accurate predictions of reservoir conditions. Furthermore, we assume that Gassman's fluid substitution model (Gassmann, 1951) produces sufficiently accurate seismic velocity predictions given the temperature, pressure and fluid saturation conditions in the reservoir.

The process starts by proposing one reservoir model realization, distinguished by its representation of permeability magnitude, anisotropy, and heterogeneity. Then, using the multi-phase/multi-component reactive transport model NUFT, we simulate CO_2/H_2O injection, HC/H_2O withdrawal, and CO_2 migration. Next, we calculate a first-order estimate of seismic response to these reservoir conditions. P-wave velocities are predicted as a function of fluid-phase saturations, pressure, and porosity using Gassmann's equation. The velocity is used together with calculated bulk densities to generate an intra-reservoir reflectivity series. These series are convolved with an estimate of the source wavelet to predict time-series waveforms that can be directly compared to observed seismic response. The likelihood for each model is computed and the Metropolis-Hastings algorithm is used to decide whether the proposed model produces relatively close agreement between predicted and observed 4D seismic reflection data. Then, a new reservoir model is proposed and the process is repeated until the process converges. This approach will produce a posterior distribution that ranks all the posed reservoir models on the basis of likelihood, which is an explicit measure of consistency between predicted and observed seismic response. We will now describe each of these steps in detail.

PROPOSE RESERVOIR MODEL REALIZATIONS

We are following a staged approach in developing the algorithm that proposes reservoir model realizations. In the initial project stages we are using a simpler algorithm that will be replaced by a more complicated and realistic algorithm in later project stages. This document will describe the simpler algorithm. In future project phases, we will use a more advanced and realistic method to propose reservoir models -- TProGS (Transition Probability Geostatistical Software). It will generate realistic 3D lithologic distributions that honor both lithologic/geophysical core data and facies transitional/juxtapositional relationships that characterize the relevant depositional environment (Carle and Fogg,

1997; Carle et al., 1998). These lithologic models will be populated with porosity/permeability data based on prior knowledge of the relationship between lithology and porosity/permeability for the specific depositional setting.

Here we describe in detail the simpler algorithm. The reservoir model proposal step starts (top left box, Figure 1) by reading in EnCana's single-permeability (matrix flow only) reservoir model as truth and propose as a series of realizations (or perturbations) about this "reality" that are bound by known/estimated uncertainties in lithologic distribution and dependent porosity/permeability.

We use available porosity/permeability data to constrain the reservoir model realizations. We have identified six layers that are included in the "reservoir box", i.e., the volume where our calculations will concentrate. The reservoir box includes: Ratcliffe formation, Midale evaporite, Midale Marly, Midale Vuggy , Frobisher Marly and Frobisher Vuggy. Currently, we have sufficient permeability/porosity data to estimate the statistical properties of the Midale Marly, Midale Vuggy, Frobisher Marly and Frobisher Vuggy. For the Ratcliffe and Midale evaporite layers we will assign values based on assumed statistical distributions.

Core porosity/permeability data for the Midale and Frobisher formations are shown in Figures 2 and 3. This data was provided by Erik Nickel, Petroleum Geology Branch, Saskatchewan Ministry of Energy and Resources.

The histograms in the top row of Figure 2 show the distributions of porosity. The histograms suggest that both formations have bimodal porosity distributions. In the Midale, the most likely mode is centered near 0.1 and the second most likely mode is centered near 0.24; these two modes correspond to the Vuggy limestone and Marly dolostone units, respectively. In the Frobisher, the most likely mode is centered near 0.07 and the second mode is centered near 0.24. The histograms suggest that smaller porosity values are more frequent than the larger values and that the porosities range from 0.03 to 0.35.

The bottom row histograms in Figure 2 show the distributions of log10 permeability. There are some large peaks for values below 10^{-2} milli-darcies that appear to be caused by permeabilities set to a default value; we will ignore these peaks. The histograms suggest that both permeability distributions are uni-modal and follow a lognormal distribution. The values are centered near 10^{-1} milli-darcies in both the Frobisher and Midale units but the values can range from 10^{-4} to 10^{4}.

We now examine the correlation trends between porosity and permeability. The top row of Figure 3 shows plots of log10 permeability plotted as a function of porosity. The plots suggest that porosity and permeability are correlated in both units. Both plots show two clusters, one centered near porosity of 0.1 (Vuggy limestone) and the other centered near porosity 0.25 (Marly dolostone). The former cluster shows a larger spread of permeability values (10^{-2} to 10^{3}) than the cluster associated with the Marly dolostone (10^{-1} to $10^{1.5}$).

In order to use these porosity/permeability trends to constrain the reservoir model realizations, we have computed the mean and standard deviation of log10 permeability as function of porosity; we used porosity window widths of 0.01 units to do the calculations. The results are shown in the bottom row of Figure 3. The red circles indicate the mean log10 permeability for all values within a window; the blue circles indicate the median value. The triangles indicate the 1 standard deviation spread about the mean value.

The Vuggy and Marly sub-units in the Midale and Frobisher formations can be approximately segregated based on porosity magnitude (Erik Nickel, personal communication, 2010). The Vuggy typically has lower porosities than the Marly. The bottom row plots in Figure 3 shows that there are also observable differences between the Vuggy and Marly porosity-permeabilty trends; the black line segments illustrate these trends. These plots show that larger porosity values are correlated with increasing permeability values and that there is change in the trend around porosity of 0.15 where the transition from Vuggy to Marly occurs. If we ignore the points in the overlap region (porosity range 0.15 – 0.21), we see that the porosity –permeability trend for each subunit is linear. We also see that each sub-unit has a characteristic slope and intercept. For example, the slopes associated with the Midale and Frobsiher Vuggy sub-units are similar. However they are different from the slopes associated with the Midale and Frobisher Marly sub-units. The plots also show the spread in permeability gets smaller and that the slope gets smaller (particularly in the Frobisher) when the porosity is > 0.15.

We use the porosity frequencies in Figure 2 together with the porosity/permeability trends in Figure 3 to propose porosity and permeability fields that honor the available core data, for each layer in the reservoir box. We now describe the algorithm that proposes the realizations of porosity and permeability. We will refer to this algorithm as the "sampler" because it produces samples of the prior distribution of models, i.e., those models that honor the data shown in Figures 2 and 3. An example of permeability realizations resulting from this algorithm is shown in Figure 4.

We assume as truth the layers in EnCana's geologic model, i.e. the lithology designations and the boundaries between layers remain fixed from one realization to the next. The porosities and permeabilities assigned to nodes within each layer are allowed to vary spatially and from one realization to the next. Each new realization consists of a perturbation about the previous realization.

Sampler algorithm (comments shown in *italics*):
1) Read in EnCana's geologic model *(developed by EnCana during Phase I of the project, provided by Barbara Dietiker, Geological Survey of Canada)*

2) Initialize layer values.
For each lithology layer:
2a) Read starting model containing lithologies, and initial porosities, and permeabilities
2b) calculate average value of porosity and permeability
2c) identify calculation grid nodes associated with each layer

2d) assign average value of porosity and permeability to each node

3) For each each realization of porosity and permeability fields:
For each lithology layer:
3a) choose at random ~ 10% of nodes to change porosity/permeability
For each node to be changed:
3a.1) randomly choose a value of porosity probability (chosenProb)
(use the probability ranges implied by the Fig. 2 histograms)
3a.2) list possible porosity values that have probability > chosenProb
3a.3) choose random porosity value from the 3a.2 list
3a.4) list possible permeability values for the porosity chosen in 3a.3
(use the porosity/permeability correlations in Fig. 3, bottom row)
3a.5) choose a random permeability from the list of possible values, (assuming a lognormal distribution).
3a.6) assign new porosity and permeability values to grid node

These realizations are used by the flow simulator (NUFT) to predict flow within the reservoir reservoir caused by injection/extraction of CO_2, H_2O and oil. The next section describes the flow simulator and how it is used in our application.

RUN FLOW SIMULATOR

We use the Nonisothermal Unsaturated-Saturated Flow and Transport code (NUFT) for reservoir-scale multiphase flow and reactive transport simulations. The NUFT code is a highly flexible software package for modeling multiphase, multi-component heat and mass flow and reactive transport in unsaturated and saturated porous media (Nitao, 1998). An integrated finite-difference spatial discretization method along with implicit time-integration scheme is used to solve mass and energy balance equations in flow and reactive transport models. At each time step the resulting nonlinear equations are solved by the Newton-Raphson method. The NUFT code, which is capable of running on PCs, workstations, and major parallel processing platforms, has been widely used for many applications such as geologic disposal of nuclear waste, CO_2 sequestration and storage, groundwater remediation, and subsurface hydrocarbon production (Buscheck et al., 2003; Glassley et al., 2003; Johnson et al., 2004, 2005).

For multiphase flow and reactive transport modeling, a 3D-domain is selected to represent a reservoir box centered at a depth of about 1400 m. This domain represents pattern 7, Phase 1A area. The modeling box includes the Ratcliffe, Midale Evaporite, Midale Marly, Midale Vuggy, Frobsiher Marly and Frobisger Vuggy units. While the top and bottom boundaries are kept impermeable the hydrostatic pressure conditions are assigned along the lateral boundary. In addition, vertical and horizontal production wells, and horizontal injection wells are included in the model.

The CO_2 is injected under supercritical conditions. Based on reservoir geologic model the reservoir domain is divided into two geologic layers, Vuggy and Marly. Overburden and underburden formations will be also considered for simulations done during later

stages of the project. We address reservoir heterogeneity by generating realizations of permeability and porosity fields within each layer. The algorithm described in the section "Propose reservoir model realizations" generates these realizations. Due to the presence of water, oil and supercritical CO_2 three-phase flow conditions are considered, and equilibrium conditions are assumed for component partitioning among three phases. The equation-of-state and viscosity of CO_2 under supercritical conditions are obtained based on the empirical equations developed by Span and Wagner (1996) and Fenghour and Wakeman (1998), respectively. The water and oil properties (pressure, volume, temperature) will be computed using the empirical formulas that can be found in most of textbooks on reservoir modeling (e.g. Chen et al., 2006). The injection-induced reservoir pressure perturbations along with CO_2 plume distributions will be predicted by the reservoir flow model, and then supplied to the algorithms that predict seismic data (described next).

CALCULATE SEISMIC LIKELIHOOD

The next step in the processing pipeline is to use a seismic forward model that predicts zero-offset seismograms using information produced by the CO_2 flow simulator (NUFT). Here we describe the various components of the seismic forward model which include (refer to Figure 5):

1) Gasmmann's petrophysical model to predict changes in bulk modulus and seismic velocities caused by changes in reservoir conditions,
2) module that predicts seismic reflectivities within the "reservoir box",
3) module that predicts zero-offset (1D) seismograms,
4) module that compares the predicted and observed seismograms to compute the likelihood that the predicted waveforms come from the "true" model.

Compute reflectivities

As the reservoir fluids, temperature and pressure conditions change during CO_2 injection, the reservoir's bulk modulus and density changes. Consequently, seismic velocities change resulting in seismic reflectivity changes within the reservoir changes. Fluid substitution techniques form an important part of the seismic modeling toolbox because they can be used to predict these changes for a range of fluid scenarios. Gassmann's equation (Gassmann, 1951) is the most common and theoretically sound fluid substitution technique at seismic frequencies (Smith et al., 2003). It relates the bulk modulus of a rock to its porosity, bulk modulus of its porous rock frame, bulk modulus of the mineral matrix, and bulk modulus of the fluids. The rock's bulk modulus can then be used to predict seismic velocities.

We next summarize our implementation of the algorithm that uses Gassmann's equation. We follow the procedure described in Smith et al., 2003 but introduce some modifications where some derived quantities are replaced with quantities measured using Weyburn core.

1. Look up mineral matrix bulk modulus (K_o) based on core measurements (Table 3.6, Wilson and Monea, 2004, and mineral moduli from Smith et al, 2003).

2. Calculate fluid bulk modulus and density based on saturations, bulk moduli and densities of constituent fluids for appropriate pressure, temperature conditions. Obtain individual bulk moduli from equations of state for each fluid. Use equation 1 (below).
3. Look-up porous rock framework bulk modulus (K^*) for appropriate pressure, temperature conditions (as determined from lab measurements, Brown, 2002).
4. Calculate moduli for saturated rock using the Gassmann equation. Use equation 4 (below) and appropriate porosity value (ϕ).
5. Calculate Vp, Vs for saturated rock using the moduli from step 4. Use equations 5-7 (below).

The equation used to calculate the composite fluid modulus is:

$$
\begin{aligned}
K_{fl} &= \frac{S_w}{K_w} + \frac{S_{oil}}{K_{oil}} + \frac{S_{co2}}{K_{co2}} \\
K_w &= brine\_\mathrm{mod}ulus \\
K_{oil} &= oil\_\mathrm{mod}ulus \\
K_{co2} &= co2\_\mathrm{mod}ulus \\
S_w &= brine\_saturation \\
S_{oil} &= oil\_saturation \\
S_{co2} &= co2\_saturation
\end{aligned}
\qquad (3)
$$

The brine modulus (K_w) is calculated using the approach of Batzle and Wang, 1992, equations 27 - 29. They use a combination of thermodynamic relationships, empirical trends and data to develop simplified relationships that produce estimates of realistic fluid properties in rock models.

The oil modulus (K_{oil}) and CO_2 modulus (K_{co2}) will be computed using the compressibilities computed by the NUFT simulator.

Gassmann's equation used to calculate saturated rock bulk moduli (K_{inj}) is:

$$
K_{inj} = K^* + \frac{(1 - \frac{K^*}{K_0})^2}{\frac{\phi}{K_{fl}} + \frac{1-\phi}{K_0} + \frac{K^*}{K_0^2}} \qquad (4)
$$

Once (K_{inj}) is calculated, we are ready to calculate the P and S seismic velocities (V_p ,V_s):

$$
V_p = \sqrt{\frac{K_{inj} + \frac{4}{3} G_{init}}{\rho_b}} \qquad (5)
$$

ρ_b is the bulk density under initial (pre-injection) conditions and,

$$G_{init} = \rho_b V_{s,core}^2 \qquad (6)$$

where $V_{s,core}$ is the shear velocity measured on rock core.

$$V_s = \sqrt{\frac{G_{init}}{\rho_b^{'}}} \qquad (7)$$

In equation 7, V_s is the shear velocity measured in the field and $\rho_b^{'}$ is the bulk density after fluid substitution.

Calculate synthetic waveforms

We are now ready to calculate the synthetic waveforms that sample the layers within the reservoir box. We have chosen the approach described by Margrave, 2003, to construct 1D seismograms. Use of a 1-D model is considered appropriate because the seismic data we will use has been migrated and the layering is approximately horizontal.

This 1D approach requires that all calculations needed to produce synthetic seismograms be done on columns of grid nodes. Calculations that make use of equations 3 – 10 are performed on grid node columns. Each set of calculations along a column is independent of node values on other node columns. Figure 5 schematically shows the operations needed to calculate the seismograms and seismic likelihoods.

Margrave's approach constructs the reflection response of a 1D medium to a unit impulse. The impulse response only considers primary reflections and ignores multiple reflections. This impulse response is then convolved with an appropriate wavelet that simulates the effects of the seismic source waveform. The resulting waveform is the synthetic 1D seismogram that will be compared with the observed 1D seismograms. This 1D modeling approach will produce seismograms that are consistent with the available, zero-offset, processed seismic data.

As indicated in Figure 5, once the seismic velocities are calculated, the reflection coefficients (*R*) within the "reservoir box" are computed. These coefficients determine the amount of reflected signal generated at each layer interface and, collectively form the impulse response along a vertical line of nodes. In equation 8, R is the reflection coefficient between two layers having seismic impedances I_1 (top layer) and I_2 (bottom layer). Seismic impedance is a function of seismic velocity and bulk density as shown in equation 9.

$$R = \frac{I_1 - I_2}{I_2 + I_1} \qquad (8)$$

$$I = V\rho_b \qquad (9)$$

Margrave indicates that synthetic seismograms intended for comparison with fully processed seismic data will usually be created using a zero-phase wavelet. The wavelet simulates the effects of an impulsive source and band-width limits the source spectrum. Two common zero-phase wavelets are the Ornsby and Ricker wavelets. Margrave suggests that the Ornsby wavelet is thought to have an unacceptable level of ripple. The Ricker wavelet has a simpler form in the time-domain though it has a broader, less controlled passband. A characteristic of this wavelet is that the higher the dominant frequency, the broader the bandwidth. We have chosen to implement the Ricker wavelet to calculate the synthetic seismograms for the purpose of algorithm testing. Eventually, a wavelet that is derived from the observed seismic data will be used in this process. Equation 10 shows the analytical form of the wavelet.

$$w(t) = [1 - 2\pi^2 f_{dom}^2 t^2] e^{-\pi^2 f_{dom}^2 t^2} \qquad (10)$$

w(t) is the amplitude as a function of time, t. f_{dom} represents the dominant frequency of the wavelet. Figure 6 shows a Ricker wavelet with a dominant frequency of 60 Hz in the time-domain and a 1 ms sampling interval. Notice that the wavelet is symmetric about 0, and thus, has a zero-phase delay.

The wavelet is then convolved with the impulse response to produce the synthetic 1D seismogram. Figure 7 helps to summarize this process. The left graph shows a 1D velocity and bulk density model used to calculate a synthetic seismogram. The middle graph shows the impulse response obtained for this 1D model. Notice that the magnitude of the response is largest where the velocity contrast between adjacent layers is largest. Also, the sign of the response depends on whether the velocity is increasing from one layer to the next (positive response) or decreasing (negative response). The impulse response would be the seismogram observed if the seismic source used to conduct the survey was perfectly impulsive and had infinite bandwidth. Real seismic sources are not perfectly impulsive and only offer limited bandwidth. Thus, to simulate a real seismic source, we need to simulate its limited bandwidth by convolving the impulse response with the wavelet described by equation 10 and shown in Figure 6. The predicted seismogram is plotted on the right side of Figure 7. Each deflection indicates the two-way travel time to a given reflecting interface. Its frequency spectrum has been band-limited by the wavelet convolution, and consequently, the pulses associated with each reflecting interface are broader (and more realistic) than the corresponding impulse response pulses.

Once seismograms are predicted for each column of grid nodes, we are ready to compare the predicted and observed seismograms.

Use Metropolis-Hastings to accept or reject proposed reservoir model

CALCULATE LIKELIHOOD FUNCTION

We now describe Metropolis-Hastings algorithm, the importance sampling algorithm we use to decide whether the proposed model should be accepted or rejected. The likelihood $L(\underline{x})$ is a measure of the degree of fit between the data predicted (i.e., the calculated

seismograms described previously) assuming model $\underline{x}$ and the observed data (fully processed, filed data derived seismograms), and *k* is a normalizing constant. In our approach, we use a likelihood function of the form:

$$L(\underline{x}) = k \exp\left(-\frac{1}{n} \sum_{i=1}^{N} \frac{\left| d(\underline{x})_{pred,i} - d_{0i} \right|^n}{\sigma_i^n} \right) \qquad (11)$$

where *N* is the number of data points, $d(\underline{x})_{pred,i}$ is the predicted data for a given model $\underline{x}$, $d_{0,i}$ is the vector of observed measurements, σ_i is the estimated data uncertainty, and $n \geq 1$. We note that most deterministic inversions also use the term in parentheses as a measure of goodness of fit.

The decision to accept or reject a proposed model is made on the basis of likelihood comparisons. Suppose that the current model of the Markov chain is $\underline{x}^{(T)}$ and that a move to an adjacent model $\underline{x}^{(T+1)}$ is proposed. If these transitions were always accepted, then our MCMC method would be sampling from the prior distribution $\rho(\underline{x})$, i.e., the observed data $\underline{d}_0$ would not influence the search.

$$\begin{aligned} &L(\underline{x}^{(T+1)}) > L(\underline{x}^{(T)}) : P_{accept} = 1 \\ &L(\underline{x}^{(T+1)}) \leq L(\underline{x}^{(T)}) : P_{accept} = L(\underline{x}^{(T+1)}) / L(\underline{x}^{(T)}) \end{aligned} \qquad (12)$$

Instead, suppose that the decision to accept the proposed transition is made as follows (see eq. 12). When the likelihood of the proposed model $L(\underline{x}^{(T+1)})$ is equal to or larger than that of the current model $L(\underline{x}^{(T)})$, the proposed transition is always accepted. If $L(\underline{x}^{(T+1)}) < L(\underline{x}^{(T)})$ but the two values are close to each other, the probability of acceptance is still around 1.0. For example, suppose that $L(\underline{x}^{(T)}) = 10$ and $L(\underline{x}^{(T+1)}) = 9$. In this case, the probability of acceptance (P_{accept}) will be 9/10 or 0.9. We then generate a uniformly random number *RN* in the range 0 to 1.0. When $P_{accept} > RN$, the transition to model $\underline{x}^{(T+1)}$ is accepted. Note that there is a high probability of accepting $\underline{x}^{(T+1)}$ because the odds are very high that $P_{accept} > RN$. Next, let's suppose that the model $\underline{x}^{(T+1)}$ is much less consistent with the data such that $L(\underline{x}^{(T+1)}) = 0.9$. In this case, P_{accept} is 0.09, the odds that $P_{accept} > RN$ are much smaller and thus, the odds of accepting the transition are a lot smaller. However, even when $L(\underline{x}^{(T+1)}) << L(\underline{x}^{(T)})$, P_{accept} is not zero. Thus, this randomized rule allows a transition to a less likely model such that the process will move out of a local extremum. Theoretically, it will never get trapped in a region of locally high likelihood as long as the likelihood of the proposed model is greater than 0.0. Then,

the randomized acceptance rule guarantees that the probability of accepting this transition will always be greater than 0.0.

Metropolis et al. (1953) proved that the samples generated through this process has a limiting distribution that is proportional to the desired posterior distribution $P(\underline{x}|\underline{d})$ -- the probability of model $\underline{x}$ being the true model of nature given that $\underline{d}$ has been measured. As a result of the randomized rule, the search tends to hover in regions of space X containing models that better fit the prior information and seismic measurements. Because of this, space X is traversed more efficiently than with traditional Monte Carlo techniques.

If the decision is to accept model $\underline{x}^{(T+1)}$, it becomes a part of the posterior distribution, and the next proposal ($\underline{x}^{(T+2)}$) will be generated as a random perturbation about $\underline{x}^{(T+1)}$. Alternatively, if the decision is made to reject $\underline{x}^{(T+1)}$, a copy of $\underline{x}^{(T)}$ is added to posterior distribution, and the next proposal ($\underline{x}^{(T+2)}$) will be generated as a random perturbation about $\underline{x}^{(T)}$. The net result of this approach is that solutions are sampled at a rate proportional to their consistency with available data.

This is a key strength of our MCMC approach. Models that are most consistent with available data observations are sampled most often, while models that are incompatible with either prior information and/or observations are rarely sampled. As a result, the frequency of models in the posterior distribution $P(\underline{x}|\underline{d})$ can be used to determine the probability that a given model is the best explanation for the available data. It can also be used to objectively rank alternative models that are consistent with the data.

Compare predicted and observed seismic waveforms

In our seismic application, we need to compare predicted and observed waveforms in order to calculate the likelihood function, specifically the numerator term in eq. 11. The algorithm we have developed follows (*comments shown in italics*).

1) Trim the observed waveforms (W_{obs}) to only include the part that corresponds to the reservoir box.

2) Scale the amplitudes of the predicted waveforms (W_{pred}) and W_{obs}. Scale such that the maximum value of each is set to 1.0. *This is done because the W_{pred} and W_{obs} data have arbitrary units and thus, need to be amplitude-matched*.

3) Cross-correlate W_{pred} and W_{obs} to determine the delay-time.

4) Time-shift W_{pred} so that it aligns with W_{obs}.

5) Compute the difference in amplitude for each waveform time-step and sum the absolute value of the differences. This corresponds to the operation shown on the right hand of eq. 11.

Analyze the posterior distribution

After convergence has been verified and pre-burn-in models discarded, the reservoir models in the posterior distribution $P(\underline{x}|\underline{d})$ can be analyzed. Our goal is to distill the relevant information in these models so that that we can infer the likely properties of the "true" reservoir model under study. The topography of $P(\underline{x}|\underline{d})$ contains multiple hills whose heights are proportional to the likelihoods for each of its member reservoir models (*this n-dimensional space is* shown schematically in Figure 8); each point in model space represents one reservoir model. The model corresponding to the peak of each hill is commonly referred to as its mode. Multiple hills indicate that the solution to the inverse problem is non-unique, the typical case for geophysical inversions. The distribution is called multi-modal when multiple hills are present and uni-modal when only a single hill is present. The width of each hill indicates that there is uncertainty in the model located at the mode; this variability may be due to factors such as measurement sensitivity or measurement error.

This complex, multi-modal structure provides a challenge when characterizing the distribution and extracting insight about the reservoir models included in $P(\underline{x}|\underline{d})$. We use a clustering approach to extract this insight. Ramirez et al., 2005, describe the clustering technique in detail; here we summarize the approach. Clustering is a standard data-mining technique used to extract structure from a collection of sampled data points – in this case sampled reservoir models. It segregates the models sampled from $P(\underline{x}|\underline{d})$ into groups of models that exhibit similar properties. In our specific example, a cluster is a group of reservoir models that show similar spatial distribution of lithology and other relevant properties such as porosity and permeability. The likelihood modes in Figure 8 represent these model clusters. The clustering process is accomplished by measuring the distance (in model space), between a model and a cluster center. A cluster's center is the model space location that best represents the central tendencies of all cluster members.

When deciding whether a reservoir model should be considered a member of a particular cluster, we measure the distance (in model space) between the candidate model and the cluster center. This distance is a measure of the dissimilarity of the sample relative to the central tendencies of all the models that are already members of the cluster. A cluster's central tendencies are represented by element-wise distribution of lithology (or other relevant properties). That is, for each element, we calculate histograms that show how frequently each of the possible lithology values appears in all models included in the cluster; these frequencies are normalized to lie between 0 and 1. The following simplified example should help clarify this method.

Suppose that each reservoir model consists of three elements and that the set of possible lithologies is: {1: dolostone, 2: limestone, 3: evaporite, 4:shale}. Suppose further that there are 100 models in a cluster with the element-wise frequency distributions shown in Table 1. The frequencies for these are calculated by dividing the number of models

showing a particular lithology index by the total number of models in the cluster. Table 1 suggests that, for element 1 there are 30 samples with lithology = 1 (dolostone -- frequency is 0.3 = 30/100), 40 with lithology = 2 (limestone), 20 with lithology = 3 (evaporite) and 10 with lithology = 4 (shale).

Table 1: Cluster lithology frequencies (crf)

lithology index	Element		
	1	2	3
1-dolostone	0.3	0.3	0.3
2-limestone	0.4	0.2	0.2
3-evaporite	0.2	0.3	0.4
4-shale	0.1	0.2	0.1

Table 2: Model lithology frequencies (mrf)

lithology index	Element		
	1	2	3
1-dolostone	0.0	0.0	0.0
2-limestone	0.0	1.0	0.0
3-evaporite	1.0	0.0	0.0
4-shale	0.0	0.0	1.0

We also need to calculate frequency histograms for the reservoir model in question so that they can be compared to the cluster's histograms. Suppose that the model being considered has lithology values of (3, 2, 4). The lithology frequencies for this model are shown in Table 2; i.e., when one of the possible lithology values is present in a element, the frequency for that lithology value is set to 1.0 and frequencies for all other possible values are set to 0.0; this is repeated for all elements.

We need to compare the frequencies in Tables 1 and 2 in order to calculate the model-cluster 'dissimilarity' *MCD*. For every Table 2 element where the frequency = 1.0, we subtract the corresponding element in Table 1. Thus, *MCD* = [(1.0 - 0.2) + (1.0 - 0.2) + (1.0 - 0.1)]/3 = 0.83. Large *MCD* values arise when the model anomaly is located in a different part of the 3D model, has different lithology indexes, or both. When MCD approaches 0.0, the reservoir model under evaluation shows a lithology distribution that is very similar to that of most reservoir models in the cluster.

The equation for *MCD* can be written as:

$$MCD_{m,n} = \frac{1}{nv} \sum_{i=1}^{nv} \sum_{j=1}^{nr} | mrf_{i,j} (mrf_{i,j} - crf_{i,j}) | \qquad (13)$$

where *m* and *n* are cluster and model identification numbers respectively, *nv* is the number of elements in one resistivity model, *nr* is the number of possible reservoir values, *mrf* is the model's lithology frequency (e.g., Table 1 values) and *crf* is the cluster's lithology frequency (e.g., Table 2 values).

The *MCD* can also be used to locate the "center state" (*CS*) for a given cluster. The *CS* is that resistivity model showing the minimum *MCD*; it is also the model that is closest to the cluster's "center of mass" and should be the one that best represents the cluster members. The *CS* and the mode refer to the same model when the cluster members are distributed symmetrically about the mode.

The algorithm we used to perform the clustering analysis is called the dynamic k-means algorithm described in detail by Ramirez et al. 2005.

Summary

We are developing a stochastic inversion tool that will explicitly integrate reactive transport modeling, facies-based geostatistical methods, and a novel stochastic inversion technique to optimize agreement between observed and predicted storage performance. Such optimization will be accomplished through stepwise refinement of: 1) the reservoir model—principally its permeability magnitude, anisotropy, and heterogeneity—and 2) geochemical parameters—primarily key mineral volume fractions and kinetic data. This document describes development activities associated with the first step: a stochastic inversion method that will identify reservoir models that optimize agreement between the observed and predicted seismic response. This document only describes the tools and approaches used for step 1.

The following models will be used to optimize reservoir model permeabilities: 1) A geostatistical model that proposes random geologic realizations of the reservoir; this module is 90% complete. 2) A flow simulator that predicts CO_2 migration and CO_2/H20 production; this module is 100% complete. 3) A seismic forward model that predicts zero-offset seismograms using information produced by the CO_2 flow simulator; this module is 90% complete. 4) The decision to accept or reject a proposal is based on the Metroplois-Hastings algorithm; this module is 100% complete.

Acknowledgements

We are grateful for the funding support provided by the Petroleum Technology Research Center, Saskatchewan, Canada. We are also grateful for the data provided by Barbara Dietiker (Geological Survey of Canada) and Erik Nickel (Saskatchewan Ministry of Energy and Resources).

Disclaimer

References:

Aines, R.D., J.J. Nitao, R.L. Newmark, S. Carle, A.L. Ramirez, D.B. Harris, J.W. Johnson, V.M. Johnson, D. Ermak, G. Sugiyama, W.G. Hanley, S. Sengupta, W. Daily, R. Glaser, K. Dyer, G Fogg, Y. Zhang, Z. Yu, R. Levine, 2002, The Stochastic Engine Initiative: Improving Prediction of Behavior in Geologic Environments We Cannot Directly Observe, UCRL-ID-148221, Lawrence Livermore National Laboratory, Livermore, CA.

Batzle, M. and Z. Wang, 1992, Seismic properties of fluids, *Geophysics*, v. 57, no. 11, pp 1396-1408.

Brown, L. 2002, *Integration of rock physics and reservoir simulation for the interpretation of time-lapse seismic data at Weyburn Field, Saskatchewan*, M. Sc. thesis, Colorado School of Mines, Golden, CO.

Buscheck TA, Glascoe LG, Lee KH, Gansemer J, Sun Y, Mansoor K, 2003, Validation of the Multiscale Thermohydrologic Model used for analysis of a proposed repository at Yucca Mountain. *J. Contaminant Hydrology*, vol. 62 no. 3, p. 421-440.

Carle, S.F.; Fogg, G.E. 1997, Modeling spatial variability with one and multidimensional continuous-lag Markov chains: *Mathematical Geology*, v. 29, n. 7, p. 891-918

Carle, S.F., Labolle, E.M., Weissmann, G.S., Van Brocklin, D, Fogg, G.E., 1998, Geostatistical simulation of hydrofacies architecture, a transition probability/Markov approach: in Fraser, GS; Davis, JM, Hydrogeologic Models of Sedimentary Aquifers, Concepts in Hydrogeology and Environmental Geology No. 1, *SEPM (Society for Sedimentary Geology) Special Publication*, p. 147-170

Chen, Z, Huan, G, Ma, Y, 2006, Computational methods for multiphase flows in porous media**.** Philadelphia: Society for Industrial and Applied Mathematics.

Fenghour, A, Wakeman, WA, 1998, The viscosity of carbon dioxide**.** *J. Phys. Chem. Ref. Data*, vol. 27, p. 31-44.

Gassmann, F., 1951, Uber die Elastizitat Poroser Medien, *Vier. Der Natur. Gesellschaftn Zurich*, 96, p.1-23.

Johnson, JW, Nitao JJ, Knauss KG, 2004, Reactive transport modeling of CO_2 storage in saline aquifers to elucidate fundamental processes, trapping mechanisms and sequestration partitioning. In: *Geological Storage of Carbon Dioxide*. Edited by SJ Baines, RH Worden. *Geological Society, London, Special Publications* vol. 223, p. 107-128.

Johnson JW, Nitao JJ, Morris JP, 2005, Reactive transport modeling of cap-rock integrity during natural and engineered CO_2 storage. In *Carbon Dioxide Capture for Storage in Deep Geologic Formations*. Edited by D. C. Thomas and S. M. Benson, vol. 2, 787-813.

Marini, L., 2007, Geological sequestration of carbon dioxide: thermodynamics, kinetics and reaction path modeling, *Developments in Geochemistry* **vol. 11**, Elsevier, New York (2007) 453 p.

Margrave, G., 2003, *Numerical Methods of Exploration Seismology with Algorithms in MATLAB*, Dept. of Geology and Geophysics, University of Calgary, Canada, CREWES Research Consortium.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys*., *1*, no. 6, 1087-1092.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems, *Journal of Geophysical Research*, vol. 100, no. B7, p12431-12447.

Nitao, J.J., 1998a, Reference manual for the NUFT flow and transport code, version 2.0: Lawrence Livermore National Laboratory, UCRL-MA-130651, 55 p.

Nitao, J.J., 1998b, User's manual for the USNT module of the NUFT Code, Version 2.0 (NP-Phase, NC-component, Thermal): Lawrence Livermore National Laboratory, UCRL-MA-130653, 76 p.

Peng, Ding-Yu, and D. B. Robinson, 1976, Anew Two constant equation of state, Ind. Eng. Chem., Fundam., Vol. 15, No. 1, p. 59 – 64.

Ramirez, A. L., J.J. Nitao, W.G. Hanley, R.D. Aines, R.E. Glaser, S.K. Sengupta, K.M. Dyer, T.L. Hickling, W.D. Daily, 2005, Stochastic Inversion of Electrical Resistivity Changes Using a Markov Chain, Monte Carlo Approach, *Journal of Geophysical Research,* Vol. 110, B02101, doi: 10.1029/2004JB003449.

Reasnor, M., 2001, *Forward modeling and interpretation of multicomponent seismic data for fracture characterization, Weyburn Field, Saskatchewan*, M. Sc. thesis, Colorado School of Mines, Golden, CO .

Simth, T. M., C. H. Sondergeld, and C. S. Rai, 2003, Gassmann fluid substitutions: A Tutorial, *Geophysics*, Vol. 68, No. 2, p. 430 - 440.

Span, R, Wagner, W, 1996, A new equation of state for carbon dioxide covering the fluid region from the triple-point temperature to 1100 K at pressure up to 800 MPa**.** *J. Phys. Chem. Ref. Data*, vol. 25, p. 1509-1596.

Spycher, N. , Pruess, K. , Ennis-King, J., 2003, CO2-H2O mixtures in the geological sequestration of CO2. I. Assessment and calculation of mutual solubilities from 12 to 100 C and up to 600 bar, *Geochimica et Cosmochimica Acta*, V. 67, no. 16, 1, pp 3015-3031, DOI: 10.1016/S0016-7037(03)00273-4.

Wilson, M. and M. Monea (eds), 2004, *IEA GHG Weyburn CO_2 Monitoring and Storage Project Summary Report 2000-2004*, Petroleum Technology Research Center, Regina, Canada.

Figures

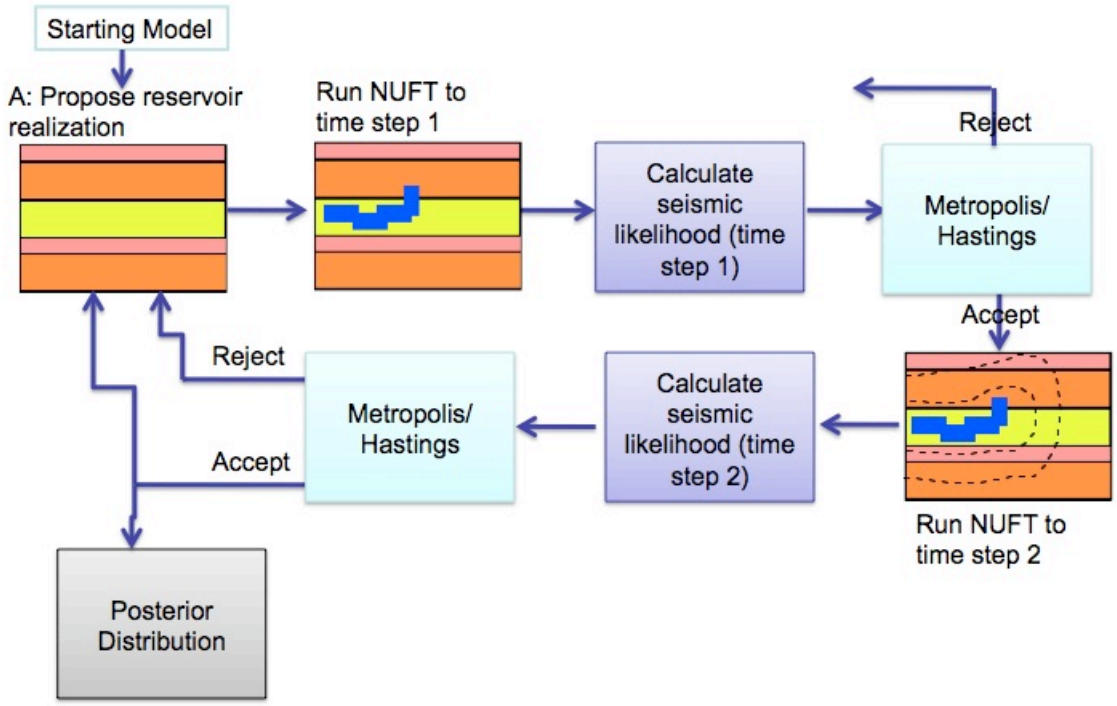


Figure 1. Flow diagram for the first inversion step using seismic data. Note that the process will use two or more seismic data sets simultaneously.

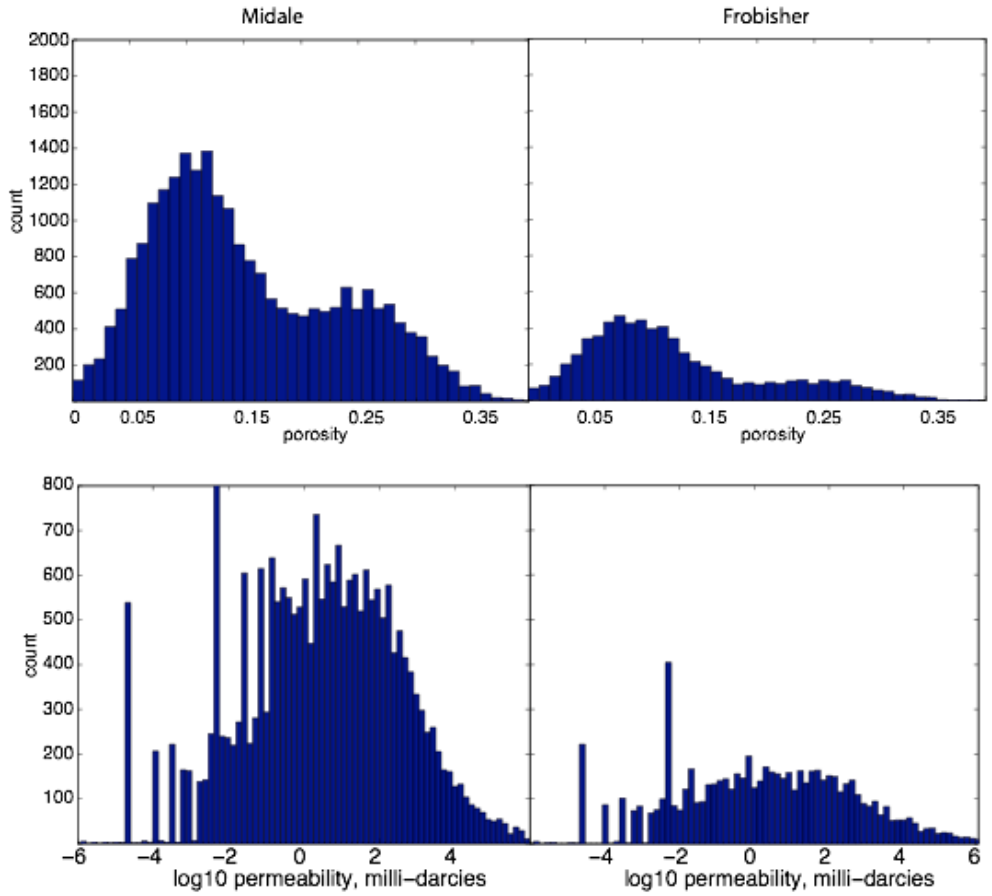


Figure 2. Histograms of core porosity and permeability, Midale and Frobisher formations. Data provided by Erik Nickel, 2009, Saskatchewan Ministry of Energy and Resources.

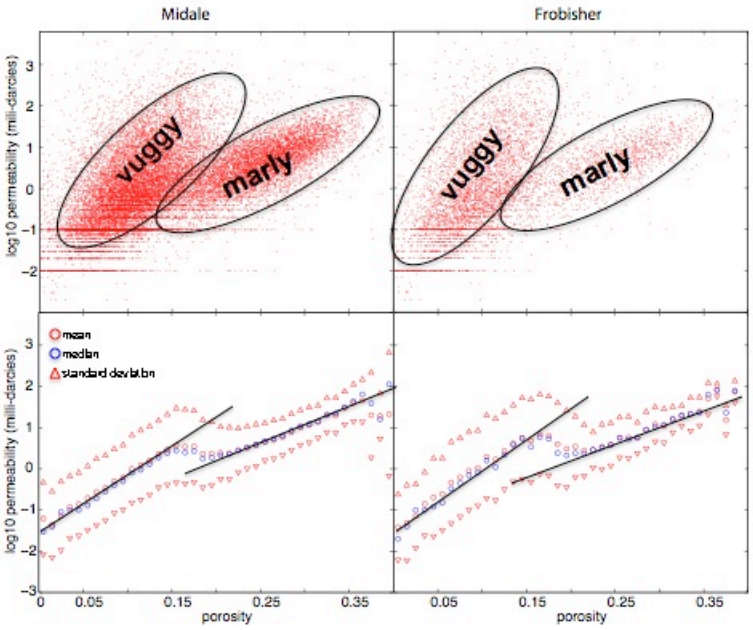


Figure 3. Permeability plotted as a function of porosity, Midale and Frobisher formations, Weyburn reservoir. The Vuggy and Marly sub-units are approximately separated based on porosity. The lower row of plots uses red circles to indicate the mean of log10 permeability for each 0.01 porosity window. The red triangles indicate the standard deviation of log 10 permeability. The black line segments represent the porosity-permeability relationships that will be used for the Vuggy and Marly subunits in each formation.

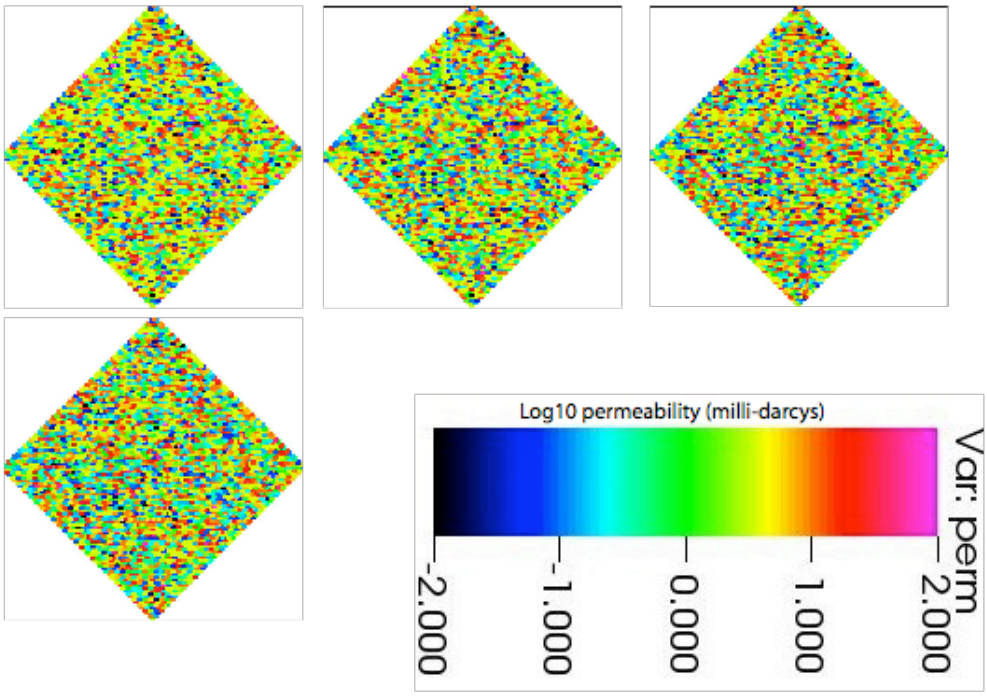


Figure 4. Permeability realizations using “simple” algorithm to randomly populate the porosity and permeability fields using the trends in Figures 2 and 3.

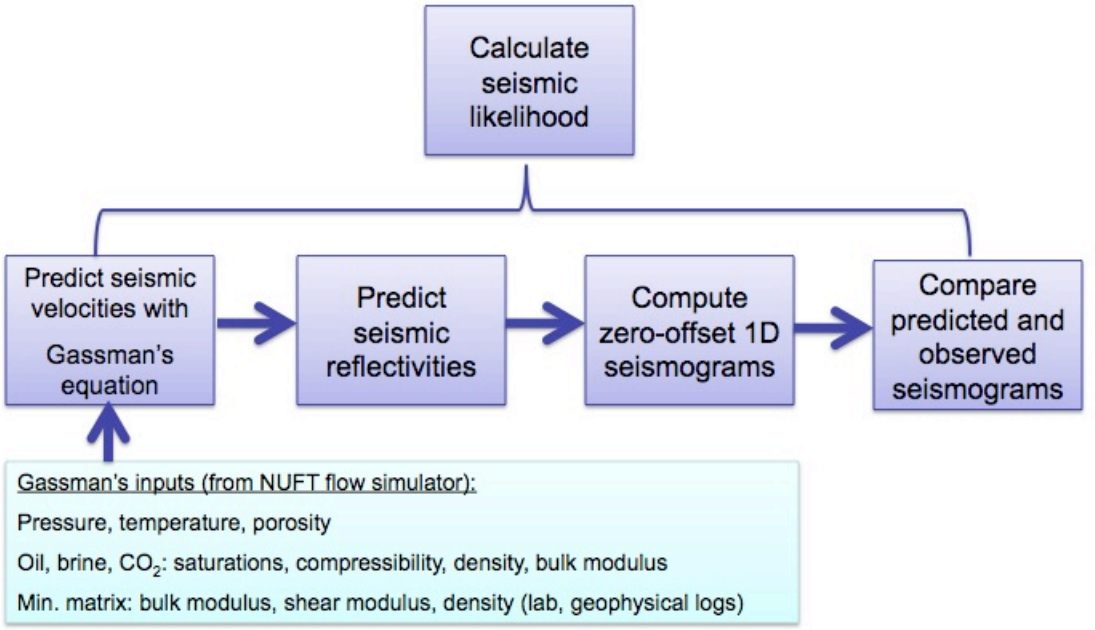


Figure 5. Flow diagram for the seismic likelihood calculation.

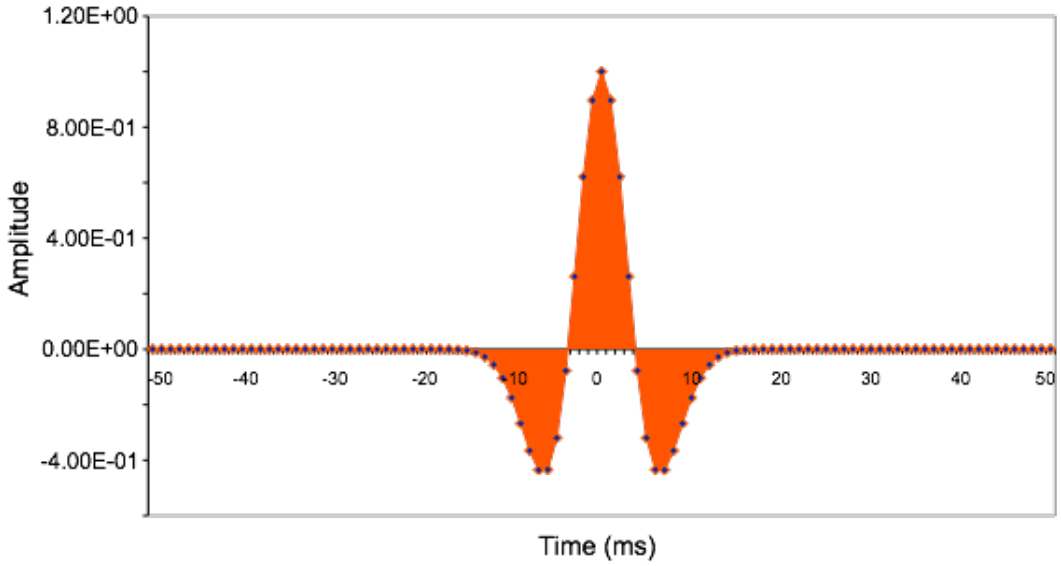


Figure 6. Amplitude of a Ricker wavelet as a function of time. The wavelet's dominant frequency is 60 Hz. The sample rate is 1 ms.

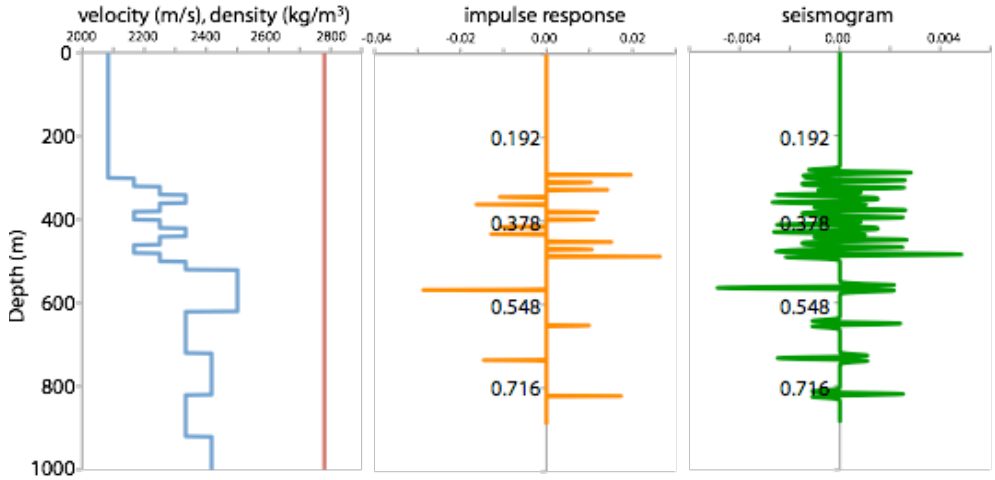
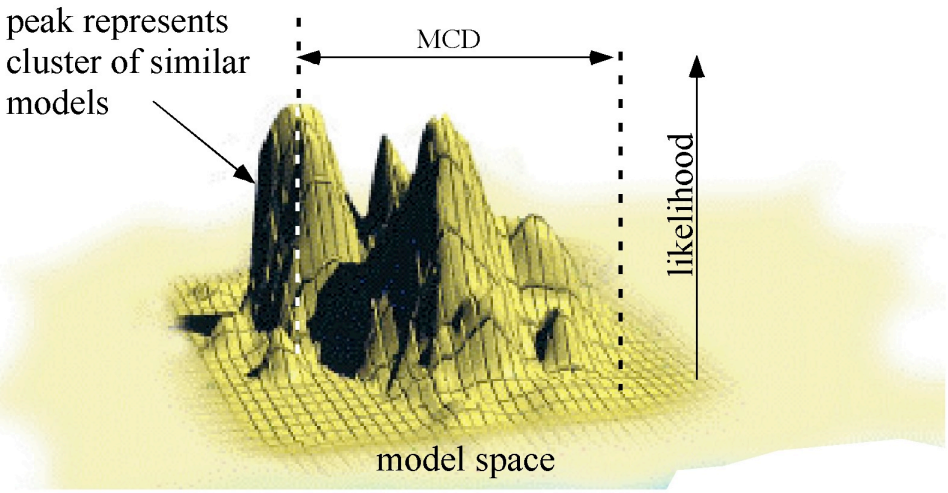


Figure 7. The left graph shows the assumed 1D velocity model (blue curve and the assumed bulk density (red curve) as a function of depth. The middle graph shows the impulse response calculated for the velocity/density model. The right graph shows the corresponding seismogram. The vertical axes of the impulse response and seismogram curves show units of two-way travel-time (seconds).



MCD = model to cluster dissimilarity

Figure 8 schematically shows the sector of model space included in the posterior distribution. Each grid node represents one 3D reservoir model. Each hill represents a cluster of reservoir models having similar properties. Multiple peaks indicate that the MCMC inversion has produced non-unique results. The taller peaks identify regions containing models that are most consistent (i.e., most probable) with the observed seismic data.