SPE 103268

Consistent Downscaling of Seismic Inversions to Cornerpoint Flow Models

Subhash Kalla, SPE, Louisiana State University, Christopher D. White, SPE, Louisiana State University, James G. Gunning, SPE, CSIRO, and Michael E. Glinsky, SPE, BHP-Billiton Petroleum, Inc.

This paper was prepared for presentation at the 2006 SPE Annual Technical Conference and Exhibition held in San Antonio, Texas, U.S.A., 24-27 September 2006.

Abstract

Reservoir simulation models are constructed from sparse well data, dense seismic data, and using geologic concepts to constrain stratigraphy and property variations. Because of the sparseness of well data, stochastically inverted seismic data offer important constraints on reservoir geometry and average properties. Although seismic data are densely distributed, they are uninformative about meter-scale features. Conversely, well data reveal fine-scale features but cannot specify intrawell geometry. To build a consistent model, conceptual stacking and facies models must be constrained by well and seismic data. Stochastic ensembles of geomodels are used to capture variability associated with seismic downscaling, lateral variability and conceptual models. The resulting geomodels must be gridded for flow simulation using methods that describe stratal architecture flexibly and efficiently.

In this paper, geomodels integrate stochastic seismic inversion results (for means and variances of "packages" of meter-scale beds), geologic modeling (for a framework and priors), rock physics (to relate seismic to flow properties), and geostatistics (for spatially correlated variability). These elements are combined in a Bayesian framework. The proposed workflow produces models with plausible bedding geometries, where each geomodel agrees with seismic data to the level consistent with the signal-to-noise ratio of the inversion. An ensemble of subseismic models estimates the means and variances of properties throughout the flow simulation grid.

Grid geometries with possible pinchouts can be simulated using auxiliary variables in a Markov Chain Monte Carlo (MCMC) method. Efficient implementations of this method require a posterior covariance matrix for layer thicknesses. Under assumptions that are not too restrictive, the inverse of the posterior covariance matrix can be approximated as a Toeplitz matrix, which makes the MCMC calculations efficient. The proposed method is validated and examined using two-layer examples. Convergence is demonstrated for a synthetic three-dimensional, 10,000 trace, 10 layer cornerpoint model. Performance is acceptable (305 s on a 2 GHz Pentium-M processor).

The Bayesian framework introduces plausible subseismic features into flow models, whilst avoiding overconstraining to seismic data, well data, or the conceptual geologic model. The methods outlined in this paper for honoring probabilistic constraints on total thickness are general, and need not be confined to thickness data obtained from seismic inversion: any spatially dense estimates of total thickness and its variance can be used, or the truncated geostatistical model could also be used without any dense constraints.

Intoduction

Problem Statement. Because reservoirs are sparsely sampled by well penetrations, seismic survey results provide essential controls for modeling. However, beds thinner than about 1/8 to 1/4 the dominant seismic wavelength cannot be resolved in these surveys.[1,2] For depths of ≈3000 m, the maximum frequency in the signal is typically about 40 Hz and for average velocities of ≈2,000 m/s this translates to best resolutions of about 10 m. Besides the limited resolution, seismic-derived depths and thicknesses are uncertain because of noise in the seismic data and uncertainty in the rock physics models.[3,4] This resolution limit and uncertainties associated with seismic depth and thickness estimates have commonly limited the use of seismic data to either inferring the external geometry or guiding modeling of plausible stratigraphic architectures of reservoirs.[5]

Our objective is to use probabilistic depth and thickness information from the layer–based seismic inversion code DELIVERY[3] as input to a downscaling algorithm operating on a cornerpoint grid that may be coarser than the geomodel. Seismic constraints and priors are modeled on the quasivertical block edges, analogous to seismic traces. Simulation at the edges preserves geometric detail in cornerpoint models.This problem fits inside a larger workflow, where this combination of the

geomodel, well data, and seismic data is referred to as "enforcement," and the associated algorithms comprise the software package known as ENFORCER. The integration of seismic inversion information with spatial correlation requirements at the meso- or layer–package scale, and the associated remapping of seismic geometry to cornerpoint geometry is performed by the DELIVERYMASSAGER.[6] Fuller integration of the geomodel and a flow simulation is a subject of ongoing work.[6,7] These coarse–scale models must be downscaled to the flow model scale, honoring well data such as layer thicknesses, porosity and permeability.[8,9] The downscaling must embrace conceptual geologic models for providing stratigraphic frameworks, especially layer correlation models between sparse conditioning points. Analogous seismic-scale frameworks are used in DELIVERY[3] for constructing prior estimates of layer locations, and are typically constructed using geomodeling software,[10] but quasi-mechanistic depositional modeling[11] or surface-oriented geostatisics algorithms[12] are possible alternatives.

The inputs to the downscaling problem are typically realisations of the seismic inversion coarse–scale model "massaged" to the edges of columns of the cornerpoint grid. These inverted models contain all the requisite coupling between geometry and rock properties which seismic inversion induces, plus the necessary spatial correlation behaviour forced by the massaging algorithm. These coarse scale models provide explicit constraints on the corresponding subgridded models, which are nontrivial to respect using conventional geostatistical algorithms for fine scale heterogeneity.

A characteristic difficulty is that parameters of the fine–scale model such as thickness may have one–sided or mixture distributions ($e.g.$, the mode of layer thickness may be zero, in a cornerpoint model). Because of constraints to be imposed, linear estimation may prove inadequate. For example, if one wishes to ensure consistency both in thickness and in average porosity in a downscaling problem consisting only of vertical gridding refinement, the following equations must be considered at column of gridblock corners:

$$\sum_{k=1}^{K} h_k = H$$

$$\sum_{k=1}^{K} h_k \phi_k = \bar{\Phi} H$$

where K is the number of layers, k indicates a particular layer, ϕ is the porosity, h is a layer thickness, H is the total thickness predicted by seismic, and $\bar{\Phi}$ is the estimated average porosity at the trace scale. If layer porosity and thickness must be jointly estimated, the problem is nonlinear.

In summary, seismic downscaling to well and stratigraphic data on an arbitrary cornerpoint grid is a difficult problem, chiefly on account of the complex constraints, but also because of significant nonlinearities.

Use of Terms. The following conventions are used.
Layers are generally not resolved by seismic data, but can be identified in wells. This terminology is illustrated in **Fig. 1**. Sublayers might exist if some geomodel layers are not resolved in the cornerpoint grid layers. In this paper, well data is used only at the layer scale – log and core data must be upscaled.
Traces are a segment of reservoir whose average properties are constrained by seismic, and will generally contain many layers. Traces correspond to the edges of the cornerpoint gridblocks[13] ($viz.$, `COORD` records).[14] Conditioning data are a type of trace; order, properties, and thickness are specified at conditioning traces.
A *Path* is a sequence in which traces (or layers, or blocks) are visited. We use a quasirandom multigrid path.
Multigrid paths are paths which preferentially visit widely spaced points early.
The *Resolution Matrix* is the inverse of the covariance matrix, and closely related to the Hessian in an optimization problem.

Problem Formulation

Our approach is to combine diverse data elements in prior and likelihood expressions to obtain a posterior probability. The overall posterior distribution is approximated by the posterior obtained by a multigrid sequential simulation passing over all columns or column–blocks of the cornerpoint grid. Each column of blocks is simulated by sampling from a Bayesian posterior distribution conditional on hard data and previously visited columns via the priors, and collocated coarse scale constraints via the likelihood. The prior distribution for each column is determined by solving an ordinary kriging system[15] using observations and previously simulated values. The seismic data are incorporated via a constraint on the sum of the layer thicknesses, which comes from a stochastic seismic inversion. In the proposed approach, layer thicknesses are modeled as truncated Gaussian processes to allow for pinchouts; this model complicates imposition of the seismic sum constraint (Sampling Approach, later). The prior data and thickness constraints are combined in a Bayesian posterior form. Finally, the posterior is sampled using Markov chain Monte Carlo methods with auxiliuary variables.[16]

An efficient approximation to the posterior covariance matrix is crucial to the success of this Bayesian approach. This posterior covariance matrix is required by the sequential simulation algorithm, and encapsulates the compromise between prior information from kriging and total thickness constraints derived from seismic information.

For simplicity, we consider systems with a single thickness constraints. More general constraints are addressed in Appendix A. Numerical methods and sampling methods are discussed in later sections.

Algorithm Outline. Before discussing details, the algorithm framework is presented (**Fig. 2**). First, the untruncated Gaussian surrogate for all conditioning data with $h = 0$ must be simulated; this can be done just once, for all sequential simulation paths and conditioning traces (Appendix B). Then, a multigrid random path for a sequential simulation is generated. At each point on the path, the prior is estimated, and the likelihood is computed using a Gibbs sampler on an auxiliary variable to treat the possibility of zero thicknesses, followed by

a Metropolis-Hastings step to propose a new thickness vector. The chain is iterated to convergence, a sample vector $\mathbf{t}$ is drawn, and the simulation then moves to the next trace in the path. Multiple paths can be used to generate multiple chains, in the same way sequential Gaussian simulations generate multiple realizations.[17]

Prior Mean and Variance by Kriging. The untruncated proxy t is kriged. The proxy t may take on negative values, whereas h is truncated at zero. The probability of $t_k \leq 0$ corresponds to the probability that layer k is absent, locally:

$$P(h_k = 0) = \int_{-\infty}^{0} dP\left(t_k\right) \tag{1}$$

A few assumptions can simplify the kriging solution, and greatly improve efficiency (Numerical Considerations, later).

1. For many block shapes and grid spacings, traces can be approximated as vertical when computing the kriging covariance matrix ($i.e.$, small lateral trace displacement compared to trace spacing). Then the areal separation between the visited trace and each of its neighbors is constant for all layers all trace-neighbor pairs.
2. If in addition the covariance models are the same for all layers, then the covariance matrices will be the same on a layer-by-layer basis as well.
3. Layer thicknesses may be $a\ priori$ uncorrelated vertically at each trace. This may be reasonable, as the lateral thickness variations are likely more informative than the thicknesses of the layers above and below. This assumption seems particularly appropriate for tubidite systems, in which meter-scale beds may correspond to individual depositional events: bed thicknesses then correlates strongly only within beds, with between-bed correlations being weak or even negative if compensatory deposition or scouring were occurring.

If all of these assumptions are reasonable, then (1) the priors for each layer can be computed separately, (2) the kriging matrices are identical for all layers, and therefore only one kriging system needs to be solved at each trace, and (3) the prior variances in each columns are then uniform. The prior means vary layer-by-layer. The tracewise-constant prior variance allows much more efficient solution methods (Numerical Considerations, later). These assumptions need not be imposed: this would make the kriging system(s) more expensive to solve, and the approximation to the posterior covariance will be much more expensive to compute.

The neighbor list is extracted from the list of conditioning data and previously simulated points using a k-d tree[18] with specifications of desired points per quadrant. This search strategy is more efficient than most alternatives, especially on irregular grids. Also, assuming only two-dimenional layer thickness correlation implies that a two-dimensional search suffices, further improving search efficiency. Cokriging or collocated kriging could be used to get prior covariances.[15] Such a result could be integrated well with the seismic data, which provide local correlated estimates of trace-scale properties.[3]

If vertical correlations are included, separate neighbor lists may be required for each of the K_ℓ layers at the trace, or a single list could be used for all layers. While the single list might require solving a larger kriging system, it would only require solving one kriging system for all K layers.

We use a Cholesky factorization to solve the kriging system. The estimated thickness is $\bar{t}_k$. The estimation variances from the kriging system, σ^2_{tk} for the thickness of layer k, will be used in the resolution matrix.

This step in the algorithm supplies prior means $\bar{t}$ and variances σ^2_t for all layers on a given trace.

The Posterior Resolution Matrix. Additional data from the seismic data are combined with the prior to obtain an unscaled posterior probability. The seismic data are incorporated as a constraint on the total thickness, $\bar{H}$, with resolution $\frac{1}{\sigma^2_H}$ obtained from a stochastic inversion using DELIVERY.[3] The stochastic inversion assumes no trace-to-trace inversion and the traces are not coincident with cornerpoint edges in the flow model. Geologically plausible lateral correlations are introduced, and seismic data are kriged to the (possibly nonvertical) cornerpoint edges using methods implemented in DELIVERYMASSAGER.[6,7]

The posterior probability for any thickness vector $\mathbf{t}$ is, from Bayes' rule,

$$\pi\left(\mathbf{t}|H, \mathbf{d}_{\ell k}\right) = \frac{p\left(H|\mathbf{t}, \mathbf{d}_{\ell k}\right) p\left(\mathbf{t}|\mathbf{d}_{\ell k}\right)}{p\left(H|\mathbf{d}_{\ell k}\right)}$$

where $\mathbf{d}_{\ell k}$ is a vector of the all neighboring conditioning or previously simulated traces in layer k in the neighborhood of trace ℓ. The likelihood $\mathcal{L}$ is proportional to the posterior, without normalizing term in the denominator, which does not depend on $\mathbf{t}$. The posterior probability of the model for $\mathbf{t}$ is

$$\pi\left(\mathbf{t}|H, \mathbf{d}_{\ell k}\right) \propto \mathcal{L}\left(H\right) p\left(\mathbf{t}|\mathbf{d}_{\ell k}\right) \tag{2}$$

where $\mathcal{L}\left(H\right) = p\left(H|\mathbf{t}, \mathbf{d}_{\ell k}\right)$ We assume that departures from the prior ($\bar{t}_k$) and updating ($\bar{H}$) data means are normally distributed with standard deviations σ_{tk} and σ_H, respectively. The assumptions apply to departures, not values, and so the resulting posterior probabilities are not assumed to be normal, as will be demonstrated in later examples. The multivariate distribution of $\mathbf{t}$ is

$$p(\mathbf{t}|\mathbf{d}_{\ell k}) = \frac{1}{(2\pi)^{\frac{K}{2}} |\mathbf{C}_p|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{t}-\bar{\mathbf{t}})^T \mathbf{C}_p^{-1} (\mathbf{t}-\bar{\mathbf{t}})\right] \tag{3}$$

where $\mathbf{C}_p$ is the prior or kriging covariance matrix, which is rank K with the kriging variances σ_{tk} along the diagonal.

Similarly, we can express the udpdating constraint H as

$$p(H|\mathbf{t}, \mathbf{d}_{\ell k}) = \sqrt{\frac{1}{2\pi\sigma^2_H}} \exp\left[-\frac{(H-\bar{H})^2}{2\sigma^2_H}\right] \tag{4}$$

where

$$H = \mathbf{t}^T \mathbf{T} \tag{5}$$

and

$$T_k = \begin{cases} 0 & \text{if } t_k < 0 \\ 1 & \text{otherwise} \end{cases} \tag{6}$$

The conditioning on $\mathbf{d}_{\ell k}$ in Eqn. (4) is indirect, due to the conditioning of $\mathbf{t}$ on $\mathbf{d}_{\ell k}$. The product of Eqns. (3, 4) is the posterior. The product can be linearized by taking the logarithm, giving

$$-2\ln\left[\pi\left(\mathbf{t}|H,\mathbf{d}_{\ell k}\right)\right] = \ln\left[\left(2\pi\right)^K\left|\mathbf{C}_p\right|\right] + \ln\left(2\pi\sigma_H^2\right) + (\mathbf{t}-\bar{\mathbf{t}})^T\mathbf{C}_p^{-1}(\mathbf{t}-\bar{\mathbf{t}}) + \frac{(\mathbf{t}^T\mathbf{T}-H)^2}{\sigma_H^2} \tag{7}$$

The Hessian, $\mathbf{G}$ of Eqn. (7) is the desired resolution matrix (which is the inverse of the posterior covariance):

$$\mathbf{G} = \mathbf{C}_p^{-1} + \mathbf{T}\mathbf{T}^T/\sigma_H^2 \tag{8}$$

Because the prior covariance matrix is diagonal, $\mathbf{C}_p^{-1}$ and $\mathbf{G}$ are easy to compute. For $T_k = 1, \forall k$, the Hessian has the form

$$\mathbf{G} = \begin{pmatrix} \frac{1}{\sigma_{t1}^2}+\frac{1}{\sigma_H^2} & \frac{1}{\sigma_H^2} & \cdots & \frac{1}{\sigma_H^2} \\ \frac{1}{\sigma_H^2} & \frac{1}{\sigma_{t2}^2}+\frac{1}{\sigma_H^2} & \cdots & \frac{1}{\sigma_H^2} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{1}{\sigma_H^2} & \frac{1}{\sigma_H^2} & \cdots & \frac{1}{\sigma_{tK}^2}+\frac{1}{\sigma_H^2} \end{pmatrix} \tag{9}$$

If the prior variances σ_{tk} are all equal, $\mathbf{G}$ is Toeplitz,[19] and in fact a particularly simple form, with all super- and subdiagonals equal. Note that the Hessian is constant except for the dependence of $\mathbf{T}$ on $\mathbf{t}$; this is the lurking nonlinearity in this problem.

Prior and Likelihood Distributions in 2D

Important features of higher-dimensional cases are easily visualized for a system with two layers (**Fig. 3**). Simple parameter choices have been made to clarify the explanation. In this example, the prior mean thicknesses sum to greater than the mean trace thicknesses, so the prior center of mass [circles in Fig. 3; Eqn. (3)] lies above the maximum likelihood line [dashed line in Fig. 3; Eqn. (4), for $t_k > 0, \forall k \in 1, K$]. Because $\bar{t}_1$ is large compared to $\bar{H}$, there is substantial prior probability and likelihood – and therefore posterior probability – that t_2 is negative, so that $h_2 = 0$.

In three dimensions, the dashed line in Fig. 3 corresponds to a triangle with vertices on each t-axis at $\bar{H}$; increasing $\bar{H}$ shifts the high-likelihood region away from the origin, but with no change in slope. Tighter seismic constraints will narrow the width of the high-likelihood region.

The assumption of equal prior variances implies the prior has the circular shape shown in Fig. 3; it would be ellipsoidal if prior variance varied. If variances were unequal, and layer means correlated, the prior would be an appropriately rotated ellipsoid. Although such priors could be sampled using methods discussed in this paper, the resolution matrices would be non-Toeplitz and the algorithms much slower.

If no layer kriging data were used and the seismic data were considered exact, any layer thickness pair (t_1, t_2) along the dashed line with 45 degree slope could be used. Conversely, in a sequential simulation not conditioned to seismic, the layer thicknesses would simply be drawn from the prior (Fig. 3).

The sampling problems and nonlinearity in this problem are caused by the bends in the likelihood where the axes are intersected by the contours of the likelihood surface (Fig. 3). This nonlinearity may dominate sampling where the priors admits significant probability of one or more thicknesses being zero (as is the case for layer 2 in Fig. 3). In higher dimensions, many layers may be pinched out at any given trace, and a method to move"around" these corners is needed (Auxiliary Variables to Treat Pinchouts, later).

Sampling Approach

Because the log–posterior surface is quadratic with constraints (Eqn. (8), the most likely $a\ posteriori$ thickness vector could be found by constrained quadratic programming.[20] However, our goal is simulation, not maximum–aposteriori estimation, so we must sample from the posterior. We use a Markov chain Monte Carlo (MCMC) method.

In this section, we focus on simulation at a given trace ℓ. The overall simulation proceeds by visiting all ℓ that are not in the conditioning data set by a specific, random, multigrid path.

Observed Thicknesses of Zero. Some layers may be absent at conditioning points, $h_k = 0$. For these points, we only know that $t_k \leq 0$ at these points, but require a particular value of t_k to use in estimating means at the traces to be simulated. One could simply draw random numbers in the range $[0, P(h_k = 0)]$ and applying an inverse normal transformation, but this decorrelates the variables. Instead, we precondition these data using a Gibbs sampler to preserve the correlation (Appendix B).

Auxiliary Variables to Treat Pinchouts. The posterior distribution has marked bends at the interfaces in parameter space where layers pinch out ($i.e.$, the hyperplanes $t_k = 0$; Fig. 3). Standard MCMC methods based on small jumping proposals will diffuse around such distributions very slowly. It has been shown that introducing auxiliary variables $\mathbf{u}$ can promote mixing, or alteration between states, in several types of difficult MCMC problems with related "configurational stiffness" characteristics.[21] Auxiliary variable methods use an augmented posterior probability:

$$\pi(\mathbf{u},\mathbf{t}) = \pi(\mathbf{t})\pi(\mathbf{u}|\mathbf{t}) \tag{10}$$

where the variables $\mathbf{u}$ are chosen so as to provide useful information for constructing jump proposals in the MCMC algorithm, and the stochastic variables $\mathbf{t}$ are obtained from the augmented-state posterior samples by simple subselection.

The term $\pi(\mathbf{u}|\mathbf{t})\ \left[= \prod_{k=1}^K \pi(u_k|t_k)\right]$ is a conditional probability for the auxiliary variables which may be constructed in any helpful way. In our case, we construct the conditional to

help detect the kinks in the posterior that occur when layers pinch out: one possible choice of a symmetric form is

$$\pi\left(u_k = 1 | t_k\right) = \begin{cases} 1 - \frac{1}{2 + t_k/\sigma_{\pi k}} & \text{if } t_k \geq 0 \\ \frac{1}{2 - t_k/\sigma_{\pi k}} & \text{otherwise} \end{cases} \tag{11}$$

where $\sigma_{\pi k}$ is a univariate approximation to the multivariate posterior covariance,

$$\frac{1}{\sigma_{\pi k}^2} = \frac{1}{\sigma_{tk}^2} + \frac{\kappa}{\sigma_H^2} \tag{12}$$

That is, $\sigma_{\pi k} \approx \sum_{j=1}^{K} G_{kj}$, [Eqn. (9)]. κ is the current number of *active* layers; $\kappa = \sum_{k=1}^{K} T_k \leq K$.

Sampling from the augmented posterior distribution is performed by alternating Gibbs samples for the auxiliary variables with Metropolis–Hastings samples for the thicknesses t_k. The Gibbs sampling scans over the layers. At each layer, a uniform $[0, 1]$ random number is drawn. If the random number is less than $\pi(u_k = 0|t_k)$, $u_k \leftarrow 0$. When the u_k for all K layers have been simulated, we construct a resolution matrix (dependent on $\mathbf{u}$) from which jumping proposals are formed which are well 'tuned' for the current configuration of the system. The auxiliary variables create an adaptively varying proposal kernel which does not break reversibility. The new kernel is used to sample a new thickness vector $\mathbf{t}$ using a Metropolis-Hastings step. The proposal kernel adapts by omitting terms with $u_k = 0$ in the resolution matrix.

The Gibbs sample gives a list of "likely" active layers at the current iterate in $\mathbf{u}$. Let the number of active layers be κ, $\kappa \leq K$. At each trace, a resolution matrix of rank K is constructed before the first iteration and its Cholesky factors are computed. Note that the "adaptive" resolution matrix $\mathbf{G}_\kappa = \mathbf{C}_p^{-1} + \mathbf{u}\mathbf{u}^T/\sigma_H^2$ is used in forming the Cholesky factors underlying the jumping proposal [Eqn. (14), later], but the likelihood evaluated at a point $\mathbf{t}$ requires the strict Hessian $\mathbf{G} = \mathbf{C}_p^{-1} + \mathbf{T}(\mathbf{t})\mathbf{T}(\mathbf{t})^T/\sigma_H^2$. The appropriate resolution and inverse matrices are computationally inexpensive for the simple Toeplitz resolution matrix used in the proposed approach (Numerical Considerations, previously).

Metropolis-Hastings Step. The Cholesky factor multiplied into a κ-long vector of random normal variables $\mathbf{r} \sim [N(0, 1)]$ produces a vector $\Delta\mathbf{t}$ of proposed changes in $\mathbf{t}$,

$$\Delta\mathbf{t} = s\mathbf{L}_{C\pi}\mathbf{r} \tag{13}$$

so that $\Delta\mathbf{t} \sim N(0, s^2\mathbf{G}_\kappa)$, where s is a scalar chosen for sampling efficiency,[16] typically $s^2 = 5.76/K$ for large K. This vector is rank κ, and the changes must be sorted back into $\mathbf{t}$ by referencing $\mathbf{u}$. We can compute the likelihood at the new point $\mathbf{t}' = \mathbf{t} + \Delta\mathbf{t}$, using Eqn. (4). The Metropolis-Hastings transition probability is then[16]

$$\alpha = \min\left(1, \frac{\pi\left(\mathbf{t}'|H, \mathbf{d}_{\ell k}\right) \prod_{k=1}^{K} \pi(u_k|t_k')}{\pi\left(\mathbf{t}|H, \mathbf{d}_{\ell k}\right) \prod_{k=1}^{K} \pi(u_k|t_k)}\right) \tag{14}$$

Equation (14) is similar to the standard Metropolis-Hastings ratio, but has been modified to include the auxiliary variables so that reversibility is maintained. The proposed transition $\Delta\mathbf{t}$ is then accepted with probability α, and the algorithm proceeds to the next Gibbs sample for the auxiliary variables.

Numerical Considerations

The Toeplitz form of the posterior resolution matrix and subsidiary assumptions simplify computations. Because of these simplifications, only two matrix solutions are required per trace: (1) a Cholesky factorization of the kriging matrix (which is dense and not Toeplitz, with rank equal to the number of neighbors used, N_ℓ), and (2) the factorization of the inverse of the Toeplitz resolution matrix (rank K_ℓ and very inexpensive). If the Toeplitz-yielding assumptions were not made, K_ℓ rank-$\sum_{k=1}^{K_\ell} N_{\ell k}$ kriging systems are required at each trace ℓ. Even more prohibitive, the posterior resolution matrix $\mathbf{G}$ would have to be refactored every time any t_k flips from a negative to non-negative state. Because this occurs deep within the sampling method (Sampling Approach, later), this would result in a remarkable loss in efficiency.

To carry out the simulation, we need the Cholesky factor $\mathbf{L}_{C\pi}$ of the posterior covariance matrix, $\mathbf{C}_\pi = \mathbf{G}^{-1}$. With $\mathbf{L}_{C\pi}$, we can generate correlated normal deviates, $\mathbf{\Delta t}$, from uncorrelated random normal input vectors, $\mathbf{r}$, $\Delta\mathbf{t} = \mathbf{L}_{C\pi}\mathbf{r}$ (Metropolis-Hastings Step, earlier).[15] For the special Toeplitz matrices, the factor $\mathbf{L}_{C\pi}$ can be computed from the Cholesky factor of the resolution matrix $\mathbf{G}$. That is, (1) Factor $\mathbf{G}$ to get $\mathbf{L}_G$, (2) invert $\mathbf{L}_G$ by backsubstitution to get $\mathbf{L}_G^{-1}$ (inexpensive because the matrix is triangular), and (3) take the persymmetric transpose of $\mathbf{L}_G^{-1}$. This is the Cholesky factor of $\mathbf{C}_\pi$, $\mathbf{L}_{C\pi}$.

The rank "downdate" from K to $\kappa < K$ is the lower $\kappa \times \kappa$ triangle of $\mathbf{L}_{C\pi}$. Thus, all the required factored correlation matrices $\mathbf{L}_{C\pi\kappa}$, regardless of the number of active layers κ, can be computed from a single factoring and inverse to get $\mathbf{L}_{C\pi}$ and taking the appropriate rank-κ submatrix.

In combination, the efficient factorization method for the posterior rank-K covariarance matrix and determination of $\mathbf{L}_{C\pi\kappa}$ for all possible pinchout combinations makes this algorithm efficient. Precise work estimates for these matrix calculations have not been done, but an upper bound is the work done for a general Toeplitz matrix,[19] inverting the resolution matrix and factoring that inverse to get $\mathbf{L}_{C\pi}$. For that less efficient approach, the inverse of the Toeplitz resolution matrix requires $W \sim K^3$ floating operations (flops), and further work $W \sim K^4$ flops is required for the factoring. In comparison, the proposed method is at worst $W \sim K^3$ for the inverse and all factors, a full order of improvement.

Simulations of Two-Layer Systems

Several two-layer simulations illustrate the behavior of the data integration algorithm. Different combinations of prior and updating data variance are considered, along with perfectly consistent versus slightly contradictory prior means and constraints. Results are summarized in **Table 1**.

Tight Sum Constraint. This case assumes the sum of the layer prior means is equal to the trace mean, but the layer thicknesses are poorly resolved (**Fig. 4**). Because the means are consistent and the constraint variance is relatively small, the simulations tightly cluster around the constraint line, and the posterior means of $\mathbf{t}$ are near their prior means, although the correlation induced by the constraint is marked (covariance column, Table 1). Moreover, many realization have $\mathbf{t}$ near $(4,0)^T$ (which is very unlikely in the prior) because of the relatively tight seismic constraint ($\sigma_t/\sigma_H = 10$). The bend in the posterior caused by the pinchout is clearly seen below $t_2 = 0$ [Fig. 4(a)]. The posterior layer variances are reduced because of the added data in the constraint (eigenvalues, Table 1). The axial (maximum) standard deviation is the same for the posterior as for the (isotropic) prior, but the transverse standard deviation is significantly reduced. The univariate histograms of t are slightly non-Gaussian, and truncation makes the histograms of h depart even more.

Loose Constraint and Prior. As for the previous case, the prior means are taken to be consistent with the seismic constraint. However, the variances of both prior and constraint are higher for this case. The data are therefore more dispersed, and it is much more likely that layer 2 is assigned a zero thickness (**Fig. 5**). As before, although t appears nearly Gaussian in the univariate histograms, h will be truncated to nonnegative values and is thus non-Gaussian, and the bend in the posterior at $t_2 = 0$ is observed.

Sum of Prior Means less than Constraint. A mismatch between the prior layer means and the thickness constraint shifts the axis of the cloud of simulations points above or below the constraint line (**Fig. 6**). In this case, both layers thicknesses are increased from their priors to better match the seismic constraint. For the moderate standard deviation and prior means much greater than zero, few truncations occur and the posteriors are nearly Gaussian. For this nearly multi-Gaussian case, the constraint has transformed the isotropic, uncorrelated prior thicknesses (Fig. 3) to a strongly correlated, more compact posterior. Because the prior and constraint variances are equal, the mean of the scatter cloud is shifted roughly one-half the distance from the prior toward the constraint, as would be expected (Table 1).[16]

Convergence. MCMC methods may converge too slowly to be practical, or may have multiple modes such that multiple chains and/or methods to switch between modes are needed. In numerical experiments undertaken so far, these potential problems do not appear to be too severe in this algorithm.

Convergence is critiqued by examining posterior distribution statistics over many iterations.[16] For a variety of cases examined, the means converge in no more than ≈ 1000 iterations, and the variances stabilize in no more than ≈ 2500 iterations. That is, some 2500 iterations are needed for the chain to begin sampling the posterior reliably; this is referred to as the "burn–in;" samples prior to burn in are discarded before the chain is used to simulate the posterior. This number of iterations, while large, is not prohibitive if the proposal method is computationally inexpensive (Numerical Considerations, previously) and the acceptance rate is not too small. For a realistic 3D synthetic problem, the proposed method attains a sampling rate of almost 200,000 iterations per second and an acceptance rate averaging ≈ 0.4, which makes such long burn-in requirements manageable (Synthetic 3D Cases, later).

Chains started in widely dispersed parts of $\mathbf{t}$-space converge to the same posterior (**Fig. 7**). This was expected, based on the relatively simple form of the posterior resolution matrix, $\mathbf{G}$. The early behavior depends on the starting point [Fig. 7(a)]: chains that move in from the flanks of the constraint (transverse paths) take large, efficient steps; those moving along the axis zig-zag and advance more slowly. The latter is the classic behavior of movement along a trough in a minimization problem where the eigenvalues of the Hessian differ markedly (Table 1). After many iterations, all chains are sampling the same region [Fig. 7(b)], and the post-burn-in chains are statistically indistinguishable.

Recapitulation. The simple two-dimensional examples indicate the algorithm is reproducing expected results in limiting cases. Extensions for multiple facies and non-Toeplitz resolution matrices are addressed in the Discussion and expanded in Appendix A.

Synthetic 3D Cases

A synthetic data 3D data set is used to test and illustrate the MCMC simulation method. Prior (range of semivariogram, R) and updating data (trends in $\bar{H}$ and σ_H) parameters are varied to illustrate behavior, and algorithm performance is discussed.

For all cases, $I \times J \times K$ is $100 \times 100 \times 10$ and the $x-y$ extent is 1000×1000 m. The framework for the reference model was created by randomly placing objects with scaled bi-Gaussian thickness variations in x and y; for the 1 km areal grid, an isotropic standard deviation, $\sigma = 500$ m, was used to compute layer thickness with $h(x,y) = h_{max} \exp\left[\frac{(x-\bar{x})^2+(y-\bar{y})^2}{\sigma^2}\right]$. This object-based method with Gaussian thickness variations is not the same as a Gaussian covariance process. The object models are used only to create conditioning data. Twenty-five traces were used in cases discussed in this section; the algorithm has also been used with no conditioning traces and with up to 200 conditioning traces.

Illustrative Cases. Four different cases show features of the data integration method (**Fig. 8**). With short ranges, termination is much more common, although the average layer thickness is similar to the longer range [Figs. 8(a,b)]. There is little noise, unlike what is commonly observed in Gaussian processeses; the layer thicknesses vary smoothly and plausibly, and near-zero thicknesses do not appear in isolated areas; this results from the truncation rules and the smooth Gaussian variogram. The pinchout pattern is clearer in the longer-range case (b). In particular, the light yellow layer near the base and the dark blue layer in the middle appear to taper and pinch out smoothly; this behavior is more characteristic of object models than most

covariance-based simulations.

Seismic data may imply a thickness trend [Fig. 8(c)]. The seismic trend will be reproduced in the simulation, with a precision conditioned on the inferred seismic thickness variance, σ_H. If the seismic variance is higher for smaller mean thickness, low thicknesses fluctuate more, as can be seen by comparing the left front edges of Figs. 8(c) and (d). For the low variance case (c), the edge panel is of nearly uniform thickness; the nonuniform variance case (d) has much greater fluctuation on the left edge.

Although based on a synthetic case, these results indicate that the proposed method can reproduce complex pinchout layering and plausible seismic trends. The number of pinchouts can be quite large in complex cornerpoint grids; 30,608 of 100,000 trace segement are zero-thickness in one of the example cases [Fig. 8(c)]. The complex pinchout structure is obtained even though the conditioning data are not especially dense [Fig. 8(d)].

Performance. For adequate performance, an MCMC simulation should converge to its target distribution in as few steps as possible. A large step size helps explore the posterior in few steps. On the other hand, large steps are more likely to rejected, "wasting" computations on a sample that is not retained. The step size is usually adjusted indirectly, by scaling the posterior covariance (which is used to generate steps; Metropolis-Hasting step, earlier). For the system examined, the covariance is not scaled; this gives a step size of the order of the square root of the smallest diagonal element in the posterior covariance matrix. In high-dimensional problems, it may be more appropriate to use $\tilde{\mathbf{C}}_\pi = \frac{5.67}{K}\mathbf{C}_\pi$ to ensure adequate acceptance rates.[16] Although the unscaled covariance yields larger steps for $K = 10$, the test cases had acceptance rates of 30 to 40 percent. This step size and acceptance rate appear to yield good convergence, thorough exploration of the posterior, and smooth posterior samples (where they should be smooth). The best choice of scaling is problem-dependent.

The computational cost of a single simulation [for the case of Fig. 8(a)] is examined component-by-component in Table 2. Several features are striking. First, 97.98 percent of the work is done in the deepest part of the sampling loop, which requires random number draws, extractions of submatrices, and multiplication of random normal vectors by lower triangular matrices (the Cholesky factor of the posterior covariance matrix, $\mathbf{L}_{C\pi\kappa}$). None of these operations is particularly expensive, but a total of 5×10^7 iterations were performed for this case ($\approx 164,000$ samples accepted per second). Because the kriging system is solved only once per trace – and is two-dimensional, with an efficient k-d neighbor search – the associated work is small, about 1.95 percent. The Toeplitz manipulations are practically cost-free, only about 0.07 percent of the total work. Finally, the overall cost of about 5 minutes on a laptop computer (for 10^5 unknowns) does not seem prohibitive.

Because it is a tracewise sequential algorithm, this MCMC method scales linearly in the number of block edges, or traces. Thus, a model with 10^6 traces and 10 layers should require approximately 8.5 hrs if attempted on a single Pentium-M processor with adequate memory: not too alarming, for a model with 10^7 unknowns. Work scales approximately with the third power of layer count (Numerical Considerations, previously), so that a model with 20 layers would take 8 times as long as the 10-layer model used in the illustrations. A crude estimate is that a $1000 \times 1000 \times 20$ model would require about 68 hours on a single 2 GHz processor (neglecting degradation to use of virtual memory).

Discussion

Sequential Methods. The most difficult aspect of these nonlinear downscaling problems is the issue of whether the overall system posterior distribution can be safely factored into the product of conditional distributions implied by the sequential pass over the columns of gridblocks. This factorisation requires computing both analytical marginal distributions (integrating over "unvisited" sites), and conditional distributions dependent only on "visited" sites. This requirement is usually met only by exponential family distribution functions. The posterior in our problem does not strictly satisfy these requirements. Nonetheless, the approximations we make can doubtless be improved by blockwise sequential schemes, though a block approach increases the dimensionality of the MCMC sampling subproblem, and the configurational complexity of handling more pinchout transitions.

Notwithstanding these concerns, we have demonstrated that using auxiliary variables greatly facilitates effective sampling of a complicated high–dimensional posterior distribution that arises in the downscaling problem we address. Similar difficulties will arise in any more or less rigorous recasting of the problem, so the technique we demonstrate should be widely applicable. Possible extensions are use of mixture–independence samplers[22] that take advantage of the piecewise quadratic form of the log–posterior function, and generalisation to multiple correlated variables in the model and associated likelihood.

Related Methods. As discussed in Simulation of Two-Layer systems, if no layers are likely to be absent, the posterior distribution remains multi-Gaussian, and simulation and estimation methods are linear. In this case, the proposed method is a variant of collocated cokriging, where the collocated data are a sum rather than a constraint on a single thickness.[15] The proposed methods are needed only when there is substantial likelihood of layers terminating laterally, in which case untruncated Gaussian models will fail.

Previous work on reservoir characterization with truncated Gaussian fields has focused on categorical simulations.[23,24] In contrast, the proposed method combines aspects of categorical and continuous simulations. The condition $t_k \leq 0$ on the thickness proxy is equivalent to setting an indicator for layer occurrence to zero. However, in the categorical case all $t_k > 0$ would be identical (for a binary case), whereas we use values $t_k > 0$ to model the continuous variable h_k. This hybrid approach could be applied without constraints, yielding sequential truncated Gaussian simulations of thickness; this corresponds closely to the cases with high σ_H presented above, and the resulting images would be similar.

The examples presented all use truncated normal distributions. The positive portion of the distribution is treated as normal, without any transformation. If the nonzero thickness histogram is not nearly a truncated normal, a normal score score transform could be used to make the positive part of the t distribution univariate normal. The transform would be applied over the cumulative probability range $[\bar{P}(h_k = 0), 1]$, where $\bar{P}(h_k = 0)$ is the areal proportion that zero thickness occurs in layer k – over all of layer k, not at a single trace..

Cornerpoint Grids. The MCMC simulation is over the block edges, or traces. This is different from many geostatistical modeling approaches, which are commonly block-centered. However, geometry – especially pinchouts or discontinuities at faults – can be modeled much more accurately using cornerpoints. The porosity and other rock properties should be simulated or estimated at the same point, because these properties are generally correlated through the rock physics model and seismic response. Even for cornerpoint grids, reservoir simulators use block centered values for rock properties such as porosity. The trace properties must be averaged appropriately to the block center. A simple mean is probably adequate for thickness and porosity-thickness. However, the permeability must be upscaled more carefully, especially for nonrectangular blocks; a good method would be to integrate the Jacobian over the half-block domains.[25] Even for uniform permeability, the Jacobian integration correctly provides face- and direction-dependent transmissibilities for a nonrectangular grid. The method could also be used to perform approximate upscaling for sublayer heterogeneities, and compute more accurate pore and bulk volumes.

Extensions. Three extensions being pursued. First, several distinct facies will be subjected to separate seismic thickness constraints. This will permit conditioning on net-to-gross as well as gross thickness. In addition, the blockiness of the constraints should allow decomposition of the multifacies problems into a series of single facies problems (Appendix A). Such block structure would also allow different facies to have distinct correlation structures and means, which is desirable. Second, weighted average constraints, of the form $\sum_{k=1}^{K} h_k \phi_k = \bar{\Phi} H$, are being investigated. Such constraints are nonlinear. Several linearization strategies are possible, including simply lagging less-variable parameters or using their mean. More general scale linkages have been implemented using Markov random fields.[26] Finally, block methods or other approaches will be considered to address difficulties with the computation of marginal distributions in non-Gaussian sequential simulation (Sequential Methods, earlier).

Conclusions

Stochastic seismic inversion computations can be integrated with a truncated Gaussian geostatistical model for layer thicknesses using a Markov chain Monte Carlo method. Truncation makes the problem nonlinear, which necessitates the introduction of auxiliary variables and a mixed Gibbs-Metropolis-Hastings sampling procedure. Under reasonable assumptions, the posterior resolution matrix is a special form of Toeplitz matrix; the special form can be exploited to make MCMC sample proposals much more efficient to evaluate. Proposal efficiency is critical to the usefulness of the method, because many thousand proposals must be evaluated at each trace for a single cornerpoint grid realization. The ability of the method to reproduce limiting case results and correctly model truncations is verifed by examining algorithm behavior in two dimensions. A synthetic 3D case demonstrates that the procedure is acceptably fast. Although many issues remain – especialy implementation of more complex constraints and integration with fine-scale geomodels – the proposed method appears to offer a foundation for further development.

Acknowledgments

BHP-Billiton funded this research by research agreements and an unrestricted gift to Louisiana State University, and research agreements with CSIRO. The authors are grateful for this support, and for permission to publish this paper.

Nomenclature

Roman Symbols

$\mathbf{C}_p$	prior covariance matrix based on kriging, m^2
$\mathbf{G}$	posterior resolution matrix or Hessian, m^{-2}
h	nonnegative layer thickness, m
H	total thickness at trace, m
$\mathcal{L}$	likelihood
p	probability density
P	probability
r	random number
R_x	covariance range parameter in direction x, m
s	scaling factor
t	Gaussian proxy for h, may be negative, m
u	auxiliary variable correlated to layer state
T	$T_k = \frac{1}{2}(\mathrm{sgn}(t_k) + 1)$
W	computational work, flops
x, y, z	coordinates, m
X, Y, Z	grid extents, m

Greek Symbols

α	Metropolis-Hastings transition probability
Δ	separation vector for variogram models, m
ϕ	layer porosity
κ	number of layers at a trace with $t_k > 0$
$\bar{\Phi}$	trace average porosity
π	posterior
σ^2	variance

Indices and Special Subscripts

D	number of nonzero conditioning data
f, F	facies index and count
i, j, k	indices in x, y and z directions
I, J, K	maximum values of i, j, and k indices
ℓ, m	indices over traces and layer-wise trace segments, respectively
L, M	total number of traces and layer-wise trace segments, respectively
m	mudstone
p	prior
s	sandstone
λ, Λ	zero thickness data index and count

Diacritical Marks

$\bar{\cdot}$	mean
$\cdot'$	proposed point, may become new point

References

1. Dobrin, M.B. and Savit, C.H.: *Introduction to Geophysical Prospecting*, McGraw-Hill Companies, New York, fourth edition (1988).
2. Widess, M.B.: "How thin is a thin bed?" *Geophysics* (December 1973) **38**, 1176.
3. Gunning, J.G. and Glinsky, M.E.: "Delivery: an open-sourced Bayesian seismic inversion program," *Computers and Geosciences* (2004) **30**.
4. Gunning, J.G. and Glinsky, M.E.: "Wavelet extractor: a Bayesian well-tie wavelet derivation program," *Computers and Geosciences* (2006) **32**.
5. Deutsch, C.V., Srinivasan, S. and Mo, Y.: "Geostatistical reservoir modeling accounting for precision and scale of seismic data," *SPE Annual Technical Conference and Exhibition*, Dallas, Texas (1996) .
6. Gunning, J.G. and Glinsky, M.E.: "*DeliveryMassager*: a tool for propagating seismic inversion information into reservoir models," (2006), submitted, *Computers and Geosciences*.
7. Glinsky, M. *et al.*: "Integration of uncertain subsurface information into multiple reservoir simulation models," *The Leading Edge* (2005) **24**, 990.
8. Behrens, R.A., MacLeod, M.K. and Tran, T.T.: "Incorporating seismic attribute maps in 3D reservoir models," *SPEREE* (April 1998) .
9. Doyen, P.M., Psaila, D.E., den Boer, L.D. and Jans, D.: "Reconciling data at seismic and well log scales in 3-D earth modeling," *SPE Annual Technical Conference and Exhibition*, San Antonio, Texas (October 1997) .
10. Schlumberger Technology Co.: *Petrel Workflow Tools, Introduction Course V.2004* (March 2005).
11. Merriam, D.F. and Davis, J.C. (eds.): *Sedsim in hydrocarbon exploration*, Kluwer Academic, Plenum Publishers, New York (2001) 71–97.
12. Pyrcz, M.J.: *The Integration of Geologic Information into Geostatistical Models*, Ph.D. thesis, University of Alberta (2004), 250 pages.
13. Ponting, D.K.: "Corner Point Grid Geometry in Reservoir Simulation," *Proc., First European Conference Math. Oil Recovery*, Cambridge (1989) .
14. Schlumberger Technology Co., Oxfordshire: *Eclipse 100 97A Reference Manual* (2004).
15. Goovaerts, P.: *Geostatistics for Natural Resources Evaluation*, Applied Geostatistics Series, Oxford University Press, Oxford (1997).
16. Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B.: *Bayesian Data Analysis*, CRC Press, Boca Raton, second edition (2003).
17. Deutsch, C.V. and Journel, A.G.: *GLSIB: Geostatistical Software Library and User's Guide*, Oxford University Press, Oxford, second edition (1998).
18. Bentley, J.L.: "Multidimensional binary search trees used for associative searching," *Commun. ACM* (September 1975) **18**.
19. Golub, G.H. and van Loen, C.F.: *Matrix Computations*, The John Hopkins University Press, Baltimore, third edition (1996).
20. Nocedal, J. and Wright, S.J.: *Numerical Optimization*, Springer Science+Business Media, Inc., New York (1999).
21. Higdon, D.M.: "Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications," *Journal of the American Statistical Association* (1998) **93**, 585.
22. Gilks, W., Richardson, S. and Spiegelhalter, D.: *Markov Chain Monte Carlo in Practice*, Chapman and Hall (1996).
23. Xu, W. and Journel, A.: "GTSIM: Gaussian Truncated Simulations of Reservoir Units in a W. Texas Carbonate Field," (1993), unsolicited manuscript, SPE paper no. 27412.
24. Matheron, G., Beucher, H., de Fouquet, C. and Galli, A.: "Conditional simulation of the geometry of fluvio-deltaic reservoirs," *SPE Annual Technical Conference and Exhibition*, Dallas, Texas (September 27-30, 1987) SPE paper no. 16753.
25. Peaceman, D.W.: "Calculation of Transmissibilities of Gridblocks Defined by Arbitrary Cornerpoint Geometry," (1993), unsolicited manuscript, SPE paper no. 37306.
26. Lee, S.H., Malallah, A., Datta-Gupta, A. and Higdon, D.M.: "Multiscale Data Integration Using Markov Random Fields," *SPE Reservoir Evaluation and Engineering* (April 2002) , 68.
27. Willis, B.J. and White, C.D.: "Quantitative outcrop data for flow simulation," *Journal of Sedimentary Research* (2000) **70**, 788.

Appendix A: Possible Extensions

Formulations for cases in which the prior variances from the kriging step are not equal for all layers *or* the thickness constraints are not applied to all layers as a group are addressed in this appendix. This situation arises if (1) the distances between each trace and its neighbors are not nearly constant ($i.e.$, steeply inclined cornerpoint grid edges), (2) prior correlation between layer thicknesses at a trace (3D kriging), (3) distinct correlation ranges in different layers, or (4) different groups of layers have distinct sum constraints.

General Case. Even if all the previous complications are in force, the sampling method outlined in the main text of the paper can be applied. However, the kriging systems may be more expensive to solve, because they are of higher rank or differ from layer to layer; however, they will still need to be solved only once at every trace. The MCMC step, however, is more problematic. The non-Toeplitz matrices will be more expensive to factor, and inverses of different rank must be recomputed as the auxiliary vector **u** changes with sampling. This will add significantly to the cost of the method.

Facies with Distinct Covariances. If multiple facies are present, the variogram model may differ from facies to facies. In addition, the seismic constraint may differ. For example, the thickness of each facies, f, of the total number of facies, F, might be informed by seismic data, $H_f \sim N\left(\bar{H}_f, \sigma_{Hf}\right)$. In addition, there may be constraints over sums of sums. For example, if facies s is sandstone and m is shale or mudstone, a constraint on net thickness applies to s only but a gross thickness constraint applies to $H_s + H_m$. Also, because each facies may be differently correlated, the prior variance in each facies could be different. Here, 3D effects are neglected, so the prior uncertainty for a given facies, σ_{hf}, is constant. Finally, note that stratigraphic order is irrelevant in such a system. We are free to group all layers of each facies together. Such a system will yield a block Toeplitz resolution matrix (Multiple Facies with More Complex Sum Constraints, later). If the constraints are only on facies thicknesses (not sums of facies thicknesses), at each trace each facies can be simulated using a distinct chain, because the overall trace resolution matrix is composed of Toeplitz blocks along the diagonal only. Although there are F chains, each block f will be smaller, and for the same total layer count K the simple multifacies system outlined here could be faster to simulate than the single facies case.

More Complex Sum Constraints. On the other hand, if the constraints are on sums of facies thicknesses (sums of sums), the problem might be approached with hierarchical sampling, bere addressing a single trace ℓ. For example, with a 4 layer system, with layers 1 and 3 sandstone, s and 2 and 4 mudstone, m. Recalling that we are free to permute layer order for convenience, the posterior resolution matrix would be (with 3 permuted to 2, and 2 to 3):

$$\mathbf{G} = \begin{pmatrix} \frac{1}{\sigma^2_{ts1}} + \frac{1}{\sigma^2_{Hs}} & \frac{1}{\sigma^2_{Hs}} & 0 & 0 \\ \frac{1}{\sigma^2_{Hs}} & \frac{1}{\sigma^2_{ts3}} + \frac{1}{\sigma^2_{Hs}} & 0 & 0 \\ \frac{1}{\sigma^2_H} & \frac{1}{\sigma^2_H} & \frac{1}{\sigma^2_{tm2}} + \frac{1}{\sigma^2_H} & \frac{1}{\sigma^2_H} \\ \frac{1}{\sigma^2_H} & \frac{1}{\sigma^2_H} & \frac{1}{\sigma^2_H} & \frac{1}{\sigma^2_{tm4}} + \frac{1}{\sigma^2_H} \end{pmatrix} \tag{A-1}$$

Although a block approach offers appealing efficiency, correct reproduction of the marginal distributions is not assured. Further investigation is needed.

Appendix B: Zero Thickness Conditioning Data

In this paper, the untruncated Gaussian proxy **t** is kriged, not the actual thickness **h**. At simulated traces, **t** is computed and stored, and only converted to **h** for output. Conditioning data present more of a challenge. If we observe some layer k on trace ℓ has $h_{\ell k} = 0$, the value of $t_{\ell k}$ is indeterminate; we only know $t_{\ell k} \leq 0$. The conditioning data might be decorrelated if we used a simple but reasonable draw such as

$$t_k = N^{-1}\left(r; \bar{t}_k, \sigma_{tk}\right), r \sim U[0, P(h_k = 0)] \tag{B-1}$$

where $P(h_k = 0)$ is given by Eqn. (1). Instead, we model the correlation as follows, with a loop over all layers.

1. Find all zero conditioning data in this layer, k; the list of the locations of zero data is indexed over $\lambda \in [0, \Lambda_k]$.The positive conditioning data in layer k are indexed by $d \in [0, D_k]$.
2. Iinitialize all Λ_k zeroes with random draws, using Eqn. (B-1).
3. Visit each point λ, forming a kriging system of size $D_k + \Lambda_k - 1$, that is, composed of all points in this layer except the current point. In the first iteration, the kriging weights are stored for reuse. Compute the mean and variance, and draw $r \sim U[0, P(h_k = 0)]$. $P(h_k = 0)$ is computed using the new mean and standard deviation of t_k. The new simulated value t_k is the inverse of $N(\bar{t}_k, \sigma_{tk})$ at cumulative probablity r.
4. Generate a chain and store.
5. Repeat $\forall k \in 1, K$

The stored chains can be used at the beginning in later simulations of layer thickness. Before simulating any new points, sets of the zero-thickness conditioning data are drawn from the stored chain.

Table 1: Parameters and results for 2-layer simulation

	Prior			Constraint		Posterior					
Case	$\bar{t}_1$	$\bar{t}_2$	σ_t	$\bar{H}$	σ_H	$\bar{t}_1$	$\bar{t}_2$	Covariance of t	Axial and transverse[a] σ_t	$\bar{H}$	σ_H
Tight	3.0	1.0	1.0	4.0	0.1	2.86	1.11	$\begin{pmatrix} 0.46 & -0.50 \\ -0.50 & 0.59 \end{pmatrix}$	$\begin{matrix} 1.01 \\ 0.14 \end{matrix}$	4.00	0.10
Loose	3.0	1.0	1.0	4.0	0.5	2.97	0.97	$\begin{pmatrix} 0.53 & -0.46 \\ -0.46 & 0.72 \end{pmatrix}$	$\begin{matrix} 1.03 \\ 0.44 \end{matrix}$	4.00	0.49
$\mathbf{T}^T\mathbf{t} < \bar{H}$	3.0	1.0	0.5	6.0	0.5	3.65	1.66	$\begin{pmatrix} 0.16 & -0.08 \\ -0.08 & 0.16 \end{pmatrix}$	$\begin{matrix} 0.49 \\ 0.28 \end{matrix}$	5.31	0.41

[a]These are the square roots of the largest and smallest eigenvalues, respectively, of the posterior covariance matrix.

Table 2: Performance summary for the 3D example (one complete simulation)[a]

Process	Work in seconds[b]
Kriging work	5.95
Toeplitz solver work	0.22
Total overhead all traces	6.17
Samples, 5000 per trace, all traces	299.20
Cost of example simulation, excluding io	305.37

[a] Model size, $100 \times 100 \times 10$; 5000 samples per trace

[b] Using a 2 GHz Pentium-M (laptop) processor with 1 GB of RAM.

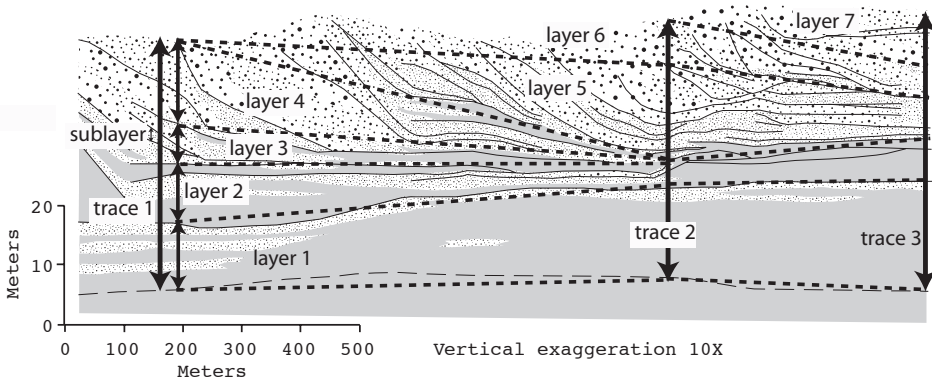


Figure 1: A trace is a line with composite properties informed by seismic data. It may comprise many layers. Sublayers are not modeled in this paper. This image is an interpreted outcrop data set.[27]

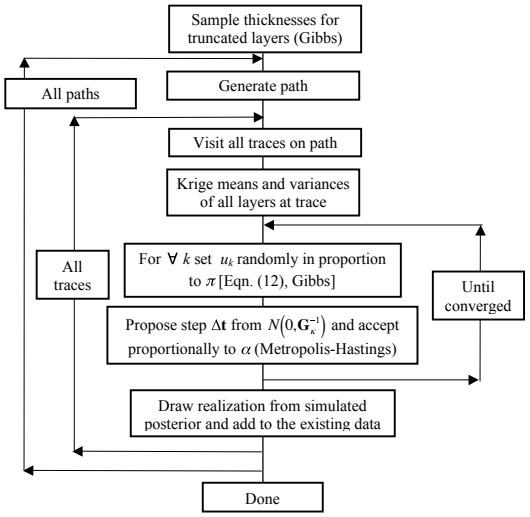


Figure 2: Flow chart for sequential simulation using Markov Chain Monte Carlo.

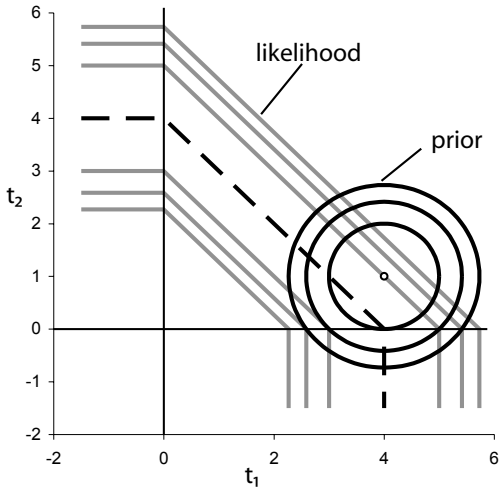
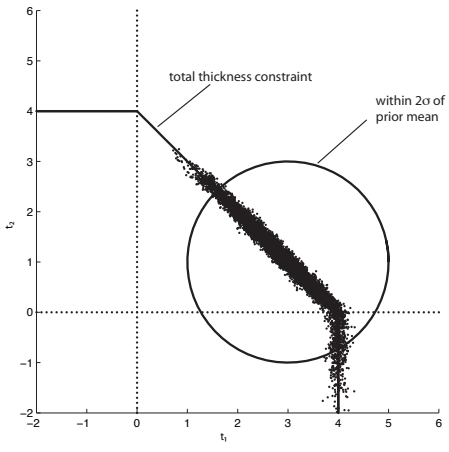
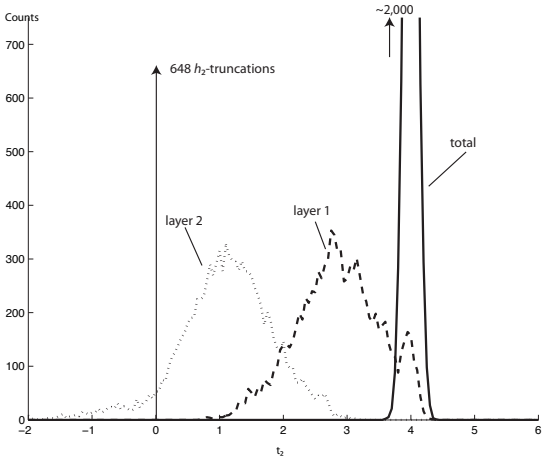


Figure 3: Contours of minus log likelihood and prior distributions for a 2-layer system, with $\bar{H} = 4, \sigma_H = 1, \bar{\mathbf{t}} = (4,1)$, and $\sigma_t = 1$. Contours are in increments of 1, with values of zero along exactly honoring the thickness sum (dashed line) and where $\mathbf{t} = \bar{\mathbf{t}}$ (small circle). Consistent units.

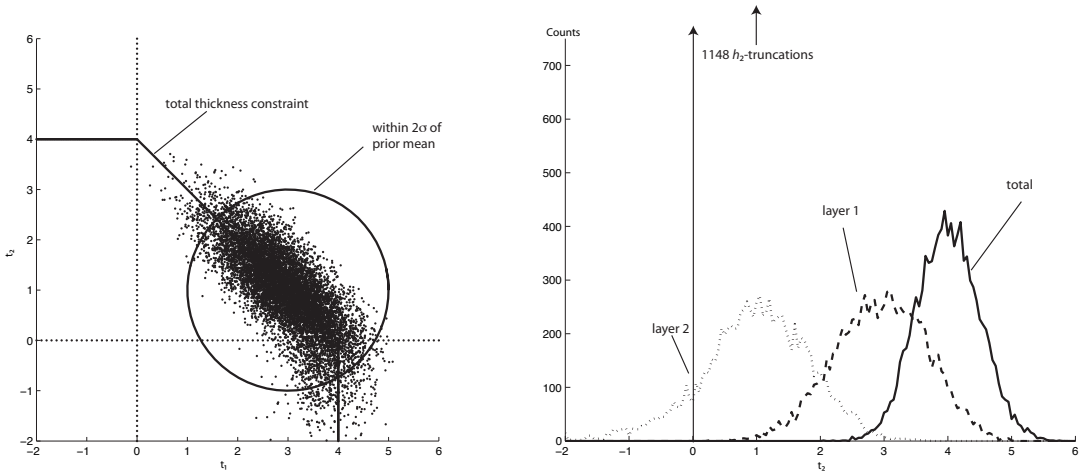


(a) Scattergram, $N = 8000$



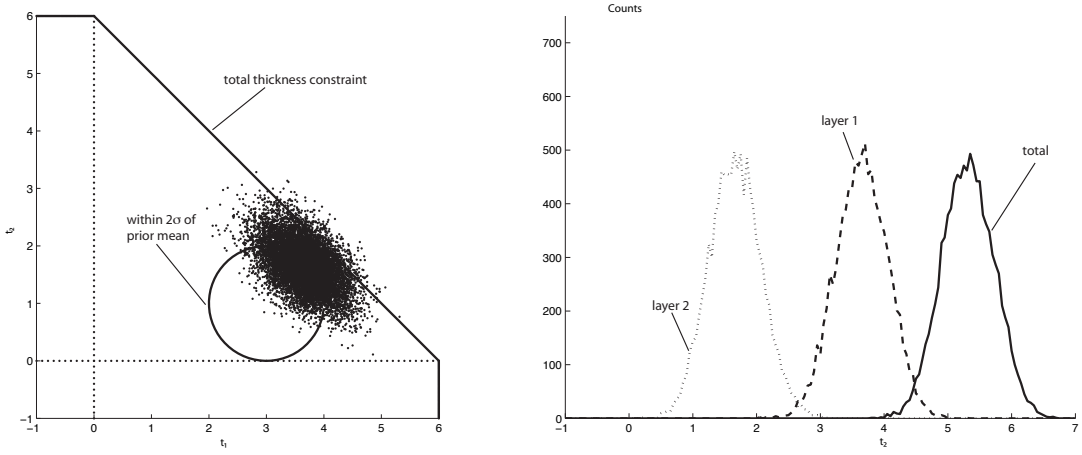
(b) Histogram, bin size $\Delta t = 0.05$

Figure 4: Simulation results for a two layer case with inaccurate layer thickness but total thickness tightly constrained. $\bar{H} = 4$, $\bar{\mathbf{t}} = (3,1)^T$, $\sigma_H = 0.1$, and $\sigma_t = 1$; consistent units.



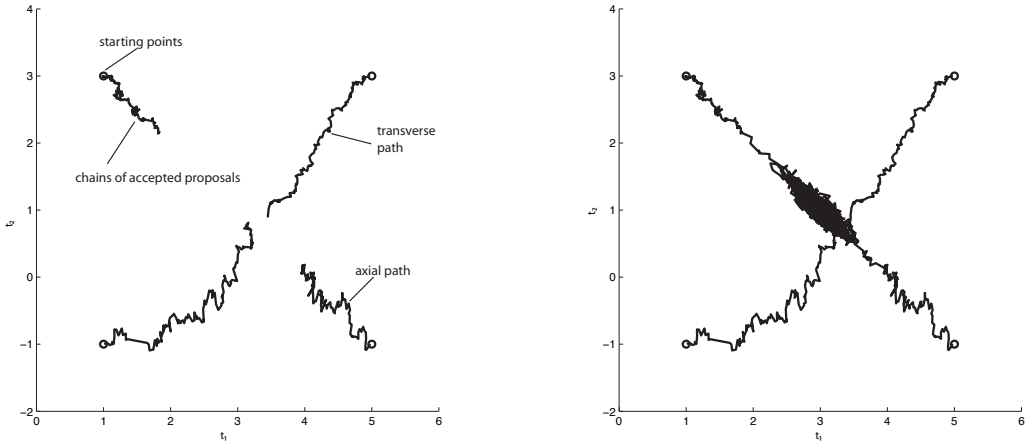
(a) Scattergram, $N = 8000$ (b) Histogram, bin size $\Delta t = 0.05$

Figure 5: Simulation results for a two layer case with inaccurate layer and total thicknesses. $\bar{H} = 4$, $\bar{\mathbf{t}} = (3, 1)^T$, $\sigma_H = 0.5$, and $\sigma_t = 1$; consistent units.



(a) Scattergram, $N = 8000$ (b) Histogram, bin size $\Delta h = 0.05$

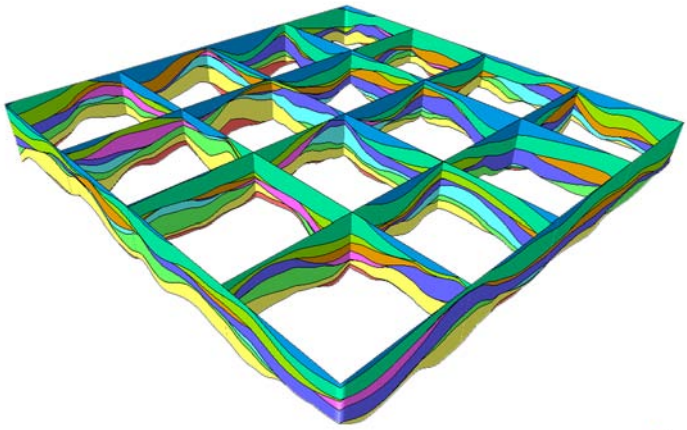
Figure 6: Simulation results for a two layer case with prior sum less than the sum constraint. $\bar{H} = 6$, $\bar{\mathbf{t}} = (3, 1)^T$, $\sigma_H = 0.5$, and $\sigma_t = 0.5$; consistent units.



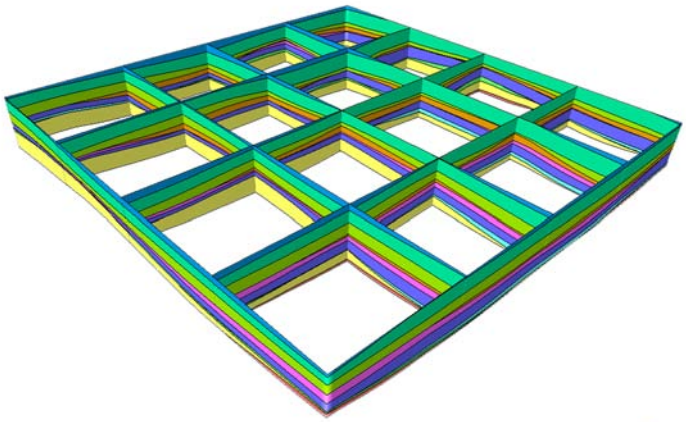
(a) First 150 samples of four chains

(b) Converging chains with 10,000 samples per chain (excluding burn-in)

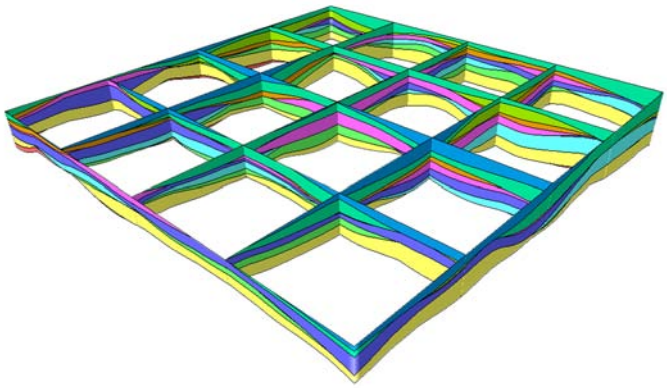
Figure 7: Four Markov chains starting from diverse points tend to migrate toward the most likely region. (a) Convergence is slower for points that must move along the axis to reach the area of the mode. (b) Results are practically identical for long chains, because the posterior is unimodal. The prior and constraint data are the same as in Fig. 4.



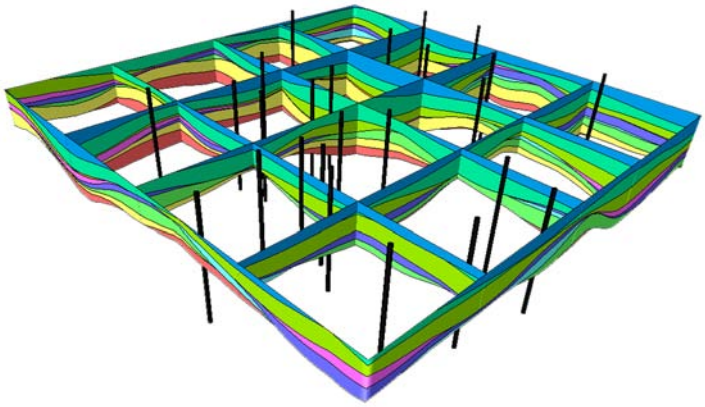
(a) Short range, $R = 200$



(b) Long range, $R = 750$



(c) Seismic thickness trend, $\bar{H} = 7 + \frac{13x}{X}$m, $R = 350$; $x = 0$ is on the left front



(d) Noise varies, $\sigma_H = 5 - \frac{3x}{X}$; R and $\bar{H}$ as in (c); $x = 0$ is on the left front

Figure 8: Simulations on $100 \times 100 \times 10$ cornerpoint grids, areal extent is $X = Y = 1000$m. Lateral extent is 1 km in x and y, and 25 conditioning traces are used. Unless otherwise noted, $\bar{H} = 20$ and $\sigma_H = 2$. All realizations use a Gaussian semivariogram with $R_x = R_y = R$, $\gamma(\Delta) = 1 - \exp\left[-\left(||\Delta||/R\right)^2\right]$, m^2. All models flattened on the topmost surface. Range, thickness, and standard deviation are in m. $5\times$ vertical exageration for all figures. Black lines in (d) are conditioning traces.