Bayesian approaches to CSEM: Resolution, Inference, and Optimization

James Gunning
(CSIRO Petroleum, Bayview Ave., Clayton, Victoria, Australia 3168
E-mail: James.Gunning@csiro.au

Summary

CSEM inverse problems are plagued by issued related to poor scaling and strong nonlinearity. Poor scaling results from the absorption of signal energy in shallow rocks, and nonlinearity is a result of a large dynamic range of electrical resistivities in the subsurface; e.g., a factor of 10^4 from seawater to some basalts.

Poor scaling means the inverse problems must be stabilized. Our preferred methods are Bayesian prior distributions on rock resistivities assembled from regional rock physics. Prior beliefs for a volume of rock (e.g. layer) used in the model should be a mixture distribution driven by depositional knowledge, rock and fluid classes, and scaling effects. Vertical correlations between layers with unknown "correlation lengths" in the prior can be used to derive smooth inverted models using Empirical Bayes methods. This provides a more satisfactory method of estimating smooth models than the "discrepancy principle" used in non-Bayesian methods. In this framework, smoothing parameters are naturally estimated at the maximum aposteriori (MAP) point of an joint model for rock-properties, correlation lengths, and possibly effective-noise parameters.

An alternative route to "simple" low resolution models is performing model-selection over a family of increasing resolution, unsmoothed models, using model-choice criteria such as the marginal model likelihood or "evidence". Our implementation is a splitting method, which is effective at placing model resolution where it is statistically justified, typically producing O(10) layers for 1D CSEM data.

Nonlinearity means that inversions built without artificial smoothing are nearly always non-unique, especially for models below natural resolution. Point estimates of models are not adequate for decision making. Strict Bayesians regard the full posterior distribution as the important quantity: in this case, the posterior is a multimodal "mixture" distribution, and planning decisions based on CSEM data must take this into account. Enumeration of the multiple modes is necessary. Sampling from the posterior is difficult using Markov Chain Monte Carlo (MCMC) on account of the poor scaling, strong correlations, and nonlinearity. Some approximate sampling using the parametric bootstrap is possible.

Bayesian formulations for a layer based 1D model (Fig. 1)

Priors

Based on regional geological knowledge, the *marginal prior* distributions for layer resistivity will be a mixture of components with weights based on prior-probabilities of types and fluids, suitably smeared from upscaling effects to reflect internal heterogeneity. We approximate this using a truncated log-normal: see Fig. 2.

For a model with "vertical correlation", we can use, e.g. the hierarchical truncated Gaussian

$$P(m) \sim \frac{\exp(-\frac{1}{2}(m-m_p)^T C_p^{-1}(\alpha)(m-m_p))}{|C_p(\alpha)|^{1/2}} H(m-m_{LB})$$

where $C_{p,ij}(\alpha)=\sigma_p^2 e^{-\alpha|i-j|}$, H the Heaviside function, and $1/\alpha$ is a correlation length.

Likelihoods

Noise in CSEM is a complex mixture of instrumental, external and modelling noise. Tractable models are needed to make progress. We expect careful preprocessing, to enable Gaussian noise models to be used. For data $\mathbf{d}$ and a noise estimate in C_d

$$L(\mathbf{d}|\mathbf{m}) \sim \frac{1}{\sigma_n^{n_d}|C_d|^{1/2}} \exp(-\tfrac{1}{2\sigma_n^2}(\mathbf{d}-\mathbf{F}(\mathbf{m}))^T C_d^{-1}(\mathbf{d}-\mathbf{F}(\mathbf{m})))$$

Here $\mathbf{F}(\mathbf{m})$ is the forward modelled response. The variance parameter σ_n is added to absorb noise-level mis-estimates.

Fig 1. Modelling setup

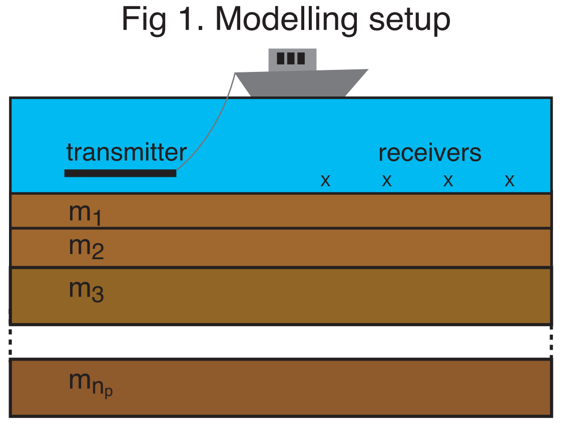
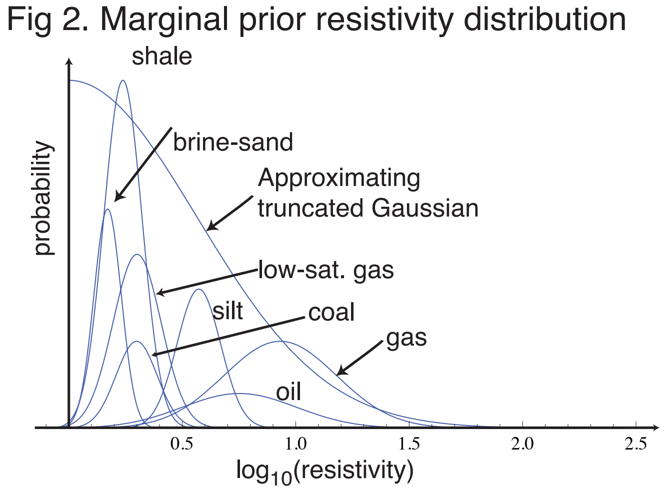


Fig 2. Marginal prior resistivity distribution



Optimisation framework for maximum aposteriori (MAP) models

MAP models are obtained by local minimization (over $\mathbf{m}$,σ_n,α) of the -ve log posterior probability:

favours large noise   favours small noise

$$\begin{aligned} -2\log(\Pi(\mathbf{m},\mu,\sigma_n|d)) &= (\mathbf{d}-\mathbf{F}(\mathbf{m}))^T C_d^{-1}(\mathbf{d}-\mathbf{F}(\mathbf{m}))/\sigma_n^2 + n_d\log(\sigma_n^2) + \log|C_d| \\ &\quad +(\mathbf{m}-\mathbf{m}_p)^T C_p(\alpha)^{-1}(\mathbf{m}-\mathbf{m}_p)) + \log(|C_p(\alpha)|) + n_p\log(2\pi) \end{aligned}$$

favours "rougher" models (large α)   a bit like: $\mu\partial^T\partial + diag(W^2)$ favours smooth models (small α)

The noise level σ_n and (possible) correlation α are determined by competition between the terms shown. Balances between the noise level and the smoothing/correlation α make good sense: as the noise level rises, the inferable resolution $1/\alpha$ broadens[2].

Optimal regularisation or smoothing can thus be obtained by standard maximum aposterior/likelihood type inference: note the objective has a term favouring "simplicity". For comparison, the usual "discrepancy principle" approach[1] uses $C_p(\alpha)^{-1} \leftarrow \mu\partial^T\partial + diag(W^2)$ (with ∂ a finite-difference derivative); there is no explicit term favouring simplicity, and we lose control over the prior marginals:

$$\chi^2 = \underbrace{(\mathbf{d}-\mathbf{F}(\mathbf{m}))^T C_d^{-1}(\mathbf{d}-\mathbf{F}(\mathbf{m}))}_{\chi^2_{\text{misfit}}} + \underbrace{(\mathbf{m}-\mathbf{m}_p)^T C_p(\alpha)^{-1}(\mathbf{m}-\mathbf{m}_p))}$$

$$\boxed{\text{Discrepancy principle: find largest } \mu \text{ so } \chi^2_{\text{misfit}} = \nu = n_d - n_p, \text{ using } C_p^{-1} \to \mu\partial^T\partial + \text{diag}(W^2)}$$

Projected-Newton bound-constraint methods for minimising a function $f(\mathbf{m})$

Noise can drive MAP type estimates towards unphysically low resistivities. Efficient bound-constraint methods are needed. We use an active set method, and projected line-search or trust-region searches towards the Gauss-Newton point.

The projection operator P "snaps" a point to within the feasible region: e.g. $P(m_i)=max(m_i,m_{LBi})$, if m_i has lower bound m_{LBi}. At stage k, the bound on m_i is termed "active" if $m_i-m_{LB,i} < \varepsilon_k$ and $\partial f/\partial m_i > 0$. The active set A is the set of active indices.

Newton steps are always of form $\mathbf{m}_{k+1}=\mathbf{m}_k - H^{-1}.\nabla f$. Active set methods separate out the inactive part: we set $H_{ij}=\delta_{ij}$, $i\in A$, with $\nabla f_i \to m_i-m_{LB,i}$. A backtracking ($\lambda$) line search towards $P(\mathbf{m}_k+\lambda\mathbf{m}_{k+1})$ then follows. Rapid convergence should ultimately emerge in the subspace of inactive constraints. The very poor scaling of CSEM problems make the implementation of this tricky. Linear-like convergence can appear for some time as the current point squeezes down narrow valleys, even for unbounded problems.

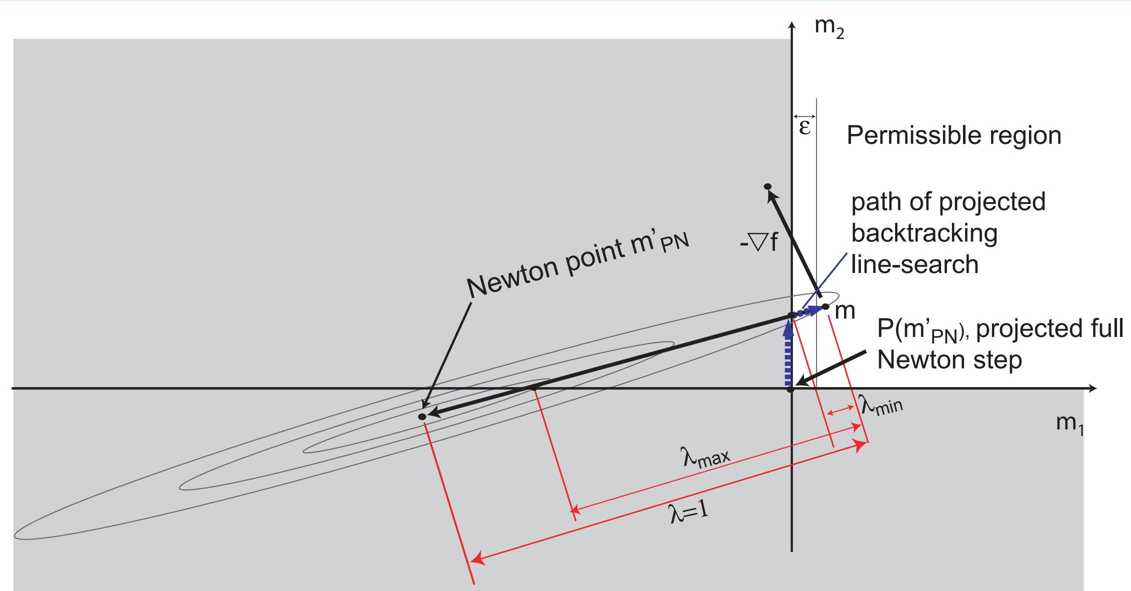


Fig 3. Backtracking Projected Newton method, just before onset of active constraint

References

1) C. Farquharson and D. Oldenburg "*A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems*", Geophysical Journal International, v156 (2004)
2) Y. Mitsuhata, "*Adjustment of regularization in ill-posed linear inverse problems by the empirical Bayes approach*" Geophysical Prospecting, v52, (2004)
3) A. Gelman et al, *Bayesian Data Analysis*, Chapman and Hall (1995)

Resolution via splitting and model-selection methods

Resolution can be inferred by performing model selection over a suite of models of variable spatial discretization. We need an efficient algorithm for visiting/generating models, and a measure of model "probability", or statistical significance. We use the marginal model likelihood (MML), or "evidence", obtained by integrating the posterior probability over all the free parameters:

Marginal model likelihood:

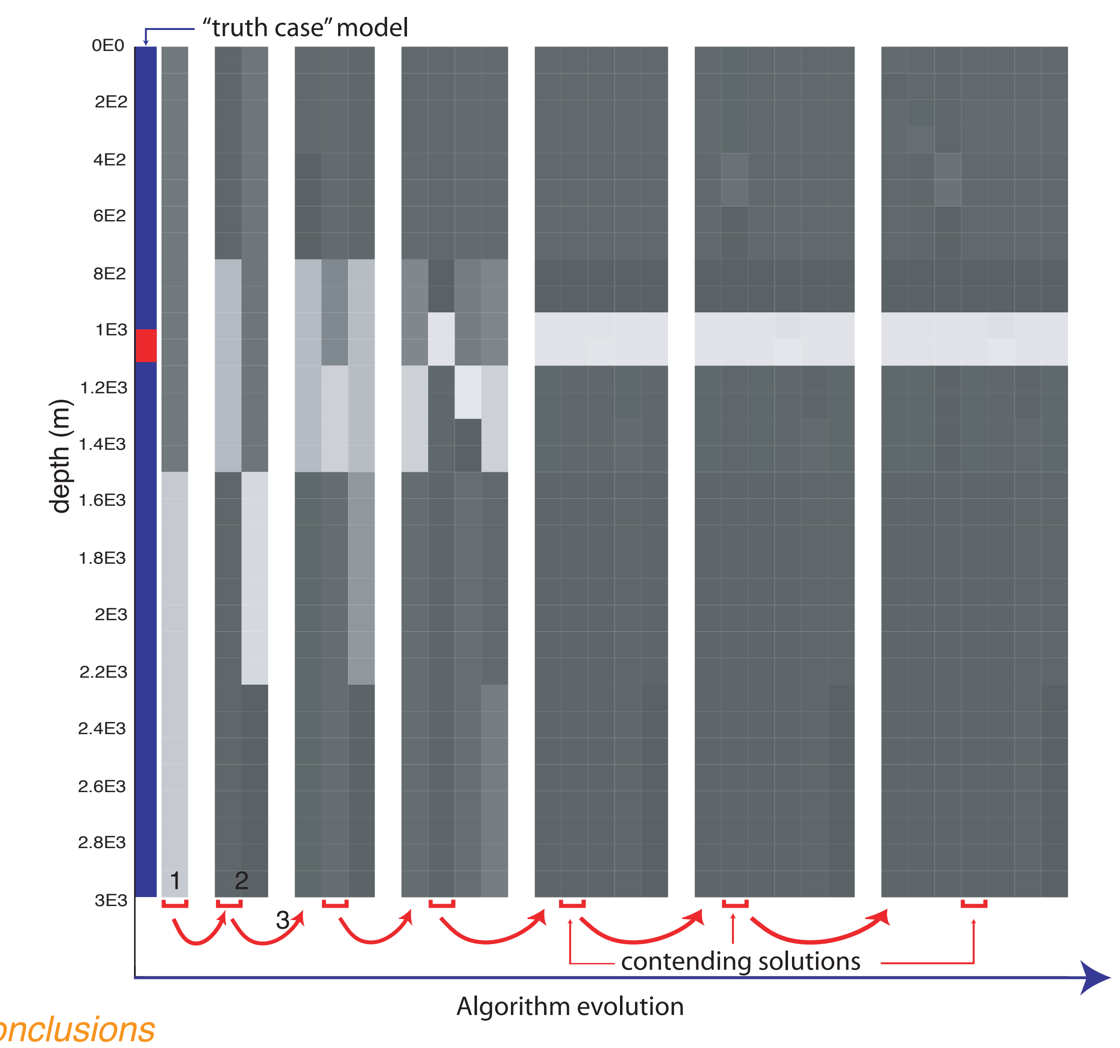
$$\pi(k) = \int L(d|m_k)p(m_k)dm_k.$$

Laplace approximation:

$$\begin{aligned} -\log(\pi(k)) &= \big[\tfrac{1}{2}(\mathbf{d}-\mathbf{F}(\mathbf{m}))^T C_d^{-1}(\mathbf{d}-\mathbf{F}(\mathbf{m}))/\sigma_n^2 \\ &\quad +\tfrac{1}{2}(\mathbf{m}-\mathbf{m}_p)^T(C_p(\alpha)^{-1})(\mathbf{m}-\mathbf{m}_p)) + \tfrac{1}{2}(\log|H|)\big]_{\mathbf{m}=\text{MAP point}} \\ &\quad +\tfrac{1}{2}n_d\log(\sigma_n^2) - \tfrac{1}{2}\log(|C_p(\alpha)^{-1}|) + \tfrac{1}{2}n_p\log(2\pi)) \end{aligned}$$

The well-known Bayes information criterion (an $exp(-n_p log(n_d)/2)$ "penalty" term in $\pi(k)$) is a further approximation of the Laplace approximation.

Based on partitions of an underlying lattice, the set of all possible models can be enormous. We explore a limited subset based on greedy recursive algorithm. One simple possibility, depicted below, is 1) Start with a 2-layer parent, 2) split each layer in the parent, invert for the MAP model, record the MML, 3) embed the best split, make the new model the parent, and recurse. Many variations are possible, but even this simple choice is fairly successful: see the figure below.



Conclusions

Resolution in CSEM problems can be inferred by either hierarchical models with meta-parameters expressing spatial correlation, or by model-selection over families of models with variable spatial discretization. In both cases, the "Occamist" virtue of parsimony is quantified, and leads naturally to models of maximum statistical significance, and thus reliability in prediction and inference.

Bayesian approaches to CSEM: Optimization, Multimodality and Uncertainty

James Gunning,
CSIRO Petroleum, Bayview Ave.,
Clayton, Victoria, Australia 3168.
E-mail: James.Gunning@csiro.au

Uncertainty evaluation

CSEM is inherently a low resolution technique. Models are often built below natural resolution, which rapidly increases the degeneracy and multimodal character of the solution space. The likelihood then is not focussed, but becomes a curving, "walnut"-like null space. Typically, Gaussian priors do not focus the posterior density at compact modes. A set of local modes is usually produced, with weak barriers separating the modes. "Geodesic"-like connections between modes are not usually straight lines in resistivity space.

The contorted shape of the posterior distribuition make uncertainty estimates difficult. Local covariance matrices estimated from the Hessian at particular modes are typically useless, except for very coarse models. Markov Chain Monte Carlo techniques are normally the tool of choice for Bayesian studies, where the posterior distribution is sampled by a random walk that "diffuses" over the support of the distribution. For problems with very strong correlations, component-wise sampling methods like the Gibbs sampler are impractical. Transformation to independent coordinates is also not practical, since the posterior is highly curved. Bootstrapping techniques, which allow the use of optimization to generate samples, have proved useful for this problem.

Multimodality. Mapping the interconnections of modes

Globalised mode enumerations strategies like least-squares with randomized starts and layer flipped starts reveal a variety of local optima. The interconnectedness between modes is interesting from the point of view of model uncertainty and possible sampling algorithms. How are modes connected? One possible approach is to minimise the line integral of the -log(posterior) between two modes A and B:

$$\Lambda_{AB} = \int_A^B \chi^2(\mathbf{M})dl$$

Minimum connecting paths should look like Fig 1 below. To find this path, a simple way is to discretize the integral using N intermediate points, add pieces to keep the segments about equal length, and minimize the sum-approximation. Computational kernels for the gradients at each segment node are simple to construct once gradients for the CSEM problem are available. Our discrete approximation is

$$\Lambda_{AB} \approx \sum_{i=0}^{i=N+1} \frac{\chi^2(\mathbf{M}_i)+\chi^2(\mathbf{M}_{i+1})}{2}||\mathbf{M}_{i+1}-\mathbf{M}_i|| + \mathcal{A}\underbrace{\sum_{i=0}^{N}(||\mathbf{M}_{i+1}-\mathbf{M}_i||-\bar{m}_S)^2}_{\text{favour equal-length segments}}$$

$$\bar{m}_S = \frac{1}{N+1}\sum_{i=0}^{N}||\Delta M_i||$$

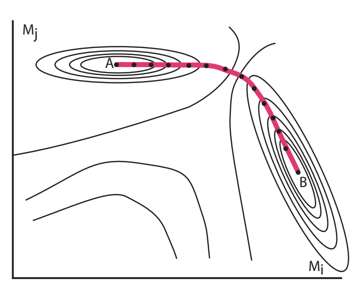


Fig 1. Typical conjectured "saddle-like" links between modes A and B in a CSEM problem. The "maximum probability interconnecting path is disctretized with N intermediate points, whose positions form a tractable optimization problem.

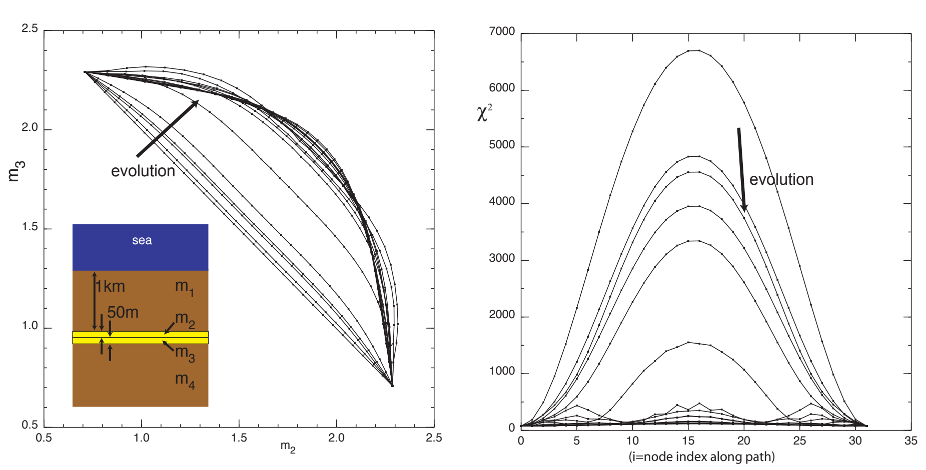


Fig 2. Evolution of mode-connection paths under BFGS for standard "split-canonical" model (deliberate split of reservoir layer into two underresolved 50m units). Left: log10(resistivity) of each reservoir layer. Right: χ^2 misfit function along saddle-link trajectory as optimization evolves. In practice, initial trajectories linear in resistivity space will be used.

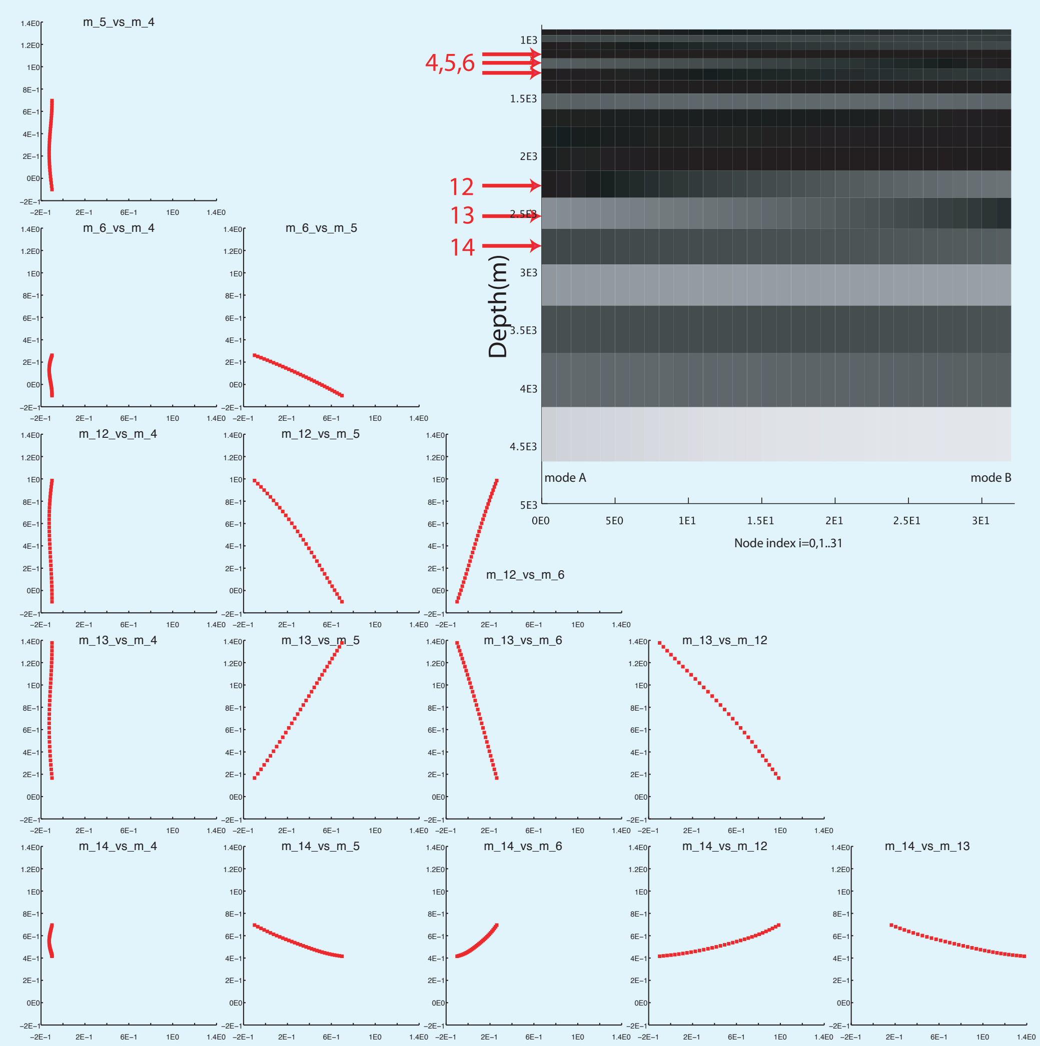


Fig 3. Typical mode-connection paths for a pair of modes in an 18-layer problem on a logarithmic grid. The inset depicts the model resistivity, "morphing" from mode A to mode B, lighter=resistive. For the 6 layers picked out in red, the scatterplot above shows the trajectory as a slice of the 18 parameter space. The exchange of resistivity near some possible reservoir target is obvious from the inset.

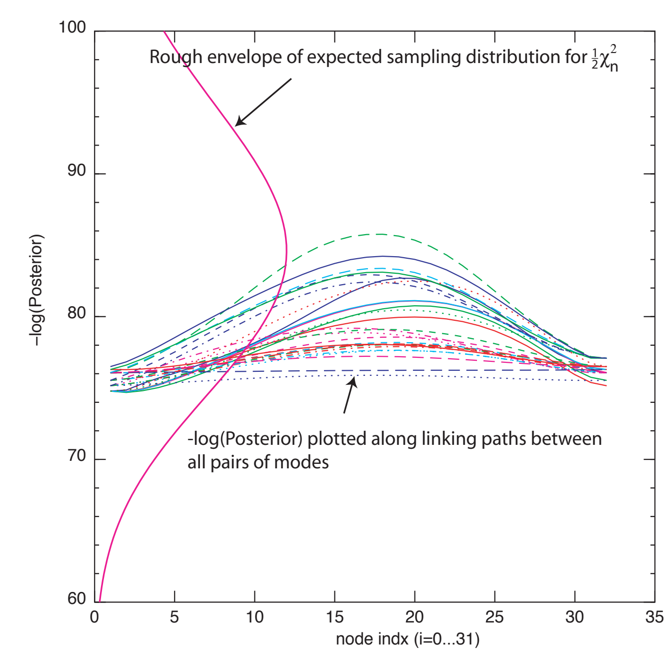


Fig 4. Example of the -log(posterior) function (roughly, $\chi^2/2$) plotted over the set of all saddle-paths connecting each pair of modes, for an unsmoothed problem with 8 modes. Each saddle path has relatively low probability barriers, so we should expect samples from the posterior to see most of these modes.

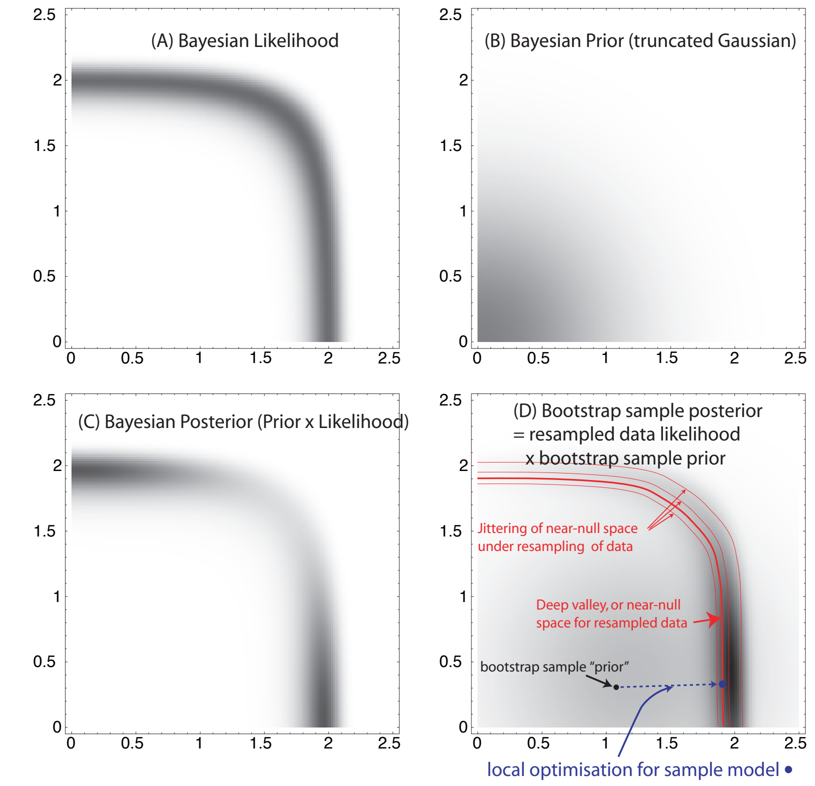
Posterior uncertainty sampling (Figs 4,5,6)

The Bayesian posterior for CSEM problems ought ideally to be sampled using some efficient MCMC method. The tight, correlated, and twisting posterior presents serious challenges to all standard methods. No simple, constant variable transformation neutralises the correlations. Strong correlations make component-wise samplers such as Gibbs or single-variable slice samplers extremely inefficient. Scaled Metropolis-Hastings samplers cannot adapt their scale to the twisting valleys without breaking reversibility.

An approximate way of sampling from the posterior is the parametric bootstrap, which is rigorous for a fully linear model, and asymptotic in large *n* for non-linear problems that are not ill-posed. Since this is a "frequentist" technique, the Bayesian prior has to be recast to look like "effective data", which is trivial[3]. We call this the Bayesian parametric bootstrap.

The method is roughly this. A MAP model is inverted for, using the true data and prior. Resampled data and resampled priors are then generated, and for each set, globalised inversions are run and the MAP model inverted from these is treated as a "posterior sample". The ability to use optimization in each bootstrap sample is a critical advantage over MCMC: the sampling is slower, but widely dispersed, converged, and independent. Fig. 4 is a cartoon explanation of how the Bayesian parametric bootstrap works, and Figs 5,6 show an example.

Fig. 4. Cartoon explanation of Bayesian parametric bootstrap. Posterior density (C) constructed from (B) prior and (A) likelihood for typical underresolved CSEM problem. Inset D shows the typical machinery of the parametric "Bayesian" bootstrap: 1) resampling the data "jitters" the null space (the red line along the valley) 2) resampling the prior "illuminates" different zones of the null space, and local optimization produces samples from the center of the "illuminated zone".



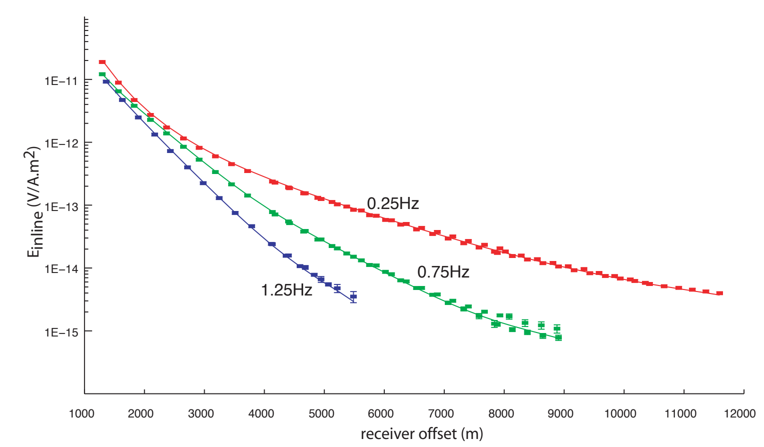


Fig. 5 Inline CMP electric field data used for the parametric bootstrap study shown in Fig 6. The inversion study uses an 18 layer logarithmic grid

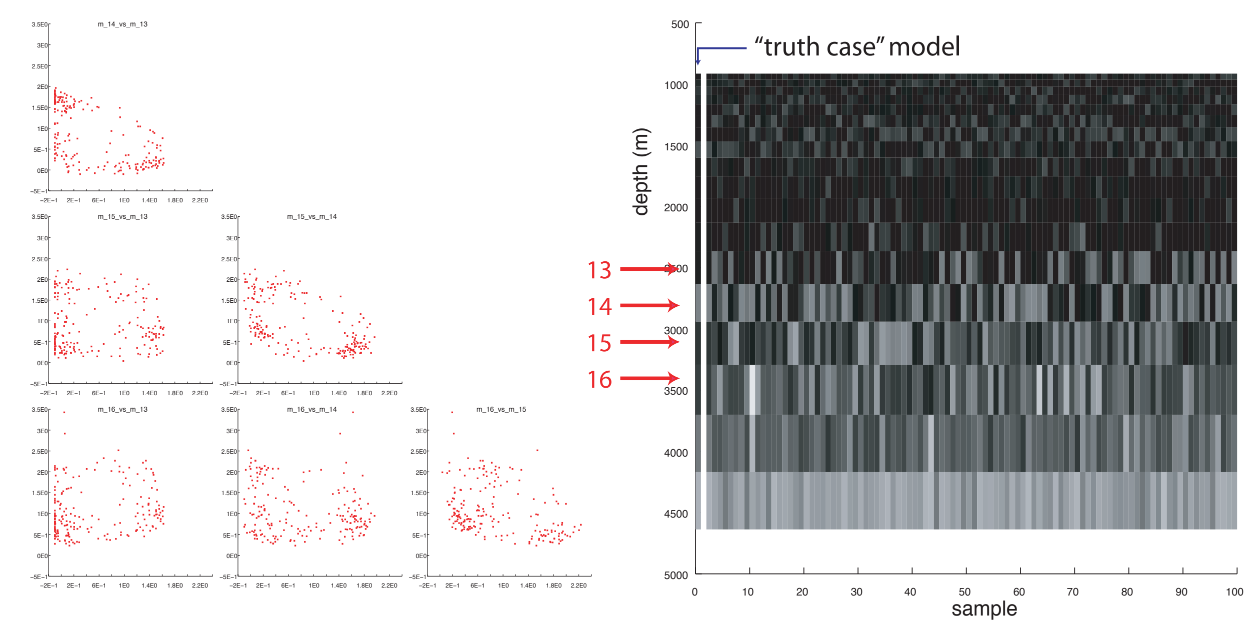


Fig. 6 Left: parametric bootstrap posterior samples as a scatterplot, on selected layers from the inversion model. Multimodality and clustering is clearly apparent. Right: posterior samples generated by the bootstrap procedure, with lighter shades more resistive. Samples are strongly independent, and layer-exchanging of resistive material is evident.

Conclusions

The uncertainty in CSEM inversion needs to take into account multimodality as well as uncertainty within modes. Diverse multiple modes can be found using standard non-linear least squares techniques, with starting points either randomized or constructed from layer flipping. Sampling the rigorous Bayesian posterior using MCMC is a difficult problem, but approximate sampling using the parametric bootstrap with globalised inversion is effective. The approximate samples are widely dispersed and independent. The posterior resistivity distribution for target layers is critical in decision making.