Reconciling data using Markov Chain Monte Carlo: An application to the Yellow Sea - Korean Peninsula region

M. E. Pasyanos, G. A. Franz, A. L. Ramirez

September 2, 2004

Disclaimer

Reconciling data using Markov Chain Monte Carlo: An application to the Yellow Sea – Korean Peninsula region

Michael E. Pasyanos, Gregory A. Franz, Abelardo L. Ramirez, Lawrence Livermore National Laboratory, Livermore, CA

Abstract. In an effort to build seismic models that are most consistent with multiple data sets, we have applied a new probabilistic inverse technique. This method uses a Markov Chain Monte Carlo (MCMC) algorithm to sample models from a prior distribution and test them against multiple data types to generate a posterior distribution. While computationally expensive, this approach has several advantages over a single deterministic model, notably the reconciliation of different data types that constrain the model, the proper handling of uncertainties, and the ability to include prior information. We also benefit from the advantage of forward modeling rather than inverting the data. Here, we use this method to determine the crust and upper mantle structure of the Yellow Sea and Korean Peninsula (YSKP) region. We discuss the data sets, parameterization and starting model, outline the technique and its implementation, observe the behavior of the inversion, and demonstrate some of the advantages of this approach.

Introduction

In the earth sciences, we often find that earth models derived using one set of geophysical data can often be inconsistent with models derived using another set of observations. Since we are interested in understanding the nature of the true earth, what we are really seeking are models that are most consistent with the full set of observations. Probabilistic inverse techniques, like the Markov Chain Monte Carlo (MCMC) algorithm, have had recent success in combining disparate data types into a consistent model. The stochastic methods that we consider here invert data by probabilistically sampling the model space, comparing the observations predicted by the proposed model to the observed data, and preferentially accepting models that produce a good fit, generating a posterior distribution of models. The model space is mapped through a series of stages comparing proposed models to data, with Bayes' Theorem relating the prior and posterior distributions. This stochastic geophysical model is able to reliably predict geophysical observations for a wide variety of data types, and provide accurate estimates of their uncertainties.

Monte Carlo integration works by drawing samples from a distribution, then forms sample averages to approximate expectations. MCMC draws these samples by running a cleverly constructed Markov chain for a long time (Gilks et al., 1996). A Markov chain is a sequence of points in the model space whose probability at a given time depends upon the value at previous points. A good example would be a random walk. The Markov chain that we construct here follows a set of rules that preferentially moves to more likely states in the model space, but sometimes moves to less likely states. The result is a chain that can sample the complete model space, efficiently inspect high-likelihood regions, but that does not get trapped in local extrema.

Both Bayes theorem and sampling methods like Monte Carlo have been extensively employed in the earth sciences. Tarantola (1987) employed a Bayesian framework for tomography, which incorporates a prior background model into the tomographic inversion. While these methods succeed at including prior information, they assume the model follows Gaussian statistics, and seek to find a single solution to the problem with uncertainties derived from this assumption. In our formulation of the problem, no assumptions have been made about the distribution of models and we find, in fact, that in many instances Gaussian distributions are not applicable. Instead of a single model, what we seek are the distributions of models that are most consistent with both our prior information and our observations.

MCMC originated in statistical physics and more has found recent application in many different fields. Application in the earth sciences has been more limited. In at least one recent study, Shapiro and Ritzwoller (2002) employed MCMC, along with a linearized inversion and simulated annealing, to invert for shear velocity structure using surface waves. The methodology proposed here differs from previous earth science applications by using multiple stages, each constraining the model with different data sets, to generate a new distribution of models. This method has recently been applied to map electrical resistivity changes (Ramirez et al., 2004). Our goal is to employ the technique to estimate regional earth structure using multiple geophysical data sets.

By ordering our stages properly, we can quickly reject models that cannot fit the observations that are easiest to calculate. The data sets that are more computationally intensive to forward model are relegated to later stages, increasing the efficiency of the inversion. Another important distinction of this methodology is that we do not seek a single best model or to simplify a distribution of models into a model with uncertainties. Rather, the final product is a posterior distribution of models that is best able to fit all of the data. As we will elaborate more in the paper, this approach has several important advantages, including less-restrictive distributions on the models, as well as the ability to reliably estimate uncertainties on new observables.

Data and Parameterization

We have selected as our study area the region of Eastern Asia in the vicinity of the Yellow Sea and Korean Peninsula (YSKP). We use seismic data from a number of sources in the region, including stations from the GSN, IMS, and CDSN networks, and data from regional broadband networks like those of KMA and KIGAM in South Korea. We also made use of two PASSCAL deployments in China and North Korea.

Two observational data types that we are currently using to construct the model are surface wave group velocities and body wave travel times. Rayleigh wave dispersion measurements have been made for thousands of paths across the region using a multiple filter analysis technique and, as shown in Figure 1a, provide excellent coverage (Pasyanos et al., 2003). The data is assembled by taking dispersion measurements at 5 s intervals of period between 10 and 100 s. The short period data is sensitive to the

shallowest earth structure (i.e. sediments), intermediate period data to deeper crustal structure (i.e. crustal velocity, crustal thickness), and long period data primarily to upper mantle structure. Combined, they do an excellent job of predicting average S-wave velocity structure. Surface wave data, however, has the downside of large tradeoffs between velocities and the depth of discontinuities, allowing multiple models consistent with the same data. The uncertainties of the group velocities are derived from the broadness of the energy arrival and generally range from 0.05 km/s to 0.10 km/s for typical paths.

The second data type that we use is body wave travel times, most notably Pn (P-velocity upper mantle head wave), Pg (crustal P-wave), Sn (S-velocity upper mantle head wave), and Lg (crustal guided S-wave), all shown in Figure 1b. Coverage is generally not as good as the coverage of surface waves. These data, however, are complementary to the surface wave data both because some phases are more sensitive to P-wave velocities, but also because they sample the earth in a much different manner. For example, Pn is sensitive specifically to the P-wave velocity in the uppermost mantle, just below the Moho. By combining these with the surface wave data, we can reduce the non-uniqueness caused by the tradeoffs discussed above. The uncertainties of the travel times are derived from the phase, and estimated qualities of the event location and pick, and reflect the ambiguity of the timing of the arrival. First arriving P-phases generally have smaller uncertainties than phases picked in the coda of earlier-arriving phases.

The first implementation step is to develop a mechanism (the base sampler) for generating model realizations from the prior distribution. We parameterize the model as a set of layers in a regular latitude-longitude grid of 2°x 2° cells. The layers represent geologic structure (water, upper sediments, lower sediments, upper crust, middle crust, lower crust, upper mantle) which are specified by their seismic parameters and thickness. This has the advantages of reducing the size of the problem, imposing realistic constraints, and making the resulting models more interpretable. The model ranges between 23° and 57° latitude and 109° and 147° longitude, for a total of 323 cells. Combined with 6 variable layers (all but water) and 4 parameters for each layer (thickness, Vp, Vs, density), this results in a total of 7752 free parameters.

The starting model selected for the region was CRUST2.0 (Gabi Laske, pers. comm.) which is similarly parameterized. From the starting model and parameter constraints, the base sampler can initially select models consistent with the a priori distribution. The seismic parameters are all selected independently which produces variations in Vp:Vs and Vp:density ratios. The ratios are checked, however, and unrealistic values of these ratios for the layer type are discarded.

Finally, in order to reduce dependence on the starting model and introduce some randomness to initial models for each chain, we have randomized the starting model for each of the two chains in our search. Chains allow us to make several forays into the model space at separate starting points, allowing a quicker sampling the model space, particularly in complex problems. This ensures coverage of the target distribution, and

can also be used to monitor convergence. Prior to convergence, the run has not had enough time to properly sample the model space. The initial period, referred to as "burn-in", is still unduly influenced by the starting model. When convergence is finally reached, each Markov chain will have had a sufficient number of steps to independently sample the model space. The chains should have similar posterior distributions, which should be stationary in the long run. Convergence can be monitored by detecting when the Markov chains are independent of their starting points, by comparing sequences drawn from different starting points and checking that they are indistinguishable (Gelman, 1996).

Methodology

The MCMC approach that we use is a derivative of the Metropolis algorithm (Metropolis et al., 1953) as described by Mosegaard and Tarantola (1995). A brief description of the methodology follows. A more detailed description of the methodology (applied to an electrical resistivity problem) can be found in Ramirez et al. (2004).

The base sampler selects proposed models from the prior distribution. By comparing the degree of fit between the data predicted assuming the model m (by forward modeling the predicted data) and the observed data, we can determine the likelihood $L(m/d)$ of the proposed model. Bayes' rule relates the prior p and posterior distributions P as follows:

$$P(m \mid d) = p(m)L(m \mid d) \qquad (1)$$

We assumed a likelihood function of the form:

$$L(m) = k \exp\left[-\sum_i \left| \left(d_{pred,i} - d_{0,i}\right)^n \middle/ \sigma_i^n \right|\right] \qquad (2)$$

where $d_{pred,i}$ is the predicted data for a given model m, $d_{0,i}$ is the vector containing the observed measurements, σ_i is the estimated data uncertainty, k is a normalization constant, and $n \geq 1$. We assumed that $n = 2$.

If these transitions were always accepted, then the simulation would be sampling from the prior distribution. Instead, suppose that the proposal transition is only accepted by computing the respective likelihoods $L(\underline{m}_i)$ and $L(\underline{m}_j)$ for both the current and proposal states $\underline{m}_i$ and $\underline{m}_j$. If $L(\underline{m}_j) \geq L(\underline{m}_i)$, then accept the proposed transition and move to state $\underline{m}_j$. If $L(\underline{m}_j) < L(\underline{m}_i)$, then use a randomized decision rule and accept the proposed transition with probability $L(\underline{m}_j)/L(\underline{m}_i) < 1$ and move to state $\underline{m}_j$. Otherwise, transition back to state $\underline{m}_i$. By allowing the walk to transition to a less likely state, the process can move out of local extrema.

The samples generated through this process will have a limiting distribution that is proportional to the desired posterior distribution $P\left(\underline{m}|d_0\right)$. That is, the search tends to

hover in regions of space M containing states with greater prior propensities and higher likelihoods.

Stochastic models have several advantages over conventional models. First, while the model is data driven, as we have seen it is very easy to incorporate prior information. This information can range from non-specific priors like our starting model to very specific prior information (such as specifying the thickness of the crust in a location determined from a refraction profile) that should be incorporated. Importantly, the staged approach presents a way of reconciling data from many sources. For example, many data types (travel times, dispersion measurements, receiver functions, waveforms, etc.) with widely different sensitivity functions can be used. The posterior distribution represents models that are most consistent with all of the observed data. The relation between the different data types is not explicitly defined since each data type is simply forward calculated from the proposed models. Predicting the data from a model, of course, is simpler than inverting the data to produce a model.

Using this technique, we obtain the full posterior model distribution, and can reliably estimate the uncertainties of model parameters, both of which will be consistent with the uncertainties of the observables. Models and the parameters that describe models (i.e. sediment thickness) are not constrained to be normally distributed but could have non-Gaussian or even multi-modal distributions. The posterior distribution of models can be used to estimate uncertainties on observable parameters, such as would be necessary for providing travel time corrections. These uncertainties reflect the uncertainties of the model parameters along with their correlations.

Results

We have made a run consisting of two chains each running 8000 iterations. Separately, we also have a two-chain run in which the models proposed by the base sampler are automatically accepted. This provides us with a prior distribution, with which we can compare and contrast the posterior distribution. In both cases, we have thrown out models from the first 3/4 of the iterations in order to ensure that we are past burn-in.

To see how the method is performing, we look at a profile in the Yellow Sea at the center of our model. Figure 2 shows the prior and posterior distributions. In both cases, we find that the two chains are well-mixed, but that the distribution of profiles is different. The range of models in the posterior distribution has decreased significantly, accepting only models which fit the data well. In the presence of data, then, the model moves away from the prior distribution. Histograms of Moho depth, sediment thickness, and Pn velocity demonstrate the same (Figure 3). Not only have the peaks changed, but the distributions have gotten tighter, indicating lower uncertainties on these parameters. We find variance reductions of the parameters of 21%, 56% and 42% for crustal thickness, sediment thickness and Pn velocity. Notice as well that the distribution of sediment thickness is decidedly non-Gaussian, appearing to be more of a Poisson distribution. In contrast, in portions of our model at the edges where we have little data, we find only minor

differences between the prior and posterior distributions of model, and the uncertainties remain high. This indicates that the prior is reflected where there is no data to drive the model, as we would have anticipated.

We can assemble the results from each point to create parameter maps. Figure 4 shows a map of crustal thickness, along with the associated uncertainties. The values were determined by the mean and standard deviation for the posterior model. The maps are generally consistent with our understanding. We find thin crust in oceanic areas like the Sea of Japan and the Pacific Ocean, but not in the continental crust of the Yellow Sea. The thickest crust in our region is found to the west. We find differences from the prior model on the order of 5 km or so. Compared to the prior model, we are finding thicker crust in the Korean Peninsula and thinner crust in the Yellow Sea. Uncertainties are generally higher in the southwestern portion of our study area, where there is no body wave data. Inconsistency between data types is also reflected in higher uncertainties.

Conclusions

Stochastic methods are an innovative technique for producing data-driven models. They have a number of advantages compared to traditional models, such as the ability to reconcile different types of geophysical data. An important component is the ability to predict new observables with proper uncertainties. We have started to apply this technique with success to the YSKP region, using Rayleigh wave group velocities and regional travel time measurements. We can now start to include other suitable data sets in the inversion, including Love wave group velocities (which could simply be added to the other group velocities), and receiver function or waveform data (both which would require new stages). Waveform modeling would probably have the strongest constraints on the model along paths, but would also be the most computationally expensive to implement. We aim to improve the model by incorporating more data sets and increasing the resolution.

References

Gelman, A., Inference and monitoring convergence, in Markov Chain Monte Carlo in Practice, Chapman & Hall, New York, 131-143, 1996.
Gilks, W.R., S. Richardson, D.J. Spiegelhalter, Introducing Markov chain Monte Carlo, in Markov Chain Monte Carlo in Practice, Chapman & Hall, New York, 1-19, 1996.
Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys., 1*, 1087-1092, 1953.
Mosegaard, K., and A. Tarantola, Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res., 100*, 12431-12447, 1995.

Ramirez, A.L., J.J. Nitao, W.G. Hanley, R.D. Aines, R.E. Glaser, S.K. Sengupta, K.M Dyer, T.L. Hickling, and W.D. Daily, Stochastic inversion of electrical resistivity changes using a Markov Chain, Monte Carlo approach, submitted to *J. Geophys. Res.*, 2004.

Shapiro, N.M., and M.H. Ritzwoller, Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int., 151*, 88-105, 2002.

Tarantola, A., *Inverse Problem Theory*, Elsevier Science, Amsterdam, The Netherlands, 1987.

Figure 1. a) Path map of surface wave measurements. b) Path map of body wave travel times. Colors indicate Pg (cyan), Pn (red), Lg (blue), and Sn (green) phases.

Figure 2. Prior (left) and posterior (right) model distributions for a point in the Yellow Sea. P-wave velocity (in km/s) is plotted as a function of depth (in km).

Figure 3. Histograms of model parameters for prior and posterior model for parameters of crustal thickness, sediment thickness, and P-velocity of the uppermost mantle.

Figure 4. Map of crustal thickness and uncertainties (in km) of the posterior model.

30.0 second Rayleigh

Pn
Sn
Pg
Lg

lat=36 lon=124

lat=36 lon=124

Depth (km)

Velocity (km/s)

Velocity (km/s)