ESTIMATING THE UNCERTAINTY AND PREDICTIVE CAPABILITIES OF THREE-DIMENSIONAL EARTH MODELS

Delaine T. Reiter[1], William L. Rodi[2], and Stephen C. Myers[3]

Weston Geophysical Corp.[1], Massachusetts Institute of Technology[2], and Lawrence Livermore National Laboratory[3]

Sponsored by the Air Force Research Laboratory and the National Nuclear Security Administration

Award Nos. FA8718-09-C-0013[1,2] and DE-AC52-07NA27344LL09[3]
Proposal No. BAA09-73

ABSTRACT

In recent years many models of 3-D seismic velocity structure in Eurasia have been developed, using a variety of techniques and data. Most of these models are not accompanied by quantitative estimates of uncertainty, either in the model parameters themselves (e.g., seismic velocities) or in geophysical observables predicted by the models (e.g., body-wave travel times). Moreover, the various 3-D models produced by these studies have not been compared to one another in their predictive capabilities in any meaningful way within the monitoring research community. To address these issues we are developing and applying evaluation metrics that robustly quantify and compare the uncertainty and predictive capabilities of 3-D seismic velocity models.

There are two major elements in our approach. First, we are performing a comprehensive evaluation of a set of velocity models for Eurasia based on previously developed data misfit and event mislocation metrics. We are using a ground-truth (GT) data set available at the International Seismic Centre (ISC). We discuss some of the difficulties involved in reconciling discrepancies between the parameterization and coverage of various models available to the research community, and compare some of the standard metrics we have been able to derive for these models.

Second, we are investigating a new approach to evaluating velocity models based on a Bayesian framework for model uncertainty in the travel-time tomography problem. Bayesian analysis provides a regimen for deriving posterior model uncertainty that combines prior model uncertainty with the information from travel-time data. Our new approach focuses on the prior model uncertainty. Following concepts used in the kriging method, we are representing prior model uncertainty with geostatistical parameters that characterize the Earth's 3-D velocity heterogeneity relative to a given reference model. These parameters include the velocity variance, which quantifies the strength of velocity heterogeneity, and vertical and horizontal correlation lengths, which quantify the spatial smoothness of velocity heterogeneity. We assume these parameters vary with position in the Earth. Our approach is to estimate the geostatistical parameters from observed travel-time residuals, relative to a reference model of interest, by fitting the parameters to the second-order sample statistics of the residuals. Our premise is that the estimated parameters will be a rigorous metric for evaluating the reference model, at least in comparison to competing reference models. Further, the estimated geostatistical parameters describing velocity uncertainty can be mapped to prediction errors on travel times and other observables, providing further model evaluation metrics. Our preliminary efforts in developing this approach consist of fitting velocity variances and correlation lengths to AK135 residual statistics derived from events in Eurasia and Africa.

OBJECTIVES

The main objective of our research project is to develop and apply meaningful measures of the predictive capabilities of 3D Earth models. We will specifically focus on seismic velocity models, although our general approach should be applicable to models of other geophysical parameters such as attenuation and density.

Our project consists of two major elements. First, we are collecting a set of regional 3D velocity models for Asia and are performing a comprehensive and methodical evaluation of them using data misfit and event mislocation metrics. These metrics are already generally accepted as informative, if not complete, measures of the prediction capability of Earth models. Metrics will be evaluated for each model based on a common ground-truth data set comprising GT5 events and their arrival picks. A large part of this effort will be to establish rigorous validation tests, employing standardized methods for model parameterization, forward modeling (e.g., ray tracing), and event relocation. The tools we need for these tests are available to us in-house from our previous tomography and location projects.

The second element of our project is the investigation of a new approach for estimating the predictive capability of 3D velocity models. Our approach will convert the uncertainty in the parameters of a 3D velocity model, as a function of position in the Earth, into the uncertainty in travel-time or other geophysical predictions. To do this requires covariance modeling capabilities we have developed for travel times in an earlier project (Rodi and Myers, 2008) and which can be extended to other observables given appropriate model sensitivities.

RESEARCH ACCOMPLISHED

Three-Dimensional Velocity Models of the Crust and Upper Mantle

One of our first tasks has been to collect 3D regionalized velocity models from their authors and format them in a way that makes comparisons straightforward. Our test-bed model for the project is JWM, the *J*oint *W*eston/*M*IT inversion model (Reiter and Rodi, 2009). JWM is result of a joint inversion of regional P travel times and Rayleigh fundamental-mode group velocities. It consists of the P and S velocities and density of the crust and upper mantle for a broad region of Asia (defined in a geographic box from 0 – 60°N, 30 – 120°E). In addition to the JWM model, we have obtained the CUB2.0 surface-wave inversion model (Ritzwoller et al., 2002) with density; the Stevens et al. (2008) surface-wave inversion model, and a new joint inversion model (LLNL G3D) from Nathan Simmons and Stephen Myers at LLNL. We also have access to several *a priori* regionalized models, such as the Unified Model from LLNL (Begnaud et al., 2004; Pasyanos et al., 2004). We continue to seek models from other researchers, but many are not available for our use at this time.

In Figure 1 we display map-view slices from four of the models in our study. The Generic Mapping Tool (GMT) package (Wessel and Smith, 1995) was used to plot the P and S velocities at a depth of 120 km, using the same color map for each model. Prior to creating this plot, the models were reformatted to a common geographic representation and then interpolated to a depth-slice grid. The different panels in Figure 1 offer a direct comparison of the models and illustrate their similarities and differences. For example, JWM and CUB2.0 are similar to each other in the strength of heterogeneity at this upper-mantle depth, particularly in the southwestern corner of the model. CUB2.0 is defined on a 2°x2° grid, which produces a smoother result compared to the other models. The LLNL G3D inversion model appears to have similar patterns of heterogeneity as the CUB2.0 and JWM models, but its heterogeneity strength is significantly smaller. The *a priori* LLNL Unified model is least like the others in pattern and heterogeneity strength – we note that this model is equivalent to the 3SMAC model (Nataf and Ricard, 1996) at this depth. We do not show the Stevens et al. (2008) model here, since we have not yet converted it to our in-house format.

While the models may be quite different in snapshot form, they can often perform like each other in travel-time prediction and location tests, as we will see in a later section. We are investigating the behavior of the 3D models using different forward modeling techniques, in an attempt to determine how and why they differ.

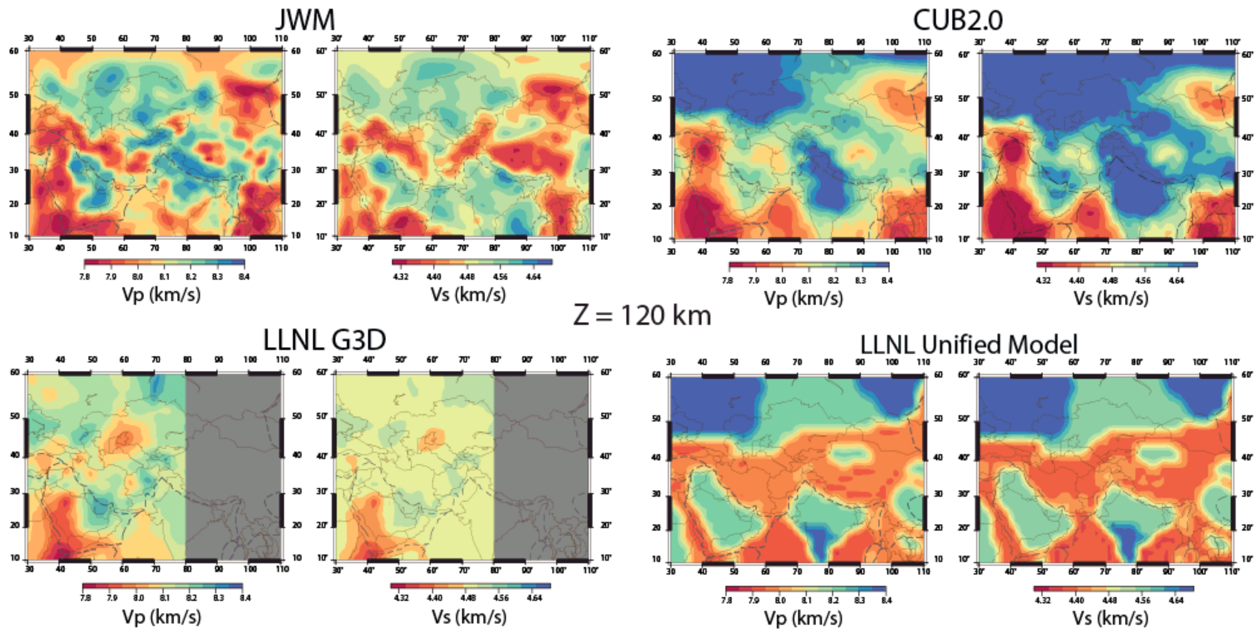
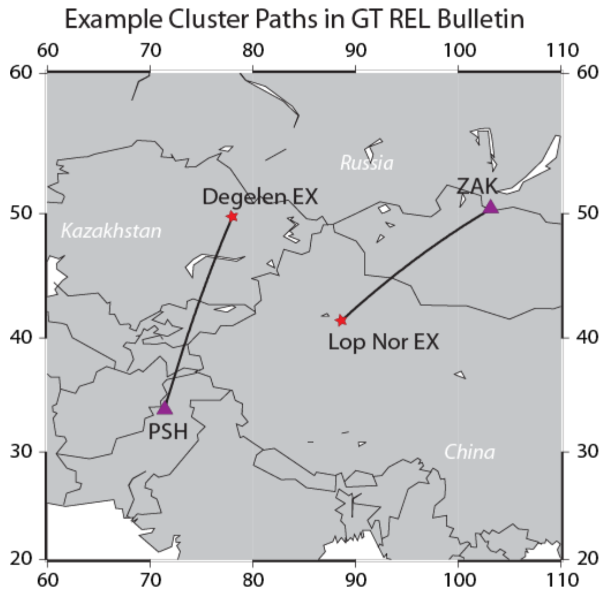


Figure 1. Map-view slices of P and S velocity from four of the regional 3-D velocity models in our study. The depth slices from each model are taken at Z=120 km depth, and the color scale is the same in each panel. The LLNL G3D model extends only to 80°E, and we plotted it to that longitude, leaving the remainder of the box empty.

Validation Data: Ground-Truth Explosions and Earthquakes

To evaluate models with respect to one another, we require a high-quality set of GT events whose epicentral locations are known precisely. We used the new catalog of GT0-5 reference events (Bondàr and McLaughlin, 2009) hosted by the ISC (http://www.isc.ac.uk). This global database includes more than 7,000 events whose epicentral location accuracy is known to at least 5 km. GT events with well-established locations and origin times have been used by multiple authors to validate velocity models (Ritzwoller et al., 2003; Flanagan et al., 2007). We filtered the new IASPEI Reference Event List (REL) for events in our study region and found 248 GT(0–5) explosions and 348 GT5 earthquakes, most of which are in specific geographic clusters. From these events we extracted the defining regional P and S arrival-time picks in the IASPEI REL that were used to develop the GT locations.

Even though the IASPEI REL bulletin represents a significant improvement over previous GT databases, we found that additional grooming was required to eliminate inconsistent arrivals. For example, we eliminated observations with clearly misidentified phase arrivals in crossover distances between various travel-time branches. In addition, we filtered out arrivals from distances less than 0.5°, since they do not sample a significant portion of our models. Following these simple filtering exercises, we examined the earthquake and explosion event clusters in the data and found repeated station-to-event phase readings that occurred over time. Most of these revealed consistent travel-time readings, especially for the first arriving P phases. However, at some distances and for some paths there were inconsistencies that made the readings unreliable. We illustrate this in Figure 2, where we plotted two examples of station-to-event paths that occur multiple times in the REL database: the Degelen nuclear explosion cluster with paths to station PSH in central Pakistan, and the Chinese Lop Nor test site explosions to station ZAK in Russia. For the Degelen example there are 27 P phase picks in the REL bulletin, with a mean epicentral distance between station and event of 16.4°. In the AK135 model the travel time for this path averages 232 s (using the exact GT hypocenters) with a standard deviation of .44 s. The mean of the observations is 231 s, but the standard deviation is 3.0 s. The same phenomenon is seen with the other example path, albeit with the S arrival and many fewer phase picks (5) in the bulletin. These large values of standard deviations in the travel times make it impossible to compare the validity of different velocity models along the paths, since many models can fit a travel time to within ±3.0 s. To date we have dealt with this problem by eliminating any phase paths in the REL arrival database that have standard deviations greater than 1.5 s for P or 2.0 s for S.



Degelen to PSH
P bulletin picks = 27

	Mean value	*Std Dev*
Δ	16.38°	0.03°
Esaz	200.0°	0.13°
Bulletin TT	231 s	3.0 s
AK TT Pred	232 s	.44 s
JWM TT Pred	230 s	.44 s

Lop Nor to ZAK
S bulletin picks = 5

	Mean value	*Std Dev*
Δ	13.4°	0.01°
Esaz	43.9°	0.06°
Bulletin TT	341 s	3.5 s
AK TT Pred	342 s	.40 s
JWM TT Pred	347 s	.40 s

Figure 2. Examples of persistent inconsistencies in the GT reference event database. On the left are two station-event paths that occur multiple times in the database; on the right are the statistics associated with the readings for the two paths. The standard deviations of the observations (Bulletin TT vs. AK TT Pred) indicate either pick errors or phase misidentifications that corrupt the GT phase bulletin.

Using the filtering criteria described above we derived a validation data set of 9,242 P-wave and 2,214 S-wave regional arrivals observed at stations within a latitude-longitude box defined by (0-60° N, 30-120° E). The great-circle raypath coverage for the P and S GT observations is shown in Figure 3. The GT data ray paths sample only a limited portion of the study region, which illustrates the difficulty of validating a model with travel times alone. However, the IASPEI REL database is currently the best available resource we have to demonstrate the performance of the 3D models.

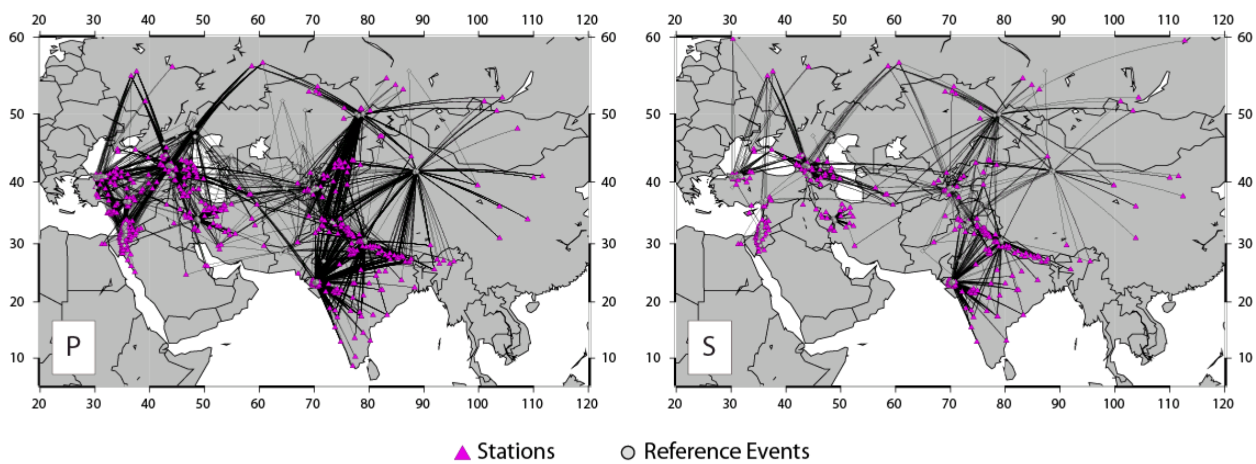


Figure 3. Great-circle raypath coverage for the P (top) and S (bottom) paths in the IASPEI REL ground-truth database. Stations are represented by purple triangles and events by gray circles. Note the sparse coverage over the area of the regional 3D models.

Validation Exercise: Regional Travel-Time Prediction

One of the first assessments we make of the 3D models involves measuring their ability to accurately model the regional travel-time data. As noted by other authors, improved travel-time prediction is the most direct way to estimate the quality of a 3D model, because it eliminates the effects of network geometry and variable pick quality.

One of the more difficult questions we are addressing in our project is how to present the model fits in a coherent and meaningful fashion. There are many ways to examine residuals; for example, on a station-by-station or event cluster basis (e.g. Flanagan et al., 2007), or through spatial patterns as a function of distance and station (e.g., Murphy et al., 2005), or by the fits to empirical phase path anomalies (e.g., Ritzwoller et al., 2003). Figure 4 represents one way to present travel-time fits to the regional P data in the GT database relative to a model. For this figure we binned the data as a function of event-to-station distance and the absolute value of the residual. By defining a visually distinct color scale and the bin sizes (here set to 0.25° in distance and 0.5 seconds in residual size), we can examine a distance-based measure of the fit; however, we remove any sense of the spatial/geographic variation or dependence of the model fits. Nevertheless, we can make some general observations about the models from these plots. The P residuals with respect to the AK135 model (Figure 4a) show a distinct pattern that is due to the tectonics of the study region. As noted in Ritzwoller et al. (2003), the increase in spread at the longer regional distances is likely due to some ray paths bottoming well below the bottom of the crust and encountering low velocity zones. The residuals plotted in Figure 4b-e show that all of the 3D models fit the GT observations better than AK135, particularly at the shorter distances (<8°). There are some differences between the 3D models at the longer distances, but this type of plot tends to mask much of the variation that is due to paths that cross differing tectonic regimes. It is remarkable that while Figure 1 implies the significant differences in the models, the travel-time predictions and subsequent residuals do not show the same variability when they are presented in this manner.

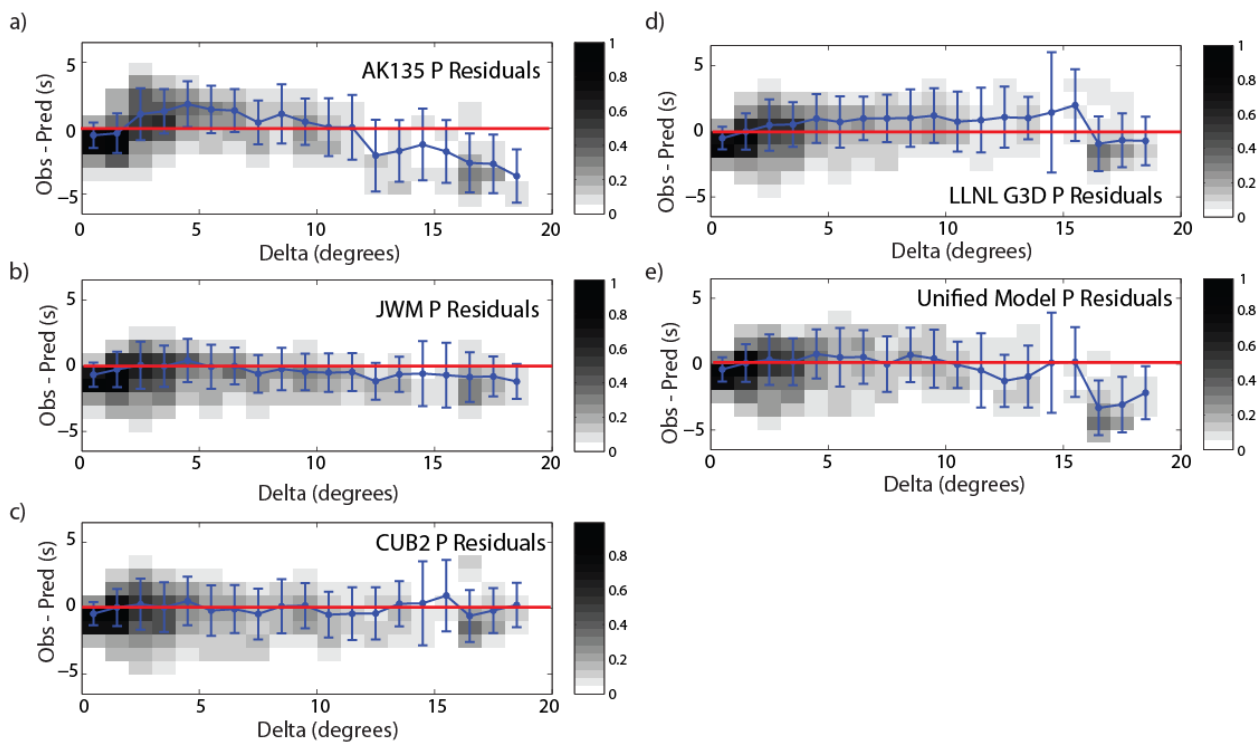


Figure 4. Ground-truth residuals for the P arrivals in our GT database, binned according to station-to-event distance and residual (intervals of 0.25° in distance and 0.5 seconds in residual size). The bin hit counts are plotted with the nonlinear color scale at the side of each subplot; the blue error bars indicate the mean and one standard deviation of the residuals at 1° intervals. Residuals with respect to five models are shown: a) AK135; b) JWM; c) CUB2.0; d) LLNL G3D; and e) Unified Model.

Evaluation of Velocity Models Based on Model Uncertainty Analysis

For the second task in our project we are investigating a new approach to evaluating velocity models based on the analysis of model uncertainty. The approach is formulated within a Bayesian framework as it applies to the travel-time tomography inverse problem. Bayesian analysis provides a regimen for deriving posterior model uncertainty that combines prior model uncertainty with the information from travel-time data. Our new approach focuses on the prior model uncertainty. Following concepts used in the kriging method, we are linking prior model uncertainty to geostatistical parameters that describe the difference between the Earth's true velocity distribution and that of a given reference model. These parameters include the velocity variance, which quantifies the strength of velocity heterogeneity, and vertical and horizontal correlation lengths, which quantify the spatial smoothness of velocity heterogeneity. We assume these parameters vary with position in the Earth.

Our approach is to estimate the geostatistical parameters from observed travel-time residuals relative to a reference model of interest. This entails fitting the parameters to the second-order statistics of the residuals, such as sample variances, covariances, variograms, etc. Our premise is that the estimated parameters will be a rigorous metric for evaluating the reference model, at least in comparison to competing reference models. The estimated geostatistical parameters would also provide a meaningful prior uncertainty for use in a tomographic inversion of the data, although this is not the primary purpose of the approach. A forward model for the estimation problem is available in the results of a previous project which developed theoretical techniques for calculating travel-time variances and covariances given assumed geostatistical parameters for velocity (see Rodi and Myers, 2008).

Our initial efforts with this approach have focused on fitting velocity variance profiles to an empirical estimate of travel-time variance vs. epicentral distance, derived from a statistical analysis of travel-time residuals relative to the AK135 reference model. Using the method of Rodi and Myers (2008), we computed theoretical travel-time variance curves for the six velocity variance profiles listed in Table 1. In conjunction with each profile, the following velocity correlation lengths were assumed: a lateral correlation length of 300 km everywhere, and a vertical correlation length of 17.5 km in the crust and 60 km in the mantle.

For instance, we can use this approach to examine the dependence of travel-time variance on distance. To illustrate this concept, we consider a linear array of 70 stations equally spaced from 0.5° to 35° in epicentral distance from a nominal event, at a common azimuth from the event. The resulting 70x70 travel-time covariance matrix reveals the dependence of travel-time variance on distance and of travel-time correlation on distance and distance separation between stations. We computed the travel-time covariances using the AK135 reference model (Kennett et al., 1995) and fixed correlation lengths (300 km horizontal, 17.5 and 60 km vertical). We then modeled six cases of slowness standard deviations in the crust and mantle, using the parameters shown in Table 1.

Table 1. Assumed velocity standard deviation for modeling travel-time variances.

	Case A	Case B	Case C	Case D	Case E	Case F
Crust	3%	3%	6%	12%	12%	12%
Moho - 210 km	1.5%	2%	2%	2%	2%	1.5%
210 km - 410 km	1.5%	2%	2%	2%	1%	1.5%
> 410 km	1.5%	1%	1%	1%	0%	0%

Figure 5 shows the calculated travel-time standard deviation for the six cases as blue circles. Each panel shows the travel-time standard deviations (blue circles), which are the square-roots of the diagonal elements of the 70x70 covariance matrix. Discontinuities in standard deviation are evident at certain distances. These are directly related to discontinuities in the ray parameter of first arriving P waves, as predicted by the AK135 model. The red line in each panel of Figure 5 is an empirical relation inferred from travel-time residual statistics derived from the LLNL ground-truth database (Ruppert et al., 2005). First-arriving P-waves were used for events meeting the GT criteria of Bondàr et al. (2004) or whose location was constrained by non-seismic means (e.g., known explosions, mine collapses, earthquakes with InSar signals). We used a methodology similar to Flanagan et al. (2007) to assess travel-time model error. We began by selecting only stations with sufficient data for distance-dependent variance estimation (at least 10 arrivals in each 1° event-station bin). For each station, we assessed pick (random) error by computing the standard deviation of residuals for clusters of events within 20 km of one another. The random error for many event clusters was used to determine pick error as a function of station-event distance. We subtracted distant-dependent pick error variance from travel-time residual variance to estimate distance-dependent model-error variance. The final curve was obtained as an average of the model-error curve for stations throughout Eurasia and Africa.

Comparing the six panels in Figure 5, it is clear that the theoretical calculations (blue circles) for Case E give the best overall fit to the empirical curve over the regional and teleseismic distance range. Large velocity heterogeneity in the crust (12%) is needed to predict the large travel-time model-error variance observed at short distances. To match the peak variance at 15° requires 2% heterogeneity at the top of the mantle (1.5% is too small: Case F). Given the high crustal heterogeneity, the relatively small variance at teleseismic distances is best fit by the velocity standard deviation below 210 km depth and setting it to zero below the 410-km discontinuity. In other words, the observations suggest that velocity heterogeneity in the Earth is highly concentrated at shallow depths.

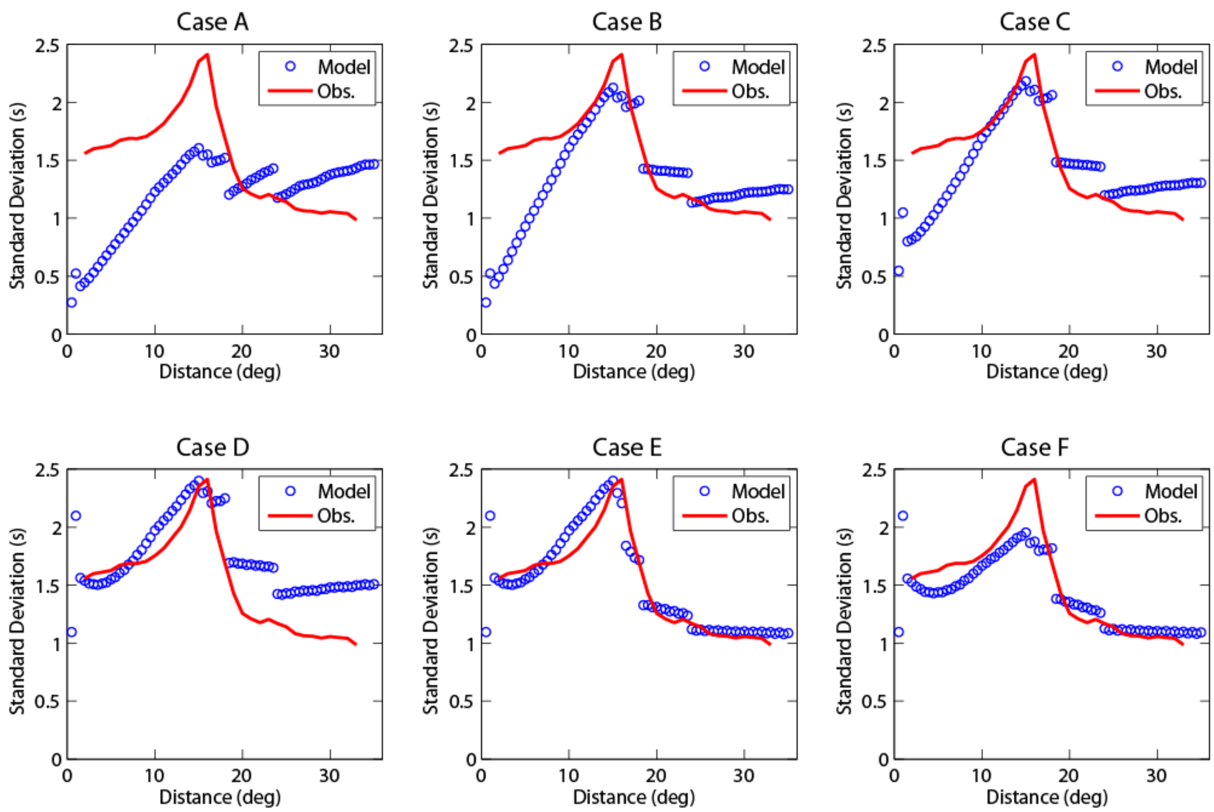


Figure 5. Standard deviation of travel-time model error versus epicentral distance for six cases of geostatistical parameters. The cases differ in the velocity standard deviations assigned to the crust and mantle, as enumerated in Table 1. The horizontal correlation length was set to 300 km; the vertical correlation length was 17.5 and 60 km in the crust and mantle, respectively.

CONCLUSIONS AND RECOMMENDATIONS

The main objective of our research project is to develop and apply meaningful measures of the predictive capabilities of 3-D Earth models. As part of this work, we are performing a comprehensive and methodical evaluation of a set of regional velocity models for Asia based on data misfit and event mislocation metrics. These metrics are already generally accepted as informative, if not complete, measures of the prediction capability of Earth models. We are currently working with five 3D regional velocity models of the crust and mantle and will develop our techniques so that they are easily applicable to others that become available.

In another task we are developing a new method to evaluate models based on the analysis of their uncertainty. We are currently formulating this estimation problem in a more formal way and reviewing statistical methodologies for variance estimation that may apply to the problem. To date, we have formulated theoretical solutions in terms of maximum-likelihood and Bayesian variance estimation but have not identified many practical numerical techniques for computing such solutions. The crux of the difficulty is that travel-time residuals mapped from velocity heterogeneity (via raypath integration) have non-stationary statistics.

REFERENCES

Begnaud, M. L., C. A. Rowe and L. K. Steck (2004). Validating three-dimensional velocity models in China and East-Asia for use in regional seismic event location, *EOS Trans. AGU*, Fall Meet. Suppl.

Bondár, I., and K. L. McLaughlin (2009). A new ground truth data set for seismic studies, *Seismol. Res. Lett.* 80: 465–472.

Bondár, I., S. C. Myers, E. R. Engdahl, and E. A. Bergman (2004). Epicentre accuracy based on seismic network criteria *Geophys. J. Int.* 156: 483–496.

Flanagan, M. P., S. C. Myers, and K. D. Koper (2007). Regional travel-time uncertainty and seismic location improvement using a three-dimensional a priori velocity model, *Bull. Seismol. Soc. Am*. 97: 804–825.

Murphy, J. R., W. Rodi, M. Johnson, D. D. Sultanov, T. J. Bennett, M. N. Toksöz, et al. (2005). Calibration of International Monitoring System (IMS) stations in Central and Eastern Asia for improved seismic event location, *Bull. Seismol. Soc. Am*. 95: 1535–1560.

Nataf, H.-C., and Y. Ricard (1996). 3SMAC: an a priori tomographic model of the upper mantle based on geophysical modeling, *Phys. Earth and Planet. Int.* 95: 101–122.

Pasyanos, M. E., W. R. Walter, M. P. Flanagan, P. Goldstein, and J. Bhattacharyya (2004). Building and testing an a priori geophysical model for Western Eurasia and North Africa, *Pure Appl. Geophys*. 161: 234–281.

Ritzwoller, M. H., N. M. Shapiro, M. P. Barmin and A. L. Levshin (2002). Global surface wave diffraction tomography, *J. Geophys. Res*., 107**,** doi: 10.1029/2002JB001777.

Ritzwoller, M. H., N. M. Shapiro, A. L. Levshin, E. A. Bergman, and E. R. Engdahl (2003). Ability of a global three-dimensional model to locate regional events, *J. Geophys. Res*. 108: 2353, doi:10.1029/2002JB002167.

Rodi, W. L., and S. C. Myers (2008). Modeling travel-time correlations based on sensitivity kernels and correlated velocity anomalies, in *Proceedings of the 30th Monitoring Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-08-05261,Vol. 1, pp. 475–484.

Reiter, D. T., and W. L. Rodi (2009). Validated 3D velocity models in Asia from joint regional body- and surface-wave tomography, *Final Report*, Contract FA8718-04-C-0027, AFRL-RV-HA-TR-2009-1009.

Ruppert, S. D., D. A. Dodge, A. B. Elliott, M. D. Ganzberger, T. F. Hauk, and E. M. Matzel (2005). Enhancing seismic calibration research through software automation and scientific information management, in *Proceedings of the 27th Seismic Research Review: Ground-based Nuclear Explosion Monitoring Technologies*, LA-UR-05-6407, Vol. 2, pp. 937–945.

Stevens, J. L., H. Xu, J. W. Given, and G. E. Baker (2008). Development of surface wave dispersion and attenuation maps and improved methods for measuring surface waves, *Final Report*, Contract FA8718-05-C-0023, AFRL-RV-HA-TR-2008-1106.

Wessel, P., and W. H. F. Smith (1995). New Version of the Generic Mapping Tools Released, *EOS Trans. AGU*, 76: 329.