# NON-LINEAR SEISMIC VELOCITY ESTIMATION FROM MULTIPLE WAVEFORM FUNCTIONALS AND FORMAL ASSESSMENT OF CONSTRAINTS

Mrinal K. Sen[1], Jay Pulliam[2], Utpal Dutta[3], Ranjana Ghosh[1], Rob J. Mellors[4], and Michael E. Pasyanos[4]

University of Texas at Austin[1], Baylor University[2], University of Alaska, Anchorage[3], and Lawrence Livermore National Laboratory[4]

Sponsored by the National Nuclear Security Administration

Award Nos. DE-AC52-09NA29327 and DE-AC52-07NA27344
Proposal No. BAA09-58

## ABSTRACT

Our goal is to jointly model surface wave dispersion, receiver functions, and characteristics of the waveform that appear in a window around the direct S arrival. Optimization is done using a three part objective function given by

$$Error = a_1 \| \mathbf{d}^{obs}_{SPL} - \mathbf{d}^{syn}_{SPL} \|_p + a_2 \| \mathbf{d}^{obs}_{dispersion} - \mathbf{d}^{syn}_{dispersion} \|_p + a_3 \| \mathbf{d}^{obs}_{receiverfucntion} - \mathbf{d}^{syn}_{receiverfucntion} \|_p$$

where $a_1$, $a_2$ and $a_3$ are the relative weights of different data and $p$ is the norm used to measure data misfit. The algorithm is very general in that the weights can change with iteration and $p$ can take different values for the three different parts of the objective function. Note that the same model is used to compute synthetic waveforms, surface wave dispersion and receiver functions. Forward modeling is done using a reflectivity algorithm and optimization is carried out using an algorithm called Very Fast Simulated Annealing (VFSA). The code runs efficiently on multiple processors using Message Passing Interface routines. In the project's first year we coded the algorithm, validated the implementation, and performed synthetic tests to confirm the method's effectiveness and tractability and to demonstrate the usefulness of associated model assessment tools.

In the past year we conducted a broad search for shear-coupled long period (SPL) waveforms among broadband seismic data available to us from the Middle East and discovered that this phase is quite rare in this region. To generate SPL a low-velocity zone is required beneath the Moho; SPL is therefore observed most commonly in shield regions. Further, sources of large-magnitude, deep-focus earthquakes concentrate at epicentral distances of ~70°, distances for which the SPL wavetrain is interrupted after one or two cycles by the arrival of strong SKS phases. Nevertheless, the Sp and SsPmP phases are occasionally prominent and modeling their characteristics, including amplitude and arrival times relative to the direct S phase, can also constrain structural models of the Earth's crust and uppermost mantle.

We will show results of both synthetic tests and applications to real broadband data from the region and use assessment tools, including Posterior Probability Density (PPD) and model parameter correlation matrices, to show the advantages of joint modeling via forward modeling with simulated annealing. These statistical tools allow us to evaluate the relative strength of constraints placed on model parameters by each type of data in addition to the portions of the model that are well, or poorly, constrained. In addition, we report on the development of a new Markov Chain Monte Carlo (MCMC) method called Hamiltonian Monte Carlo (HMC) approach for joint inversion of surface wave dispersion and receiver function data.

## OBJECTIVES

Our goal is to develop optimal procedures for the use of multiple datasets. Due to the inherent variability, inconsistency, and peculiarities of disparate datasets and the well-known nonlinearity and non-uniqueness associated with geophysical modeling, such procedures must include methods for evaluating the performance and contribution of each dataset to the final results.

We use quantitative assessment tools and a formal Bayesian approach to explore and evaluate each step of the modeling process, rather than to simply toss all constraints into a simultaneous fitting procedure to find the single solution that satisfies particular criteria. The procedure we propose is best characterized as velocity analysis via optimization; it is analogous to velocity analysis in exploration seismology, rather than "inversion". It provides quantitative error measures of structural parameter estimates that can then be translated to earthquake location errors, and thus guide seismologists toward the most effective and efficient ways to improve model reliability.

## RESEARCH ACCOMPLISHED

We have added an additional functional, S-wave Receiver Functions, to the three functionals considered previously. Any of these four (P-wave receiver functions, surface wave dispersion, waveform modeling, and, now, S-wave Receiver Functions) can now be modeled individually or jointly via global optimization. Our method also computes the Posterior Probability Density (PPD) function and correlation matrix, to evaluate the uniqueness of the resulting models, and the trade-offs between individual model parameters therein.

The addition of S-wave Receiver Functions is significant for the region we are targeting (the Middle East) because our exhaustive search of available data turned up very few examples of shear-coupled PL and only a handful of examples of Sp and SsPmP phases with good signal-to-noise characteristics. We have shown previously that shear-coupled PL provides helpful constraints on P velocities in the crust and uppermost mantle and on impedance contrasts across the Moho. Sp and SsPmP can also help constrain the depth to Moho and P velocities in the crust. The fact that these latter two have low SNR suggests that forming S-wave receiver functions and modeling those, perhaps after stacking, instead of the raw waveforms themselves, may be a fruitful approach. Deconvolving the vertical component from the radial, for example, will also remove systematic sources of noise and stacking can be expected to boost the SNR.

Not finding SPL, however, is a significant drawback. The reason for the dearth of SPL, particularly, probably has to do with upper mantle structure. SPL typically requires a low-velocity zone beneath the Moho for strong propagation and such LVZs are more often found in older cratons, not in more tectonically active regions such as the Middle East. For example, we previously applied the code to determine the crust and upper mantle structure beneath permanent broadband seismic stations in Africa using teleseismic earthquakes (M 5.5-7.0 and 200-800 km focal depth). We modeled the S, SP, SsPmP, and shear-coupled PL waves from these earthquakes and our P- and S-wave velocity models compare well with, and in some cases improve upon the models obtained from other existing methods. We obtained P- and S-wave velocities simultaneously and our use of the shear-coupled PL phase wherever available improved constraints on the models of the lower crust and upper mantle (Gangopadhyay et al., 2007).

In the past year we did, however, find a few examples in the broader region (Ethiopia and Cyprus) and our modeling efforts for stations in those regions are shown here. Figure 1 displays the study area and locations of two stations DESE and GVD. We show results from full waveform inversion at these two stations in Figures 2 and 3. In each figure we show data fit, best model, marginal PPD and correlation plots. Note that at station DESE, we obtain good fit to the data and well-resolved layer parameters. On the other hand, at station SVD, data fit is poor and the correlation matrix shows significant off-diagonal values, indicating significant tradeoff between model parameters. In particular, we notice tradeoff between velocities and thicknesses of some of the layers.

### Joint Modeling of Multiple Datasets

There are several advantages to jointly modeling multiple datasets. First, each data functional has unique sensitivities to Earth structure. For example, receiver functions are primarily sensitive to shear wave velocity contrasts and vertical travel times while surface wave dispersion measurements are sensitive to vertical shear wave velocity averages (Julia et al., 2000). Full waveform modeling of S, Sp, SsPmP phases, in contrast, are more sensitive to compressional wave velocity contrasts and vertical travel times; adding SPL to the modeling improves senstivity to the uppermost mantle shear wave structure and to velocity contrasts across the Moho (Gangopadhyay et al., 2007). The amplitudes and signal-to-noise characteristics of waves that produce these data functionals depend on several factors, including epicentral distance, event focal depth, fault mechanism and radiation patterns, source time function, properties of the intervening Earth structure (including attenuation, low velocity zones, velocities, heterogeneity, and anisotropy), and characteristics of the recording seismometer. Since regions of high seismicity

are highly restricted, many stations are not well situated to record appropriate events. An algorithm that incorporates multiple data types is more widely applicable than one that relies solely on a single type.

However, with these advantages come some disadvantages. For example, combining disparate data types requires great care in their treatment and assessment (Roy et al., 2005). Benefits of additional data may be null if the method used to model them preferentially fits one type. Or, worse, minimizing an inappropriate criterion in conjunction with incompatible data may "split the difference" between them to choose a model that is wholly inaccurate and inappropriate for its intended purpose. If the model is to be used for regional or local earthquake locations, for example, it would be a mistake to rely on the best fit to surface wave dispersion. The judgment and experience of seismologists who keep a clear eye on their goal is critical, and this experience must be combined with rigor in the computational modeling. This experience and judgment can be incorporated after the fact, as is sometimes the case with, for example, the smoothing applied to 3D tomographic models but it is usually better to acknowledge the "prior" explicitly at the outset. This is one advantage of the Bayesian approach that we propose here. While priors are sometimes referred to pejoratively as "bias", their explicit statement during the formulation of the modeling algorithm forestalls serious criticism and enables a clear, quantitative discussion of "bias".

Next, as a general rule of thumb, a greater number of data functionals incorporated into modeling will result in a broader range of model parameter sensitivities, and it will be less likely that a linear inversion approach will be adequate. This is unfortunate because linear approaches are much more tractable and straightforward than nonlinear methods. But only a broad search of the model space will demonstrate whether a linear approach is valid. Non-linear global optimization algorithms require no change in the algorithm to include multi-part objective function with different norms. A variety of methods for nonlinear inversion are now available and the only real cost for conducting a thorough search and finding the single best-fitting model is in computational effort and complexity. The downside is that theoretically exact methods for assessing model uncertainties and model reliability are not generally tractable, and approximate methods result in significantly greater cost and complexity than is required to find the best-fitting model alone. Nevertheless, our previous work has demonstrated that useful methods are indeed tractable, and the method we propose will add only marginal increases in computation time.

While surface wave dispersion and receiver functions have been modeled jointly (Ammon et al., 2005; Ammon et al., 2004; Cakir and Erduran, 2004; Chang et al., 2004; Dugda and Nyblade, 2006; Herrmann et al., 2001; Julia et al., 2000; Julia et al., 2005; Lawrence and Wiens, 2004; Ozalaybey et al., 1997; Tkalcic et al., 2006), no study has, to our knowledge, incorporated waveform constraints such as S, Sp, SsPmP, and Shear-coupled PL. Further, none has conducted a thorough, nonlinear assessment of the constraints provided by each functional.

**Multi-Objective Optimization (MOO) for Multiple Datasets**

The primary goal of our current project is to develop a tool for estimating crustal structure that can explain surface wave dispersion, receiver function and full waveform modeling. Thus our optimization algorithm is expected to minimize misfit of each of the datasets using three different objective or cost functions. The process of optimizing systematically and simultaneously a collection of objective functions is called multi-objective optimization (MOO) or vector optimization. The general MOO is posed as follows:

$$\underset{\mathbf{m}}{\textit{Minimize}}\ \boldsymbol{F}(\mathbf{m}) = \left[\boldsymbol{F}_1(\mathbf{m}), \boldsymbol{F}_2(\mathbf{m}), ..., \boldsymbol{F}_{\boldsymbol{k}}(\mathbf{m})\right]^{\boldsymbol{T}},$$

where **m** is the model vector, and *k* is the number of objective functions. The objective function is minimized subject to some constraints. In our application, we use standard constraints such as bounds and negativity.

Contrary to single-objective optimization, an MOO may be considered more of a concept than a definition. Typically, with noisy data there is no single global solution and it is often necessary to determine a set of points that all fit a predetermined definition for an optimum. This is commonly done using what is known as Pareto optimality (Pareto, 1906) defined as follows.

*Definition*: A model **m*** belonging in the model space is said to be Pareto optimal if and only if there does not exist any other model **m** in the model space such that $\boldsymbol{F}(\mathbf{m}) \le \boldsymbol{F}(\mathbf{m}^*)$, and $\boldsymbol{F}_{\boldsymbol{i}}(\mathbf{m}) \le \boldsymbol{F}_{\boldsymbol{i}}(\mathbf{m}^*)$ for at least one function.

We often use a *global criterion* in search for models belonging to Pareto set, which is a scalar function that mathematically combines multiple objective functions; it may or may not involve preference of one or the other.

**Objective Function**

We use the following objective function which outputs a single scalar value.

$$Error = a_1 \| \mathbf{d}^{obs}_{SPL} - \mathbf{d}^{syn}_{SPL} \|_p + a_2 \| \mathbf{d}^{obs}_{dispersion} - \mathbf{d}^{syn}_{dispersion} \|_p + a_3 \| \mathbf{d}^{obs}_{receiverfucntion} - \mathbf{d}^{syn}_{receiverfucntion} \|_p .$$

This general objective function allows for different measures of error and different weights to the datasets. There are at least three fundamental issues that we need to address

1. What measures of error do we use? Can they be different for different datasets?
2. Can we use simple differences?
3. How do we choose the three weights?

Question (3) does not generally have a straightforward answer. The weights have to be determined by trial and error. One important consideration in defining the global objective function for MOO is to ascertain that no individual objective function introduces more impact than the others unless so desired. To address this, we use the following normalized objective function defined in Sen and Stoffa (1996)

$$\boldsymbol{Error} = 1 - \frac{2\sum \left| \boldsymbol{d}^{\boldsymbol{i}}_{\boldsymbol{obs}} - \boldsymbol{d}^{\boldsymbol{i}}_{\boldsymbol{syn}} \right|^{\alpha}}{\sum \left| \boldsymbol{d}^{\boldsymbol{i}}_{\boldsymbol{obs}} + \boldsymbol{d}^{\boldsymbol{i}}_{\boldsymbol{syn}} \right|^{\alpha} + \sum \left| \boldsymbol{d}^{\boldsymbol{i}}_{\boldsymbol{obs}} - \boldsymbol{d}^{\boldsymbol{i}}_{\boldsymbol{syn}} \right|^{\alpha}}$$

where the sum is taken over all the data points and the parameter $\alpha$ is equivalent to norm.

Initially we chose $\alpha$=1 for all three objective functions. The objective function corresponding to the surface wave dispersion is smoothly varying, unlike those for the functionals thatdepend on the waveforms, and we therefore chose $\alpha$=0.5 to introduce more sensitivity. With these choices of $\alpha$ values and normalized objective functions as defined in the above equation, we employ equal weights to all three parts of the objective function.

**Optimization and Parallelization**

Having defined the objective function, we employ Very Fast Simulated Annealing (VFSA) for optimization (Sen and Stoffa, 1996; Pulliam and Sen, 2005). Note that VFSA causes no difficulties since we do not require computing derivatives. As used in Pulliam and Sen (2005), we use a parallel optimization code for MOO as well. Note that the most expensive computation element is that of synthetic waveform by the reflectivity method and the reflectivity computations are therefore distributed over multiple nodes. Since surface wave dispersion and receiver function computations are fairly fast, we distribute those to one node each and thus achieve reasonable load balance.

**Incorporation of S-wave Receiver Functions into MOO Modeling Package**

Modeling S-to-P (commonly called "S-wave") receiver functions (SRFs) can, in principle, incorporate many of the same constraints found in the broadband waveform windowed around the direct S arrival, including Sp and SsPmP. Shear-coupled PL, however, would not be included in S-wave receiver functions, so one would forego the constraints they impose on the lower crust, uppermost mantle, and the impedance contrast across the Moho. SPL is rarely observed in the Middle East, however, and an additional complication arises because the epicentral distance of the majority of events appropriate for such modeling (M 5.5-7.0 and 200-800 km focal depth,) are located at epicentral distances near 70°, at which distance SKS interferes with the SPL wavetrain. We prefer SPL observations to be in the range $30° \leq \Delta \leq 60°$ so we can see, and model, several cycles of SPL.

S-wave receiver functions, on the other hand, have reverse moveout compared to P-wave receiver functions, because S-to-P converted waves arrive before direct S. This, combined with the minimization of source effects, due to deconvolution, and the possibility of stacking to improve signal-to-noise characteristics make SRFs attractive options for modeling jointly with the other three functionals, or in place of the waveform fitting if necessary. Their arrivals avoid the interference by SKS described above and thus extend the range of useable epicentral distances and thus the quantity of useable data. Adding SRFs as and additional modeling option therefore produces a more robust and more broadly applicable modeling package.

We are in the process of incorporating S-wave receiver function modeling into our code and now have a subroutine that computes the SRF from a stack of layers and complementary pre-processing of observed data to be modeled. When this addition is finalized we will be able to use the modeling assessment tools developed previously to quantify the strength of constraints imposed by each data functional on the model parameters.

**Development of a Hamiltonian Monte Carlo Modeling Technique**

We completed a new code that utilizes a combination of local and global searches, using the VFSA algorithm. The new code is computationally faster than a standalone VFSA and overcomes the primary limitation of the gradient-based inversion algorithm, that it fails to search the model space broadly, while benefiting from gradient-based method's greatest asset, its fast convergence to a solution. The approach, called Hamiltonian Monte Carlo Method (HMC), is a rigorous sampling algorithm that is able to draw samples from the PPD efficiently. The current version of the code includes surface wave dispersion and receiver functions only. The algorithm combines the random walk with a deterministic search based on local gradient of the cost function.

Results from HMC inversion of surface wave dispersion and receiver function data are displayed in Figure 4. We generated synthetic surface wave dispersion and P-wave receiver functions for a simple six layer model. This dataset is then used as observation for optimization. Our models search parameters are Vs, Poisson's ratio, density, and layer thickness for the six layers. We use search bounds for the model parameters as indicated in Figure 4(c).

Top left panel shows a fit between synthetic receiver function data and the data generated by the best fit model. Top right panel shows a fit between synthetic dispersion data (used as observation) and the data generated by the best fit model. Bottom left panel displays marginal PPDs for Vp and Vs – it also shows the search window used in this problem. The panel at bottom right shows a plot of posterior correlation matrix. The correlation matrix shows several significant off-diagonal elements. In particular, note high negative correlation between velocity and thickness of one of the shallow layers indicating that this layer is poorly resolved by the data

## CONCLUSIONS AND RECOMMENDATIONS

We have developed a parallel VFSA-based multi-objective optimization code that can be used to obtain crustal velocity structures by modeling broadband waveform, receiver function, and surface wave dispersion data. The code has been applied successfully to synthetic examples and to stations from the broader Middle East region.

In the past year the code has been extended to include a hybrid global/local search algorithm that combines the advantages of both methods: first searching the model space broadly in order to find the single best-fitting solution and also characterize the strength of constraints imposed by the data on model parameters and then refining the best-fitting solution efficiently via local search. A second innovation, not yet complete, is the addition of S-wave receiver function modeling to the MOO package. The option to model S-wave receiver functions can be used in addition to waveform modeling, or in its place when, for example, shear-coupled PL phases do not appear (so the constraints offered by this phase are not available) , epicentral distances are near 70° (so SKS interferes with SPL), source characteristics (such as moment tensor or depth or time function) are inaccurate, or the signal-to-noise ratio is poor.

In the coming year we will complete the S-wave receiver function modeling, extend the HMC method to real data and apply the MOO modeling package to more realistic synthetics and data recorded in the Middle East.

## REFERENCES

Ammon, C.J. (1991). The isolation of receiver effects from teleseismic P waveforms. *Bull. Seismol. Soc. Am*. 81: 2504–2510.

Ammon, C.J. et al. (2005). Simultaneous Inversion of Receiver Functions and Surface-Wave Dispersion Measurements for Lithospheric Structure Beneath Asia and North Africa, 27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies, LA-UR-05-6407, Vol.1, pp. 3–12.

Ammon, C.J. et al. (2004). Simultaneous Inversion of Receiver Functions, Multi-mode Dispersion, and Travel-time Tomography for Lithospheric Structure Beneath the Middle East and North Africa, in *Proceedings of the 26th Seismic Research , Review - Trends in Nuclear Explosion Monitoring*, LA-UR-04-5801 Vol. 1, pp. 17–28.

Cakir, O. and Erduran, M. (2004). Constraining crustal and uppermost mantle structure beneath station TBZ (Trabzon, Turkey) by receiver function and dispersion analyses. *Geophys. J. Int*. 158(3): 955–971.

Chang, S.-J., Baag, C.-E., and C. Langston (2004). Joint analysis of teleseismic receiver functions and surface wave dispersion using the genetic algorithm. *Bull. Seismol. Soc. Am*. 94(2): 691–704.

Dugda, M. and Nyblade, A.A. (2006). New constraints on crustal structure in eastern Afar from the analysis of receiver functions and surface wave dispersion in Djibouti. *Geological Society of London Special Publications*, 259: 239–251.

Gangopadhyay, A., Pulliam, J. and Sen, M.K., 2007. Waveform Modeling of Teleseismic S, SP, SsPmP, and Shear-Coupled PL Waves for Crust and Upper Mantle Velocity Structure Beneath Africa. *Geophys. J. Int*., 170: 1210–1226.

Herrmann, R.B. and Ammon, C.J. (2002). Computer Programs in Seismology

Herrmann, R.B., Ammon, C.J. and Julia, J. (2001). Application of joint receiver-function surface-wave dispersion for local structure in Eurasia, in *Proceedings of the 23rd Seismic Research Review: Worldwide Monitoring of Nuclear Explosions,* LA-UR-01-4454, pp. 46–54.

Julia, J., Ammon, C.J., Herrmann, R.B. and Correig, A.M. (2000). Joint inversion of receiver function and surface wave dispersion observations. *Geophys. J. Int*., 143(1): 99–112.

Julia, J., Ammon, C.J. and Nyblade, A.A. (2005). Evidence for mafic lower crust in Tanzania, East Africa, from joint inversion of receiver functions and Rayleigh wave dispersion velocities, *Geophys. J. Int*. 162: 555–569.

Lawrence, J.F. and Wiens, D.A. (2004). Combined Receiver-Function and Surface Wave Phase-Velocity Inversion Using a Niching Genetic Algorithm: Application to Patagonia. *Bull. Seismol. Soc. Am*. 94(3): 977–987.

Ozalaybey, S., Savage, M.K., Sheehan, A.F., Louie, J.N. and Brune, J.N. (1997). Shear-wave velocity structure in the northern Basin and Range province from the combined analysis of receiver functions and surface waves. *Bull. Seismol. Soc. Am.* 87(1): 183–199.

Pareto, V. 1906: Manuale di Economica Politica, Societa Editrice Libraria. Milan; translated into English by A.S. Schwier as Manual of Political Economy, edited by A.S. Schwier andA.N. Page, 1971. New York: A.M. Kelley.

Pulliam, J. and Sen, M.K., 2005. Assessing Uncertainties in Waveform Modeling of the Crust and Upper Mantle, in *Proceedings of the 27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-05-6407, Vol. 1, pp. 152–160.

Pulliam, J., Sen, M.K. and Gangopadhyay, A. (2006). Determination of Crust and Upper Mantle Structure Beneath Africa Using a Global Optimization-Based Waveform Modeling Technique,in *Proceedings of the 28th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies*, LA-UR-06-Vol. 1, pp. 196–208.

Roy, L., Sen, M.K., Stoffa, P.L., McIntosh, K. and Nakamura, Y. (2005). Joint inversion of first arrival travel time and gravity data. *J. Geophys. Engin*., 2: 277–289.

Sen, M.K. and Stoffa, P.L. (1996)a. Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. *Geophys. Prospect*., 44: 313–350.

Tkalcic, H. et al., 2006. A multistep approach for joint modeling of surface wave dispersion and teleseismic receiver functions: Implications for lithospheric structure of the Arabian Peninsula. *J. Geophys. Res*., 111(B11311): 10.1029/2005JB004130.

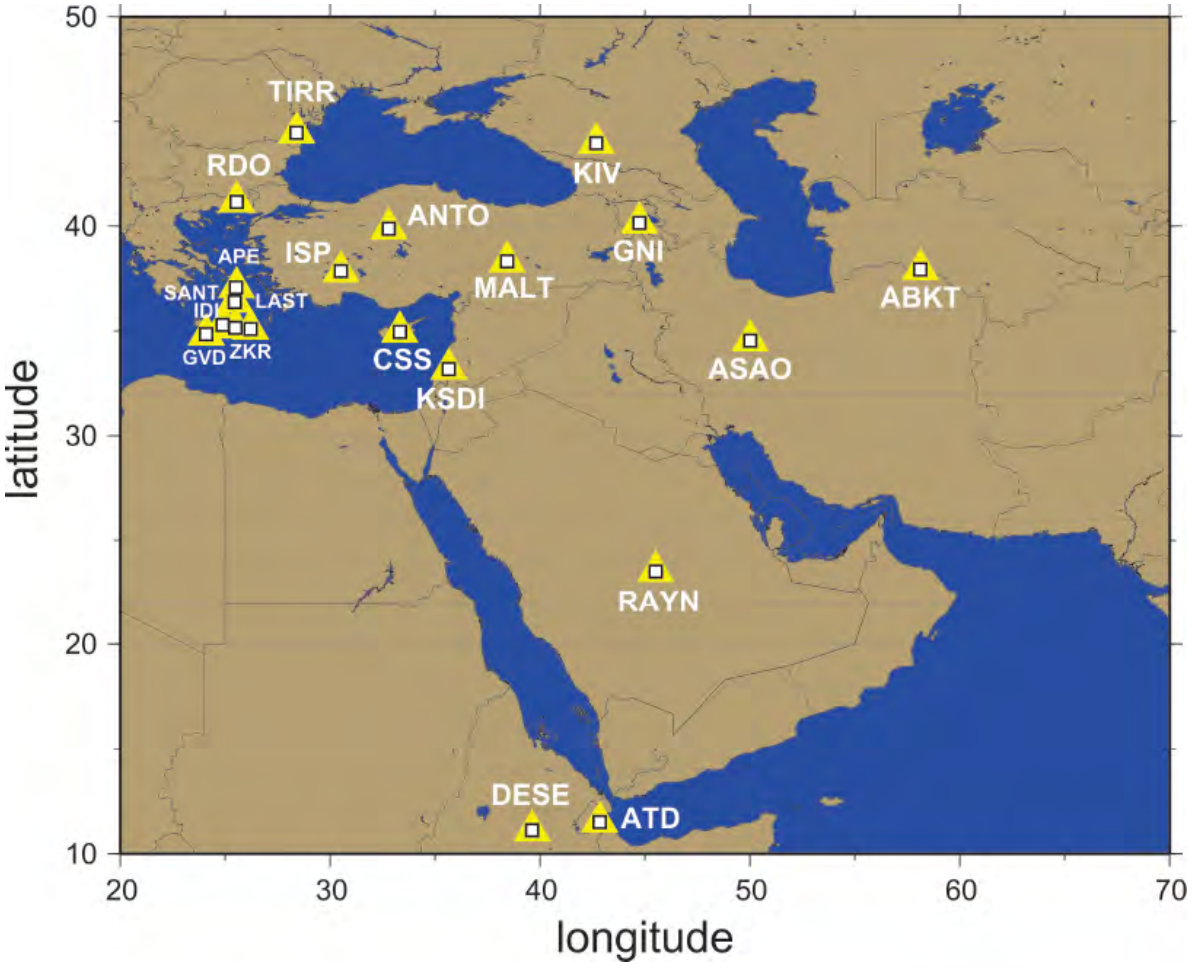


**Figure 1. Map of broadband seismic stations in the Middle East.**

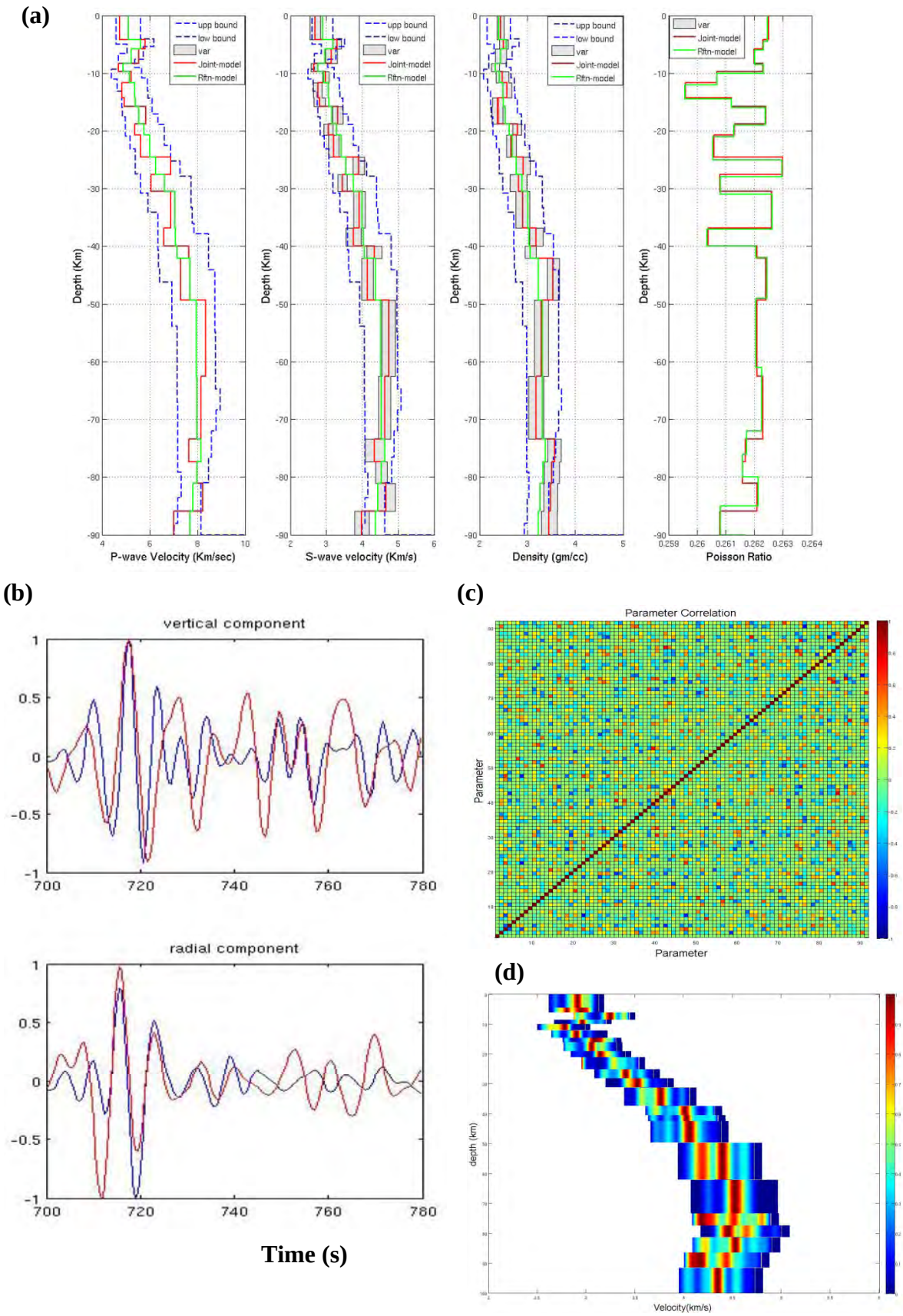


**Figure 2. Station DESE: (a) Model obtained from inversion, (b) waveform fitting (data-blue, synthetic-red), (c) correlation matrix, (d) PPD plot (S-wave velocity).**

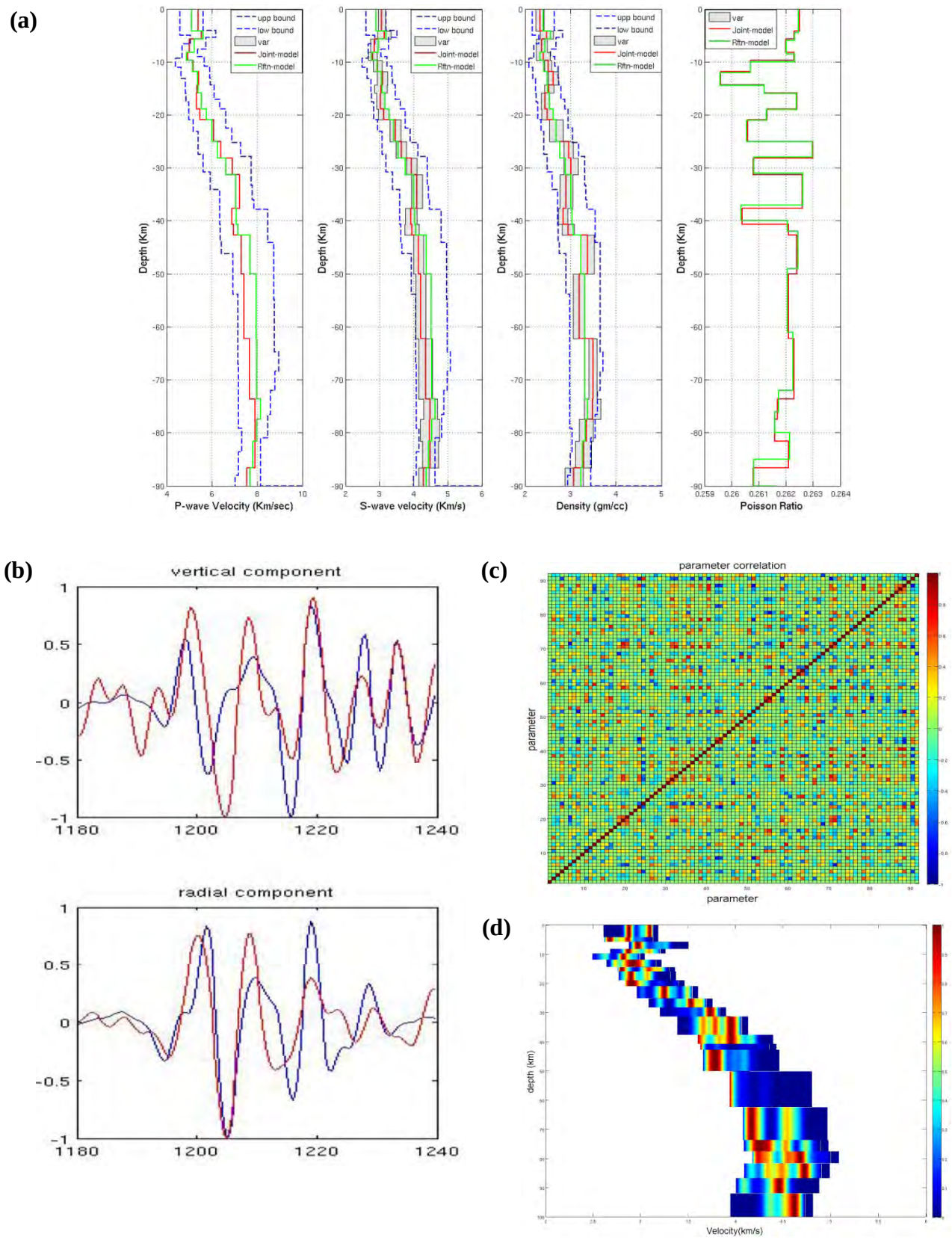


**Figure 3. Station GVD: (a) Model obtained from inversion, (b) waveform fitting (data-blue, synthetic-red), (c) correlation matrix, (d) PPD plot (S-wave velocity).**

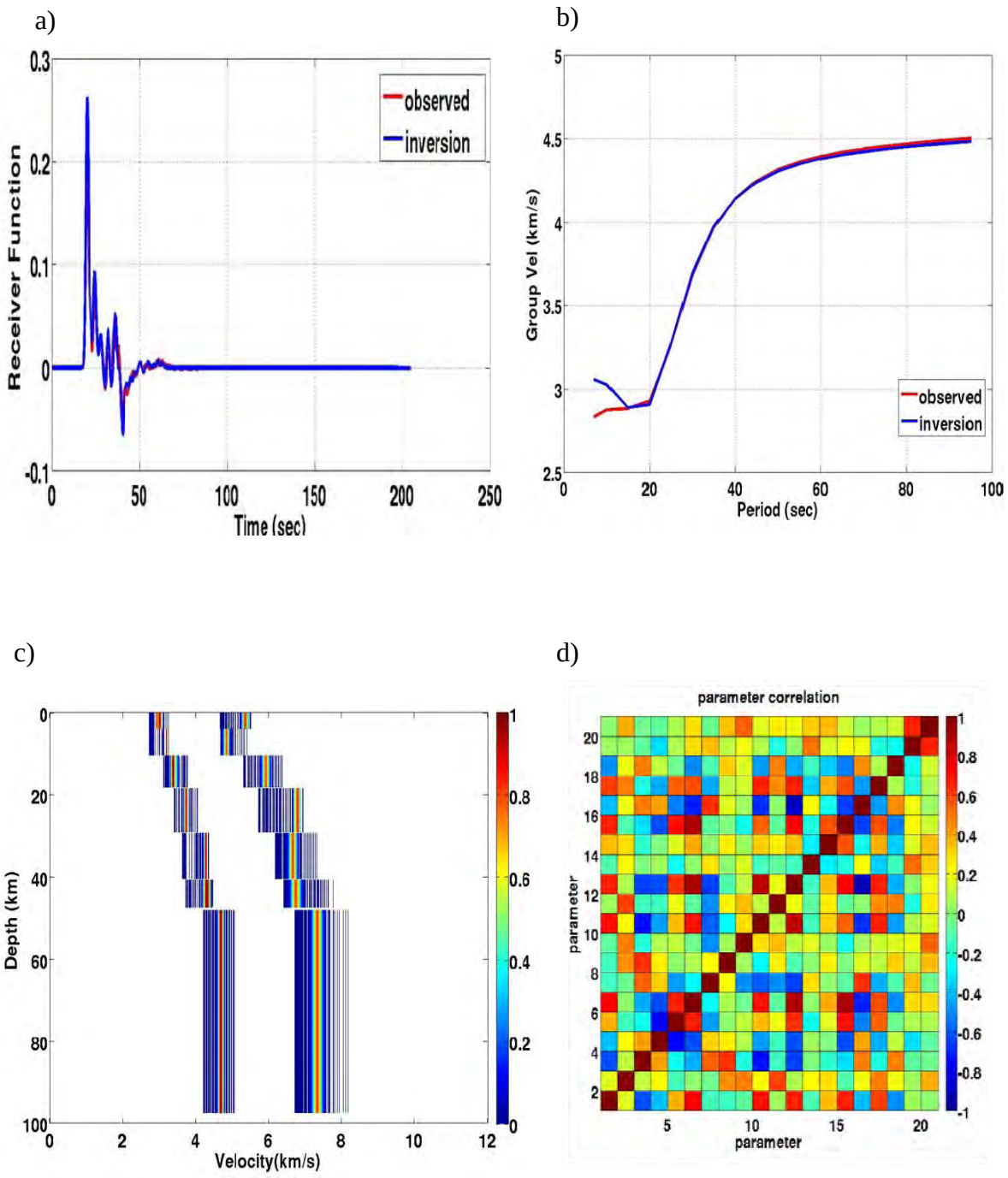


**Figure 4. Results from HMC sampling application in joint inversion of surface wave dispersion and receiver function. Top left panel (a) shows a fit between synthetic receiver function data (used as observation) and the data generated by the best fit model. Top right panel (b) shows a fit between synthetic dispersion data (used as observation) and the data generated by the best fit model. (c) The panel at bottom left displays marginal PPDs for Vp and Vs ; it also shows the search bounds used in this optimization problem. (d) The panel at bottom right shows a plot of posterior correlation matrix.**