Stochastic modeling of the Earth's crust and upper mantle with Genetic Algorithms: Binary or real-valued algorithms?

Boomer, KB
474 Olin Building
Bucknell University
Lewisburg, PA, USA 17837
kb.boomer@bucknell.edu

Brazier, Richard
Penn State University DuBois Campus
174 Smeal Building
College Place
DuBois, PA, USA 15801
rab27@psu.edu

Introduction

A primary focus of seismic analysis is the accurate modeling of the geologic composition of the Earth's crust and upper mantle as well as the speed with which seismic waves travel through these layers. Velocity modeling is a multi-objective optimization problem with competing objectives typically approached via a single or joint inversion of objective(s), resulting in a single estimated velocity structure.

Stochastic modeling of the velocity structure is a relatively new approach in seismology, and multi-objective genetic algorithms offer a stochastic forward modeling approach, resulting in a set of plausible solution velocity models. Benefits include the ability to increase the number of objectives at relatively low cost and to incorporate multiple types of data including magnetotelluric and gravity data; this is difficult to accomplish with a joint inversion.

When using a genetic algorithm one choice to be made is whether to represent the velocity model in terms of a binary digit, as has been done historically, or as a real-valued variable. We explore the how choice of a binary coding or real-valued representation of the velocity structure impacts the resulting models.

Seismic Application

Seismology is the study of how waves (surface waves, body waves, etc.) travel through the Earth. The velocity structure, describing the speed with which seismic waves travel, provides insight to the Earth's dynamics, allowing short-term and long-term risk assessment for future earthquakes. Each geologic layer has its own velocity (in km/s), correlated with the velocity in the adjacent geologic layers and varying with layer depth (Figure 1). The velocity model is a solution vector of velocity estimates for each layer in the region of interest. In this paper, the region of interest is the Archaean Kaapvaal Craton found in southern Africa.

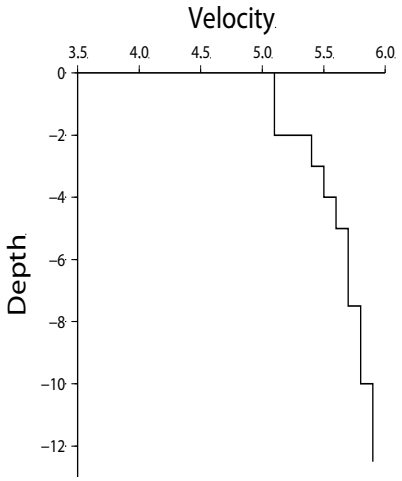

Figure 1: A typical velocity structure

In general, data come from seismic events (typically earthquakes or explosions) as travel times from the source event to the recording seismograph, as the seismic waves pass through the heterogeneous crust and upper mantle. Waves traveling around the surface of the Earth get dispersed into different frequencies, traveling at different speeds. A surface wave dispersion curve plots absolute velocities (averaged across geologic layers) by wave frequency (Figure 2). De-convolving the vertical and horizontal ground motion from a wave

isolates where seismic waves convert from compressional to shear waves, indicating a change in the velocity structure. These de-convolutions, classified by wave frequency and travel path, are plotted as a function of time, which is termed a receiver function. A receiver function (Figure 2) highlights sharp contrasts in velocity beneath the seismograph location, relative to the first signal that arrives. Four receiver functions were used in these analyses. The first (i.e. top) and second receiver functions measured higher frequency waves than the third and fourth receiver functions. The first and third receiver functions traced wave paths closer to the source whereas the second and fourth receiver functions traced wave paths that were further from the source.

In the multiple objective optimization problem of modeling the velocity structure, the objective functions are the minimization of the deviation of the estimated surface wave dispersion curve and the receiver functions from the observed data. The combination of these multiple objectives can determine the layered velocity structure (Julià *et al.* 2000).

The Genetic Algorithm

The genetic algorithm (GA) is a stochastic search and optimization method based on the notion of evolution, in which initial solutions evolve toward the ideal solution (De Jong *et al.*, 1997). The algorithm begins with an initial *population* of *individuals* (i.e. initial velocity models) in which each initial velocity model $\mathbf{X}_i^0, \quad i = 1, \ldots, \mu$, is randomly generated. At each *generation* (i.e. iteration), new individuals are obtained. Using terminology from evolution, the current individuals *mate* (via recombination) and *mutate* to form the next generation. In binary tournament selection, with some probability, two velocity models (now called *parents*) are selected and with a crossover probability p_c, two new individuals (called *children* or *offspring*) are calculated from the parent solutions, as described below. With probability 1-p_c, the two parent solutions enter into the next generation unchanged. Further, with probability p_m, the children mutate. In the case of velocity modeling, the velocity parameter at any one layer may be changed by randomly moving its current numerical value closer to the upper or lower bound of plausible values (provided by the user). The mating and mutation ensure that a diverse population of velocity models is evaluated, enhancing the efficiency of the search mechanism.

In each generation *t*, the individual models $\mathbf{X}_i^{(t)}$, i=1,…, μ, are used to produce a synthetic surface wave dispersion curve and synthetic receiver functions to be compared to the observed data. The *fitness* or *objective* functions, $f_k\left(\mathbf{X}_i^{(t)}\right) = \sum \left| g\left(\mathbf{X}_i^{(t)}\right) - g\left(\mathbf{X}_i^{obs}\right) \right|$, k=1,2,…n, (summed over j=1,2,…,l layers) are assessed using Pareto dominance and optimality. A velocity model solution $\mathbf{X}_i^{(t)}$ is considered better or dominant over another solution $\mathbf{X}_{i'}^{(t)}$ if $f_k\left(\mathbf{X}_i^{(t)}\right) \leq f_k\left(\mathbf{X}_{i'}^{(t)}\right)$, for all k, and, for one objective function, $f_{k'}\left(\mathbf{X}_i^{(t)}\right) < f_{k'}\left(\mathbf{X}_{i'}^{(t)}\right)$.

The Pareto optimal solution is the set of all solutions that are non-dominated. Ranking models relative to the Pareto optimal front, the best models are selected to be parents for the next generation. A new generation is created by recombination and mutation and the best picked from these. The process is continued until the Pareto optimal front no longer changes. This version of the GA is called the Non-dominated Sorting Algorithm II (NSGA-II, Deb *et al.*, 2002).

Comparison of Binary and Real-Valued Representations

A principle decision to be made when using GA, including the NSGA-II algorithm, is whether to represent the solution (e.g. a velocity model) in binary format or using a real-valued representation. GA were initially developed based on a binary representation and expanded to real-valued representation later. The choice of representation impacts several aspects of the algorithm as well as the results. Most published work applying genetic algorithms to geophysical modeling, including velocity modeling, use binary representation. Our interest here is whether the results differ based on the choice of representation.

Generation of the initial population

Using binary coding, the velocity in each layer (as in Figure 1) is represented by a 5 bit binary digit

(termed a *chromosome*). The range of velocities for our model is 3 to 6.2 km/s; a 5 bit chromosome can represent 32 possibilities and thus velocity estimates at a specific layer can change by a step size of 0.1 km/s. In Figure 1, the velocity at the top layer at the crust is 5.1 km/s and is represented by 10101. The initial model randomly draws each bit from a Bernoulli distribution with p=0.5. In the real valued representation, the velocity at each structure is randomly drawn from a Uniform distribution with range 3 km/s to 6.2 km/s.

Recombination and Mutation

In the recombination phase, a proportion p_c of pairs of parents are randomly selected from the population for mating; the remainder passes unchanged to the next generation. Recombination and mutation are calculated differently for binary and real-valued representations. In the real-valued representation, the children are generated as an average of the parents (indexed by i) weighted towards the better solution; each layer j is calculated as:

$$x_{ij}^{(t+1)} = \frac{x_{ij}^{(t)} + x_{i+1,j}^{(t)}}{2} + \alpha^{\beta},\ x_{i+1,j}^{(t+1)} = \frac{x_{ij}^{(t)} + x_{i+1,j}^{(t)}}{2} - \alpha^{\beta}$$

where α^{β} is a random weighted distance from the more dominant parent model. The result is a model with velocities at each layer bounded by the parent velocities at that layer.

In the binary case the recombination is performed within the chromosome by "crossing over" bits. The child begins with the same digits as one of the parent. At randomly selected bit "b", the child crosses over to the other parent to obtain the remaining digits. For example, for parents 01101 (i.e. 4.3 km/s) and 11000 (i.e. 5.4 km/s) and a randomly selected crossover point of 3, the child's velocity is calculated by crossing over the last 3 digits, producing 01000 (3.8 km/s). Unlike the real-value representation, binary recombination can and does produce velocities outside the parents' range.

Mutation is used to diversify the population to prevent getting caught in local minima. Mutation is essential for real-value representation and useful for binary representation. A proportion p_m of the children have every layer perturbed. In binary coding, the method is similar to recombination in that a portion of each chromosome is randomly replaced with digits from a Bernoulli distribution with probability p=0.5. The real-value representation is a random perturbation of the layer value.

Results

We have applied the NSGA II algorithm to the velocity model beneath the BOSA station on the Kaapvaal Craton in southern Africa. The velocity model obtained via a joint inversion (Kgaswane *et al*., 2009) is used as a comparison. Consistent with the Kgaswane model, we modeled the velocity at each of 53 layers. The population size in all runs was μ=50. The recombination probability (p_c) and the mutation probability (p_m) were varied as summarized in Table 1.

Table 1: Summary of conversion results for binary and real-valued representations

p_m	p_c	Binary Representation		Real-Valued Representation	
		Number of Iterations Until Convergence	Percent of Reasonable Objective Functions	Number of Iterations Until Convergence	Percent of Reasonable Objective Functions
0	0.3	4	33%	9	68%
0	0.5	8	16%	7	72%
0	0.8	21	80%	16	70%
0	0.9	34	50%	17	64%
0.02	0.9	75	90%	23	74%
0.05	0.3	54	98%	16	14%
0.05	0.5	No convergence		20	78%
0.05	0.8	85	96%	23	72%
0.05	0.9	75	94%	12	68%

On average, the real valued representation methods converged 2.6 times faster than the binary coding, particularly in the presence of mutation. This is not unexpected as the binary coding has five times as many parameters to estimate. In a 53 layered model, a five bit representation requires 265 estimated values whereas the real-valued representation only requires 53 values.

Within each solution set of 50 models, not all velocity models correspond well to the observed surface wave dispersion curve or the receiver functions. These are the standard comparisons used by geophysicists to evaluate the goodness of fit for estimated velocity models. Such measures evaluate the entire velocity model (as opposed to individual layers) and models with large misfits (objective functions) between the model-predicted and observed data are suspected to be poor and not trustworthy. Examples of large misfits are show in Figure 2b, c. These examples also demonstrate that velocity modeling incorporates competing objectives, for which GAs are well-suited.

Figure 2: Three sets of four receiver functions and one surface wave dispersion curve: (a) both the estimated receiver functions and the surface wave dispersion curve match the observed data well; (b) the surface wave dispersion curve indicates the velocity model overestimated the observed velocity, though the receiver functions fit well; (c) the receiver functions indicate excessive noise, suggesting estimated arrival times may be inaccurate.

For both the binary and real-valued representations, adding a small probability of mutation improved the percent of models with reasonable objective functions, possibly because a larger portion of the model space is searched with mutation. Further, binary representation yielded more reasonable objective functions than the real-valued representations. This can be best seen by the final Pareto optimal sets (Figure 3). The Pareto fronts from a representative run (crossover probability of 0.90, mutation probability of 0.05) show that the binary coding produces a more compact set of objective functions, particularly with respect to the surface wave dispersion curve, representing more accurate velocity estimates (averaged across the 53 geologic layers modeled). Large values for the surface wave dispersion objective function for the real-valued representation suggests a bias in estimated velocities. Receiver functions 1 and 2, describing the high frequency waves, favor the real-value representation, with the largest misfit functions observed for binary models. Receiver functions 3 and 4 favor the binary representation. In particular, receiver function 3 (low frequency, close paths) reveals several extreme real-valued objective function values.

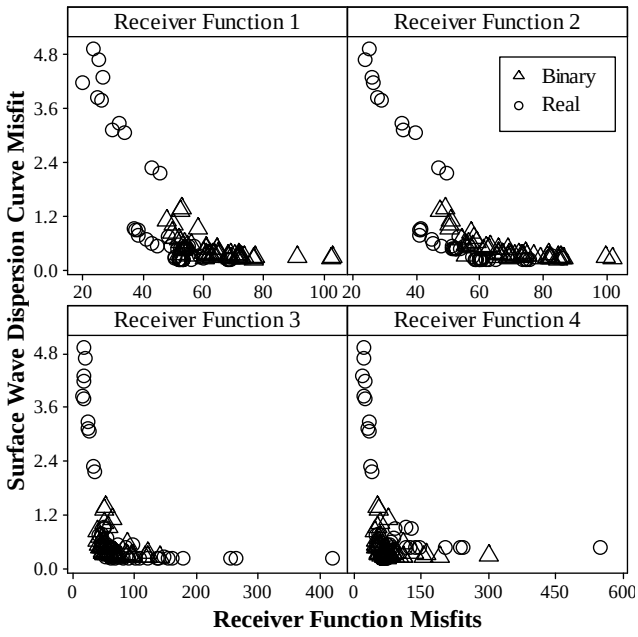

Figure 3: Comparison of the objective functions corresponding to the final Pareto optimal set of velocity models for the binary coding and real-valued representations (p_c=0.9, p_m=0.05)

Finally, the set of 50 solutions in the Pareto optimal set are plotted by depth (km). This comparison shows layer-by-layer variability in velocity estimates. Figure 4 shows an overlay of the 50 solutions using a crossover probability of 0.9 and a mutation probability of 0.05, for both the (a) binary coding and (b) real-valued representations. The velocity structures are compared to the Kgaswane velocity model (shown as a line), which is the currently accepted model for the region. A more compact distribution is observed for the real-valued representation; a bias towards higher velocities is evident, consistent with the large surface wave dispersion curve misfit functions shown in Figure 3. The distribution of solutions from the binary representation shows a much wider and more uniform spread.

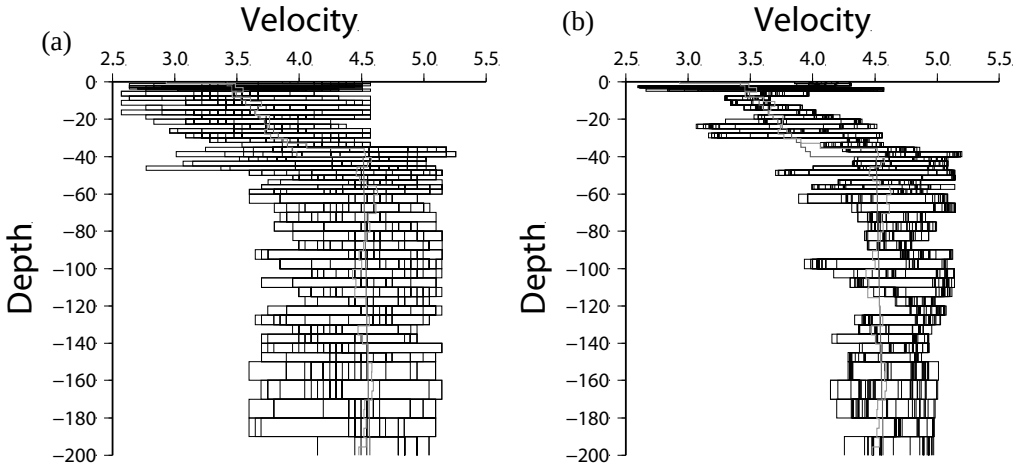

Figure 4: Distribution of velocity models from the Pareto optimal set for (a) binary and (b) real-valued representation (p_c=0.9, p_m=0.05). The solid line is the Kgaswane model.

Discussion and Summary

Genetic algorithms applied to problems with a continuous model space either incorporate a binary coding of the real-valued measures or directly use the real-valued variables. Genetic algorithms used in the geosciences often employ a binary coding (e.g. Sambridge and Kennett, 1996; Mackenzie *et al*., 2001; Dal Moro and Pipan, 2007). Our question is how the choice of binary coding or real-valued representation impacts the resulting Pareto optimal solution set. Preliminary results suggest that the real-valued representation converges faster than the binary coding, and that a small proportion of mutation is advantageous.

Comparisons based solely on the objective functions suggest that the binary coding is preferable. In the real-valued representations, large misfits are observed between the model-predicted and observed surface wave dispersion curve and receiver functions. Commonly, such large misfits would indicate that the corresponding velocity model is not physically plausible. For example, misfit errors observed with receiver functions as in Figure 2c suggest a rapid rate of change in velocity not physical observed in practice. However, when examining the velocity models in the Pareto optimal solution, as in Figure 4, the real-valued representation shows less variability in estimated velocities at each layer. In fact, the variability in velocities resulting from the binary coding is too large to yield usable results.

Despite being less than when using binary coding, the within-layer variability in estimated velocities using the real-valued representation remains too great to interpret the geologic structure. Additional constraints placed on the largest surface wave dispersion curve misfits may improve the precision in estimating the true velocities.

Binary genetic algorithms limit the resolution by forcing the parameters into 5-bit (or similar) chromosomes. Given it is not uncommon to have 30 or more geologic layers, the binary coding increases the number of parameters to be estimated and computation time.

REFERENCES

Dal Moro, G., and Pipan, M. (2007) Joint inversion of surface wave dispersion curves and reflection travel times via multi-objective evolutionary algorithms, *Journal of Applied Geophysics*, 61, pp. 56-81.

De Jong, K., Fogel, L., and Schwefel, H., eds. (1997) *The Handbook of Evolutionary Computation*, Oxford University Press, Oxford. 997 pp.

Deb K. and Agrawl, A. (1994) Simulated Binary Crossover For Continuous Search Space, *Indian Institute of Technology, Department of Mechanical Engineering, Kanpur, SMD-94027,* 34 pp.

Deb K., Agrawl, A., Pratap, A. and Meyarivan, T. (2002) A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, *IEEE Transactions on Evolutionary Computation*, 6, pp. 182-197.

Kgaswane, E, Nyblade, A.A, Julià, J. and Dirks, P. (2009) Shear wave velocity structure of the lower crust in Southern Africa: Evidence for compositional heterogeneity within Archaean and Proterozoic terrains, *Journal of Geophysical Research*, 114 B12.

Julià J., Ammon, C., Herrmann, R., and Corrieg, A. (2000) Joint inversion of receiver functions and surface wave dispersion observations, *Geophysical Journal International,* 143, pp. 99-112.

Mackenzie G., Maguire, P., Denton, P., Morgan, J., and Warner, M. (2001) Shallow seismic velocity structure of the Chicxulub impact crater from modelling of Rg dispersion using a genetic algorithm, *Tectonophysics*, 338, pp. 97-112.

Sambridge M., and Kennett, B. (1996) Genetic algorithm inversion for receiver functions with application to crust and uppermost mantle structure beneath Eastern Australia, *Geophysical Research Letters*, 23(14), pp. 1829-1832.