Velocity inversion of a seismic trace with a micro genetic algorithm

Gabriel Alvarez[1]

ABSTRACT

An ever present goal of seismic processing and inversion is to extract meaningful geology information from seismic data. The simplest possibility is to invert for seismic impedances, or, considering density constant, invert for velocities. Here I use a micro genetic algorithm program to achieve this goal. I use a real sonic log to compute a synthetic seismic trace and then use that trace as input to the micro genetic algorithm program to invert for the velocity log. Without further input, the velocity inversion cannot hope to recover the general velocity-depth trend because this information corresponds to the very low frequencies which are absent in the seismic data. In order to achieve a good match between the real and the inverted sonic log, the inversion must be suppled with an estimate of the velocity-depth trend. Here I use a 33-point Savitzky-Golay filter with a sixth-degree polynomial to smooth the velocity log and show that with this trend the velocity log is reasonably well recovered. I used both L_1 as well as L_2 norm of the sample-to-sample difference between the reference and the inverted trace and show that the results with both norms are similar.

INTRODUCTION

Genetic algorithms have always been recognized as powerful tools for inverting complex object functions with complex constraints either continuous or discrete (Goldberg, 1989; Haupt and Haupt, 1998; Faulkenauer, 1998; Beasley et al., 1993a,b) but there is always a question mark about their robustness in handling high-dimensional problems with reasonable accuracy and speed. In a companion paper in this report (**?**) I show that genetic algorthms, and in particular micro-genetic algorithms (Krishnakumar, 1989), can be used to solve a relatively high-dimensional problem. In this paper I give the details of that problem: inverting a seismic data trace for the underlying interval velocities assuming a layer-cake velocity model. This is a relatively simple non-linear problem that has been investigated for at least 20 years (Lindseth, 1982). Several techniques have been proposed and various commercial software packages exist with high levels of sophistication, using deternministic inversion or simulated annealing.

[1] **email:** gabriel@sep.stanford.edu

If we can use genetic algorthms to efficiently solve this kind of problems, we can then take advantage of the simplicity with which genetic algorithms handle all kinds of continuous or discrete constarints to go beyond the inversion of seismic data for interval velocities. Multiples, wave mode conversion and other such complications may be equally handled.

In this paper I employ a micro-genetic algorithm to invert a synthetic seismic trace for the underlying velocities that produced it. The seismic trace is created from a real well log assuming normal-incidence, a layer-cake velocity model and no multiples or absorption. I show that when the inversion is supplied with an estimate of the velocity-depth trend obtained from the well log with a 33-point, sixth order Savitzky-Golay filter (Press et al., 1992), the inversion yields a very reasonable estimation of the input sonic log.

PREPROCESSING OF THE WELL LOG

The well log used in this project is from Colombia. A 1000-ft long segment of the sonic log from depths 3600 to 4600 feet was selected for the study. This depth interval corresponds to a predominantly shally sequence with a marked transition to sandstone at the bottom. Originally the sonic log was sampled at 0.5 feet but I sub-sampled it to 10 feet intervals using Backus average method (Mavko et al., 2000). The subsampled log thus contains 99 samples. The left panel of Figure 1 shows the subsampled sonic log as it was used to create the synthetic trace.

SYNTHETIC SEISMOGRAM

From the sonic log, assuming constant density, normal incidence, no multiples or reverberations and no absorption, a synthetic seismic trace was created. The reflection coefficients for the trace were simply computed with the recursion:

$$RC_i = \frac{V_i \rho_i - V_{i-1}\rho_{i-1}}{V_i \rho_i + V_{i-1}\rho_{i-1}} \tag{1}$$

where RC_i is the reflection coefficient at the ith interface and, as usual, V and ρ represent velocity and density respectively. The density was assumed to be constant. The depth to time conversion was simply done by integrating the vertical travel time

$$t_i = \sum_{j=1}^{i-1} \frac{2 * \Delta z_j}{V_j} \tag{2}$$

where Δz_j represents the thickness of the j-th layer. In this case the layer thickness was kept constant at 10 feet.

Once converted to time, the reflection-coefficient trace was interpolated to a constant time-sampling interval using an 8-point sinc interpolator and convolved with a Ricker wavelet. The right panel of Figure 1 shows the resulting synthetic trace.

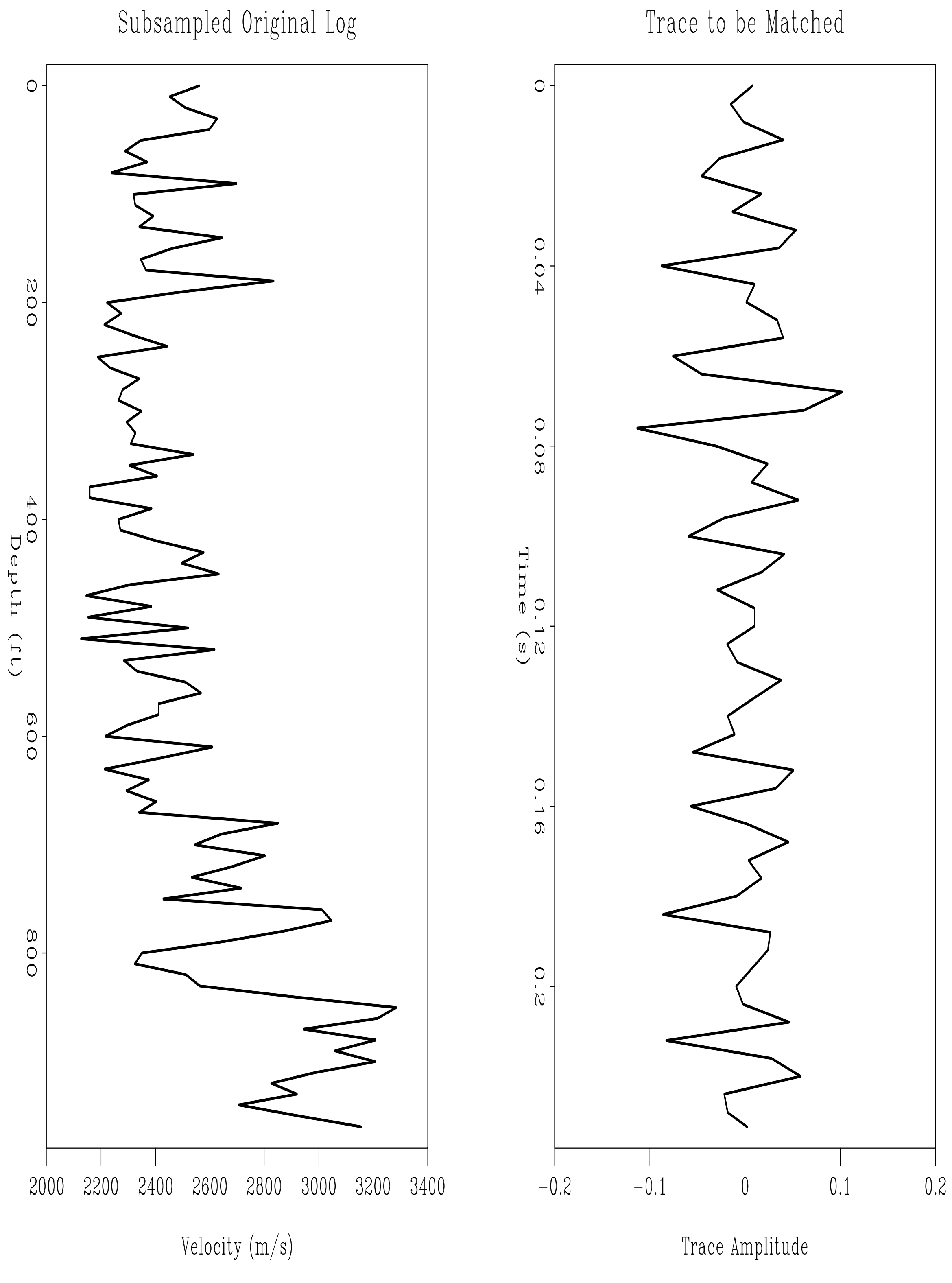

Figure 1: Left panel: Subsampled sonic log used to generate the synthetic trace on the right panel. input [ER]

VELOCITY INVERSION

The next step is to invert the seismic trace for the sonic velocities. This is done by matching candidate traces to the synthetic trace obtained from the log (right panel in Figure 1). I used a micro-genetic algorithm since it proved to be superior to a standard genetic algorithm for this problem (**?**). To perform the inversion, the micro genetic algorithm starts with a small random population of synthetic sonic logs from which trial solutions, in the form of synthetic seismic traces, are generated using the procedure described in the previous section. Each of these traces is matched to the synthetic trace obtained from the real log and a figure of merit (fitness value) assigned according to the "goodness" of the match. Two possible choices of fitness functions are described below. The genetic algorithm "evolves" the solution according to the rules of evolution and survival of the fittest (Goldberg, 1989) until a satisfactory match is achieved between the synthetic seismic trace from the real log and a trial synthetic trace. Once this match is achieved, the corresponding well log will hopefully be a good match to the real one (given the restrictions of no multiple reflections and no attenuation).

This is the ideal situation but in reality we must deal with the important issue of the impossibility of directly recovering the general velocty-depth trend present in the sonic log. This happens because the seismic data lack the very low frequencies which are the ones that could record that information. This issue is obviously very important and will be addressed in some detail below.

Description of the Micro Genetic Algorithm Program

The micro genetic algorithm program used in this study is generation-based (that is, children replace the parents each generation except for those parents selected by elitism), uses binary encoding with a tournament selection for choosing the parents for the next generation and uniform crossover. Other operators such as jump and creep mutation and niching were not used, following the results in (**?**). Table 1 summarizes the parameters of the micro genetic algorithm.

Table 1: Summary of the micro genetic algorithm parameters

Population size	5	Crossover rate	0.95
Mutation rate	0.0	Creep mutation rate	0.0
Minimum parameter value	2200	Maximum parameter value	3200
Elitism	Yes	Niching	No
Selection strategy	Tournament	Number of children	1

Parameter Encoding

Although the samples of a sonic log may not be considered statistically uncorrelated, I will assume that the samples of the subsampled well log are completely independent from one another. There are, therefore, 99 unknowns in our problem, one for each sample of the subsampled sonic log we wish to recover. Each model parameter, i.e. sample of the trial sonic log, was coded as a binary number (called a "gene") and the 99 parameters were stringed together into a long binary string (a "chromosome" in the jargon of genetic algorithms). Although the model paramneters can in principle assume any real value between some reasonable maximum and minimum values, for the purpose of representing them as binary numbers we need to discretize them. Here I used 1024 possible values for each parameter, between 2200 and 3200 m/s. This means a resolution just under 1 m/s. This in turn means that the number of bits per parameter is 10 and the number of bits per chromosome is 990.

Fitness function

For the fitness function I tested both an L_1 and an L_2 norm of the sample-wise difference between the two traces as the criterion to measure the fit. That is, I used

$$f_j = \sum_{i=0}^{N} |x_i - y_{j,i}| \tag{3}$$

and

$$f_j = \sqrt{\sum_{i=0}^{N} (x_i - y_{j,i})^2} \tag{4}$$

where j represents the jth individual and N is the number of time samples. These fitness functions consider each sample in the subsampled well log as being completely independent of any other. Notice that I am using "fitness" here to actually mean the opposite of its usual meaning. That is, an individual will be considered to be highly fit if its fitness value, as defined above, is very low, i.e. it is a good match to the original trace.

Convergence Criteria

The ideal convergence criterion for a genetic algorithm would be one that guaranteed that each and all of the parameters converge independently (Goldberg, 1989; Beasley et al., 1993a). However, this may be too demanding or may result in too many iterations, so more relaxed convergence criteria are usually employed. Here I used four convergence criteria:

1. The fitness function value must be below a given threshold value.
2. The difference between the best and the average fitness is less than a given fraction of the fitness of the average individual.

3. The difference between the best individual of the current population and the best individual so far must be very small (even zero). This means that the most fit individual has converged even if the population itself has not.

4. the number of iterations (generations of the sample population) exceeds a given limit. This prevents the algorithm from spending to much time refining an existing solution.

The combination of these criteria is intended to guarantee that the solution is not due to a lucky guess of the random generator but to a comprehensive search of the model space.

Inversion Constraints

Constraints are critical in non-linear problems to limit the number of possible solutions and to increase the chances of finding a solution that satisfies all aspects of the problem. An obvious constraint to be imposed on the inversion is that the computed velocities be within a reasonable interval. In this case I chose $2200 < V < 3200$ in m/s. This constraint, however, is not enough because we need to account for the velocity-depth trend present in the sonic log. It is a well-known fact that in general wave propagation velocities increase with depth due to several factors such as compaction, cementation, closing of small fractures due to overburden, etc. There may be effects such as overpressure that locally change this behavior, but in average we expect a trend of velocities increasing with depth. When inverting seismic data for interval velocities, it is disappointing to find that this general trend is not recovered. The reason is that the information in the velocity trend would correspond to very low frequencies (lower than, say, 5 Hz) and these low frequencies are not recorded in the standard seismic method. Furthermore, with land data we usually attenuate or eliminate low frequencies because of the detrimental effect of ground-roll on the seismic data. Since this trend is not resolvable by the direct inversion of the seismic trace, it is necessary to impose a second constraint to honor it. This constraint will be discussed in more detail below.

RESULTS

The left panel of Figure 2 shows a comparison between the synthetic trace obtained from the real sonic log, and the one obtained with the genetic algorithm using the L_1 norm. The right panel shows a similar comparison between the original and the inverted log. Although the match of the seismic traces is not perfect, the main features in the original trace were remarkably well recovered. The result for the sonic logs, however, is not so auspicious because the lack of the general velocity-depth trend makes it very difficult to evaluate the “goodness” of the match. If we forget about the trend and examine the details in the logs, we can see that most features of the log were indeed recovered, but the overall result is not satisfactory.

Figure 3 shows a similar comparison as the one just described in Figure 2 but now using the L_2 norm. The results are similar: good match of the seismic traces and poor match of the sonic logs because of our inability to recover the velocity-depth trend.

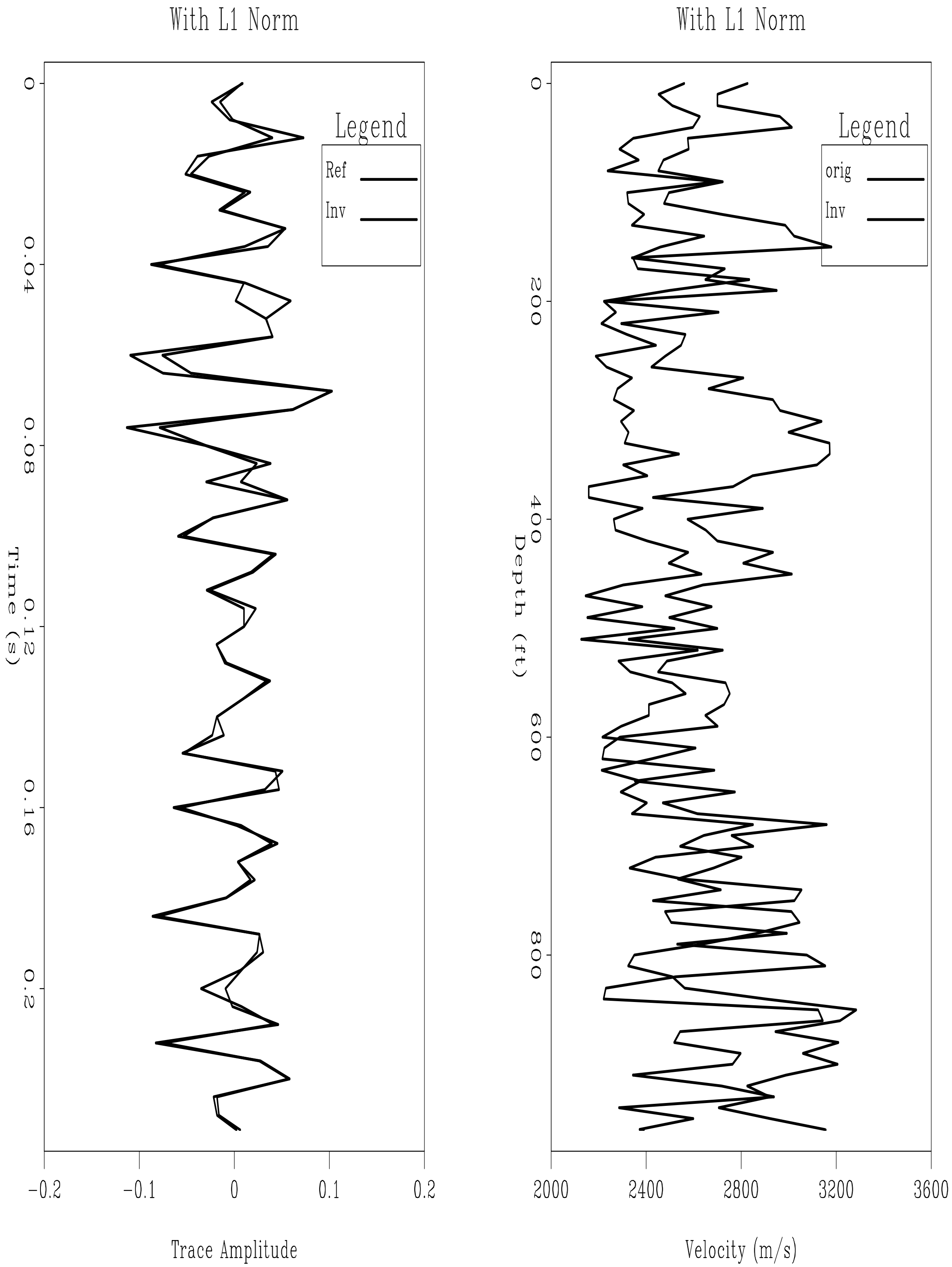

Figure 2: Comparison of the synthetic trace and the inverted trace (left panel) and the original log and the inverted log (right panel) using the L_1 norm. L1_inv_no_trend [ER]

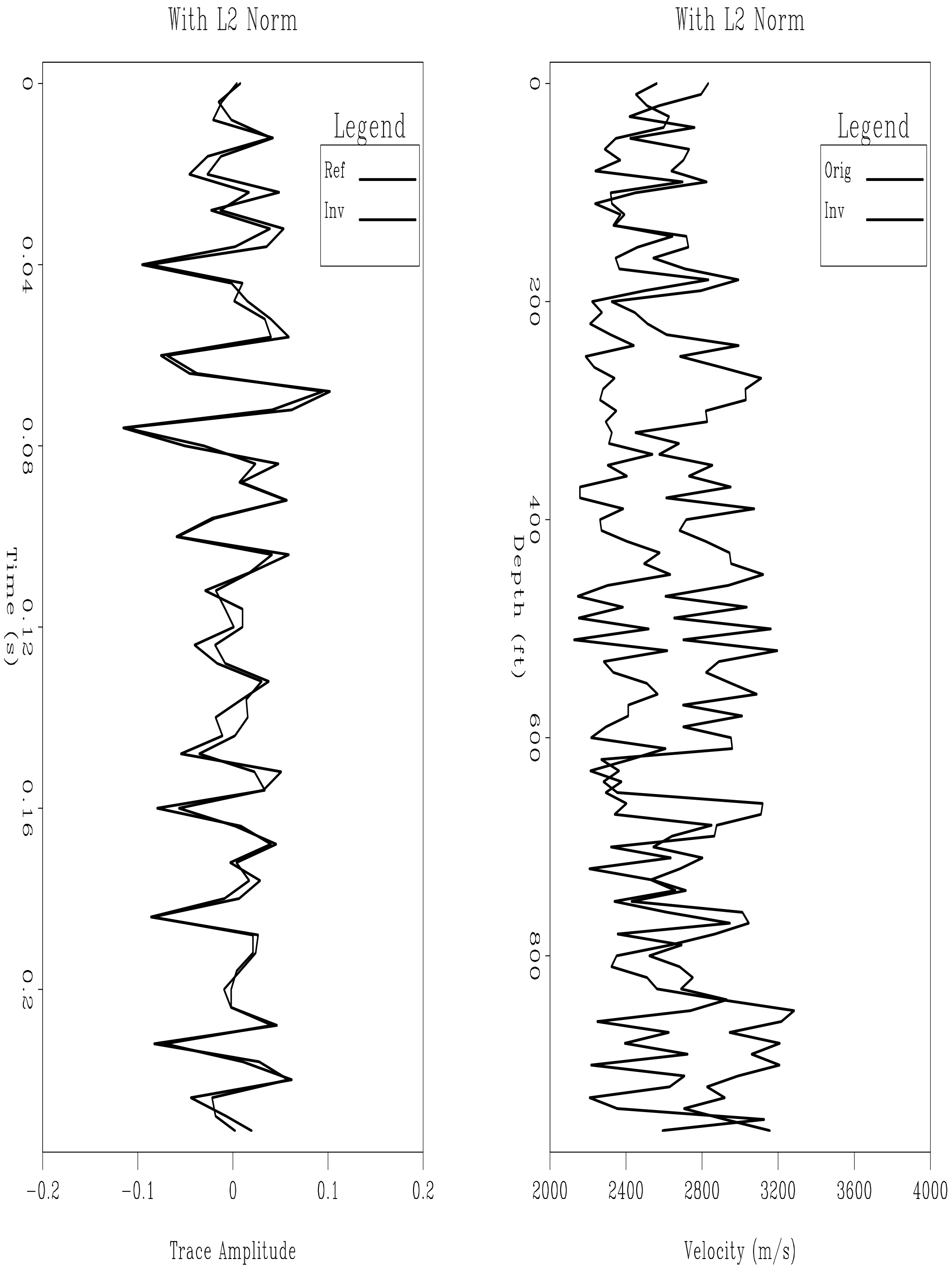

Figure 3: Comparison of the synthetic trace and the inverted trace (left panel) and the original log and the inverted log (right panel) using the L_2 norm. L2_inv_no_trend [ER]

VELOCITY-DEPTH TREND

The results of the previous section illustrate the need to consider the velocity-depth trend as a constraint to the inversion. I used a Savitzky-Golay smoothing filter (Press et al., 1992) to compute the trend from the sonic log. This filter is particularly well-suited for this purpose because it is easy to control the number of samples of the input log that are used to compute a sample of the smoothed log and the degree of the smoothing polynomial.

Figure 4 shows the original subsampled log (top) and the computed velocity-depth trend (bottom) after applying a Savitzky-Golay filter with 33 points (16 to each side) with a sixth-order smoothing polynomial. Extrapolation was used in the original log to 16 points off each of the ends of the log to avoid end-effect problems with the filter. The velocity-depth trend is well recovered, in particular for the deeper samples. This trend will be used as a constraint to the inversion.

INVERSION WITH THE VELOCITY-DEPTH TREND

The velocity-depth trend of Figure 4 was used to constrain the possible values of velocity for each sample of the sonic log. Figure 5 shows a comparison similar to that of Figure 2 but considering the velocity-depth trend. The match for the seismic traces did not change much, since it was already very good without the trend. The match of the sonic log, however, changed significantly, because now, without the distraction of the lack of the trend, we can see that the inverted sonic log matches the original log very well, except for depths between about 700 and 800 feet where the match is a little off in depth.

Similar situation can be seen in Figure 6 which shows the same comparison when the L_2 norm was employed. The conclusion is similar as for the L_1 norm, that is, the match of the sonic logs increases greatly, although it is still far from perfect, illustrating the limitations of the method.

Dicussion

I just showed that the inclusion of the velocity-depth trend incresead significantly the match between the original and the inverted sonic log. The question now is: Why didn't we get an even better match in the sonic logs? An obvious explanation is that the sonic log in this case was subsampled to 10 feet which is still a very thin interval when it comes to the depth resolution that we can achieve with typical seismic data. In order to get a better match we would need higher frequencies, so there is always a limit in the quality of the match that we can reasonably expect to obtain in the inversion of the sonic log. To illustrate this point, Figure 7 shows the same comparison as in Figure 6 except that the peak frequency of the Ricker wavelet was halved. Clearly, the lower frequencies have a very detrimental effect in the quality of the match of the sonic logs. See in particular the depth intervals between about 300-400 ft, 450-550 ft and 700-800 ft.

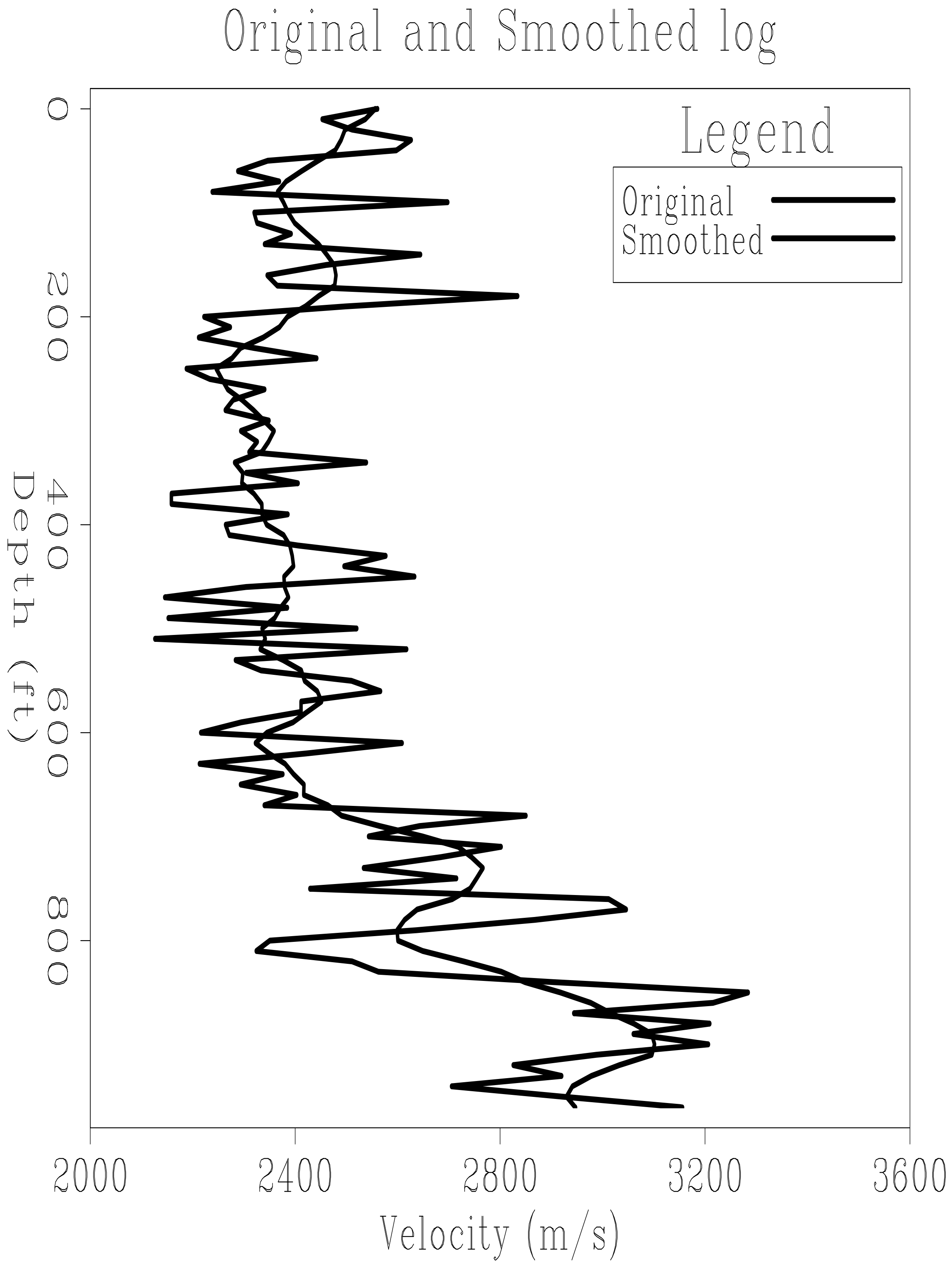

Figure 4: Subsampled original sonic log and velocity-depth trend computed with 33-point sixth order Savitzky-Golay filter. velocity_depth_trend [ER]

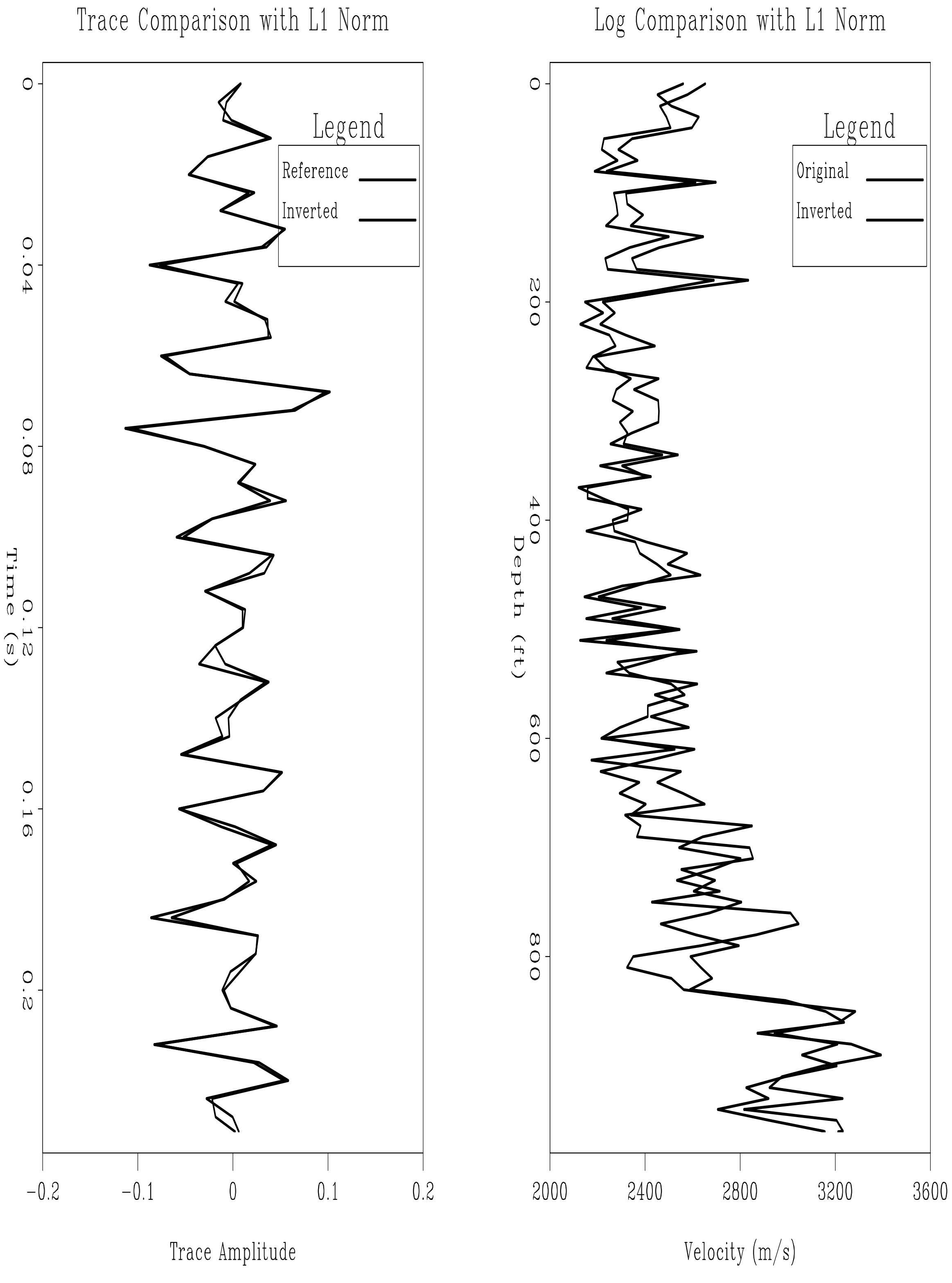

Figure 5: Comparison of reference and matched trace (left panel) and of original and inverted log using the L_1 norm when the velocity-depth trend was taken into account. L1_inv_with_trend [ER]

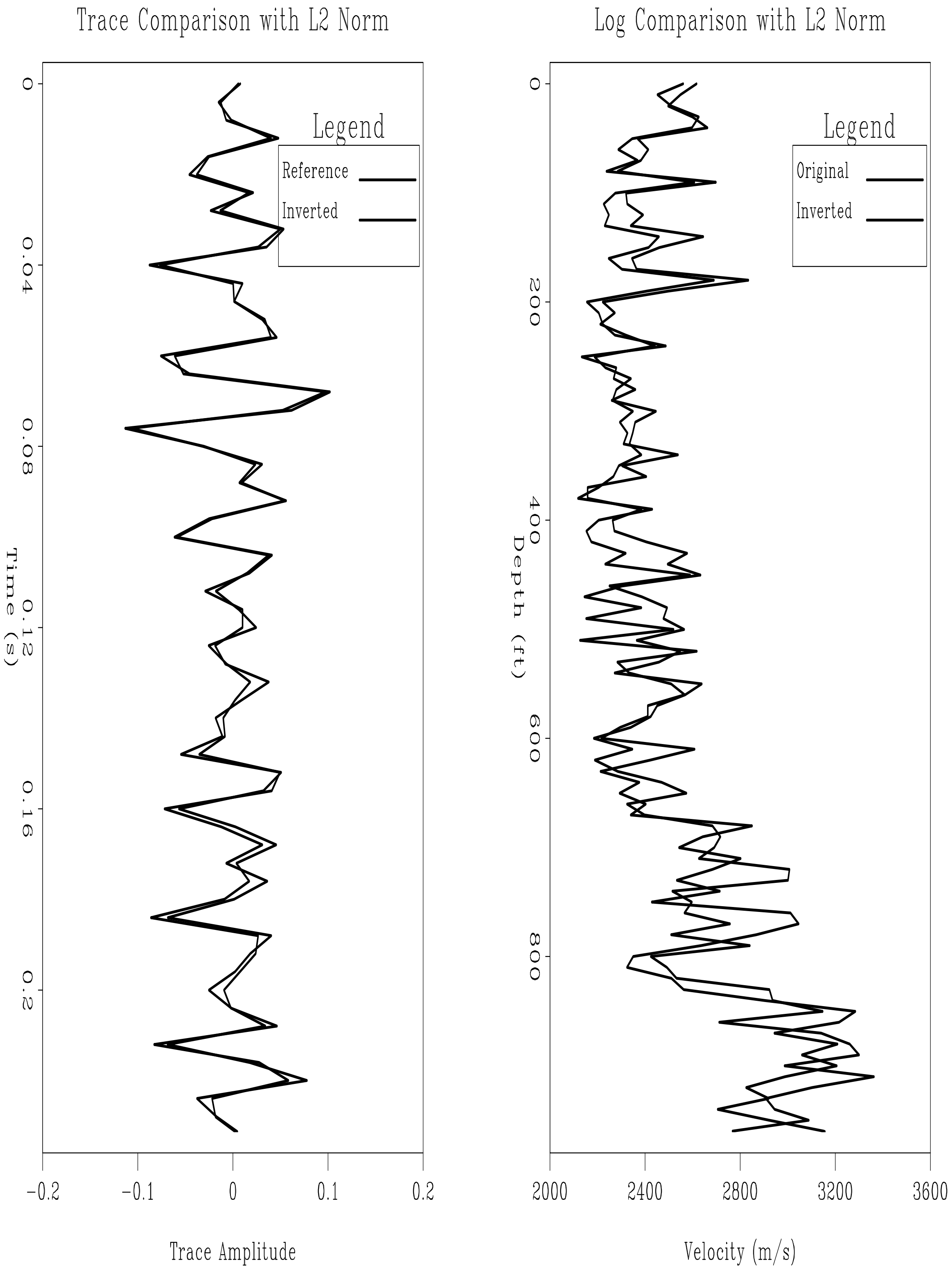

Figure 6: Comparison of reference and matched trace (left panel) and of original and inverted log using the L_2 norm when the velocity-depth trend was taken into account. L2_inv_with_trend [ER]

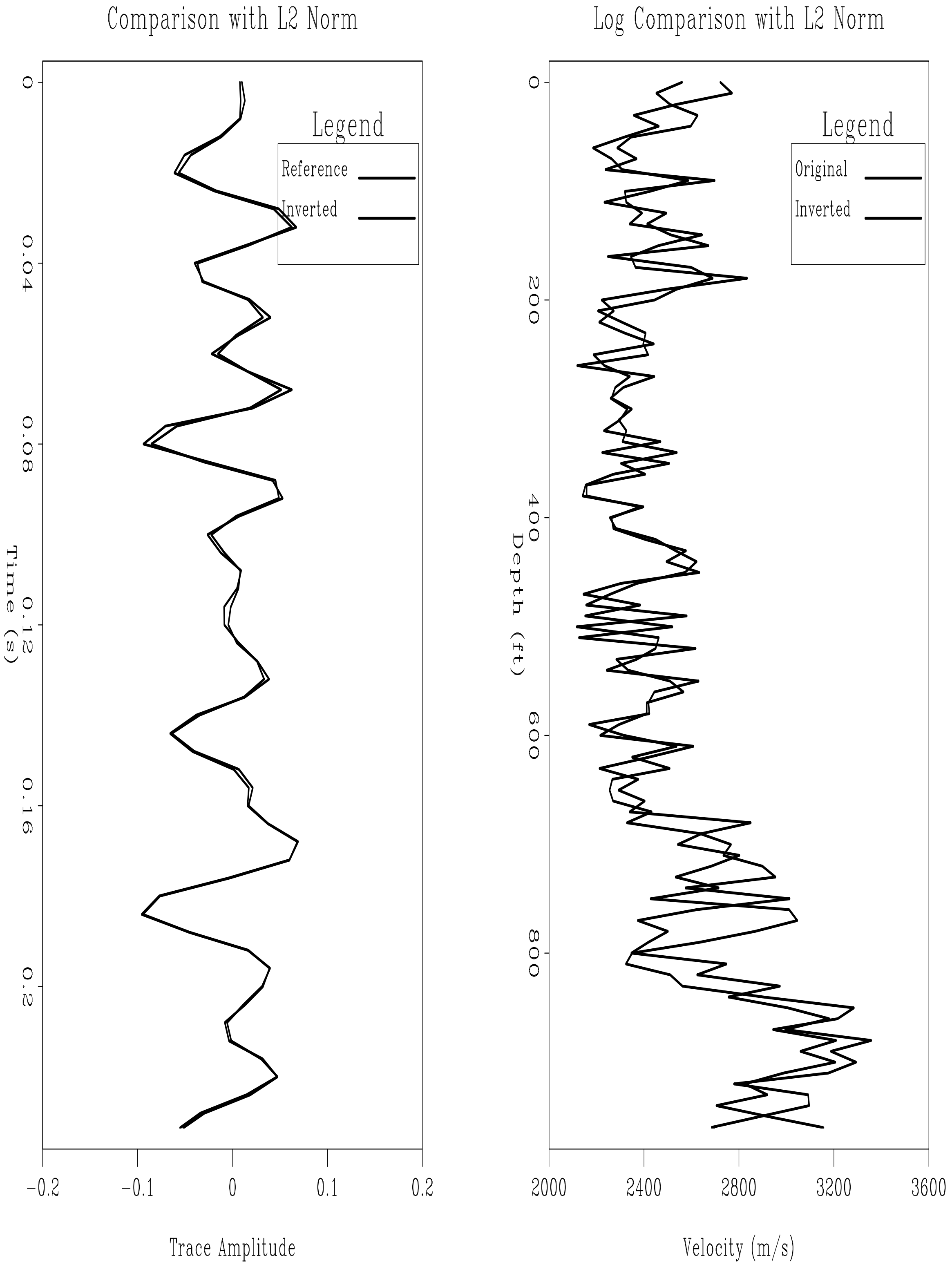

Figure 7: Comparison of reference and matched trace (left panel) and of original and inverted log using the L_2 norm when the velocity-depth trend was taken into account and the peak frequency of the Ricker wavelet was halved. L2_inv_trend_low_freq [ER]

It should also be noted that the example here is not very realistic because I used the exact same approach to compute the seismic data to be matched that I used to compute the trial solutions. In a real situation the reference trace will be obtained from a real seismic survey whereas the trial solutions will still be computed with a procedure similar to the one used here (perhaps inproved by allowing random noise or multiples or attenuation). Therefore, ther will be reasons other than the velocity-depth trend or the frequency content of the data that will play a rol. The estimation of the seismic wavelet, for example, is well-recognized as a very important issue when inverting for real data.

CONCLUSIONS AND FUTURE WORK

From the point of view of the genetic algorithm inversion some lessons have been learned after extensive parameter testing: use of a micro-genetic algorithm with uniform cross-over and no mutation emerges as the best option for this problem (as opposed to a standard genetic algorithm with single-point cross-over and jump mutation). A micro-genetic-algorithm population of 5 or 7 individuals with a cross-over probability of 0.9-0.95 seems to be optimum for this problem.

An important issue to be further analyzed is that of the multi-modality of the search space. It is clear in this case that there is a single global minimum, namely matching the original trace sample-by-sample. This, however, does not guaratee a simialr sample-to-sample match in the original log, which is a consequence of the non-linearity of the problem.

I have found that once I get close enough to this global minimum it takes a large number of iterations to scape local minima (many traces match “almost exactly” the original). My present convergence criteria do not allow for checking of convergence of individual parameters so I have to investigate alternative options. Another issue is the convenience of using the floating point representation of the model parameters directly for the inversion rather than their binary representation. This may increase the resolution of the model parameters but may make it more difficult for the micro-genetic algorithm to converge.

REFERENCES

Beasley, D., Bull, D., and Martin, R., 1993a, An overview of genetic algorithms: Part 1, fundamentals: University Computing, **15**, no. 2, 58–69.

Beasley, D., Bull, D., and Martin, R., 1993b, An overview of genetic algorithms: Part 2, research topics: University Computing, **15**, no. 4, 170–181.

Faulkenauer, E., 1998, Genetic algorithms and grouping problems: John Wyley and sons.

Goldberg, D., 1989, Genetic algorithms in search for optimization and machine learning: Addison-Wesley Pub. Co.

Haupt, R., and Haupt, S., 1998, Practical genetic algorithms: John Wyley and sons.

Krishnakumar, K., 1989, Micro-genetic algorithms for stationary and non-stationary function optimization.: SPIE: Intelligent control and adaptive systems, **1196**.

Lindseth, R., 1982, Digital processing of geophysical data - areview: Soc. of Expl. Geophys.

Mavko, G., Dvorkin, J., and Mukerji, T., 2000, Rock physics handbook.: Cambridge University Press.

Press, W., Teukolsky, S., Vettering, W., and Flannery, B., 1992, Numerical recipes in fortran 77. the art of sicentific computing: Cambridge University Press.