DETERMINATION OF THE STACKING VELOCITY FIELD VIA OPTIMIZATION METHODS

Danian Steinkirch de Oliveira[1], Milton José Porsani[2] and Paulo Eduardo Miranda Cunha[3]

ABSTRACT. We developed a strategy for automatic Semblance panels pick, that uses Genetic Algorithm optimization method. In conjunction with restrictions and penalties set from *a priori* information it's obtained as a result a nonlinear fit of time interval velocities, that when converted at root mean square (RMS) velocity, better maximizes the sum of the Common Mid Point (CMP) group, corrected with normal moveout (NMO). Currently, a good imaging of deep reflectors, especially in Brazilian basins, below the salt layer, has proved to be a major challenge. Obtaining a seismic velocity field corresponding to the subsurface geology and resulting in a focused seismic image is the main target of seismic processing. In the last decade, the reflection tomography has established itself as one of the main methods of velocity model construction for seismic data migration. On the other hand the full waveform inversion (FWI), taken forward due to recent advances in computing, become feasible in inversion of 2D and 3D velocity models. Despite the stacking velocity analysis be, among these, the less accurate method for generating velocity fields, it is still used on a large scale by the oil and seismic processing companies, because of its low cost and can provide a good initial velocity field for tomography and FWI.

Keywords: Genetic Algorithm, velocity analysis, Semblance.

RESUMO. Foi desenvolvida uma estratégia de *pick* automático dos painéis de *Semblance*, que usa o método de otimização Algoritmo Genético. Em conjunto com restrições e sanções estabelecidas a partir de uma informação *a priori*, foi obtido como resultado um ajuste não-linear de velocidades intervalares em tempo, que quando convertidas em velocidade RMS, melhor maximiza a soma do grupo CMP, corrigida de NMO. Atualmente, provou ser um grande desafio a geração de uma boa imagem de refletores profundos, especialmente em bacias brasileiras abaixo da camada de sal. A obtenção de um campo de velocidades sísmica correspondente à geologia do subsolo, resultando em uma imagem sísmica focada é o principal alvo de processamento sísmico. Na última década, a tomografia de reflexão estabeleceu-se como um dos principais métodos de construção de modelo de velocidade de migração de dados sísmicos. Por outro lado, a inversão de onda completa (FWI) tomou a frente, devido aos seus excelentes resultados de inversão de modelos de velocidade 2D e 3D, que se tornaram viáveis somente pelos recentes avanços na computação. Apesar da análise de velocidade de empilhamento ser, entre estes, o método menos preciso para gerar campos de velocidade, ainda é utilizada em larga escala pelas companhias de petróleo e processamento sísmico, por causa do seu baixo custo e por poder proporcionar um bom campo de velocidade inicial para tomografia e FWI.

Palavras-chave: Algoritmo Genético, análise de velocidade, *Semblance*.

[1]Department of Geophysics, Petrobras's Research Center, Av. Horácio Macedo, 950, Ilha do Fundão, 21941-915 Rio de Janeiro, RJ, Brazil. Phone: +55(21) 2162-6465
– E-mail: danian@petrobras.com.br

[2]Universidade Federal da Bahia, Institute of Physics/CPGG, 123, Department of Geophysics of Earth and Enviroment, Campus Universitário da Federação, 40170-290
Salvador, BA, Brazil. Phone: +55(71) 3283-8521; Fax: +55(71) 3283-8551 – E-mail: porsani@cpgg.ufba.br

[3]Department of Geophysics, Petrobras's Research Center, Av. Horácio Macedo, 950, Ilha do Fundão, 21941-915 Rio de Janeiro, RJ, Brazil. Phone: +55(21) 2162-4525
– E-mail: pemcunha@petrobras.com.br

INTRODUCTION

Obtaining the seismic velocity field became the main target of subsurface imaging in oil exploration. Seismic data migration is the process of depropagation of the wave field recorded in the receivers, in order to recover it at the exact moment of the scattering occurred during acquisition.

It can be understood as a mapping of the data recorded in time for a seismic section in time or in depth, with the correct position and amplitude of the reflectors. A good velocity field is able to focus seismic reflectors. With a good image the geophysicist can interpret the geological features of an oil field.

The objective of this research is to analyze the response of differents optimization methods, global and local, facing the inverse problem of the velocity model estimation which better corrects the NMO (normal moveout) and stacks CMP (Common Mid Point) groups. It will be shown a strategy of inversion that performs automatic pick of Semblance panels and generates an initial model of interval velocities in time. This strategy was based on Lumley (1997), which used the Monte Carlo method for generating random velocity field models. The global search method, Genetic Algorithm, used in this work, proved to be superior than purely statistical methods such as Monte Carlo method by using a rule of probability to guide their search among the models parameters (Sen & Stoffa, 1995).

Although the stacking velocity analysis be the less accurate among the methods of inversion of the velocity field, it is still used on a large scale by major petroleum and seismic processing companies because it is cheap and can provide a good initial velocity field.

THE SEISMIC VELOCITY FIELD

The good imaging of subsurface using seismic reflection is essential for the correct interpretation of geological features and construction of the reservoir model that will be the target of exploitation in oil industry.

However, the imaging is attached to the velocity field, in other words, a good image can be obtained only with a knowledge of the seismic velocity model. A good strategy for determining the subsurface velocity becomes essential (Kosloff et al., 1996). The seismic velocities play two major roles in exploration geophysics (Gao et al., 2007). The first one, as mentioned above, says that the fields obtained by migration or stack velocity analysis, are used to enhance the images of the subsurface seismic reflectors. However, these generally do not represent the true velocity in the subsurface, especially in areas with complex structures.

The velocity field used for depth migration in 3D should be identical to the subsurface geology, or unless a low pass filtered version of it. The second function of velocity fields requires that models, wich approximate the true velocity distribution in the subsurface, contains important information about the lithology and location of fluids.

This idea is supported by Rowbotham & Pratt (1997) who argues that the velocity models are essential in detecting subtle changes in lithology, porosity and permeability.

The subsurface information that we possess are basically transit times and amplitudes. The transit time of the wave fronts generally provide information of low frequency components (background) parameters of the medium, while the amplitudes of the wave fronts carry information of high frequency components (Clayton & Stolt, 1981). It is important to use both information contained in the seismic data to have a velocity model as complete as possible, with the low and high frequency components. There are two main techniques used in the construction of the velocity field, the seismic tomography and FWI (Full Waveform Inversion). The first is based on geometrical optics, while the second uses the wave equation.

The tomography technique consists in picking the transit times of seismic events, which constitute the observed data. Ray tracing methods are used then to predict the transit times of candidates velocity models (Varela et al., 1998). In the last decade, the reflection tomography has established itself as one of the main methods of velocity model construction for time or depth migration of seismic data and tomography of first arrival, applied to large-aperture seismic data, proved to be a interesting tool in the investigation of the Earth's interior structures (Delost et al., 2008).

In the last decade the FWI has gained increasing attention due to its success in geologically complex scenarios (Pratt et al., 1998; Jaiswal et al., 2009), where conventional processing has shown limited results. The FWI is an extremely powerful technique capable of constructing high-resolution images whose aim is to estimate the velocity model that most closely approximates the true wave field recorded (Valenciano et al., 2009), making an adjustment of real seismic data (seismograms) and the data obtained by modeling the wave field. This nonlinear inverse problem is solved comparing the data recorded with the seismic data modeled by the wave equation, and through various iterations, improve the initial velocity model (Charara et al., 2000). Recent advances in high-performance computing allow the feasibility of this technique in 3D acoustic models (Virieux & Operto, 2009). The knowledge of the optimization methods solution of linear systems stands as central point to understanding these two inversion techniques.

BASIC CONCEPTS OF THE INVERSION THEORY

A key role of geophysics is to make quantitative inference of the physical properties of the Earth's interior from surface data and indirect measurements, which is an inverse problem or a process of inductive reasoning. The seismic inversion extracts information about the subsurface by adjusting the model parameters and predicted data, by direct modeling (simulation), with the observed data on the surface or in the pit (Symes et al., 2011). Most geophysical problems are nonlinear problems that may be expressed by the operator

$$d = F(m), \tag{1}$$

where m is some function (or vector) describing the model parameters, d is a data set which can also be characterized as a function of the observation point (in the case of continuous observations) or a vector (in the case of discrete observations) and F a function that relates m with d. For now, we will discuss linear discrete version of the inverse problem

$$\mathbf{d} = \mathbf{Gm}, \tag{2}$$

where $\mathbf{d}$ and $\mathbf{m}$ are, respectively, characterized by vectors

$$\overline{d} = \begin{bmatrix} d_1 \\ \vdots \\ d_M \end{bmatrix} \qquad \overline{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_N \end{bmatrix}$$

and $\mathbf{G}$ is a matrix $M \times N$ which relates the model space with observed data space, also called the system geometry.

$$\mathbf{G} = \begin{bmatrix} g_{11} & \cdots & g_{1N} \\ \vdots & \ddots & \vdots \\ g_{M1} & \cdots & g_{MN} \end{bmatrix}$$

Solving an inverse problem is to determine a estimated model, used in direct modeling, that has a minimum error when compared the predicted data with the observed data (Sen & Stoffa, 1995).

$$\Phi(\mathbf{m}^{est}) = \sum_{i=1}^{M} (d_i^{obs} - d_i^{prev})^2 \tag{3}$$

Take for example the Φ function in equation (3) with L_2 norm, where the observed and predicted data can be described as

$$\mathbf{d}^{obs} = \mathbf{Gm}^{verd} \tag{4}$$

and

$$\mathbf{d}^{prev} = \mathbf{Gm}^{est}. \tag{5}$$

The true model of the Earth, where the observed data was recorded is called m^{verd} while m^{est} is the model used to calculate the predicted data. This function has several names, including the name of the objective function is the most common. This function can also purchase other forms with standard Lp (Porsani et al., 2000), for example.

$$\Phi(\mathbf{m^{est}}) = 1 - \frac{2\sum_{i=1}^{M} \left| d_i^{obs} - d_i^{prev} \right|^p}{\left[\sum_{i=1}^{M} \left| d_i^{obs} + d_i^{prev} \right|^p + \sum_{i=1}^{M} \left| d_i^{obs} - d_i^{prev} \right|^p \right]} \tag{6}$$

If we consider $\mathbf{m}^{est} = \tilde{\mathbf{m}}$ and $\mathbf{d}^{prev} = \tilde{\mathbf{d}}$, the vector $\mathbf{e}$ become a representation of the error between the observed data and the predicted data $[\mathbf{d}^{obs} - \mathbf{d}^{prev}]$ or simply $[\mathbf{d} - \tilde{\mathbf{d}}]$

$$\mathbf{e} = [\mathbf{d} - \tilde{\mathbf{d}}]. \tag{7}$$

The error is measured according to the norm used. The L_1 norm is more robust because it is less sensitive to measures that are very nonstandard (outliers). A L_2 norm, also called the Euclidean norm, tries to adjust the models to these measures (outliers), but when you're close to the solution, the L_2 norm is better because it is infinitely differentiable (Menke, 1989).

OPTIMIZATION METHOD – GENETIC ALGORITHM

The inverse problem consists in recover the model parameters by fitting the observed seismic data with the synthetic modeled data. However, the data predicted by the solution of the direct problem are generally different from observed data. So the strategy is to use iterative methods that minimize the error between observations and predictions by adjusting the parameters of the models at each iteration (Amundsen & Ursin, 1991).

The optimization methods for solving inverse problems are recursive methods, which aim at each iteration k, compute a new model $\mathbf{m}^k$, whose error $\mathbf{e}^k = [\mathbf{d} - \mathbf{d}^k]$ is less than the error in the previous iteration ($\mathbf{e}^k < \mathbf{e}^{k-1}$).

A major goal of the inversion is to find the minimum of an error function or the maximum of a correlation function. Due to the fact that such functions may have several minima and maxima of different heights, a linear gradient-based method may not be the most appropriate inversion technique, unless the reference model be adequately close to the true solution.

The local scope optimization methods does not prevent the convergence of function due to the limited accuracy of the initial model, the lack of low frequencies, the presence of noise and approximations made during direct modeling (Virieux & Operto, 2009).

Thus, in order to avoid non-convergence of the local scope optimization methods, we choose a gobal search method or a random search technique, such as the Monte Carlo method (Sen & Stoffa, 1995). Other methods based on random search but using a probability rule to guide their search parameters among models, such as Simulated Annealing and Genetic Algorithm, instead purely statistical methods such as Monte Carlo.

The Genetic Algorithm (GA) is a technique used in functions defined in spaces of models with large dimensions. It simulates the biological processes of evolution by crossing, selection and mutation, to improve performance (fitness) of a given population models (Hong & Sen, 2009). The performance (φ) of a model can be calculated in different ways (Sen & Stoffa, 1995; Porsani, 2008), for example

$$\varphi(\mathbf{m^k}) = \frac{\mathbf{d}^T\mathbf{d}^k}{(\mathbf{d}^T\mathbf{d})^{\frac{1}{2}}(\mathbf{d}^{(k)T}\mathbf{d}^k)^{\frac{1}{2}}}, \qquad (8)$$

where, if $d = d^k \longrightarrow \varphi(\mathbf{m}) = 1$. The performance or fitness for a particular model quantifies how good is the fit of the observed data, $\mathbf{d}$, with calculated data, $\mathbf{d}^k$, from a given model $\mathbf{m}^k$ under the F operator,

$$\mathbf{d}^k = F(\mathbf{m}^k).$$

The first step is to randomly select the parameters that will generate a population of models. To illustrate the GA method, let us consider the hypothetical example with three parameters, velocity v, the gradient layer α and depth of the interface h. The parameters are vectors that have a minimum, a maximum and a sampling interval value, defined based on prior knowledge of the geological and geophysical of the area of interest.

$$\{\mathbf{m}_{1k}\}_{k=1}^{N_v} = \{\mathbf{v}_k\}_{k=1}^{N_v} = \{v_1, \ldots, v_{Nv}\}$$
$$\{\mathbf{m}_{2k}\}_{k=1}^{N_a} = \{\alpha_k\}_{k=1}^{N_a} = \{\alpha_1, \ldots, \alpha_{Na}\}$$
$$\{\mathbf{m}_{3k}\}_{k=1}^{N_h} = \{\mathbf{h}_k\}_{k=1}^{N_h} = \{h_1, \ldots, h_{Nh}\}$$

We may relate the position k of the vector with the parameter value $\mathbf{m}_{jk}$, through

$$\mathbf{m}_{jk} = m_{jMin} + (k-1) * \Delta m_j, \qquad (9)$$

where m_{jMin} is the initial value and Δm_j is the sampling interval of the model $\mathbf{m}_j$. A unique feature of Genetic Algorithm is that it is able to work simultaneously with a population with various models coded. The next step consists in designing a binary encoding scheme (Sen & Stoffa, 1995), that represents models. Thus, each bit corresponds to a gene, and each model in the population is described by a bit string or chromosomes. Conventionally, the chromosomes in the GA method has a fixed size. The size of the binary string is equal to

$$N_j = 2^{n_{bits}} - 1 \qquad (10)$$

where N_j is the total number of samples of model's parameter vector $\mathbf{m}_j$. There are several ways to code models. Tables 1 and 2 exemplify the encoding used in this research where the binary values are always related to the position k on the parameter vector (Sen & Stoffa, 1995)

The initial models with binary coding are gradually optimized by maximizing the fitness function, with the three evolutionary operators until the best model is obtained. In Selection step, models will be chosen to replace the current population. The probability of model being selected is proportional to its performance. Thus models with higher performance are more likely to be selected. This is followed by crossing, where some models are paired and information is exchanged between each pair based on a probability function. Finally, the mutation operator disturbs all parameters with a low probability. The mutation causes diversity in models which prevents premature convergence and helps to jump out of the local optimum. The probability of mutation must be very low, otherwise good models can be destroyed.

At the end of the mutation step, the new generation candidates are accepted or rejected based on the Metropolis algorithm, which is followed by a new cycle of genetic operations until a certain convergence criterion is satisfied. The GA is quite robust, and its main advantage is that it always converges to models with higher performance with a reasonable number of crossover and mutation (Hong & Sen, 2009).

Genetic Algorithm Steps (Porsani, 2008).

1. Generate an initial population at iteration $k = 0$.

2. Evaluate the fitness of the population.

3. Apply the genetic operator.

Table 1 – Binary encoding example of a parameter with 3 values.

Parameter	Value	Position in vector	Binary value
v	1500 m/s	1	00
	2000 m/s	2	01
$N_v = 3$	2500 m/s	3	11

Table 2 – Binary encoding example of a parameter with 5 values.

Parameter	Value	Position in vector	Binary value
α	5°	1	000
	10°	2	001
	15°	3	011
	20°	4	100
$N_a = 7$	25°	5	101
	30°	6	110
	35°	7	111

3.1. Randomly select two parents.

3.2. Make crossing.

3.3. Compare the child with the worst fitness model.

4. Generate a new population descending

5. Set $k = k + 1$, go to step 2 and repeat until reach a stopping criterion which is satisfied by the descendant population.

Inversion example using Genetic Algorithm

To better exemplify the Genetic Algorithm optimization method for solving nonlinear inverse problem, we take the analytical solution of the transit times for a simple model with a sloping interface of inclination θ to the horizontal, and depth h on the zero-offset position. The time travel ($\overline{srx_i}$) from a particular source positioned at the point s whose wave propagates in the layer V_1 (we will call it v), reflects in the interface V_1/V_2 and is received in the geophones $\mathbf{x}$, are calculated by

$$\mathbf{t}(h, v, \theta, \Delta\mathbf{x}) = \frac{1}{v}\left\{\left[2hcos(\theta)\right]^2 + \left[\Delta\mathbf{x} + 2hsen(\theta)\right]^2\right\}^{\frac{1}{2}} \qquad (11)$$

where $\Delta\mathbf{x}$ is the difference between the position of receivers ($\mathbf{x}_i$) and the position of the source (s).

$$\Delta\mathbf{x} = (\mathbf{x}_i - s) \qquad (12)$$

It can be seen from equation (11) that the geometry ($s, \mathbf{x}$) can not be separated from model $\mathbf{m}(h, \theta, v)$, as it was done earlier in the linear problem. Then, to describe this problem, we use the equation (1) modified

$$t = F(\mathbf{x}, \mathbf{m}) \qquad (13)$$

where

$$\mathbf{m} = \begin{bmatrix} v \\ \theta \\ h \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}. \qquad (14)$$

Through a modification of equation (3) it is possible to generate the objective function for all space models

$$\phi(\mathbf{m}^{est}) = \sum_{i=1}^{M}\left[\mathbf{d}(x_i)^{obs} - F(\mathbf{m}^{est}, x_i)\right]^2, \qquad (15)$$

where $\mathbf{d}(x_i)^{obs}$ is the observed data vector in the x_i position and the error vector is equal to

$$\Delta\mathbf{d} = \left[\mathbf{d}(\mathbf{x})^{obs} - F(\mathbf{m}^{est}, \mathbf{x})\right] \qquad (16)$$

or

$$\Delta\mathbf{d}(\tilde{\mathbf{m}}, \mathbf{x}) = \Delta\tilde{\mathbf{d}}$$
$$= \left[F(\mathbf{m}, \mathbf{x}) - F(\tilde{\mathbf{m}}, \mathbf{x})\right]. \qquad (17)$$

The reflection travel times calculation of various source-receiver pairs, simulating a common group of shooting were done using equation (11). An acquisition was simulated with 100 shots

spaced 25 meters. The first shot has coordinated x equal to 3025 meters. Each group has 241 channels also spaced 25 meters, disposed in a split-spread arrangement.

The travel times were used as observed data for inversion. The model to be inverted has three parameters, the slope, the velocity and depth of the initial layer, defined by the equation (13). It was made a inversion of the parameters that define the model described above. Details of the binary encoding of the model can be seen in Table 1. We used the objective function of equation (6) with L_p and $p = 1/2$. It was set to the true model the following parameters:

- v_0 = 1500 m/s

- z_0 = 100 m

- α = 6 degrees

The parameters are vectors that have a minimum, a maximum and a sampling interval value, which suggests the need for a prior knowledge of the geological and geophysical properties of the true model. The global search method Genetic Algorithm, does not need that the candidate model be in the local minimum, just close to it, contained in an interval $(P\_min : P\_max, \delta P)$, as shown by the parameterization below. So the best model could be contemplated during the iterations of the GA.

- v_0 = [1200:18000,100] (m/s)

- z_0 = [80:120,100] (m)

- α = [4:8,1] (degrees)

The inverted model

- v_0 = 1500 m/s

- z_0 = 105 m

- α = 6 degrees

had a premature convergence due to the number of parameters that define the model space, contrasted with the large number of individuals (models) of the initial population. The depth initial z_0 assumed this value due to the regularization parameters for the binary encoding.

SEISMIC PROCESSING IN VELOCITY ANALYSIS

During the stage called direct modeling, which calculates the predicted data for the inversion strategy used in this study, was applied several techniques that are quite common in seismic processing.

One technique used in this work is the conversion of interval velocity fields in depth to time and to RMS (root mean square). The relationship between the interval velocity field in depth and interval velocity field in time is given only by a change of variables, where

$$\mathbf{V_{int}}(t) = \mathbf{V_{int}}(z) \qquad (18)$$

and

$$t = \frac{z}{\mathbf{V}(z)} \qquad (19)$$

The equation that relates time interval velocity field with the time RMS velocity (Thomas, 2001) is given by

$$\mathbf{V}_{RMS}(t_n) = \frac{\sqrt{\sum_{i=1}^{n} \mathbf{V_{int}}^2(t_i)}}{\sum_{i=1}^{n} t_i} \qquad (20)$$

The seismic data organized in CMP groups are used to make the stacking velocity analysis and construction of the zero-offset section, where the position of the shot is the same of the receptor. To generate the zero-offset section, CMP groups are corrected from normal moveout (NMO) time and stacked toward the receivers. Normal moveout times are calculated by the formula (Castle, 1994)

$$\Delta\mathbf{t}_{nmo} = \mathbf{t}_{nmo} - \mathbf{t}_0, \qquad (21)$$

where

$$\mathbf{t}_{nmo} = \mathbf{t}_0 \sqrt{1 + \left(\frac{\Delta\mathbf{x}}{\mathbf{vt}_0}\right)}, \qquad (22)$$

An undesirable event of NMO correction is the stretching of the wavelet. The stretching can be automatically muted, giving a maximum value in percentage for the ratio between the period of the stretched pulse (T) and the original pulse (T_0):

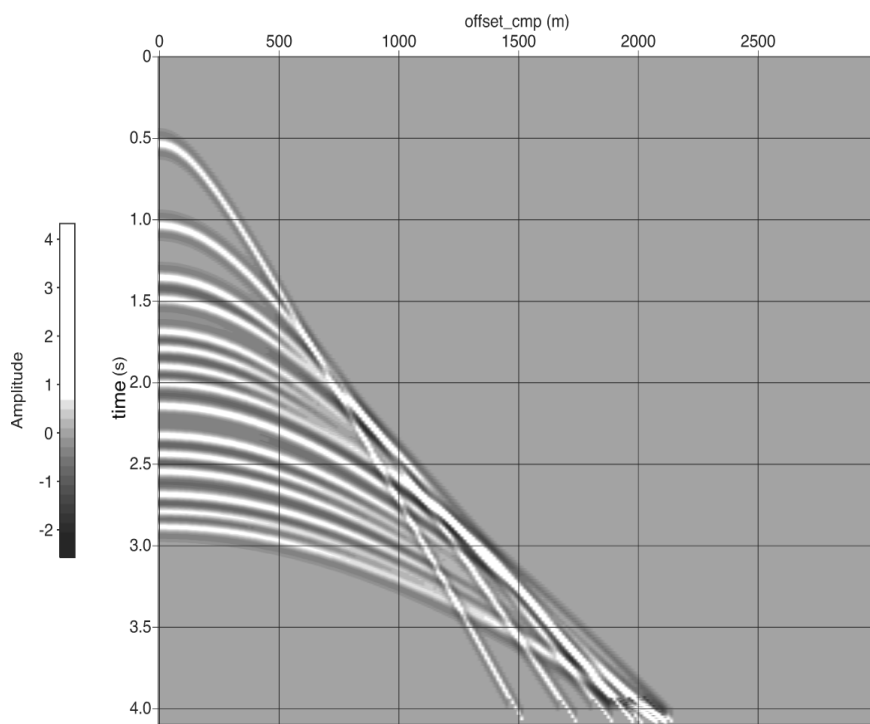$$\frac{T}{T_0} = \frac{\Delta t_{nmo}}{t_0} \qquad (23)$$

After NMO correction, the traces of the CMP group are summed toward the receiver, according to the equation

$$\mathbf{d}(g_0)^{prev} = \sum_{i=1}^{M} [\mathbf{c}(g_i)] \qquad (24)$$

where $\mathbf{d}(g_0)^{est}$ is the seismic trace at offset equals to zero, generated from the stacking of seismic traces $\mathbf{d}(g_i)^{obs}$ of a CMP group corrected with the NMO ($\mathbf{c}(g_i)$), and M is the number of traces of the CMP group.
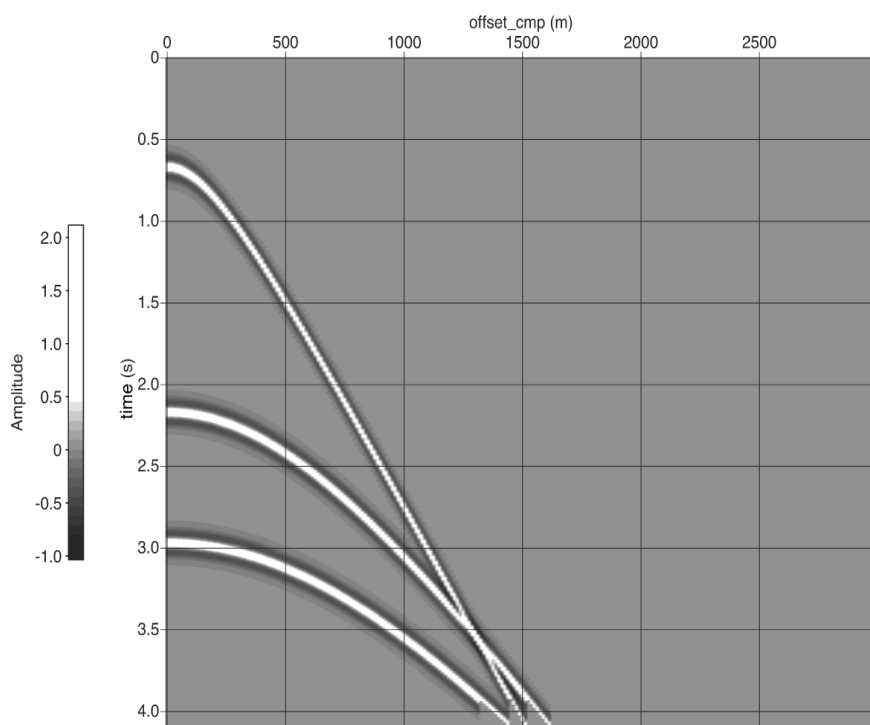
(a) 04 layer model

(b) 16 layer model

Figure 1 — Synthetic CMP group.

Semblance panels are generated by normal moveout correction and stacking of CMP groups with different constant velocities. These amplitudes values are normalized by Semblance equation (Key & Smithson, 1990),

$$NE = \frac{1}{m} \frac{\sum_{n=1} \left(\sum_{i=1} \mathbf{f}_{i,n} \right)^2}{\sqrt{\sum_{n=1} \sum_{i=1} \mathbf{f}_{i,n}^2}} \qquad (25)$$

Semblance panels are widely used in seismic processing, in the stacking velocity analysis step for constructing an initial velocity model.

LINEAR AND NONLINEAR SEMBLANCE FIT

Although there are more advanced techniques to build the velocity model for migration, the stacking velocity analysis, by interpreting the Semblance panels, is still widely used because it is a method of low computational cost. But picking the Semblance panels still require a huge human effort.

Where the geology is not to complex it is assumed that the stacking velocity is the RMS velocity, used for migration of seismic data. An algorithm capable of automatic interpretation of the stacking velocity of Semblance panels can minimize the cost of this processing step.

The main problem to be solved is around Semblance panels. The global search optimization method, Genetic Algorithm, used for the automatic interpretation of Semblance, randomly generates an initial population of time interval velocities models and converts the values of the parameters of model space in binary, where the number of bits of the binary string has a relation (equation 8) with the number of model parameters.

On the initial models, is applied three GA procedures, which aims to generate descendants models better than the previous population. Therewith, a rapid convergence to the solution was obtained. The three processes are: selection, crossover and mutation. To validate the strategy of Semblance fit using the Genetic Algorithm, this methodology was applied in two synthetic CMP groups (Figs. 1a and 1b).

The evaluation of the models of the initial population as well as descendants of the models is called fitness. This value corresponds to the sum of the values of the Semblance that cross the RMS velocity curve, and controls whether a descending model is accepted or rejected. Therefore, it was necessary to convert the time interval velocity models for RMS. This procedure was used in two steps called linear and nonlinear fit of RMS velocity.

Linear fit

A first step, called linear inversion, adjusts a function of RMS velocities (equation 26) that maximizes the Semblance integration.

$$v_{RMS} = v_0 + \alpha t^\beta \qquad (26)$$

The values to be inverted are v_0, which is the initial or the water surface velocity, the gradient of the velocity (α) and β, which can be understood as the curvature of the velocity function. The initial parameterization

- $v_0 = [1450{:}1550{,}25]$ (m/s)

- $\alpha = [100{:}1000{,}100]$ (m/s^2)

- $\beta = [100{:}2000{,}100]$

defines the space of models. From this space is generated a population of RMS velocities. The initial population was evaluated (fitness calculation for each model) and passed through the processes of selection, crossover and mutation. The best models are used as parents of the next generation. After n iterations we will have a linear fit of Semblance, represented by Figures 2a and 2b.

The black curves represent the RMS velocities, adjusted in the process of inversion and the red curves represent the time interval velocities generated by the RMS conversion. This step is very important because the velocity obtained becomes the guide function that defines the boundaries of the model space for nonlinearfit.
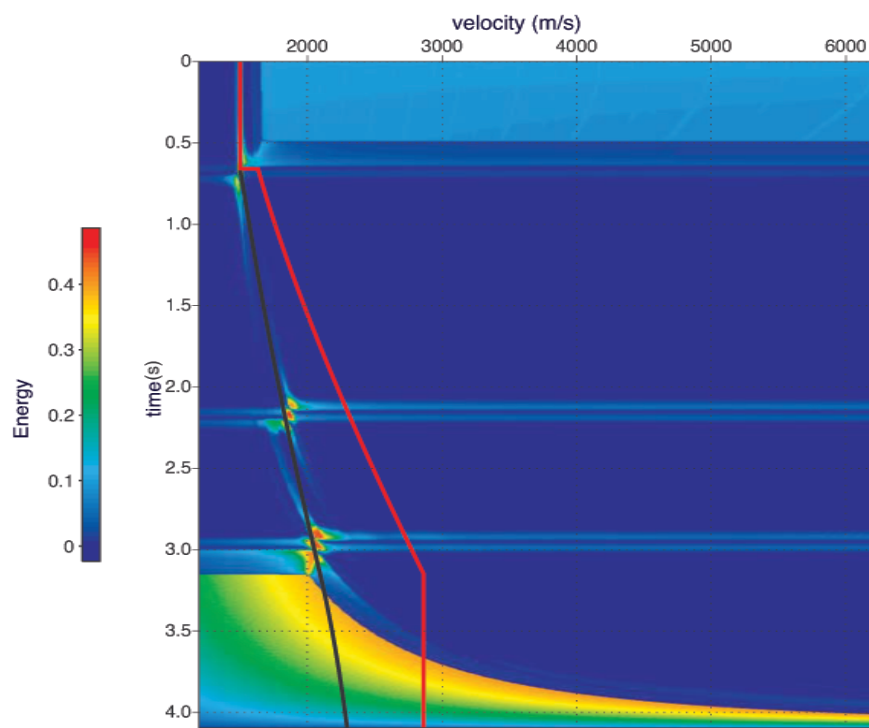
Nonlinear fit

The second step of the inversion performs nonlinear fitting (Figs. 4a and 4b). The inverted RMS velocity in the first step sets the guide function of this nonlinear step. Along this prior information, maximum and minimum limits are defined. After that, models are generated to compose the initial population, which can be seen both in the RMS and interval domain. The RMS shows the main values of the Semblance that are being contemplated for the model space, while the interval consider whether the velocity are consistent with the geological reality of the area.

Clapp et al. (1998) makes an interpolation of the velocity field RMS interpreted in the conventional way, using Semblance as weights for valuing the consistent events. Similarly, Semblance was used as the guide function conditioner to restrict the number of null space solutions to this nonlinear inverse problem. If the function guide is near the high values of Semblance, model space size is decreased (Figs. 3a and 3b).

From this space, multiple random models of time interval velocities are generated. These models has restrictions defined
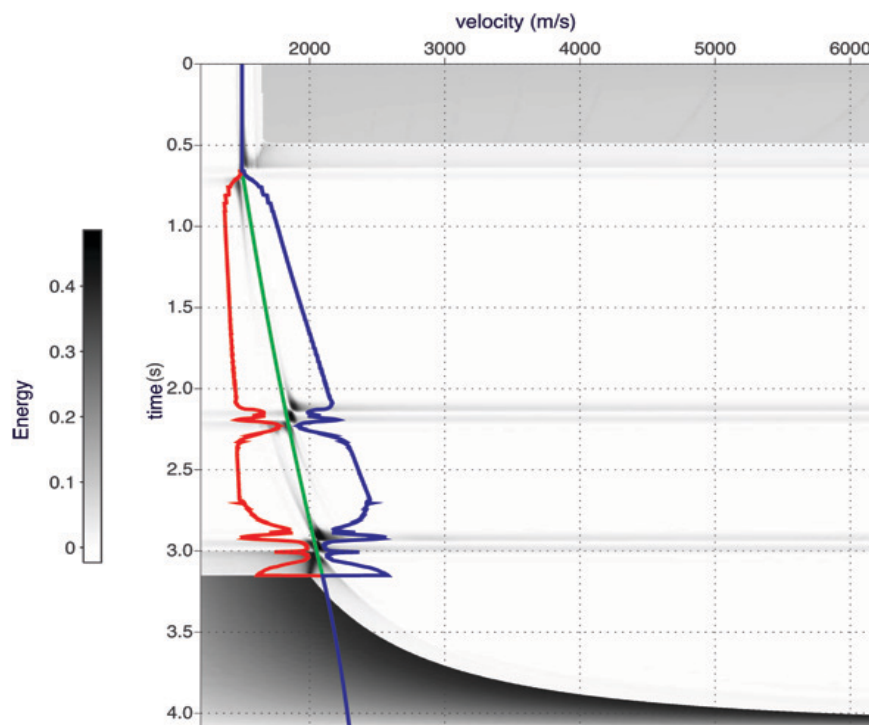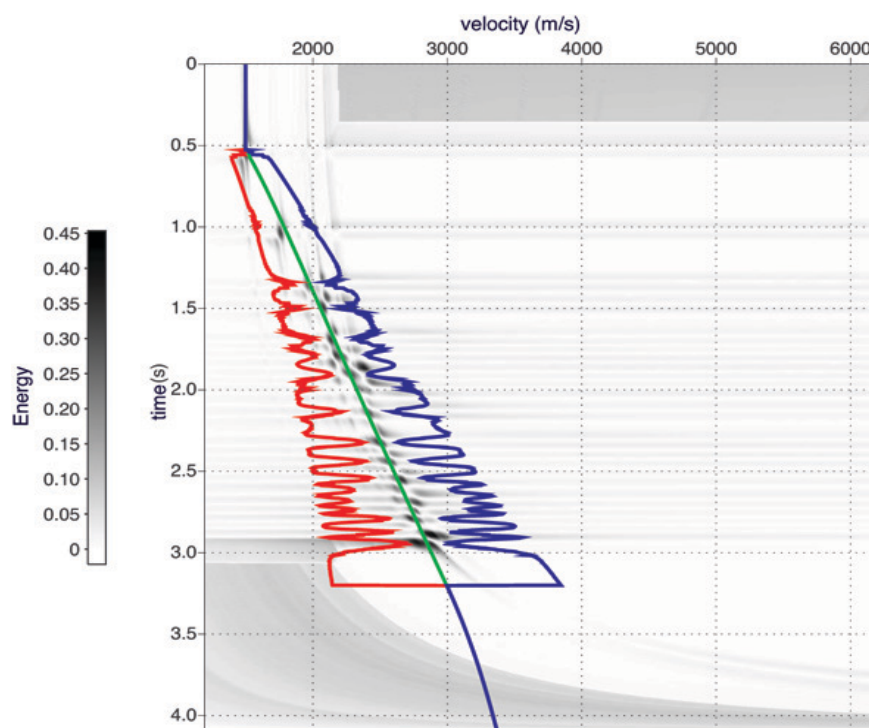
(a) 04 layer model
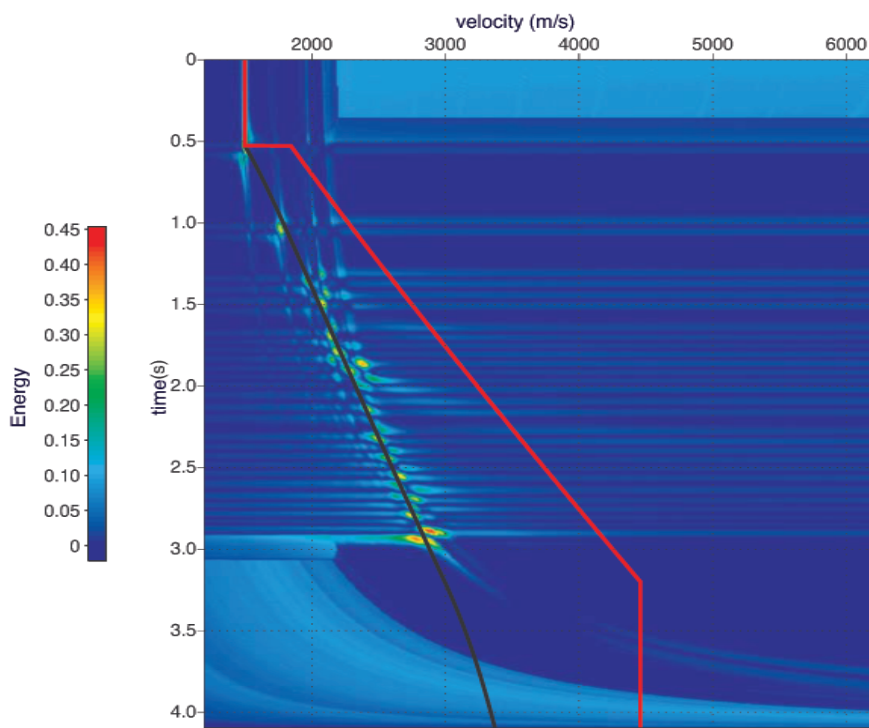
(b) 16 layer model

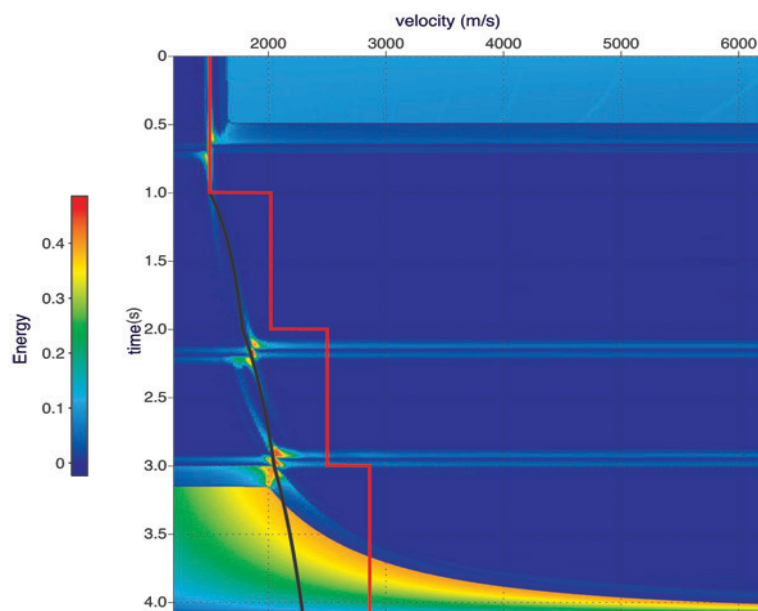**Figure 2** — Linear fitting of RMS velocity in Semblance panels using Genetic Algorithm.
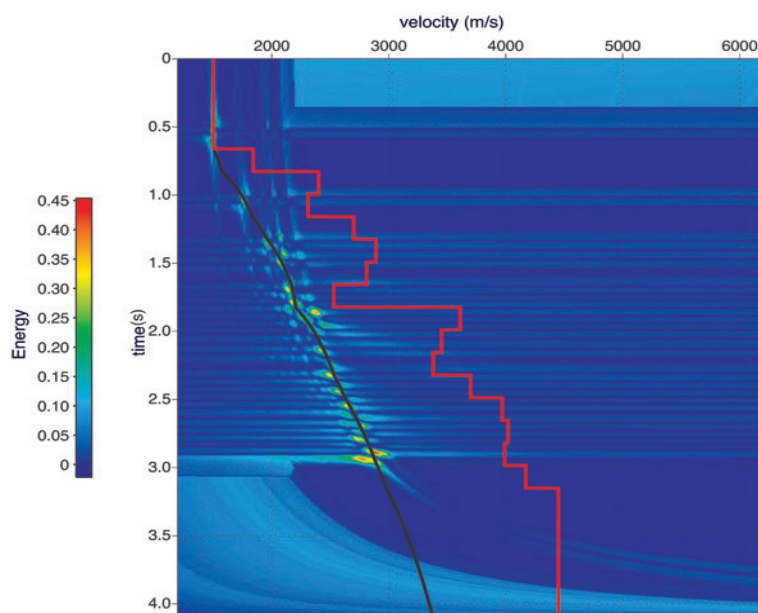
(a) 04 layer model

(b) 16 layer model

Figure 3 — Space models weighted by Semblance, defined by the limits blue curve (minimum) and red (maximum) and the guide function of RMS velocity in time (green).

(a) 04 layer model

(b) 16 layer model

Figure 4 — Nonlinear fit RMS velocity function in time (black curve) of the maximum values of Semblance using Genetic Algorithm. The red curve represents the best model of interval velocities in time.

a priori that maintains the geological consistency. It was made the conversion of these interval models for RMS and measure the initial population fitness. Then, these models suffer the processes of Genetic Algorithm in n iterations until some stopping criterion is satisfied. The result is a nonlinear fit of RMS velocities under the maximum values of the Semblance panel.

The nonlinear fit performed in these models (Fig. 4) uses two *a priori* information: the time of zero-offset reflection of the seafloor and basement (last event of reflection) and the number of interval velocities that want to invert. To better compare the inverted velocities with the true ones was maintained the real number of layers as input parameter.

The time interval velocities inverted (red line on Fig. 4) can be directly compared with the true velocities (Fig. 5). The convergence curves of Figure 6 show the normalized fitness of the population generated in each iteration. The red curve represents the best models while the blue curve, the lower fitness models, and the green curve is the average of the previous two, conducted throughout 400 iterations (generations).

Jequitinhonha's Basin real seismic data

At this stage of the work we perform the automatic velocity analysis in real 2D seismic data of Jequitinhonha's Basin. The Jequitinhonha's Basin is located on the east bank of the Brazilian coast. It occupies an area of approximately 25,685 km^2, of which only 5,535 km^2 are emerged. It lies between Almada basins on north and Cumuruxatiba on south. It comprises a rift basin that evolved into a typical passive margin (Rangel & Caixeta, 2007).

The seismic line used was acquired by the team 214 under the number 2660. Was acquired 1577 shots spaced 25 meters, with almost 40 km long. Each shot has 120 channels spaced 25 meters, with a 150 meters of minimum offset and maximum of 3125 meters, recorded in 1571 samples with sampling interval of 4 milliseconds, totaling 7 seconds of recording time.

Reordering of the data for the CMP domain was performed. There are 1517 CMP groups with 120 traces in maximum mul-
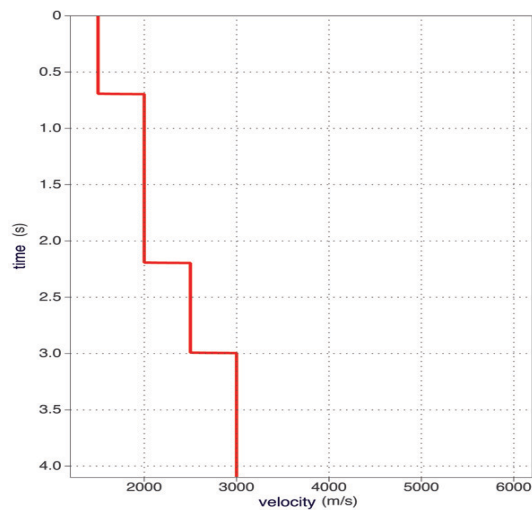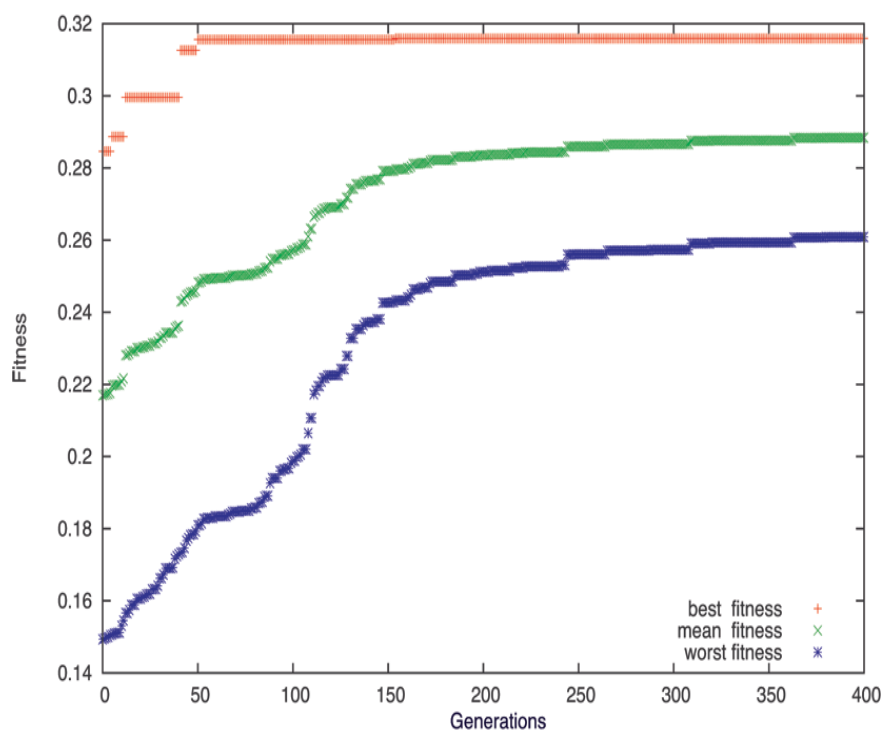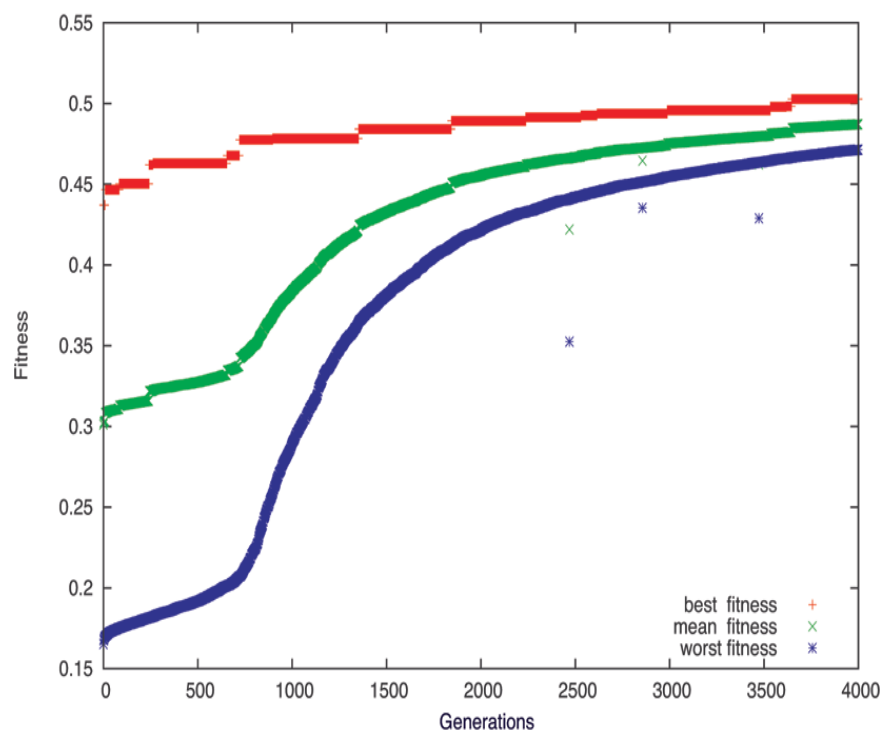

(a) 04 layer model
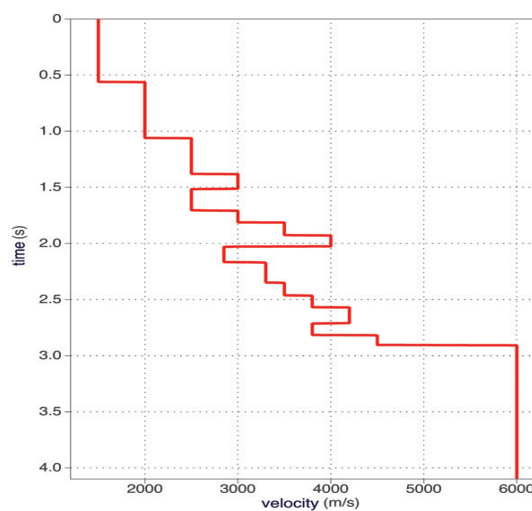
(b) 16 layer model

Figure 5 — True interval velocity function in time.

(a) 04 layer model

(b) 16 layer model

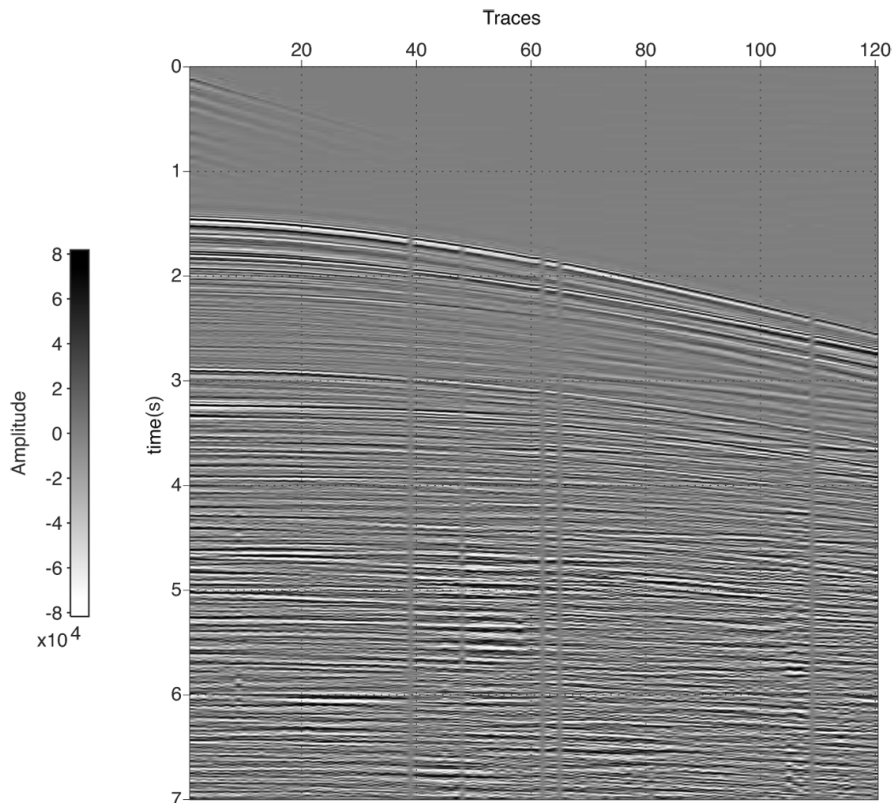**Figure 6** — Convergence curve of nonlinear fitting.

Figure 7 — 1061 CMP group of Jequitinhonha's Basin seismic line.

tiplicity, spaced 25 meters, whose position of the first CMP corresponds to shot number 60. All the pre-processing was carried out in software *LANDMARK SEISSPACE®*, which includes seismic data loading, geometry and deconvolution. We selected the 106 CMP group as an example for displaying the Semblance fit through the proposed method.

1061 CMP group

In this example we used the 1061 CMP group (Fig. 7) to perform the linear fit (Fig. 9a), which selects the best model of RMS velocity. This model is the guide function that defines the RMS space models (Fig. 9b). The guide function and the limits of space models are converted to interval, in which these last two, are conditioned by Semblance. It's set the model space for nonlinear fit whose result is expressed in Figure 10a.

The convergence curve (Fig. 8) shows a maximum fitness of 34%, in a total of 500 generations. This group has deep reflexion events almost horizontals (slightly hyperbolic) with high values of Semblance. This leads to relatively low overall fitness. The 1061 CMP group embraces important reflectors in the middle portion of the seismic section.

Jequitinhonha's 2D seismic section

After analyzing the individual CMP to find the best parameterization, we apply the methodology across the 2D line. The best model time interval velocities for each CMP, the final result of automatic nonlinear fit of Semblance panels is represented by Figure 12.

The conversion of the time interval velocity field to RMS was done. The result is a stacking velocity field obtained by automatic nonlinear fit of Semblance panels (Fig. 11). This field is used to make the NMO correction and stacking of 1517 CMP groups to generate stacked section of Jequitinhonha's Basin (Fig. 13).

CONCLUSIONS

We presented a methodology for automatic generation of stacking velocity field through nonlinear fit of Semblance panels. The inverted velocity can be compared with the true velocity functions.

Despite the limitations of the method, the results show that the algorithm used are robust and efficient to achieve the proposed objective. The 2D model inversion of Jequitinhonha's Basin totaled 1517 Semblance panels interpreted.
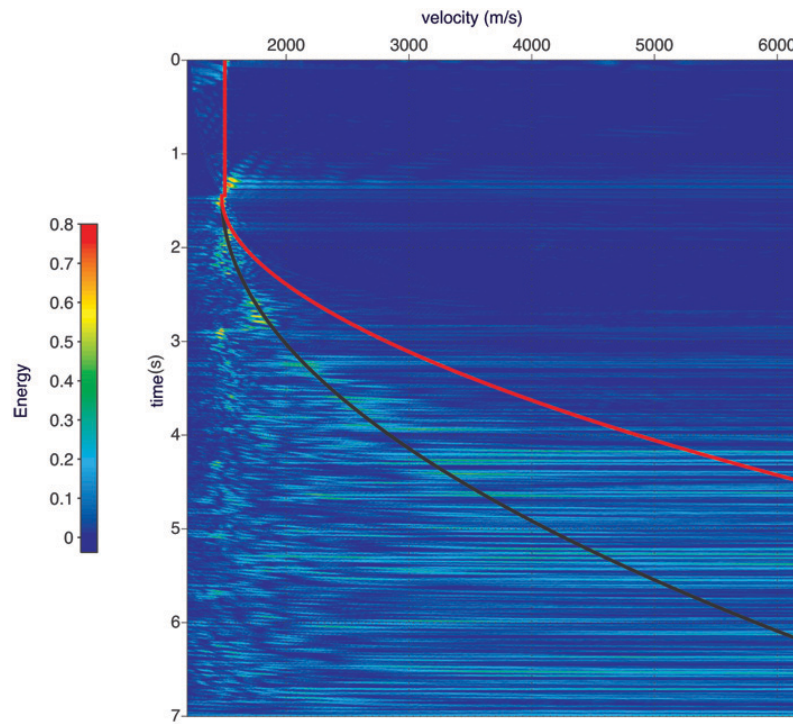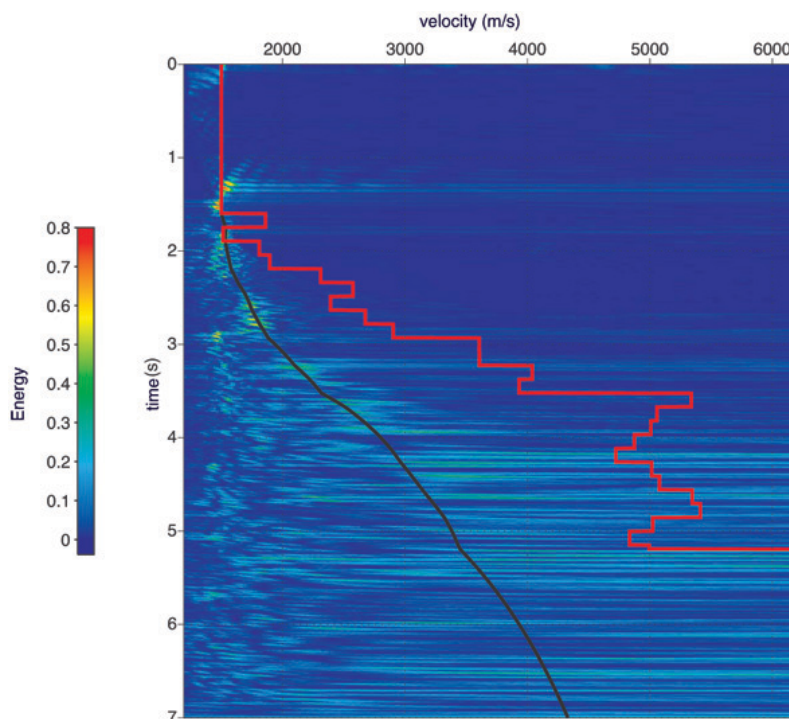
(a) The RMS velocity in black line and the interval in red line.
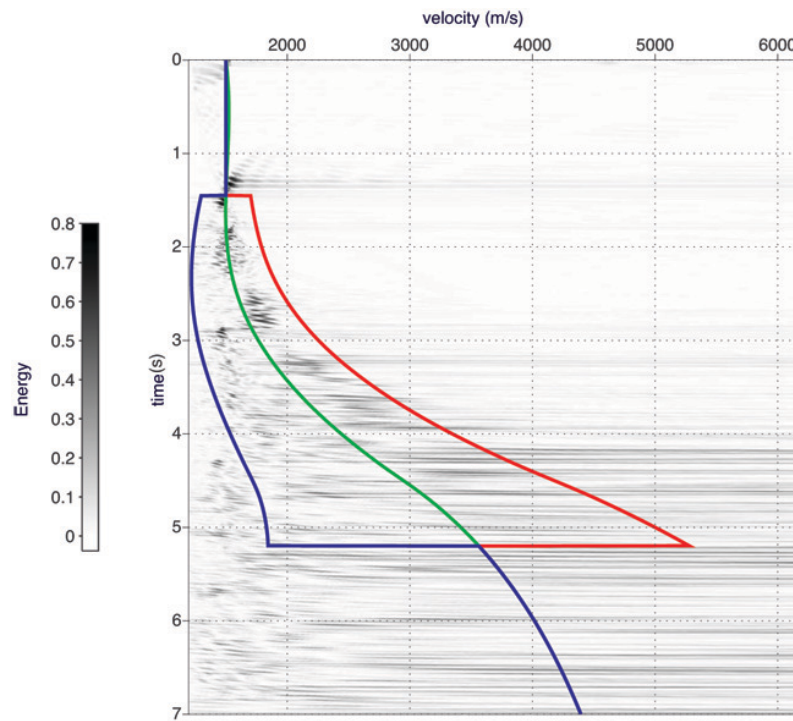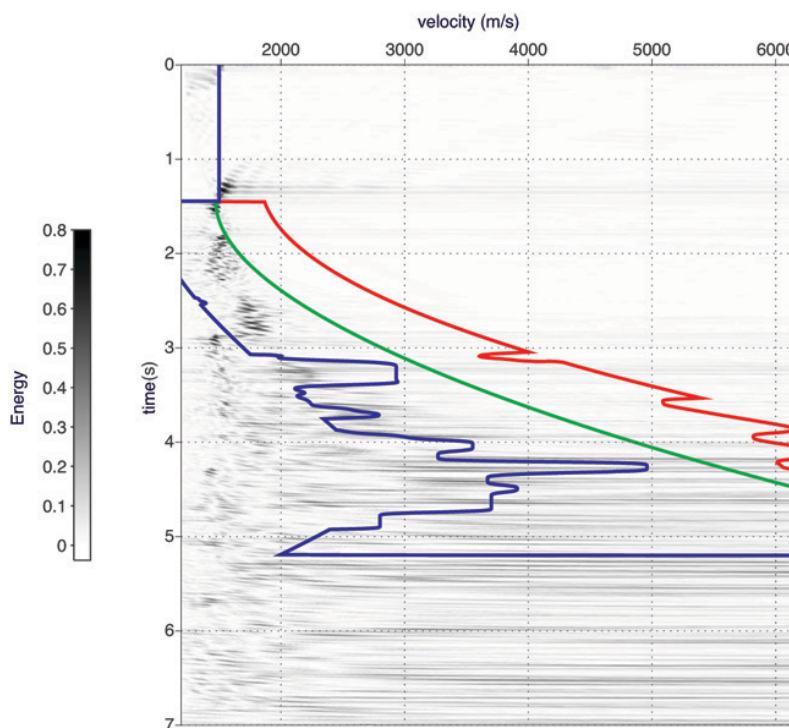

(b) Upper bound in red, lower bound in blue and guide function in red.

Figure 8 — 1061 CMP of Jequitinhonha's Basin linear fit of Semblance (a) and space model generated from the guide function RMS (b).

(a) The RMS velocity in black line and the interval in red line.

(b) Upper bound in red, lower bound in blue and guide function in red.

Figure 9 — 1061 CMP of Jequitinhonha's Basin linear fit of Semblance (a) and space model generated from the guide function RMS weigthed by the Semblance (b).
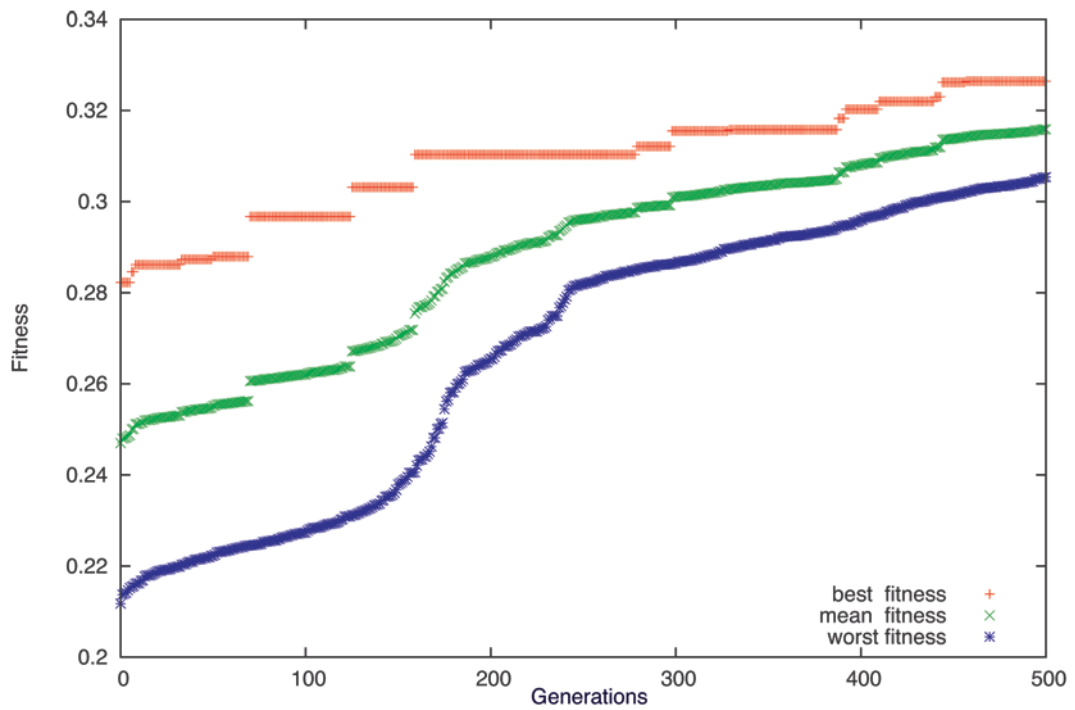
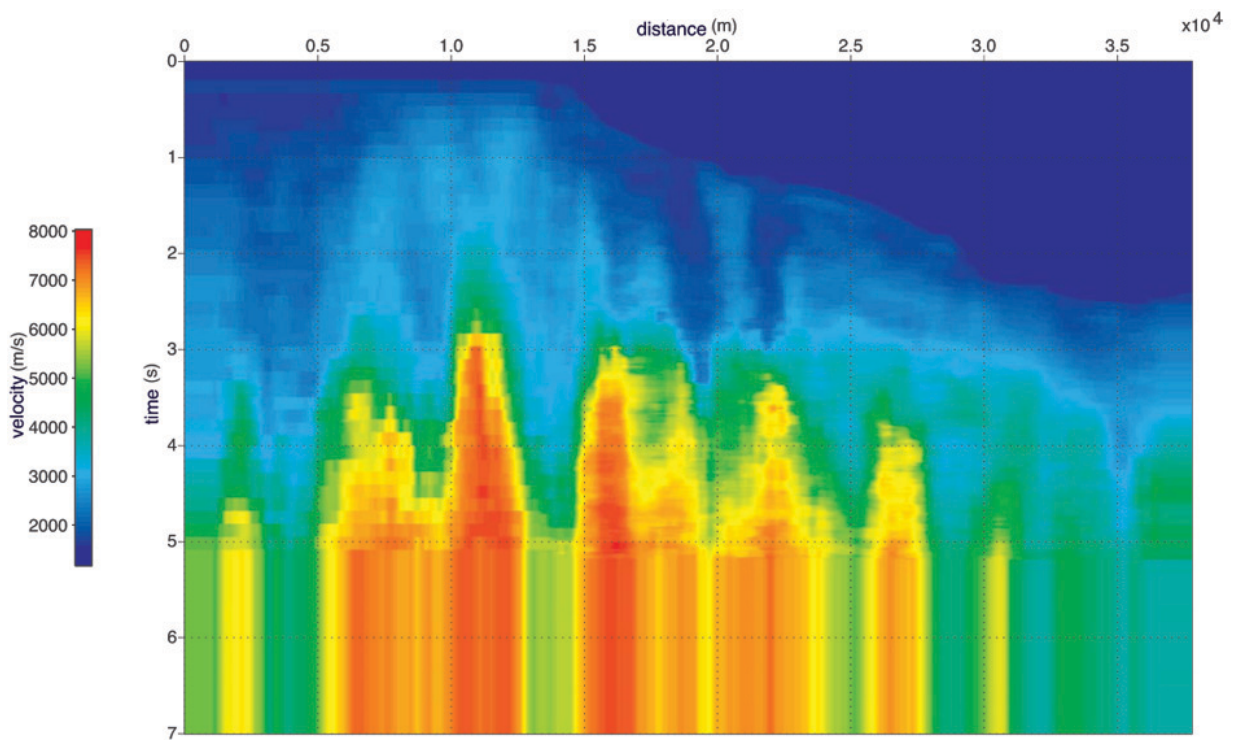**Figure 10** — Convergence curve of 1061 CMP group of Jequitinhonha's Basin.

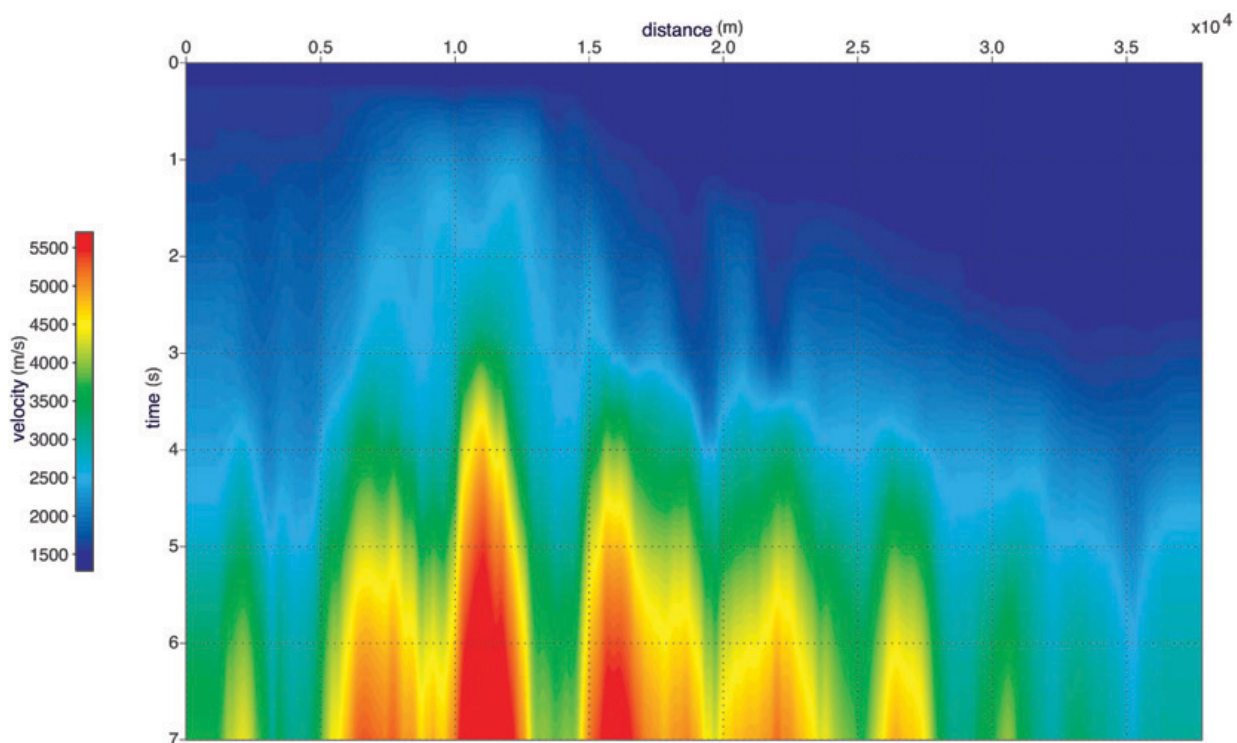**Figure 11** — Jequitinhonha's Basin inverted time interval velocity.

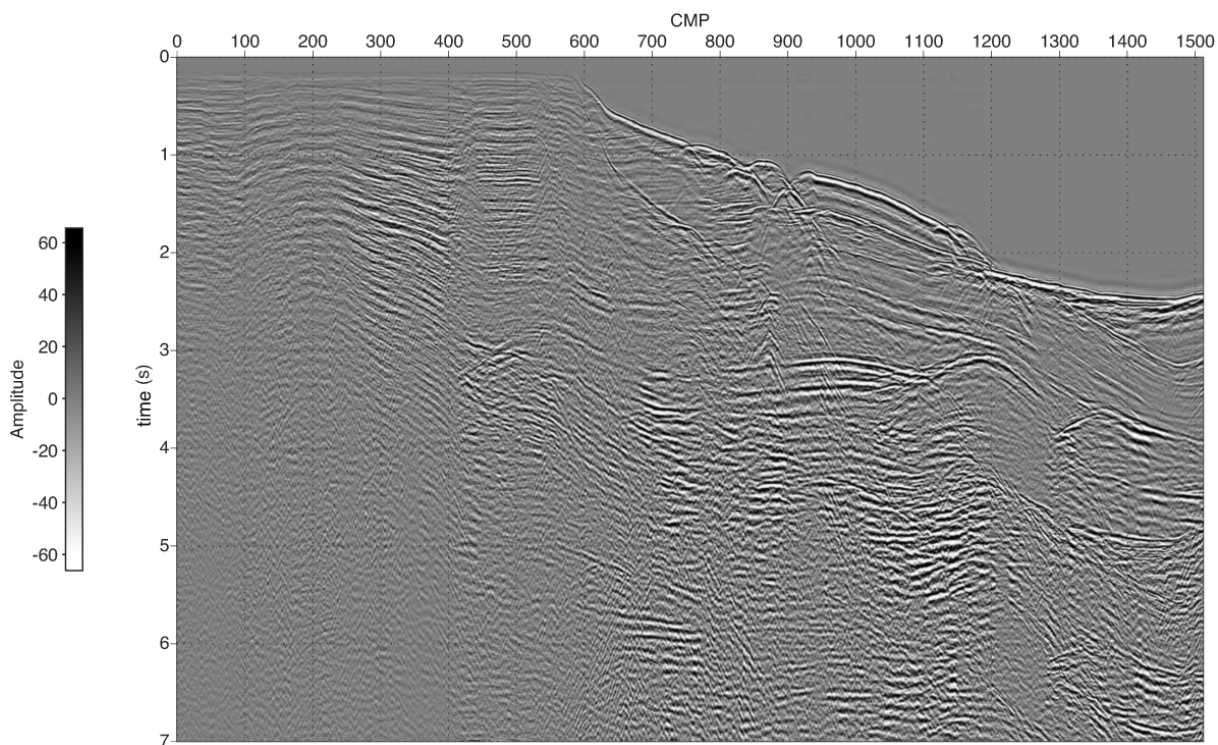**Figure 12** — Jequitinhonha's Basin inverted RMS velocity.

Figure 13 — Zero-offset section of Jequitinhonha's Basin generated with inverted RMS velocity.

In the initial population, 200 models and 800 iterations, or generations were used. This counterpoised the computational cost by the quality of the inverted interval velocity.

We can allow us to infer that the estimated velocity is close to the true result, or at least close to the global minimum.

ACKNOWLEDGEMENTS

REFERENCES

AMUNDSEN L & URSIN B. 1991. Frequency-wavenumber inversion of acoustic data. Geophysics, 56: 1027–1039.

BILLETTE F & BRANDSBERG-DAHL S. 2005. The 2004 BP velocity benchmark. In: 67th Annual Internat. Mtg., EAGE, Expanded Abstracts, B035, EAGE.

CASTLE RJ. 1994. A theory of normal moveout. Geophysics, 59: 983–999.

CHARARA M, BARNES C & TARANTOLA A. 2000. Full waveform inversion of seismic data for a viscoelastic medium. In: Lecture Notes in Earth Science: Springer. Volume 92, pp. 68–81.

CLAPP RG, SAVA P & CLAERBOUT JF. 1998. Interval velocity estimation with a nullspace. SEP Report, 97: 147–157.

CLAYTON RW & STOLT RH. 1981. A Born-WKBJ inversion method for acoustic reflection data. Geophysics, 46: 1559–1567.

DELOST M, VIRIEUX J & OPERTO S. 2008. First-arrival traveltime tomography using second generation wavelets. Geoph. Prosp., 56(4): 505–526.

GAO F, LEVANDER A, PRATT RG, ZELT CA & FRADELIZIO GL. 2007. Waveform tomography at a groundwater contamination site: Surface reflection data. Geophysics, 72(5): G45–G55.

HONG T & SEN M. 2009. A new MCMC algorithm for seismic waveform inversion and corresponding uncertainty analysis. Geophys. J. Int., 177: 14–32.

JAISWAL P, ZELT CA, DASGUPTA R & NATH KK. 2009. Seismic imaging of the Naga Thrust using multiscale waveform inversion. Geophysics, 74(6): WCC129–WCC140.

KEY SC & SMITHSON SB. 1990. New approach to seismic-reflection event detection and velocity determination. Geophysics, 55: 1057–1069.

KOSLOFF D, SHERWOOD J, KOREN Z, MACHET E & FALKOVITZ Y. 1996. Velocity and interface depth determination by tomography of depth migrated gathers. Geophysics, 61(5): 1511–1523.

LUMLEY DE. 1997. Monte Carlo automatic velocity picks. SEP Report, 75: 1–25.

MENKE W. 1989. Geophysical data analysis. Discrete inverse theory. volume 45 of International Geophysics Series: Academic Press Inc.

PORSANI MJ. 2008. Uma formulação de otimização para os algoritmos gradiente e gradiente conjugado pré-condicionados. Lecture Notes. Universidade Federal da Bahia, Brazil. 34 pp.

PORSANI MJ, STOFFA PL & SEN M. 2000. Fitness functions, genetic algorithms and hybrid optimization in seismic waveform inversion. J. Seis. Expl., 9: 143–164.

PRATT RG, SHIN C & HICK GJ. 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. Geophys. J. Int., 133(2): 341–362.

RANGEL HD, OLIVEIRA JLF & CAIXETA JM. 2007. Bacia de Jequitinhonha. Boletim de Geociências da Petrobras, 15: 475–483.

ROWBOTHAM PS & PRATT RG. 1997. Improved inversion through use of the null space. Geophysics, 62: 869–883.

SEN M & STOFFA PL. 1995. Global optimization methods in geophysical inversion, volume 4 of Advances in Exploration Geophysics. Elsevier.

SYMES WW, SUN D & ENRIQUEZ M. 2011. From modelling to inversion: designing a well-adapted simulator. Geoph. Prosp., 59(5): 814–833.

THOMAS JE. 2001. Velocidades sísmicas. Universidade Petrobras.

VALENCIANO AA, BIONDI BL & CLAPP RG. 2009. Imaging by target-oriented wave-equation inversion. Geophysics, 74(6): WCA109–WCA120.

VARELA CL, STOFFA PL & SEN MK. 1998. Background velocity estimation using nonlinear optimization for reflection tomography and migration misfit. Geoph. Prosp., 46(1): 51–78.

VIRIEUX J & OPERTO S. 2009. An overview of full-waveform inversion in exploration geophysics. Geophysics, 74(6): WCC1–WCC26.

NOTES ABOUT THE AUTHORS

Danian Steinkirch de Oliveira. BS in Geology (UFPR, 2002). M.Sc. in Geophysics (UFBA, 2013). Works as a geophysicist in the research center of Petrobras with processing and imaging seismic data since 2003. Currently, is doing Ph.D. at COPPE/UFRJ, in the Department of Naval and Ocean Engineering. Area of interest: processing, inversion and imaging of seismic reflection data.

Milton José Porsani. BS in Geology (USP, 1976). M.Sc. in Geophysics (UFPA, 1981). Ph.D. in Geophysics (UFBA, 1986). From 1986 to the present is a researcher CPGG-UFBA. Post doctorate in Geophysics at the Institute for Geophysics at the University of Texas, the September/92 – October/93. Professor of Geology and Geophysics of IGEO-UFBA (1990). Professor in the field of oil exploration (2000). Researcher at CNPq. Coordinator of the oil exploration CPGG-UFBA. Area of interest: development of methods and algorithms for mono and multichannel filtering, seismic processing, inversion of geophysical data.

Paulo Eduardo Miranda Cunha. BS in Physics (UFMG, 1978). M.Sc. in Geophysics (UFBA, 1997). Ph.D. in Geophysics (COPPE, 2004) with a stage of one year in Steklov Institute of Mathematics, Russian Academy of Sciences, St. Petersburg, Russia. From 2004 to the present works as a geophysicist in the research center of PETROBRAS. Area of interest: paraxial and gaussian beams methods, tomographic inversion o velocity field and imaging of seismic reflection data.