Katherine-A., Flórez; Sergio-Alberto, Abreo; Ana-B., Ramírez
OPTIMAL CODING OF BLENDED SEISMIC SOURCES
FOR 2D FULL WAVEFORM INVERSION IN TIME

ecoPETROL

OPTIMAL CODING OF BLENDED SEISMIC SOURCES FOR 2D FULL WAVEFORM INVERSION IN TIME

CODIFICACIÓN ÓPTIMA DE FUENTES SISMICAS MEZCLADA PARA UNA INVERSIÓN EN TIEMPO DE ONDA COMPLETA 2D

Flórez, Katherine-A. [a]*, Abreo, Sergio-Alberto[a], Ramírez, Ana-B.[a]

ABSTRACT

Full Waveform Inversion (FWI) schemes are gradually becoming more common in the oil and gas industry, as a new tool for studying complex geological zones, based on their reliability for estimating velocity models. FWI is a non-linear inversion method that iteratively estimates subsurface characteristics such as seismic velocity, starting from an initial velocity model and the preconditioned data acquired.

Blended sources have been used in marine seismic acquisitions to reduce acquisition costs, reducing the number of times that the vessel needs to cross the exploration delineation trajectory. When blended or simultaneous without previous de-blending or separation, stage data are used in the reconstruction of the velocity model with the FWI method, and the computational time is reduced. However, blended data implies overlapping single shot-gathers, producing interference that affects the result of seismic approaches, such as FWI or seismic image migration.

In this document, an encoding strategy is developed, which reduces the overlap areas within the blended data to improve the final velocity model with the FWI method.

RESUMEN

La inversión de onda completa (FWI, por sus siglas en inglés) ha llamado la atención de la comunidad de exploración de gas y petróleo, como una nueva herramienta para el estudio de zonas geológicas complejas, en donde es necesario el desarrollo de técnicas para la estimación de modelos de velocidad confiables. La inversión de onda completa es un método de inversión no lineal que iterativamente estima características del subsuelo como la velocidad sísmica, partiendo de un modelo inicial de velocidad y el dato adquirido en campo.

Las fuentes blended o simultáneas han sido usadas en la adquisición sísmica marina con el fin de reducir los costos de adquisición, disminuyendo el número de veces que el buque de exploración debe pasar por la trayectoria delimitada de exploración. Cuando los datos blended son utilizados en la reconstrucción del modelo de velocidad empleando el método FWI, evitando la etapa previa de de-blending o separación, el tiempo de procesamiento es reducido.

Sin embargo, un dato blended implica la superposición entre los disparos individuales contenidos en él, produciendo interferencia que afecta a el resultado final de la FWI. En este documento, se desarrolla una estrategia de codificación que disminuye las zonas de superposición dentro del dato blended con el fin de mejorar el modelo de velocidad final.

AFFILIATION

[a] CPS Research Group, Depart. Electrical, Electronic and Telecom. Eng- Universidad Industrial de Santander, carrera 27 calle 9 Bucaramanga, Colombia
*email: katherine.florez12@gmail.com

1 INTRODUCTION

Seismic exploration of oil and gas in complex geological areas requires the development of techniques for the estimation of reliable velocity models, to generate a successful migration of seismic traces. Full waveform inversion (FWI) has been used to estimate high-resolution velocity models [1], and it reaches more detailed models than those attained by traditional inversion methods, as sparse-spike methods, recursive inversion, among others [2].

In marine seismic exploration, the Towed Streamer is the most widely used geometry, where one or more streamers containing the acquisition transducers are located to obtain the desired coverage. Another geometry used in marine acquisition, is the 2D Towed Streamer blended (blended geometry) [3]. It consists in a single cable or streamer that is dragged behind the exploration vessel together with various sound sources [4]. Subsurface reflections are below the line drawn by the movement of the vessel, resulting in a two-dimensional blended shot gather (horizontal and vertical).

A blended shot gather involves temporary overlapping of multiple shots randomly located in the same acquisition. This is an emerging technology that reduces acquisition costs by decreasing the number of recorded shots, as it performs simultaneous source shooting (i.e. super-shot), as shown in **Figure 1**. It also decreases the number of times that the ship must pass through the traced path, thus reducing the seismic acquisition time and costs [5].

Most of the scientific community has focused its efforts on the de-blending stage, which is intended separate the blended sources to obtain non-overlapping shot-gathers using filters, deconvolution, muting, etc … The generation of coded sources has been performed using only one encoding parameter such as time delay [5]; phase [6] or weight [7]. In those cases, the encoding process is associated with different probability distributions or empirical tests.

We present herein a formula for the acquisition of super-shots for FWI, without the de-blending stage. The encoding methodology is based on the minimum coherence problem and the estimation of the velocity models is performed via FWI, using blended data. Simulations and results are used to examine the proposed blending methodology from a practical point of view.
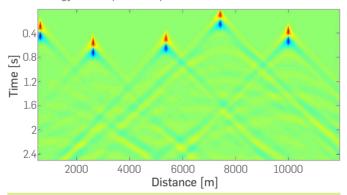

Figure 1. Blended Shot-Gather

2. THEORETICAL FRAME

Blended data implies the acquisition of multiple shots at the same time (super-shot), producing an overlap between the traditional single shot-gathers. This overlap generates interference that affects the result of seismic approaches, such as FWI or seismic image migration.

A mathematical expression for a single shot, *src*, of the wavefield measured at the surface in the time domain can be obtained with the numerical solution of the 2D acoustic wave equation with constant density given by

$$src(x_0, t_0) = \frac{\partial^2 p(x,z,t)}{\partial x^2} + \frac{\partial^2 p(x,z,t)}{\partial z^2} - \frac{1}{m^2(x,z)} \frac{\partial^2 p(x,z,t)}{\partial t^2} \quad (1)$$

where, *x* and *t* are the spatial position and the time delay for the single source, respectively; $m \in R^{N_x \times N_z}$ (for the 2D case) is the acoustic velocity of the medium as a function of space location (*x,z*) and *p* is the scalar pressure field for different time steps, *t*. The observed blended data at the surface, p_{bl}, is given by the linear sum of the *N* shots with encoding time delay, t_n, and position, x_n, as

$$p_{bl} = \sum_{n=1}^{N} p(x - x_n, z = 0, t - t_n) \quad (2)$$

where, $t_n \in Q_t$ and $x_n \in Q_x$ are the set of feasible values for both variables.

BLENDED CODED SOURCES

A blended data or super-shot p_{bl} in terms of the pressure wavefield at the surface of a single shot located in the position x_0 and the time delay t_0, $p(x-x_0, z=0, t-t_0)$, can be defined as the convolution (*) between the pressure wavefield and the matrix *A*, as

$$p_{bl} = A * p(x - x_0, z = 0, t - t_0) \quad (3)$$

where $A_N \in R^{N_x \times N_t}$ is the binary coded matrix for each N super-shot as shown in **Figure 2**. The position and the rows represent the time delay when the source is activated. Each source is represented by one (1) in the coordinate (x_n, t_n). It should be noted that Equation 3 is equivalent to the expression in the Equation 2.

Each binary coded matrix A_N is obtained distributing *k* sources within the super-shot with a minimum offset without overlapping between the intersection point of direct waves, (x_p, t_p), and the position of each source, as **Figure 3** shows. The intersection point is defined as the crossing point of direct arrivals, which are propagated from source to receiver only in water. In real life, the water layer is known from the preliminary studies that delineate the acquisition trajectory and the type of vessel used.
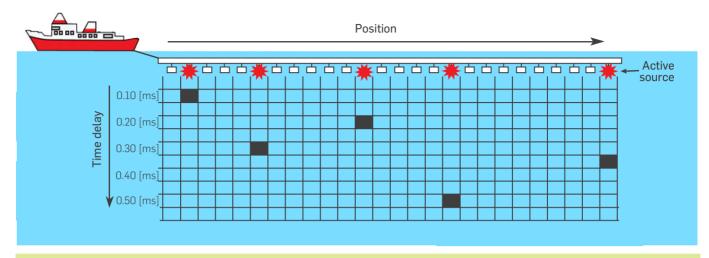
Figure 2. Physical representation of coded matrix A for one super-shot with five simultaneous sources, where the columns are the source position and the rows represent the time delay when the source is activated. The black cells indicate the activation moment of each source
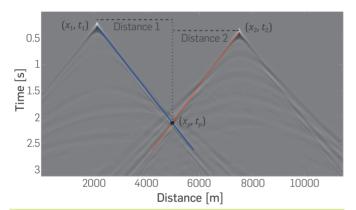

Figure 3. Representation of the intersection point (x_p, t_p), with the cross point between the direct wave of first shot (blue line) and the direct wave of second shot (red line). The distance between the position of each source and the intersection point (distance 1 and distance 2) is compared with the minimal offset clear for determine if this distribution fixes.

Direct waves are modeled using the equations for ray tracing in laterally homogeneous Earth models [8]. Being the ray parameter p defined as

$$p = u \, sin\theta \tag{4}$$

where $u=1/v$ is the slowness, with v as the velocity of the layer and θ is the ray incidence angle (from vertical). The expression for the surface-to-surface travel time $T(p)$ and the total distance $X(p)$ for a constant layer are,

$$X(p) = 2p \frac{\Delta z_i}{(u_i^2 - p^2)^{\frac{1}{2}}} \tag{5}$$

$$T(p) = 2 \frac{u_i^2 \Delta z_i}{(u_i^2 - p^2)^{\frac{1}{2}}} \tag{5a}$$

being Δz_i the thickness of the layer, in this case the water layer.

Encoding N blended shots require to optimize each coded matrix A_i with the aim to each one acquires data from different areas. Each matrix $Ai \in R^{N_A \times 1}$ becomes a column of matrix $B \in R^{N_A \times N}$ with $N_A = N_x \cdot N_t$, as shown in **Figure 4**. An optimal coding is obtained finding the minimal mutual coherence between the columns of the design matrix B. The minimal mutual coherence implies linearly independent columns of the design matrix B, which involve different coded matrices with the least interference between super-shots.
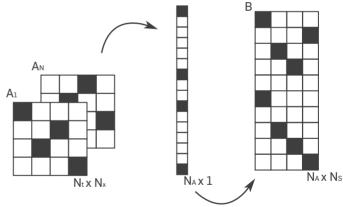

Figure 4. Process to generate the design matrix B where each column is the vectorization of matrix A_i

The mutual coherence of matrix B is the largest absolute normalized inner product between different columns of B [9], and it is given by

$$\mu(B) = \max_{1 \leq i \neq j \leq m} \frac{|b_i^T b_j|}{||b_i||_{l2} \, ||b_j||_{l2}} \tag{6}$$

where b_i is the i-th column of B, being each b_i column function of the variables x_n and t_n.

The minimization problem of mutual coherence of the design matrix B as

$$\underset{x_n, t_n, B}{argmin} \; \mu(B) \, s.t. \, b_i = f_i(x_n, t_n), x_n \in Q_x, t_n \in Q_t \tag{7}$$

To simplify the problem, the modification of the optimization problem showed by Obermeier [10] can be used. It is possible to use only the upper triangular components of the matrix $\mathbf{U}=\mathbf{B}^T\times\mathbf{B}$ to compute the mutual coherence, where the elements of the upper and lower triangular matrix are the same, Equation 8.

$$\mathbf{U}=\begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,N-1} & u_{1,N} \\ u_{1,2}^* & u_{2,2} & \cdots & u_{2,N-1} & u_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ u_{1,N-1}^* & u_{2,N-1}^* & \cdots & u_{N-1,N-1} & u_{N-1,N} \\ u_{1,N}^* & u_{2,N}^* & \cdots & u_{N-1,N}^* & u_{N,N} \end{pmatrix} \quad (8)$$

The matrix $\mathbf{U}$ can be expressed as the vector $u=(u_{1,2},u_{1,3},...,u_{1,N},u_{2,3},...,u_{N-1,N})$ to compactly represent these elements. We use the notation $u_{i,j}$ to represent the element in the vector u that corresponds to the value of U at the location (i,j) and the definition of $f(x_n,t_n)$ is expanded as follows:

$$f(x_n,t_n)=(f_1(x_n,t_n),f_2(x_n,t_n),...,f_N(x_n,t_n)) \quad (9)$$

Where $f_i(x_n,t_n)$ computes the i-th column of B.

With these modifications, the equivalent mutual coherence minimization problem can be formulated as

$$\underset{x_n,t_n,u}{argmin}||u||_{l2} \, s.t. \, u_{i,j}, x_n \in Q_x, t_n \in Q_t \quad (10)$$

$$where \; u_{i,j}=\frac{|f_i(x_n,t_n)^T f_j(x_n,t_n)|}{||f_i(x_n,t_n)||_{l2}\,||f_j(x_n,t_n)||_{l2}}, 1\leq i\neq j\leq m$$

This optimization problem can be formulated without constraints using the Augmented Lagrangian method with equality constraints [11],

$$L_B(x_n,t_n,u,\lambda;\rho)=||u||_{l2}+$$

$$\sum_{1\leq i\neq j\leq N}\frac{\beta}{2}\left|u_{i,j}-\frac{|f_i(x_n,t_n)^T f_j(x_n,t_n)|}{||f_i(x_n,t_n)||_{l2}\,||f_j(x_n,t_n)||_{l2}}+\frac{\lambda_{i,j}}{\rho}\right|^2 (11)$$

where ρ is a positive scaling factor that increases at each iteration to ensure that the algorithm converges, and λ is the Lagrange multipliers that are updated in each k iteration as

$$\lambda_{i,j}^{k+1}=$$

$$\lambda_{i,j}^k-\rho^k\left(u_{i,j}^k-\frac{|f_i(x_n^k,t_n^k)^T f_j(x_n^k,t_n^k)|}{||f_i(x_n^k,t_n^k)||_{l2}\,||f_j(x_n^k,t_n^k)||_{l2}}\right) \quad (12)$$

The coherence minimization procedure is summarized in Algorithm 1 (see **Table 1**).

Table 1. Algorithm 1. Augmented Lagrangian method for coherence minimization problem

```
Choose initial values → x⁰ₙ,t⁰ₙ,ρ¹;

Set → u⁰ᵢ,ⱼ = |fᵢ(xₙ,tₙ)ᵀ fⱼ(xₙ,tₙ)| / ||fᵢ(xₙ,tₙ)||₁₂||fⱼ(xₙ,tₙ)||₁₂ , λ¹ᵢ,ⱼ = 0;

for l = 1:k  do
        solve → (x¹ₙ,t¹ₙ,u¹)= argmin Lв(xₙ,tₙ,u,λ;ρ);
                              xₙ,tₙ,u
        update → xₙ,tₙ,u ;
        update the langrange multipliers→

        λᵏ⁺¹ᵢ,ⱼ = λᵏ⁺¹ᵢ,ⱼ -ρᵏ(uᵏᵢ,ⱼ - |fᵢ(xᵏₙ,tᵏₙ)ᵀ fⱼ(xᵏₙ,tᵏₙ)| / ||fᵢ(xᵏₙ,tᵏₙ)||₁₂||fⱼ(xᵏₙ,tᵏₙ)||₁₂);

        update→ ρˡ⁺¹>ρˡ;

end(for)
```
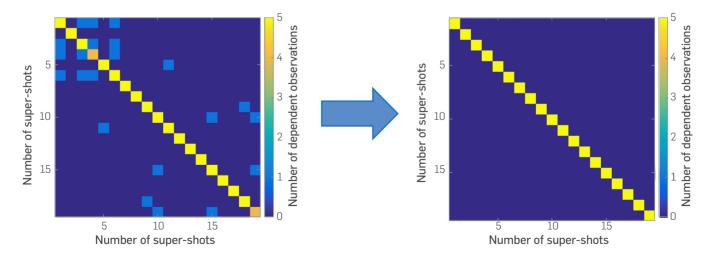

Figure 5. Representation of the minimization problem to diagonalize the initial matrix **U** in a diagonal matrix. The final matrix **U** represents 20 super-shots and 5 simultaneous sources within each super-shot. The diagonal represents 5 different coded sources in each super-shot and the zero values in the triangular matrix a zero correlation with each acquisition source.

The minimization problem, with Obermeier modification [10], can be seen as to diagonalize the matrix $U = B^T \times B$ that contains non-zero values outside the diagonal in a diagonal matrix, as shown is **Figure 5**. Where the numerical value of the diagonal elements represents the number of simultaneous shots with different codification within each super-shot and the absence of values outside the diagonal represents a zero correlation between super-shots, that is, a source distribution where the encoding parameters are different in each super-shot and, in turn, they are different with respect to the rest of the shots.

FULL WAVEFORM INVERSION

Full Waveform Inversion (FWI) is a processing tool that estimates the subsurface parameters from an initial model and the acquired data at the surface. FWI updates the model by decreasing iteratively the misfit function given by the squared l_2-error norm between the acquired data, d_{obs}, and the modeled data, d_{mod}.

The misfit function [12] is given by

$$J(m) = \frac{1}{2}||d_{mod} - d_{obs}||_2^2 \tag{13}$$

It can be minimized using a gradient descent method [1] by updating iteratively the velocity model as

$$m_{k+1} = m_k + \Delta m \tag{14}$$

where

$$\Delta m = -\alpha_k [H(m_k)]^{-1} G(m_k) \tag{15}$$

being Δm the search direction where $G(m_k)$ is the gradient that represents the evolution direction in the misfit function, $H(m_k)$ is the Hessian, which represents the magnitude of this direction, both evaluated at m_k and α_k is a scalar step length value.

The velocity gradient for the 2D acoustic and isotropic wave equation is computed using the first order adjoint state method by [13]

$$G(m_k) = \sum_s \frac{-2}{m_k^3} \int_0^T q_s(x, z, T - t) \frac{\partial^2 p_s(x, z, t)}{\partial t^2} dt \tag{16}$$

where T is the recorded time and t is the time variable, p_s is the forward wavefield generated by the source $src(x,z)$; q_s is the backward pressure wavefield generated by the residual data (the difference between the modelled and acquired data); s represents a specific shot; each shot produces its own gradient and therefore, the final gradient per iteration is the summation of all the gradients.

The Hessian matrix is not directly computed, but an approximation of the product between the inverse of the Hessian matrix and the gradient $([H(v^k)]^{-1} g(v^k))$ using $L\text{-}BFGS$ method was defined by Liu and Nocedal [14]. This approximation uses the last n gradients and velocity models to compute the step forward and it needs at least two gradients and two velocity models to obtain an approximation of the search direction $\Delta m = r$.

Algorithm 2 (see **Table 2**) shows the pseudocode of this method, where $s_k = m_{k+1} - m_k$, $y_k = g_{k+1} - g_k$ and $\sigma_k = 1 / y_k^t s_k$. The matrix D_k^0 is approximated by a diagonal $D_k^0 = \gamma_k I$, with

$$\gamma_k = \frac{s_{k-1}^t y_{k-1}}{y_{k-1}^t y_{k-1}} \tag{17}$$

Table 2. Algorithm 2. L-BFGS algorithm pseudocode.

```
q ← G(mₖ);
for i=k-1:-1:k-n
        εi ← σᵢsᵢᵀq;
        q ← q-εᵢyᵢ;
end(for)
r ← Dₖ⁰q;
for i=k-n:+1:k-1
        β ← σᵢyᵢᵀr;
        r ← sᵢ(εᵢ-β);
end(for)
```

The step length factor is set to $\alpha_k = 1$, and the condition

$$J(m_{k+1}) < J(m_k) \tag{18}$$

is tested, where $J(m_k)$ is the misfit value obtained from the wavefield propagation in the velocity model m_k at the *kth* iteration and $J(m_{k+1})$ is the misfit value obtained with the updated velocity model given by the L-BFGS method.

When the condition given in Equation 18 does not hold, then the step length α_k decreases by half until the condition is satisfied.

Blended data implies that in the FWI methodology the number of observed data is less than the number used traditional implementation, where we acquire each shot individually. This represents the FWI method less gradient computation (Equation 16), which decreases the processing time by making fewer calculations.

3. EXPERIMENTAL DEVELOPMENT

The Ricker wavelet was used to generate both, the observed data and the modeled data at each FWI iteration. The mathematical expression used to obtain the source is

$$S = (1 - 2\pi^2 f^2 (t - t_0)^2) e^{-\pi^2 f^2 (t-t_0)^2} \tag{19}$$

where f is the central frequency, t is the time window and t_0 is a time delay.

The FWI multiscale approach is a strategy that allows to first recover the background of the velocity model and then go detailing the structures using an inversion in which a sweep of the central frequency of the wavelet is made from low frequencies to higher frequencies. For our experiments, the first central frequency is 3[Hz] [15].

On the implementation, the Finite Difference in Time Domain method (FDTD) is used to discretize the wave equation. The time and spatial discretization use a second order and eight order stencils, respectively. The CPML zone uses 20 points.

The Marmousi velocity model that was created in 1988 by the Institut Francais du Pétrole (IFP), is based on a profile through the North
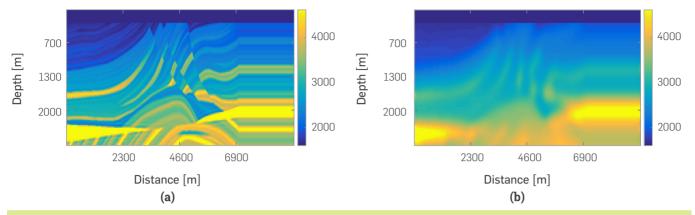
(a)　　　　　　　　　　　　　　　　　　**(b)**

Figure 6. (a) Marmousi model size 9.2km x 3km as an original velocity model. (b) Marmousi smooth model size 9.2km x 3 km as an initial velocity model.

Quenguela trough in the Cuanza Basin in Angola and it contains many reflectors, steep dips and strong velocity gradients in both vertical and lateral directions [16].

It used a Marmousi model with a distance of 9.2 km and a depth of 3km on a grid of 25 meters discretized in a grid of 368x120 points (see **Figure 6a**) and a minimum offset of 750[m] of data without overlapping; the maximum number of simultaneous sources within the super-shot is 5.

Three different sets of sources are generated: twenty coded super-shots with 5 simultaneous sources using the optimal coded implementation proposed in this paper; twenty randomly coded super-shots with 5 simultaneous sources; and 100 equally spaced traditional sources equivalent to the twenty blended sources. All of them with 328 receivers spaced evenly in the acquisition area.

The performance of the three different source distributions in the FWI is measured using the processing time and the PSNR value of the obtained velocity model against the real model using the different source distributions. PSNR allows quantifying the quality of the velocity model **Figure 6b** comparing this with the original velocity model.

Other important parameters of the experiments are shown in **Table 3**.

Table 3. Main parameters of FWI used to run the experiments.

Parameter	Value
Model Distance	368 [pt.]
Model Depth	120 [pt.]
Grid spacing	25 [m]
Time step	1 [ms]
Recording time	2 [s]
Number of sources	100
Number of receivers	828
Size of absorbing boundaries	20 [pt.]
Number of frequency steps	3, 6 and 9 Hz
Number of iterations per frequency	20

MUTUAL COHERENCE MINIMIZATION

For the case study, 20 initial random coded matrices are generated (matrix A_1 to A_{20}) where the plausible range of data for the encoding of the parameters is: $Q_T \in [0.1, 0.7]$ seconds, with a time step of 1[ms], and $Q_X \in [21, 347]$ points with a spatial step of 1 point.

It generated 20 coded matrices A of size 340x70 with 5 simultaneous inner sources. By vectorizing each encoding matrix as a column of the design matrix **B**, we obtain a design matrix **B** of size 20x23800. Being the initial U matrix used for the minimization problem Equation 8, U=$B^T B$ of size 20x20.

Figure 7a represents the initial random coded of each source, where the X-axis is the horizontal position of each source and the Y-axis the time delay. The initial code of sources have large empty areas highlighted in red and areas of sources agglomeration marked in blue, where more than one source has the same code. These problems in the initial code are represented in the matrix **U Figure 7b**, where values in the diagonal other than 5 mean that there is at least one source with the same code, super-shot 4 and 20, and non-zero values in the triangular matrices shows dependent observations between super-shots, i.e. shots with equal codes.

Starting from the random matrix **U**, the mutual coherence for the initial distribution of sources is $\mu(B)$=0.78, obtaining after 25 iterations a coherence of $\mu(B)$=0.12.

Figure 8a represents the final code of each source, where a more homogeneous distribution of sources is observed. **Figure 8b**, presents the final U matrix where the diagonal takes the expected value of 5, showing that each super-shot has 5 different simultaneous shots; the reduction of non-zero values in the triangular shows the improvement in the encoding distribution of sources.

TRADITIONAL DISTRIBUTION

For the traditional distribution of sources, the 100 sources are located equally spaced in the acquisition area (inner square area **Figure 9**, where the borders are the CPML absorbing zone of 20 points), each source (red starts in **Figure 9**) is activated and acquired individually obtaining 160 observed data. There are 328 receivers equally spaced with 160 sources placed every 75 [m].

RANDOM BLENDED DISTRIBUTION

For the random blended source distribution, 20 super-shots within 5 simultaneous sources are located in the acquisition area with 20 points of CPML absorbing zone, **Figure 10** shows an example of the distribution of 3 of the super-shots, where we can see that some sources of each super-shot are the same for other super-shot (source 1 of super-shot 1 and 2 and source 5 of super-shot 2 and source 4 of super shot 3).
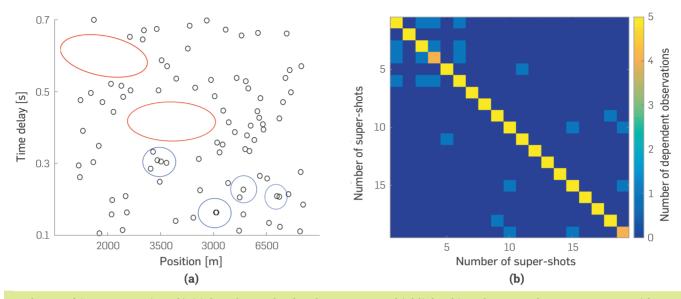
(a)

(b)

Figure 7. (a) Representation of initial random code of each source, areas highlighted in red presents large empty areas without sources and blue areas represent an agglomeration of sources. (b) initial matrix $\hat{U}$ for the minimization problem $\mu(B)$=0.78
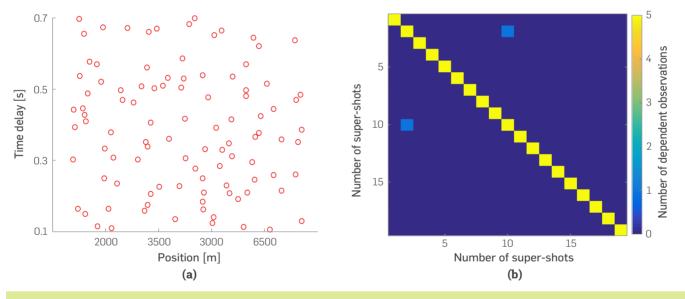

(a)

(b)

Figure 8. (a) Representation of final coded of each source, (b) final matrix U, $\mu(B)$=0.12
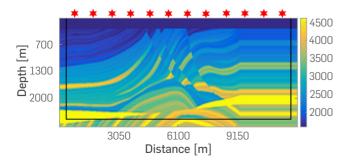

Figure 9. Traditional distribution of sources represented as red stars spaced evenly along the acquisition area (inner area of the figure), with 20 points of CPML frontier (black of the figure).

The codification is obtained using a random uniform distribution in the spatial parameter, and random Gaussian distribution for the time delay parameter. There are 328 receivers ,equally spaced, and the minimum and maximum values of separation between sources are 25 [m] and 6525[m], respectively.

OPTIMAL CODED BLENDED DISTRIBUTION

For optimal blended distribution of sources, 20 super-shots within 5 simultaneous sources are in the acquisition area with 20 points of CPML frontier producing 20 observed data sets, obtaining the codification with the methodology proposed hereunder. Figure 11 shows an example of the distribution of 3 of the super-shots, where each acquired data from different areas. There are 328 equally spaced receivers and the minimum and maximum value of separation between sources are 750 [m] and 3225[m] respectively.
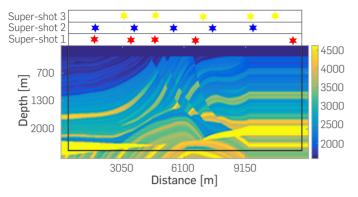
Figure 10. Representation of three Random Blended distribution of super-shots as red stars, for the first super-shot; blue starts, for the second one; and orange starts for the third one. The code for each super-shot is obtained using a random uniform distribution for the spatial parameter and Random Gaussian distribution for the time delay parameter. Along the acquisition area (inner area of the figure), with 20 points of CPML frontier (black borders of the figure).
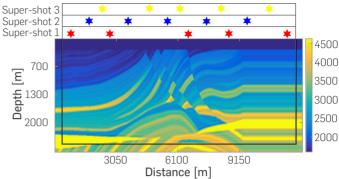

Figure 11. Representation of three Optimal Blended distribution of super-shots as red stars, for the first super-shot; blue starts, for the second one; and orange starts for the third one. The code for each super-shot is obtained using the methodology proposed in this document, along the acquisition area (inner area of the figure), with 20 points of CPML frontier (black borders of the figure).

4. RESULTS ANALYSIS

To compare the FWI algorithm for traditional, random blended distribution, and optimal blended distribution proposed in the document, three frequencies are used during the inversion 3, 6 and 9 [Hz], the number of iterations per frequency is set to 20, and the propagation time is set to 2 sec. The initial velocity model is depicted in **Figure 5b**, which is a smoothed version of the original Canadian model **Figure 5a**.

For each of the three experiments explained in the last section, we started with the smooth Marmousi velocity model (**Figure 5b**) as the initial velocity model in the first step of frequency for the multiscale FWI. **Figure 12** shows the results of 20 iterations using the three different distributions of sources. **Figure 12a**, traditional distribution of sources, **Figure 13b**, random blended distribution of sources and **Figure 12c** optimal blended distribution proposed in the document.

The optimal blended distribution shows a better delimitation of the first velocity zone of 4000 [m/s] than the random blended distribution (yellow zones in **Figure 12b** and **Figure 9a**). Being the PSNR value at least 2[dB] upper in the optimal distribution than the random distribution.

For the next frequency step (6 Hz), the initial velocity models for each experiment corresponds to the final velocity model of the previous frequency step shown in **Figure 12**, using the three different distributions of sources. **Figure 13a**, traditional distribution of sources, **Figure 13b**, random blended distribution of sources and **Figure 13c** optimal blended distribution proposed in the document. Final velocity models of the traditional and optimal distribution of sources look similar, being the PSNR value a metric to determine how close to our real model each one is, as presented in the second line of **Table 4**. It may be observed that final velocity models with the optimal blended acquisition, **Figure 14c**, presents artifacts near the borders. The random blended distribution of sources generates less defined structures backed up with the lower PSNR value.
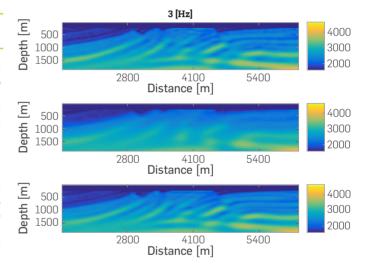

Figure 12. Final velocity models for the first stage (central frequency 3 Hz) in the multiscale FWI, using the smooth Canadian model, Figure (5b) as initial velocity model, and the three sets of observed data. Figure (a) Velocity model with a traditional distribution of sources. (b) Velocity model with a randomly blended distribution of sources and (c) Velocity model with an optimal blended distribution of sources.

For the last step of frequency (9 Hz), the initial velocity models for each experiment corresponds to the final velocity model of the previous frequency step shown in **Figure 13**, using the three different source distributions. **Figure 14a**, traditional source distribution, **Figure 14b**, random blended distribution of sources and **Figure 14c** optimal blended distribution proposed in the document.

The third line of **Table 4** shows the PSNR value for the last frequency step, and the traditional distribution is the model with the highest PSNR value as each source is acquired individually without interference of the others. However, this type of acquisition requires the ship passing more times through the exploration area and,
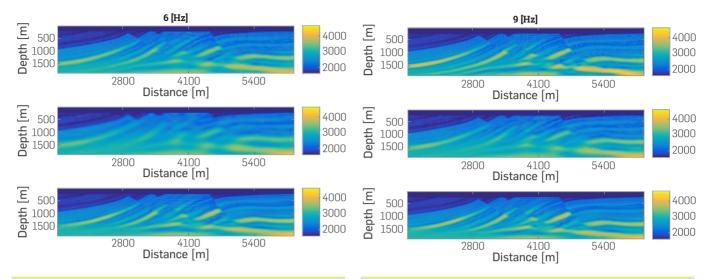
6 [Hz]

9 [Hz]

Figure 13. Final velocity models for the second stage (central frequency of 6 Hz) in the multiscale FWI, using for each experiment the final velocity model of the previous frequency step shown in Figure 9 as initial velocity model, and the three sets of observed data. Figure (a) Velocity model with a traditional distribution of sources. (b) Velocity model with a randomly blended distribution of sources, and (c) Velocity model with an optimal blended distribution of sources.

Figure 14. Final velocity models for the second step of frequency, 9 Hz, in the multiscale FWI, using for each experiment the final velocity model of the previous frequency step shown in Figure 10 as initial velocity model, and the three sets of observed data. Figure (a) Velocity model with a traditional source distribution. (b) Velocity model with a randomly blended source distribution, and (c) Velocity model with an optimal blended source distribution.

therefore, being more expensive than the proposed methodology that acquires more of one source simultaneously. The proposed codification reduces the processing time in the computation of FWI 5.23 times with traditional FWI, obtaining a final velocity model with a PSNR value higher than the random distribution.

The optimal blended distribution obtains in the first frequency step (3[Hz]) a PSNR value closer than the PSRN of random blended distribution in the second step (6[Hz]). This behavior continues in the next pair of frequency, 6[Hz] for optimal blended distribution and 9[Hz] for random blended distribution.
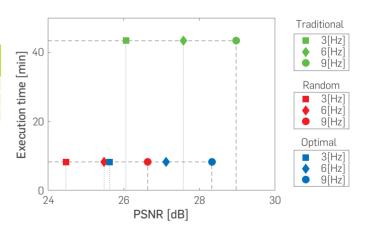
Table 4. PSNR for the multiscale FWI approach using a different distribution of sources

Freq. [Hz]	Traditional PSNR	Random Code PSNR	Optimal code PSNR
3	26.06[dB]	24.47[dB]	25.63[dB]
6	27.58[dB]	25.48[dB]	27.12[dB]
9	28.97[dB]	26.63[dB]	28.33[dB]

Figure 15 shows the evolution of each experiment comparing the execution time versus the PSNR value of each distribution of sources and each frequency step. Being traditional distribution (green marks), random blended distribution (red marks) and optimal blended distribution (blue marks). We use three different figures to show each frequency step, squares, diamonds and circles for 3[Hz], 6[Hz] and 9[Hz] respectively.

The execution time for each frequency step, in the three experiments, is approximately the same, being 8.32 [min] for both blended FWI (random and optimal) and 43.40[min] for the traditional case.

The progression can be observed in a better velocity model (with higher PSNR value) of each frequency step, it being the best velocity model obtained by traditional source distribution (green marks), followed closely by Optimal Blended distribution (blue marks), where the difference between traditional PSNR and optimal PSNR value for each frequency decreases in every frequency step, with a final difference of 0.6 [dB].

Figure 15. Execution time per frequency step versus PSNR of each velocity models, with squares being the first frequency step (3[Hz], diamonds the second frequency step (6[Hz]) and circles the last frequency step (9[Hz]). Each color represents one of the three experiments, traditional distribution of sources, (green), random blended distribution of sources (red), and optimal blended distribution of sources (blue).

CONCLUSIONS

In this work, a methodology for optimal coded blended sources has been presented with the aim of improving the results of 2D blended FWI in contrast to the randomly coded blended sources and the traditional equally spaced sources.

The minimization problem developed can be seen to diagonalize an initial matrix U that contains non-zero values outside the diagonal in a diagonal matrix, where the numerical value of the elements of the diagonal represents the number of simultaneous shots with different codification within each super-shot and the absence of values outside the diagonal represents a zero correlation between super-shots, that is, a distribution of sources in which the encoding parameters are different in each super-shot, and in turn they are different with respect to the rest of the shots.

The methodology improves the final velocity model, which is closer to the model obtained with the traditional acquisition, with a difference of about 0.6[dB], in contrast to the 2[dB] obtained by the randomly blended sources. Furthermore, the processing time is the same for the two blended FWI with an acceleration of near 5.23 times in computation time if both are compared with the traditional FWI. Optimal blended distribution obtains a balance between less processing time and quality of the image with a higher PSNR value than random blended distribution, observed in **Figure 15**, where the difference between the PSNR value of traditional and optimal blended distribution is lower in each frequency step.

Blended data implies that in the FWI methodology the number of observed data is less than that used in the traditional implementation, where we acquire each shot individually. This represents in the FWI method less gradient computation, which reduces the processing time by making fewer calculations.

Traditional acquisition produces a better velocity model after the FWI process, but it requires one image for each source increasing the acquisition time, the number of times that the vessel goes through the exploration trajectory, increasing the acquisition cost. The processing time is higher than blended FWI due to it is need of calculating more velocity gradients.

FUTURE WORK

It is necessary to analyze the effect of blended acquisition interference with other measures to quantify the difference of data sets in a real scenario. PSNR is only plausible in a synthetic case, where we know the original velocity model. In a real case cycle skipping of traces could be used as quality control, which measures the similarity as a function of the displacement of one trace relative to the other.

REFERENCES

[1] Virieux, J., and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6), WCC1-WCC26.

[2] Russell, B. (1988). *Introduction to seismic inversion methods.* Society of Exploration Geophysicists Tulsa.

[3] Beasley, C., Chambers, R., and Jiang, Z. (1998). A new look at simultaneous sources. *SEG Technical Program Expanded Abstracts*, 17(1), 133-135.

[4] International Association of Oil and Gas Producers (OGP). 2011. An overview of marine seismic operations., Report No. 448, April.

[5] Anagaw, A., and Sacchi, M. (2012). Full waveform inversion using blended acquisition geometry with different frequency strategies. *74th EAGE Conference and Exhibition incorporating EUROPEC*, Copenhagen, Denmark.

[6] Romero, L., Ghiglia, D., Ober, C., and Morton, S. (1999). Phase encoding of shot records in prestack migration. *Geophysics*, 65(2), 426-436.

[7] Haber, E., Van Den Doel, K., and Horesh, L. (2015). Optimal design of simultaneous source encoding. *Inverse Problems in Science and Engineering*, 23(5), 780-797.

[8] Shearer, P. (2009). Introduction to seismology. Cambridge University Press.

[9] Bruckstein, A., Donoho, D., and Elad, M. (2009).From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1), 34-81.

[10] Obermeier, R., and Martinez-Lorenzo, J. (2017) Sensing Matrix Design via Mutual Coherence Minimization for Electromagnetic Compressive Imaging Applications. *IEEE Transactions on Computational Imaging*, 3(2), 217229.

[11] Nocedal, J., and Wright, S. (2006) Numerical optimization. *Springer science & Business Media*,

[12] Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8), 1259-1266.

[13] Plessix, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2), 495-503.

[14] Liu, D., and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1-3), 503-528.

[15] Mao, J., Wu, R.-S., and Wang, B. (2012). Multiscale full waveform inversion using GPU. *SEG Technical Program Expanded Abstracts 2012*, 1-7.

[16] Versteeg, R. (1994). The marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge*, 13(9), 927-936.