

Article

Constrained Full Waveform Inversion for Borehole Multicomponent Seismic Data

Marwan Charara ^{1,*}  and Christophe Barnes ² ¹ Skolkovo Institute of Science and Technology, Moscow 121205, Russia² Department of Geosciences and Environment, University of Cergy-Pontoise, 95000 Neuville-sur-Oise, France; christophe.barnes@u-cergy.fr

* Correspondence: m.charara@skoltech.ru; Tel.: +7-495-280-1481

Received: 2 December 2018; Accepted: 25 December 2018; Published: 16 January 2019



Abstract: Full-waveform inversion for borehole seismic data is an ill-posed problem and constraining the problem is crucial. Constraints can be imposed on the data and model space through covariance matrices. Usually, they are set to a diagonal matrix. For the data space, signal polarization information can be used to evaluate the data uncertainties. The inversion forces the synthetic data to fit the polarization of observed data. A synthetic inversion for a 2D-2C data estimating a 1D elastic model shows a clear improvement, especially at the level of the receivers. For the model space, horizontal and vertical spatial correlations using a Laplace distribution can be used to fill the model space covariance matrix. This approach reduces the degree of freedom of the inverse problem, which can be quantitatively evaluated. Strong horizontal spatial correlation distances favor a tabular geological model whenever it does not contradict the data. The relaxation of the spatial correlation distances from large to small during the iterative inversion process allows the recovery of geological objects of the same size, which regularizes the inverse problem. Synthetic constrained and unconstrained inversions for 2D-2C crosswell data show the clear improvement of the inversion results when constraints are used.

Keywords: constrained inversion; elastic full waveform inversion; multicomponent seismic data; borehole seismic

1. Introduction

Full-waveform inversion of seismic data allows one to obtain an image of the subsurface through the determination of a certain number of physical parameters. Borehole data (VSP and crosswell) are of special interest because of the geometry of acquisition, which provides more informative signals on the medium properties than surface seismic data (high frequency signal, energetic P-S conversions). However, the lack of seismic data redundancy (e.g., few shots) renders the inversion problem underdetermined. In general, full-waveform inversion is an ill-posed problem, in the sense that an infinite number of models match the data [1]. It is classically solved with local optimization schemes and is therefore strongly dependent on the starting model definition. This starting model should predict arrival times with errors less than half of the period to cancel the cycle-skipping ambiguity [1]. The multiscale strategy performed by moving from low to high frequencies during the inversion allows reduction of the nonlinearities and cycle skipping issues of the inversion and helps convergence toward the global minimum. Regularizations are conventionally applied to the inversion in the model space to make it better posed [2,3]. Tikhonov and Arsenin [4] have proposed a regularization strategy within the optimization step to find the smoothest model that explains the data. Preconditioning techniques acting as a smooth operator on the model update [5] may add strong prior features of the expected structure through directive Laplace preconditioning, as in Guitton et al. [6].

Regularization schemes that preserve edges and contrasts have also been developed for specific full waveform inversion applications through a ℓ_1 model penalty [7] or through a multiplicative regularization [8]. Regularization can also be expressed in the curvelet or wavelet domains [9,10]. In such domains, the ℓ_1 norm minimization is generally preferred for the model term penalty because it ensures sparsity in the model space. All the previous regularization techniques allow stabilization of the inversion scheme by assuming a particular representation or structure of the velocity model (smoothness, sparsity, and so on). However, geological information, taking into account prior model information, is generally not used in classical full waveform inversion implementation; however, there are some examples [11,12]. In addition, most of the weighting operators or covariance matrices associated with the model parameters are set to identity or in the best cases to a diagonal matrix. The same applies to the weighting operators or covariance matrices associated with the data.

Weighting operators on the data and models or inverse of the data and model covariance operators for least squares in the frame of the Bayesian formulation [3] are introduced in the misfit function. In this paper, we will show the benefit of introducing covariance matrices in the data and model space. On the one hand, we will illustrate the impact on constraining the seismic inverse problem on the data space for the case of a synthetic two-component borehole seismic data by using the polarization analysis to fill a block diagonal data space covariance matrix. On the other hand, we will illustrate the benefit of constraining the seismic inverse problem on the model space by performing a crosswell synthetic experiment. The horizontal and vertical spatial correlations using Laplace distribution are used to fill the model space covariance matrix in order to introduce an a priori solution, favoring a tabular medium whenever it doesn't contradict the information contained in the data.

2. Constrained Least Squares Inversion

Using the probabilistic formalism developed by Tarantola [3] and considering only the data space term in the equation, we rewrite the misfit function in the vicinity of the current model \mathbf{m}_n :

$$S(\mathbf{m}_n + \delta\mathbf{m}) = (\delta\mathbf{d}_n - \mathbf{G}_n\delta\mathbf{m})^T \mathbf{C}_D^{-1} (\delta\mathbf{d}_n - \mathbf{G}_n\delta\mathbf{m}) + (\Delta\mathbf{m}_n - \delta\mathbf{m})^T \mathbf{C}_M^{-1} (\Delta\mathbf{m}_n - \delta\mathbf{m}), \quad (1)$$

where:

- $\delta\mathbf{m}$ is the perturbation in the vicinity of the current model \mathbf{m}_n ,
- $\delta\mathbf{d}_n$ are the data residuals for the model \mathbf{m}_n ,
- $\Delta\mathbf{m}_n$ is the difference between the current model \mathbf{m}_n and the a priori model \mathbf{m}_{prior}
- \mathbf{G}_n is the linear function tangent to \mathbf{g} at the model \mathbf{m}_n ,
- \mathbf{g} is the function mapping the model space $\mathbf{m} \in M$ into the data space $\mathbf{d} = \mathbf{g}(\mathbf{m}) \in D$
- \mathbf{C}_D is the covariance matrix on the data space.
- \mathbf{C}_M is the covariance matrix on the model space (defining the Gaussian a priori probability density).

Applied to the waveform inversion problem of estimating the elastic parameters and the density of the earth, the minimization of the misfit function can be solved by iterative gradient methods.

The iterative process of a nonlinear inversion can be summarized as follows [3]: for a given iteration k , from the current model \mathbf{m}_k (i.e., the discretized physical fields characterizing the medium), we perform simulations of the wave in order to obtain the synthetic data $\mathbf{d}_k^{cal} = \mathbf{g}(\mathbf{m}_k)$. Thus, comparing them to observed data \mathbf{d}^{obs} , we obtain the residuals $\delta\mathbf{d}_k = \mathbf{d}^{obs} - \mathbf{d}_k^{cal}$. These residuals are weighted $\hat{\delta}\mathbf{d}_k = \mathbf{C}_D^{-1}\delta\mathbf{d}_k$ (the hat denotes the dual space) and back-propagated using the wave propagation equation again to obtain the gradient in the dual model space $\hat{\gamma}_k = \mathbf{G}_k^T \mathbf{C}_D^{-1} \hat{\delta}\mathbf{d}_k$ (where \mathbf{G}_k is the derivative of the function \mathbf{g} over the model space at the point \mathbf{m}_k and where T denotes the transpose operation) and consequently, the gradient in the model space $\gamma_k = \mathbf{C}_M \hat{\gamma}_k$. The gradient indicates the steepest ascent direction of the misfit function in the model space. We modify this direction using the conjugate gradient algorithm $\varphi_k = \gamma_k + \alpha_k \varphi_{k-1}$, where φ_k is the conjugate

gradient and α_k is defined from the gradient [13]. Finally, using an additional simulation, we optimize the step length μ_k [14] in the upgrading equation $\mathbf{m}_{k+1} = \mathbf{m}_k - \mu_k \boldsymbol{\varphi}_k$ and we update the model.

3. Constraints on the Data Space

The quantification of uncertainties in seismic data is often neglected, because in many cases it is a difficult task. For instance, data errors generated by receiver response are correlated in time, which means accounting for them is not straightforward. To illustrate the possible contribution of constraints provided by the analysis of uncertainties of data, we have performed two inversions from the synthetic data experiment. The first inversion is done without any data constraints, whereas in the second one we have incorporated polarization wave analysis in the covariance matrix \mathbf{C}_D . This matrix treats the uncertainties on particle velocities according to the eigenbasis of the particle motion.

For the case of multi-component data, the uncertainty of a component may be correlated to the uncertainty of the other components. For each time for each of the receiver components, we can define a 2×2 matrix for 2-component data (as in the example bellow), or a 3×3 matrix for 3-component data. In that case, the covariance matrix on the data space \mathbf{C}_D is not a diagonal matrix. In order to take into account polarization, we consider that the uncertainties are proportional to the local cross-correlation matrix of the signal (the data) and that the signal is locally stationary (over 2 periods). The cross-correlation matrix between components i and j of the signal $s_{r,t}$ is computed with zero lag ($\tau = 0$) for any receiver r and any time t :

$$C_{r,t}^{ij}(\tau) = \int_{-\infty}^{+\infty} s_{r,t}^i(u) s_{r,t}^j(u + \tau) du, \tag{2}$$

where:

$$s_{r,t}^i(u) = d_r^i(u) w_t(u)$$

with $d_r^i(u)$ being the trace for the component i of the receiver r of the observed data and $w_t(u)$ a Hamming time window centered on time t with a length of typically 2 periods. One can use other time tapering windows classically used in spectral analysis (Blackman, Nuttall, etc).

3.1. Cross-Correlations and Polarization

The polarization of the multicomponent signal is estimated from this matrix $C_{r,t}^{ij}(\tau = 0)$ and the matrix of the uncertainties has the same eigenvectors (polarization directions) than the cross-correlation matrix. The uncertainty has the same shape as the polarization of the signal. The ratios among the eigenvalues are the same, in other words, the rectilinearity and planarity of the polarization [15] is conserved.

The polarized signal for rectilinear polarization can be modeled by:

$$\mathbf{d}(t) = \mathbf{a} \cdot s(t) + \mathbf{n}(t), \tag{3}$$

where $\mathbf{d}(t)$ is the 3C signal, \mathbf{a} is the polarization vector, $s(t)$ the scalar signal along the polarization vector, and $\mathbf{n}(t)$ the 3C Gaussian noise.

In the frequency domain, the equation can be expressed as:

$$\hat{\mathbf{d}}(f) = \mathbf{a} \cdot \hat{s}(f) + \hat{\mathbf{n}}(f), \tag{4}$$

where the hat denotes the frequency domain.

Let us consider the spectral matrix of the 3C signal defined by:

$$\begin{aligned}
 \hat{\mathbf{S}}(f) &= \overline{\hat{\mathbf{d}}(f)} \cdot \hat{\mathbf{d}}^T(f) \\
 &= \overline{(\mathbf{a}\hat{s} + \hat{\mathbf{n}})} \cdot (\mathbf{a}\hat{s} + \hat{\mathbf{n}})^T \\
 &= \mathbf{a} \cdot \mathbf{a}^T \cdot \overline{\hat{s}}\hat{s} + \overline{\hat{\mathbf{n}}} \cdot \hat{\mathbf{n}}^T \\
 &= \mathbf{a} \cdot \mathbf{a}^T |\hat{s}(f)|^2 + \hat{\mathbf{N}}(f),
 \end{aligned}
 \tag{5}$$

where the upper bar denotes the complex conjugate, T denotes the transpose operator and $\hat{\mathbf{N}}(f)$ being the spectral matrix of the noise. Note that $\overline{(\mathbf{a}\hat{s})} \cdot \hat{\mathbf{n}}^T = \overline{\hat{\mathbf{n}}} \cdot \mathbf{a}^T \hat{s} = 0$ because noise and signal are considered uncorrelated. The polarization vector is an eigenvector of the spectral matrix (for isotropic noise) for any frequency f :

$$\begin{aligned}
 \hat{\mathbf{S}} \cdot \mathbf{a} &= (\mathbf{a} \cdot \mathbf{a}^T |\hat{s}|^2 + \hat{\mathbf{N}}) \mathbf{a} \\
 &= |\hat{s}|^2 \mathbf{a} \cdot \mathbf{a}^T \mathbf{a} + \hat{\mathbf{N}} \mathbf{a} \\
 &= |\hat{s}|^2 \mathbf{a} + (|\hat{\mathbf{n}}|^2 \mathbf{I}) \mathbf{a} \\
 &= (|\hat{s}|^2 + |\hat{\mathbf{n}}|^2) \mathbf{a}.
 \end{aligned}
 \tag{6}$$

When integrating the spectral matrix over \mathbb{R} , one obtains the cross-correlation matrix $\mathbf{C}(\tau = 0)$ with zero lag with the same eigenvectors.

The structure of the covariance matrix on the data space \mathbf{C}_D is derived from the previous equations. For instance, for a 2C receiver, a given source or receiver couple, and at any time t , the corresponding 2×2 submatrix is defined as:

$$\mathbf{C}_D = \mathbf{V} \cdot \sigma_D^2 \begin{pmatrix} 1 & 0 \\ 0 & \frac{\lambda_{II}}{\lambda_I} \end{pmatrix} \cdot \mathbf{V}^{-1}$$

where \mathbf{V} is the eigenvectors matrix (column), σ_D is the standard deviation of the (isotropic) noise for this couple, and λ_I and λ_{II} are the max and min of the two positive eigenvalues of the spectral matrix, respectively.

Let us consider the sub-matrix for the first data 2C-trace of the covariance matrix on the data space:

$$\mathbf{C}_D(rec = 1) \propto \mathbf{C}_D(r = 1, t, \tau = 0)$$

$$\begin{bmatrix}
 \sigma_{1,X}^2 & \rho_{1,X} \sigma_{1,X} \sigma_{1,Z} & 0 & 0 & \dots & \dots & 0 & 0 \\
 \rho_{1,X} \sigma_{1,X} \sigma_{1,Z} & \sigma_{1,Z}^2 & 0 & 0 & \dots & \dots & 0 & 0 \\
 0 & 0 & \sigma_{2,X}^2 & \rho_{2,X} \sigma_{2,X} \sigma_{2,Z} & \dots & \dots & 0 & 0 \\
 0 & 0 & \rho_{2,X} \sigma_{2,X} \sigma_{2,Z} & \sigma_{2,Z}^2 & \dots & \dots & 0 & 0 \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\
 0 & 0 & 0 & 0 & \dots & \dots & \sigma_{nt,Z}^2 & \rho_{nt,X} \sigma_{nt,X} \sigma_{nt,Z} \\
 0 & 0 & 0 & 0 & \dots & \dots & \rho_{nt,X} \sigma_{nt,X} \sigma_{nt,Z} & \sigma_{nt,Z}^2
 \end{bmatrix}$$

where ρ_i , $\sigma_{i,X}$, and $\sigma_{i,Z}$ are, respectively, the correlation coefficient, the standard deviation for the X component, and the standard deviation for the Z component for a given time sample i , and nt being the total number of time samples. The matrix \mathbf{C}_D is a block diagonal matrix and its inverse (i.e., \mathbf{C}_D^{-1} , is another block diagonal matrix), composed of the inverse of each block.

3.2. An Offset VSP Synthetic Example

In this numerical example, the acquisition geometry is an offset VSP with a unique source at 500 m offset and the antenna is between 580 and 970 m depth in a vertical well (see Figure 1 for more details). The isotropic elastic model depends only on the depth (1D model). The “true” model is defined in Figure 1. The horizontal and vertical components of the “observed” data obtained from the true model

are displayed in Figure 1. The direct problem (seismic modeling) is based on discretizing the wave equation by the finite differences method (FDM) [16].

The inverted parameters are the P-wave velocity, the S-wave velocity, and the density. The starting models are simple, the water layer is considered as known, and the other layers are replaced by a constant vertical gradient for each parameter (gray solid lines in Figure 2). The test consists of comparing the results of inversion when data uncertainties are isotropic and when the polarization is included in the data uncertainties.

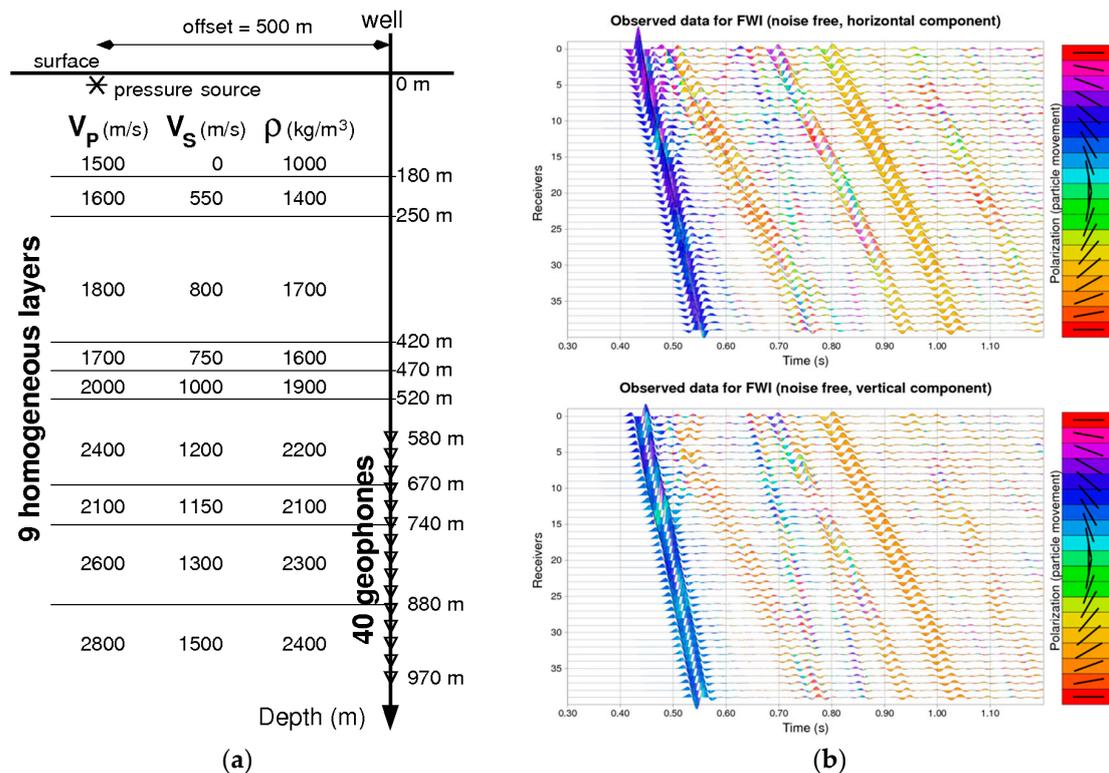


Figure 1. Offset Vertical Seismic Profile (OVSP) numerical experiment: (a) Synthetic experiment configuration where the asterisk near the surface denotes the source position for the unique shot point, whereas the triangles in the well denote the location of the 40 two-component geophones. The “true” model is composed of nine homogeneous layers, the first one being the water layer. This true model is used both to model the observed data and as a reference to check inversion results; (b) the horizontal (top) and vertical (bottom) components of the ground velocity field recorded at geophones—the fields are obtained by modeling the wave equation in the true model (finite difference approximation). The color represents the local polarization estimated using a 2-period time window.

As shown in Figure 2, the inversion results exhibits in both cases a good fit in the receiver zone, while the estimation is not good in the upper part of the model. One can clearly note that the S-wave velocity model is recovered with more detail than the P-wave velocity model. This is due to the fact that converted S-waves are present in borehole seismic data and also because the S-waves have a shorter wavelength than P-waves, hence, a better spatial resolution. The final residuals for both inversion experiments are small; a few percent of the misfit. The inversion results when polarization is included are significantly better than for the isotropic uncertainties. More precisely, for the case of polarization, both P-wave and S-wave estimated velocity fields are well fitted in the antenna zone.

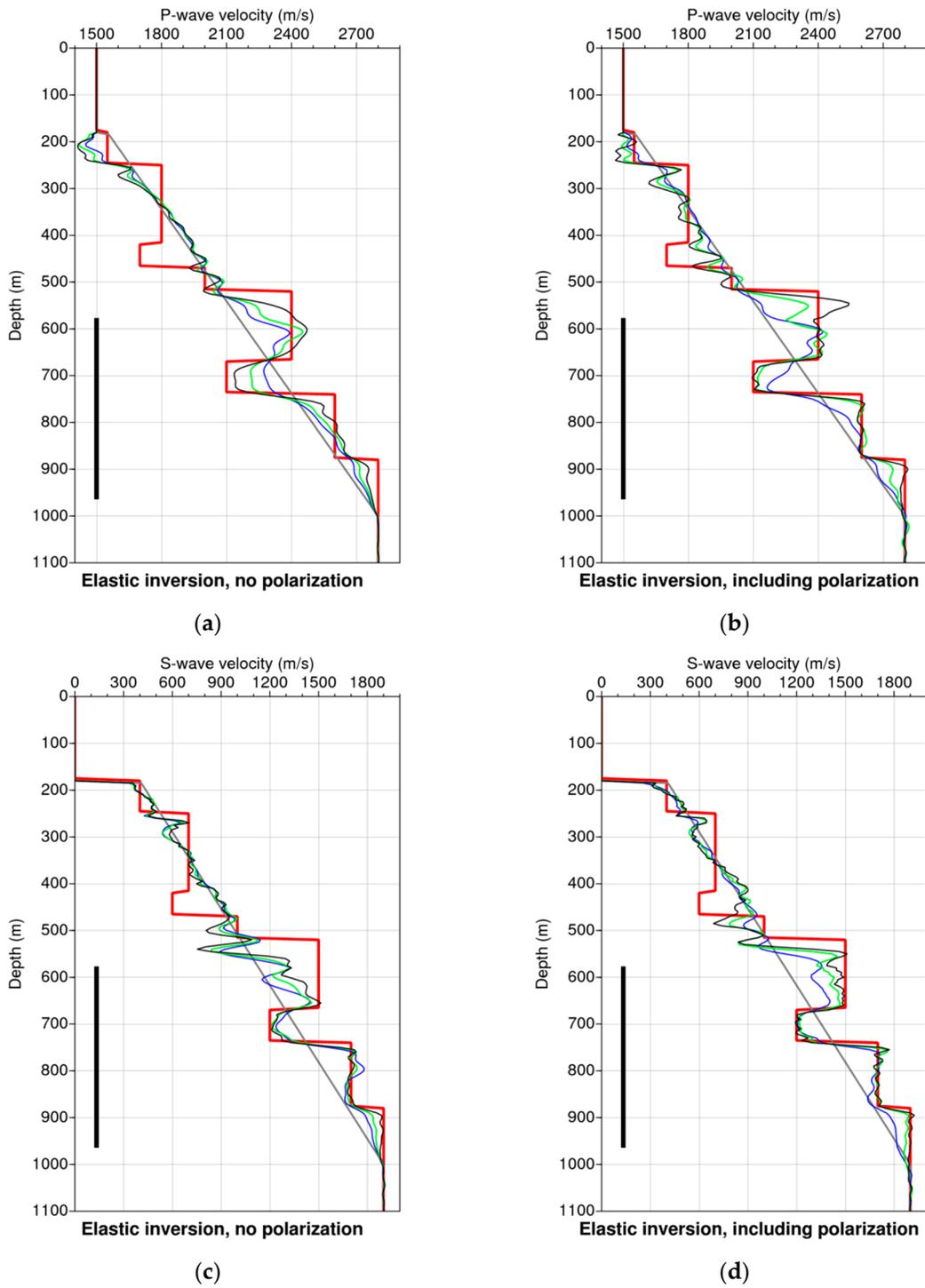


Figure 2. Two inversion experiments with and without polarization constraints. (a) Without polarization constraints for P-wave velocity estimated model. (b) With polarization constraint for P-wave velocity estimated model. (c) Without polarization constraints for S-wave velocity estimated model. (d) With polarization constraints for S-wave velocity estimated model. In all plots: red thick solid line denotes the true model we want to retrieve by inverting the observed data; gray solid line denotes the starting model in the inversion process; blue, green, and black solid lines denote estimated models from inversion, respectively, at iterations 10th, 30th, and 80th. The vertical thick black solid line indicates the level of the receiver antenna. The constrained inversion contributes to the recovery of the sharp interfaces, especially at the level of the receiver antenna.

4. Constraints on the Model Space

We can distinguish two kinds of geological a priori information that can be incorporated in an inversion process. On the one hand, information obtained by measurements (well logging), that we consider as “objective” information, allows the definition of an a priori model at the well vicinity. On the other hand, information coming from geological interpretations, such as geological layer dips, that we qualify as “subjective” information, can be simplified and introduced into the covariance matrix \mathbf{C}_M .

4.1. Quantification of Number of Degrees of Freedom

In order to evaluate the importance of constraints on an inversion, it is necessary to quantify the number of degrees of freedom. This number can be estimated from the set of parameters as a function of existing correlations among them and independently of their standard deviations. The number N of free parameters is equal to the number n of parameters in the absence of correlations, and it is reduced to 1 in the case of perfect correlations.

4.1.1. General Case

Let us consider a 1D Gaussian random field \mathbf{X} , the domain is discrete and finite consisting of a set of n points; this field is equivalent to the set of random variables (X_1, X_2, \dots, X_n) . It can be characterized by the following probability density function (or p.d.f.):

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \det(\mathbf{C})^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1}(\mathbf{x} - \mathbf{m})\right), \quad (7)$$

where \mathbf{x} is a realization, \mathbf{m} is the expectation of the field ($E(\mathbf{X}) = \mathbf{m}$), and \mathbf{C} is the covariance matrix of the field.

The covariance matrix can be rewritten with the following form = \mathbf{SRS} , where \mathbf{R} is the correlation matrix and \mathbf{S} the diagonal matrix of the standard deviations.

If there is no correlation among the random variables X_i , then $\mathbf{R} = \mathbf{I}$ and the covariance matrix is diagonal and the number of free parameters is then n . If a correlation exists, we then use the LU decomposition method to estimate the number of free parameters (see Appendices A and B for more details). We can write $\mathbf{R} = \mathbf{MM}^T$ where \mathbf{M} is the left lower triangle matrix of the LU decomposition of the matrix \mathbf{R} . The number of free parameters is then reduced to:

$$N = \text{trace}(\mathbf{M}). \quad (8)$$

4.1.2. Spatial Correlation on the Model Space

For illustration purposes, let us consider the special case where the spatial correlations are described by an exponential model. Thus, for any pair of points M_1 and M_2 (belonging to the same horizontal line), the correlation ρ between the pair of random variables X_1 and X_2 corresponding to the points M_i can be expressed as a function of the distance:

$$\rho(X_1, X_2) = \exp\left(-\frac{d(M_1, M_2)}{r}\right), \quad (9)$$

where $d(\cdot)$ is a distance and r is the range of the correlation.

By setting the variance for each variable to be constant $\sigma_i = \sigma_1$, the p.d.f. of Equation (7) reduces to,

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \cdot \sigma^n \sqrt{\det(\mathbf{R})}} \exp\left(-\frac{1}{2} \left(\frac{\mathbf{x} - \mathbf{m}}{\sigma}\right)^T \mathbf{R}^{-1} \left(\frac{\mathbf{x} - \mathbf{m}}{\sigma}\right)\right), \quad (10)$$

This simplifies, as we will show below, the study of the relation between the matrix **R** and the degrees of freedom using an exponential correlation function.

Let us consider the same Gaussian field **X**, but this time defined on a defined 1D domain regularly sampled (as for a grid) containing *n* points. The distances between points are multiples of the step Δx and the correlation matrix has the following form (for ordered points):

$$\mathbf{R} = \begin{pmatrix} 1 & a & a^2 & a^3 & a^4 & \dots & a^n \\ a & 1 & a & a^2 & a^3 & \dots & a^{n-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a^n & a^{n-1} & a^{n-2} & a^{n-3} & a^{n-4} & \dots & 1 \end{pmatrix}, \tag{11}$$

where

$$a = \exp\left(-\frac{\Delta x}{r}\right) \leq 1. \tag{12}$$

One interesting property of the **R** matrix is that the **M** matrix of the LU decomposition has a simple form

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ a & b & 0 & 0 & 0 & \dots & 0 \\ a^2 & ab & b & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a^n & a^{n-1}b & a^{n-2}b & a^{n-3}b & a^{n-4}b & \dots & b \end{pmatrix}, \tag{13}$$

with $b = \sqrt{1 - a^2}$. This means that, in terms of a sequential simulation, the $n - 1$ variables, X_2 to X_n , have the same degree of freedom: *b*. The total degrees of freedom *N* of the field **X** is then:

$$N = 1 + (n - 1) \cdot b, \tag{14}$$

where for *n* parameters (or *n* grid points), the number of free parameters, considered as the degrees of freedom, of the inversion is *N*. Figure 3 shows the relation between the range of the correlation (expressed in number of steps, Δx) and the values of *b*.

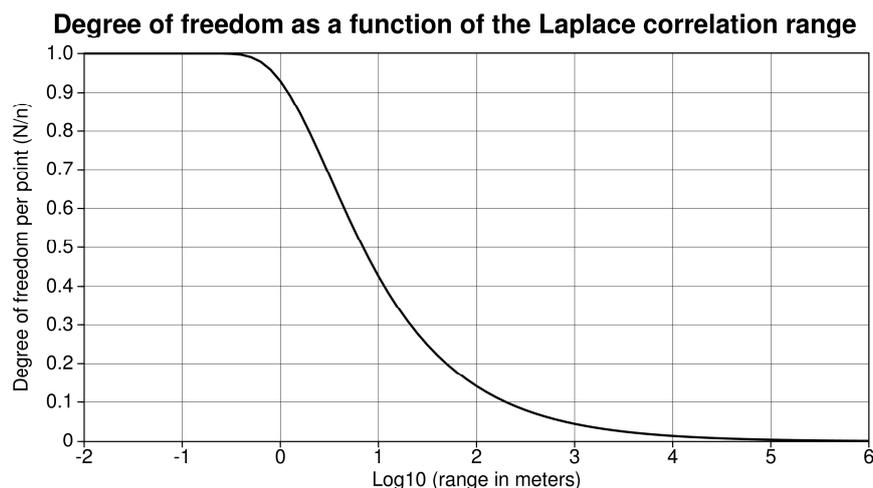


Figure 3. Degree of freedom per point (for a regular grid) as a function of the range (expressed in multiples of the grid step) of the correlation for a Laplace correlation function.

4.2. Seismic Crosswell Numerical Experiment Example

This numerical example is an illustration how to incorporate the “subjective” geological information that our medium is mainly tabular but might contain complex geological structures using the covariance matrix on the model space.

4.2.1. Description of the Crosswell Experiment and Observed Data

The reference subsurface models are shown in Figure 4. The word “reference” indicates that this model is used to generate the synthetic seismic data used as observed data (see Figure 5) during the inversion process. It is also the model to be retrieved by applying the full-wave inversion method to the observed data. The medium for wave propagation is elastic isotropic. The parameters to be inverted are discretized physical fields: vertical P-wave and S-wave velocities. For simplicity, the density (ρ) is not inverted—the field is constant with $\rho = 2500 \text{ kg}\cdot\text{m}^{-3}$ and it fixed at the correct value in the different inversions.

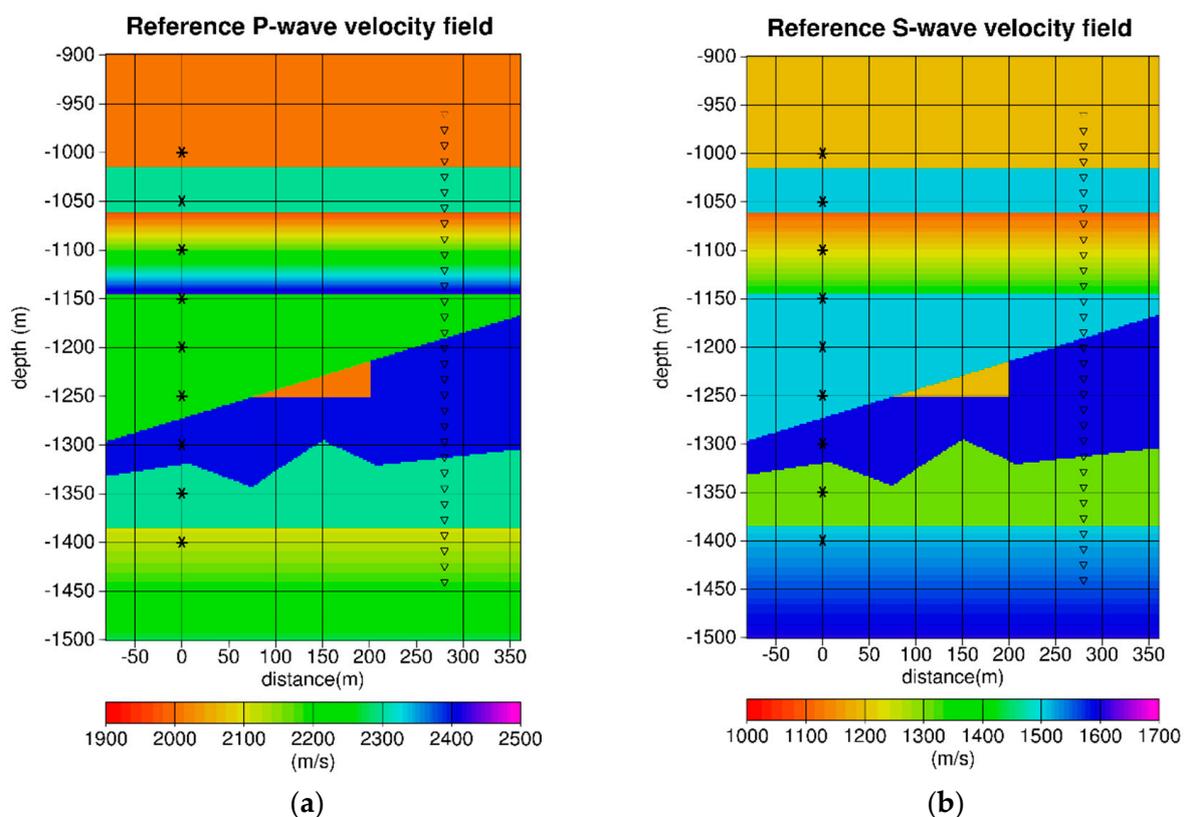


Figure 4. The reference velocity models: (a) Reference P-wave velocity field; (b) reference S-wave velocity field. The source borehole is located at 0 m distance and the nine asterisks (*) denote the source locations for the nine shots. The receiver borehole is located at 280 m offset and the small triangles denote the locations of the receivers. One can notice that the structures of these reference velocity fields are the same for P- and S-wave. It is composed of 1D regions (invariance along the horizontal axis) in the upper part of the model and at the bottom with homogeneous layers and other layers with a velocity increasing linearly with depth (vertical constant gradient). The middle part of the model exhibits complex 2D structures.

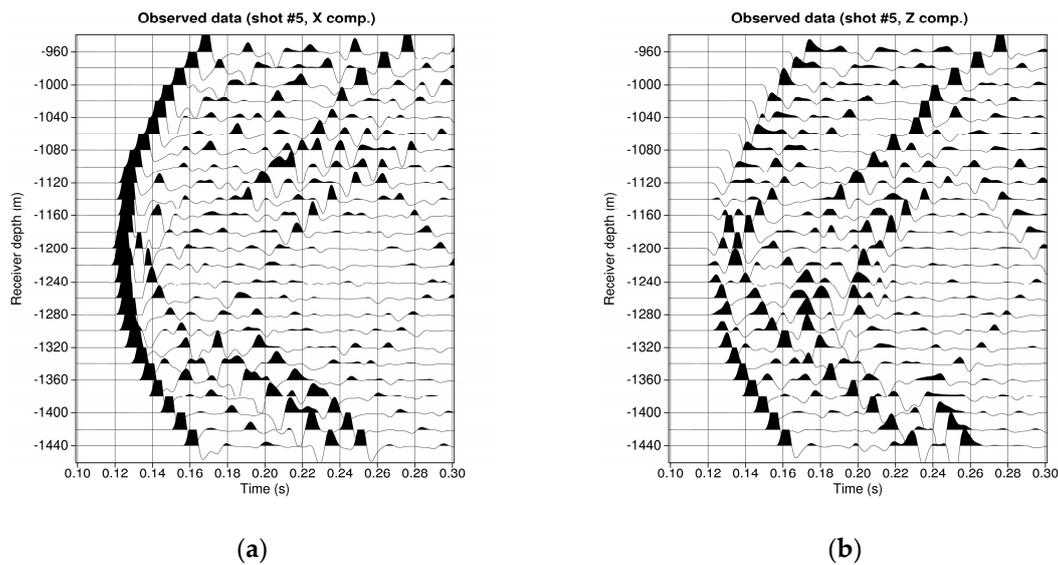


Figure 5. The 5th shot, at 1200 m deep, of the observed data obtained by full-wave modeling from the reference models: (a) The horizontal component; (b) the vertical component. The source function is a first derivative of a Gaussian with a central frequency of 100 Hz and it acts as a stress source with the following radiation coefficients $\sigma_{xx} = 1$, $\sigma_{zz} = 0.5$, $\sigma_{xz} = 0$. These data are used as observed data in the multiscale inversion process.

4.2.2. Multiscale Constrained Inversion

The inversion process consists of four successive inversions using the same observed data but with decreasing correlation ranges in horizontal and vertical directions; therefore, it is a multiscale inversion and each inversion is for a given “scale”. The model parameter results of each inversion are used as initial model parameters for the next inversion.

Uncertainties on the model space are required for the definition of C_M^{-1} . It can be defined by specifying the spatial correlation ranges and the uncertainties on the parameters. Here the uncertainties on the parameters are kept constant: $\sigma_{Vp} = 120$ m/s and $\sigma_{Vs} = 100$ m/s. The ranges of the horizontal and vertical correlations do not depend on the parameters but on the spatial location. The horizontal correlation depends on depth and both horizontal and vertical ranges depend on the inversion scale. We define three types of a priori regions with respect to depth (see Table 1): “Quasi 1D” with large correlation ranges, “2D” with short correlation ranges (where we expect to have complex geological objects, see Figure 4), and “Transition”, with correlation ranges linearly varying with depth to match the ranges of the two adjacent regions.

Table 1. A priori geological region type with respect to depth.

Depth (m)	Region Type	Number of Vertical Points
800–1120	Quasi 1D	161
1120–1180	Transition	29
1180–1340	2D	81
1340–1400	Transition	29
1400–1600	Quasi 1D	101

Table 2 lists the different correlation ranges associated with the four successive inversion scales and for the a priori region types. The four successive inversion scales are denoted by (b), (c), (d), and (e) in relation to the results displayed in Figures 6 and 7. The correlation ranges are gradually decreasing with the successive inversion scales in order to solve first the large wavelengths in the model before resolving the small wavelengths (i.e., the small model structures).

Table 2. Horizontal and vertical correlation ranges for the different a priori region types and inversion scales.

Region Type	Correlation Ranges (m)	Inversion Scale (b)	Inversion Scale (c)	Inversion Scale (d)	Inversion Scale (e)
Quasi 1D	(r_x, r_z)	(1000, 24)	(1000, 10)	(1000, 6)	(1000, 2)
2D	(r_x, r_z)	(320, 24)	(80, 10)	(20, 6)	(5, 2)

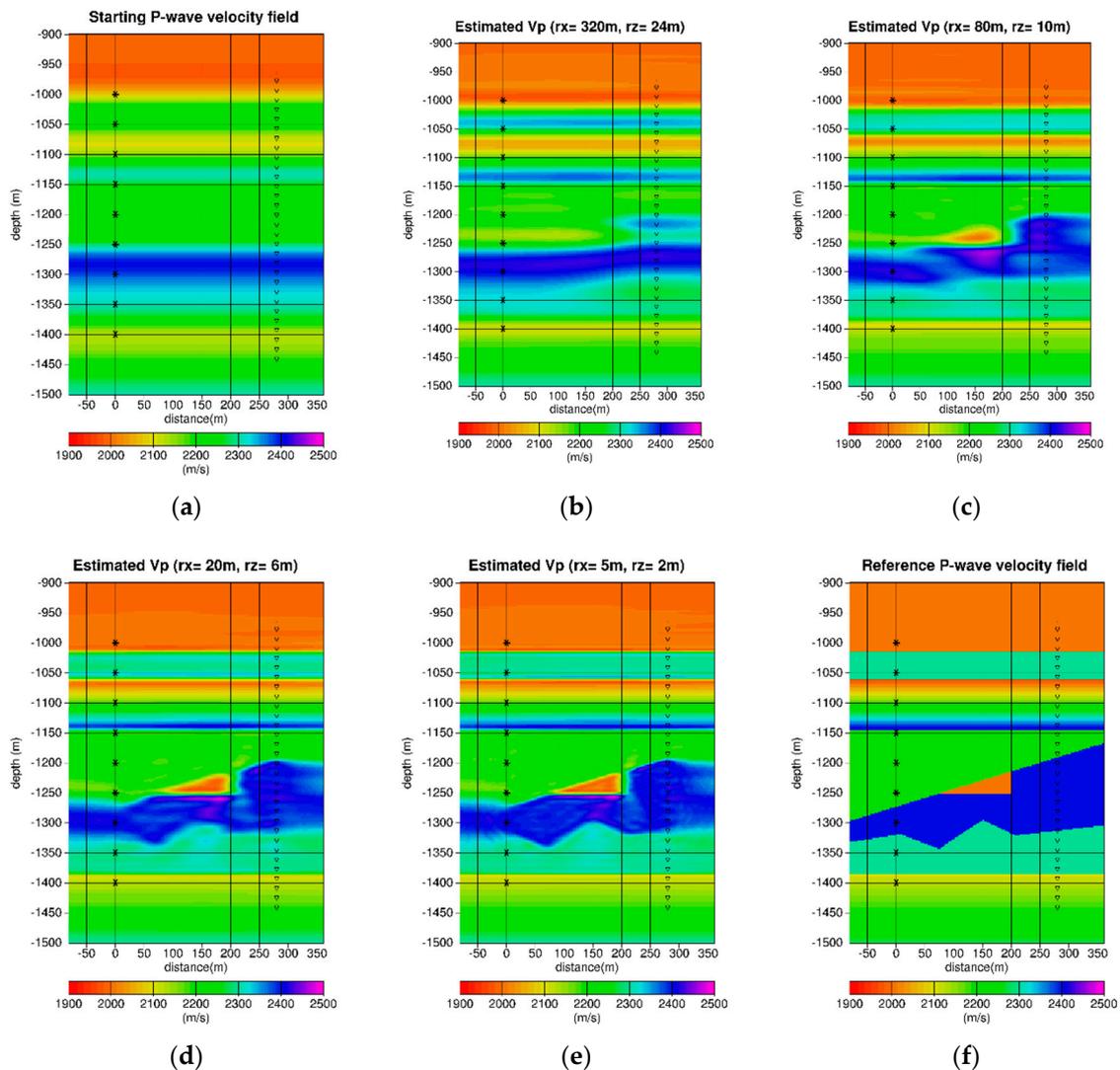


Figure 6. The results of the multiscale inversion process for the P-wave velocity field. Stars denote the source location while the small triangles denote the receiver location: (a) The P-wave velocity starting model; (b) the estimated P-velocity field obtained from the first inversion scale using correlation ranges of $r_x = 320$ m and $r_z = 24$ m in the 2D region; (c) the estimated P-velocity field obtained from the second inversion scale using correlation ranges of $r_x = 80$ m and $r_z = 10$ m in the 2D region; (d) the estimated P-velocity field obtained from the third inversion scale using correlation ranges of $r_x = 20$ m and $r_z = 6$ m in the 2D region; (e) the estimated P-velocity field obtained from the fourth and last inversion scale using correlation ranges of $r_x = 5$ m and $r_z = 2$ m in the 2D region; (f) the reference P-wave velocity model. The results for the first scale of the inversion do not allow recovery of the 2D structures, as the correlation ranges are greater than the size of the objects. The 2D structures start to appear from the second scale, improving from the third scale, and are well recovered for the last scale.

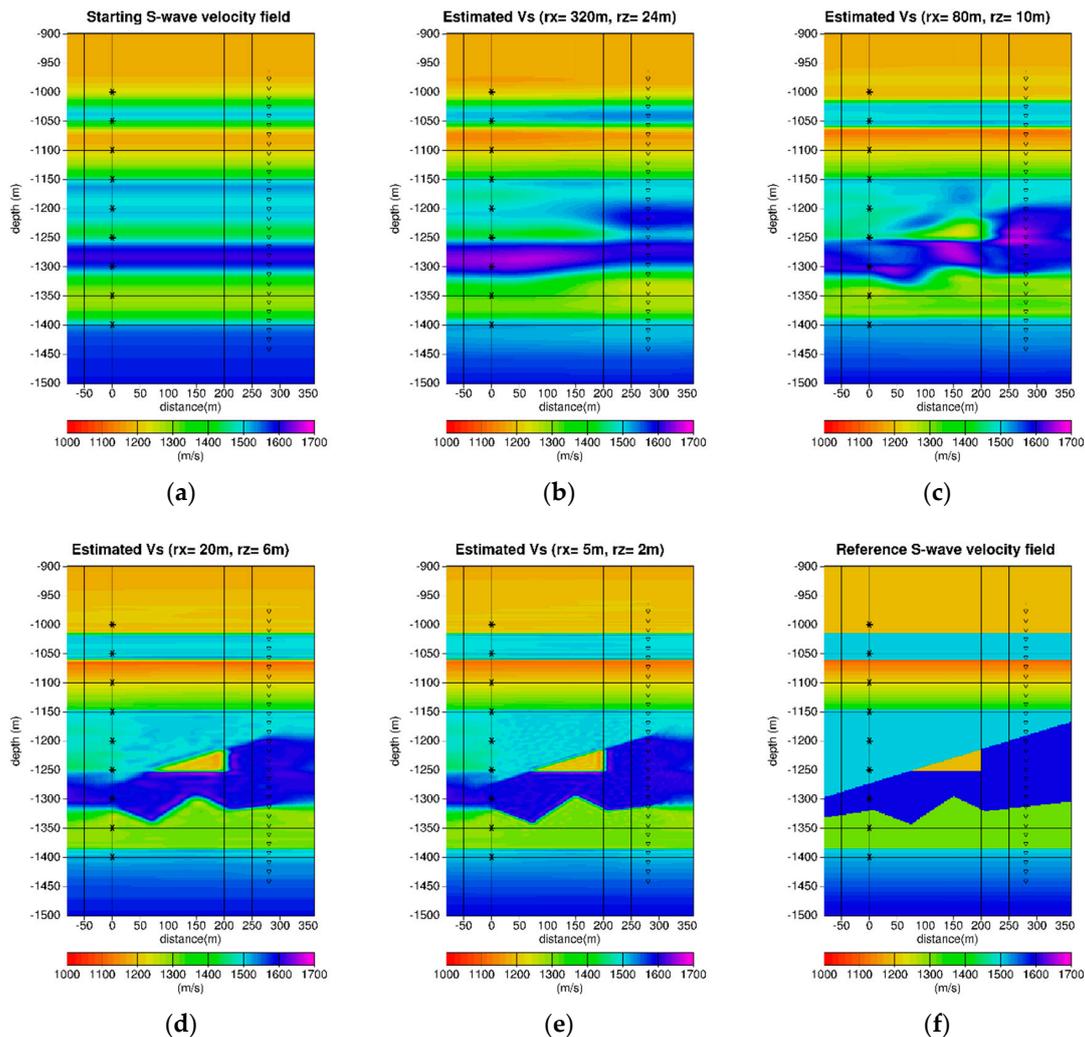


Figure 7. The results of the multiscale inversion process for the S-wave velocity field. Stars denote the source location while the small triangles denote the receiver location: (a) The S-wave velocity starting model; (b) the estimated S-velocity field obtained from the first inversion scale using correlation ranges of $r_x = 320$ m and $r_z = 24$ m in the 2D region; (c) the estimated S-velocity field obtained from the second inversion scale using correlation ranges of $r_x = 80$ m and $r_z = 10$ m in the 2D region; (d) the estimated S-velocity field obtained from the third inversion scale using correlation ranges of $r_x = 20$ m and $r_z = 6$ m in the 2D region; (e) the estimated S-velocity field obtained from the fourth and last inversion scale using correlation ranges of $r_x = 5$ m and $r_z = 2$ m in the 2D region; (f) the reference S-wave velocity model. The S-wave estimated field exhibits a better spatial resolution than the P-wave (see Figure 6) due to the smaller wavelength content associated with this mode. Small artifacts can be observed around the 2D triangle structure, pointing out the fact that the problem is less constrained in this part of the model.

The results of the multiscale inversion process for the P-wave velocity field are displayed in Figure 6. The P-wave velocity starting model is obtained by a low-frequency full waveform inversion. It is also used as the starting model for the unconstrained inversion; such starting models can also be obtained from traveltime inversion, especially for crosswell data. The estimated P-velocity field is obtained from the first inversion scale, i.e., using correlation ranges of 320 m and 24 m in the 2D region, respectively, for the horizontal and vertical correlations. One can notice that the 2D structure cannot be resolved, as the horizontal correlation range is larger than the size of the triangular object. The 1D structures are partially retrieved. For the inversion scale (c), the estimated P-velocity field is, obtained from the inversion scale (b), i.e., using correlation ranges of 80 m and 10 m in the 2D region,

respectively, for the horizontal and vertical correlations. One can notice that the 2D structure start to be resolved as both correlation ranges are smaller than the size of the triangle. The 1D structures are well estimated, with some vertical smoothing due to the vertical correlation range of 10 m. For the inversion scale (d), the estimated P-velocity field is obtained from the third inversion scale, i.e., using correlation ranges of 20 m and 6 m in the 2D region, respectively, for the horizontal and vertical correlations. One can notice that the 2D structure is getting better spatial resolution, for both the triangle and the lower zigzag interface. The 1D structures are well resolved. For the inversion scale (e), the estimated P-velocity field is obtained from the fourth and last inversion scale, i.e., using correlation ranges of 5 m and 2 m in the 2D region, respectively, for the horizontal and vertical correlations. One can notice that the delineation of the 2D structure is close to the reference model. As expected, the estimated field outside the zone between the two wells is not reliable.

The results of the multiscale inversion process for the S-wave velocity field are displayed in Figure 7. The results are similar to the P-wave velocity field (Figure 6). The main difference is clear at the fourth scale (e): the spatial resolution is almost perfect for the S-wave due to smaller wavelength content associated to this mode. However, the inversion process provides an overestimation of velocity contrast near the interfaces and some small oscillations in the estimated field clearly visible for the light blue domain above the central triangle. It means that the number of degrees of freedom is too high regarding the amount of information provided by the data.

The misfit function for the successive multiscale inversion continuously decreases from one inversion scale to another (see Table 3).

Table 3. The misfit and iteration number for the different inversion scales.

Inversion Parameters	Inversion Scale (b)	Inversion Scale (c)	Inversion Scale (d)	Inversion Scale (e)
Initial Misfit	45.5%	27.9%	6.55%	0.81%
Final Misfit	27.9%	6.55%	0.81%	0.14%
Iteration #	30	30	50	50

The number of degrees of freedom per point are provided in Table 4 for the different inversion scales. These numbers of freedoms are calculated from Equation (14). The ratio of the number of degrees with respect to the number of model parameters (i.e., grid points), in other words the number of degrees of freedom per point, is provided. It is increasing slower for the quasi 1D region than for the 2D region because only the vertical correlation range is decreasing, while both horizontal and vertical ranges are decreasing in the 2D region. These ratios are pointing out that the degrees of freedom are significantly lower in the early scales of the multiscale inversion, allowing better constraint of the problem and improving the stability by defining an overdetermined system. During the last scale stages, the constraints are relaxed, allowing one to obtain a better resolution while avoiding artifacts in the estimated fields. Even for the last scale, the ratio indicates about 5 times less free parameters than in the unconstrained inversion.

Table 4. Number of degrees of freedom (DoF) per point for the successive multiscale inversions.

Region Type	Inversion Scale (b)	Inversion Scale (c)	Inversion Scale (d)	Inversion Scale (e)
Quasi 1D	1919	2812	3416	4554
Transition	267	468	622	857
2D	1022	2922	6792	15,725
All	3208	6202	10,830	21,136
DoF/point	2.8%	5.5%	9.6%	18.8%

For comparison purposes with the results of the constrained inversion, we performed an unconstrained inversion. In the Figure 8, we display the results of the multiscale constrained inversion with respect to the unconstrained inversion for the same misfit of 0.3%. We can notice that the unconstrained inversion results are noisier than the constrained inversion. The main artifacts

can be found near the sources (classical problem in borehole seismic full waveform inversion), below the triangle structure, above the source or receiver zones (migration smile type artifacts), and in the constant gradient 1D layers. These artifacts in the estimated fields are the consequence of the instability of the least-squares inversion problem when the problem is not correctly determined for all parameters, in other words, when it is partially underdetermined. The number of degrees of freedom is too large regarding the information provided by the data. Reducing the number of degrees of freedom of the problem taking into account prior geological information (i.e., constraining the fields using spatial statistic where we have more prior information) allows one to overcome this problem, reducing drastically the number and the magnitude of the artifacts.

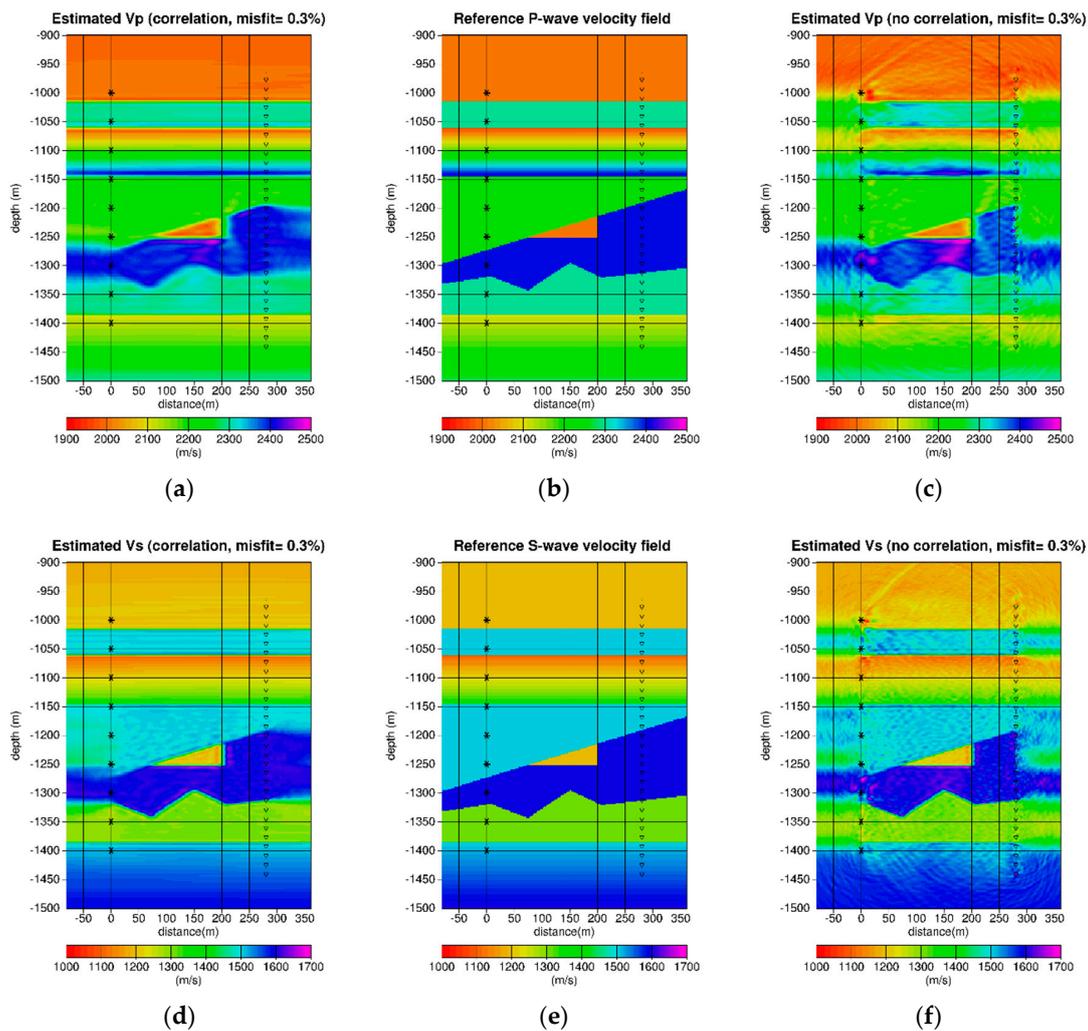


Figure 8. Comparison between results from the constrained inversion using the multiscale inversion process results of the multiscale inversion process and results from the unconstrained inversion: (a) The estimated P-wave velocity field for the multiscale inversion for a misfit of 0.3%; (b) the reference P-wave velocity field; (c) the estimated P-wave velocity field for the unconstrained inversion for a misfit of 0.3%; (d) the estimated S-wave velocity field for the multiscale inversion for a misfit of 0.3%; (e) the reference S-wave velocity field; (f) the estimated S-wave velocity field for the unconstrained inversion for a misfit of 0.3%. Both P- and S-wave estimated fields show that the multiscale inversion results are smoother than the unconstrained inversion results, with well artifact and better reliability. The main artifacts are near the sources, below the triangle structure, above the source or receiver zones and below (less visible due to the color range), and in the constant gradient 1D layers. As expected, there is no additional information outside the square delimited by the wells for the case of unconstrained inversion.

5. Conclusions

In this paper, we have illustrated how to introduce, in the inverse problem, constraints consistent with the least squares formalism (through covariance matrices) on both the data and the model space.

Constraining the seismic inverse problem in the data space is not straightforward but when possible, it is a very powerful tool. In the case of a multicomponent borehole seismic data, using the polarization in the inversion process allowed better recovery of the sharpness of the interfaces at the level of the receivers.

Constraining the seismic inverse problem in the model space can be achieved by defining a priori model parameters and evaluating the model uncertainties associated with this model. Horizontal and vertical spatial correlation using Laplace distribution can be used to fill the model space covariance matrix in order to introduce an a priori information favoring the tabular region versus more complex regions, by varying accordingly the ranges of these correlations. The inversion favors solutions are consistent with our a priori, whenever it does not contradict the information contained in the data. Moreover, by adopting a multiscale type of inversion by relaxing the correlation ranges, it regularizes the inverse problem.

The incorporation of all our a priori knowledge of the parameters and all statistical studies on the data, allows not only the algorithmic stabilization of the inversion process, but also the reduction of the solution set for an underdetermined problem, with the purpose being not necessarily to converge quickly towards a good model (in term of residuals), but to prospect regions of the model space populated by models that are sensible, a priori, and also yielding the lowest possible misfit.

Author Contributions: Conceptualization, M.C. and C.B.; methodology, M.C. and C.B.; validation, M.C. and C.B.; formal analysis; writing—original draft preparation, M.C. and C.B.; writing—review and editing, M.C. and C.B.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this appendix, we will explicitly state the relation between the correlation matrix and the degrees of freedom.

Appendix A.1 Generalities

Let us consider a 1D Gaussian random field \mathbf{X} defined in section *Quantification of number of degrees of freedom* and characterized by its p.d.f. given by Equation (7).

If the covariance matrix \mathbf{C} is diagonal (i.e., no correlation between the random variables X_i), Equation (7) can be rewritten:

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^{n/2} \prod_i \sigma_i} \exp\left(-\frac{1}{2} \sum_i \left(\frac{x_i - m_i}{\sigma_i}\right)^2\right) \\ &= \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2} \sum_i \left(\frac{x_i - m_i}{\sigma_i}\right)^2\right) \\ &= \prod_i f_i(x_i), \end{aligned} \quad (\text{A1})$$

where σ_i is the standard deviation of the Gaussian random variable X_i for the i^{th} point, and $f_i(x_i)$ the p.d.f. associated with this variable X_i . For a Gaussian field, when correlation coefficients are null, the variables X_i are independent and as a consequence, the p.d.f. of the field is the product of the marginal p.d.f. of the variables, the f_i .

Assuming for simplicity that the m_i and the σ_i are constant, the field is made of n independent variables, identically distributed. Then, considering we have in this case n free parameters, we will study the effect of the non-independence of variables on the number of the free parameters, incorporating correlation between variables through a non-diagonal covariance matrix.

Appendix A.2 Sequential Simulation

Whatever the discrete random field, the knowledge of the joint p.d.f. (like f defined in Equation (7) for the Gaussian random field \mathbf{X}) allows simulation of a realization of the field sequentially, using the marginal-conditional decomposition (correct whatever the order of the indexes):

$$f(x_1, x_2, \dots, x_n) = g_1(x_1) \cdot g_2(x_2|x_1) \cdot g_3(x_3|x_1, x_2) \cdot \dots \cdot g_n(x_n|x_1, x_2, \dots, x_{n-1}), \tag{A2}$$

where g_i is the marginal-conditional p.d.f. of the random variable X_i , independently of the random variables X_j for $j > i$ and conditionally to the random variables X_k for $k < i$. More precisely, g_i is defined by the relation:

$$g_i(x_i|x_1, x_2, \dots, x_{i-1}) = \frac{\int dx_{i+1} \int dx_{i+2} \dots \int dx_n f(x_1, x_2, \dots, x_n)}{\int dx_i \int dx_{i+1} \dots \int dx_n f(x_1, x_2, \dots, x_n)}. \tag{A3}$$

Using Equation (A2), it is possible to simulate a realization of the field sequentially. First, we use $g_1(x_1)$ to generate a realization of the random variable X_1 , independently of all other variables. Once $x_1 = x_1^0$ is determined, we use $g_2(x_2|x_1^0)$ to generate a realization x_2^0 of the random variable X_2 , dependent to the value x_1^0 , but independent of the other random variables X_3 to X_n . We repeat this procedure using g_3 , then g_4 , and so on, until the last value x_n^0 is determined, thereby obtaining a realization \mathbf{x}^0 of the random field \mathbf{X} .

We can link the degree of freedom of the variable X_i , when the values x_i to x_{i-1} are known, to the standard deviation σ_i^c of the Gaussian defined by the p.d.f. $g_i(x_i|x_1, x_2, \dots, x_{i-1})$. To account for the correlation between the variables X_i , independently of the values of the standard deviations σ_i of the marginal p.d.f. $f_i(x_i)$, we can then define the degrees of freedom of the variable X_i (when x_1 to x_{i-1} are known) as the following ratio:

$$d = \frac{\sigma_i^c}{\sigma_i}, \tag{A4}$$

which equals one when X_i is not correlated to the variables $(X_1, X_2, \dots, X_{i-1})$ and zero when the correlation is perfect.

Appendix A.3 The LU Simulation Technique

In order to simplify the calculation of the degrees of freedom, we introduce here the LU decomposition technique for Gaussian field simulation, as this method can be interpreted in term of sequential simulation using Equation (7).

A way to simulate a Gaussian discrete random field \mathbf{X} defined by the p.d.f. in Equation (A2) (n points with expectation \mathbf{m} and covariance matrix \mathbf{C}) is the use of the LU decomposition of the covariance matrix \mathbf{C} :

$$\mathbf{C} = \mathbf{LU} = \mathbf{LL}^T, \tag{A5}$$

where \mathbf{L} denotes the left lower triangle matrix and \mathbf{U} the upper right one.

Considering a random field \mathbf{E} constituted by n random variables, independently and identically distributed following a Gaussian law with zero mean and unit standard deviation; the field $\mathbf{m} + \mathbf{LE}$ is then equivalent to \mathbf{X} . Therefore, to simulate a realization \mathbf{x} of the random field \mathbf{X} , we can use a realization \mathbf{e} of the field \mathbf{E} and write:

$$\mathbf{x} = \mathbf{m} + \mathbf{Le}. \tag{A6}$$

Appendix A.4 Interpretation of the LU Decomposition of the Covariance Matrix

Let l_{ij} be the elements of the matrix \mathbf{L} . The relation given by Equation (A6) can be interpreted in term of sequential simulation: the first line gives $x_1 = m_1 + l_{11}e_1$, meaning that we simulate the realization of X_1 independently of the other variables, as if we used the 1D marginal p.d.f. $g_1(x_1)$ defined in Equation (A2):

$$g_1(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\left(\frac{x_1 - m_1}{\sigma_1}\right)^2\right) = f_1(x_1). \tag{A7}$$

Once the value of x_1 is fixed, the second line of the Equation (15) means that the second random variable x_2 can be written:

$$X_2 = m_2 + l_{21}e_1 + l_{22}E_2 = m_2 + l_{21}\left(\frac{x_1 - m_1}{l_{11}}\right) + l_{22}E_2, \tag{A8}$$

Meaning that we simulate the realization of X_2 independently of X_3, \dots, X_n but dependent of the value $X_1 = x_1$, as if we used the conditional marginal p.d.f. $g_2(x_2|x_1)$ (see Equation (A2)). A given line i of the Equation (A6) means we can simulate a random variable X_i conditionally to the determined values of the preceding variables x_1 to x_{i-1} , as if we used the proper marginal-conditional p.d.f. $g_i(x_i|x_1, x_2, \dots, x_{i-1})$ (see Equation (A2)).

Let us come back to the estimation of the degrees of freedom. The standard deviations σ_i^f are given here by the i^{th} diagonal element l_{ii} of the matrix \mathbf{L} . This standard deviation is lesser than the standard deviation of the variable X_i , independently of the other variables (i.e., the standard deviation σ_i of the Gaussian defined by the marginal p.d.f. $f_i(x_i)$). When correlation between X_i and the variables X_1 to X_{i-1} is strong, the ratio $\frac{l_{ii}}{\sigma_i}$ is weak. When X_i is independent of the variables X_1 to X_{i-1} , the ratio equals 1, meaning the freedom of the variable is total. We can say that when variables are independent (no correlation), the number of the free parameters is $n = \sum_i \frac{1}{1}$ and when correlation exists, it is reduced to:

$$N = \sum_i \frac{l_{ii}}{\sigma_i}. \tag{A9}$$

Appendix A.5 The LU Decomposition of the Correlation Matrix \mathbf{R}

The covariance matrix can be written with the following form:

$$\mathbf{C} = \mathbf{SRS}, \tag{A10}$$

where \mathbf{R} is the correlation matrix and \mathbf{S} the diagonal matrix of the standard deviations;

$$\mathbf{S} = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n \end{pmatrix}. \tag{A11}$$

The LU decomposition of \mathbf{R} gives: $\mathbf{R} = \mathbf{MM}^T$, with $\mathbf{L} = \mathbf{SM}$. Therefore, the field \mathbf{X} can be written (see Equation (20)): $\mathbf{X} = \mathbf{m} + \mathbf{LE} = \mathbf{m} + \mathbf{SME}$.

The number of free parameters is then:

$$N = \sum_i \frac{l_{ii}}{\sigma_i} = trace(\mathbf{M}). \tag{A12}$$

Appendix B

Let us consider a pair of random variables (X, Y) characterized by the following p.d.f.:

$$f(x, y) = \frac{1}{2\pi\sqrt{\det(\mathbf{C})}} \exp\left(-\frac{1}{2} \begin{pmatrix} x - m_x \\ y - m_y \end{pmatrix}^T \mathbf{C}^{-1} \begin{pmatrix} x - m_x \\ y - m_y \end{pmatrix}\right), \tag{A13}$$

where \mathbf{C} is the covariance matrix, we have:

$$\mathbf{C} = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}, \det(\mathbf{C}) = \sigma_x^2\sigma_y^2(1 - \rho^2), \mathbf{C}^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} \frac{1}{\sigma_x^2} & \frac{-\rho}{\sigma_x\sigma_y} \\ \frac{-\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{pmatrix},$$

where ρ is the correlation between the random variables X and Y . An example is given in Figure A1, with the following values: $m_x = m_y = 0, \sigma_x = \sigma_y = 1, \rho = 4/5$.

Equation (16) of the decomposition of the p.d.f. is written:

$$f(x, y) = f_1(x) \cdot f_2(y|x), \tag{A14}$$

where

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{1}{2} \left(\frac{x - m_x}{\sigma_x}\right)^2\right) \\ f_{y|x}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2} \left(\frac{y - m_y - \rho\sigma_y\left(\frac{x - m_x}{\sigma_x}\right)}{\sigma_y\sqrt{1 - \rho^2}}\right)^2\right). \tag{A15}$$

We can note that the characteristics of these p.d.f. are given by the elements of the left triangle matrix of the covariance LU decomposition:

$$\mathbf{L} = \begin{pmatrix} \sigma_x & 0 \\ \rho\sigma_y & \sigma_y\sqrt{1 - \rho^2} \end{pmatrix}. \tag{A16}$$

Indeed, the first line corresponds to the standard deviation of the p.d.f. $f_x(x)$. The conditional p.d.f. $f_{y|x}(y|x)$ is plotted on Figure A1. It is characterized by its mean $m_c = m_y + \rho\sigma_y\left(\frac{x - m_x}{\sigma_x}\right)$ and its standard deviation $\sigma_c = \sigma_y\sqrt{1 - \rho^2}$. In the second row, the first element corresponds to the scale factor of the correction of the mean of the conditional p.d.f. $f_{y|x}$ and the second element to the standard deviation of the conditional p.d.f. In summary, the diagonal elements are the standard deviations of the random variables X and $Y|X$. The effect of the correlation ρ is to reduce, by the factor $\sqrt{1 - \rho^2}$, the degree of freedom of the second variable Y when X is known, as we see in Figure A1. We then note that the total degree of freedom of this pair is (taking as a reference the uncorrelated pair characterized by the diagonal covariance matrix, $\rho = 0$):

$$d = \frac{\sigma_x}{\sigma_x} + \frac{\sigma_y\sqrt{1 - \rho^2}}{\sigma_y} = 1 + \sqrt{1 - \rho^2}. \tag{A17}$$

We can note that $d = 1$ when the correlation is perfect ($\rho = \pm 1$) and $d = 2$ when the correlation is inexistent ($\rho = 0$).

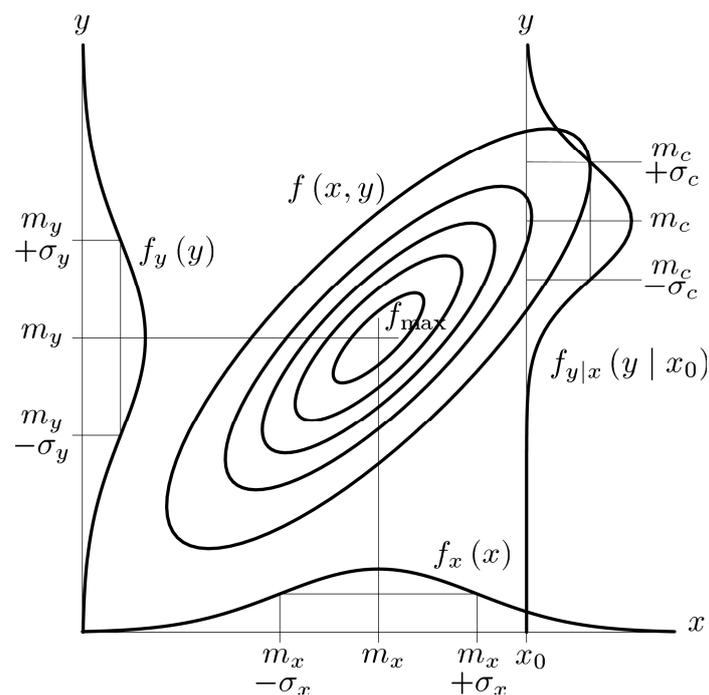


Figure A1. The joint probability density $f(x, y)$ of the couple of random variables (X, Y) is plotted using contour levels (at level $0.9 \cdot f_{max}$, $0.7 \cdot f_{max}$, $0.5 \cdot f_{max}$, $0.3 \cdot f_{max}$ and $0.1 \cdot f_{max}$), the marginal probability densities for the variables themselves ($f_x(x)$ for X and $f_y(y)$ for Y) are plotted on the x and the y axes. The value $x_0 = 1.5$ has been fixed and the conditional probability density $f_c(y) = f_{y|x}(y | x_0)$ is plotted, the mean $m_x = m_y = 0$ and m_c has been reported as the standard deviation $\sigma_x = \sigma_y = 1$ and σ_c , the correlation $\rho = 0.8$ appears only in the ellipticity of the ellipses for the joint p.d.f. f . As Equation (A17) shows, the number of free parameters of the couple (X, Y) is $1 + \frac{\sigma_c}{\sigma_y} = 1 + \sqrt{1 - \rho^2} = 1.6$.

References

- Virieux, J.; Operto, S. An overview of full waveform inversion in exploration geophysics. *Geophysics* **2009**, *74*, WCC1–WCC26. [[CrossRef](#)]
- Menke, W. *Geophysical Data Analysis: Discrete Inverse Theory*; Academic Press, Inc.: Orlando, FL, USA, 1984.
- Tarantola, A. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation: Discrete Inverse Theory*; Elsevier Science Publ. Co., Inc.: New York, NY, USA, 1987.
- Tikhonov, A.; Arsenin, V. *Solution of Ill-Posed Problems*; Winston: Washington, DC, USA, 1977.
- Operto, S.; Virieux, J.; Dessa, X.; Pascal, G. Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: Application to the eastern Nankai trough. *J. Geophys. Res.* **2006**, *111*, B09306. [[CrossRef](#)]
- Guittou, A.; Ayeni, G.; Gonzales, G. A preconditioning scheme for full waveform inversion. In *SEG Technical Program Expanded Abstracts 2010*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2010; pp. 1008–1012. [[CrossRef](#)]
- Guittou, A. A blocky regularization scheme for full waveform inversion. In *SEG Technical Program Expanded Abstracts 2011*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2011; pp. 2418–2422. [[CrossRef](#)]
- Abubakar, A.; Hu, W.; Habashy, T.M.; van den Berg, P.M. Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data. *Geophysics* **2009**, *74*, WCC47–WCC58. [[CrossRef](#)]
- Herrmann, F.J.; Erlangga, Y.A.; Lin, T.T.Y. Compressive simultaneous full-waveform simulation. *Geophysics* **2009**, *74*, A35–A40. [[CrossRef](#)]
- Loris, I.; Douma, H.; Nolet, G.; Daubechies, I.; Regone, C. Nonlinear regularization techniques for seismic tomography. *J. Comput. Phys.* **2010**, *229*, 890–905. [[CrossRef](#)]
- Asnaashari, A.; Brossier, R.; Garambois, S.; Audebert, F.; Thore, P.; Virieux, J. Regularized seismic Full-waveform inversion with prior model information. *Geophysics* **2013**, *78*, R25–R36. [[CrossRef](#)]

12. Asnaashari, A.; Brossier, R.; Garambois, S.; Audebert, F.; Thore, P.; Virieux, J. Time-lapse seismic imaging using regularized full-waveform inversion with a prior model: Which strategy? *Geophys. Prospect.* **2015**, *63*, 78–98. [[CrossRef](#)]
13. Polak, E.; Ribière, G. Note sur la convergence de directions conjuguée. *Rev. Fr. Inform. Rech. Oper. 3e Année* **1969**, *16*, 35–43.
14. Crase, E. Robust Elastic Nonlinear Inversion of Seismic Waveform Data. Ph.D. Thesis, University of Houston, Houston, TX, USA, 1989.
15. Jurkevics, A. Polarization analysis of the three component array data. *Bull. Seismol. Soc. Am.* **1988**, *78*, 1725–1743.
16. Virieux, J. *P-SV* wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics* **1986**, *51*, 889–901. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).