Local optimization approaches for simultaneous AVO inversion based on re-parameterized Zoeppritz equations

Mariana Lume and Kris Innanen

ABSTRACT

Linearized AVO inversion methods, such as the weighted stacking approach, are based on approximations of the Zoeppritz equations subject to several assumptions, including the limitation of incidence angles to 35-40°, typically smaller than the critical one. Thus, in long-offset acquisitions these approaches fail. This study focuses on developing a nonlinear inversion appropriate under these circumstances, based on re-parameterized Zoeppritz equations in terms of the fractional density and compressional and shear impedances. To achieve this, P-P and P-S datasets with different characteristics of noise and frequency are considered. In general, conditioned to a good initial model, three different local optimization algorithms demonstrate a superior performance respect to the simultaneous weighted stacking inversion, producing more accurate results for most elastic parameters, and the inclusion of noise information during the inversion improves the fractional shear impedance and the fractional Vp/Vs ratio, promising a better rock properties estimation.

INTRODUCTION

Amplitude versus offset (AVO) is an interpretation tool relevant for reservoir description, which provides complementary information to the conventional stacked seismic (Downton et al., 2000). The inverse problem consists in determining elastic parameters, assuming that the data can be reasonably forward modeled by elasticity theory. Conventional approaches to invert pre-stack data seek linearized approximations of the Zoeppritz equations (Grossman, 2003). One successful method is the weighted stacking technique, proposed by Smith and Gidlow (1987) and expanded by Stewart (1990) and Larsen (1999). This is based on the minimization of the misfit between the observed data and its theoretical description using the Aki-Richards approximations; then, the resulting pair of linear equations is solved, obtaining expressions for the fractional compressional and shear impedances, encompassing the multiplication of P-P and P-S reflectivities and time and angle-variant weights (Larsen, 1999). This method produces accurate estimates as long as its underlying assumptions are met, i.e., weak contrast of elastic parameters, incidence angles smaller than the critical one, incidence angles smaller than 35° thus very small values of $\Delta\rho/\rho$ are assumed, and values of the Vp/Vs ratio between 1.5 and 2 (Smith and Gidlow, 1987; Larsen, 1999).

The traditional procedure for angles larger than the critical one is to limit them, but this compromises the reliability of the estimates since they depend on the range of angles, and in some cases the critical angle may be as low as 25° (Downton and Ursenbach, 2005). Additionally, there is a growing interest in seismic acquisitions with long-offset ranges for different industry purposes, such as studying reservoirs with strong contrast interfaces or using streamer recordings to analyze the missing converted S-waves energy, occurring at wider angles of incidence (Skopintseva et al., 2011; Williams et al., 2001).

Since the linearized inversion would not produce accurate estimations if the available angles are higher than 35-40°, this study aims for developing an inversion strategy appropriate for scenarios with long-offsets/large incidence angles. The Zoeppritz equations are not mathematically conditioned to any particular range of angles, hence the proposed approach is based on their re-parameterization in terms of the fractional impedances and density. However, in reality a plane wave approximation would be made, since seismic data is not produced by plane but spherical waves. In order to gain general insights, the nonlinear inversion is evaluated against the weighted stacking approach, after considering synthetic P-P and P-S reflectivities generated from a two-layer model which does not produce a critical angle; thus, only effects of increasing the incidence angles will be accounted for. To achieve this, three local optimization methods are tested, namely the Pure Gauss-Newton, Steepest Descent, and Levenberg-Marquardt and it is also studied how the selection of the initial model affects the nonlinear inversion.

FORWARD MODELING

To perform the forward problem, plane wave Zoeppritz equations are considered:

$$\mathbf{P}\begin{bmatrix} R_{pp} \\ R_{ps} \\ T_{pp} \\ T_{ps} \end{bmatrix} = \mathbf{b} \tag{1}$$

where:

$$\mathbf{P} = \begin{bmatrix} -X & -\sqrt{1-B^2X^2} & CX & \sqrt{1-D^2X^2} \\ \sqrt{1-X^2} & -BX & \sqrt{1-C^2X^2} & -DX \\ 2B^2X\sqrt{1-X^2} & B(1-2B^2X^2) & 2AD^2X\sqrt{1-C^2X^2} & AD(1-2D^2X^2) \\ -(1-2B^2X^2) & 2B^2X\sqrt{1-B^2X^2} & AC(1-2D^2X^2) & -2AD^2X\sqrt{1-D^2X^2} \end{bmatrix}$$

and:

$$\mathbf{b} = \begin{bmatrix} X \\ \sqrt{1-X^2} \\ 2B^2X\sqrt{1-X^2} \\ 1-2B^2X^2 \end{bmatrix}$$

and where:

$$A = \frac{\rho_2}{\rho_1},\ B = \frac{\beta_1}{\alpha_1},\ C = \frac{\alpha_2}{\alpha_1},\ D = B\frac{\beta_2}{\beta_1},\ X = sin\theta_1$$

here, α, β and ρ are the P-wave velocity, S-wave velocity and density, respectively.

Equation 1 is re-parameterized in terms of the upper medium properties and jumps in the impedances and density, which is familiar from AVO analysis. To accomplish this, terms A, C, and D are modified according to equation 2 commonly used when linearizing problems.

$$\frac{Y_2}{Y_1} = \left(1 + \frac{1}{2}\frac{\Delta Y}{Y}\right) \Big/ \left(1 - \frac{1}{2}\frac{\Delta Y}{Y}\right) \tag{2}$$

Since C and D depend on α and β, respectively, these expressions are altered in terms of the impedances *I*=$\alpha\rho$ and *J*=$\beta\rho$:

$$C = \frac{I_2}{I_1}\frac{\rho_1}{\rho_2};\ D = B\frac{J_2}{J_1}\frac{\rho_1}{\rho_2} \tag{3}$$

Applying equation 2 to term A and to terms C and D of equation 3 produces:

$$A = \left(1 + \frac{1}{2}\frac{\Delta\rho}{\rho}\right) \Big/ \left(1 - \frac{1}{2}\frac{\Delta\rho}{\rho}\right) \tag{4}$$

$$C = A^{-1}\left(1 + \frac{1}{2}\frac{\Delta I}{I}\right) \Big/ \left(1 - \frac{1}{2}\frac{\Delta I}{I}\right) \tag{5}$$

$$D = BA^{-1}\left(1 + \frac{1}{2}\frac{\Delta J}{J}\right) \Big/ \left(1 - \frac{1}{2}\frac{\Delta J}{J}\right) \tag{6}$$

Now, each element of **P** is a nonlinear function of $\Delta I/I$, $\Delta J/J$, and $\Delta\rho/\rho$.

By inverting **P** and multiplying it on both sides of equation 1, the solution of the four reflection coefficients is obtained for a chosen P-wave incidence angle θ_1. To work with P-P and P-S data simultaneously, the raytracing methodology applied by Larsen (1999) was followed in this study, i.e., different incident angles for P-P and P-S mode energy conversions were computed, in order each pair of plane waves reach the same receiver (offset). Hence, the forward problem needs to be performed twice, one using the P-P incidence angles and another using the P-S incidence angles. Subsequently, the first coefficient from the P-P forward problem corresponds to R_{pp}, while the second one from the P-S forward problem corresponds to R_{ps}.

ITERATIVE NONLINEAR SIMULTANEOUS INVERSION

The vector of coefficients **u** contains four elements per incidence angle, but a sampling operator **S** can be used to only extract the P-P and P-S data:

$$\mathbf{d_{pred}} = \mathbf{Su} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} R_{pp} \\ R_{ps} \\ T_{pp} \\ T_{ps} \end{bmatrix} = \begin{bmatrix} R_{pp} \\ R_{ps} \end{bmatrix} \tag{7}$$

Counterpart to the predicted data are real or observed data, d_{obs}, which are elements of the same space that can be compared through subtraction. When performing an unconstrained optimization, an objective function (without restrictions on the values that the variables can take) is minimized (Nocedal and Wright, 2006). This objective function is constructed measuring the difference between predicted and observed data vectors, i.e., from the L2 norm:

$$\phi(\mathbf{m}) = \frac{1}{2}\sum_{j=1}^{N}\Big(\mathbf{Su}(\mathbf{m},\theta_j) - \mathbf{d_{obs}}(\theta_j)\Big)^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\mathbf{W}\Big(\mathbf{Su}(\mathbf{m},\theta_j) - \mathbf{d_{obs}}(\theta_j)\Big) \tag{8}$$

where **m** corresponds to the model parameter vector:

$$\mathbf{m} = \begin{bmatrix}\frac{\Delta I}{I} & \frac{\Delta J}{J} & \frac{\Delta \rho}{\rho}\end{bmatrix}^{\mathsf{T}} \tag{9}$$

and **W** is a diagonal matrix with form:

$$\mathbf{W} = \begin{bmatrix}1/\sigma_{pp} & 0\\ 0 & 1/\sigma_{ps}\end{bmatrix} \tag{10}$$

where σ is the standard deviation of the dataset. In this case, $\mathbf{W}^{\mathsf{T}}\mathbf{W}$ is the inverse of the data covariance matrix. The elements of this matrix act as weighting factors and this is applied because in practice some measurements are made with more accuracy than others, then it is desirable to assign greater weights to the more accurate ones (Zhdanov, 2002). If **W** is not known it can be replaced by the identity matrix.

The iterative approach involves derivatives of the objective function with respect to all the model parameters. The gradient is a vector in model space pointing in the direction of most rapid ascent in the objective function:

$$\mathbf{g} = \begin{bmatrix}\frac{\partial\phi(\mathbf{m})}{\partial m_1} & \frac{\partial\phi(\mathbf{m})}{\partial m_2} & \frac{\partial\phi(\mathbf{m})}{\partial m_3}\end{bmatrix}^{\mathsf{T}} \tag{11}$$

and:

$$\mathbf{g} = \sum_{j=1}^{N}\mathbf{J}_j^{\mathsf{T}}\mathbf{W}^{\mathsf{T}}\mathbf{W}\Big(\mathbf{Su}(\mathbf{m},\theta_j) - \mathbf{d_{obs}}(\theta_j)\Big) \tag{12}$$

J is the 2x3 Jacobian matrix; it is formed by derivatives of the predicted data with respect to each model parameter. The elements of the Jacobian matrix, per incidence angle, can be expressed as:

$$J_\mu^i = S_K^i\frac{\partial u^K}{\partial m^\mu}; \qquad i = 1,2;\ \mu = 1,2,3;\ K = 1,2,3,4 \tag{13}$$

After differentiating both sides of equation 1 with respect to each model parameter and inverting the matrix **P**, it is obtained:

$$\frac{\partial u^K}{\partial m^\mu} = -(P^{-1})_L^K\frac{\partial P_M^L}{\partial m^\mu}u^M; \qquad M = L = 1,2,3,4 \tag{14}$$

When computing **J**, the main new task is to determine all 48 elements of $\partial P_M^L/\partial m^\mu$ in equation 14, which is a straightforward, yet laborious process.

The Jacobian takes this matrix form:

$$\mathbf{J} = \begin{bmatrix} J_{pp}^{m_1} & J_{pp}^{m_2} & J_{pp}^{m_3} \\ J_{ps}^{m_1} & J_{ps}^{m_2} & J_{ps}^{m_3} \end{bmatrix} \tag{15}$$

but there are some factors that need to be applied to equation 14 to consider the raytracing approach: (1) $\partial u^K/\partial m^\mu$ is a 4x3 matrix. Its rows are formed by the derivative of a coefficient with respect to each model parameter. (2) To construct this matrix, $\mathbf{u}$ is computed twice, one per each type of incidence angle, in order to use $\mathbf{u}_{pp}$ and $\mathbf{P}_{pp}$ to calculate the first and third rows of $\partial u^K/\partial m^\mu$ and $\mathbf{u}_{ps}$ and $\mathbf{P}_{ps}$ for the second and fourth rows. (3) The construction of $\partial u^K/\partial m^\mu$ is repeated per each pair of incidence angles, forming N Jacobian matrices.

On the other hand, the Hessian is a symmetric and orthogonal matrix of second derivatives of the objective function which can be computed through:

$$\mathbf{H} = \sum_{j=1}^{N} \mathbf{J_j}^\intercal \mathbf{W}^\intercal \mathbf{W} \mathbf{J_j} \tag{16}$$

The resulting 3x3 Hessian matrices per incidence angle are summed between each other.

Features of local optimization methods

A local minimization requires a reasonable starting point, thus the algorithm iterates many times, lowering the cost function, until either no more progress is made or the solution is approximated with a given accuracy (Nocedal and Wright, 2006). To approach the local minimum, the algorithm fixes a direction $\Delta\mathbf{m}$ and a distance α to move. This distance, namely step length, can be found after solving:

$$\min_{\alpha>0} f(\mathbf{m_k} + \alpha \Delta \mathbf{m_k}) \tag{17}$$

However, this is an expensive process and a one-dimensional inexact minimization can be carried out, sacrificing accuracy but conserving computation time (Luenberger and Ye, 2008). For this, a number of candidate steps lengths are generated until finding one that approximates the minimum of expression 17, while enforcing the sufficient decrease and curvature conditions (Nocedal and Wright, 2006).

The local gradient and/or curvature of the objective function is used to compute an update of the model parameters (Sen and Stoffa, 2020). The search direction or model update can be computed with:

$$\Delta\mathbf{m} = -\mathbf{B}^{-1}\mathbf{g} \tag{18}$$

where **B** is adapted depending on the applied method:

- **B= I** $\rightarrow$ Steepest Descent
- **B= H** $\rightarrow$ Gauss-Newton
- **B= H** + $\lambda diag(\mathbf{H}) \rightarrow$ Levenberg-Marquardt modification

Steepest Descent methods always guarantee a descent direction, linear convergence rate with zigzag movements, and robustness; thus, they are computationally inexpensive, but their convergence is slow. On the other hand, for Newton's methods, the convergence rate is quadratic, but there is a lack of robustness and the Hessian could be indefinite far away from the solution; in addition, since the inverse Hessian is needed, the computational cost is higher (Eriksson, 1996; Zou, 2020). The Levenberg-Marquardt method is a regularization of the Gauss-Newton's expression, which is useful when the Hessian is ill-conditioned; the Hessian is damped by a diagonal cut off, ensuring a descent direction and global convergence (Eriksson, 1996; Lu et al., 2018). The value of the damping parameter λ is adjusted per iteration and it is equivalent to selecting a step size; when the solutions are far away from the optimal values, λ is large and the optimization behaves as Steepest Descent, and as they get closer, λ is small and the optimization behaves as the Gauss-Newton's method (Tolle, 2003).

After computing the model updates, values of the model parameters will be modified per iteration (k) according to:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha \Delta \mathbf{m}_k \tag{19}$$

COMPARISON BETWEEN OPTIMIZATION METHODS

A two-layer model of solid units in welded contact was designed with elastic properties that do not produce a critical angle. A range of offsets was assumed from 0m to 4000m with intervals of 80m and incidence angles were determined through raytracing. This model meets most of the assumptions underlying the accurate performance of the weighted stacking inversion, i.e., small contrasts across the interface and a value of Vp/Vs ratio of 1.9. However, this linearized inversion is well-suited only for angles smaller than 35°, but the reflectivities of Figure 1 have a maximum incidence angle of approximately 53° for the P-P dataset and 65° for the P-S data. Therefore, it is expected that the weighted stacking approach does not produce as good estimates as the nonlinear inversion.

For the nonlinear approach, the initial model was constructed from weak perturbations of the true values of α, β, and ρ. However, since the fractional impedances result from multiplying values of velocities and densities, the initial perturbations of the model parameters were not that small. Figure 2 compares the true and the initial model.

Each set of reflectivities was inverted with the Gauss-Newton, Steepest Descent, and Levenberg-Marquardt methods. For the Gauss-Newton optimization, a fixed step length α=1 was considered and since it took the unit value, the method is called "Pure Gauss-Newton". For the Steepest Descent, a backtracking line search was applied to compute

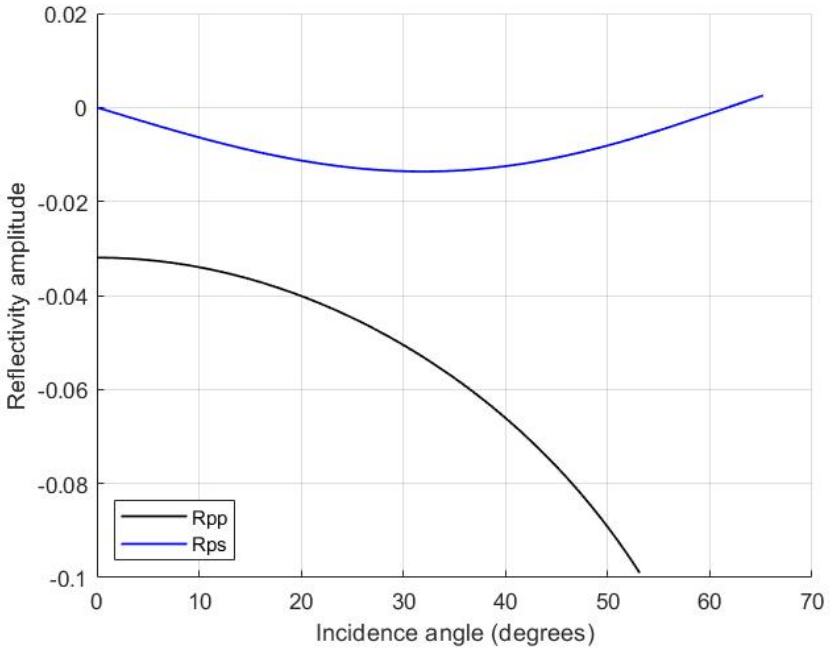

FIG. 1. P-P and P-S reflection coefficients.

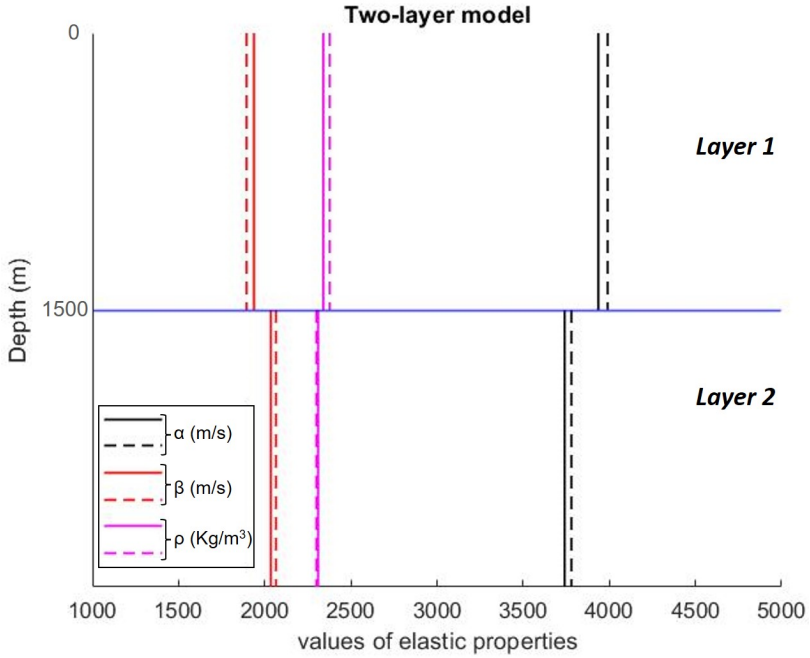

FIG. 2. True (continuous lines) and initial model (dashed lines) used for the nonlinear pre-stack inversion.

the appropriate step length per iteration, starting with α=1. Moreover, for the Levenberg-Marquardt method, the initial value of the damping parameter λ was set equal to 40. The performance of the optimization problems was evaluated through plots of the model values and data misfits per iteration. As applied by Liu and Liu (2016), the data misfit was

computed from the RMS error between the observed and predicted reflectivities:

$$E = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\Big(\mathbf{d}_{\mathbf{pred}}(\theta_j) - \mathbf{d}_{\mathbf{obs}}(\theta_j)\Big)^2} \tag{20}$$

Additionally, the weighted stacking inversion was performed to estimate values of $\Delta I/I$ and $\Delta J/J$. However, for both types of inversion, values of the fractional Vp/Vs ratio ($\Delta q/q$) were computed through:

$$\frac{\Delta q}{q} = \frac{\Delta I}{I} - \frac{\Delta J}{J} \tag{21}$$

Estimations were evaluated with accuracy tests, applying:

$$\%\ \text{error} = \frac{\left.\frac{\Delta(\text{I,J,q or }\rho)}{(\text{I,J,q or }\rho)}\right|_{\text{CALCULATED}} - \left.\frac{\Delta(\text{I,J,q or }\rho)}{(\text{I,J,q or }\rho)}\right|_{\text{TRUE}}}{\left.\frac{\Delta(\text{I,J,q or }\rho)}{(\text{I,J,q or }\rho)}\right|_{\text{TRUE}}} \times 100 \tag{22}$$

Broadband and noise free reflectivities

When inverting these reflectivity sets, all the optimization methods reached convergence at a different iteration number, as can be seen in Figure 3. Pure Gauss-Newton had the fastest convergence rate, finding invariable values at the second iteration. However, Steepest Descent and Levenberg-Marquardt had a slower performance, finding the minimum point at the 17 and 10 iteration, respectively. In addition, the estimated parameters were very close to the true values with almost zero data residuals. Particularly for Steepest Descent, the RMS error of P-P reflectivities increased in early iterations, demonstrating to be somewhat more unstable than the other methods. Nonlinear and linearized inversions are compared in Figure 4. Results from all the nonlinear algorithms got trapped into the same local minimum. The fractional impedances, as well as the fractional Vp/Vs ratio were more accurately estimated than the fractional density. Furthermore, the performance of the optimization methods was significantly superior than the weighted stacking inversion, evidencing differences of around 2.5% between both approaches and allowing the estimation of density values.

Broadband and noisy reflectivities

Identical and independently distributed (IID) noise was randomly generated from a Gaussian distribution about zero with a standard deviation computed according to an indicated signal-to-noise ratio (SNR) and the RMS power of the reflection coefficients. Subsequently, this noise was added to the synthetic reflectivities, with a SNR of 8 for the P-P dataset and 4 for the P-S dataset. Since the observed data can be treated as a random variable, because measurements always contain random noise (Zhdanov, 2002), each nonlinear and linearized inversion was repeated 5000 times to analyze the results statistically, comparing the maximum likelihood solutions of both types of inversion. Moreover, each

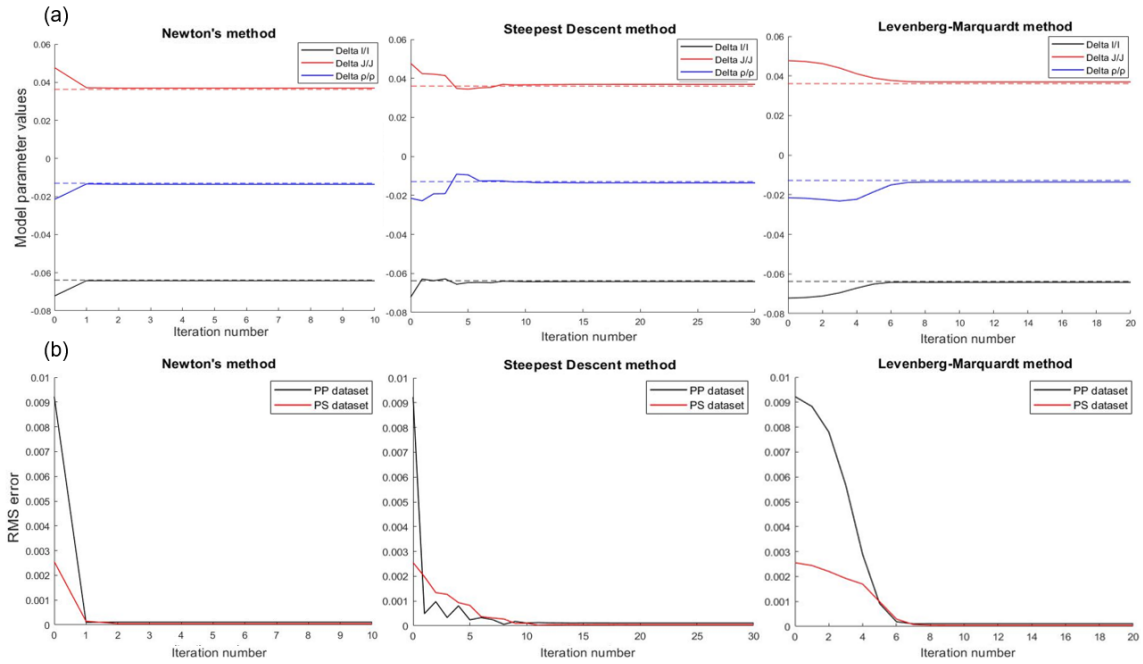

FIG. 3. Convergence of optimization problems after inverting broadband and noise free reflectivities. In Figure (a) dashed lines are the true values.

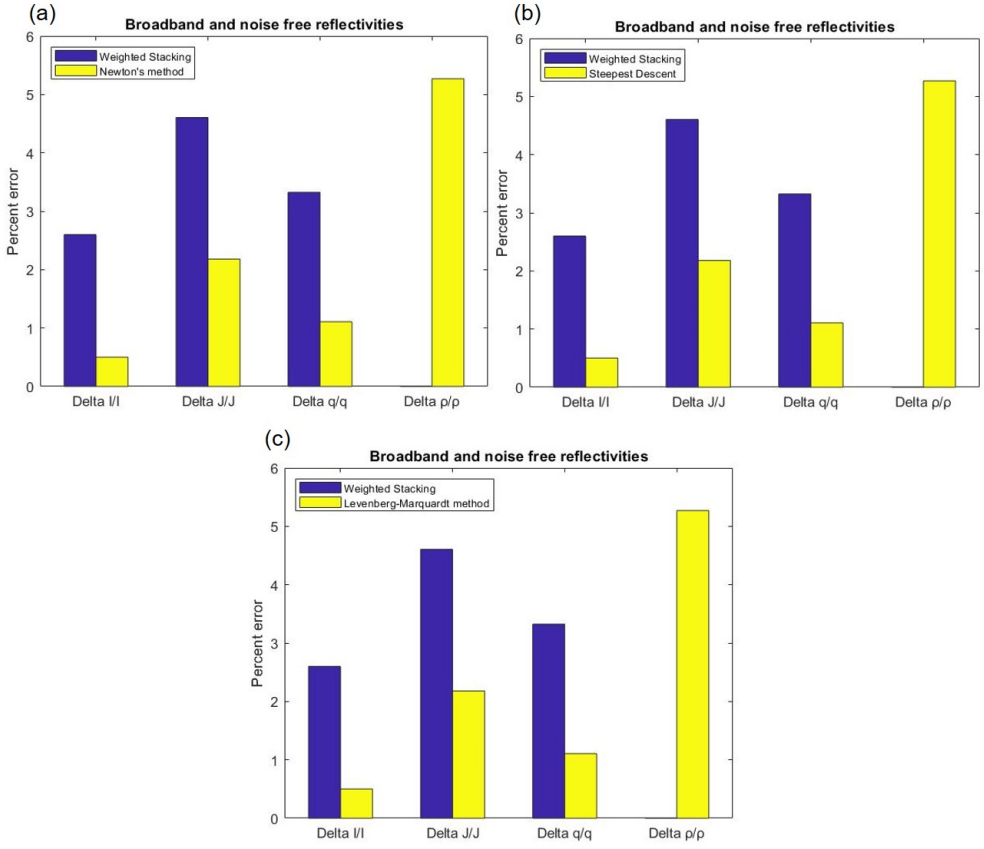

FIG. 4. Comparisons between nonlinear and linearized inversion results after using broadband and noise free reflectivities.

optimization was done with and without the data covariance matrix in order to understand the effects after considering noise information.

Particularly, the $\mathbf{W}^{\intercal}\mathbf{W}$ matrix assigned more importance to the P-S reflectivities, because although both standard deviations were of the order of magnitude 10^{-3}, σ_{ps} was smaller than σ_{pp}. However, these variances actually indicated that the P-S data was noisier than the P-P data, because the order of magnitude of both synthetic measurements was different (10^{-2} for P-P data and 10^{-3} to 10^{-5} for P-S data). Hence, it is expected that these values of quality help to improve the estimation of those parameters mainly affected by R_{ps}.

The accuracy of the maximum likelihood solutions is studied in Figure 5. In general, results from each optimization were almost identical, estimations resembled to those of the noise free case, and the fractional density had the highest percent error of the entire set. Although the linearized inversion helps to suppress noise due the nature of the stacking procedure (Larsen, 1999), the nonlinear approach did a better estimation with and without using the data covariance matrix. On the other hand, when including noise information, the error of $\Delta I/I$ and $\Delta\rho/\rho$ slightly increased, while the error of $\Delta J/J$ and $\Delta q/q$ had a subtle decrease, showing that, effectively, the improved parameters were those mainly influenced by the P-S reflectivities. Thus, noise information positively impacts the estimation of some model parameters, promising a more accurate reservoir characterization.

Band-limited and noisy reflectivities

Both synthetic noise free P-P and P-S datasets were filtered using an Ormsby wavelet but applying a different range of frequencies depending on the type of data. For the P-P reflectivities the set of frequencies used was 5-10-60-75 Hz, and for the P-S dataset 5-10-40-55 Hz. The NMO corrected synthetic CMP/CCP gathers are shown in Figure 6. Later, each band-limited dataset was contaminated with IID random noise assigning the same characteristics of the previous section. Each optimization was statistically performed with 5000 samples and the evaluation of the estimates was done using equation 22 and considering the band-limited true values. Figure 7 illustrates the accuracy tests related to each maximum likelihood solution. Once more, the nonlinear inversion did a better estimation than the linearized approach for most model parameters. Nevertheless, only for $\Delta I/I$ the optimization methods yielded to a slightly more erroneous solution.

Additionally, it was observed similar results between optimization methods and improvements brought by the $\mathbf{W}^{\intercal}\mathbf{W}$ matrix over $\Delta J/J$ and $\Delta q/q$. These enhancements were more notorious than those of the previous section, passing from errors of around 9% to 5.5% for $\Delta J/J$ and from 4.5% to 3.5% for $\Delta q/q$. On the other hand, the accuracy of all estimates was smaller than the corresponding to the broadband reflectivities, but for most elastic parameters, the errors stayed within an acceptable range. For instance, when the $\mathbf{W}^{\intercal}\mathbf{W}$ matrix was not included, errors were smaller than 10% and smaller than 6% when it was. Nevertheless, density exhibited a dramatic percent error, reaching a value of around 57% when the data covariance matrix was not considered, and approximately 67% when it was. Therefore, P-P and P-S datasets have little sensitivity to the density, not doing a proper constraint of its band-limited true value.

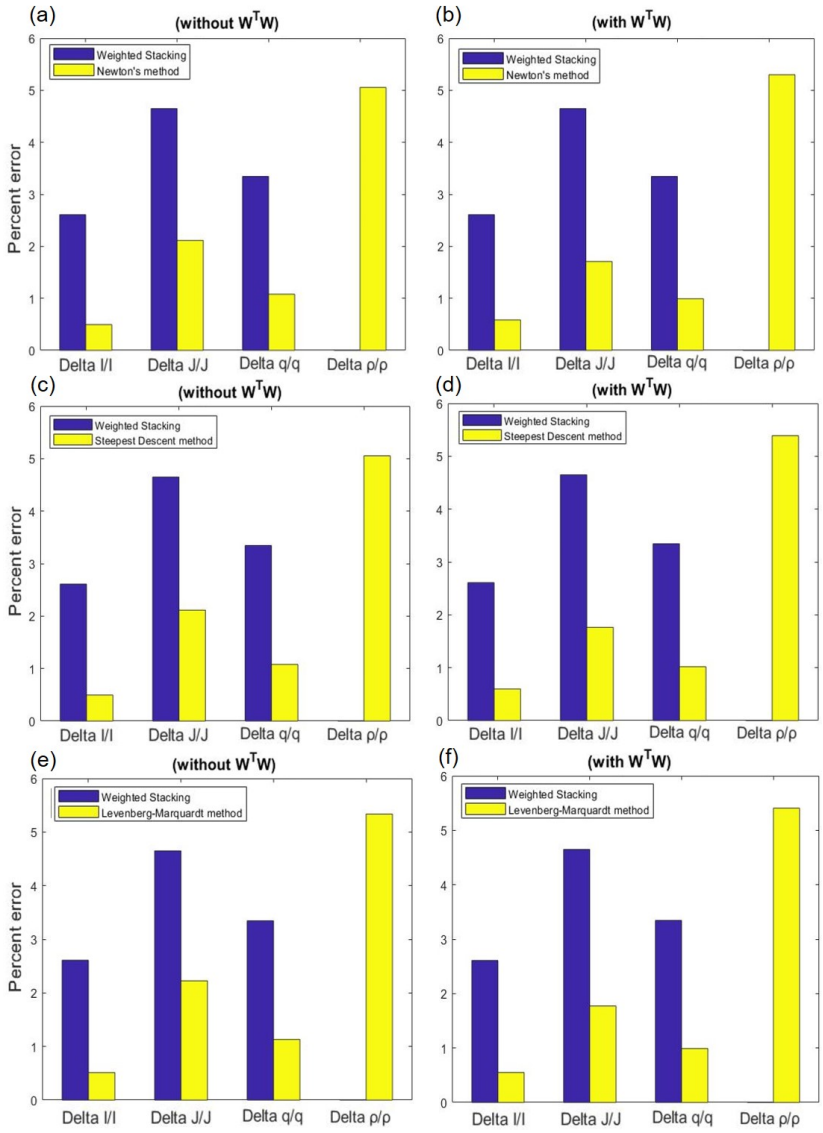

FIG. 5. Nonlinear and linearized inversion results after using broadband and noisy reflectivities.

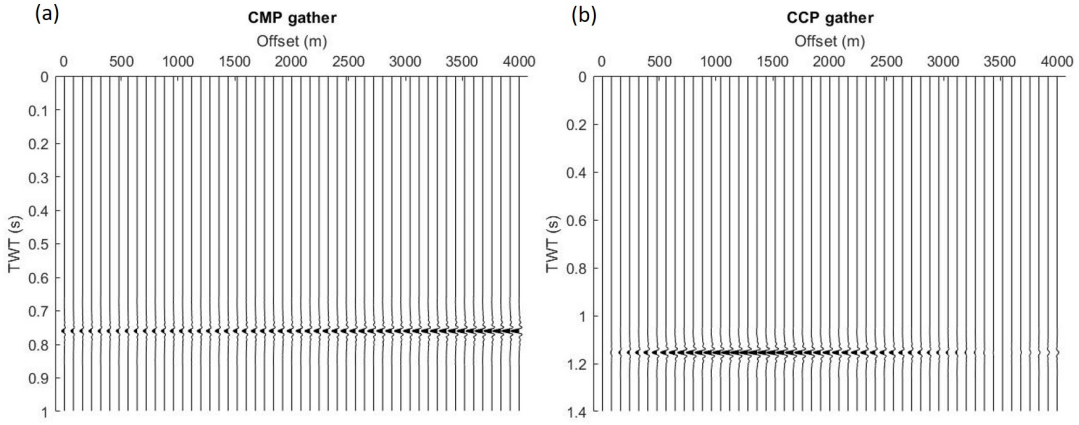

FIG. 6. Syntethic seismograms for the a) P-P dataset and b) P-S dataset.

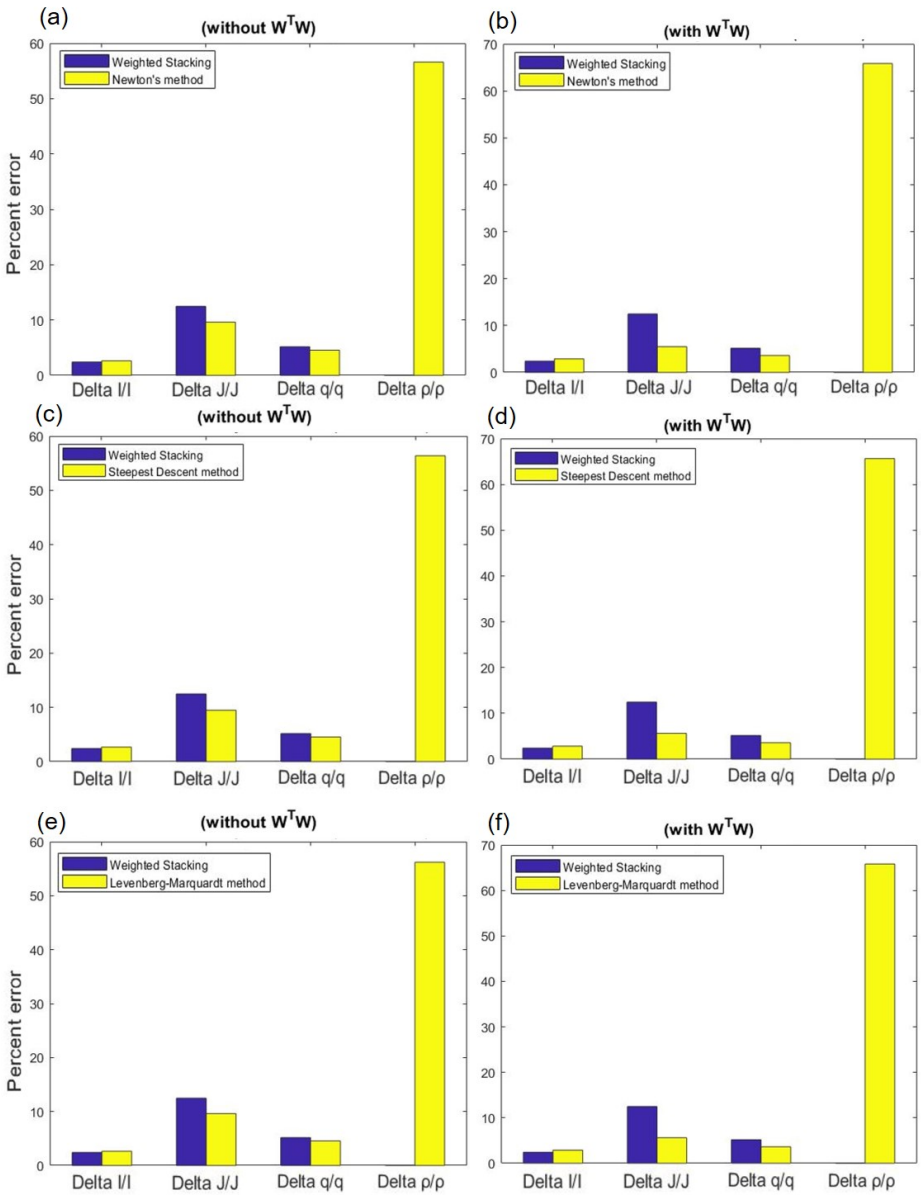

FIG. 7. Nonlinear and linearized inversion results after using band-limited and noisy reflectivities.

EFFECTS OF THE INITIAL MODEL

To recognize the impact of initial models in the results, only one initial parameter was strongly perturbed and the others were left invariable respect the initial model of Figure 2. The magnitude of the perturbations was comparable between model parameters. These effects were studied with the Pure Gauss-Newton method on broadband and noise free data and results were contrasted to those of Figure 4a. Figure 8a indicates that after only modifying the initial α, errors increased around 0.75% for $\Delta I/I$, 4.3% for $\Delta J/J$ and 10.8% for $\Delta\rho/\rho$. Figure 8b shows the effects after changing the initial β; all percent errors increased by almost 1.5 times the result from Figure 8a. However, after varying the initial ρ, no difference was observed between Figure 8c and Figure 4a. This suggests that initial values of α and β mainly affect the estimations, and higher discrepancies of the initial β

respect the true model produce less accurate results.

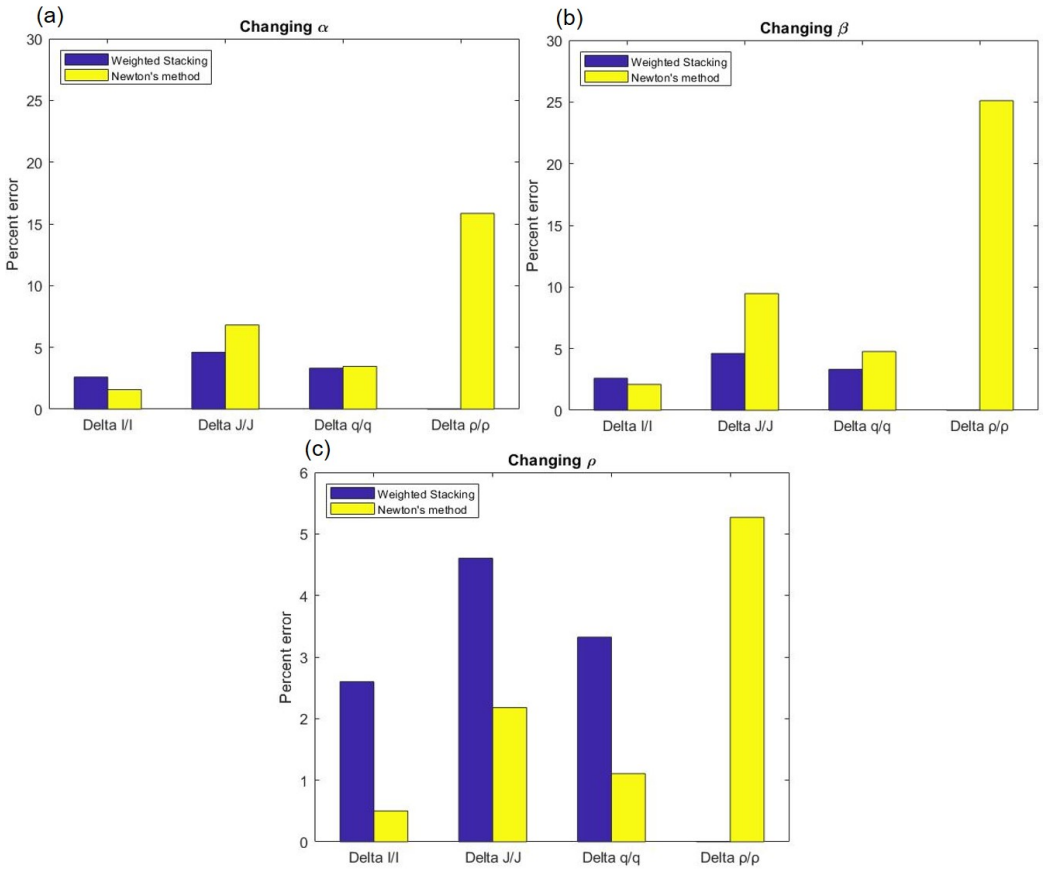

FIG. 8. Accuracy tests after changing only one initial model parameter for the Pure Gauss-Newton method.

A good initial model is necessary in order the nonlinear inversion outperforms the linearized approach. This finding is supported in the study presented by Liu and Liu (2016), where numerical tests of P-P inversion based on Zoeppritz equations demonstrated that effectively the perturbation level of the initial models as well as the noise level in the observed data strongly impact the convergence of the local optimization algorithm. To understand how good the initial model needs to be so the outperformance of the nonlinear inversion could be guaranteed, several initial models were constructed with different relationships between the P-wave and S-wave velocity perturbations, as well as features of the elastic contrasts. Figure 9 shows that one of the most important characteristics to be met is the closeness of the initial elastic contrasts to the true contrasts to obtain remarkably low or adequate percent errors. These contrasts need to be the closest possible to reality, since differences of even 100 units could yield to failure of the method respect the linearized inversion for some model parameters, as illustrated in Figure 9f. However, establishing an appropriate initial model is complex since it has to satisfy some relationships between the initial perturbations of α and especially β.

Figure 9a and Figure 9b demonstrate that if the perturbation of α doubled the perturbation of β, the percent errors for all the model parameters were almost zero. Moreover, other acceptable relations for α:β were 1.5:1 and 3:1, as observed in Figures 9c and 9d.

However, Figure 9e indicates that if the perturbation of α is 6 times the perturbation of β, the percent errors of the fractional shear impedance and fractional density significantly increase. Hence, the initial value of β and its relationship with α has a strong influence in the converge of the algorithm. Moreover, the compliance of these relationships allowed to supply initial values for α and ρ with even 300 units of perturbation respect their true values, as seen in Figure 9a. Finally, if α and ρ are close to the true values but β is strongly perturbated, then the estimations would not be very accurate.

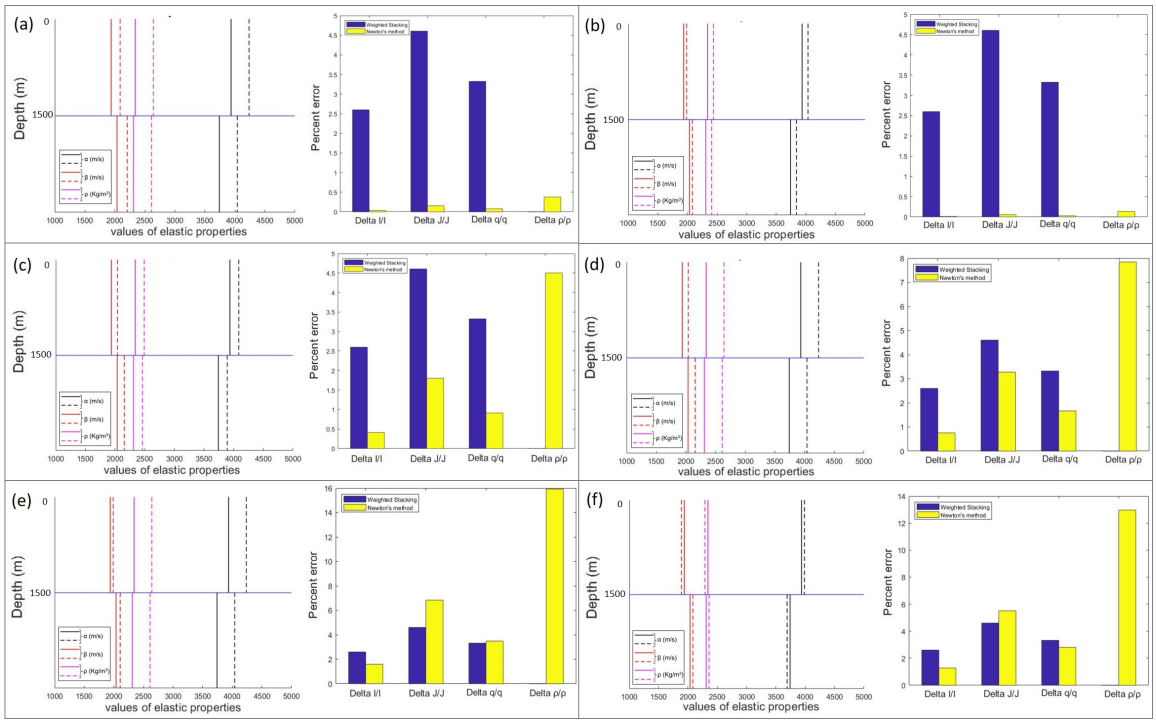

FIG. 9. Different initial models and the resulting accuracy tests. Elastic contrasts close to reality along with an α:β relation of (a) 2:1, (b) 2:1, (c) 1.5:1, (d) 3:1, (e) 6:1. Figure (f) shows elastic contrasts with 100 units of difference respect the true value.

ADVANTAGES OF THE SIMULTANEOUS NONLINEAR INVERSION

The benefits of incorporating the P-S reflectivities in the inversion process can be understood after comparing the joint P-P and P-S inversion results against the ones obtained with the conventional P-P inversion. To achieve this, the Pure Gauss-Newton algorithm was used for the P-P inversion, considering only the P-P incidence angles for the computation of the forward model and Jacobian matrix, and modifying the sampling operator **S** according to:

$$d_{\text{pred}} = \mathbf{Su} = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} R_{pp} \\ R_{ps} \\ T_{pp} \\ T_{ps} \end{bmatrix} = R_{pp} \tag{23}$$

In this case, the condition number of the Hessian matrix was approximately 300, while the condition number calculated for the simultaneous technique was approximately 24. Thus, similar to the analysis shown by Larsen (1999), including P-S reflectivities helped to stabilize the inversion. Moreover, Figure 10 illustrates that the conventional P-P approach had an extraordinary performance for $\Delta I/I$ and $\Delta\rho/\rho$ when inverting a broadband and noise free dataset, estimating these parameters with an almost perfect accuracy. However, the estimation of $\Delta J/J$ and $\Delta q/q$ had higher percent errors than the simultaneous inversion. In addition, as noise and band-limitations were included, the method had a worse behavior than the simultaneous technique, exhibiting errors higher than 50% for $\Delta J/J$ and $\Delta\rho/\rho$ in Figure 10c, but always producing a very accurate result for $\Delta I/I$.

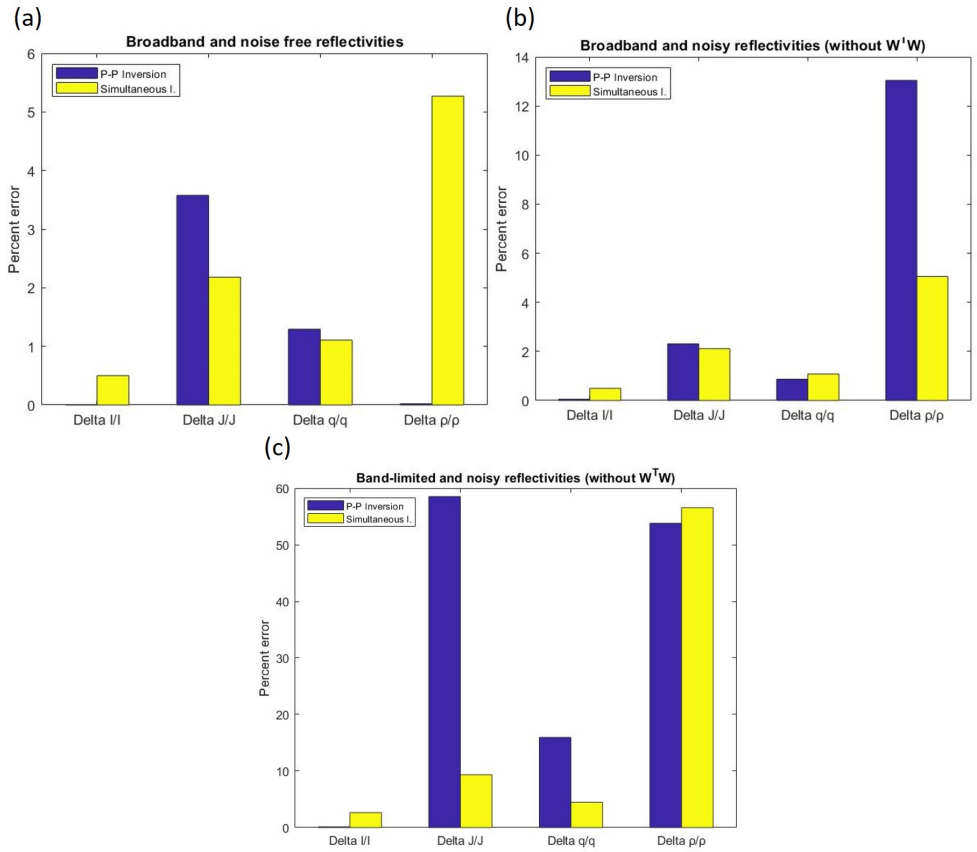

FIG. 10. Comparison of accuracy tests obtained from reflectivities with different characteristics of noise and frequency content.

Figure 11 shows normalized histograms constructed with model estimates obtained after statistically inverting band-limited and noisy datasets with the conventional P-P and the simultaneous inversion. It is observed that although the P-P inversion generated an almost perfect maximum likelihood solution for $\Delta I/I$, this method did not have an adequate performance for the remaining model parameters, exhibiting a lack of accuracy and precision. Nevertheless, the simultaneous inversion did a better job overall because although it sacrificed a bit of accuracy for $\Delta I/I$, the maximum likelihood solution of $\Delta J/J$ and $\Delta q/q$ had a much smaller percent error, and estimations of all the model parameters were much more precise, hence there is more probability to compute better estimates applying this approach. Unfortunately, in this scenario, both types of inversions could not generate a good

maximum likelihood solution for the fractional density, but statistically the joint P-P and P-S inversion produced a more precise result.

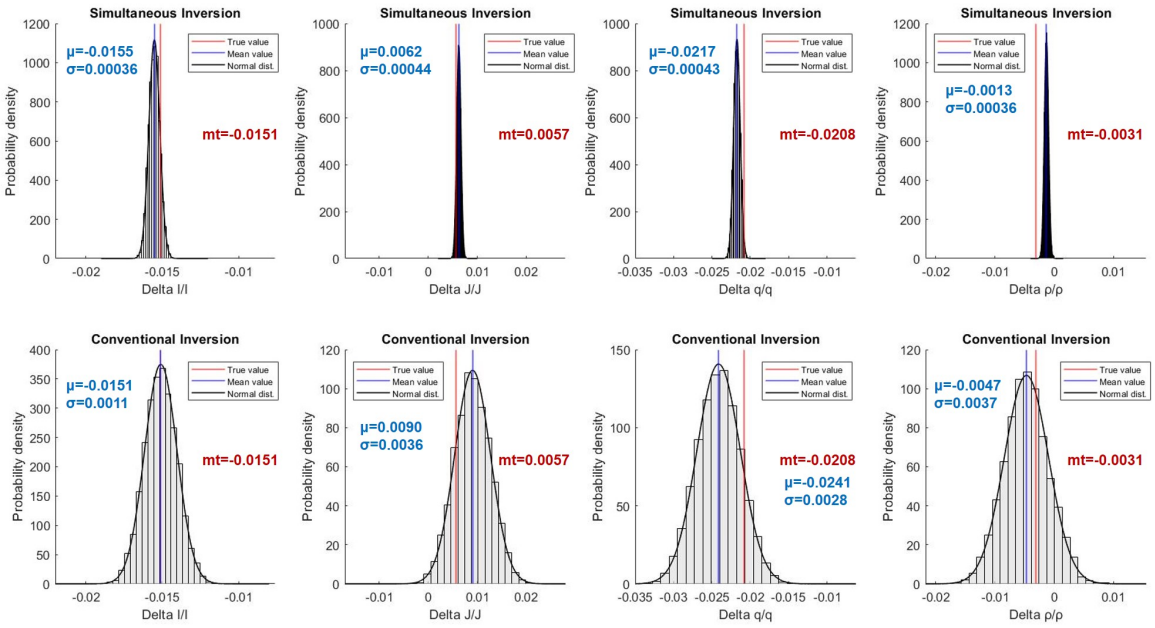

FIG. 11. Posterior uncertainty distributions of each estimate obtained from band-limited and noisy reflectivities; “mt” corresponds to the band-limited true fractional value.

CONCLUSIONS

Synthetic data allowed to confirm the outperformance of the simultaneous nonlinear AVO inversion over the weighted stacking approach under scenarios of long-offsets. Conditioned to a sufficiently good initial model, estimations of the fractional compressional and shear impedances, as well as the fractional Vp/Vs ratio, from local optimizations, were very accurate with remarkably smaller percent errors. Additionally, regardless the optimization algorithm applied, convergence was reached to the same minimum point, but exhibiting different convergence features, being Pure Gauss-Newton the fastest method. On the other hand, the inclusion of noise information through a data covariance matrix led to improve the estimation of the parameters $\Delta J/J$ and $\Delta q/q$, especially when the datasets were band-limited, which is beneficial for reservoir characterization. Moreover, advantages of applying a simultaneous nonlinear inversion over a conventional P-P technique were demonstrated by the high accuracy and precision of the results. Finally, the nonlinear inversion had a negative impact when estimating $\Delta\rho/\rho$ from band-limited reflectivities, indicating the inadequacy of the dataset to produce a proper constraint to the right answer. Hence, in future work, it is important to determine improvements for a better inversion of this elastic parameter.

ACKNOWLEDGMENTS

The authors would like to thank the sponsors of the CREWES project as well NSERC (Natural Science and Engineering Research Council of Canada) under the grant CRDPJ 543578-19 for making this work possible through their financial support.

REFERENCES

Downton, J. E., Russell, B. H., and Lines, L. R., 2000, AVO for managers: pitfalls and solutions: CREWES Research Report, **12**, 46.1–46.21.

Downton, J. E., and Ursenbach, C., 2005, Linearized AVO inversion with supercritical angles: SEG Technical Program Expanded Abstracts 2005.

Eriksson, J., 1996, Optimization and regularization of nonlinear least squares problems: Ph.D. thesis, Umea University.

Grossman, J. P., 2003, AVO and AVA inversion challenges: a conceptual overview: CREWES Research Report, **15**, 44.1–44.7.

Larsen, J. A., 1999, AVO analysis by simultaneous P-P and P-S weighted stacking: M.Sc. thesis, University of Calgary.

Liu, Y., and Liu, X., 2016, Application and performances of unconstrained optimization methods in seafloor AVO inversion: Arabian Journal of Geosciences, **9**, 1–7.

Lu, J., Wang, Y., Chen, J., and An, Y., 2018, Joint anisotropic AVO inversion of PP and PS seismic data: Geophysics, **83**, No. 2, N31–N50.

Luenberger, D. G., and Ye, Y., 2008, Linear and Nonlinear Programming: Springer.

Nocedal, J., and Wright, S. J., 2006, Numerical Optimization: Springer.

Sen, M. K., and Stoffa, P. L., 2020, Inverse theory, global optimization, *in* Gupta, H. K., Ed., Encyclopedia of Solid Earth Geophysics: Springer Netherlandss, 625–632.

Skopintseva, L., Ayzenberg, M., Landrø, M., Nefedkina, T., and Aizenberg, A. M., 2011, Long-offset AVO inversion of PP reflections from plane interfaces using effective reflection coefficients: Geophysics, **76**, No. 6, C65–C79.

Smith, G. C., and Gidlow, P. M., 1987, Weighted stacking for rock property estimation and detection of gas: Geophys. Prosp, **35**, 993–1014.

Stewart, R. R., 1990, Joint P and P-SV inversion: CREWES Research Report, **2**.

Tolle, J. W., 2003, Nonlinear programming, *in* Meyers, R. A., Ed., Encyclopedia of Physical Science and Technology: Academic Press, 583–596.

Williams, R. G., Roberts, G., and Hawkins, K., 2001, Long offset towed streamer recording - a cheaper alternative to multi-component OBC for exploration?: Exploration Geophysics, **32**, 316–319.

Zhdanov, M., 2002, Geophysical Inverse Theory and Regularization Problems: Elsevier Science.

Zou, X., 2020, Atmospheric Satellite Observations: Academic Press.