Model misspecification and bias in the least-squares algorithm: Implications for linearized isotropic AVO

Henning Hoeber[1]

https://doi.org/10.1190/tle40090646.1

Abstract

When inversions use incorrectly specified models, the estimated least-squares model parameters are biased. Their expected values are not the true underlying quantitative parameters being estimated. This means the least-squares model parameters cannot be compared to the equivalent values from forward modeling. In addition, the bias propagates into other quantities, such as elastic reflectivities in amplitude variation with offset (AVO) analysis. I give an outline of the framework to analyze bias, provided by the theory of omitted variable bias (OVB). I use OVB to calculate exactly the bias due to model misspecification in linearized isotropic two-term AVO. The resulting equations can be used to forward model unbiased AVO quantities, using the least-squares fit results, the weights given by OVB analysis, and the omitted variables. I show how uncertainty due to bias propagates into derived quantities, such as the χ-angle and elastic reflectivity expressions. The result can be used to build tables of unique relative rock property relationships for any AVO model, which replace the unbiased, forward-model results.

Introduction

Many geophysical inversion workflows rely on a comparison of estimated model parameters to their forward-model equivalents. Such workflows assume that the inverted model parameters are *unbiased* estimates of the model values. This means that were we to average an increasing number of repeated measurements, the averages of their model parameters would converge to the true model values. A key assumption for this to hold in least-squares theory *with nonorthogonal basis functions* is the correct specification of the model in the inversion. *Model misspecificatio*n, such as when fitting a linear model to data with quadratic variability, results in biased model parameter estimates. In such cases, the least-squares algorithm is able to find a better fit to the data, and a lower least-squares value, by adjusting the model parameters away from their unbiased values. Model misspecification is common in linearized amplitude variation with offset (AVO) (Causse et al., 2007; Ball et al., 2014b; Thomas et al., 2016), where we often fit the data to linearized and simplified versions of the Knott-Zoeppritz equations (Knott, 1888; Zoeppritz, 1919)[2]. First, there is the linearization error itself, which is introduced by assuming small contrasts across the boundaries. Second, much of AVO fitting is performed with truncated models where only the first two dominant terms of the linearized AVO equations are modeled. Linearized AVO of prestack data has limited information content, which means that estimates of three or more terms can be numerically unstable (de Nicolao et al., 1993; Ursin and Tjaland, 1993). In addition, the recorded data at larger angles are less reliable due to processing and imaging issues (Connolly, 2017). Different linearized versions of the Knott-Zoeppritz equations, such as those commonly called the Shuey (Wiggins et al., 1983; Shuey, 1985) and Fatti (Gidlow et al., 1992; Fatti et al., 1994) two-term AVO equations, have different third terms so that their truncation errors due to the omission of this term are different. Within linear least squares, an analysis of bias due to model misspecification exists; it is called omitted variable bias (OVB) (Greene, 2003). It allows us to analytically forward model bias due to the omission of variables based on the known and data-independent design matrix $\boldsymbol{A}$ of the fully specified linear model, and the values of the omitted variables. The theory is general, but I could find no mention of it in the geophysical literature. Here, I show how to calculate exactly the bias due to model misspecification in linearized isotropic two-term AVO. I base the analysis on the pseudo-quadratic expansion of the Knott-Zoeppritz equation due to Wang (1999) and Mallick (1993), which adds a single higher-order term to the three-term Aki-Richards expression of linearized AVO (Aki and Richards, 1980). This allows us to see the relative importance of two omitted variables: the respective third terms and the quadratic term. Extensions to other cases are easily derived from the general theory. I show how to use the analytic bias expressions to model the systematic uncertainty in least-squares two-term AVO parameters and how this uncertainty propagates into derived quantities, such as the χ-angle and elastic reflectivity expressions (Ball et al., 2014a; Connolly, 2019).

Key concepts using a toy model

Consider an AVO toy model[3] in which the reflectivity data as a function of incidence angle are described by a two-term Shuey model (Wiggins et al., 1983; Shuey, 1985), linear in the $\sin^2(\theta)$ values: $R(\theta) = R(0) + \sin^2(\theta)\, G$. Assume that the acquired data follow this model but with added Gaussian homoskedastic noise. A least-squares fit of the data to the two-term model $d = m_0 + \sin^2(\theta) m_1$ gives unbiased estimates $\hat{m}_0 = \hat{R}(0)$ and $\hat{m}_1 = \hat{G}$ of the model Shuey intercept $R(0)$ and model Shuey gradient G, with the variance of the estimated model parameters dependent on the measurement noise (Aster et al., 2005; Menke, 2012). As is customary, I put a hat on the least-squares estimates of the model parameters to distinguish them from their forward-model equivalents. If we were able to repeat the experiment many times, yielding many noisy realizations of the data (as in a Monte Carlo simulation of

---

[1]CGG, Crawley, UK. E-mail: henning.hoeber@cgg.com.
[2]The authors of Ball et al. (2014b) call this truncation bias, but I find this term is used differently in the literature (Greene, 2003).
[3]Toy model: A simple model, with details removed, to help explain the key concepts of the general theory.

the data), we would find that the average values of the fit parameters to the noisy data agree with the true model values.

Suppose now we misspecify the model and fit a single term to the data, $d = m_0$. We may hope to obtain an unbiased estimate of the two-term model Shuey intercept $\hat{m}_0 = \hat{R}(0)$ since $R(0)$ is the first term in the two-term forward model. But this is not what happens in least-squares fitting or inversion. The cost function of the least-squares algorithm adjusts the remaining model parameter to minimize the misfit of the predicted *data model* and the input data. A much smaller overall misfit is obtained by setting $\hat{m}_0$ to the average data value. The result of the one-term fit is therefore an estimate $\hat{m}_0 = \hat{S}$ of the model stack S, and hence a mix of the Shuey intercept and the Shuey gradient, where the gradient term is weighted with the average $\langle \sin^2(\theta) \rangle$ of the $\sin^2(\theta)$ values. The omitted variable in the one-term model, the gradient, leaks into the estimate of the intercept $\hat{m}_0$. Comparing the biased one-term intercept $\hat{m}_0$ to a forward-modeled Shuey intercept $R(0)$ is a category error (Causse et al., 2007; Thomas et al., 2016), in effect a comparison of apples and oranges. The one-to-one correspondence between the two fit parameters $\hat{m}_0$ and $\hat{m}_1$ and the two Shuey model parameters $R(0)$ and G only holds when the data variability is fully described by the model used.

The bias of the one-term intercept estimate $\hat{m}_0$ is the product of a data-independent weight term $\langle \sin^2(\theta) \rangle$ and the omitted variable G. This structure always holds for least-squares model parameters in misspecified models:

$$\begin{aligned} &\text{biased variable} = \text{unbiased variable} + \text{bias} \\ &\text{bias} = \text{data independent weight term} \times \text{omitted variable(s)}. \end{aligned} \tag{1}$$

In OVB, the least-squares fit result in the reduced model is rewritten in terms of the least-squares fit result in the full, unbiased model, as in equation 1. This means the notation needs to track the order of the fit for each least-squares model parameter. For example, in the toy model, we saw that the one-term and two-term least-squares intercept estimates are $\hat{m}_0^{(1t)} = \hat{S}$ and $\hat{m}_0^{(2t)} = \hat{R}(0)$. The superscripts $1t$ and $2t$ on the left-hand side denote the order of the fit. The terms on the right-hand side of these equations, without reference to the order of the fit, and written in AVO Shuey notation, imply that these are unbiased estimates. One-term AVO returns an unbiased estimate of the stack, and two-term AVO, in our toy model where the full data variability is given by two terms, returns unbiased estimates of the Shuey intercept $\hat{R}(0)$ and the gradient $\hat{m}_1^{(2t)} = \hat{G}$.

Combining those two results, we can write:

$$\begin{aligned} \hat{m}_0^{(1t)} &= \hat{S} \\ &= \hat{R}(0) + \text{bias} \\ &= \hat{m}_0^{(2t)} + \langle \sin^2(\theta) \rangle \hat{G}. \end{aligned} \tag{2}$$

In the presence of bias, the one-term least-squares estimate of the intercept $\hat{m}_0^{(1t)}$ cannot be used in subsequent analysis, for example using a rock-physics template, as if it were an estimate of the Shuey model intercept $\hat{R}(0)$. Thomas et al. (2016) formulated this slightly differently: "Quantitative relationships between true elastic properties are not always applicable to inverted elastic properties." This is due to bias, in effect a misinterpretation of the least-squares model parameters in a misspecified model.

We can further use the toy model to explore strategies for obtaining less biased or even unbiased least-squares model parameters. First, we may use less of the data and fit the one-term model only to that part of the data where we think it best specifies the data. This keeps the data-independent weight in the bias term small. The price we pay for this strategy is a much larger variance as a trade-off for less bias. OVB tells us precisely the relation between the systematic error due to OVB and the maximum fit angle.

An alternative strategy is to reparameterize the model in *orthogonal variables*. In orthogonal basis systems, the bias weights are zero. None of the model parameter estimates exert any influence over the others, and we can fit each term independently of the others. In seismic processing, we make use of this whenever we stack the data: the stack is orthogonal to all higher-order terms and hence unbiased. All the basis functions for all the included and omitted variables must be orthogonal for the least-squares estimates of the included model parameters of the misspecified model to be unbiased. For example, fitting two-term AVO with stack and gradient as model parameters is not sufficient to remove gradient bias due to the curvature component because gradient and curvature are not orthogonal variables, even if stack and gradient as well as stack and curvature are.

A third strategy consists of finding AVO models in which the omitted variable, rather than the bias weight, is small. For example, the Fatti AVO parameterization has a density contrast as the third term, whereas in the Shuey AVO model the third term relates to the P-wave velocity contrast. This suggests that the two-term Fatti model has smaller bias than the two-term Shuey model corresponding to the same three-term linearized AVO.

Lastly, consider the outcome of the two one-term regressions in the toy model with respect to the quality control (QC) we use to investigate the efficacy of the least-squares fitting. A standard QC is given by a plot of the residuals versus the offsets or angles. Both one-term parameterizations yield an estimate of the data average, so the fit residuals of these models are identical. The residual plots show that neither of the one-term models fully represent the data variability since the residuals correlate as a function of incidence angle. But the residual QC does not tell us if the model parameter in one or the other model is biased. The residuals only tell us if the data variability is explained with the model parameter $\hat{m}_0^{(1t)}$ used.

To recap the key findings, the least-squares algorithm (fitting or inversion) with nonorthogonal basis functions only gives unbiased model parameters when the model fully explains the data variability. When the model is misspecified, the least-squares algorithm finds an optimal fit of the data to the erroneous model by adjusting, and hence biasing, the reduced number of model fit parameters. Fit QCs based on the mismatch of the data to the model only tell us if the model parameters explain the data variability and not if the parameters are biased. Bias propagates into subsequent analysis; therefore, the possible systematic error due to bias must be understood and accounted for. To this end, I now develop OVB in the context of linearized AVO fitting.

Bias in linearized AVO

We start with a linearized version of the Knott-Zoeppritz equations due to Aki and Richards (1980). Following Wiggins et al. (1983) and Shuey (1985) it can be rearranged to emphasize angle ranges:

$$R_{\text{Shuey}}(\theta)=\left\{\frac{d\rho}{2\rho}+\frac{dV_{\text{P}}}{2V_{\text{P}}}\right\}+\left\{\frac{d\rho}{2\rho}-4\frac{V_{\text{S}}^2}{V_{\text{P}}^2}\left(2\frac{dV_{\text{S}}}{2V_{\text{S}}}+\frac{d\rho}{2\rho}\right)\right\}\sin^2(\theta)+\frac{dV_{\text{P}}}{2V_{\text{P}}}\sin^2(\theta)\tan^2(\theta). \quad (3)$$

I will call this the Shuey approximation. Another rearrangement due to Gidlow et al. (1992) and Fatti et al. (1994) is:

$$R_{\text{Fatti}}(\theta)=\left(1+\tan^2(\theta)\right)\frac{dI_P}{2I_P}-8\frac{V_{\text{S}}^2}{V_{\text{P}}^2}\frac{dI_S}{2I_S}\sin^2(\theta)-\left\{\tan^2(\theta)-4\frac{V_{\text{S}}^2}{V_{\text{P}}^2}\sin^2(\theta)\right\}\frac{d\rho}{2\rho}\ . \quad (4)$$

The Shuey parameterization is more often written using the three reflectivity shortcuts

$$\begin{aligned}
R(0)&=\left\{\frac{d\rho}{2\rho}+\frac{dV_{\text{P}}}{2V_{\text{P}}}\right\}=R_\rho+R_{V_{\text{P}}}=R_{I_P}\ ,\\
G&=\frac{dV_{\text{P}}}{2V_{\text{P}}}-2\gamma^2\left\{\frac{d\rho}{\rho}+2\frac{dV_{\text{S}}}{V_{\text{S}}}\right\}=R_{V_{\text{P}}}-4\gamma^2\left(2R_{V_{\text{S}}}+R_\rho\right)=R_{I_P}-8\gamma^2R_{I_S}+R_\rho\left(4\gamma^2-1\right),\\
C&=\frac{dV_{\text{P}}}{2V_{\text{P}}}=R_{V_{\text{P}}}=R_{I_P}-R_\rho,
\end{aligned} \quad (5)$$

where $\gamma = V_{\text{S}}/V_{\text{P}}$ and for any variable x with values x_{uppper} and x_{lower} above and below the reflecting interface

$$\begin{aligned}
dx&=x_{\text{lower}}-x_{\text{upper}},\\
x&=\frac{x_{\text{lower}}+x_{\text{upper}}}{2},\\
R_{x=(V_{\text{P}},V_{\text{S}},\rho)}&=\frac{dx}{2x}\ .
\end{aligned} \quad (6)$$

When working only to lowest order, the contrasts of all elastic parameters are assumed small relative to their averages across the boundary. Wang (1999) and Mallick (1993) showed that at the next highest order a pseudo-quartic expansion of the Knott-Zoeppritz equations adds a single term:

$$\gamma^3\sin^2(\theta)\cos(\theta)\left\{\frac{d\rho}{\rho}+2\frac{dV_{\text{S}}}{V_{\text{S}}}\right\}^2\ . \quad (7)$$

For example, the linearized Shuey AVO model becomes:

$$R_{\text{Wang-Mallick}}(\theta)=R(0)+G\ \sin^2(\theta)+C\ \sin^2(\theta)\tan^2(\theta)+\frac{1}{4\gamma}(C-G)^2\ \sin^2(\theta)\cos(\theta). \quad (8)$$

At both orders three and four, the AVO models of Shuey and Fatti are identical.

Figure 1 shows the full Knott-Zoeppritz model as well as the Shuey and Fatti models at orders two, three, and four. An efficient way to understand the model mismatch is to plot their respective residuals relative to the full Knott-Zoeppritz model, as shown in Figure 1b. In the development of OVB, where we require a linearized model, I often treat the Wang-Mallick model as the ground truth. I find the higher-order correction to be significant in many published AVO examples. The validity of models always needs to be verified with a full modeling analysis.

Now, as in the toy model, we consider what happens when we fit the data to misspecified models. We would like the least-squares fit at a given order to approximate the corresponding AVO model at that order. However, as seen in the toy model, this is not what the least-squares algorithm is designed to do. The least-squares algorithm attempts to match the data as best as possible, and it freely adjusts the available model parameters to do this.

Figure 2 shows this principle in action. In the first example (Figure 2a), the experimental data are the noise-free Wang-Mallick approximation of the Knott-Zoeppritz equation. In the second example, the model is the noise-free three-term Aki-Richards equation. In both cases, least-squares fitting with a two-term Shuey model gives good fits to the respective models[4]. This means that the model parameters predict data models for which they were not originally intended. In both cases, the fits are biased.

The least-squares algorithm estimates model parameters by projecting the data onto the Moore-Penrose generalized inverse (Aster et al., 2005; Menke, 2012)[5]:

$$\hat{\boldsymbol{m}}=\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{d}=\boldsymbol{A}^{-g}\boldsymbol{d}\ . \quad (9)$$

[4]Of course, a single AVO model proves little, hence I provide a Jupyter notebook with this paper.
[5]To keep the equations uncluttered, I will ignore the data noise term.

The only condition for a solution to exist is that the covariance matrix (A^TA) is invertible.

The matrix $\boldsymbol{A}(\boldsymbol{\theta})$ consists of column vectors of the basis functions with which we express the data vector $\boldsymbol{d}(\boldsymbol{\theta})$. For example, we can write the Wang-Mallick four-term AVO design matrix in Shuey notation as

$$A_{\text{Wang-Mallick}} = \begin{pmatrix} 1 & \sin^2(\theta_{\min}) & \sin^2(\theta_{\min})\tan^2(\theta_{\min}) & \sin^2(\theta_{\min})\cos(\theta_{\min}) \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ 1 & \sin^2(\theta_{\max}) & \sin^2(\theta_{\max})\tan^2(\theta_{\max}) & \sin^2(\theta_{\max})\cos(\theta_{\max}) \end{pmatrix}. \quad (10)$$

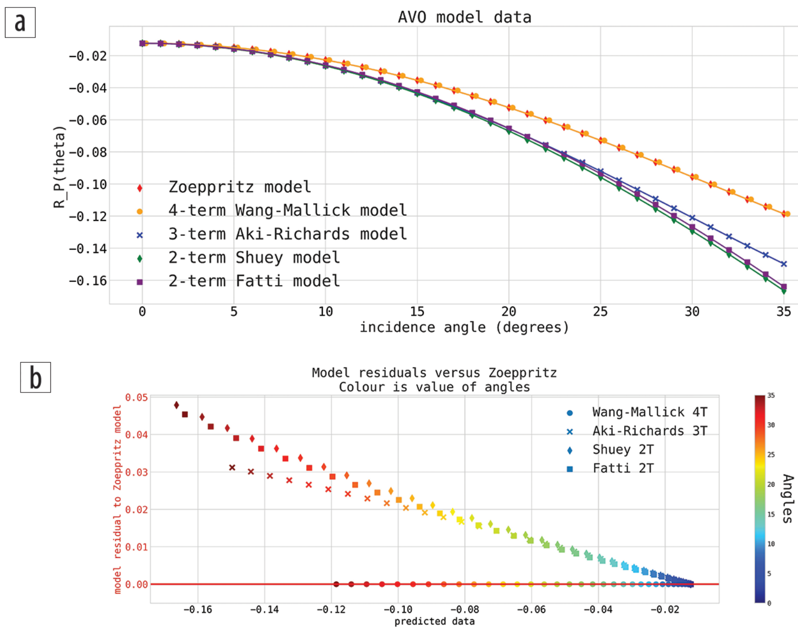

Figure 1. (a) The main AVO models used in this paper. Since OVB needs a linearized model, I use the four-term Wang-Mallick model to approximate the Knott-Zoeppritz equation. (b) Residuals of the models relative to the Zoeppritz model. Note: I plot some data points slightly offset in angle to make them more visible.

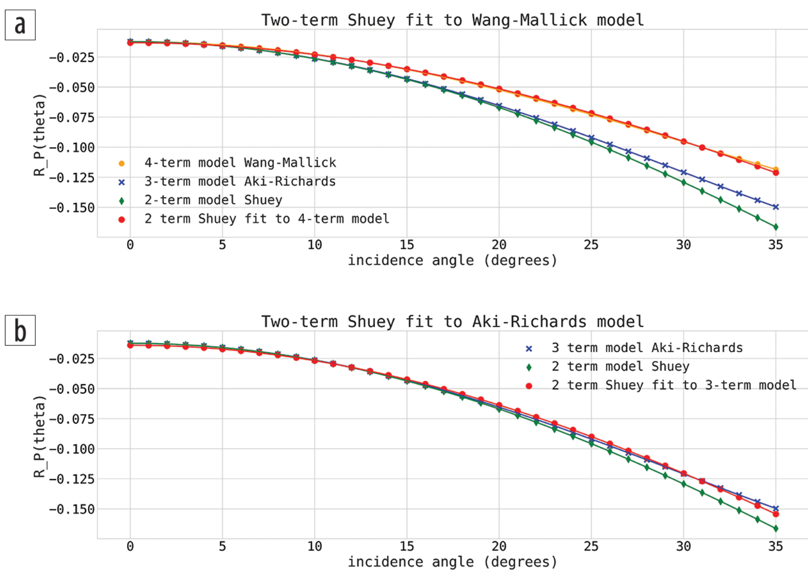

Figure 2. The two-term Shuey fit using a maximum angle of 35° matches (a) the four-term AVO model or (b) the three-term AVO better than the two-term AVO model. This is an expression of bias: To match the data, the two model parameters in the misspecified least-squares model adjust from their unbiased values to fit the data model with four or three parameters.

Fitting the data with less than these four basis vectors means choosing a design matrix omitting columns starting from the right. Since our goal is to compare the inversion with the full design matrix $\boldsymbol{A}$ to one with a reduced set of variables, we identify these two sets of parameters with subscript i, for included, and o, for omitted, and split the design matrix and model vector into two corresponding parts:

$$\boldsymbol{d} = \boldsymbol{A}\boldsymbol{m} = \boldsymbol{A}_i\boldsymbol{m}_i + \boldsymbol{A}_o\boldsymbol{m}_o = \begin{pmatrix} \boldsymbol{A}_i & \boldsymbol{A}_o \end{pmatrix}\begin{pmatrix} \boldsymbol{m}_i \\ \boldsymbol{m}_o \end{pmatrix}. \quad (11)$$

We derive the least-squares solution of the model parameters with the full design matrix, in this notation, via the normal equations $(\boldsymbol{A}^T\boldsymbol{A})\boldsymbol{m} = \boldsymbol{A}^T\boldsymbol{d}$:

$$\begin{pmatrix} \boldsymbol{A}_i^T\boldsymbol{A}_i & \boldsymbol{A}_i^T\boldsymbol{A}_o \\ \boldsymbol{A}_o^T\boldsymbol{A}_i & \boldsymbol{A}_o^T\boldsymbol{A}_o \end{pmatrix}\begin{pmatrix} \boldsymbol{m}_i \\ \boldsymbol{m}_o \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_i^T\boldsymbol{d} \\ \boldsymbol{A}_o^T\boldsymbol{d} \end{pmatrix}. \quad (12)$$

This gives two least-squares equations for the components $\boldsymbol{m}_i$ and $\boldsymbol{m}_o$ that make up the full unbiased model vector. The equations are coupled via the off-diagonal terms, given by the vector products, the correlations, of the included and omitted basis vectors. When all basis vectors are orthogonal, these off-diagonal contributions vanish, in which case the solution for the included and omitted variables decouples, and we can fit the terms for the included and omitted variables independently of each other. The solution for the first set of parameters is given by

$$\hat{\boldsymbol{m}}_i = \boldsymbol{A}_i^{-g}\boldsymbol{d} - \boldsymbol{A}_i^{-g}\boldsymbol{A}_o\hat{\boldsymbol{m}}_o. \quad (13)$$

In our toy model, this corresponds to $\hat{R}(0) = \hat{S} - \langle \sin^2(\theta) \rangle \hat{G}$. The first term on the right-hand side is the solution to the regression with the subset of the variables denoted by subscript i. This is a regression with the incomplete model, such as a two-term AVO fit to data given by the Aki-Richards or Wang-Mallick equations. The second term on the right-hand side is the influence that the remaining variables with subscript o exert over the first set of variables with subscript i in the regression with the full design matrix. If we only regress with the terms of the matrix $\boldsymbol{A}_i$, we obtain $\boldsymbol{A}_i^{-g}\boldsymbol{d}$, which is a partial, and hence generally biased, solution to the unbiased least-squares regression in the full model. Using the notation and color code for bias introduced earlier, we find that the result of such a partial regression is given by:

$$\begin{aligned} \boldsymbol{A}_i^{-g}\boldsymbol{d} &= \hat{\boldsymbol{m}}_i + \boldsymbol{A}_i^{-g}\boldsymbol{A}_o\hat{\boldsymbol{m}}_o \\ &= \text{unbiased variable} + \text{bias} \\ &= \text{unbiased variable} + \text{data independent weight term} \times \text{omitted variable(s)}. \end{aligned} \tag{14}$$

The model parameters estimated in the reduced regression (left-hand side) are equal to the unbiased least-squares estimate of the model parameters in the full model (first term on right-hand side) plus a bias term. The bias term splits into a data-independent weight term, in the following also called *bias weights*, multiplied with the omitted variables. The weight term (matrix) $\boldsymbol{A}_i^{-g}\boldsymbol{A}_o$ has one entry for each combination of included and omitted variable.

As an example, solving equation 14 analytically for the Shuey two-term bias due to the omission of the curvature term, I find:

$$\begin{aligned} \boldsymbol{A}_{i=(2t)}^{-g}\boldsymbol{d} &= \begin{pmatrix} \hat{R}(0)^{(2t)} \\ \hat{G}^{(2t)} \end{pmatrix} \\ &= \begin{pmatrix} \hat{R}(0) \\ \hat{G} \end{pmatrix} + \frac{1}{\operatorname{var}\left(\sin^2(\theta)\right)} \begin{pmatrix} \langle\sin^4(\theta)\rangle\langle\sin^2(\theta)\tan^2(\theta)\rangle - \langle\sin^2(\theta)\rangle\langle\sin^4(\theta)\tan^2(\theta)\rangle \\ \operatorname{cov}\left(\sin^2(\theta), \sin^2(\theta)\tan^2(\theta)\right) \end{pmatrix} \hat{C}\,. \end{aligned} \tag{15}$$

$\boldsymbol{A}_{i=(2t)}^{-g}$ is the design matrix for two-term AVO, consisting of the first two left columns of the four-term Wang-Mallick matrix equation 10. The angular brackets denote the average over all angles. The result is written in terms of a variance and covariance of the elements of the basis functions[6]. Figure 3 shows the weight terms due to the omission of the respective third term, in the case of Fatti and Shuey two-term AVO to data from a three-term Aki-Richards or from a four-term Wang-Mallick model. Both AVO models have relatively small bias weights for the P-impedance contrast. Gradient bias weights, however, are significantly larger than the bias in the Fatti shear-impedance reflectivity. To obtain the bias, following equation 14, the Shuey bias weights are multiplied with the P-velocity contrast, whereas the Fatti bias weights are multiplied with the density contrast.

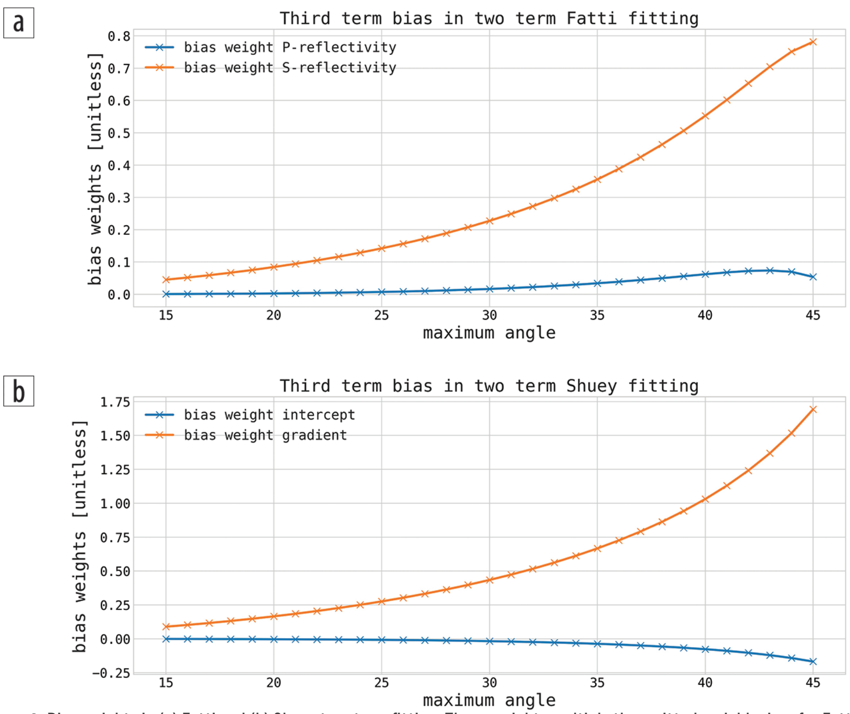

Figure 3. Bias weights in (a) Fatti and (b) Shuey two-term fitting. These weights multiply the omitted variable, i.e., for Fatti the density reflectivity and for Shuey the curvature (the P-wave velocity reflectivity), to give the total bias of this variable due to the omission of the third term. These weights are independent of the AVO model. Note the difference in scale.

[6]The covariance of X and Y is $\operatorname{cov}(X, Y)\ E((X - E(X))E((Y - E(Y)))$. Covariance is positive when X and Y are above or below their expected values in unison, and negative if one of them having a positive fluctuation corresponds with the other having a negative fluctuation. If X and Y are independent, the covariance is zero; the reverse does not hold. The variance of X is given by $\operatorname{var}(X) = \operatorname{cov}(X, X)$.

Systematic versus statistical errors in least-squares AVO fitting

Equation 14 can be rewritten to express the unbiased model estimates as a function of the biased least-squares model parameters from the fit of the reduced model and the unknown omitted variables. Figure 4a shows this for a reduced two-term Shuey model. I model the Shuey parameters $\hat{R}(0)$ and $\hat{G}$ as a function of the two-term model estimates and the two omitted variables for a Wang-Mallick data model with no noise. I have extended the model space for the omitted variables, the curvature and the γ value, well beyond physically meaningful values, since two- and three-term Shuey fitting impose such harsh and unphysical constraints on the solution space.

The results for the two- and three-term least-squares estimates fall along the curve with $\gamma \to \infty$. In this limit, the quadratic correction term vanishes. In addition, the two-term model forces the solution of Shuey two-term AVO to $C = 0$. The two-term least-squares estimate of the gradient is significantly overestimated, whereas the intercept is underestimated; this anticorrelation is due to the off-diagonal element in the two-by-two covariance matrix of the Shuey two-term fit, which we also observe in AVO crossplots.

Such forward modeling is not restricted to the Shuey AVO parameters. Because we know the relation of the Shuey parameters to the reflectivities of other elastic quantities, the modeling can be rewritten to visualize the impact on any two reflectivities. Figure 4b, for example, shows the dependency of the P- and S-impedances on the two biased least-squares fit parameters in the Shuey model and the unknown two values of γ and density reflectivity.

To put the systematic error into perspective, I compare it to the statistical error for the gradient. I run 2000 realizations of the data model with different homoskedastic random noise. In Figure 5, I plot the distributions of the two-term fits with maximum angles of 25° and 35°; for comparison, I also show the gradient distribution of a three-term fit to 35°. In this case, and with the chosen noise variance (a user parameter in the forward modeling), the systematic error due to the higher-order correction dominates the error analysis.

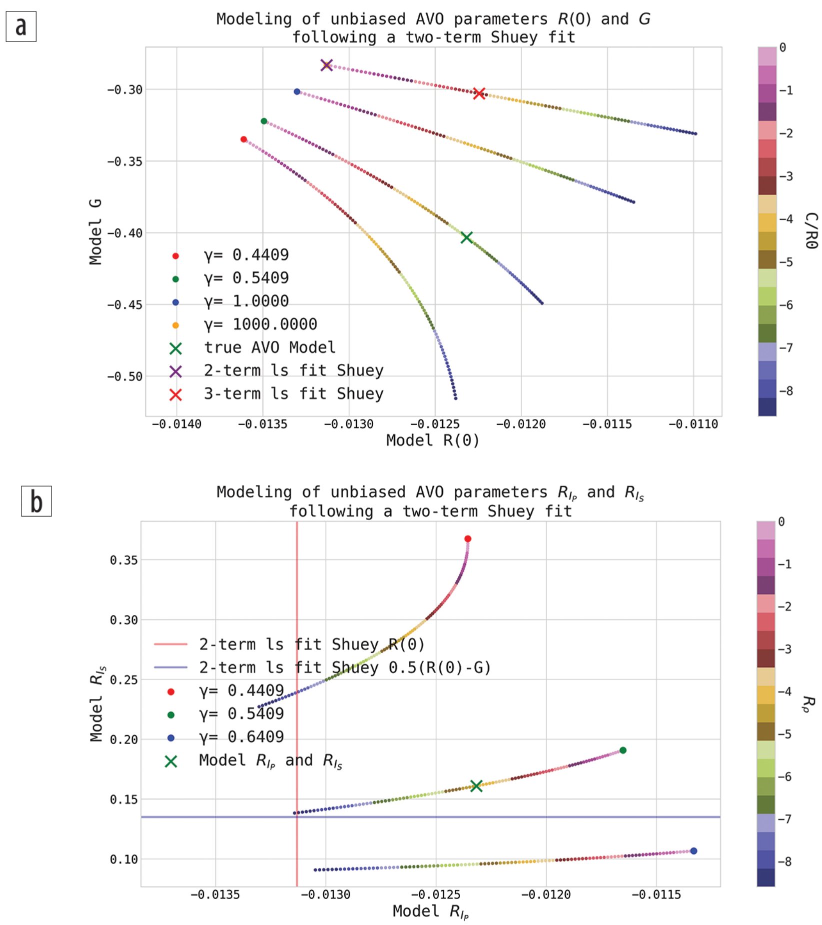

Figure 4. A two-dimensional uncertainty analysis of two-term AVO Shuey fitting when the true data variability is given by the Wang-Mallick model. The least-squares fit results are obtained with a maximum angle of 35°. (a) Following the fit with two terms, I model unbiased values of the Shuey intercept and gradient as a function of γ and $C = R(0)$, using equation 23 to solve for $\hat{m}_i$. In (b), the modeling is performed for the P- and S-wave reflectivities, with γ and density reflectivity R_ρ as parameters. Four values of γ are modeled (indicated by the colored circles), one of which $\gamma = 1000$ (only for panel [a]) mimics the case of infinitely large γ and hence corresponds to vanishing higher-order correction. Both two- and three-term fit results fall on this curve. In addition, the two-term fit result must satisfy $C = 0$.

Elastic reflectivities with bias

In the presence of bias, we must adapt AVO workflows comparing forward-modeled data with AVO results from least-squares fitting. Consider, for example, the calculation of χ-angles in fluid-lithology analysis, and assume, for simplicity, that the three-term Aki-Richard model describes the full variability of the data. While the model gradient is perpendicular to the model intercept, $\chi_G = 90°$, this is no longer true for biased least-squares estimates of the gradient, due to the leakage of omitted variables $\hat{G}^{(2t)} = \hat{G}^{(2t)} + \text{bias}_{\hat{G}} C$. Using a generic Gardner relation (Gardner et al., 1974), $C = c\hat{R}(0)^{(2t)}$, the χ-angle for the least-squares gradient estimate with bias is given by:

$$\tan\left(\chi_{\hat{G}^{(2t)}}\right) = \frac{1}{c\,\text{bias}_{\hat{G}}}. \qquad (16)$$

With a typical value of $c = 0.8$ and a conservative bias weight of 0.25, the gradient is rotated toward the intercept by more than 10°.

To derive the most general elastic reflectivity in three-term Shuey space, we write the linearized three-term Shuey AVO in terms of the reflectivities of P-wave velocity, S-wave velocity, and density, but with the addition of the bias:

$$\begin{pmatrix}\hat{R}(0)\\ \hat{G}\\ C\end{pmatrix} = \begin{pmatrix} 1+\text{bias}_{\hat{R}(0)} & 0 & 1\\ 1+\text{bias}_{\hat{G}} & -\frac{8}{\gamma^2} & -\frac{4}{\gamma^2}\\ 1 & 0 & 0\end{pmatrix}\begin{pmatrix}R_{V_P}\\ R_{V_S}\\ R_\rho\end{pmatrix}. \qquad (17)$$

The bias terms are the weights previously calculated. We need the inverse of this relation:

$$\begin{pmatrix}R_{V_P}\\ R_{V_S}\\ R_\rho\end{pmatrix} = \begin{pmatrix} 0 & 0 & 1\\ -\frac{1}{2} & -\frac{\gamma^2}{8} & \frac{1}{2}\left(1+\frac{\gamma^2}{4}\right)+\frac{1}{2}\left(\frac{\gamma^2}{4}\text{bias}_{\hat{G}}+\text{bias}_{\hat{R}(0)}\right)\\ 1 & 0 & -1-\text{bias}_{\hat{R}(0)}\end{pmatrix}\begin{pmatrix}\hat{R}(0)\\ \hat{G}\\ C\end{pmatrix}. \qquad (18)$$

Any elastic reflectivity can be written as a linear combination of the three fundamental elastic parameters (Ball et al., 2014a; Connolly, 2019):

$$R_x = \begin{pmatrix} c_1 & c_2 & c_3\end{pmatrix}\begin{pmatrix}R_{V_P}\\ R_{V_S}\\ R_\rho\end{pmatrix}. \qquad (19)$$

Now insert the expression we already calculated, equation 18, and keep simplifying to find

$$R_x = \hat{R}(0)\left(-\frac{1}{2}c_2+c_3\right)+\hat{G}\left(-\frac{\gamma^2}{8}c_2\right)+C\left(c_1+c_2\frac{1}{2}\left(1+\frac{\gamma^2}{4}\right)-c_3\right)$$
$$+\,C\left(c_2\frac{1}{2}\left(\frac{\gamma^2}{4}\text{bias}_{\hat{G}}+\text{bias}_{\hat{R}(0)}\right)-c_3\text{bias}_{\hat{R}(0)}\right). \qquad (20)$$

Equation 20 is the generalization to arbitrary elastic reflectivities of the relation between unbiased and biased parameters in two-term AVO fitting. It expresses the unbiased elastic reflectivity R_x in terms of the biased least-squares estimates of the Shuey intercept and gradient, the bias weights calculated analytically from the design matrix of the AVO fit, and the two unknowns C and γ. Setting the bias weights to zero for both parameters recovers equations from forward modeling. Extensions to the method are of course possible, such as expressing the unbiased elastic reflectivities R_x in other AVO models. This can also be used to create analytic relations between biased fit parameters in different AVO models.

To express equation 20 in terms of a χ-angle we use a two-dimensional AVO space and write R_x as a weighted sum of Shuey intercept and gradient. We eliminate the curvature via a generalized Gardner's relation (Gardner et al., 1974) $C = c\hat{R}(0)$, relating the curvature to the least-squares estimate of the intercept. Then,

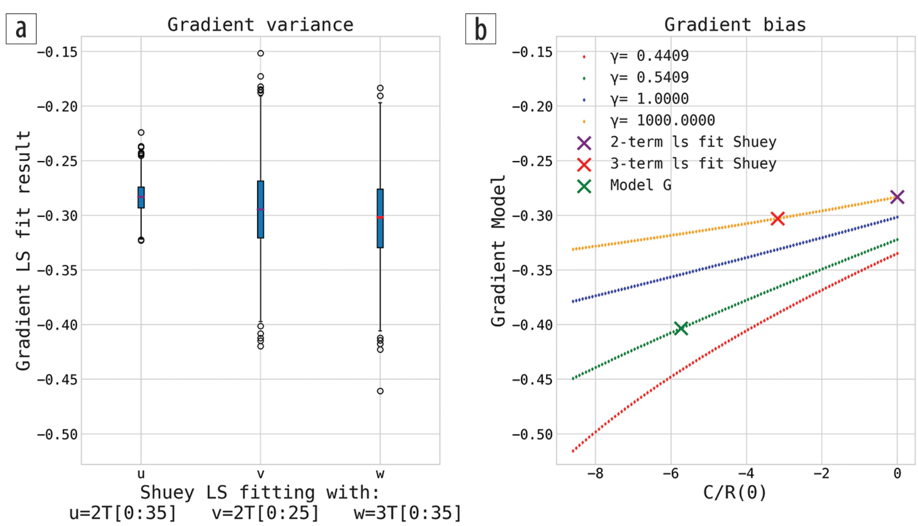

Figure 5. Analysis of gradient bias with homoskedastic noise: The plot in (a) shows boxplots of the gradient values from a Monte Carlo run of two- (2T) and three-term (3T) least-squares fitting with a Shuey AVO model. In all cases, the data are modeled with a four-term Wang-Mallick linearized AVO model to which I add Gaussian noise. I include a two-term fit with a smaller angle range to 25°. In (b), I show the bias of the gradient in the noise-free case similar to the analysis of Figure 4; each point corresponds to a model gradient based on the two-term least-squares fit result to 35° and with the omitted variables γ and $C/R(0)$ as parameters. Four values of γ are modeled, as before, and color coded. Due to the importance of the higher-order corrections for this AVO model, all of these fits are significantly biased and disagree with the model result, even taking statistical errors into account.

$$\tan(\chi_x) = \frac{-\frac{\gamma^2}{8}c_2}{-\frac{1}{2}c_2 + c_3 + c\left\{c_1 + c_2\frac{1}{2}\left(\left(1+\text{bias}_{\hat{R}(0)}\right) + \frac{\gamma^2}{4}\left(1+\text{bias}_{\hat{G}}\right) - c_3\left(1+\text{bias}_{\hat{R}(0)}\right)\right)\right\}} \ . \quad (21)$$

As for the reflectivity, setting the bias weights to zero recovers the conventional χ-angles from forward modeling.

For example, using the simple case where $\gamma = 2$, the χ-angle for the shear-impedance reflectivity R_{I_s} in terms of the biased model estimates is

$$\begin{aligned} \begin{pmatrix} c_1 & c_2 & c_3 \end{pmatrix}_{I_s} &= \begin{pmatrix} 0 & 1 & 1 \end{pmatrix} , \\ R_{I_s} &= \frac{1}{2}\hat{R}(0)\left(1 + c\left(\text{bias}_{\hat{G}} - \text{bias}_{\hat{R}(0)}\right)\right) - \frac{1}{2}\hat{G} \ , \\ \tan(\chi_{I_s}) &= \frac{-1}{1 + c\left(\text{bias}_{\hat{G}} - \text{bias}_{\hat{R}(0)}\right)} \ . \end{aligned} \quad (22)$$

Since the gradient bias is significantly larger than the intercept bias, the χ-angle for R_{I_s} is rotated toward the intercept axis, like the gradient example above. Elastic reflectivities R_x with positive c_2 weight are rotated toward the intercept axis, a negative c_2, such as for V_P/V_S and λ, means a correction to higher angles relative to $R(0)$.

Equation 22 shows how the model relation between R_{I_s} and the Shuey parameters, $2R_{I_s} = R(0) - G$, valid for $V_P/V_S = 2$, proposed by Wiggins et al. (1983), is altered in least-squares model parameter space. This was also observed by Causse et al. (2007) and by Thomas et al. (2016). The solution proposed by Thomas et al. (2016) is subtly different from mine. Their analysis relates least-squares model parameters in different AVO models, both potentially biased. Just like the bias weights, the conversion is given in terms of the design matrices (Ball et al., 2014b; Thomas et al., 2016):

$$\hat{\boldsymbol{m}}_{\text{model b}} = \boldsymbol{A}^{-g}_{\text{model b}}\boldsymbol{A}_{\text{model a}}\hat{\boldsymbol{m}}_{\text{model a}} \ . \quad (23)$$

These authors suggested to use this relation to convert highly biased two-term model estimates, such as a two-term Shuey AVO fit to large maximum angle, to "a low-residual domain such as the Fatti domain" (Ball et al., 2014b). This works well when the chosen reference AVO model and the relevant AVO model parameters $\hat{\boldsymbol{m}}_{\text{model b}}$ are near bias free. OVB and equation 20, by contrast, relate the biased parameters to the bias-free model and hence provide a unique relationship between any linearized AVO parameterization and the unbiased model space.

Discussion and conclusion

This paper shows how to calculate exactly the bias due to misspecified models in least-squares parameter estimation. To do this, I introduced OVB, a technique well known in least-squares analysis in the context of econometric data analysis (Greene, 2003) but that I have not yet seen applied to geophysical data analysis. I showed how OVB can be applied to analysis of linearized isotropic AVO models, both analytically and numerically. For misspecified models, such as two-term AVO fitting with large angle range or with large contrasts, OVB provides relations between the biased and unbiased least-squares model parameters. Using relative rock physics, I showed how bias propagates into other elastic reflectivities. Equation 20, or its corresponding version for other AVO models, can be used to build tables of unique relative rock property relationships for any AVO model, which replace the unbiased, forward-model results. The resulting equations can be used to forward model unbiased AVO quantities, using the least-squares fit results, the weights given by OVB analysis, and the omitted variables. This analysis can also be used to make informed decisions on maximum AVO fit angles. However, to do so in a meaningful manner, other sources of bias, such as angle errors, ignoring anisotropy, and wavelet variations with offset, need to be accounted for — a task beyond the scope of this paper.

Various orthogonal AVO schemes have been proposed, such as with orthogonal polynomials (Johansen et al., 1995), using data-driven principal component analysis (PCA) (Saleh and de Bruin, 2000; Cambois and Herrmann, 2001) or using model-based PCA (Causse et al., 2007). In these schemes, the orthogonality of the covariance matrix decouples the least-squares model parameters, making them bias free and allowing us to fit consecutive higher-order terms without having to redo a full parameter fit. The decorrelation also makes orthogonal AVO parameters better suited to AVO crossplot analysis.

In any of these orthogonal AVO parameterizations, the model fit parameters are different from the ones originally proposed in the AVO linearized models. This was evident in the toy model, where the first term in the orthogonalized Shuey scheme is the stack and hence a mix of the original Shuey parameters of all orders in the AVO model. Following an orthogonal fit, we may choose to transform back to the original AVO model of choice. However, transforming from the orthogonal parameters to the original parameters reintroduces bias *unless the full variability of the data has been explained in the orthogonal fit*.

If we wish to project into a lower-dimensional AVO space, such as χ-angle space, we need to take into account uncertainties in the missing variables, as shown in Figure 4. At this stage, additional constraints, and prior information from rock physics, relating to the omitted parameters, could also be introduced and used to further derisk the AVO inversion. At the very least, as shown in the discussion of the impact of bias on χ-angles, the OVB analysis provides guides for the direction of the bias as it propagates from our fit parameters to other elastic reflectivities. It tells us which derived reflectivities are systematically under- or overestimated. These results should be valid even if further limitations, such as those of the Knott-Zoeppritz model itself, or other sources of bias, such as heteroskedastic noise, come into play.

I applied OVB to isotropic linearized AVO. However, the theory of OVB is general and applicable whenever a misspecified model is used in least-squares fitting or inversion. It therefore should be of use in other areas of seismic imaging and inversion. TLE

Acknowledgments

I thank my CGG colleagues Jon Downton, Céline Lacombe, Jeremy Palmer, and Brian Russell for discussions and encouragement during the development of this work. All remaining errors are mine alone.

Data and materials availability

A Jupyter notebook is included as supplemental material at https://github.com/cgg-com/TLE2021_AVO and via Binder at https://mybinder.org/v2/gh/cgg-com/TLE2021_AVO/HEAD?urlpath=lab, allowing the user to reproduce the figures in the paper with their own AVO models. The AVO model values used to create the figures in this paper are from a class 3 AVO event with ($V_{P,upper}$ = 2550 m/s; $V_{P,lower}$ = 2880 m/s; $V_{S,upper}$ = 1100 m/s; $V_{S,lower}$ = 1810 m/s; ρ_{upper} = 2.35 g/cm^3; ρ_{lower} = 1.99 g/cm^3).

Corresponding author: henning.hoeber@cgg.com

References

Aki, K., and R. G. Richards, 1980, Quantitative seismology: Theory and methods: W. H. Freeman and Company.

Aster, R. C., B. Borchers, and C. H. Thurber, 2005, Parameter estimation and inverse problems: Elsevier Academic Press.

Ball, V., J. P. Blangy, C. Schiøtt, and A. Chaveste, 2014a, Relative rock physics: The Leading Edge, **33**, no. 3, 276–286, https://doi.org/10.1190/tle33030276.1.

Ball, V., L. Tenorio, J. P. Blangy, M. Thomas, and C. Schiøtt, 2014b, Uncertainty quantification of two-term relative elastic inversion: Second Integrated Reservoir Modelling Conference, EAGE, https://doi.org/10.3997/2214-4609.20147465.

Cambois, G., and P. Herrmann, 2001, Statistically uncorrelated AVO attributes: 63rd Conference and Exhibition, EAGE, https://doi.org/10.3997/2214-4609-pdb.15.P187.

Causse, E., M. Riede, A. J. van Wijngaarden, A. Buland, J. F. Dutzer, and R. Fillon, 2007, Amplitude analysis with an optimal model-based linear AVO approximation: Part I — Theory: Geophysics, **72**, no. 3, C59–C69, https://doi.org/10.1190/1.2668600.

Connolly, P., 2019, Elastic reflectivity vectors and the geometry of intercept-gradient crossplots: The Leading Edge, **38**, no. 10, 762–769, https://doi.org/10.1190/tle38100762.1.

Connolly, P. A., 2017, Chi: 79th Conference and Exhibition, EAGE, https://doi.org/10.3997/2214-4609.201700825.

De Nicolao A., G. Drufuca, and F. Rocca, 1993, Eigenvalues and eigenvectors of linearized elastic inversion: Geophysics, **58**, no. 5, 670–679, https://doi.org/10.1190/1.1443451.

Fatti, J. L., G. C. Smith, P. J. Vail, P. J. Strauss, and P. R. Levitt, 1994, Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the Geostack technique: Geophysics, **59**, no. 9, 1362–1376, https://doi.org/10.1190/1.1443695.

Gardner, G. H. F., L. W. Gardner, and A. R. Gregory, 1974, Formation velocity and density — The diagnostic basics for stratigraphic traps: Geophysics, **39**, no. 6, 770–780, https://doi.org/10.1190/1.1440465.

Gidlow, P. M., G. C. Smith, and P. Vail, 1992, Hydrocarbon detection using fluid factor traces: A case history: 3rd SAGA Biennial Conference and Exhibition, https://doi.org/10.3997/2214-4609-pdb.224.041.

Greene, W. H., 2003, Econometrics analysis: Prentice Hall.

Johansen, T. A., L. Bruland, and J. Lotro, 1995, Tracking the amplitude versus offset (AVO) by using orthogonal polynomials: Geophysical Prospecting, **43**, no. 2, 245–261, https://doi.org/10.1111/j.1365-2478.1995.tb00134.x.

Knott, C. G ., 1888, Reflexion and refraction of elastic waves, with seismological applications, Read at Tokyo before the Seismological Society of Japan, 23 February 23 1888, and published in their Transactions of that year. Available in Pelissier et al. (2007).

Mallick, S., 1993, A simple approximation to the P-wave reflection coefficient and its implication in the inversion of amplitude variation with offset data: Geophysics, **58**, no. 4, 544-552, https://doi.org/10.1190/1.1443437.

Menke, W., 2012, Geophysical data analysis: Discrete inverse theory: Elsevier.

Pelissier, M. A., H. Hoeber, N. van de Coevering, and I. F. Jones, eds., 2007, Classics of elastic wave theory: Geophysics Reprint Series 24, SEG, https://doi.org/10.1190/1.9781560801931.

Saleh, S. J., and J. A. de Bruin, 2000, AVO attribute extraction via principal component analysis: 70th Annual International Meeting, SEG, Expanded Abstracts, 126–129, https://doi.org/10.1190/1.1815626.

Shuey, R. T., 1985, A simplification of the Zoeppritz equations: Geophysics, **50**, no. 4, 609-614, https://doi.org/10.1190/1.1441936.

Thomas, M., V. Ball, J. P. Blangy, and L. Tenorio, 2016, Rock-physics relationships between inverted elastic reflectivities: The Leading Edge, **35**, no. 5, 438–444, https://doi.org/10.1190/tle35050438.1.

Ursin, B., and E. Tjaland, 1993, The accuracy of linearised elastic parameter estimation: Journal of Seismic Exploration, **2**, 349–363.

Wang, Y., 1999, Approximations to the Zoeppritz equations and their use in AVO analysis: Geophysics, **64**, no. 6, 1920–1927, https://doi.org/10.1190/1.1444698.

Wiggins, R., G. S. Kenny, and C. D. McClure, 1983, A method for determining and displaying the shear-velocity reflectivities of a geologic formation: European Patent Application 0113944.

Zoeppritz, K., 1919, Erdbebenwellen VIIb. Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen: Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse, 66–84. Available in English in Pelissier et al. (2007).