Quantifying uncertainties in AVO forward modeling

Ayon K. Dey, Chandra Rai[*], and Carl Sondergeld[*]

INTRODUCTION

A common tool used in the exploration of hydrocarbons is Amplitude-Versus-Offset (AVO) analysis. This particular method has been very effective in identifying major natural gas deposits. Sengupta et al., 1997 shows that the AVO response from data is sensitive to uncertainties in various rock properties. Specifically, the work of Sengupta et al., 1997 shows how the uncertainties in compressional-wave velocities (V_p), shear-wave velocities (V_s), and densities (ρ) project themselves as uncertainties in an AVO response. That is to say, these authors quantify how sensitive the AVO response is to uncertainties in the rock properties.

In this report there is a more detailed analysis of the uncertainties in the rock properties and its effect on the AVO response. The rock properties V_p, V_s, and ρ are used in AVO forward modeling and the results are analysed. The focus of this work is to define confidence regions for various exploration scenarios and to examine how these different regions interact. The approach taken is to create an elliptical area that contains a certain amount of scatter data (parameters b_0 and b_1, in this case) and to examine to what degree the various areas overlap.

BACKGROUND

This project breaks down into two theoretical areas. The first theoretical area is that of estimating a confidence region by ellipse determination; the second is the determination of whether these confidence regions interact and, if so, to what degree. This section discusses the major issues in the theory for both of these areas. The next section outlines how the theory is resolved for this particular problem.

The general second-degree equation has the form

$$Ax^2+2Bxy+Cy^2+2Dx+2Ey+F=0$$

and the entire family of conic sections is defined by it. That is, parabolas, hyperbolas, circles, and ellipses are all defined by the general equation. These conics relate to the general equation by the following theorem.

THEOREM 1: Consider the equation $Ax^2+2Bxy+Cy^2+2Dx+2Ey+F=0$. If A, B, and C are not all 0 and if the graph is not degenerate, then the graph is:

- *a circle or an ellipse if $B^2-4AC<0$ (in a circle, B=0 and A=C), or*

- *a parabola if $B^2-4AC=0$, or*

[*] Tulsa Technology Center, Amoco Corporation.

Contents

- *a hyperbola if B^2-4AC>0.*

Hence, the magnitude and the sign of the coefficients *A*, *B*, and *C* will determine what curve is produced from the general equation. Moreover, the existence of coefficients *B*, *D*, and *E* that are not all zero will cause the conic section to deviate from the standard position. Things can be simplified somewhat by the following two theorems (Hart, 1950).

THEOREM 2: If B^2-4AC≠0, then the curve(s) defined by the general second-degree equation has a unique center (h,k), and translation of axes to this point as origin gives a transformed equation without linear terms,

$$Ax'^2+2Bx'y'+Cy'^2+F=0.$$

THEOREM 3: If B≠0, then rotation of axes through an angle θ will remove the xy term from the general second-degree equation if θ satisfies the equation

$$cot(2\theta) = \frac{A\text{-}C}{2B}.$$

Theorems 2 and 3 state that any form of the general equation can be reduced to a standard form by eliminating *x*, *y*, and *xy* terms via translation and rotation. All conic equations can reduce to the form:

$$A''x''^2+C''y''^2+F''=0,$$

where

$$x'' = (x-h)\cos(\theta)+(y-k)\sin(\theta),$$

and

$$y'' = -(x-h)\sin(\theta)+(y-k)\cos(\theta).$$

The reduced form of the general equation states that only the signs of *A* and *C* are relevant in determining which curve results. The equation defines an ellipse if *-4A″C″<0* and it defines a hyperbola if *-4A″C″>0*.

Fitting ellipses to data is a fundamental problem in various branches of science. Most widely employed methods use one of two general techniques: clustering and least-squares fitting. Clustering methods are based on mapping sets of points to the parameter space. These methods have great advantages (see Fitzgibbon, 1996) but are computationally complex and yield non-unique solutions. The least-squares approach (also see Fitzgibbon, 1996) traditionally attempts to fit an ellipse to the general second-degree equation by estimating the coefficients and then rejecting non-elliptical fits. These methods are cheap and effective if the data is relatively noise free and lies on an elliptical arc. However, if the data is not strictly elliptical and there are significant amounts of noise, then an unbounded hyperbolic fit will result.

Contents

After determining the ellipses for various reflection scenarios, one is concerned with the overlap of various ellipses. When ellipses have been assigned to two interfaces, it is of great interest to know how much separation exists between the two ellipses. At the most general level, this problem is one of classical calculus, involving the determination of the area of intersection between two curves. The procedure is to determine the points of intersection and to integrate over this domain. While it seems that there is an analytical approach to resolving this issue, a significant problem manifests itself in the translation from the analytical method to the computational method.

To determine the points of intersection between two curves, the usual approach is to set the analytical description of each curve to zero and solve the system simultaneously. In order to see why this method is problematic consider the following system:

$$f_1(x,y)=A_1x^2+2B_1xy+C_1y^2+2D_1x+2E_1y+F_1=0,$$

$$f_2(x,y)=A_2x^2+2B_2xy+C_2y^2+2D_2x+2E_2y+F_2=0.$$

Without loss of generality, assume that A, B, and C are such that the equations above define ellipses. The solution to this system will find the points of intersection (if there are any) between two ellipses f_1 and f_2. This system is also a non-linear system of two equations in two unknowns. Both f_1 and f_2 have zero contour lines that divide the (x,y) plane into regions where their respective equation is positive or negative. These zero contour lines are of interest as the goal is to determine those points common to the zero contours of f_1 and f_2. Generally, the equations have no relation and there is nothing unique about a common zero point that can be exploited. To find all the solutions to the non-linear system of equations, the full zero contours of both functions must be mapped out. Moreover, these contours will, in general, consist of an unknown number of disjoint closed curves. To numerically find these roots, one must have further insight or knowledge of the particular problem. Often it is required that there is knowledge of the number of roots and the neighbourhood of each root.

One of the most common methods to solve such systems numerically is to implement some form of the *Newton-Raphson* method for a non-linear system of equations. The details of this method are beyond the scope of this report. However, it is a very efficient means of converging to a root, if a "sufficiently good" initial guess is known. In order to automate the process, it is required that the number of existing roots are known and that a neighbourhood about each root is known. It is impossible to know, *a priori*, all of this information since the system may have anywhere from zero to infinitely many solutions. Furthermore, the neighbourhood about each root must not be too large, for if it is, the method will not converge. This proves nothing except that the initial guess was poor. A common method is to generate random initial guesses and proceed until convergence is achieved. This leads to enormous computational costs.

Contents

METHODOLOGY

Rather than attempt to implement a general method for ellipse fitting that is complex and not well understood a choice is made to use a different method. The method employed in this investigation is to fit a *standard deviation ellipse* to the scatter data in order to summarise the dispersion in the point pattern and follows the method outlined in Ebdon, 1977.

A standard deviation ellipse is an ellipse centred about the mean centre of a data set, with its long axis in the direction of maximum dispersion and the small axis in the direction of minimum dispersion. In order to fit such an ellipse, the following information must be known:

> 1. the length of the short axis,
>
> 2. the length of the long axis, and
>
> 3. the orientation of the ellipse.

The length of the long axis and the length of the short axis are the values of the data's standard deviation in the x-direction and the y-direction, respectively. Data contained in such an ellipse is interpreted as being within at least *K* standard deviations of the mean center. The step-wise method is as follows.

- Calculate the co-ordinates $(\bar{x}, \bar{y})$ of the mean centre. Note that summations run from 1 to *n*; where *n* is the total number of data points under consideration. By default, the value of *n* is identical in both the $\bar{x}$ and $\bar{y}$ summations.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{i=n} x_i,$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{i=n} y_i.$$

- Calculate the co-ordinates of the data in the translated system by subtracting the mean from each of the original data co-ordinates.

$$\forall x, x' = x - \bar{x}.$$

$$\forall y, y' = y - \bar{y}.$$

- Calculate the tangent of the angle of rotation, θ. Note that from this point forward, all summations can be assumed to go from *i=1* until *i=n*.

Contents

$$\tan(\theta) = \frac{\left(\sum x'^2 - \sum y'^2\right) + \sqrt{\left(\sum x'^2 - \sum y'^2\right)^2 + 4\left(\sum x'y'\right)^2}}{2\sum x'y'} \; .$$

- The arctangent of this value will give the rotation, or orientation, angle θ that is between the translated y-axis and the y-axis of the ellipse. The angle is measured *clockwise* from the translated y-axis. Note that this formulation to calculate the tangent of the orientation angle can give a negative result. If the tangent value happens to become negative, then the correct measure for the orientation angle θ is $\theta = 90° - \arctan(-\tan(\theta))$.

- Calculate the standard deviation along the x-axis of the ellipse. This is a measure of the dispersion in the *x*-direction for the scatter data.

$$\sigma_x = \sqrt{\frac{\sum\left(x'\cos(\theta) - y'\sin(\theta)\right)^2}{n}} \; .$$

- Calculate the standard deviation along the y-axis of the ellipse. This is a measure of the dispersion in the *y*-direction for the scatter data.

$$\sigma_y = \sqrt{\frac{\sum\left(x'\sin(\theta) + y'\cos(\theta)\right)^2}{n}} \; .$$

Expanding the quadratic expressions in the standard deviation formulae can reduce computational time. This reduces the computation time since it avoids multiplying each x' and y' by $cos(\theta)$ and $sin(\theta)$ individually. These formulae will generate a 1-σ ellipse. That is, the data contained in the ellipse will be within one standard deviation of the mean. By introducing a user defined constant *K*, and multiplying the standard deviations in the *x* and *y* directions by it, the algorithm can fit a *K*-σ ellipse. Therefore, ellipses of any size that the user wants can be fit to the data at hand.

Following this method will yield the information necessary to plot an ellipse. Data that lies within this ellipse will be within *K* standard deviations of the mean centre. Often, the ellipse will be translated and rotated away from the origin. By using σ_x, σ_y, and θ, the points on the ellipse can be generated by rotating the system through 360°. To do this, determine a point on the ellipse using the computed information and then rotate the point and translate it to the correct reference frame. Figure 1, below, outlines various standard deviation ellipses and the amount of scatter data that they contain.

Contents
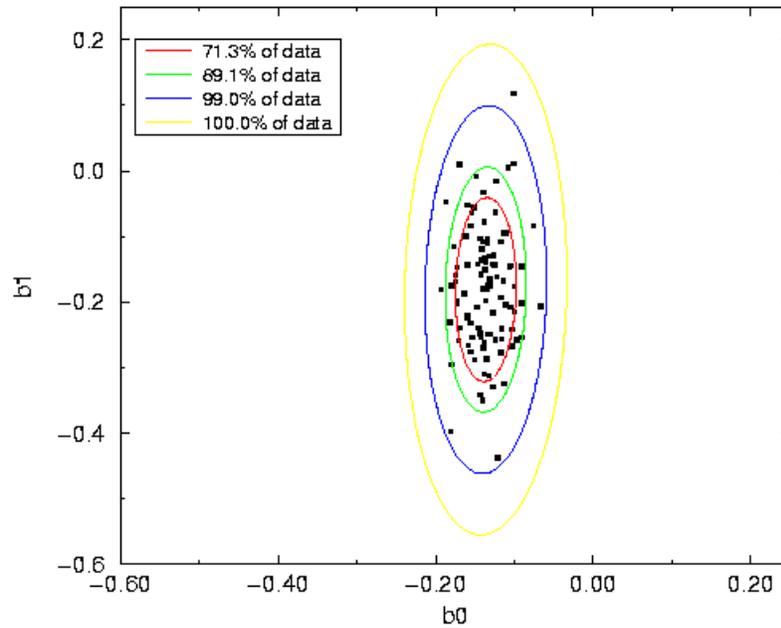
Figure 1: 1.5-σ, 2-σ, 3-σ, and 4-σ ellipses for a representative scatter data set.

The approach taken to analyse the overlap of ellipses from different reflection interfaces is far less elegant. Rather than develop a method that solves for the points of intersection and then implement some type of numerical integration routine, a *brute force* method is chosen. This method exploits the fact that the regions of interest are ellipses and it is computationally cheap because the sizes of the data sets are small (in terms of computation time required). Quite simply, the method examines each point in the two data sets of interest and determines whether or not the point is within both ellipses. If it is within both ellipses, then it means that the point is within *K* standard deviations of both mean centres and cannot be definitely assigned to one region or the other. The algorithm is given below.

- $\overline{x_1}, a_1, \overline{y_1}, b_1, \theta_1, \overline{x_2}, a_2, \overline{y_2}, b_2, \theta_2$ are values determined from ellipse fitting for the two interfaces. The a_i and b_i values are the σ_x and σ_y values, respectively.

- For each point *(x,y)* in both data sets, compute:

$$e_1 = \frac{\left[(x - \overline{x_1})\cos(\theta_1) + (y - \overline{y_1})\sin(\theta_1)\right]^2}{a_1^2} + \frac{\left[-(x - \overline{x_1})\sin(\theta_1) + (y - \overline{y_1})\cos(\theta_1)\right]^2}{b_1^2}$$

and

$$e_2 = \frac{\left[(x - \overline{x_2})\cos(\theta_2) + (y - \overline{y_2})\sin(\theta_2)\right]^2}{a_2^2} + \frac{\left[-(x - \overline{x_2})\sin(\theta_2) + (y - \overline{y_2})\cos(\theta_2)\right]^2}{b_2^2}$$

- If $e_1 < 1$ and $e_2 < 1$, then the data point *(x,y)* is within both ellipses.

Contents

By bookkeeping the number of points that fall within both ellipses, the percentage of total data that lies in an area common to both confidence regions can be computed. The following figures show the various kinds of overlap that can occur.
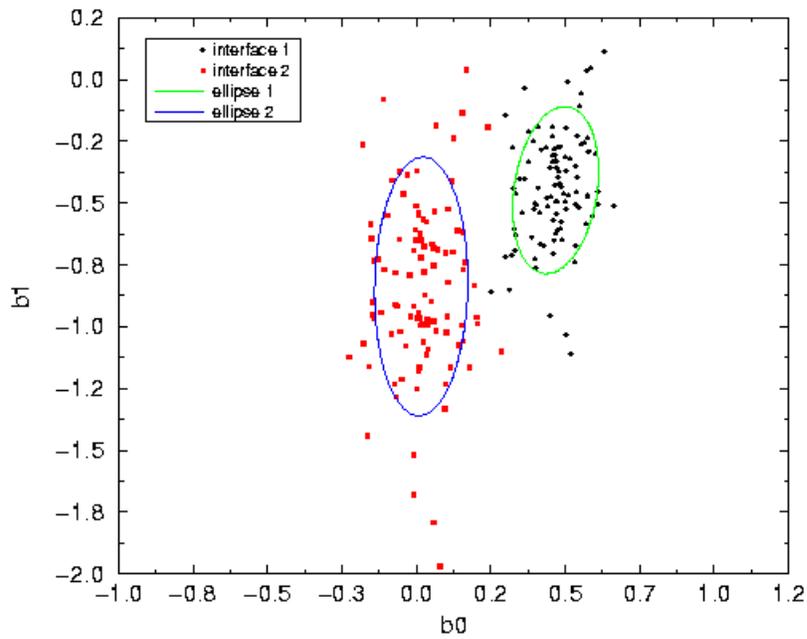
Figure 2: No interaction between confidence regions (0% overlap of contained data).
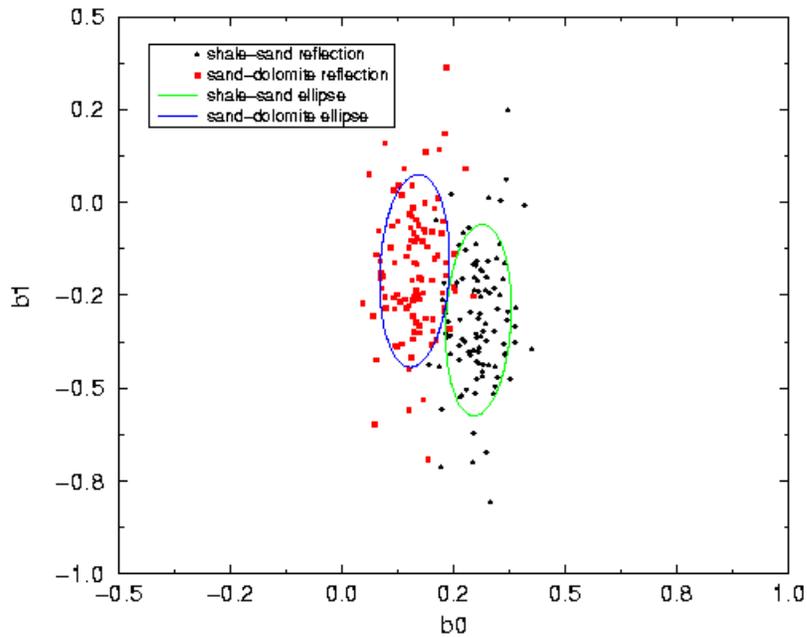
Figure 3: Insignificant interaction between confidence regions (0.5% overlap of data).

Contents
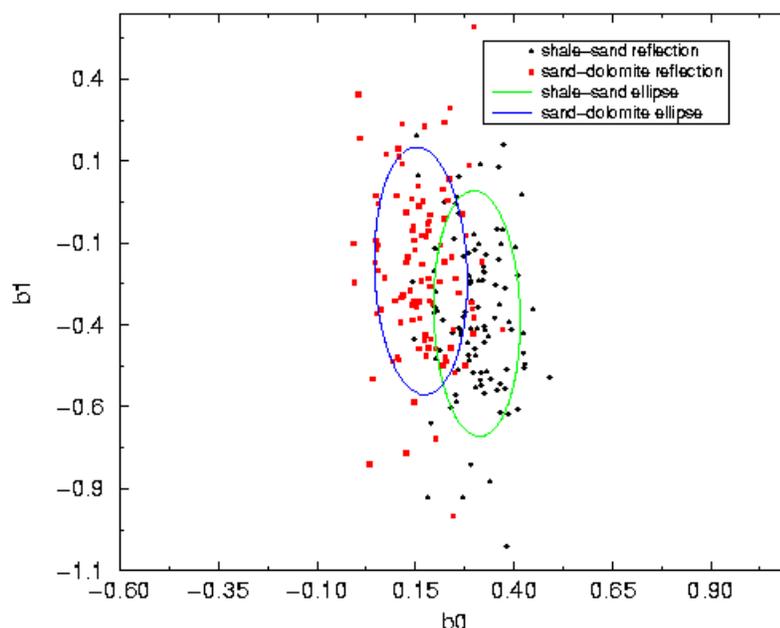
Figure 4: Significant interaction between confidence regions (21.3% overlap of data).

RESULTS

This section shows the results of a number of different tests on the ellipse generation algorithm and the overlap algorithm. Figures 5-7 show scatter plots of b_0 versus b_1 with 20% uncertainty in varying values. Test results when the uncertainties in all the values are the same but the values themselves vary in different combinations are shown in Figures 8, 9, and 10. The final six figures (Figures 12-17) show results for investigations into what occurs when there is a scenario of a shale overlying a sandstone that is saturated with various fluids. Outlined in the table below are the modelling parameters that are used to generate the scatter data seen in Figures 5 through 7. All the ellipses shown have $K=1.5$ as a constant and, hence, all the data within an ellipse is 1.5 standard deviations from the mean center.

Table 2: The values and their uncertainties used to model the data shown in Figures 5-7.

Figure	Layer	V_p *(km/s)*	V_s *(km/s)*	ρ *(gm/cc)*
5	1	$4.875 \pm 20\%$	$2.5 \pm 20\%$	$1.77 \pm 0\%$
	2	$5.940 \pm 20\%$	$3.0 \pm 20\%$	$2.76 \pm 0\%$
6	1	$4.875 \pm 20\%$	$2.5 \pm 0\%$	$1.77 \pm 20\%$
	2	$5.940 \pm 20\%$	$3.0 \pm 0\%$	$2.76 \pm 20\%$
7	1	$4.875 \pm 0\%$	$2.5 \pm 20\%$	$1.77 \pm 20\%$
	2	$5.940 \pm 0\%$	$3.0 \pm 20\%$	$2.76 \pm 20\%$

Contents
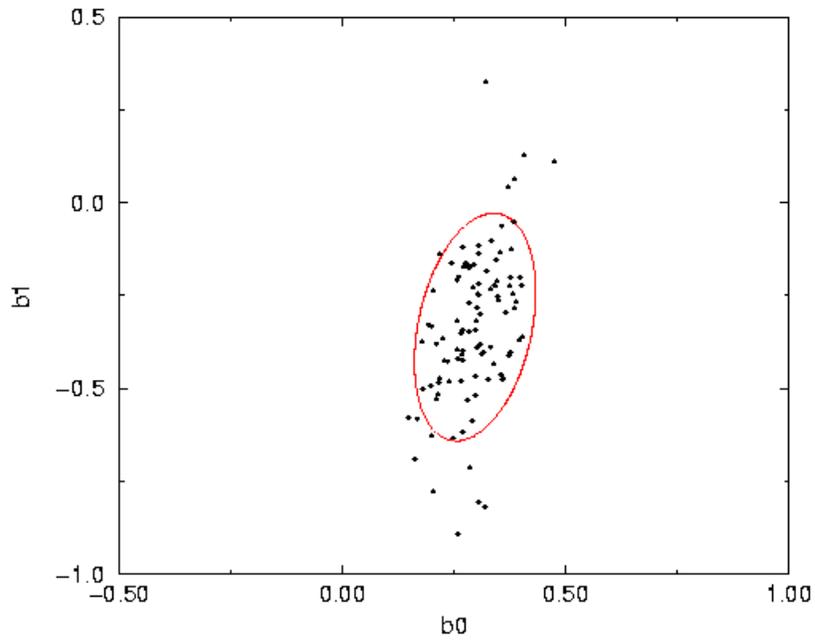
Figure 5: Scatter data with an upwardly right trend and its 1.5-σ confidence region.
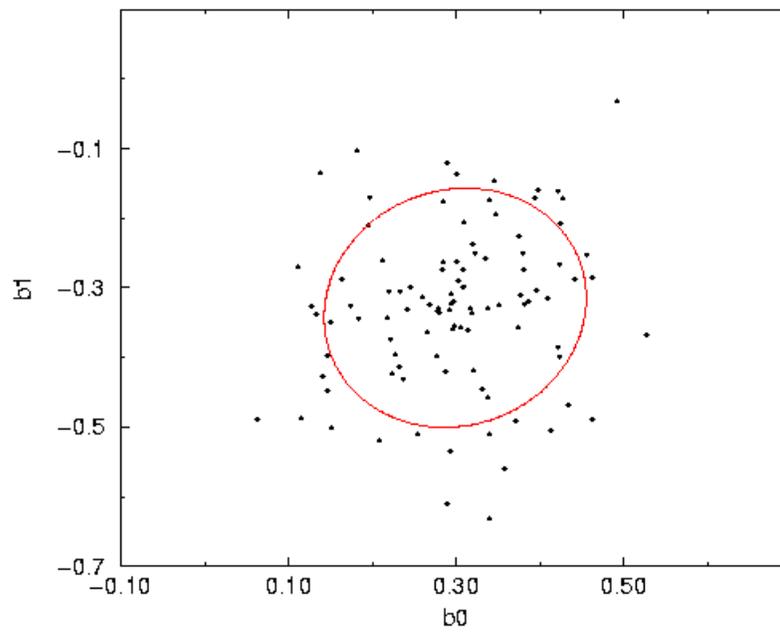
Figure 6: Scatter data with no appreciable trend and its 1.5-σ confidence region.

Contents
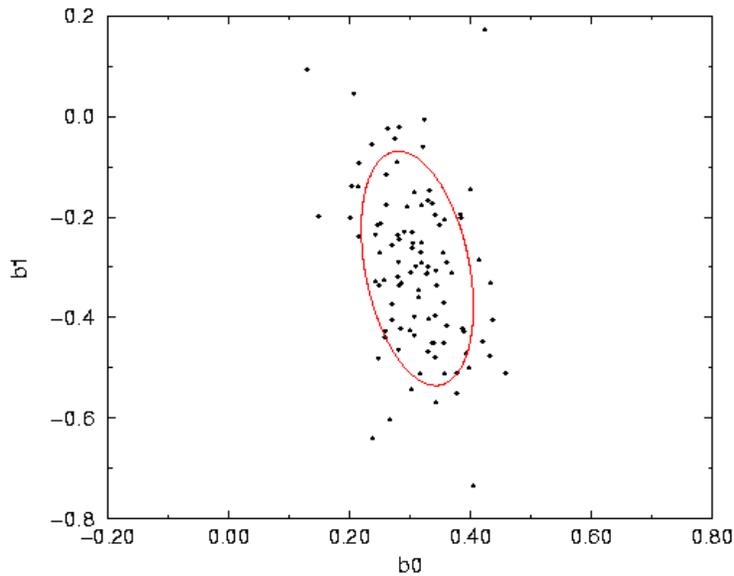
Figure 7: Scatter data with an upwardly left trend and its 1.5-σ confidence region.

These three figures show exactly what is to be expected from the ellipse fitting method. Each of the ellipses has the correct orientation and qualitatively contains the proper amounts of data. There was no actual computation to see how much data lies in each ellipse. This calculation is not difficult. A variation on the method used to calculate the overlap between two ellipses could be used to compute exactly how much data falls within a particular ellipse. It is interesting to note that, for the scatter data used in this report, these various standard deviation ellipses seem to correspond to what would intuitively be a best-fit ellipse to the data. This lends support to the claim that fitting these specific ellipses is a reasonable approximation to the general best-fit ellipse and these ellipses can be generated with far less computational costs and complexity.

Results are now shown for the cases where there is a known, constant uncertainty in the parameter values but the values themselves have variations across the interface. That is to say, the uncertainties in V_p, V_s, and ρ are the same but the values of each can change over the interface of interest. Table 3 outlines the scenarios shown in Figure 8, Figure 9, and Figure 10.

Contents

Table 3: The values and their uncertainties used to model the data shown in Figures 8-10.

Figure	Interface	Layer	V_p *(km/s)*	V_s *(km/s)*	ρ *(gm/cc)*
8	i	1	$4.875 \pm 20\%$	$2.5 \pm 20\%$	$1.77 \pm 20\%$
		2	$5.940 \pm 20\%$	$3.0 \pm 20\%$	$2.76 \pm 20\%$
	ii	1	$4.875 \pm 20\%$	$2.5 \pm 20\%$	$1.77 \pm 20\%$
		2	$5.940 \pm 20\%$	$2.5 \pm 20\%$	$2.76 \pm 20\%$
9	i	1	$4.875 \pm 20\%$	$2.5 \pm 20\%$	$1.77 \pm 20\%$
		2	$5.940 \pm 20\%$	$3.0 \pm 20\%$	$2.76 \pm 20\%$
	ii	1	$4.875 \pm 20\%$	$2.5 \pm 20\%$	$1.77 \pm 20\%$
		2	$4.875 \pm 20\%$	$3.0 \pm 20\%$	$1.77 \pm 20\%$
10	i	1	$4.875 \pm 20\%$	$2.5 \pm 20\%$	$1.77 \pm 20\%$
		2	$5.940 \pm 20\%$	$2.5 \pm 20\%$	$2.76 \pm 20\%$
	ii	1	$4.875 \pm 20\%$	$2.5 \pm 20\%$	$1.77 \pm 20\%$
		2	$4.875 \pm 20\%$	$3.0 \pm 20\%$	$1.77 \pm 20\%$
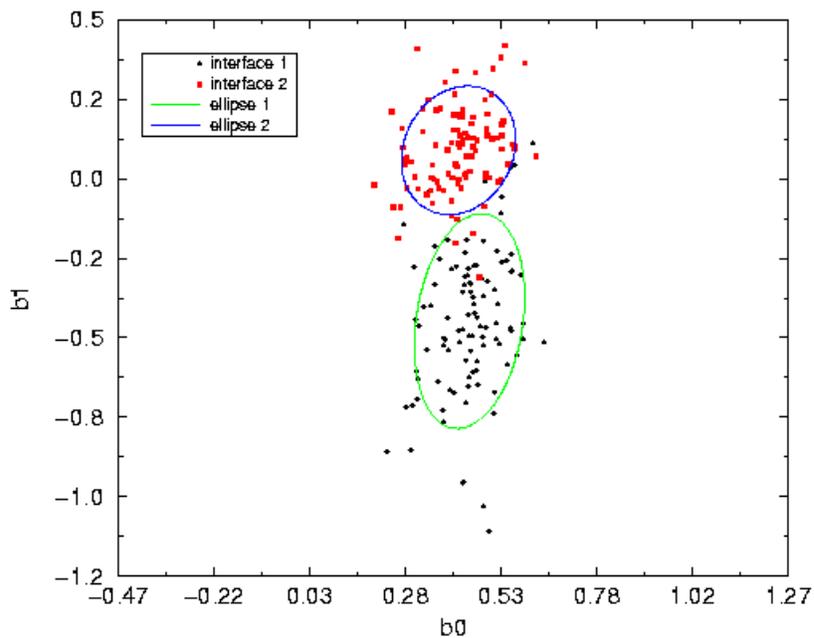
Figure 8: Cross-plot of b_0 versus b_1 for the two forward modelled AVO responses. There is no overlap between the ellipses and, hence, no interaction between the confidence regions.

Contents
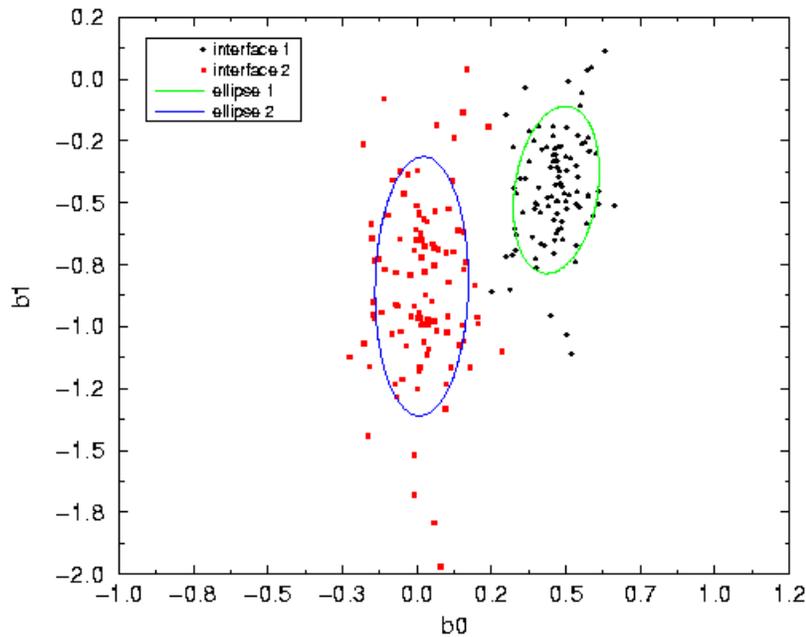
Figure 9: Cross-plot of b_0 versus b_1 for the two forward modelled AVO responses. There is no overlap between the ellipses and, hence, no interaction between the confidence regions.
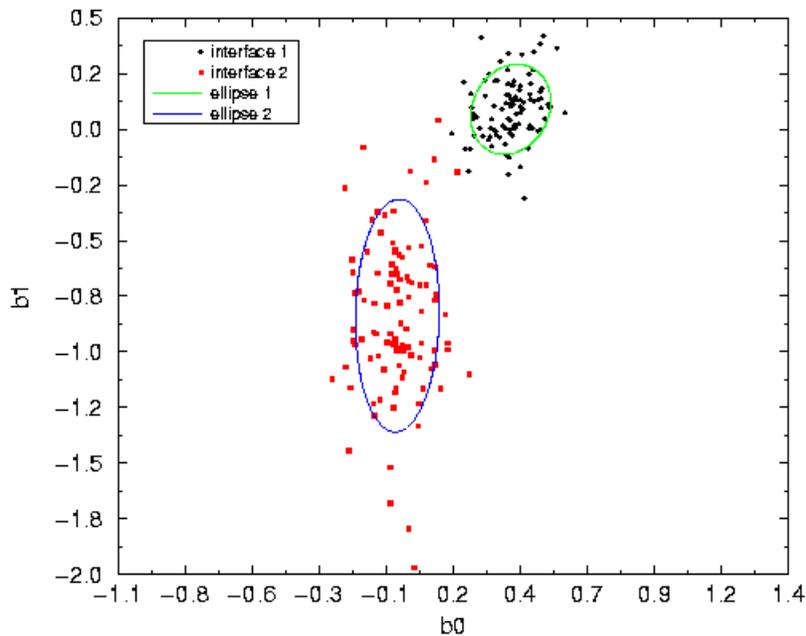
Figure 10: Cross-plot of b_0 versus b_1 for the two forward modelled AVO responses. There is no overlap between the ellipses and, hence, no interaction between the confidence regions.

These plots are representative of three common hydrocarbon exploration areas. Figures 8, 9, and 10 show that for the model parameters in Table 3, there is nice, clean separation between the data of the two interfaces. While this is a pleasing result, more tests need to be done before this can be declared to be the expected result.

Contents

Now, a simple stratigraphic scenario is considered. An isotropic, elastic wavefield's AVO response is forward modelled as it encounters a shale-sandstone interface. The sandstone itself is saturated with various fluids. This is geologically simple in that there are only two layers and there is no dip. This scenario is schematically represented in the diagram below. Figures 12 through 17 show examples that serve to illustrate the AVO sensitivity to pore fluids. They display how this method can be used to determine whether or not AVO will be successful tool in discriminating between different fluid saturations. The plots all have 1.5-σ ellipses that contain approximately 68% of the total data. Table 4 outlines the AVO parameters used.

shale layer
sandstone layer

Figure 11: Simple stratigraphy through which the AVO response of an elastic and isotropic wavefield is modeled.

Table 4: Uncertainties in the parameter values for the shale-sandstone models.

Layer/Saturation Fluid	$V_p \pm 5\%$ *(ft/s)*	$V_s \pm 10\%$ *(ft/s)*	$\rho \pm 5\%$ *(ft/s)*
shale	8700	4000	2.15
brine	8510	4320	2.08
100% black oil	7385	4430	1.98
50% black oil	7568	4374	2.03
20% lean gas condensate	7028	4372	2.03
100% lean gas condensate	7213	4600	1.83

Contents
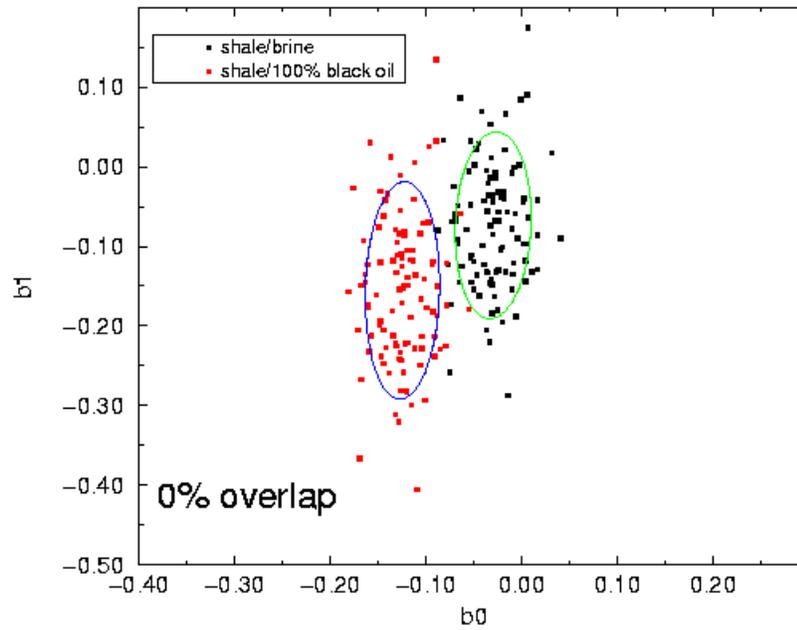
Figure 12: Cross-plot of b_0, b_1 AVO parameters for brine and 100% black oil saturation in the sandstone layer. There is complete separation between confidence regions and, therefore, no interaction between them.
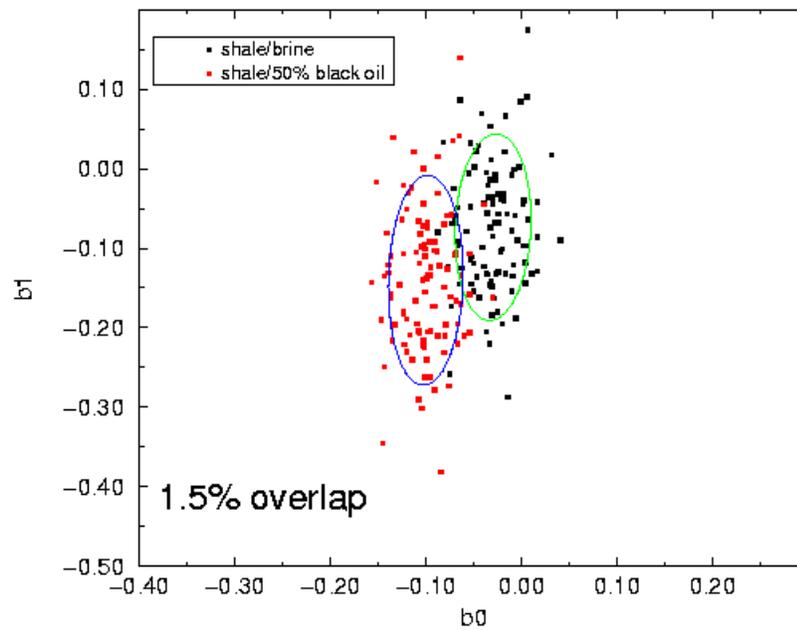
Figure 13: Cross-plot of b_0, b_1 AVO parameters for brine and 50% black oil saturation in the sandstone layer. There is a large degree of separation between confidence regions with very minor interaction between the regions.

Contents
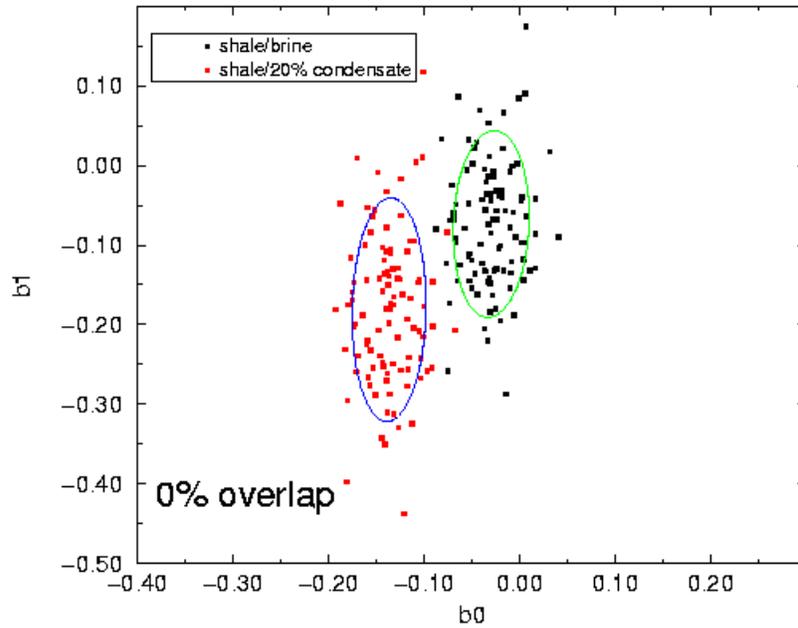
Figure 14: Cross-plot of b_0, b_1 AVO parameters for brine and 20% lean gas condensate saturation in the sandstone layer. There is complete separation between confidence regions and, therefore, no interaction between them.
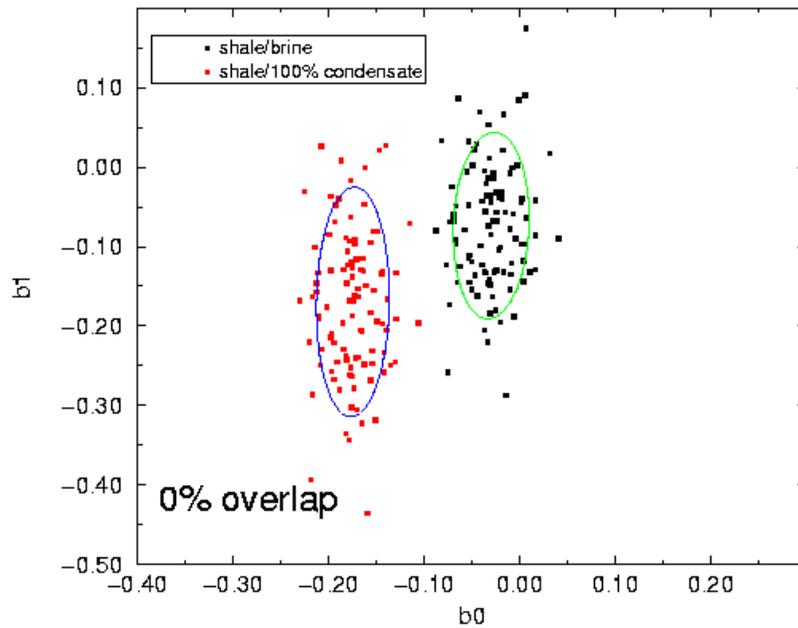
Figure 15: Cross-plot of b_0, b_1 AVO parameters for brine and 100% lean gas condensate saturation in the sandstone layer. There is complete separation between confidence regions and, therefore, no interaction between them.

Contents
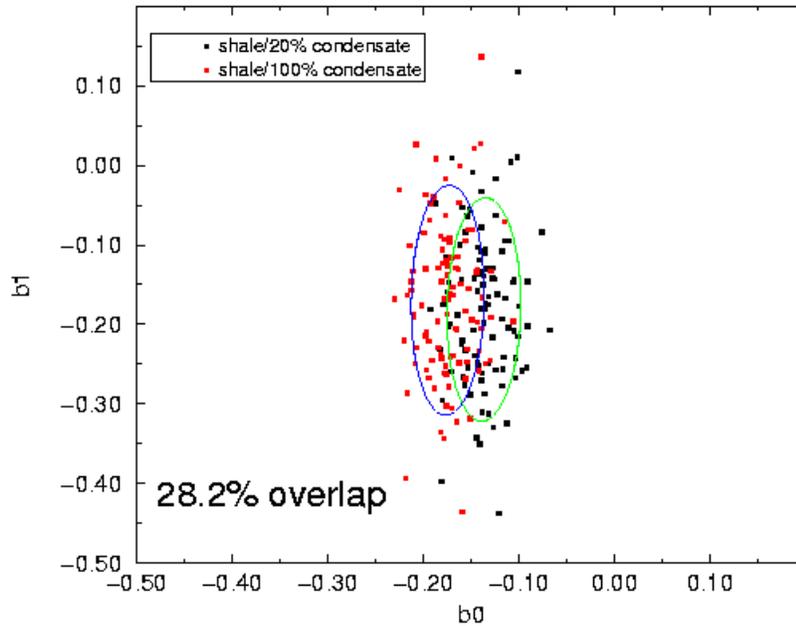
Figure 16: Cross-plot of b_0, b_1 AVO parameters for 20% lean gas condensate and 100% lean gas condensate saturation in the sandstone layer. There is an overlap between confidence regions with a significant amount of interaction.
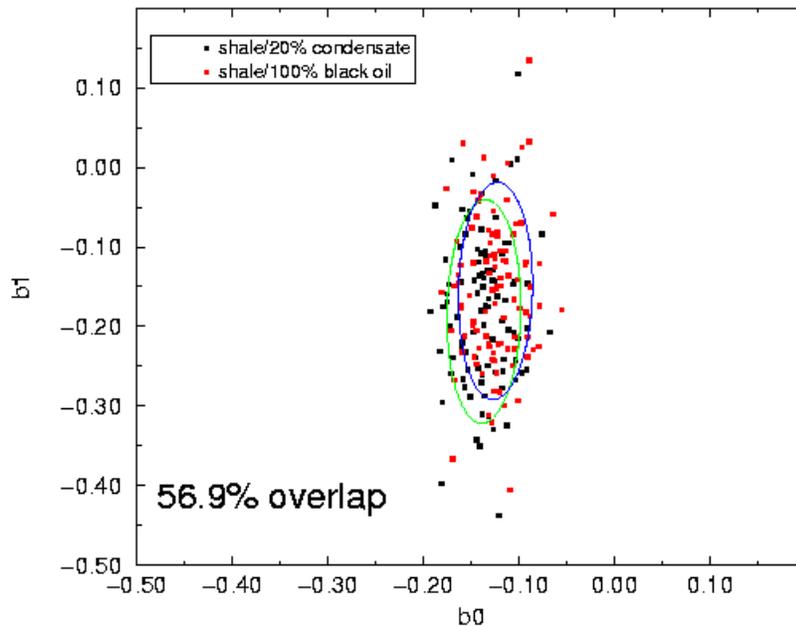
Figure 17: Cross-plot of b_0, b_1 AVO parameters for 20% lean gas condensate and 100% black oil saturation in the sandstone layer. There is a large overlap between confidence regions with a strong amount of interaction.

Figures 12 through 17 indicate that the type of fluid saturation in the sandstone can greatly impact the amount of overlap and interaction that occurs between the confidence regions for various AVO responses. Specifically, AVO can be a useful tool in discriminating between a brine saturation and black oil or lean gas condensate saturation. This, however, is not the case when trying to discriminate between

Contents

different condensate responses or condensate and black oil responses. In these cases, there is a significant amount of overlap between the confidence regions and the interaction between them must be accounted for.

CONCLUSIONS

This project yields results that are consistent with some of the intuitive notions that arise from AVO forward modelling. That is, the best-fit estimate to a confidence region for b_0-b_1 scatter data is an ellipse. Furthermore, the specific fitting of standard deviational ellipses to various scatter data quite adequately approximates the general best-fit ellipse to the data and at far less cost. In addition, one can exploit the geometric properties of these ellipses to quickly and cheaply test to see if a data point has an equal probability of belonging to either response of interest. That is, the data can be efficiently divided into two clear groups; those that unambiguously belong to a certain response and those that have an equal probability of belonging to either response being considered. In addition, the entire process can be applied to any set of data where the underlying objective is to quantify the observed dispersion. Overall, this method can provide a quantitative index for AVO risk assessment. The ultimate goal is to create a general method to best-fit ellipses and calculate overlap that is quick and stable. While there is an elegant and simple analytical theory to achieve this, current technology does not allow this theory to be effectively translated into the discrete language of digital computing.

FURTHER DIRECTIONS

There still exist several avenues that can be explored with this research. A simple but important change that can be made to the current programs is the incorporation some error checking blocks (i.e. insure that the *a*-axis and the *b*-axis are not 0). This will allow for the method to be more numerically stable and computationally sound. A more significant path of investigation is the whole problem of *best-fitting* an ellipse to the scatter data. As stated before, this is not a straightforward problem. Many theoretical and practical problems can arise by trying to use traditional methods. Fitzgibbon, 1996 develops a new method for fitting ellipses to scatter data. Their claim is that the new method:

- is ellipse-specific so that even ill-conditioned data will return an elliptical fit,

- can be naturally solved by a generalised eigensystem ,

- and is extremely easy to implement.

ACKNOWLEDGEMENTS

Contents

(IC^3). In particular, I am grateful to my supervisors Chandra Rai and Carl Sondergeld for this project and all their advice. In closing, I have to thank Ewoud, Richard, Cory, Sverre, Leo, Minh-na, Ki, Eleana, and everybody else at Amoco Summer Camp 1998.

REFERENCES

Davis, P. J. 1973. *The Mathematics of Matrices, Second Edition*. New York: John Wiley and Sons, Incorporated.

Fitzgibbon, A. W., M. Pilu and R. B. Fisher. 1996. Direct Least Squares Fitting of Ellipses. Departement of Artificial Intelligence, University of Edinburgh.

Ebdon, D. 1977. Statistics in Geography: A Practical Approach. Oxford: Basil Blackwell.

Hart, W. L. 1950. *Elements of Analytic Geometry*. Boston: D. C. Heath and Company.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. *Numerical Recipes in FORTRAN: The Art of Scientific Computing, Second Edition*. New York: Cambridge University Press.

Sengupta, M., C. Rai, and C. Sondergeld. 1997. Sensitivity Studies in AVO Forward Modeling. EPTG - Rock Properties, AMOCO CORPORATION.

Sheriff, R. E. and L. P. Geldart. 1995. *Exploration Seismology, Second Edition*. New York: Cambridge University Press.

Telford, W. M., L. P. Geldart, and R. E. Sheriff. 1990. *Applied Geophysics, Second Edition*. New York: Cambridge University Press.

Contents