Selecting velocity models using Bayesian Information Criterion

Tomasz Danek¹, Bartosz Gierlach¹, Ayiaz Kaderali², Michael A. Slawinski² and Theodore Stanoev²*

¹Department of Geoinformatics and Applied Computer Science, AGH University of Science and Technology, Krakow, 30-059, Poland, and ²Department of Earth Sciences, Memorial University of Newfoundland, St. John’s, NL A1C 5S7, Canada

Received February 2021, revision accepted September 2021

ABSTRACT
We present a strategy for selecting the values of model parameters by comparing walkaway vertical seismic profiling data with a multilayered model in the context of Bayesian information criterion. We consider \(P\)-wave traveltimes and assume elliptical polar velocity dependence. A model with different propagation speeds, depending on the angle of propagation, can be a good approximation for a medium composed of thin layers. While elliptical anisotropy in a one-layer model yields good results, an efficient tool for multilayer modelling would provide improved inversion results. To obtain the proper set of velocity values for specific parameterizations, we require two steps of optimization. In the first step, we find the signal trajectory; in the second step, we obtain parameter values by minimizing the misfit between the model and the data. By comparing models and data, we choose the best model in the sense of the Bayesian information criterion.

Key words: Anisotropy, Borehole geophysics, Inverse problem.

1 INTRODUCTION

In this paper, we use Bayesian information criterion (BIC) to select a justifiable parameterization of a model (Schwarz, 1978). To restrict the parameterization from an infinity of models, we use explicit and implicit selection criteria. For the former, as traveltime inversion is one of the most important techniques for extracting information on the Earth’s properties (e.g., Aki and Richards, 2002, Section 9.4), we select four traveltime parameterizations, each with a varying arrangement of layer anisotropy, to be considered in BIC. For the latter, we impose a range of model-parameter values that are consistent with sedimentary basins. We use ray theory to solve a two-stage optimization problem to obtain the model parameters of a multilayer medium consisting of anisotropic vertically inhomogeneous layers. Herein, the anisotropy results in an elliptical polar velocity dependence of a wavefront, which implies azimuthal isotropy about the symmetry axis (Thomsen, 1986). It does not, however, imply transverse isotropy, in the strict sense of a symmetry class. The dataset contains \(P\)-wave traveltimes for a wide range of offsets, which is necessary to examine anisotropy.

The objective of this paper is to establish a strategy for determining the optimal parameterization using BIC for a given model with a specific dataset. The methods described, herein, comprise a fast and reliable approach for seismic model verification by determining the degree of model complexity based on information criteria.

Quantitative analysis of seismic wave propagation is essential in seismic interpretation. Such an analysis is complicated even for relatively simple cases, such as horizontally layered media. Ray theory, which is invoked in this work, provides mathematical tools that simplify the analysis (e.g., Keller, 1978; Červený, 1985; Shearer and Chapman, 1988; Slawinski and Webster, 1999; Wang, 2014; Slawinski et al., 2003, 2004).

The other existing options for inversion are \(P\)-wave only, \(P\)- and SV-wave Vertical Seismic Profiling inversion, but they...
do not provide a good platform for automatic parameterization selection (e.g., Grechka and Mateeva, 2007). The described method can be used to infer anisotropy, wherein only traveltimes are used in computations. Many models can be tested efficiently resulting in a quantitative choice of an optimal model in the context of available data.

2 THEORY

2.1 Elliptical velocity dependence

We obtain signal traveltime in an anisotropic vertically inhomogeneous medium by considering stationary traveltimes within a given velocity model. We consider elliptical anisotropy, which is tantamount to

\[ v(\theta) = \sqrt{v_x^2 \sin^2 \theta + v_z^2 \cos^2 \theta} = v_x\sqrt{(1 + 2\chi)^2 \sin^2 \theta + \cos^2 \theta}. \] (1)

Equation (1) is the elliptical velocity dependence of a wavefront (Slawinski et al., 2004), where

\[ \chi = \frac{v_x^2 - v_z^2}{2v_x^2}, \] (2)

and \( v_x \) and \( v_z \) are the horizontal and vertical speeds. For \( v_x = v_z, \chi = 0 \) and, hence, the wavefront velocity is isotropic. Furthermore, we consider vertical inhomogeneity \( v_\perp := a + bz \), where \( a \) and \( b \) are constant and \( z \) is the depth.

From a point along the upper interface of a layer, at \((x, z = 0)\), to a point along the lower interface, at \((x, z)\), the traveltime is (Rogister and Slawinski, 2005)

\[ t = \frac{1}{b} \left[ \arctanh \left( p bx - \sqrt{1 - (1 + 2\chi)p^2 x^2} \right) \right. \]
\[ + \arctanh \left( \sqrt{1 - (1 + 2\chi)p^2 x^2} \right), \] (3)

where

\[ p = \frac{2\chi}{\sqrt{x^2 + (1 + 2\chi)z^2} \left[ (2a + bz)^2 + (1 + 2\chi)(1 + 2\chi) + b^2 x^2 \right]} \] (4)

is the ray parameter, which is a conserved quantity along the ray. In keeping with SI units, the units for \( a \) and \( b \) are m/s and 1/s, respectively, for speed are m/s, for traveltime are s, and for the ray parameter are s/m. Since \( p^2 = 1/v^2 \), the eikonal equation for elliptical velocity dependence and linear inhomogeneity is

\[ (a + bz)^2 \left[ (1 + 2\chi)p_x^2 + p_z^2 \right] = 1, \] (5)

where \( p_x \) and \( p_z \) are the horizontal and vertical slownesses. If anisotropy vanishes, the problem reduces to a traveltime and horizontal-slowness relationship (Červený, 2001; Slawinski, 2020).

2.2 Ray optimization in multilayered media

We consider the aforementioned model with layer interfaces based on vertical seismic profile measurements (Kaderali, 2009). Each layer is characterized by the values of \( a, b, \chi \). In each layer, the traveltime along a ray is given by expression (3).

We consider a two-step optimization. First, the signal trajectory is optimized for each source-receiver pair to obey Fermat’s principle, for a set of the \( a, b, \chi \) values. Second, these values are adjusted to minimize the misfit between the modelled and measured traveltimes. These steps are repeated until the misfit is at a minimum; the misfit is used in the Bayesian information criterion (BIC) context.

Both steps are performed using the Nelder–Mead simplex method, which is a local optimization. Since it is not based on the gradient, it can be used for nondifferentiable functions. The method works for functions of \( n \) variables, whose values are calculated at \( n + 1 \) points in an \( n \)-dimensional solution space. These points are the vertices of a polyhedron called a simplex. Successive steps of optimization consist of adjusting these vertices according to specific rules. The detailed description of the method can be found in Nelder and Mead (1965).

2.3 Bayesian information criterion

The optimization requires setting the number of parameters \( a \) priori. This number should be chosen to match the resolving power of the data. For that purpose, we use BIC, whose most general form is (Kass and Raftery, 1995, equation 23)

\[ \text{BIC} = -2 \ln L + k \ln M, \] (6)

where \( L \) is maximized likelihood, \( k \) is the number of model parameters and \( M \) is the number of data points, which, herein, is the number of traveltimes. According to Priestley (1982, pp. 375—376), the same minimum value is obtained by minimizing

\[ \text{BIC} = M \ln \hat{\sigma}^2 + k \ln M, \] (7)

where \( \hat{\sigma}^2 \) is the error variance, which, herein, is the normalized mean of squared differences between measured and modelled traveltimes. The model whose BIC value is the least is considered best in terms of balance between agreement with measurements and model complexity. Compared to other criteria, such as Akaike information criterion, BIC results in a bigger penalization for additional parameters (Kass and Raftery, 1995). The BIC method is commonly used in similar studies (e.g., Guo et al., 2011; Danek and Slawinski, 2012).
3 RESULTS

3.1 Data and initial models

The dataset used in this paper consists of Vertical Seismic Profiling measurements from offshore Newfoundland (Kaderali, 2009). The walkaway vertical seismic profile (VSP) data are the basis for the inversion; the zero-offset VSP is used to get the initial model.

For the zero-offset VSP, receivers are in the entire well at 30 m intervals. For the walkaway VSP, there are 200 source locations with 25 m intervals along the NW–SE line. The maximum offset is 4000 m, in the NW direction, and 1000 m, in the SE direction. The receiver array consists of five geophones at depths between 1980 m and 2020 m, with respect to mean sea level. In accordance with Kaderali (2009), the near-offset data, up to 300 m, are removed to avoid problems with solution stability for near-vertical rays. For both types of VSP data, we consider $P$-wave traveltimes only.

Figure 1 is the plot of the smoothed interval velocities, from which we infer a three-layer model from the three distinct velocity gradients. The interfaces, at 1300 m and 1750 m, are used in all computations.

Our simplest model is inhomogeneous, but isotropic, and consists of six parameters $a_i$, $b_i$, where $i = 1, 2, 3$. The most complicated model is inhomogeneous and anisotropic and consists of nine parameters $a_i$, $b_i$ and $\chi_i$, where $i = 1, 2, 3$. In between these two extremes, we consider a seven-parameter model, where the third layer is anisotropic, and an eight-parameter model, where the second and third layers are anisotropic. A similar analysis – for synthetic data – is described by Gierlach and Danek (2018).

3.2 Inversion models

We obtain consistent results for each of the models regardless of their complexities. For example, in Figure 2, we see the results for a seven-parameter model.

Since the chosen optimization method is local, the parameters obtained depend on the initial model. To obtain consistent results, we use a multistart procedure for a wide range of initial values. In other words, the inversion is performed numerous times with randomly chosen initial-model values. The final results correspond to the least misfit. Also, we used the multistart analysis to examine interdependences between parameters, as illustrated in Figure 3.

The multistart analysis diminishes the dependence of results on the initial model. Thus, values obtained can be treated as global extrema. A reasonable initial model is a set of values obtained from the zero-offset VSP with the addition of small elliptical anisotropy for the third layer.

Using expression (7), we calculate the Bayesian information criterion value for each model. As illustrated in Figure 4, the least value is obtained for a seven-parameter model.

4 DISCUSSION

As illustrated in Figure 4, the best model is composed of seven parameters for which only the third layer is anisotropic. With fewer parameters, the Bayesian information criterion (BIC) value increases substantially. An improvement of solution due to more parameters is not sufficient to justify their addition.
Figure 3 Crossplots of values for parameters obtained for the seven-parameter model. The left- and right-hand plots correspond to the first and third layers, respectively; in the latter, \(a\) corresponds to speed at its upper interface. Plotted data points correspond to the top 25\% of results with respect to the residual sum of squares.

Figure 4 BIC values for four models; the model with the lowest value is selected. The values along the vertical axis result from expression (7), whereas the values along the horizontal axis refer to the number of parameters in a model.

It is important to emphasize that our choice of three layers is based on well-log information. One of the features of BIC is that if the true model is within the test set, the probability of choosing it tends to unity while \(n \to \infty\) (Burnham and Anderson, 2002), but in practical applications, with the relatively low number of data points, it is important to limit the set of tested models to avoid spurious solutions.

Furthermore, we can observe relations between certain parameters. In the left-hand plot of Figure 3, the values of \(a\) and \(b\) of the top layer exhibit the correlation coefficient of \(-0.9994\). Consequently, it is impossible to retrieve their individual values, since many of their pairs produce very similar traveltimes. The same phenomenon appears if we assume anisotropy in the top layer. In the right-hand plot of Figure 3, the values of \(a\) and \(\chi\) in the third layer exhibit a correlation coefficient of \(-0.7729\), which makes it impossible to retrieve individual values. The same thing occurs with anisotropy in all layers even though the dependency is weaker.

Anisotropy at specific depths is illustrated by ellipses. In Figure 5, we show such ellipses for different parameterizations. We note that the shapes of these ellipses are identical for the seven-, eight-, and nine-parameter models, due to the consistency of results for each of the models with anisotropy in the bottom layer. This supports an inference about the third layer exhibiting anisotropy, in contrast to \(\chi\) being only a fitting – without a physical meaning – parameter for other layers.

Let us comment on our use of BIC in selecting a model parameterization. In the context of traveltimes alone, a possible parameterization would be a three-layer isotropic model, which would consist of six parameters. However, beyond seismic measurements, we also have geological information, namely the subsurface is composed of shale. Hence, we also include anisotropy in our model. With the introduction of an extra parameter, \(\chi\), BIC suggests that anisotropy should be included in the third layer, within a three-layer model, which results in a seven-parameter model. We accept this model for the following reasons. The isotropy of the first layer can be justified by its position. The shallowest layer is not subject to significant compaction, which results in preferential alignments and, hence, in anisotropy; the compaction increases with overburden, as a function of depth. In contrast, the anisotropy of the third layer can be justified because, as the deepest layer, it is subjected to the most significant compaction due to overburden.
5 CONCLUSIONS

To conclude, let us emphasize that, to a large extent, our model is an epistemological analogy to account for observations. Intrinsically – without external interpretation – there is no ontological claim. In general, Bayesian information criterion (BIC) provides a satisfactory model to account for measurements, but not necessarily a model that corresponds to the physical reality. BIC provides the most empirically adequate model, according to its criteria, even though a more complex model would fit the data better. This is the very purpose of BIC to ensure that the model complexity does not surpass the accuracy of data.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Elena Patarini for her graphical support, David Dalton for reviewing and editing, and the reviewers, Igor Ravve and Alexey Stovas, for their valuable comments that have improved the manuscript. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada, grant 202259 and AGH University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, project 11.11.140.613.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study were derived from the following resources available in the public domain: https://research.library.mun.ca/9189/.

ORCID

Tomasz Danek https://orcid.org/0000-0001-8101-7469
Bartosz Gierlach https://orcid.org/0000-0001-5668-8797
Theodore Stanoev https://orcid.org/0000-0001-8698-2831

REFERENCES

Aki, K. and Richards, P. G. (2002) Quantitative Seismology, 2nd edition. University Science Books.
Burnham, K. P. and Anderson, D. R. (2002) Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd edition. Springer-Verlag.
Červený, V. (1985) The application of ray tracing to the numerical modelling of seismic wave fields in complex structures. Seismic Shear Waves, 15, 1–124.
Červený, V. (2001) Seismic Ray Theory, 1st edition. Cambridge University Press.
Danek, T. and Slawinski, M. A. (2012) Bayesian inversion of VSP traveltimes for linear inhomogeneity and elliptical anisotropy. Geophysics, 77, R239–R243.
Gierlach, B. and Danek, T. (2018) Inversion of velocity parameters in multilayered elliptical anisotropy medium – synthetic data example. E3S Web of Conferences, 66, 01017.
Grechka, V. and Mateeva, A. (2007) Inversion of P-wave VSP data for local anisotropy: theory and case study. Geophysics, 72, D69–D79.
Guo, R., Dosso, S. E., Liu, J., Dettmer, J., and Tong, X. (2011) Nonlinearity in Bayesian 1-D magnetotelluric inversion. Geophysical Journal International, 185, 663–675.
Kaderali, A. (2009) Investigating anisotropy and inhomogeneity using tomographic inversion of VSP traveltimes: validation of analytic expressions for linearly inhomogeneous elliptically-anisotropic models. Master’s thesis, Memorial University of Newfoundland, St. John’s, Canada.
Kass, R. and Raftery, A. (1995) Bayes factors. Journal of the American Statistical Association, 90, 773–795.
Keller, J. B. (1978) Rays, waves and asymptotics. Bulletin of the American Mathematical Society, 84, 727–750.
Nelder, J. A. and Mead, R. (1965) A simplex method for function minimization. The Computer Journal, 4, 308–313.
Priestley, M. B. (1982) Spectral Analysis and Time Series. Academic Press.
Rogister, Y. and Slawinski, M. (2005) Analytic solution of ray tracing equations for a linearly inhomogeneous and elliptically anisotropic velocity model. Geophysics, 70, D37–D41.
Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6, 461–464.
Shearer, P. M. and Chapman, C. H. (1988) Ray tracing in anisotropic media with a linear gradient. Geophysical Journal, 94, 575–580.
Slawinski, M. A. (2020) Waves and Rays in Elastic Continua, 4th edition. World Scientific.
Slawinski, M. A., Lamoureux, M. P., Slawinski, R. A., and Brown, R. J. (2003) VSP traveltime inversion for anisotropy in a buried layer. Geophysical Prospecting, 51, 131–139.
Slawinski, M. A. and Webster, P. S. (1999) On generalized ray parameters for vertically inhomogeneous and anisotropic media. Canadian Journal of Exploration Geophysics, 35, 28–31.
Slawinski, M. A., Wheaton, C. J., and Pwojowski, M. (2004) VSP traveltime inversion for linear inhomogeneity and elliptical anisotropy. Geophysics, 69, 373–377.
Thomsen, L. (1986) Weak elastic anisotropy. Geophysics, 51, 1954–1966.
Wang, Y. (2014) Seismic ray tracing in anisotropic media: a modified Newton algorithm for solving highly nonlinear systems. Geophysics, 79, T1–T7.