Journal of Applied Geophysics 117 (2015) 13–22

Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: www.elsevier.com/locate/jappgeo

Seismic velocity estimation from well log data with genetic algorithms in comparison to neural networks and multilinear approaches

Mattia Aleardi

University of Pisa, Earth Sciences Department, Via S. Maria 53, 56126 Pisa, Italy

A R T I C L E I N F O

Article history:
Received 25 September 2014
Received in revised form 13 March 2015
Accepted 14 March 2015
Available online 17 March 2015

Keywords:
Missing log data prediction
Genetic algorithms
Gibbs sampling
Neural networks
Multilinear regression

A B S T R A C T

Predicting missing log data is a useful capability for geophysicists. Geophysical measurements in boreholes are frequently affected by gaps in the recording of one or more logs. In particular, sonic and shear sonic logs are often recorded over limited intervals along the well path, but the information these logs contain is crucial for many geophysical applications. Estimating missing log intervals from a set of recorded logs is therefore of great interest. In this work, I propose to estimate the data in missing parts of velocity logs using a genetic algorithm (GA) optimisation and I demonstrate that this method is capable of extracting linear or exponential relations that link the velocity to other available logs. The technique was tested on different sets of logs (gamma ray, resistivity, density, neutron, sonic and shear sonic) from three wells drilled in different geological settings and through different lithologies (sedimentary and intrusive rocks). The effectiveness of this methodology is demonstrated by a series of blind tests and by evaluating the correlation coefficients between the true versus predicted velocity values. The combination of GA optimisation with a Gibbs sampler (GS) and subsequent Monte Carlo simulations allows the uncertainties in the final predicted velocities to be reliably quantified. The GA method is also compared with the neural networks (NN) approach and classical multilinear regression. The comparisons show that the GA, NN and multilinear methods provide velocity estimates with the same predictive capability when the relation between the input logs and the seismic velocity is approximately linear. The GA and NN approaches are more robust when the relations are non-linear. However, in all cases, the main advantages of the GA optimisation procedure over the NN approach is that it directly provides an interpretable and simple equation that relates the input and predicted logs. Moreover, the GA method is not affected by the disadvantages that characterise gradient descent techniques such as the NN method.

1. Introduction

The determination of a reliable velocity–depth trend in a particular well location is crucial in exploration geophysics. For example, reliable compressional velocity estimates are needed to tie well data to seismic data (Herron, 2014) or to derive a low-frequency trend for estimating the absolute acoustic impedance from inverted post-stack data (Morozov and Ma, 2009). Velocity logs are needed for efficient amplitude versus angle (AVA) modelling (Mazzotti, 1990; Aleardi and Mazzotti, 2014a) and to calibrate PP and PS wave reflections (Stewart et al., 2002; Gaiser and Van Dok, 2005; Zhang and Wang, 2009). Finally, P- and S-wave velocities are crucial for building rock-physics templates for facies and lithology classification (Avseth et al., 2005; Dvorkin et al., 2014). However, sonic and shear sonic logs are often recorded over limited depth intervals along the well path due to the limited budget for acquiring such logs. Therefore, it is important to reliably estimate P-wave and S-wave velocities in missing log intervals. This task can be accomplished by determining a specific relation that links the velocity logs with other recorded logs (e.g., gamma ray, resistivity, density logs). Once these relations are known, they can be used to predict seismic velocities for the depth intervals in which data are missing or, assuming negligible lateral variations in petrophysical properties, to predict seismic velocities in nearby wells. However, the non-linearities in the relations that link seismic velocities to the other log data often make explicit evaluations difficult. To address this issue, linear relations between the seismic velocities and the other rock properties are usually assumed (Pickett, 1963; Han et al., 1986; Castagna et al., 1993; Mazzotti and Zamboni, 2003) and linear regression methods have been also used for estimating missing well log data (Jain and deFigueiredo, 1982). However, the linearity assumption is frequently violated in real cases being the relationships between different log data or petrophysical rock properties, in most of the cases, nonlinear. For example, it is well known that non-linear relations link the P-wave velocity to water saturation and the P-wave velocity to clay content (Eberhart-Philipps et al., 1989; Avseth et al., 2005).

The recent development of computer-based intelligence methods has enabled researchers to accurately address the non-linearity that characterise these optimisation problems. In particular, the artificial

E-mail address: mattia.aleardi@dst.unipi.it.

http://dx.doi.org/10.1016/j.jappgeo.2015.03.021
0926-9851/

neural network (NN) method has received the most attention in exploration geophysics. In general, NN technology helps to identify nontrivial correlations between geophysical data. This method has been widely applied in many geophysical problems, such as wavelet estimations (Wang and Mendel, 1992), velocity analysis (Calderón-Mac ı´ as et al., 1998), automatic horizon picking (Huang, 1997), seismic facies classification (Coléou et al., 2003; Herrera et al., 2006; Marroquín et al., 2008) and to relate seismic-derived attributes to reservoir properties (e.g., to relate seismic attributes to porosity logs as in Dorrington and Link, 2004). The NN method has also been widely applied for predicting missing log intervals, namely, to identify a specific relation that links a set of input logs with other logs (Arandia et al., 2001; Poulton, 2002; Srisutthiyakorn, 2012). However, notwithstanding its many successful applications in exploration geophysics, the classic NN approach suffers from many limitations and drawbacks. In particular, as discussed in Saggaf et al. (2003) and Van der Baan and Jutten (2000), the NN method is primarily limited by its gradient-based nature.

In this study, I attempt to overcome the drawbacks and limitations of the NN method by applying a different method, the genetic algorithm (GA), to determine the linking relation between a set of input logs and a desired log. In particular, I focus on predicting P- and S-wave velocities from a set of input logs. Similar to the classical NN approach, the proposed methodology is based on the fact that seismic velocities are related to the rock's petrophysical properties, such as texture, mineralogy, saturation and pore fluid content. However, other well logs are also dependent on these petrophysical properties, and thus, the GA can identify quantitative relations between the recorded logs without requiring any a priori information. Therefore, linear and/or exponential relations are assumed to relate the seismic velocities to a set of input logs, and the coefficients that characterise these relations are determined by performing a GA optimisation.

The aim of any inversion or optimisation process should not only be to find an optimal solution but also to quantify the uncertainties that affect the final result (Sen and Stoffa, 2013). Geophysical inverse problems suffer from non-uniqueness; that is, many solutions fit the observed data equally well (Tarantola, 2005). To address this problem, geophysical inverse problems are often cast in a statistical framework (Duijndam, 1988) in which the solution to an inverse problem is not only a single set of predicted model parameters but is also represented by a posterior probability density function in the model space. However, similar to other global search algorithms, the GA is not a Markov Chain Monte Carlo (MCMC) method and does not honour the principle of importance sampling (Rubinstein and Kroese, 2011). Therefore, a biased posterior probability distribution is estimated if it is computed directly from the set of GA sampled models and their associated likelihoods (Sen and Stoffa, 1996). In particular, the GA method has been shown to underestimate the variance and thus the uncertainty that is associated with each inverted parameter (Sen and Stoffa, 1996; Aleardi and Mazzotti, 2014b). As discussed in Sen and Stoffa (1996), MCMC methods can only be applied in cases with a limited number of unknowns (no more than four or five) due to their high computational cost. Therefore, several methods have been developed to obtain a reliable and unbiased estimate of the posterior distributions after a GA inversion (e.g., Sen and Stoffa, 1996; Mallick, 1999; Hong and Sen, 2009). In the present study, I follow the strategy proposed by Sambridge (1999) in which a stochastic global search algorithm is combined with a subsequent resampling of the explored portion of the model space according to an MCMC method such as the Gibbs sampler (Geman and Geman, 1984). Therefore, this approach attempts to combine the speed of GA in finding an optimal solution with the accuracy of a subsequent GS to obtain a reliable estimate of the posterior probability distribution. Once the uncertainties in the coefficients in the estimated relation have been derived, the uncertainties are then propagated into the predicted seismic velocities via a Monte Carlo simulation (see Avseth et al., 2005 for many examples of Monte Carlo simulations in exploration geophysics).

A brief theoretical overview of genetic algorithms introduces the proposed methodology. In the section that follows, neural networks and multilinear methods are succinctly described as their results will be compared to those of the proposed GA methodology. Three real log data examples, discussed in detail, constitute the main body of the paper. The final section contains the comparison of the GA results with the NN and the classical multilinear regression results.

2. A brief introduction to genetic algorithms

Genetic algorithms are global stochastic methods that were developed by Holland (1975) and belong to a larger class of evolutionary algorithms. GAs are based on the mechanics of natural selection and evolution (the “survival of the fittest” Darwinian principle) to search a model space for optimal solutions. The optimisation process is driven by three main genetic operators: mutation, cross-over and selection. In a genetic algorithm, a population of strings (called chromosomes), which encode candidate solutions (called individuals or phenotypes) to an optimisation problem, evolves towards better solutions during the evolution process, which starts from a population of randomly generated individuals. In each generation, the fitness (the goodness of each possible solution) of each individual is evaluated, and multiple individuals are then stochastically selected from the current population based on their fitness. These individuals are then modified (using crossover and mutation operators) to form a new population, which is used in the next iteration. The algorithm terminates when either a maximum number of generations have been produced or a satisfactory fitness level has been reached in the current population. In the subsequent tests, the number of individuals in the initial population is set to ten times the number of unknowns, and the maximum number of iterations is 50. In each iteration 90% of the parents are selected for reproduction and mutation. The mutation rate (the probability of mutating a variable) is fixed at 10%, whereas the stochastic universal sampling selection method and a linear ranking are used for the selection. A fitness-based elitist reinsertion method is applied to replace the parents with the generated offspring. More information and details about GA can be found in Goldberg (1989) and Mitchell (1996).

3. The proposed methodology

Shear and compressional wave velocities depend on rock properties, such as effective pressure, water saturation, mineralogical composition, porosity, and other parameters. Well logs inherently contain information describing the reservoir properties (e.g., gamma ray and resistivity logs provide lithological and pore fluid type information, respectively). Thus, these logs can be used to estimate the P- or S-wave velocities. Moreover, if we assume that the S- or P-wave velocities are related to each log by a linear or exponential relation, we can write:

$$Velocity = \sum_{n=1}^{k} a_n Log_n{}^{b_n} \tag{1}$$

where the *Log* variable represents the *nth* log used in the prediction procedure. The weight of each input variable is given by the coefficient *a*, whereas the exponent *b* is used to reproduce the effects of variations in the input log on the seismic velocity. Eq. (1) can be seen as a generalization of classical depth-trends (for example the well-known Gardner equation; Gardner et al., 1974) and similar equations have also been used by other authors such as Banchs et al. (2001).