PETROPHYSICS, VOL. 59, NO. 6 (DECEMBER 2018); PAGES 826–840; 14 FIGURES; 4 TABLES. DOI: 10.30632/PJV59N6-2018a7

Comparative Study of Shallow Learning Models for Generating Compressional and Shear Traveltime Logs[1]

Jiabo He[2], Siddharth Misra[2], and Hao Li[2]

ABSTRACT

Compressional and shear traveltime logs (DTC and DTS, respectively) acquired using sonic logging tools are used to estimate connected porosity, bulk modulus, shear modulus, Young's modulus, Poisson's ratio, brittleness coefficient, and Biot's constant for purposes of geomechanical characterization. We propose a data-driven technique to synthesize DTC and DTS logs in the absence of a sonic logging tool. Six shallow learning methods, namely ordinary least squares (OLS), partial least squares (PLS), least absolute shrinkage and selection operator (LASSO), elastic-net regularization, multivariate adaptive regression splines (MARS) and artificial neural network (ANN), suitable for function approximation problems, are trained and tested to synthesize DTC and DTS logs. To that end, the six shallow learning models process 13 conventional and easy-to-acquire logs, namely lithology, gamma ray, caliper, density porosity, neutron porosity, photoelectric factor, bulk density, and resistivity at six depths of investigations. A total of 8,481 observations along a 4,240-ft depth interval in a shale reservoir were available for the proposed data-driven application. The ANN algorithm performs the best among the six algorithms. ANN-predicted DTC and DTS logs have coefficients of determination (R^2) of 0.87 and 0.85, respectively, for Well 1. The next best prediction performance is provided by the MARS-predicted DTC and DTS logs, with accuracies of 0.85 and 0.83, respectively. PLS- and OLS-predicted DTC and DTS have accuracies measured by R^2 of 0.83 and 0.80, respectively, whereas the LASSO- and elastic-net-predicted DTC and DTS have accuracies of 0.79 and 0.75, respectively. The prediction performances of the six algorithms for DTC are always better than those for DTS. The ANN model trained in Well 1 is deployed in Well 2, which was drilled in the same reservoir. The ANN-predicted DTC and DTS logs in the 1,460-ft depth interval of Well 2 have R^2 of 0.85 and 0.84, respectively.

INTRODUCTION

Sonic logging tools transmit sonic waves that propagate from one or more sources to multiple receivers. Compressional and shear slowness logs (DTC and DTS, respectively) can be estimated from the waveforms recorded at the receiver, and both the logs have a high degree of correlation (Willis and Toksoz, 1983). Compressional waves are longitudinal waves while shear waves are transverse waves. The compressional-wave traveltime depends upon the elastic properties of the rock (matrix plus fluid) and varies, depending upon the composition and microstructure of the matrix, the type and distribution of the pore fluid and the porosity of the rock. The shear wave cannot propagate in fluids, as fluids do not behave elastically under shear deformation. DTC and DTS values are most affected by porosity and tend to have large values in formations with large porosity. DTC and DTS are used to estimate various reservoir properties for geomechanical characterization, such as porosity, various moduli (bulk modulus, shear modulus and Young's modulus), and Poisson's ratio.

Many researchers predicted DTS (or V_s, the reciprocal of DTS) using DTC (or V_p, the reciprocal of DTC) and other conventional logs as inputs. Empirical equations were used to estimate DTS from lithology, porosity, and S_w and estimate DTC from two sets of logging data in offshore Gulf of Mexico clastics with an accuracy of 0.81 and 0.76, in terms of R^2, respectively (Greenberg and Castagna, 1992). Intelligent systems, including fuzzy logic (FL), neuro-fuzzy (NF) and artificial neural network (ANN) algorithms predicted DTS successfully with neutron porosity (NPOR), DTC, gamma ray (GR), bulk density (RHOZ), and deep laterolog resistivity as inputs in the sandstone reservoir of Carnarvon Basin in Australia (Rezaee et al., 2007) with a

Manuscript received by the Editor May 23, 2018; revised manuscript received July 28, 2018, manuscript accepted September 5, 2018.
[1]Originally presented at the SPWLA Spring Topical Conference on Petrophysical Data-Driven Analytics: Theory and Applications, Houston, Texas, USA, April 15–17, 2018.
[2]Mewbourne School of Petroleum & Geological Engineering, University of Oklahoma, Norman, Oklahoma 73019-1003, USA.

mean squared error (MSE) of 0.05. Empirical correlations and machine-learning methods were used to predict DTS in a carbonate reservoir in Iran (Maleki et al., 2014). Empirical correlations use DTC to calculate DTS, such as the Castagna equation, the Brocher equation and the Carroll equation, with accuracies of 0.92, 0.86 and 0.88 in terms of R^2, respectively. Support-vector machine (SVM) and ANN predicted DTS using DTC, RHOZ, GR, effective porosity (PHIT), true formation resistivity (R_t) and DCAL as inputs with accuracies of 0.94 and 0.88 in terms of R^2, respectively.

Few researchers have simultaneously predicted DTC and DTS. Empirical equations were implemented to derive DTC and DTS using GR, lithology and porosity in the Oregon basin, Wyoming (Iverson and Walker, 1992). The Xu-White model was proposed to estimate the velocity of the compressional wave and shear wave (V_p and V_s) of shaly sandstones from porosity and shale content (Xu and White, 1995). A key step in Xu-White model requires the estimation of dry-rock bulk and shear moduli for the sand/shale mixture, and its computation time can be improved by assuming constant Poisson's ratio (Keys and Xu, 2002). Numerical experiments showed a close match, less than 2% in terms of relative error, between velocities obtained with the approximations and those computed with the original differential-effective-medium method in Xu-White model. In another work, the Thomas-Stieber approach to petrophysical analysis of thin beds and the Dvorkin and Gutierrez sand/shale rock physics models were applied to predict V_p and V_s in thin beds using mineral properties, rock porosity and shale volume (V_{sh}) (Baines et al., 2008). In the Asmari carbonate reservoir, a committee machine, together with intelligent systems (CMIS) was used to simultaneously predict V_p, V_s, and Stoneley-wave velocity (V_{st}) by processing NPOR, RHOZ, R_t and V_{sh} (Asoodeh and Bagheripour, 2012). Fuzzy-logic (FL), NF, and ANN models were included in CMIS and the system predicted V_p, V_s and V_{st} with accuracy, in terms of R^2 equal to 0.93, 0.89 and 0.74, respectively.

To the best of our knowledge, there has not been a comparative study of prediction performances of shallow learning models for sonic log synthesis. In this paper, six shallow learning models were used to process 13 conventional and easy-to-acquire logs to synthesize DTC and DTS logs in a shale reservoir. In comparison to deep-learning methods (Li and Misra, 2017a, 2017b), the proposed task is suitable for shallow learning models because of limited data size and limited dimensionality of the synthesis task. We study the sensitivity of the prediction performances of the models to noise in inputs, noise in outputs, missing data, and the size of training data. Ordinary-least-squares (OLS), partial-least-squares (PLS), least absolute shrinkage and selection operator (LASSO), and elastic-net models are four linear-regression models that exhibit lower prediction accuracy for the proposed sonic log synthesis, while multivariate adaptive regression splines (MARS) and ANN take nonlinearities into consideration and exhibit higher prediction accuracy. This study will enable engineers and geoscientists to obtain improved geomechanical characterization when a sonic logging tool is not available due to operational or financial constraints. Importantly, it can be used as a tool to extract the hidden relationship between conventional 'easy-to-acquire' measurements and the shear/compressional travel times.

DATA PREPARATION AND PROPROCESSING

Data preparation

In this paper, six shallow learning models are applied to a shale reservoir. Gamma-ray log (GR), caliper log (DCAL), density-porosity log (DPHZ), neutron-porosity log (NPOR), photoelectric factor log (PEFZ), bulk-density log (RHOZ), and laterolog resistivity logs at six depths of investigation (RLA0, RLA1, RLA2, RLA3, RLA4, RLA5) are selected as the easy-to-acquire conventional logs fed into the six machine-learning models. Those input logs (Fig. 1, Tracks 2 to 6,) and DTC and DTS logs (Fig. 1, Track 7) are used to train and test these models. One synthetic discrete-valued log, referred to as lithology flag, is obtained from the lithology information of specific depth intervals. The formation is divided into several intervals according to the dominant lithology. The flag is assigned an integer value between 1 and 13 along the 4,240-ft depth interval. In summary, 13 inputs and 2 output logs were used for the training and testing purposes. There are two main criteria for selecting the logs as inputs: (1) the input logs and sonic logs should share similar petrophysical properties that govern them, and (2) the easy-to-acquire input logs should be available in most of the wells as a part of the conventional logging plan. For example, DPHZ and NPOR are sensitive to the porosity of the reservoir, which also affects the DTC and DTS logs. More information regarding the logs is available in Tathed et al. (2018).

Data Preprocessing

Before using logging data to train models, data preprocessing is necessary to make input and output logs more suitable for purposes of prediction. No obvious outliers are detected in logging data. Normalization is necessary to transform all the data to the range of -1 to 1. This preprocessing method guarantees stable convergence of weights and biases of models. Normalization is performed using the following equation:

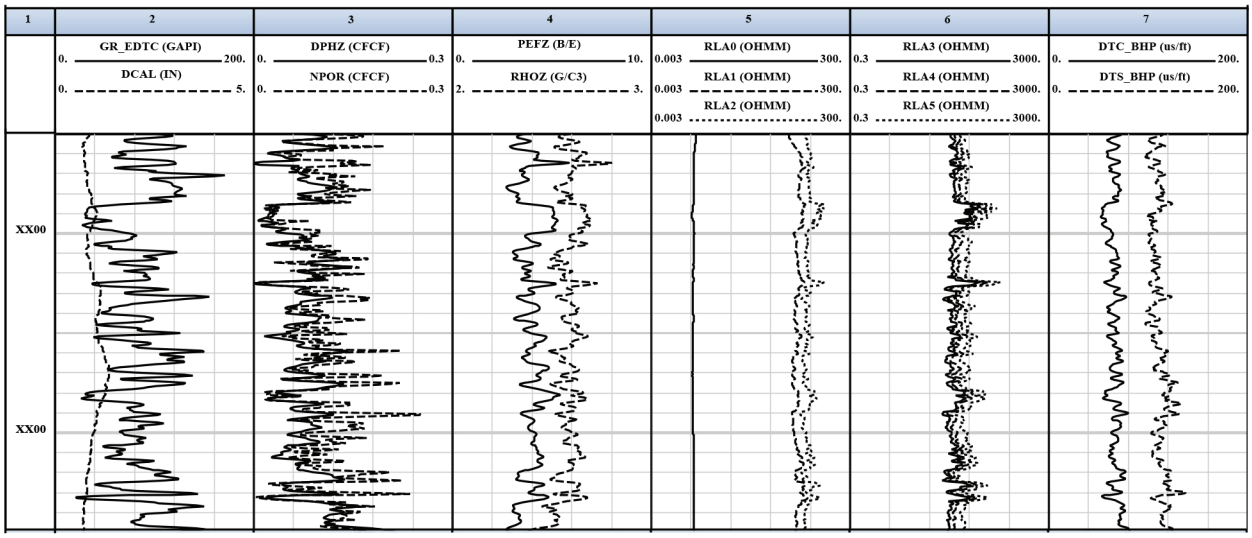

Fig. 1— These logs are used for training and testing the machine-learning models. Track 1, depth; Track 2, gamma ray and caliper logs; Track 3, density-porosity and neutron-porosity logs; Track 4, formation photoelectric factor and bulk-density logs; Track 5, laterolog resistivity logs at shallow depths of investigation (RLA0, RLA1, RLA2); Track 6, laterolog resistivity logs at deep depths of investigation (RLA3, RLA4, RLA5); and Track 7, DTC and DTS logs for a 200-ft section of the formation.

$$y'_i = 2\frac{y_i - y_{min}}{y_{max} - y_{min}} - 1, \quad (1)$$

where y_i is the original value of a log response and y'_i is the normalized value of the log response at each depth i. In this study, the models are developed (i.e., trained and tested) using well logs acquired in Well 1, and after model development, the models are deployed in Well 2 to demonstrate the prediction performance on a new dataset from a different data source. We use the well-log data in Well 1 to train and test the models, and use well logs from Well 2 to deploy and validate the trained models. When building the predictive models, the dataset from Well 1 is split into two parts: training data and testing data, such that 80% of data are randomly selected as training data and the remaining 20% form the testing data for all the models investigated in this paper. This is a common practice when model training and testing are based on a supervised learning technique. With the well logs from Well 2, we perform blind-test validation of the trained models.

Measurement of Prediction Performance

The correlation coefficient (R^2) is used to compare the prediction performance of all models, which is formulated as

$$R_j^2 = 1 - RSS_j/TSS_j, \quad (2)$$

where

$$RSS_j = \sum_{i=1}^{n}\left(y_{pi,j} - y_{mi,j}\right)^2, \quad (3)$$

and

$$TSS_j = \sum_{i=1}^{n}\left(y_{pi,j} - \bar{y}_j\right)^2, \quad (4)$$

such that n is the number of depths for which prediction needs to be performed, $j = 1$ indicates the DTC log and $j = 2$ indicates the DTS log, $y_{pi,j}$ is the sonic response predicted at depth i, $y_{mi,j}$ is the log j response measured at depth i, and $\bar{y}_j$ is the mean of log j responses measured at all depths for which training or testing are being performed. RSS_j is the sum of squares of the residuals and TSS_j is the total sum of squares proportional to the variance of the corresponding log j responses.

Feature Selection

We applied a univariate feature selection process to statistically evaluate the relationship between the input logs and output sonic logs. The feature selection process calculates the p-value between each input logs and the output sonic logs. The p-value indicates the probability that the output is under null hypothesis. In other words, if the p-value is high, it indicates the information contained in the output sonic logs is very likely not contained in the input log. The p-values calculated between each input log and two output sonic logs are shown in Fig. 2.

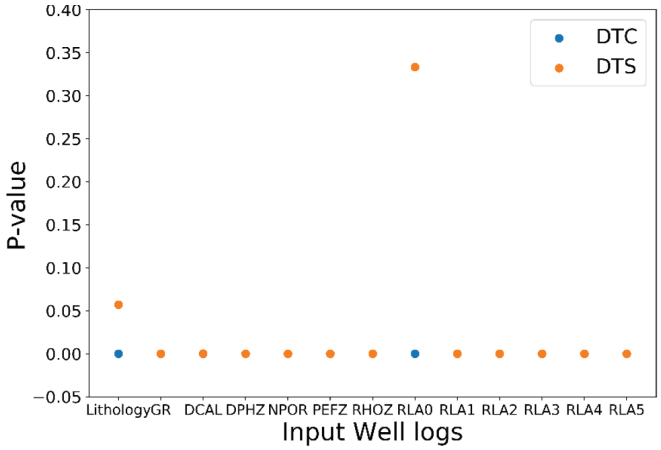

Fig. 2—*P*-values between each input log and two output sonic logs. The *p*-value indicates the possibility of a null hypothesis, such that the input logs with low *p*-values are important for the proposed synthesis of DTC and DTS logs.

For DTC, all the *p*-values of input logs are near zero. This indicates all the input logs contain information required for accurate reconstruction of DTC. For DTS, lithology log and RLA0 input logs the *p*-value are 0.06 and 0.33, respectively. The *p*-values for other logs are near zero. Although RLA1 has relative high *p*-value, its absolute value is still low. To keep a consistent input dimension when predicting both DTC and DTS, we chose to keep all the 13 input logs.

METHODOLOGY

Six shallow learning models are implemented for the comparative study of their prediction performances when generating the DTC and DTS logs by processing easy-to-acquire conventional logs. The first five models are regression models with different learning algorithms. The sixth model is a simple neural network model.

Ordinary-Least-Squares (OLS) Model

The OLS model is a linear-regression model that optimizes the coefficient β_i corresponding to each input x_{ij} (Draper and Smith, 2014). Coefficients are calculated by minimizing the sum of squared errors (SSE) between original outputs y_i and predicted outputs $\hat{y}_i$ of the model. The function of the OLS model is formulated as

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad (5)$$

where y_i are original outputs of the model, i is an integer ranging from 1 to n indicating distinct depth points in the given log dataset, n is the total number of observations (depth points along the length of the well where the logs were acquired), x_{ij} is the j-th input log, which will be processed by the OLS model, measured at depth i, p is the number of input logs processed by the OLS model (easy-to-acquire conventional logs), j ranges from 1 to p, β is coefficient vector of the OLS model, and ε is the error term. β and ε are determined by the OLS algorithm in the model. The loss function SSE is formulated as

$$\text{SSE} = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2, \quad (6)$$

$$\hat{y}_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}, \quad (7)$$

where $\hat{y}_i$ is predicted output of the model for the depth i. A smaller SSE leads to better prediction performance. OLS model is such a simple model that people can use it for well log prediction easily and quickly. However, due to its simplicity, it also has few weaknesses, as described in the result section. The OLS model will be unduly influenced by outliers if they cannot be deleted before model construction. Besides, it is impossible for the OLS model to detect correlations between inputs and to reduce the dimensionality of the problem before the model construction; thereby, adversely affecting the computation time.

Partial-Least-Squares (PLS) Model

The PLS model is also a regression model that first constructs latent structures explaining the most variation in dependent variables and then uses the latent structures to explain the maximum variation in the dependent variables (Draper and Smith, 2014), thereby minimizing the covariance of input variables. New latent structures are combined with the original variable to form components. The number of components m (no more than the number of inputs) is chosen to maximally summarize the covariance with outputs. The determination of m builds the PLS model with fewer inputs than the OLS model, and is suited when there are high correlations among inputs. The removal of redundant inputs can save computational time for the training the model while simplifying the model's structure. The flowchart of the PLS algorithm design is shown in Fig. 3.

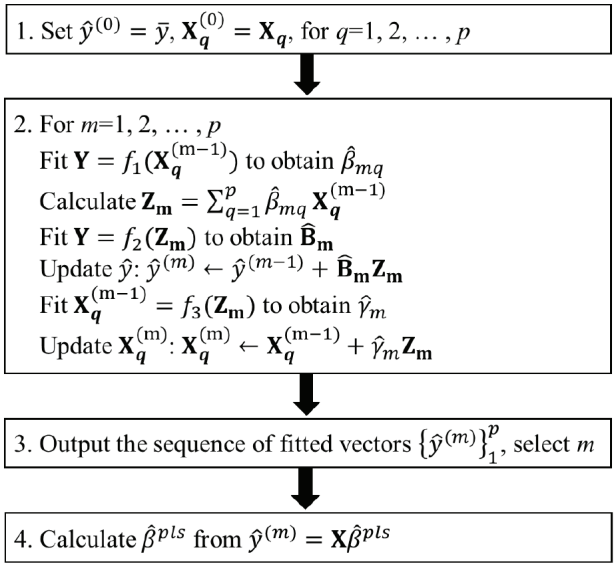

Fig. 3—Flowchart of the PLS method to find the best representation of the variation of the output logs in terms of the input logs in a latent space.

In Fig. 3, **X** is the input logs matrix, **Y** is the original sonic output matrix, m is the number of components to be generated, p is the number of inputs, β is the coefficient of the PLS regression model, Z_m is the iteration term, and $\hat{\beta}$ *and* $\hat{y}$ are the iteration coefficients. In the PLS algorithm, $\hat{\beta}$ and $\hat{y}$ are optimized for each m and the prediction performance for each m is measured and recorded for comparison. The smallest m with the best prediction performance is used to build the model. In the PLS model, sonic logs are generated with constructed components, which are combinations of latent structures of original inputs. The latent structure corresponding to the most variation in **Y** is extracted and then explained using a latent structure in **X**.

The PLS model will also outperform other regression models, in some cases where the number of inputs is larger than that of observations (Draper and Smith, 2014). However, it may lose useful information for prediction when constructing the components, which leads to poorer prediction performance than other models. For purposes of our study, Fig. 4 shows the result of tuning the number of components. When the number of components is equal to 13, the best prediction performance is obtained. The results show that there are no obvious redundant inputs in the PLS model. When m = 13, the prediction accuracy is the same as that of the OLS model, which is 0.83 and 0.80 for DTC and DTS predictions, respectively, in terms of R^2. For a lower number of components, when m = 8, the prediction accuracy of the PLS model will drop to 0.80 and 0.78 for DTC and DTS predictions, respectively.

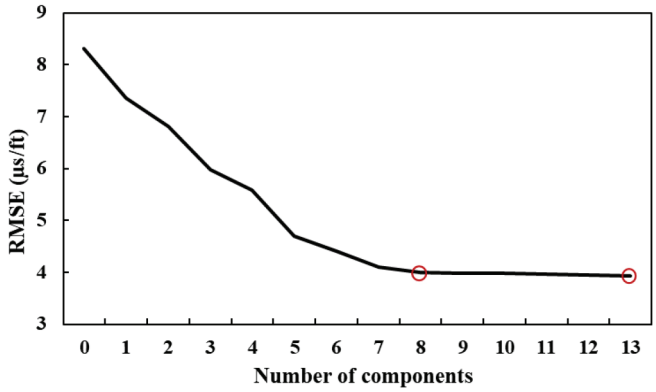

Fig. 4—Change in performance of the PLS model when tuning the number of components in the model. For a number of components higher than eight (m > 8), the RMSE flattens out and the performance of the model does not drastically improve with further increase in number of components.

Least Absolute Shrinkage and Selection Operator (LASSO) model

The LASSO model is a regression model that combines the SSE with a penalty term λ (Kuhn and Johnson, 2013), which is introduced to constrain the number of nonzero coefficients. The loss function of LASSO model is formulated as

$$L = \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda\sum_{q=1}^{p}|\beta_q| = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda\sum_{q=1}^{p}|\beta_q|. \quad (8)$$

As λ increases, the number of nonzero coefficients in the coefficient vector $\beta = \{\beta_1,\beta_2,...\beta_p\}$ corresponding to the input vector x_i will decrease, which eliminates the negative effect of redundant inputs when building the model. More information regarding the penalty term and regularization is available in Han and Misra (2018). Unlike the PLS model, which constructs new components for better prediction performance, the LASSO model sets the coefficients of a few of the redundant input variables to zero. Redundancy in inputs (multicollinearity) results in coefficient estimates for the model to be erratically sensitive to the changes in model and data, without necessarily affecting the model predictions. However, when there are no redundant inputs, the LASSO model can fail to fully use all inputs for the best prediction performance unless λ is tuned to zero. Decreasing λ will increase $\Sigma_{q=1}^{p}|\beta_q|$ in the model (Fig. 5). In this case, when λ < 4.83, $\Sigma_{q=1}^{p}|\beta_q|$ will increase significantly. So, we choose an optimum value of λ of 4.83. For λ = 4.83, the prediction accuracy of the LASSO model in terms of R^2 is 0.79 and 0.75 for DTC and DTS predictions, respectively. Notably, for the shallow and deep resistivity logs the coefficients are zero, indicating redundancy in the resistivity logs of various depths of investigation. Similar observations can be made for the density log that tends to correlate well with the neutron-porosity log.

Table 1—Estimates of Coefficient β_q in the LASSO Model When λ = 4.83

Lithology	GR	DCAL	DPHZ	NPOR	PEFZ	RHOZ	RLA0	RLA1	RLA2	RLA3	RLA4	RLA5
0.31	0.06	-1.29	0.00	41.63	0.60	-55.51	0.00	0.00	-1.01	-1.05	0.00	0.00

Elastic-Net Model

The elastic-net model is a generalization of the LASSO model (Kuhn and Johnson, 2013), which introduces two penalty terms (λ_1 and λ_2) in the loss function

$$L = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \lambda_1 \sum_{q=1}^{p}|\beta_q| + \lambda_2 \sum_{q=1}^{p}\beta_q^2. \quad (9)$$

The elastic-net model is a better model than the LASSO model for high-dimensional data with highly correlated variables when certain groups of correlated variables cannot be neglected. Unlike the elastic-net model, the LASSO model neglects all correlated variables in a group except one. Two parameters should be tuned when building this model. In this case, λ_1 is tuned to be 4.8, which is close to λ in the LASSO model and λ_2 is tuned to be 0.1. The two parameters are tuned by selecting the best performing combination out of the several different combinations of values. The prediction accuracy of the elastic-net model in terms of R^2 is 0.79 and 0.75 for DTC and DTS predictions, respectively.

Multivariate Adaptive Regression Splines (MARS) Model

The four models discussed so far are linear-regression models that linearly combine inputs using a single-regression model, whereas the MARS model builds multiple linear-regression models across the range of input values (Kuhn and Johnson, 2013). The MARS model fits separate linear regressions for distinct continuous regions of the input space. The number of distinct regions (partitions) where regressions need to be performed is estimated in the model. The combination of linear regressions for the different regions generates the predicted output; thereby better accounting for the nonlinearities compared to the four previously discussed models. The function of the MARS model is formulated as

$$\hat{y}_i = \sum_{q=1}^{p} \alpha_q B_q(x_{qi}), \quad (10)$$

where x_{qi} is the value of q-th input log x_q at the i-th depth point, $B_q(x_{qi})$ is a hinge function and α_q is the coefficient of $B_q(x_{qi})$. A hinge function B_q has the form: max $(0, x_{qi} - C_q)$ and max $(0, C_q - x_{qi})$, where C_q is called the knot of the hinge

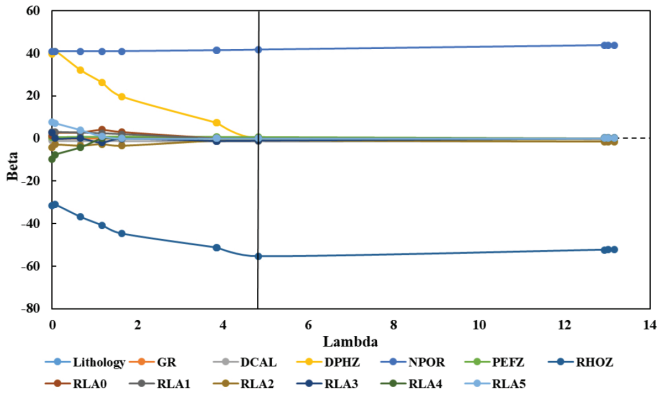

Fig. 5—Variation in the coefficient values (β_q) for various input features of the LASSO model with increase in the penalty parameter (λ) of the regularization term. Tuning λ in the LASSO model results in removing colinear features by pushing the corresponding coefficient values to zero. The optimum value of λ is selected as 4.83 for the LASSO model.

function, which is a constant for a specific hinge function. There is one term C_q for each hinge function B_q belonging to a specific partition of the input space. MARS model estimates the number of partitions k, α_q and knot C_q of a hinge function for the best prediction performance in the case. There is one term (the constant) for the input instead of the hinge function if the input is not used in the MARS model and there are two terms (α_{q1} max (0,x− C_q) or α_{q2} max (0, C_q − x) for the input if it is used in the MARS model. The MARS model performs better than linear-regression models when there are strong nonlinearities between inputs and outputs. However, it can also be negatively influenced by outliers.

The MARS model for our study implements 21 terms (two terms for each hinge function corresponding to an input log; 10 input logs; one constant term for the model, and nine partitions) are tuned for the highest prediction accuracy. The MARS model estimates DTC and DTS at a prediction accuracy measured by R^2 of 0.85 and 0.83, respectively. Although the three inputs, namely RLA0, RLA4, and RLA5, have high correlations with other inputs, namely RLA1, RLA2, and RLA3, and are not used in the MARS model.

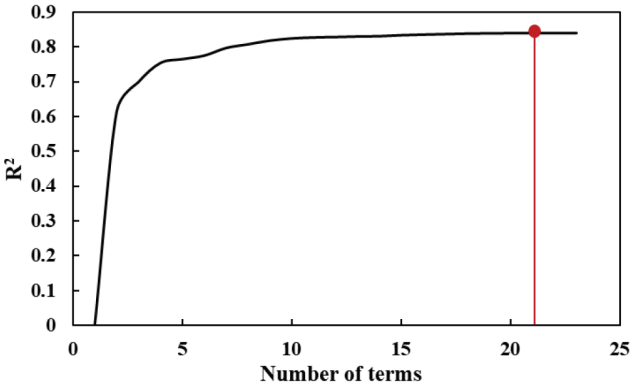

Fig. 6—Variation of the MARS model performance when tuning the number of terms in the MARS model. The red dot is the highest R^2 value achieved by the model beyond which there is no significant improvement in performance. The red line shows the corresponding number of terms, equal to 21, for which the best predictive performance was achieved.

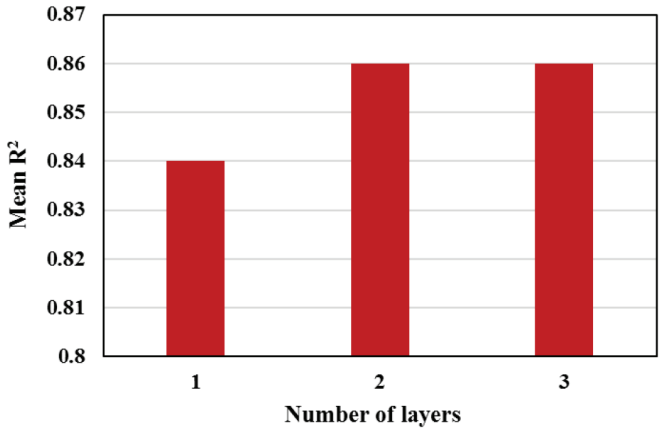

Fig. 7—Comparison of prediction performances of various ANN models with different number of hidden layers. The accuracy is measured by the coefficient of determination, R^2. ANN models with two and three hidden layers have similar predictive performances.

Artificial Neural Network (ANN) Model

The ANN model is a widely used machine-learning model. ANN models implement various architectures and learning algorithms. In the ANN model, the first layer is an input layer and the last layer is an output layer. The number of hidden layers in the middle and the number of neurons in each hidden layer should be decided according to the number of inputs, outputs and the complexity of the problem. An arithmetic sequence of the number of neurons in each hidden layer generates high prediction accuracy (Hagan at al., 2014). In our case, there are 13 inputs and two outputs for sonic log prediction, two hidden layers is set in our model, which is enough for function approximation problems. There are nine neurons in the first hidden layer and five neurons in the second hidden layer. The specific numbers of neurons in the hidden layers are chosen to ensure a smooth transition from input layer to output layer. The ANN model with three hidden layers does not outperform the one with two hidden layers, as shown in Fig. 7. There is not any mandatory number of neurons; based on our experience with the proposed task, we chose the specific neural network architecture and the arrangement of neurons that provided reliable results. A user can choose any other combination of neurons in the two hidden layers but needs to test the reliability of the chosen architecture.

Several algorithms can be used as the training functions to adjust weights and biases of the neurons for minimizing certain loss functions of the ANN model. The best-performing ANN model requires the minimization of the loss functions using the training functions. Levenberg-Marquardt (LM) backpropagation (Chamjangali et al., 2007) and conjugate gradient (CG) backpropagation (Cheng et al., 2005) are the two most widely used training functions. CG backpropagation is applied as the training function after a comparison of training time with the two algorithms. The CG algorithm is more efficient than LM, and thus is implemented in this study. Overfitting can occur when minimizing the loss function. The ANN model cannot identify generalizable relationships between inputs and outputs because of overfitting. The following loss function is designed to avoid overfitting, which is expressed as

$$L = \sum_{i=1}^{n}(y_{ij} - \hat{y}_{ij})^2 + \lambda \sum_{j=1}^{2} \sigma_j^2, \quad (11)$$

where n is the number of depths in a well, j = 1 and 2 represents DTC and DTS logs, respectively, λ is the penalty parameter, y_{ij} is the sonic log j measured at depth i, $\hat{y}_{i,j}$ is the of output log j predicted at depth i, and σ_j^2 is the variance of the predicted output log j, such that

$$\sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_{ij} - \mu_j)^2, \quad (12)$$

where μ_j is the mean of the predicted output log j. A regularization method is used to avoid overfitting and ensure a balanced bias-variance tradeoff. λ is tuned to be 0.01 by numerical experiments. In this case, the prediction performance of the ANN model in terms of R^2 is 0.87 and 0.85 for DTC and DTS log predictions, respectively. Although it is more accurate than the MARS model, the ANN model takes more computation time, i.e., the computational time measured during the training phase of the algorithm. When the model is deployed, the algorithm does not need to be retrained.

CASE STUDY

Prediction Results of Six Models

The six models were built using data from Well 1. After the models were trained, data from Well 2 were used to simulate the deployment of the models. The comparison of the prediction performances of the six models is shown in Table 2 and visualized in Fig. 8. In the 4,240-ft depth interval of Well 1, the ANN model performs the best when predicting DTC and DTS logs, with R^2 of 0.87 and 0.85, respectively (Fig. 9). MARS performs second best among all models, followed by OLS and PLS. The LASSO and elastic-net models perform the worst. Cross-validation was performed to ensure robustness of model predictions. The six models were then deployed in a 1,460-ft depth interval of Well 2, which located in the same reservoir. The prediction accuracy ranking does not change much. However, the PLS model performs worse than OLS and the elastic-net model performs worse than LASSO. Each model suffers a distinct drop in accuracy when trained in Well 1 and deployed in Well 2 (Fig. 10). The ANN model performs the best in both Well 1 and Well 2 (Fig. 10). The accuracy drop of the PLS model is much larger than that of others, especially the LASSO and elastic-net models. The computation time required by the six models are compared in Table 3. The ANN model required the most computation time. The PLS model required the second most computation time because it requires identification of latent structures. The remaining models have low computation time requirements.

Table 2—Prediction Performances of the Six Models Trained and Tested in Well 1 and Deployed in Well 2

		Accuracy of Model Predictions					
		OLS	PLS	LASSO	ElasticNet	MARS	ANN
Well 1	DTC	0.830	0.830	0.791	0.791	0.847	0.870
	DTS	0.803	0.803	0.756	0.753	0.831	0.848
Well 2	DTC	0.804	0.790	0.778	0.774	0.816	0.850
	DTS	0.794	0.769	0.763	0.755	0.806	0.840

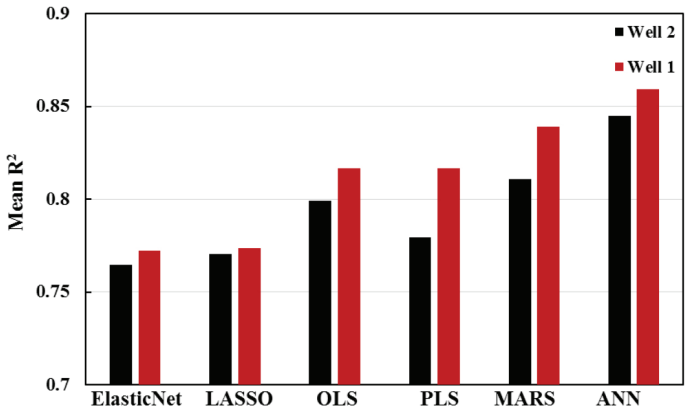

Fig. 8—Comparison of prediction performances of six models trained and tested in Well 1 and deployed in Well 2.

Table 3—Comparison of the Computation Time Required by the Six Models During the Training Phase in Well 1.

Model Computation Time (sec)					
ANN	PLS	MARS	OLS	LASSO	Elastic Net
4.17	3.57	0.14	0.09	0.03	0.04

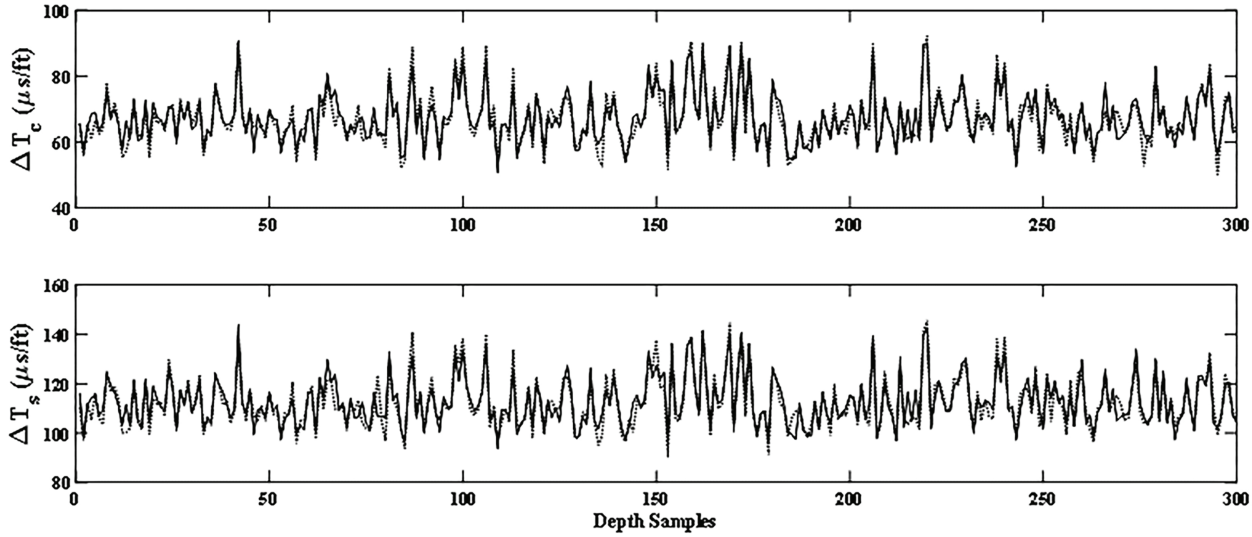

Fig. 9—Comparisons of original (dashed) and predicted (solid) DTC and DTS logs in Well 1, when the ANN model is trained and tested in Well 1.

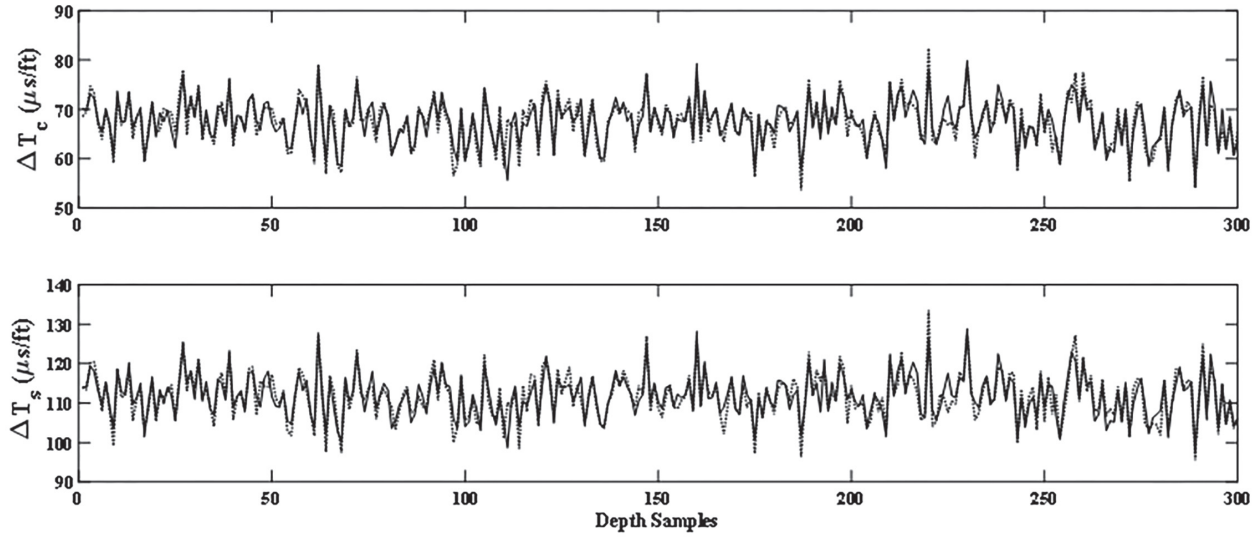

Fig. 10—Comparison of original (dashed) and predicted (solid) DTC and DTS logs in Well 2, when the ANN model is trained and tested in Well 1 and deployed in Well 2 to generate the DTC and DTS logs.

Analysis of Results

The ANN model performs the best among the six models, followed by the MARS, OLS, PLS, LASSO, and elastic-net models, in order of accuracy from high to low. The ANN model with multiple neurons and layers is most suited for handling nonlinearity in trends and relationships among the six models tested in this study. Cross-validation was done to ensure the model generalization. The MARS model performs better than the other four regression models because the MARS model takes nonlinearities into consideration through multiple linear regressions over the partitions, whereas the other four models are single linear-regression models. The OLS and PLS models have similar prediction performances because m in the PLS model is close to p, which implies that there are no obvious redundant inputs. The accuracies of the LASSO and elastic-net model predictions are the same because the optimum penalty parameters are close to each other ($\lambda \approx \lambda_1$ and $\lambda_1 \approx 0$). The OLS and PLS models outperform the LASSO and elastic-net models because the former two use all inputs to build models, while the latter two penalize the correlated inputs. The prediction performances of the six models for DTC are always better than those for DTS. One possible reason is that the higher speed of the compressional wave leads to

cleaner acquisition of DTC without the interference of shear waves. Another possible reason is that DTC is less affected by cracks, unconsolidated formations and borehole fluids as compared to DTS, which makes DTC predictions more reliable with less noise and uncertainty (Paillet et al., 1992).

Comparison of Prediction Performances of the Six Models in Well 1

In this section, relative error (RE) is used to evaluate the prediction performance. RE is formulated as

$$RE = \frac{|P-M|}{M}, \tag{13}$$

where P is the predicted value and M is the measured value of either DTS or DTC at any given depth i. Mean REs for both DTC and DTS logs are calculated at each depth, resulting in one RE value for each depth that is further averaged over 50-ft depth intervals. When the mean RE for a depth interval is less than 0.1, between 0.1 and 0.2, or larger than 0.2, the depth interval is categorized as good, intermediate, and poor prediction performance, respectively. All models exhibit poor prediction performances from around 1,250 ft to 1,800 ft below the top of the formation depth under investigation and medium prediction performances from depths around 3,800 to 4,240 ft.

In this section, we provide statistical and petrophysical reasons for aforementioned behaviors of the learning models. The 4,240-ft depth interval in Well 1 can be divided into four large sections based on the RE of the depth intervals, namely Section I, 0 to 1,250 ft; Section II, 1,250 to 1,800 ft; Section III, 1, 800 to 3,800 ft; and Section IV, 3,800 to 4,240 ft (also illustrated in Fig. 11). Sections I and III exhibit good prediction performances, Section II exhibits poor prediction performance, and Section IV belongs to intermediate prediction performance. Three statistical parameters were selected for the analysis of prediction performance in Well 1, which includes mean (μ), coefficient of variation (S_d/μ) and skewness, which are listed in Table 4 for each interval. S_d/μ is standard deviation over the mean value; S_d/μ and skewness are not highly correlated with the prediction performance.

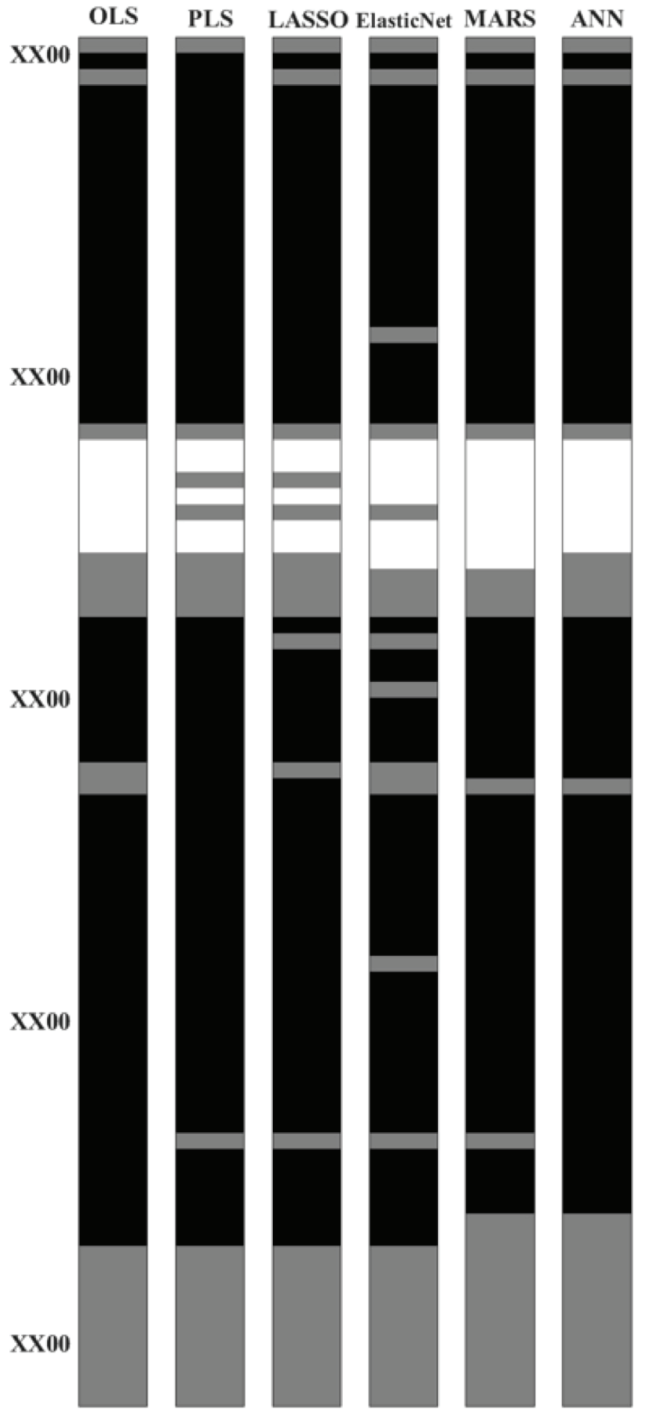

Fig. 11—Comparison of RE in the DTS and DTC predictions generated by the six models for the 4,240-ft depth interval in Well 1, where black-, gray-, and white-colored intervals represent depths in which the learning models exhibit good, intermediate, and poor prediction performances, respectively.

Table 4—Statistical Description of all Inputs and Outputs for the Depth Sections in Well 1

	μ				S_d/μ				Skewness			
	I	II	III	IV	I	II	III	IV	I	II	III	IV
GR	68.217	95.620	70.666	83.113	0.430	0.361	0.364	0.259	1.430	0.402	0.616	0.779
DCAL	1.909	0.786	0.765	0.120	0.543	0.132	0.412	1.565	1.511	-0.166	1.332	0.222
DPHZ	0.089	0.141	0.064	0.082	0.468	0.353	0.550	0.347	0.183	-0.533	-0.191	-1.214
NPOR	0.095	0.177	0.092	0.116	0.427	0.410	0.513	0.312	0.237	-0.403	0.166	0.396
PEFZ	3.194	3.303	3.467	2.923	0.305	0.184	0.243	0.147	1.668	1.043	1.278	1.763
RHOZ	2.566	2.483	2.607	2.578	0.026	0.032	0.022	0.018	-0.183	0.533	0.191	1.213
RLA0	0.031	0.035	0.033	0.033	0.120	0.024	0.059	0.033	-0.075	-0.667	0.481	0.049
RLA1	11.670	15.717	23.007	16.533	0.618	0.412	0.476	1.101	0.703	0.754	-0.168	2.279
RLA2	33.677	47.158	86.168	21.517	1.369	0.940	0.932	1.714	2.378	1.310	1.865	5.541
RLA3	61.603	82.055	143.065	23.881	1.825	1.222	1.148	1.967	3.193	1.956	1.929	6.489
RLA4	134.719	117.122	228.613	24.890	2.781	1.555	1.828	2.190	5.710	3.326	4.705	7.641
RLA5	187.924	129.367	247.366	24.707	2.894	1.472	2.084	2.108	6.013	3.541	6.018	7.229
DTC	64.040	78.444	64.504	71.172	0.080	0.132	0.103	0.064	-0.065	-0.753	-0.187	-0.572
DTS	109.783	126.857	108.676	118.544	0.080	0.100	0.076	0.058	1.251	-0.419	0.191	-0.423

Section code: I = 0 to 1,250 ft; II = 1,250 to 1,800 ft; III = 1,800 to 3,800 ft; IV = 3,800 to 4,240 ft

Depths with high skewness of porosity correspond to zones of poor prediction performance. Skewness indicates the direction and relative magnitude of how far a distribution deviates from normal distribution, which has skewness of zero. Skewness is the measure of asymmetry of the probability distribution about its mean either due to long tail or fatter tail. Skewness of porosity indicates the asymmetric distribution of the porosity such that the majority of layers in the formation tend to have either large or small pores with outliers biased towards either large or small pores, as quantified in terms of the skewness. When the formation properties are not uniform and tend to be skewed, the prediction accuracy for parameter values in the shorter and thinner tail will be low because the machine-learning models tend to be biased towards the fatter tails and adversely affected by the outliers in the longer tail. In this case, depth intervals with large of GR, porosity, DTC, and DTS logs and small of bulk density correspond to zones of poor prediction performance. High-GR, high-porosity, and low-density formation layers exhibit different geomechanical properties in comparison to the rest of the formation layers. Poor prediction accuracy arises when the machine-learning models are not sufficiently exposed to or not well trained on certain statistical distributions of inputs and outputs or certain statistical correlations between the inputs and outputs. We observed that the formation layers with large GR, large porosity and low density are related to low prediction performance of the machine-learning models. High GR is related to thick shale layers. High porosity and low density tend to be associated with unconsolidated or porous formations.

SENSITIVITY OF THE MODEL PREDICTIONS

Effect of Noise in the Input and Output Logs on ANN Model Predictions

Learning models are affected by the noise in the data. We added 10% Gaussian noise separately to each of the input logs, except the flags. Adding 10% Gaussian noise to an input log means that the standard deviation of the added noise is equal to 10% of the input log. The magnitude of the added noise at any depth is randomly selected from the Gaussian noise distribution. Parallel computing is used to set random noise on the input logs for 50 times and average the resulting predictions. The effect of adding noise to each input log on the prediction accuracy of the ANN model was recorded and the input logs were ranked in accordance of the reduction in accuracy (Fig. 12).

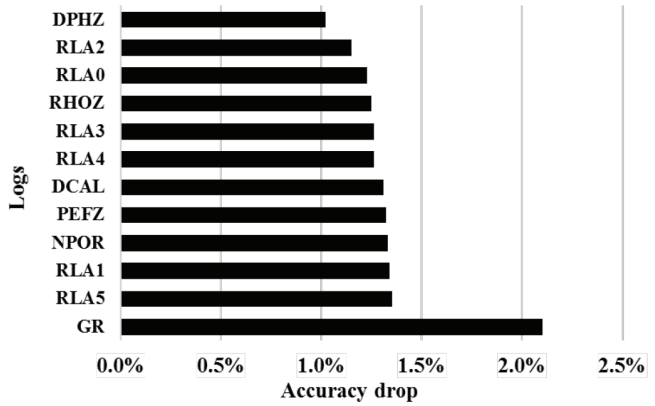

Fig. 12—Comparison of the reductions in ANN model prediction performances when 10% Gaussian noise is added to the input logs one at a time.

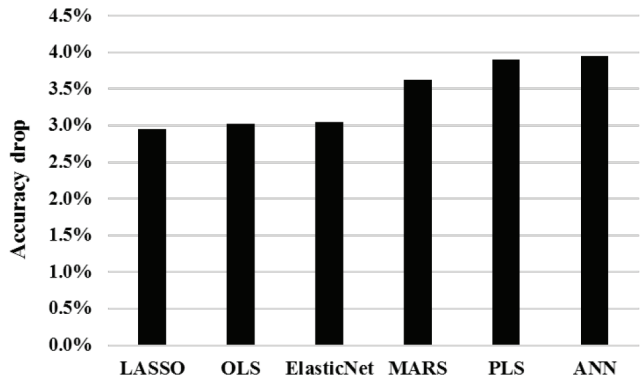

Fig. 13—Comparison of reductions in prediction performances of the six learning models when 10% Gaussian noise is simultaneously added to all the input logs.

Noise in the GR has the most effect on the prediction. This does not mean that the GR is the most important input log compared to the rest. For example, RLA0 to RLA5 have a high correlation among them; consequently, adding noise to one of them will not significantly change the prediction performance. Similarly, correlations between DPHZ and NPOR compensate for the noise in one of the logs. We then add 10 and 20% Gaussian noises separately to the output logs, namely DTC and DTS logs. When noise is added to the DTC log, the prediction accuracy of DTS does not drop, and vice versa. When 10% noise is added to the DTC and DTS logs, prediction accuracy drops by 41.4 and 52.8%, respectively. Twenty-percent noise is then added to DTC and DTS logs, resulting in 73.3 and 81.0% accuracy drops, respectively. In conclusion, prediction accuracy is more sensitive to noise in the output logs than that in the input logs. DTS is more sensitive to noise as compared to DTC.

Comparison of the Effect of Noise in the Input Logs on the Performances of the Six Learning Models

Ten-percent Gaussian noise is simultaneously added to all the input logs before training the six learning models. Accuracy drops of the OLS, LASSO and elastic-net models are lower than those of the PLS, MARS and ANN models (Fig. 13). The former three models (OLS, LASSO and elastic-net) are less sensitive to noise because they are all simple linear regressions with simple model structures. The MARS and ANN models both consider nonlinearities during the training step. For simple problems, the ANN and MARS models will tend to overfit in the presence of noise, resulting in a larger accuracy drop. Notably, the ANN model predictions exhibit close to 2% accuracy drop when 10% Gaussian noise is added only to GR logs, as compared to a 4% accuracy drop when 10% noise is simultaneously added to each of the input logs.

Comparison of the Effect of Quantity of Data on the Performances of the Six Learning Models

Three different sizes of datasets are used to train and test the six models. The entire data in Well 1 contains 8,481 depth points in the approximately 4,200-ft depth interval under investigation. About half of the 8,000 depth points are randomly selected to create the second dataset. One half of the 4,000 depth points in second dataset are randomly selected to build the third dataset. Accuracy, in terms of of the six models, decreases with the decrease in the size of the dataset available for the training and testing of the six models. The OLS, LASSO and elastic-net models are less sensitive to dataset size, while the PLS, MARS and ANN models are more affected by the reduction in dataset size. Two thousand data samples are sufficient for the linear regression problem with 13 inputs and two outputs. As a result, the prediction performances of the former three models does not change much when more data are available for the training and testing. However, the prediction performance of the latter three mentioned models significantly improves when dataset increases because more data are needed to build robust nonlinear models and to build the linear model where latent structures are constructed from the data prior to regression.

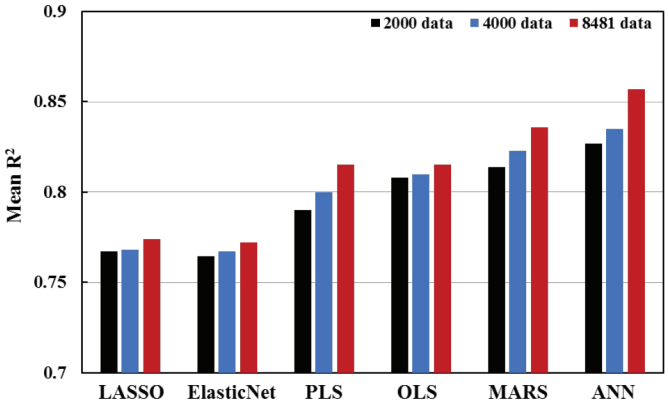

Fig. 14—Comparison of reductions in prediction performances of the six learning models when different dataset sizes are used for the training and testing of the models.

CONCLUSIONS

For purposes of shale reservoir geomechanical characterization, six shallow learning models successfully predict compressional-wave (DTC) and shear-wave (DTS) traveltime logs in two wells by processing 13 conventional easy-to-acquire logs from two wells in a shale reservoir. We implemented four linear-regression-based learning models, namely ordinary-least-squares (OLS), partial-least-squares (PLS), LASSO and elastic-net models, and two nonlinear learning models, namely multivariate adaptive regression splines (MARS) and artificial neural networks (ANN). These data-driven methods were compared in terms of prediction performance, sensitivity to noise, and sensitivity to dataset size. The ANN model performs the best in both the wells. Simple linear-regression models, such as the OLS, LASSO and elastic-net models, are more robust to noise and the size of the training dataset but fail to handle nonlinearities. The learning models tend to exhibit poorer prediction performance in depth intervals exhibiting high gamma ray, high porosity, and large absolute values of skewness of porosity. Prediction performance for the DTC log is always better than that for DTS logs. This study will enable improved geomechanical characterization under data constraints by developing and identifying the best performing machine-learning models for the syntheses of DTC and DTS logs, when a sonic logging tool is not available due to operational or financial challenges.

NOMENCLATURE

Abbreviations

ANN = artificial neural network
AT10 = induction resistivity logs at 10-in. spacing
AT90 = induction resistivity logs at 90-in. spacing
CG = conjugate gradient algorithm
CMIS = committee machine with intelligent systems
DCAL = caliper log
DD = dielectric dispersion
DPHZ = density-porosity log
DTC = Δt compressional slowness log
DTS = Δt shear slowness log
GR = gamma ray log
LASSO = least absolute shrinkage and selection operator
LM = Levenberg-Marquardt algorithm
MARS = multivariate adaptive regression splines
NMR = nuclear magnetic resonance
NPOR = neutron-porosity log
NRMSE = normalized root mean square error
OLS = ordinary least squares
PEFZ = photoelectric factor log
PLS = partial least squares
RE = relative error
RHOZ = bulk-density log
RLA0-5 = laterolog resistivity at different depths of investigation
RSS = sum of squares of residuals
SSE = sum of squared errors
TOC = total organic carbon
VCL = volume of clay log

Symbols

μ = mean
S_d = standard deviation
λ = penalty term
σ = variance
y_{ij} = log j measured at depth i
$\hat{y}_{ij}$ = mean of output log j
x_{ij} = j-th input log at depth i

REFERENCES

Asoodeh, M., and Bagheripour P., 2012, Prediction of Compressional, Shear, and Stoneley Wave Velocities From Conventional Well Log Data Using a Committee Machine With Intelligent Systems, *Rock Mechanics and Rock Engineering*, **45**(1), 45–63. DOI: 10.1007/s00603-011-0181-2.

Baines, V., Bootle, R., Pritchard, T., Macintyre, H., and Lovell, M.A., 2008, Predicting Shear And Compressional Velocities In Thin Beds, Paper I, *Transactions, SPWLA 49th Annual Logging Symposium*, Edinburgh, Scotland, UK, 25–28 May.

Chamjangali, M.A., Beglari, M., Bagherian, G., 2007, Prediction of Cytotoxicity Data (CC_{50}) of Anti-HIV 5-Pheny-l-Phenylamino-1H-Imidazole Derivatives by Artificial Neural Network Trained With Levenberg–Marquardt Algorithm, *Journal of Molecular Graphics and Modelling*, **26**(1), 360–367. DOI: 10,1016/jmgm.2007.01.005.

Cheng, C., Chaum K., Sun, Y., and Lin, J., 2005, Long-term Prediction of Discharges in Manwan Reservoir Using Artificial Neural Network Models, *in* Wang, J., Liao, X.F., and Yi, Z., Editors, *Advances in Neural Networks*–ISNN 2005, Lecture Notes in Computer Science, **3498**, 975–975. DOI: 10.1007/11427469_1.

Draper, N.R., and Smith, H., 2014, *Applied Regression Analysis*, 3rd Edition (eBook), Wiley Series in Probability and Statistics. ISBN: 978-1118625682.

Greenberg, M.L., and Castagna, J.P., 1992, Shear-Wave Velocity Estimation in Porous Rocks: Theoretical Formulation, Preliminary Verification and Applications, *Geophysical Prospecting*, **40**(2), 195–209. DOI: 10.1111/j.136502478.1992.tb00371.x.

Hagan, M.T., Demuth, H.B., Beale, M.H., and De Jesus, O., 2014, *Neural Network Design*, 2nd Edition. http://hagan.okstate.edu/NNDesign.pdf. Accessed October 25, 2018.

Han, Y., and Misra, S., 2018, Joint Petrophysical Inversion of Multifrequency Conductivity and Permittivity Logs Derived From Subsurface Galvanic, Induction, Propagation, and Dielectric Dispersion Measurements, *Geophysics*, **83**(3), D97–D112. DOI: 10.1190/geo2017-0285.1

Iverson, W.P., and Walker J.N., 1992, Shear and Compressional Logs Derived From Nuclear Logs, *The Log Analyst*, **33**(1), 12–21.

Keys, R.G., and Xu, S., 2002, An Approximation for the Xu-White Velocity Model, *Geophysics*, **67**(5), 1406–1414. DOI: 10.1190/1.1512786.

Kuhn, M., and Johnson, K., 2013, *Applied Predictive Modeling*, Springer. ISBN: 978-1-4614-6848-6.

Li, H., and Misra, S., 2017a, In-situ Wireline-Log-Derived Indices for Miscible Light-Hydrocarbon-Injection Recovery in Bakken Petroleum System, *SEG Technical Program Expanded Abstracts 2017*, 3453–3457. DOI: 10.1190/segam2017-17796067.1

Li, H., and Misra, S., 2017b, Prediction of Subsurface NMR T2 Distributions in a Shale Petroleum System Using Variational Autoencoder-Based Neural Networks, *IEEE Geoscience and Remote Sensing Letters*, **14**(12), 2395–2397. DOI: 10.1109/LGRS.2017.2766130.

Maleki, S., Moradzadeh, A., Riabi, R.G., Gholam,i R., and Sadeghzadeh, F., 2014, Prediction of Shear Wave Velocity Using Empirical Correlations and Artificial Intelligence Methods, *NRIAG Journal of Astronomy and Geophysics*, **3**(1), 70–81. DOI: 10.1016/j.nrjag.2014.05.001.

Paillet, F.L., Cheng, C.H., and Pennington W.D., 1992, Acoustic-waveform Logging—Advances in Theory and Application, *The Log Analyst*, **33**(3), 239–258.

Rezaee, M.R., Ilkhchi, A.K., and Barabadi, A., 2007, Prediction of Shear Wave Velocity From Petrophysical Data Utilizing Intelligent Systems: An Example From a Sandstone Reservoir of Carnarvon Basin, Australia, *Journal of Petroleum Science and Engineering*, **55**(3-4), 201–212. DOI: 10.1016/j.petrol.2006.08.008.

Tathed, P., Han, Y., and Misra, S., 2018, Hydrocarbon Saturation in Bakken Petroleum System Based on Joint Inversion of Resistivity and Dielectric Dispersion Logs, *Fuel*, **233**, 45–55. DOI: 10.1016/j.fuel.2018.06.019.

Willis, M.E., and Toksoz, M.N., 1983, Automatic P and S Velocity Determination From Full Waveform Digital Acoustic Logs, *Geophysics*, **48**(12), 1631–1644. DOI: 10.1190/1.1441444.

Xu, S., and White, R.E., 1995, A New Velocity Model for Clay-Sand Mixtures, *Geophysical Prospecting*, **43**(1), 91–118. DOI: 10.1111/j.1365-2478.1995.tb00126.x.

ABOUT THE AUTHORS

Jiabo He received an MS degree in petroleum engineering from the Mewbourne School of Petroleum and Geological Engineering at The University of Oklahoma. He is currently pursuing a PhD in computer science and machine learning at the University of Melbourne.

Hao Li received BS and MS degrees in petroleum engineering from China University of Petroleum, Beijing, China, in 2013 and 2016, respectively. He is currently a PhD candidate at the Mewbourne School of Petroleum and Geological Engineering, the University of Oklahoma, Norman, Oklahoma, USA. His research interests include petrophysics and machine learning.

Siddharth Misra is Assistant Professor in Mewbourne School of Petroleum and Geological Engineering. He holds PhD and MSE degrees in Petroleum Engineering from the University of Texas at Austin. Dr. Misra completed his BTech degree in Electrical Engineering at the Indian Institute of Technology Bombay, India. He was recently awarded the Department of Energy Early Career Award and American Chemical Society New Investigator Award.