Theory-guided data science-based reservoir prediction of a North Sea oil field

Jonathan E. Downton[1], Olivia Collet[2], Daniel P. Hampson[1], and Tanya Colwell[1]

https://doi.org/10.1190/tle39100742.1

Abstract

Data science-based methods, such as supervised neural networks, provide powerful techniques to predict reservoir properties from seismic and well data without the aid of a theoretical model. In these supervised learning approaches, the seismic-to-rock property relationship is learned from the data. One of the major factors limiting the success of these methods is whether there exists enough labeled data, sampled over the expected geology, to train the neural network adequately. To overcome these issues, this paper explores hybrid theory-guided data science (TGDS)-based methods. In particular, we build a two-component model in which the outputs of the theory-based component are the inputs in the data science component. First, we simulate many pseudowells based on the well statistics in the project area. The reservoir properties, such as porosity, saturation, mineralogy, and thickness, are all varied to create a well-sampled data set. Elastic and synthetic seismic data are then generated using rock physics and seismic theory. The resulting collection of pseudowell logs and synthetic seismic data, called the synthetic catalog, is used to train the neural network. The derived operator is then applied to the real seismic data to predict reservoir properties throughout the seismic volume. This TGDS method is applied to a North Sea data set to characterize a Paleocene oil sand reservoir. The TGDS results better characterize the geology and well control, including a blind well, compared to a solely theory-based approach (deterministic inversion) and a data science-based approach (neural network using only the original wells). These results suggest that theory and data science can complement each other to improve reservoir characterization predictions.

Introduction

Machine learning algorithms employing neural networks have been used for some time in geophysics to quantitatively predict elastic and rock properties from seismic data (Hampson et al., 2001). These supervised learning techniques derive a statistical relationship between the log and seismic data at the well locations. This relationship is then applied to the seismic data to estimate the log properties at other locations in the seismic survey. The limiting factor in performing this analysis is that sufficient labeled data (i.e., well control) are needed to train and validate the relationship. More recent neural network architectures, such as deep neural networks (DNNs) (Goodfellow et al., 2016), have been adapted for geophysical reservoir characterization. Downton and Hampson (2018) use a DNN, while Zheng and Zhang (2018) and Das et al. (2019) use convolutional neural networks (CNNs) to predict elastic parameters. These networks place even greater demands on the amount of data necessary to perform the analysis. The image-recognition CNN of Krizhevsky et al. (2012) is trained using millions of labeled images. The training wells available in typical geophysical reservoir characterization projects number in the tens rather than millions. Halevy et al. (2009) show that increasing the amount of data is often more important than the choice of algorithm, which is why this paper focuses on how to deal with this lack of training data.

Many different approaches have been taken to increase the amount of training data. K. Wang et al. (2019) perform a series of transformations on the labeled data to increase the amount available. Y. Wang et al. (2019) use a 1D cycle-consistent generative adversarial network to help stabilize the problem. Downton and Hampson (2018), Zheng and Zhang (2018), and Das et al. (2019) all generate synthetic data to increase the amount of training data. A number of important considerations determine whether there is enough training data. First, there needs to be sufficient data to calculate statistically reliable operators in the presence of noise. Second, in order to derive relationships that generalize satisfactorily, the well data need to sample the range of the expected geology for the project area. For example, to predict the fluid saturation, both hydrocarbon- and brine-filled training examples are needed. If only brine-filled reservoir examples are available, an interpreter might perform fluid substitution and amplitude variation with offset (AVO) modeling to understand the seismic response to the other fluid. Other important variables such as the reservoir thickness, porosity, and mineralogy can also be varied to understand their responses. These all complicate the mapping between the seismic and the desired reservoir properties, often in a nonlinear and nonunique manner. In this way, rock physics and seismic theory can be used to create training data that span the range of the expected geology. This is the approach of Downton and Hampson (2018) and of this paper. Methods that solely transform the original labeled data or are based on the statistics of the original labeled data do not address this second consideration.

This use of theory to generate training data within a data science methodology is a form of hybrid theory-guided data science (TGDS) (Karpatne et al., 2017) and hence is the terminology adopted by this paper. The theory component consists in using rock-physics relationships to simulate a large idealized set of pseudowell logs and AVO modeling to generate synthetic gathers, which we will refer to as the synthetic catalog (Dvorkin et al., 2014). This synthetic catalog, which contains both the target log

[1]CGG GeoSoftware, Calgary, Alberta, Canada. E-mail: jon.downton@cgg.com; dan.hampson@cgg.com; tanya.colwell@cgg.com.
[2]CGG GeoSoftware, Massy, France. E-mail: olivia.collet@cgg.com.

property and input seismic attributes, is then used to train and validate the DNN. Finally, the trained DNN is applied to the real seismic data set on a trace-by-trace basis to obtain 3D estimates of elastic and petrophysical properties.

In this paper, we first describe the key steps used in generating the synthetic catalog. Then, we describe the DNN training and validation process. Finally, we show elastic and petrophysical properties estimated by applying the DNN. The TGDS method is demonstrated on a North Sea commercial oil field data set. The analysis is focused on Paleocene reservoirs located in remobilized injectite sands cross-cutting various stratigraphies at steep angles. Those injectite sands, although difficult to image, could provide well-connected, high-porosity, and high-permeability reservoirs. To evaluate the robustness of the methodology, we compare the TGDS results at a blind well location to those obtained by deterministic inversion and by a data science-based approach in which the DNN is solely trained using the original well control. In both cases, the TGDS results are more consistent with the geologic control.

Synthetic catalog generation

The synthetic catalog is composed of a series of pseudowells and synthetic seismic gathers. To make the synthetic pseudowells realistic, we use prior geologic knowledge of the area to model various "what-if" scenarios. We also perform a detailed statistical analysis of the available well control. Typically, the statistics of each well are nonstationary, so each well is broken up into a series of layers, or lithofacies, that share common rock property statistics. The resulting statistical parameterizations are then used to simulate rock property values. Rock physics and seismic theory are used to model the corresponding elastic and seismic data for each pseudowell. This requires the interpreter to choose a rock-physics model (RPM) appropriate to the project area and then to calibrate it using the available well control. It is important that the chosen RPM accurately reconstructs the density, P-wave, and S-wave velocity logs so that we have confidence that the RPM generalizes well to other combinations of rock properties not sampled by the original well control. The use of the RPM necessitates performing a petrophysical analysis of the well-log data to supply the necessary input. In summary, the key steps in generating the synthetic catalog are:

1) performing a petrophysical analysis to generate the rock properties necessary for the RPM
2) selecting and calibrating the RPM
3) determining the nonstationary well-log statistics, which requires breaking up each well into a series of lithofacies and then determining their statistical parameterization
4) generating a series of pseudowells based on the previous analysis, where the rock properties are simulated based on the statistical analysis while the elastic properties are calculated using the calibrated RPM
5) calculating seismic gathers for each pseudowell

Ideally, this workflow should be applied to all the wells in a study area, but there are always data and human constraints. In order for the wells to be analyzed, they need to have the appropriate data logged (e.g., S-wave logs), and the data need to be of a sufficient quality. The petrophysical analysis takes time and again limits the number of wells analyzed. Given these constraints, we select a subset of the wells to analyze. Ideally, we want to choose wells that capture the natural variability of the reservoir in the study area (e.g., different fluids, porosity, clay volume, etc.). Note that the synthetic catalog will be used to train the DNN and derive the relationship between the seismic and well data so that the analyzed wells do not need to directly tie with the seismic. For the North Sea example, we analyze four wells. An additional well is reserved to serve as a blind test of the methodology.

Petrophysical analysis. To generate realistic synthetic log data, a petrophysical analysis is first performed to calculate the necessary input to the RPM, such as porosity, clay volume, and water saturation. The petrophysical analysis needs to identify and correct spurious log measurements on all the logs used in the following workflow, including the density, P-wave, and S-wave velocity logs. Smith (2011) summarizes some important considerations in performing a petrophysical analysis with the goal of seismic reservoir characterization.

RPM calibration. Next, we establish the RPM linking the calculated petrophysical properties to the elastic properties. The choice of the RPM largely depends on its effectiveness, which includes the ability to calibrate the model, using easily acquired data and observations, and the ability to use it in reverse modeling to accurately predict petrophysical properties of interest from elastic properties (Mur and Vernik, 2019). For this study, we use a granular model that combines Hertz-Mindlin contact theory and Hashin and Shtrikman (1963) bounds. This model, which best describes unconsolidated (soft) sandstones (Dvorkin and Nur, 1996), was further extended to stiffer sandstones by Allo (2019) through the matrix stiffness index (MSI). In Figure 1, we show the rock-physics templates for various MSI values ranging from 0 to 1. The fluid effects are accounted for using Gassmann (1951) fluid substitution. To calibrate the RPM and fit the log measurements, we invert for two MSI values: one that fits the S-wave modulus and one that fits the P-wave modulus. After

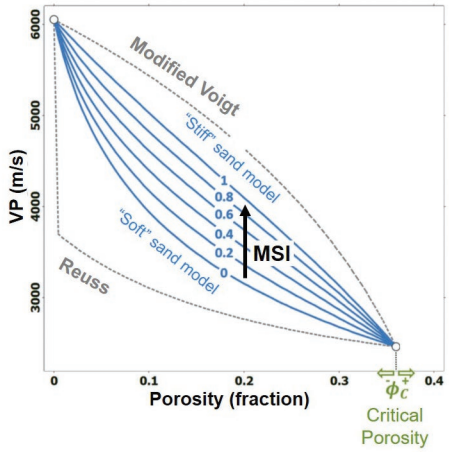

Figure 1. Rock-physics template of the extended unconsolidated sandstone model used to relate the petrophysical properties to the elastic properties. The MSI is used as a fitting parameter to calibrate the RPM. Adapted from Allo (2019).

inverting for these parameters, we are able to reconstruct the P- and S-wave velocities measured at the wells.

Statistical analysis. To obtain the parameters needed to generate the pseudowell simulations, we next perform a statistical analysis on the rock property curves (i.e., the petrophysical and MSI curves) for each well. For various reasons, including the fact that the earth is composed of a series of stratigraphic layers with different lithologies, the rock property values are nonstationary. Hence, the first step we take in our statistical analysis is to separate the well into a series of vertically contiguous lithofacies that share common statistics (Figure 2). Following Dvorkin et al. (2014), we assume that the rock properties of each lithofacies are described by a multivariate Gaussian distribution parametrized in terms of a mean vector and a covariance matrix. The diagonal of the covariance matrix describes the variability of each of the rock properties within the layer. The off-diagonal elements of the covariance matrix capture the correlations between the different rock properties and help differentiate between lithofacies. For example, dispersive and layered shale sands have different porosity versus clay relationships, implying different correlations. Lastly, the mean describes the average value of the background trend for the layer. Specifying a constant mean leads to a blocky layer model. In practice, this is inadequate since most rock properties exhibit some depth dependence. Instead, we define a linear background trend for each rock property within each lithofacies (Figure 3a).

The determination of the lithofacies is one of the key steps in the synthetic catalog. Since defining the lithofacies intervals manually can be quite time-consuming, we use a two-step semisupervised approach to automate this procedure. In this method, we first classify each log sample into different lithology classes based on a categorization of the input attributes, in this case, porosity, clay volume, and water saturation. For our case study example, the porosity attribute is split into two categories (e.g., high or low porosity) while the clay volume and water saturation attributes are split into three categories, which yields 18 lithology categories (i.e., $2 \times 3 \times 3 = 18$). Some of the resulting lithology layers do not have enough samples to calculate reliable statistics. So, in the second step, we agglomerate the thin lithology beds into adjacent beds sharing similar multivariate Gaussian statistics. The final lithofacies are shown in Figure 2. After building the lithofacies log, we calculate the background trend and covariance matrix for each lithofacies interval.

Lastly, we need to account for the fact that the logs within each well are vertically correlated. If this vertical correlation is ignored, the simulations exhibit a sample-to-sample variability that is inconsistent with the original well-log data. The spatial correlation is modeled using an exponential variogram. This variogram mainly acts as a high-cut filter, with the range of the variogram being the key parameter controlling the vertical variability of the simulated log for each lithofacies, as illustrated in Figure 4. In the next step, we combine the covariance matrix

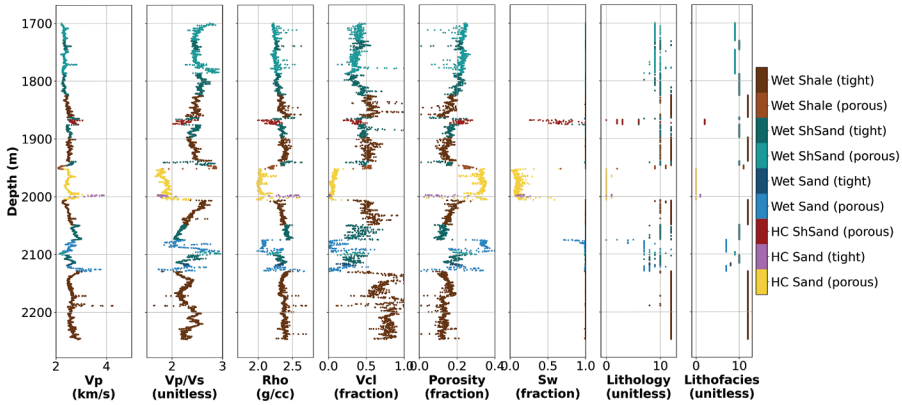

Figure 2. Elastic properties (P-wave velocity, V_P/V_S ratio, and density) and petrophysical properties (clay volume, porosity, and water saturation) color coded by the lithofacies log. HC and ShSand respectively refer to hydrocarbon-saturated and shaly sand categories. The lithology log is obtained using a decision tree logic based on thresholds applied to the petrophysical logs. The lithology log is then upscaled to the lithofacies log based on statistical considerations.

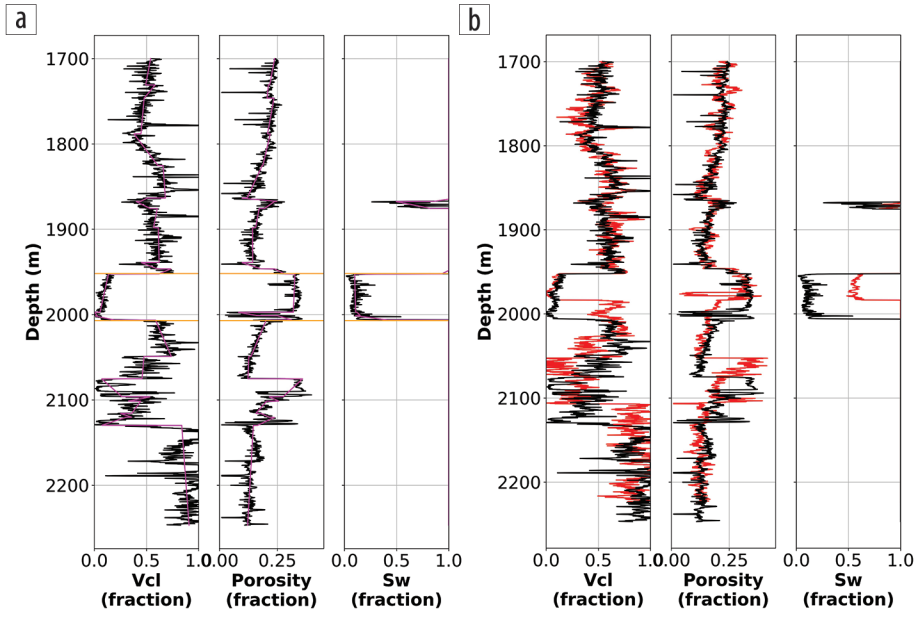

Figure 3. (a) Clay volume, porosity, and water saturation measured in one of the wells (black) and corresponding linear background trends computed on each lithofacies (magenta). The orange lines delineate the bottom reservoir, of which the thickness and water saturation trend are modified for the simulation shown in (b). (b) Initial measured logs (black) and example of simulation (red) obtained after modifying the thickness and increasing the water saturation in the bottom reservoir layer.

calculated from the exponential variogram with the covariance matrix of the rock properties using the Kronecker product (Dvorkin et al., 2014) to derive the covariance matrix used to generate the simulations.

Simulations. From results obtained in previous steps, it is now possible to create many pseudowells that are representative of the geology. We use Gaussian simulations to generate the rock properties. These Gaussian simulations can be superimposed to what-if scenarios applied to specific layers to generate training examples spanning the range of the expected geology. After creating the rock properties, we derive the corresponding elastic properties from the calibrated RPM.

For each well, the rock property curves (i.e., the petrophysical and MSI curves) are simulated based on the Gaussian statistics for each lithofacies or layer. The correlation matrix imposes a relationship between the different rock properties, while the range of the variogram controls the vertical variability of the curves. To ensure the simulated properties are physical, we impose bounds for each property and reject simulations for which the simulated properties do not fall between the bounds. The simulations are performed independently for each layer to account for the different statistics of the lithofacies. Next, the density, P-wave, and S-wave velocity logs are calculated from the simulated rock properties using the previously calibrated RPM. An important QC is to ensure that the simulated logs accurately capture the important geologic characteristics of the original logs. The mean of the simulations should follow the background trend of the original well control, while the random realizations model the natural variability of the geology. Figure 5 shows 10 simulations superimposed on the original log curves. This QC may reveal some inadequacies in the lithofacies classification. For example, after running this QC we adjusted some of the parameters used in the initial lithofacies classification in order to isolate the high-velocity calcareous streaks observed in the original logs at 2000 m depth.

Then, to ensure that the synthetic data sample the range of possible geologic scenarios, we systematically perturb the reservoir and surrounding layer properties. For example, the porosity, mineralogy, fluid content, and saturation are all varied around the original values. This involves shifting the background trends of the relevant properties. In addition, the thickness of the reservoir layers is varied to understand the impact of tuning. Figure 3b displays an example of a simulation obtained by decreasing the thickness by a factor of two and increasing the water saturation. Once again, the calibrated RPM is used to calculate the density and velocity curves. This is similar to AVO modeling studies in which different parameters are systematically varied to understand the interrelationship of these parameters. In this study, we introduce five variations for each of the following reservoir properties: porosity, water saturation, and thickness. This resulted in 125 combinations per well for the four wells, giving rise to a total of 500 pseudowells. This high number of reservoir property combinations far exceeds the number of different reservoir property scenarios that would normally be sampled at the existing well control.

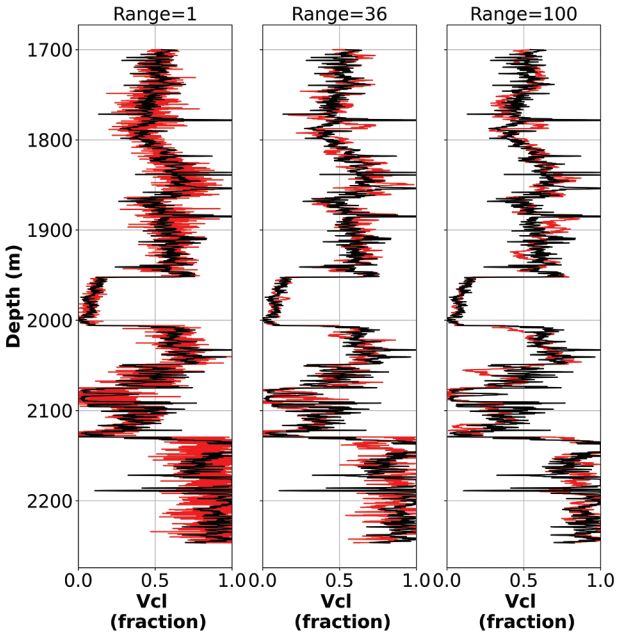

Figure 4. Effect of the range of the variogram on the simulated curves (in red). The range is the key parameter to adequately model the vertical variability of the original log curves (in black). The range value is expressed in number of samples with a depth sampling of 0.1524 m.

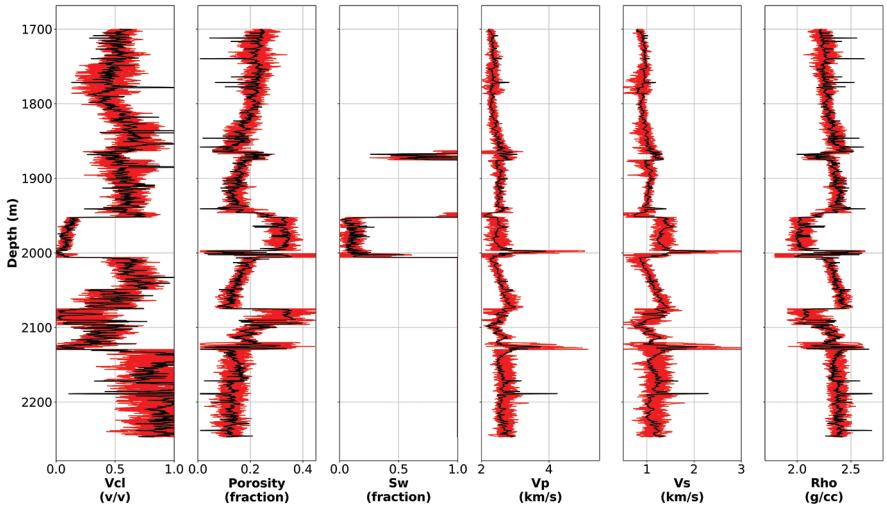

Figure 5. The 10 log simulations in red capture important geologic features of the original well logs shown in black. Note that the high-impedance, low-porosity calcareous streak is accurately modeled at 2000 m. This required adjusting the lithofacies parameterization.

Synthetic seismic gathers. Finally, we perform AVO modeling to create synthetic seismic gathers from each of the pseudowells. Because one of the objectives of this study is to compare the results with other methods, such as deterministic inversion, the seismic data were processed following a controlled amplitude processing sequence (Soroka

et al., 2002) suitable for P-wave AVO analysis. This assumes that only compressional wave reflectivity is considered. Transmission losses, converted waves, and multiples are not incorporated in this model and so must be addressed through prior processing.

To be consistent with these assumptions, the synthetic gathers are calculated using the P-wave reflection coefficients calculated using the Zoeppritz (1919) equations and a convolutional model using a wavelet extracted from the real seismic data. Figure 6 shows a subset of these synthetic gathers generated from eight pseudowells. The synthetic gathers are subsequently processed in a manner similar to the real seismic data, generating a series of angle stacks. The resulting synthetic angle stacks serve as the input to the neural network analysis.

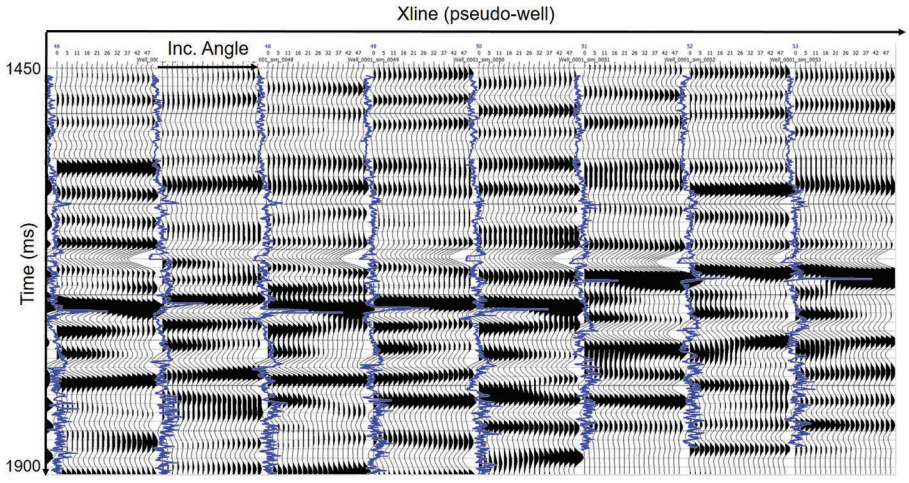

Figure 6. Subset of AVO synthetic gathers generated from the pseudowells. Curves in blue show the P-wave velocities in the pseudowells.

It is worth noting that multiples and converted waves could be incorporated into the analysis similar to the modeling shown by Simmons and Backus (1994). This would involve performing reflectivity modeling (Kennett, 1984) and processing the real seismic data in a manner consistent with these different assumptions.

DNN

The synthetic catalog is then used to train the DNN. The target property can be any of the simulated logs (e.g., P-wave and S-wave impedance, density, porosity, saturation, and lithology). The input features to the neural network are the synthetic seismic data or seismic attributes derived from the synthetic seismic. Spatially smoothed background trends can also be input as an attribute to provide missing low frequencies to the result. We follow the methodology described in Hampson et al. (2001) with the exception that the synthetic catalog is used to train the DNN. The synthetic catalog is randomly split into a training data set comprising 85% of the wells and a validation data set made up of the remaining 15% of the wells. Following the methodology, stepwise regression (Draper and Smith, 1998) is used to select the set of attributes that best correlates to the target property. For example, Table 1 lists the attributes in order of importance targeting the P-wave impedance. Note that training on the synthetic catalog and the original wells results in different lists, the implications of which are discussed in the results section.

The selected attributes are input into a fully connected DNN. The network is deep in the sense that it has multiple hidden layers (Figure 7). The number of hidden layers and number of hidden nodes per hidden layer are treated as hyperparameters that are optimized for each target property. In the P-wave impedance example, we use three hidden layers and 20 nodes. The number of input nodes corresponds to the number of input features. To account for the resolution differences between the seismic and log data we use a convolutional operator on the input layer, as described in Hampson et al. (2001). Note that this results in a different architecture than what is commonly known as a CNN.

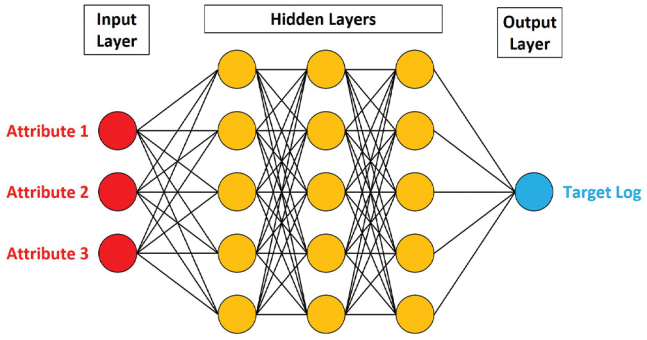

Figure 7. Example of DNN architecture with three attributes constituting the input layer nodes, three hidden layers made of five nodes each, and a single node in the output layer that corresponds to the target log.

Table 1. Stepwise regression results for four initial wells and the synthetic catalog.

Attribute set 1 (stepwise regression on four initial wells)	Attribute set 2 (stepwise regression on synthetic catalog)
Low-frequency P-impedance model	Low-frequency P-impedance model
Quadrature trace (ultra-far-angle stack)	Derivative (near-angle stack)
Filter 5/10–15/20 (far-angle stack)	Second derivative (near-angle stack)
Instantaneous phase (far-angle stack)	Filter 45/50–55/60 (mid-angle stack)
Amplitude-weighted cosine phase (far-angle stack)	Second derivative (mid-angle stack)

Lastly, the output layer contains a single node corresponding to the target property.

The DNN operator is then tested on the validation data set to ensure that the nonlinear relationship generalizes well to data not used in the training process. Note that neither the training process nor the validation process includes the original wells and the real

3D seismic data. DNN operators are designed for each target property from the synthetic data and applied to the actual 3D seismic data. After training the DNN operator for a specific target property, we apply it to each trace of the real 3D seismic volume to obtain 3D estimates of the target property. Similar to prestack elastic inversion, the real seismic data are scaled prior to applying the DNN operator in order to account for the amplitude differences between the real and synthetic seismic.

Results

To evaluate whether the TGDS approach is superior to a solely theory-based approach or a solely data science-based approach we first compare the TGDS P-wave impedance estimates with those of deterministic inversion. Then, to demonstrate that the method generalizes well, we display the TGDS P-wave impedance, S-wave impedance, and porosity results at a blind well location. Lastly, we compare the TGDS result (DNN trained using the synthetic catalog) with the data science solution (DNN trained only on data from four original wells).

Figure 8 shows the P-wave impedance estimated from the prestack inversion (Hampson et al., 2005) compared to the TGDS result along an arbitrary line through the key well locations. The inputs to both analyses are five angle stacks and a low-frequency P-wave impedance model. The objective is to resolve a remobilized sand injectite cross-cutting a range of stratigraphy at very steep angles. Wells A, C, and D are producing from these sands while well B missed the sand injectite and is wet. At first glance, the TGDS result exhibits better lateral continuity and images some of the thin layers better than the deterministic inversion. Examination of the results reveals that the TGDS predictions distinguish the producing wells D and A from the wet well B and better separate the oil-saturated sand from the upper wet sand in well C. Note that both methods share the same low-frequency P-wave impedance model and rely on 1D operators so that the apparent lateral continuity differences are not a function of spatial operators. Quantitatively, the correlations between the TGDS-estimated P-impedance and the P-impedance measured at the wells range from 74% to 90%, while they range from 63% to 87% for the deterministic P-impedance inversion results.

To show that the TGDS method generalizes satisfactorily, we examine the TGDS P- and S-impedance estimates at a blind well location. Unlike the four other wells shown on the previous P-impedance section, this blind well was not used to generate the pseudowells and synthetic seismic data that were input to the DNN training. Figure 9 displays the TGDS-estimated P- and S-impedance sections on an inline passing through the blind well and well C. The TGDS method accurately estimates the P- and S-wave impedances measured at the blind well, with correlations of 87% and 89%, respectively. As shown in Figure 10, these results are comparable to the inversion results for which the correlations reach 89% and 91% for the P-impedance and S-impedance, respectively.

As mentioned earlier, any simulated log can serve as a target, which allows the TGDS methodology to solve other geophysical inversion problems, such as porosity estimation. The TGDS porosity estimates are shown in Figure 11 along the inline passing through the blind well and well C and an arbitrary line passing through the four wells that were used to generate the pseudowells. Again, we notice a good match between the TGDS predictions and the porosities estimated at the wells. Since petrophysical properties, such as porosity, clay volume, and water saturation, are not routinely included in the forward model for seismic model-based inversions, the present technique could prove useful in estimating reservoir properties without relying solely on a rock-physics transform linking inverted elastic attributes and rock properties.

Finally, we compare the TGDS P-wave impedance estimates obtained by training the DNN on the synthetic catalog (Figure 12c) to results obtained by training the DNN only on data from the four original wells. The TGDS estimate (Figure 12c) matches the well control better, in particular the high-impedance tight streaks. Further, the DNN trained on the original wells (Figure 12a) is more discontinuous and noisier. The difference is a result of two factors, the selection of attributes and, secondly, the impact of the greater fold and sampling on the operator design.

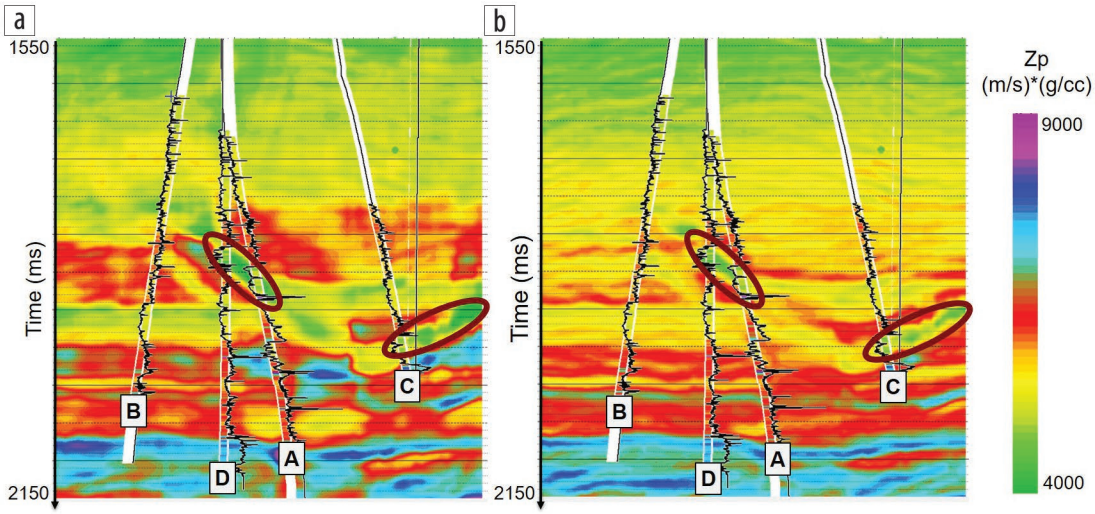

Figure 8. P-wave impedance results obtained through (a) simultaneous inversion and (b) the DNN methodology on an arbitrary line going through the four wells that were used to generate the pseudowells. The brown ellipses indicate the injectite sand reservoirs.

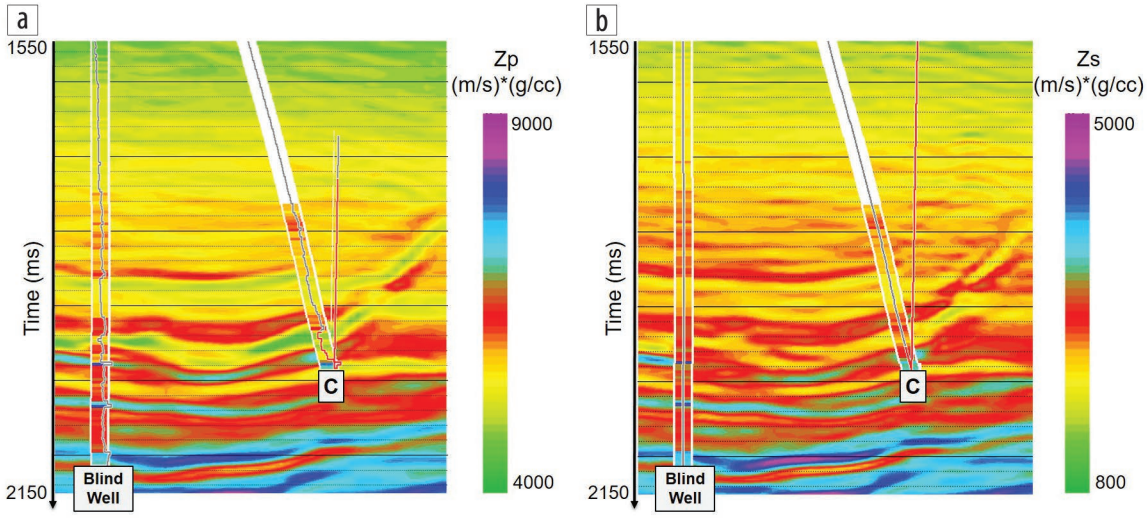

Figure 9. (a) P-wave and (b) S-wave impedance DNN results at an inline going through a well (blind well) that was not used to generate the pseudowells.

Recall from the DNN section that there are two training steps: first, we use stepwise regression to identify the attributes that serve as input features to the neural network; second, we train the DNN to find a nonlinear operator linking these input features to the target log. Table 1 shows the attributes identified through stepwise regression calculated on the four original wells (attribute set 1) and the synthetic catalog (attribute set 2). The instantaneous phase attribute is responsible for the poor continuity in Figure 12a and is probably selected due to a spurious correlation. This highlights the fact that the stepwise regression may not identify the best attributes to describe the target log if run only on a small amount of data.

To make a fairer comparison, Figure 12b shows the DNN estimate trained on the original wells using the same attributes as the synthetic catalog result (attribute set 2). The two results are closer now, but the DNN trained on the synthetic catalog (Figure 12c) is still more laterally continuous and matches the well control better than the DNN trained using the original well control (Figure 12b). In summary, Figure 12 shows that the synthetic catalog improves the result by improving the selection of features and by improving the DNN operator design.

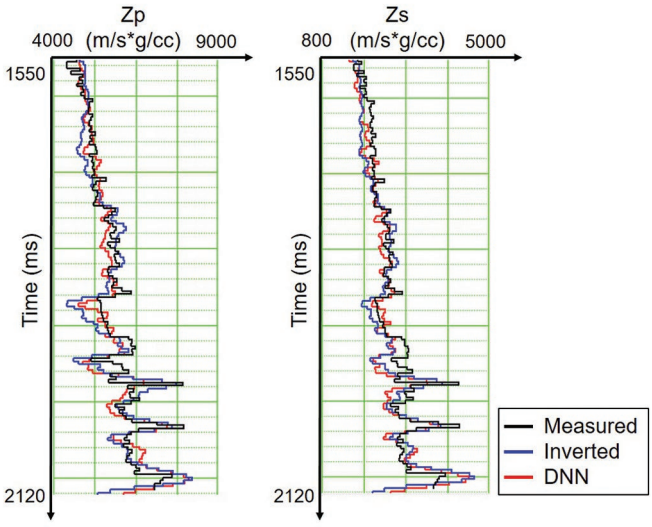

Figure 10. Comparison of the DNN (red) and simultaneous inversion (blue) P-wave and S-wave impedance estimates with the well-log values (black) at the blind well location. The correlations of the P-wave and S-wave impedance DNN estimates with respect to the well logs are 87% and 89%, respectively. These results are similar to the simultaneous inversion correlations of 89% and 91%.

Conclusions

In this study, we have presented a new TGDS-based approach for reservoir characterization. The methodology consists of first simulating many rock property logs based on the statistics of the available well control. Then, rock-physics theory is used to model the corresponding elastic response. Finally, we perform AVO modeling to generate synthetic seismic data. The pseudologs and synthetic data are then used to train a DNN. The trained DNN operators are then applied to the real seismic volume to obtain 3D volumes of reservoir properties.

The key step in the workflow is to use theory to generate training data for geologic situations not sampled by the original well control — for example, using fluid substitution to model hydrocarbon-filled reservoirs when only brine-filled reservoir examples are available. This ability to generate more training examples becomes more important as the reservoir depends on more variables. The existing well control will never sample all the possible geologic scenarios. Incorporating theory makes it possible to generate better sampled training data, resulting in DNN operators which generalize better.

This TGDS approach was applied to a North Sea data set to estimate a variety of different reservoir parameters, including P-wave impedance, S-wave impedance, and porosity, demonstrating the flexibility of the approach. In each case, the TGDS results obtained at a blind well location proved that the method generalizes quite well away from the original well control that is used to generate the synthetic catalog.

The TGDS P-wave impedance results compare favorably to those obtained by the theory-based deterministic inversion. The

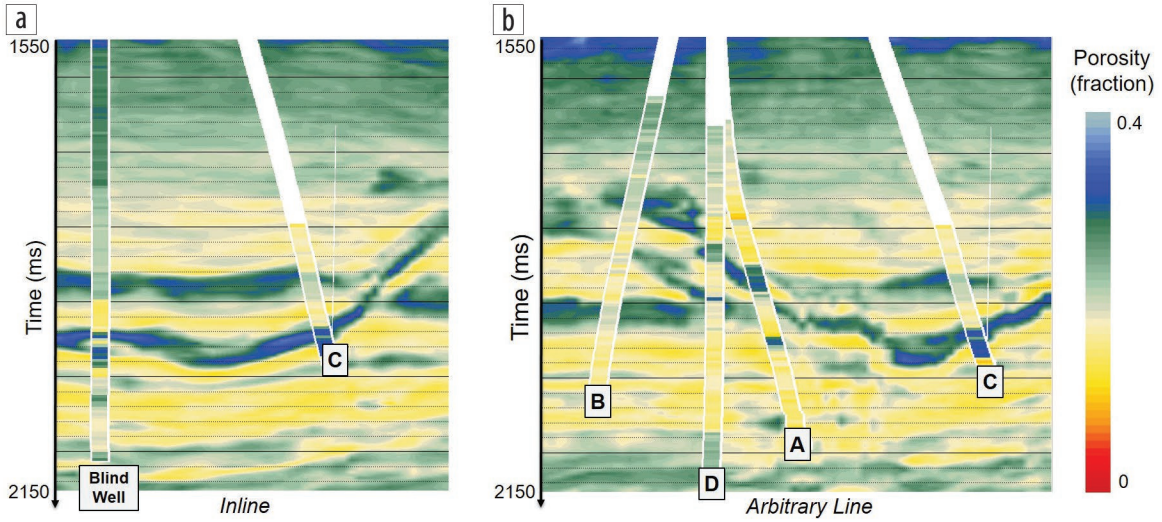

Figure 11. DNN porosity estimates at (a) the inline going through the blind well and (b) the arbitrary line going through the four wells that were initially used to generate the pseudowells.

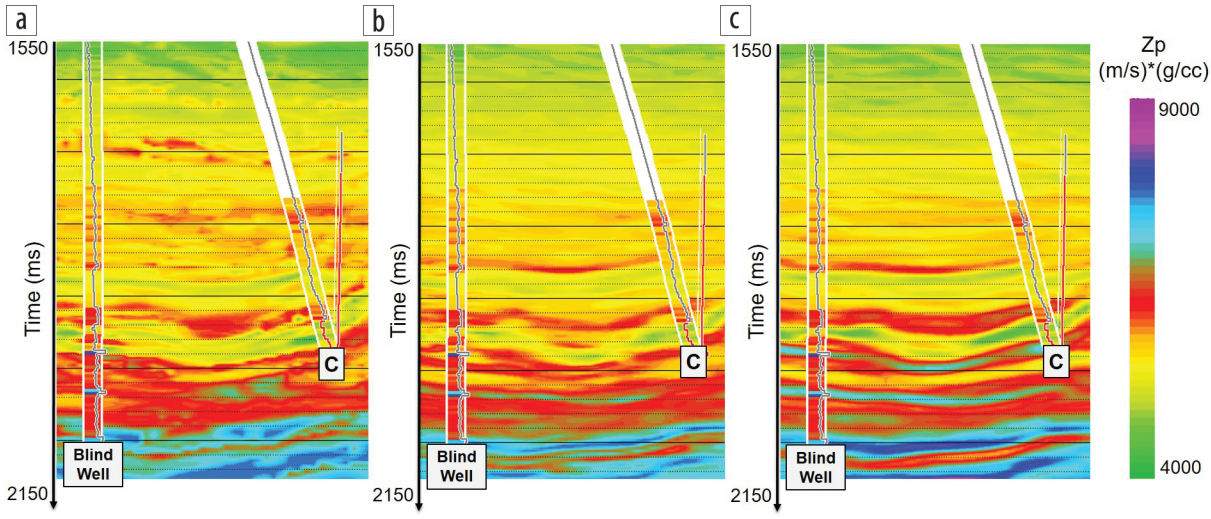

Figure 12. (a) P-wave impedance estimates from DNN trained on four initial wells using attributes derived from stepwise regression on the four initial wells (attributes set 1). (b) P-wave impedance estimates from DNN trained on four initial wells using attributes derived from stepwise regression on the synthetic catalog (attributes set 2). (c) P-wave impedance estimates from DNN trained on the synthetic catalog using attributes set 2. Note the improved lateral continuity and prediction accuracy at the well locations when using the synthetic catalog.

TGDS correlation coefficients to the well control range from 0.74 to 0.90, while those of the deterministic inversion range from 0.63 to 0.89. The TGDS results were superior to those obtained using a purely data science-based approach where the DNN was trained only from the original well control. This is due to the fact that the synthetic catalog increases the amount of training data and also allows us to model scenarios that may not be encountered at the well locations. This contributes to improving the selection of features used to train the DNN and the accuracy and adequacy of the DNN operator. This has practical implications for data science-based reservoir interpretation projects, suggesting it would be worthwhile incorporating synthetic data into the training. The TGDS method has been successfully applied to multiple data sets, including one that had only one well available in the study area. TLE

Acknowledgments

We wish to acknowledge Aker BP for its collaboration and permission to show its data. We are also grateful to our CGG colleagues, in particular Brian Russell, for their support and contribution to this work.

Data and materials availability

Data associated with this research are proprietary client data. Access could be granted, pending client approval.

Corresponding author: jon.downton@cgg.com

References

Allo, F., 2019, Consolidating rock-physics classics: A practical take on granular effective medium models: The Leading Edge, **38**, no. 5, 334–340, https://doi.org/10.1190/tle38050334.1.

Das, V., A. Pollack, U. Wollner, and T. Mukerji, 2019, Convolutional neural network for seismic impedance inversion: Geophysics, **84**, no.6, R869–R880, https://doi.org/10.1190/geo2018-0838.1.

Downton, J. E., and D. P. Hampson, 2018, Deep neural networks to predict reservoir properties from seismic: Presented at GeoConvention 2018, CSPG CSEG CWLS.

Draper, N. R., and H. Smith, 1998, Applied regression analysis: John Wiley & Sons.

Dvorkin, J., and A. Nur, 1996, Elasticity of high-porosity sandstones: Theory for two North Sea data sets: Geophysics, **61**, no. 5, 1363–1370, https://doi.org/10.1190/1.1444059.

Dvorkin, J., M. A. Gutierrez, and D. Grana, 2014, Seismic reflections of rock properties: Cambridge University Press, https://doi.org/10.1017/CBO9780511843655.

Gassmann, F., 1951, Über die elastizität poröser medien: Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich, **96**, 1–23.

Goodfellow, I., Y. Bengio, and A. Courville, 2016, Deep learning: MIT Press.

Halevy, A., P. Norvig, and F. Pereira, 2009, The unreasonable effectiveness of data: IEEE Intelligent Systems, **24**, no. 2, 8–12, https://doi.org/10.1109/MIS.2009.36.

Hampson, D. P., J. S. Schuelke, and J. A. Quirein, 2001, Use of multiattribute transforms to predict log properties from seismic data: Geophysics, **66**, no. 1, 220–236, https://doi.org/10.1190/1.1444899.

Hampson, D. P., B. H. Russell, and B. Bankhead, 2005, Simultaneous inversion of pre-stack seismic data: 75th Annual International Meeting, SEG, Expanded Abstracts, 1633–1637, https://doi.org/10.1190/1.2148008.

Hashin, Z., and S. Shtrikman, 1963, A variational approach to the theory of the elastic behavior of multiphase materials: Journal of the Mechanics and Physics of Solids, **11**, no. 2, 127–140, https://doi.org/10.1016/0022-5096(63)90060-7.

Karpatne A., G. Atluri, J. H. Faghmous, M. Steinbach, A. Banerjee, A. Ganguly, S. Shekhar, N. Samatova, and V. Kumar, 2017, Theory-guided data science: A new paradigm for scientific discovery: IEEE Transactions on Knowledge and Data Engineering, **29**, no. 10, 2318–2331, https://doi.org/10.1109/TKDE.2017.2720168.

Kennett, B. L. N., 1984, Seismic wave propagation in stratified media: Cambridge University Press.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012, ImageNet classification with deep convolutional neural networks: Proceedings of the 25th International Conference on Neural Information Processing Systems, 1097–1105.

Mur, A., and L. Vernik, 2019, Testing popular rock-physics models: The Leading Edge, **38**, no. 5, 350–357, https://doi.org/10.1190/tle38050350.1.

Simmons, J. L., and M. M. Backus, 1994, AVO modeling and the locally converted shear wave: Geophysics, **59**, no. 8, 1237–1248, https://doi.org/10.1190/1.1443681.

Smith, T. M., 2011, Practical seismic petrophysics: The effective use of log data for seismic analysis: The Leading Edge, **30**, no. 10, 1128–1141, https://doi.org/10.1190/1.3657071.

Soroka, W. L., T. J. Fitch, K. K. Van Sickle, and P. D. North, 2002, Successful production application of 3-D amplitude variation with offset: The lessons learned: Geophysics, **67**, no. 2, 379–390, https://doi.org/10.1190/1.1468598.

Wang, K., L. Bandura, D. Bevc, S. Cheng, J. DiSiena, A. Halpert, K. Osypov, B. Power, and E. Xu, 2019a, End-to-end deep neural network for seismic inversion: 89th Annual International Meeting, SEG, Expanded Abstracts, 4982–4985, https://doi.org/10.1190/segam2019-3216464.1.

Wang Y., Q. Ge, W. Lu, and X. Yan, 2019b, Seismic impedance inversion based on cycle-consistent generative adversarial network: 89th Annual International Meeting, SEG, Expanded Abstracts, 2498–2052, https://doi.org/10.1190/segam2019-3203757.1.

Zheng, Y., and Q. Zhang, 2018, Pre-stack seismic inversion using deep learning: Presented at 1st EAGE/PESGB Workshop on Machine Learning.

Zoeppritz, K., 1919, Erdbebenwellen VIII B, Über die Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen: Gottinger Nachr., **1**, 66–84.