geoconvention
Calgary • Canada • June 20-22 2022

Machine learning and lessons learned to improve Castagna's mudrock, Gardner's density, and Faust's velocity estimation.

David J. Emery, Marcelo Guarido*, Brian Russell*^, and Daniel Trad**
**University of Calgary, CREWES Consortium Data Science Initiative*
^ GeoSoftware, HampsonRussell

Summary

Determination of seismic lithology, porosity and pore fluid requires detailed modeling of petrophysical logs to improve the correlation with a seismic AVO response. Unfortunately, acquiring a full set of logs for all wells in a seismic survey is unpractical, and estimation of sonic, shear and density using empirical relationships is the standard approach. While these empirical relationships have worked for recon analysis, they have generally not given the details needed for accurate geophysical analysis. Machine Learning has given us a new way of investigating these relationships, and this presentation will provide insights we learned from analysis using over 138 wells with DT, Vs & RHOB logs from the North Sea, Australia, and Canada.

This paper build on work first published in the 2021 Calgary Geoconvention and presented in the SEG Machine Learning workshop in the spring of 2022 on a lithological determination using machine learn (Emery et al. 2021), using the FORCE 2020 well data from the North Sea (FORCE 2020), and then tested against wells in the Jeanne d'Arc Basin (Emery, 2006) & from the Poseidon 3D in NW Australia (Naeine & Prindle, 2018).

Introduction

A petrophysical log suite generally covers about 85% of a geological section, but full log sets, including sonic (DT), shear (DTS) and density (RHOB), typically make up less than 25%. The most significant data gap is the top hole but the increased use of logging while drilling (LWD) has making this problem worst as DT logs are generally not acquired as part of a LWD program.

The regression method (Faust 1951 & 1953, Gardner 1974, Castagna 1985) became the standard method to creating missing logs up and through the 1990s (Figure 1). These empirical relationships worked well for post-stack interpretation (Tanner 1977, Mitchell 1977, Wagoner 1988, Brown 1999) and inversion (Lindseth 1977, Aki 1980, Ostrander 1984, Shuey 1985, Russell & Hampson 1991, Castagna 1993).

The adoption of AVO analysis (Russell 1997, Goodway 1997, Connolly 1998, Castagna 1998, Hampson 2001 & 2005) as standard practice during the late '90s started to emphasize several shortcomings for these basic estimations. Applying rock physics standards (Mavko 1998, Avseth 2004, Emery 2006, Lee 2006, Russell 2013, Downton 2020) has significantly improved geophysical interpretation, providing knowledge of lithology and fluid types are known before estimating missing velocity and density information.

Machine Learning has recently become popular in petrophysics with multiple publications outlining its uses in lithology prediction (Hall 2016, Hall & Hall 2017, Guarido 2019, Emery 2020). Regrettably, most of these solutions require complete log data. Machine Learning has also been used to improve log prediction using support-vector machines, or SVM (Adeniran 2019, Anemangely 2019, Liu 2021, Sebtosheikh 2015), neural networks (Iwuoha 2019, Rezaee 2008), both (Mehrad 2022), or combinations of multiple methods (Azadpour 2020, Gupta 2018, Pan 2022, Zhang 2020). The results from these machine learning predictions have been impressive compared to estimates using older empirical relationships. However, the solutions have not been generalized nor they are as understandable as solutions derived from rock physics. The primary Machine Learning difficulties consist of the ill-conditioning of well logs, uneven sampling, finding the correct features engineering solutions, the need to impute missing values and overfitting.

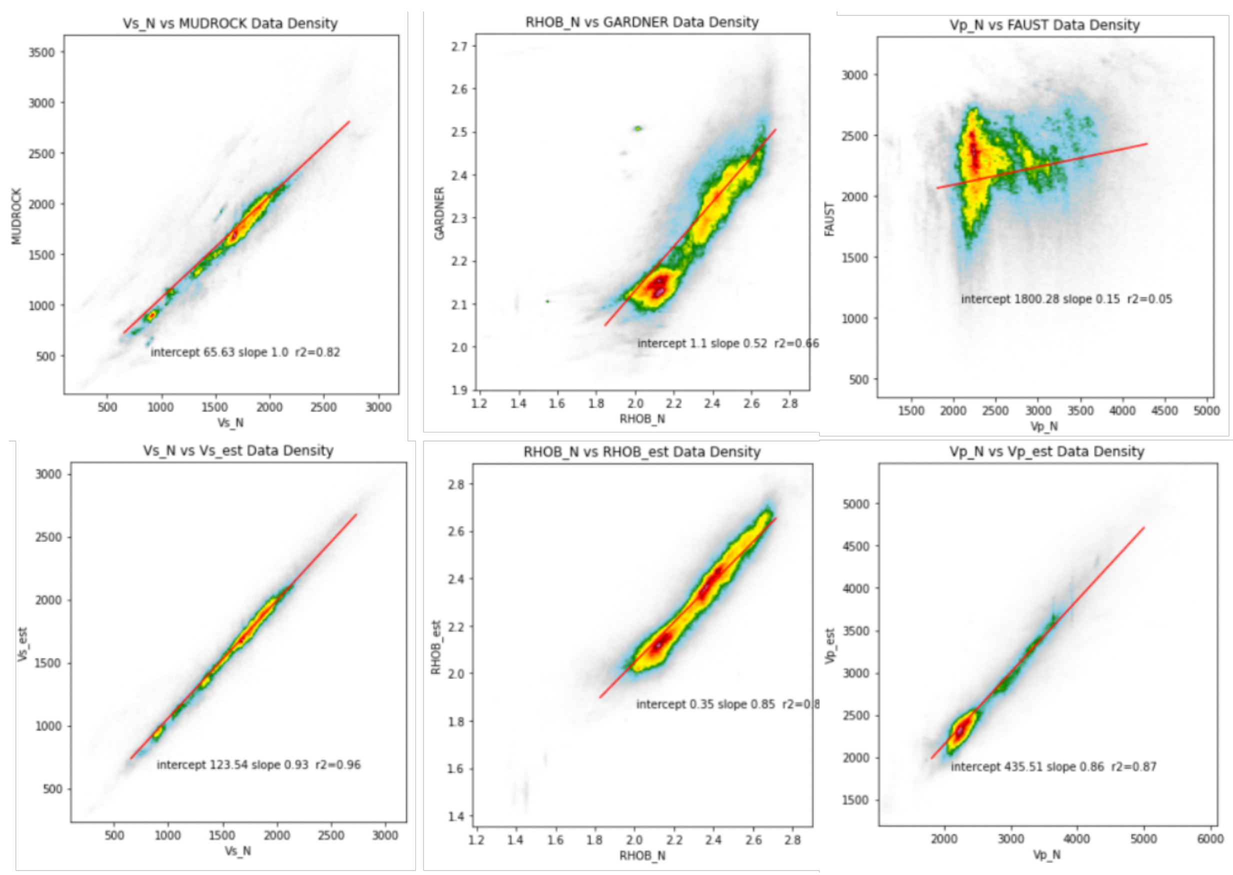

Figure 1: Algebraic Relationships (Mudrock left, Gardner center, Faust right) vs Machine Learn (Bottom)

During this presentation, we intend on sharing our observation using log normalization & depth trend removal on DT, RHOB & DTS estimation using XGBoost (Chen & Guestrin, 2016), one of the few Machine Learning solutions that don't require completed datasets.

Theory – Algebraic Relationships

Empirical relationships have generally been used to estimate missing log information. Estimating DT from resistivity (RES) was proposed by Hacirkoyulu (2006) after Faust (1953) using a combination of depth, geological age, and resistivity variation, and is written

$$V_p = \gamma (ZF)^{1/6} \; where \; F = {R_t}/{R_w} \qquad (1)$$

where γ = 2.2888, V_p is in km/s and Z in km. F, the formation factor, is defined as the formation resistivity (R_t) over the background water resistivity (R_w).

When the density logs are missing or questionable, it is common to use Gardner's relationship to estimate density values from the sonic log. The Gardner relationship (Gardner, Garner & Gregory 1974) is generalized for clastics as:

$$\rho = cV_P^{.25} \quad (2)$$

where c =1.741, if Vp is in km/s and ρ, is in g/cm3.

Shear (DTS) logs were not widely acquired until recently, and S-wave velocities are generally estimated from the P-wave sonic log. Several empirical Vp-Vs relationships have been proposed (Pickett, 1963; Tatham, 1980; Tosaya, 1982; Eastwood and Castagna, 1983; Castagna et al., 1985: Castagna et al., 1993; Greenberg and Castagna, 1992; Mavko, 1998) with the Castagna et al. (1985) mud-rock line being the most widely referenced relationship between P-wave and S-wave velocities. Castagna et al. (1985) proposed the mudrock line as an approximation to relate P-wave to S-wave values.

$$V_S = 0.8621 \times V_P - 1172 \qquad \text{Mud Rock Line – Castagna (1985)}$$

All three relationships, while helpful, were derived for log, seismic and laboratory measurements for Gulf Coast clastics and are reasonable first-order approximation for brine-wet shales and sandstones formations in similar depositional environments. Several modifications, using different experimental data sets, of the mudrock line have been proposed (Tosayn and Nur, 1982; Castagna, 1993; Han, 1986; and Mavko, 1998).

Machine Learn Method

For this work, we selected gradient boosting (Friedman 2001) as our machine learning engine since as an ensemble method it combines results from a series of decision trees, each solving for the residual error from the previous steps with different decision trees. The technique uses a logistic regression approach which produces a dichotomous outcome (yes/no). We use the implementation known as XGBoost (Chen & Guestrin, 2016) which conveniently adds the handling of missing data and increases speed by using parallel processing. In our test, this library has outperformed other tools.

As XGBoost builds multiple trees sequentially, each new tree corrects some errors made by a previous tree, the model can become computational expensive for large datasets with many features. Hyperparameter optimization is a crucial step to improve machine learning metrics but can also result in overfitting the training data making the solution area-specific (Emery 2022). For example, a geological framework has been shown to dominate both lithofacies and mineralogy classification (Emery 2021 & 2022). To make the XGB regression more general for the estimation of missing log values, we have removed all location and formation interpretations.

Feature Engineering

For this analysis, as training data we used 118 wells from the North Sea, made available by the Norwegian Petroleum Directorate for the FORCE 2020 competition. For testing datasets, we used 12 wells from Jeanne d'Arc Basin in Canada (Emery, 2006) and 8 wells in NW Australia (Naeine & Prindle, 2018). Three sub-datasets were generated from the 138 wells for DT, RHOB, & DTS to create 3 test and validation sets (118/20, 114/20, 46/20). All the wells were from offshore settings, and future work will incorporate Western Canada data.

The logs used for the analysis are DEPTH, gamma ray (GR), deep resistivity (DRES), median resistivity (MRES), shallow resistivity (SRES), neutron porosity (NPHI), density (RHOB), slowness (DT), shear slowness (DTS), photoelectric absorption (PEF), spontaneous potential (SP), caliper (CAL), and borehole size (BS). As the training data was 90% siliciclastic and dominantly shale, and we aimed to create a general solution, we chose to estimate a single global trend and analyze the residuals. While a more detailed porosity relationship (Ehrenberg 2009) could be appropriate, for simplicity, we used a linear depth trend estimated for each log type. For SP, where mud conditions dominate the response, a local trend was used to create the residual.

In addition, as compaction trends are referenced to burial depth (below seafloor), correcting for variation in water depth is essential. Again, for simplicity, the strategy used was to apply a bulk shift using the residual for an individual log from the background trend. This solution worked well except for the NW Australia data, that was more a carbonate system than siliciclastic, where we need to reduce the bulk shift for the resistivity and slowness logs.

Standard Petrophysical sub-products were also created: resistivity crossover (RDEP-RMED), average resistivity (RDEP+RMED, RDEP only RMED not available), NPHI-DPHI crossover, and the impedance (Vp x RHOB). To guarantee input independence for the XGBoost solution, testing was done using either the slowness (DT, DTS), the velocity (Vp, Vs), or the Vp & Vp/Vs ratio.

Multiple combinations of XGBoost tests were conducted using the raw input data, water bottom correction, and residual after removing the depth trend.

Results

As DTS is the most under-sampled petrophysical log, we will first cover its results first (Figure 2). XGBoost models were built using the FORCE2020 dataset and evaluated against the Jeanne d'Arc and NW Australia data. A feature importance analysis (Figure 3) for Vs estimation indicates a dominance of Vp followed by NPHI and then RDEP. A surprising outcome was the relatively low importance of GR and RHOB.

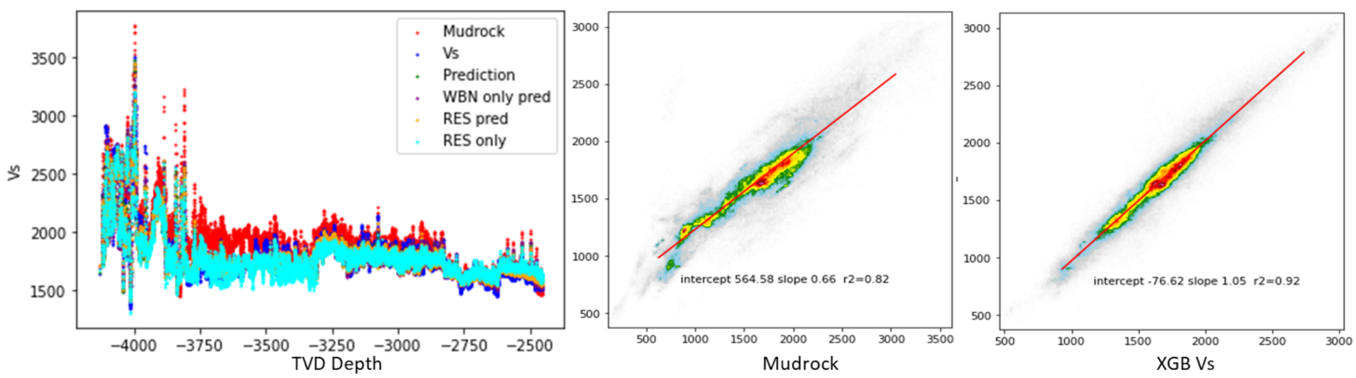

Figure 2: Vs prediction versus Mudrock & measure log (left), Mudrock vs real value (center), XGB prediction vs real (right)

Evaluation of RHOB indicates a high importance (Figure 3 center panel) again for Vp but modifying the role of NPHI and, to a less extent, RDEP. Surprisingly, GR seemed to have low importance in the estimations. The RHOB estimation shows a more balance feature importance, with Vp still being the most significant but a more balanced RDEP, NPHI, and GR. Minor importance was found from the PEF, RESdiff logs, and the lowest importance was found for the SP log.

In most circumstances, when DT needs to be estimated, RHOB and Vs would also be missing, and therefore the input log suite was reduced to GR, RDEP, RESdiff, NPHI, SP & PEF. The Faust approximation (Figure 1 – upper right) showed poor correlation and the result from XGBoost (Figure 1 - lower right) showed a much higher correlation. The feature importance indicates that Vp can be estimated from NPHI and DRES (Figure 3) and to a lesser GR and RESdiff.

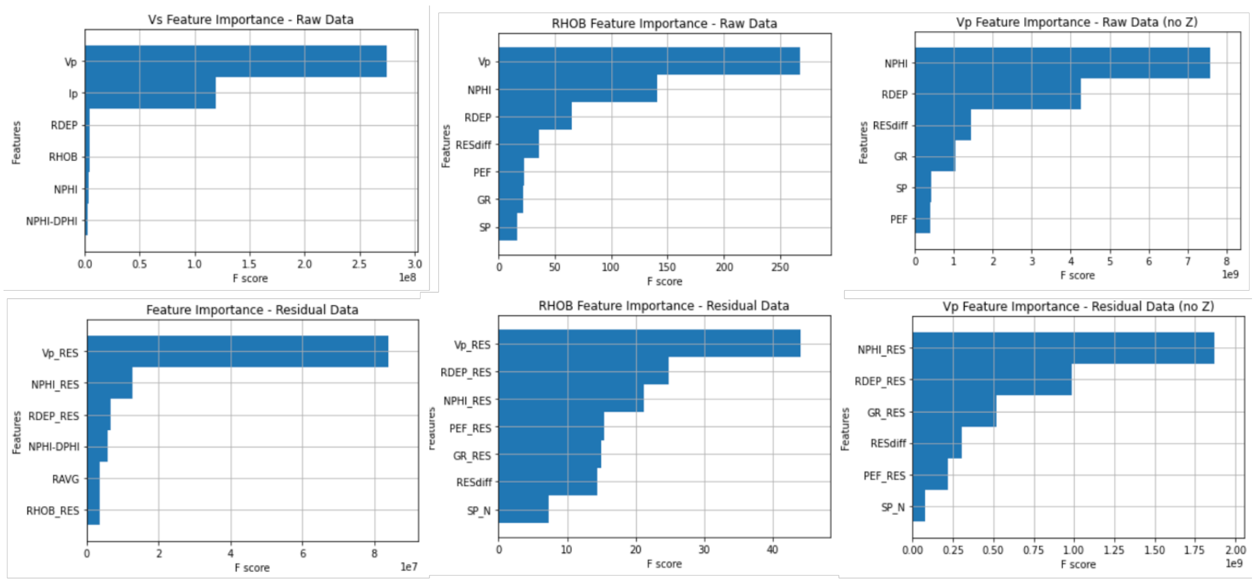

Figure 3: Feature Important (normal – top, residual – bottom) for Vs (left), RHOB (center), Vp (right)

Observations

The evaluation against the wells (Figure 4) in the Jeanne d'Arc Basin & NW Australia showed a mixed response. Machine learning estimation of Vp over Faust was superior (note the higher correlation). RHOB over Gardner was good, but Vp over Castagna was only fair. The amount of available log data and the degree of feature importance had a relationship to the overall performance of machine learning.

The Faust estimation of Vp has almost no correlation (R2 = 0.05), while machine learning shows a significant correlation (R2 = 0.76). The Vp estimation using only RDEP and GR reduces R2 (0.71), but the regional trend dominates this. Estimating changes from the residual show more of the difference in R2 (0.57 and 0.40 LWD) with an improved RMSE.

Predicting RHOB with the addition of RDEP & NPHI again show an improvement compared to Gardner's (R2 0.76 from 0.60, RMSE 26217 compared to 51009). A baseline shift error appears still to exist, and additional work will be required

Vp dominates the Vs estimated through machine learning and, as such, only shows a moderate difference from the Castagna mudrock calculation (R2 0.61 from 0.53, RMSE 53329 compared to 64406).

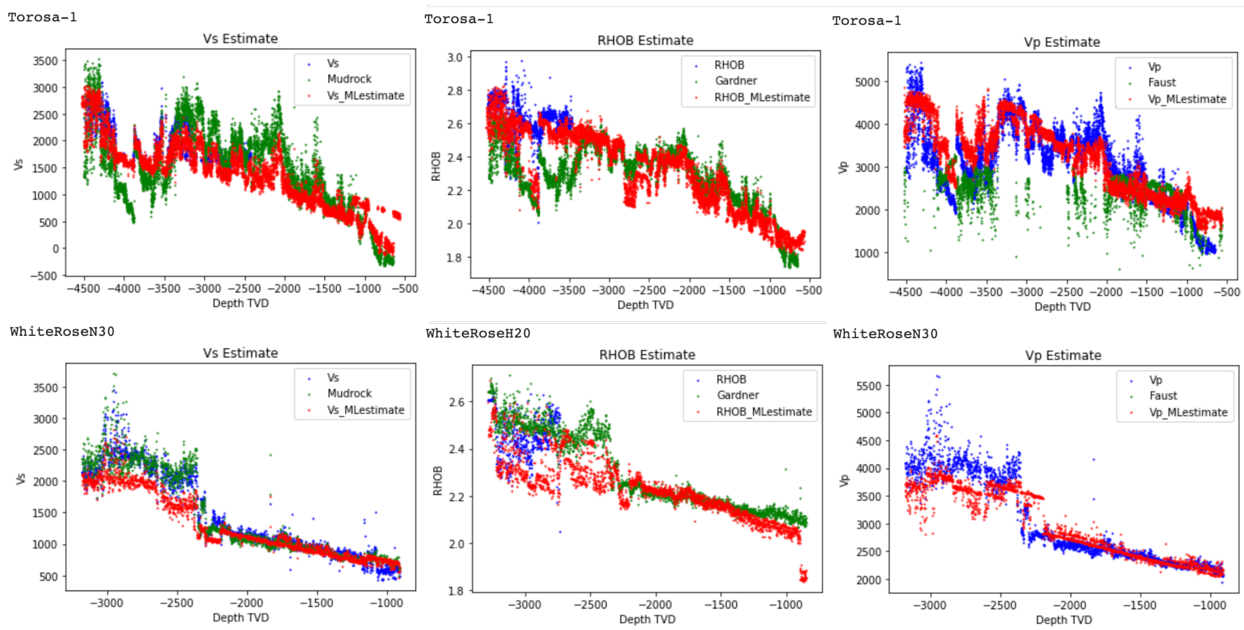

Figure 4: Example blind estimation, blue real, green empirical relationship & red result from machine learning

Conclusions

Machine Learning analysis for estimation of missing logs shows a significant improvement over previous approaches using empirical relationships. From the three relationships investigated, Faust, Gardner and Castagna, we found that the Castagna relationship was the closest to the machine learning results. While additional work is required, machine learning shows promise in estimating petrophysical logs and discovering new insights into the importance of the various parameters.

Acknowledgements

The authors thank the sponsors of CREWES for their continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13 and CRDPJ 543578-19, and the financial support from Canada First Research Excellence Fund.

References

Aki, K. and Richards, P.G., 1980, Quantitative Seismology: W.H. Freeman.

Abdurrachman, M.F., & Muhsin, MD, 2021, Reducing Uncertainties in Shear Wave Petrophysical Log Prediction by using Deep Neural Network and Machine Learning Methods; Indonesian Petroleum Association, 45 Convention

Adeniran, A.E., and Aizebeokhai, A.P., Compressional-Shear Velocity Model of "Toki" Field using Support Vector Regression, Offshore Niger Delta; 2019 J. Phys.: Conf. Ser. 1299 012085

Anemangely, M., Ramezanzaheh, A., Amiri, H., Hoseinpour, S., 2019, Machine learning technique for the prediction of shear wave velocity using petrophysical logs; Journal of Petroleum Science and Engineering, 174, 306-327

Avseth, P. A., and E. Odegaard, 2004, Well log and seismic data analysis using rock physics templates: First Break, 22, 37–43,

Azadpour, M., Saberi, M.R., Javaherian, A., & Shabani, M., 2020, Rock physics model-based prediction of shear wave velocity utilizing machine learning technique for a carbonate reservoir, southwest Iran; Journal of Petroleum Science and Engineering, 195, 15 pages

Brown, A.R., 1999, Interpretation of 3-dimensional seismic data (5'h ed.). AAPG Memoir 42, 341 p.

Castagna, J.P., Batzle, M.L., and Eastwood, R.L., 1985, Relationship between compressional-wave and shear-wave velocities in clastic silicate rocks: Geophysics, Vol. 50, p. 571-581

Castagna, J.P., Batzle, M.L., and Kan, T.K., 1993, Rock physics – The link between rock properties and AVO response: in Offset-Dependent Reflectivity – Theory and Practice of AVO Analysis, J.P. Castagna and M. Backus, eds. Investigations in Geophysics, No. 8, Society of Exploration Geophysicists, Tusa, Oklahoma, 135-171.

Castagna, J.P., Swan, H.W., and Foster, D.J., 1998, Framework for AVO gradient and intercept interpretation: Geophysics, Vol. 63, p. 948-956

Chen, T. & Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. New York, NY, USA: ACM, pp. 785–794.

Connolly, P. 1998, Elastic impedance. The Leading Edge; 18 (4): 438

Downton J.E, Collet, O., Hampson D.P, and Colwell, T., 2020, Theory-guided data science-based reservoir prediction of a North Sea oil field: The Leading Edge, October 2020, 742-749.

Eastwood, R.L. and Castagna, J.P., 1983. Basis for interpretation of Vs/Vp ratios in complex lithologies: Proceedings of 24th SPWLA annual logging symposium, paper G.

Ehrenberg, S.N., Nadeau, P.H., Steen, O., 2009. Petroleum reservoir porosity versus depth: Influence of geological age: AAPG Bulletin, v. 93, no. 10, pp. 1281–1296

Emery, D.J., Steward, R.R, 2006., S Using VP/VS to explore for sandstone reservoirs: well log and synthetic seismograms from the Jeanne d'Arc basin, offshore Newfoundland: CREWES Research Report, Vol. 18.

Emery, D.J., Guarido, M., Trad, D. O., and Innanen, K. A. H., 2021, Lessons Learned, Pitfalls and Feature Engineering for FORCE 2020: Log Facies Classification using Machine Learning; Geoconvention, Abstract

Emery, D.J., Guarido, M. and Trad, D.O., 2022, The Pitfalls and Insights of Log Facies Classification using Machine Learning: SEG Research Workshop: Data Analytics & Machine Learning for Exploration & Production, Abstract

Faust, L.Y. 1951, Seismic Velocity as a Function of Depth and Geologic Time, Geophysics, XVI, 192-206

Faust, L.Y. 1953, A Velocity Function including Lithologic Variation, Geophysics, XVIII, 271-288

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics, 1189–1232.

FORCE 2020, (2022, Feb). Results of the FORCE 2020 lithology machine learning competition. https://www.npd.no/en/force/Previous-events/results-of-the-FORCE-2020-lithology-competition/

Gardner, G.H.F., Garner, L.W., and Greogory, A.R., 1974, Formation velocity and density – The diagnostic basics for stratigraphic traps: Geophysics, 39, 770-780

Garotta, R., 1987, Two-component acquisition as a routine procedure: in Danborn, S.H. and Domenico, S.N., eds., Shear-wave exploration: Soc. Expl. Geophys., 122-136

Gassmann, F., 1951, Elastic waves through a packing of spheres: Geophysics, 16, 271-288

Goodway, W., Chen, T., and Downton, J., 1997, Improved AVO fluid Detection and Lithology Discrimination Using Lame Petrophysical Parameters: Extended Abstracts, SEG 67th Annual International Meeting, Denver.

Greenberg, M.L. and Castagna, J.P.,1992. Shear-wave velocity estimation in porous rocks: theoretical formulation, preliminary verification and applications: Geophys. Prospecting, 40, 195-209

Guarido, M., 2019, Machine Learning Strategies to Perform Facies Classification: GeoConvention, Abstract

Guarido, M., Emery, D. J., Macquet, M., Trad, D. O., and Innanen, K. A. H., 2020, The Pitfalls and Insights of Log Facies Classification for a Machine Learning Contest: CREWES Research Report, 32, 18.

Hall, B., 2016, Facies classification using machine learning: The Leading Edge, 35, No. 10, 906–909.

Hall, M., and Hall, B., 2017, Distributed collaborative prediction: Results of the machine learning contest: The Leading Edge, 36,No. 3, 267–269.

Hampson, D.P., Russell, B.H., and Bankhead, B., 2005, Simultaneous inversion of pre-stack seismic data: Extended Abstract, SEG Technical Program.

Hampson, D., Schuelke, J.S., and Quirein, J.A., 2001, Use of multi-attribute transforms to predict log properties from seismic data: Geophysics, 66, 220-231.

Hacikoylu, P., Dyorkin, J., and Mavko, G., 2006, Resistivity-velocity transforms revisited: The Leading Edge, 1006-1009.

Iwuoha, S.C, Per K. Pedersen, P.K., Clarkson, C.R., Gates, I.D., 2019, A working method for estimating dynamic shear velocity in the montney formation; Methods X, G, 1876-1993

Lee, M.W., 2006, A simple method of predicting S-wave velocity: Geophysics ,Vol. 71, No. 6 , 161-164.

Lindseth, R.O., 1976, Synthetic sonic logs-a process for stratigraphic interpretation: Geophysics, 44, 3-26.

Liu, Y., 2021, Petrophysical log recovering and core-log calibration using Machine Learning: Geoconvention, abstract

Mavko, G., Mukerji, T., and Dvorkin, J., 1998, The Rock Physics Handbook: Tools for seismic analysis in porous media. Cambridge University Press, 325 pgs.

Mitchum, R.M., Vail, P.R., and Thompson, S ., 1977, The depositional sequence as a basic unit for stratigraphic analysis . In, C.E. Payton (ed.), Seismic Stratigraphy - Applications to hydrocarbon exploration, AAPG Memoir 26, p . 53-62 .

Naeine, E.Z. and Prindle, K., 2018. Machine learning and learning from machines: The Leading Edge, 886-893

Nur, A., Mavko, G., Dvorkin, J., and Gal, D., 1995. Critical porosity: the key to relating physical properties to porosity in rocks: in Proc., 65th Ann. Int. Meeting, Soc. Expl. Geophys., 878.

Ostrander, W.1., 1984, Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence: Geophysics, 49, 1637-1648

Pickett, G.R., 1963. Acoustic character logs and their applications in formation evaluation: Journal of Petroleum Technology, 15, 650-667

Russell, B. and Hampson, D., 1991, A comparison of post-stack seismic inversion methods: Annual Meeting Abstracts, Society Of Exploration Geophysicists, 876-878.

Russell, B., Hampson, D ., Schuelke, J ., Quirein, J., 1997. Multiattribute seismic analysis . The Leading Edge, v. 16, p .1439-1443 .

Russell, B., & Lines, L., 2013. A Gassmann-consistent rock physics template; CSEG Recorder

Shuey, R. T., 1985, A simplification of the Zoeppritz-equations: Geophysics, Soc. of Expl. Geophys., 50, 609-614.

Tanner, M.T., and Sheriff, RE, 1977, Application of amplitude frequency and other attributes to stratigraphic and hydrocarbon determination, In, C.E. Payton (ed.), Seismic stratigraphy - applications to hydrocarbon exploration, AAPG Memoir 26, Tulsa, p . 301-327.

Tatham, R.H., 1980, Vp/Vs and lithology: 50th Ann. Internat. Mtg: Soc. of Expl. Geophys., Session:R.33.

Tosaya, C., and Nur, A., 1982. Effects of diagenesis and clays on compressional velocity in rocks: Geophs. Res. Lett., 9, 5-8

Walls, J.D., Taner, M.T., Guidish, T., Taylor, G., Duman, D., & Derzhi, N., North Sea reservoir characterization using rock physics, seismic attributes, and neural networks; a case history; Rock Solid Images, Report, 4 pages

Zhang, Y., Zhong, H.R., Wu, Z.Y., Zhou, H., Ma, Q., 2020, improvement of petrophysical workflow for shear wave velocity prediction based on machine learning methods for complex carbonate reservoirs; Journal of Petroleum Science and Engineering, 192, 14 pages