

Machine Learning Mineralogy Classification Comparison to Empirical Log Relationship and Implication for Physics Informed Modeling

David J. Emery, Marcelo Guarido, and Daniel O. Trad

ABSTRACT

Determination of seismic lithology, porosity and pore fluid requires detailed modelling of petrophysical logs to improve the correlation with a seismic AVO response. Unfortunately, acquiring a complete set of logs for all wells in a seismic survey is unpractical, and estimating 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. By analyzing over 138 wells with DT, Vs & RHOB logs from the North Sea, Australia, and Canada, we could generate synthetic Vp, Vs, and RHOB using traditional and the XGBoost regressor, where the latter showed to work better in this data.

INTRODUCTION

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

The regression method (Faust, 1951; Faust, 1953; Gardner, Garner, & Greogory, 1974; Castagna, Batzle, & Eastwood, 1985) became the standard method for creating missing logs up and through the 1990s (**Error! Reference source not found.**). These empirical relationships worked well for post-stack interpretation (Tanner & Sheriff, 1977; Brown, 1999) and inversion (Lindseth, 1976; Aki & Richards, 1980; Ostrander, 1984; Shuey, 1985; Russell & Hampson, 1991; Castagna, Batzle, & Kan, 1993).

The adoption of AVO analysis (Russell B. , Hampson, Schuelke, & Quirein, 1997; Goodway, Chen, & Downton, 1997; Connolly, 1998; Castagna, Swan, & Foster, 1998; Hampson, Schuelke, & Quirein, 2001; Hampson, Russell, & Bankhead, 2005) as standard practice during the late '90s started to emphasize several shortcomings for these basic estimations. Applying rock physics standards (Mavko, Mukerji, & Dvorkin, 1998; Avseth & Odegaard, 2004; Emery & Steward, 2006; Lee, 2006; Russell & Lines, A Gassmann-consistent rock physics template, 2013; Downton, Collet, P, & Colwell, 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 B. , 2016; Hall & Hall, 2017; Guarido, 2019; Emery D. J., Guarido, Trad, & Innanen, 2021). 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 (Aderiran & Aizeneokhai, 2019; Anemangely, Ramezanzahed, Amiri, & Hoseinpour, 2019; Liu, 2021), neural networks (Iwuoha, Pedersen, Clarkson, & Gates, 2019), or combinations of multiple methods (Azadpour, Saber, Javaherian, & Shabani; Zhang, Zhong, Wu, Zhou, & Ma, 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 are they as understandable as those 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.

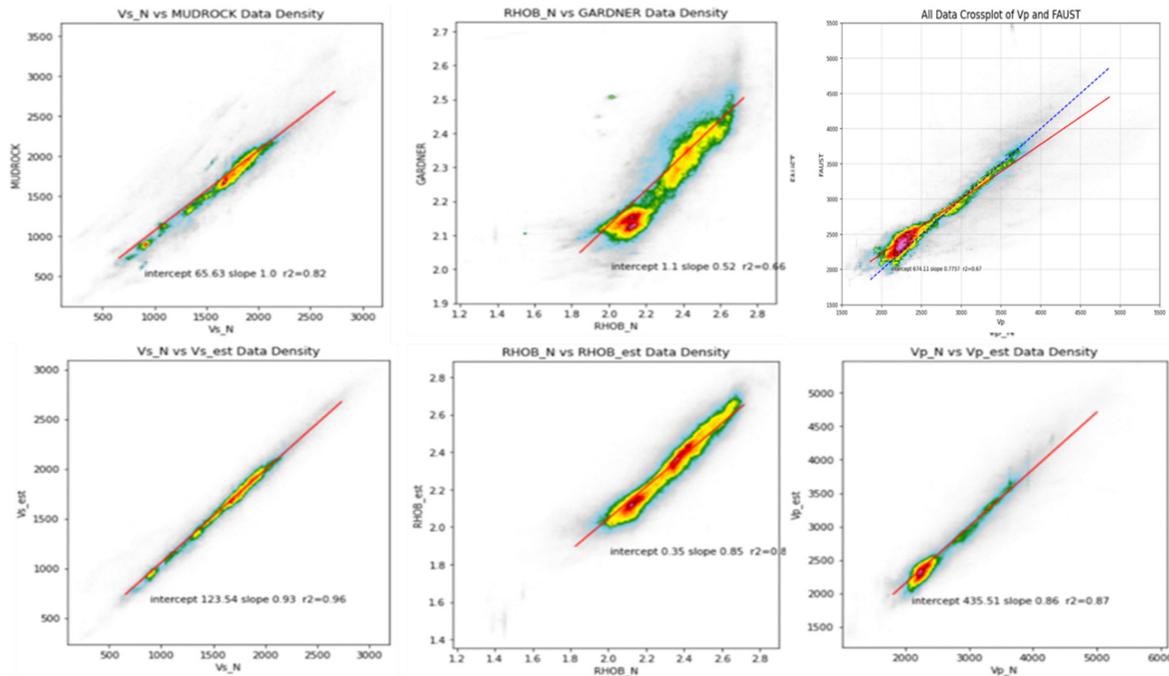


FIG. 1. Algebraic Relationships (Mudrock left, Gardner center, Faust right) vs XGBoost Machine Learn Estimation (Bottom).

In the first part of this report, we intend to share our observation following log normalization & depth trend removal on DT, RHOB & DTS estimation using XGBoost (Chen & Guestrin, 2016), one of the few Machine Learning solutions that do not require completed dataset. In the second part of this report, we will look at using these empirical relationships to evaluate mineralogy. In the final part of the report, we will share some of the observations of Vp, Vs and RHOB's relationship with mineralogy.

THEORY – ALGEBRAIC RELATIONSHIPS

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

$$V_p = \gamma(ZF)^{\frac{1}{6}} \text{ where } F = R_t/R_w \quad (1)$$

where $\gamma = 2.2888$, 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, & Greogory, 1974) is generalized for clastic as:

$$\rho = cV_p^{2.5} \quad (2)$$

where $c = 1.741$ if V_p 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 V_p - V_s relationships have been proposed (Pickett, 1963; Tatham, 1982; Tosaya & Nur, 1982; Eastwood & Castagna, 1983; Castagna, Batzle, & Eastwood, 1985; Castagna, Batzle, & Kan, 1993; Greenberg & Castagna, 1992; Mavko, Mukerji, & Dvorkin, 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:

$$V_s = 0.8621 \times V_p - 1172 \quad (3)$$

All three relationships, while helpful, were derived for log, seismic and laboratory measurements for Gulf Coast clastics at reservoir depth and are a reasonable first-order approximation for brine-wet shales and sandstone formations in similar depositional environments. Several modifications, using different experimental data sets, of the mudrock line have been proposed (Tosaya & Nur, 1982; Castagna, Batzle, & Kan, 1993; Mavko, Mukerji, & Dvorkin, 1998), but again, all concentrate on standard reservoir depths.

More recent proposed methods for estimating V_p , V_s and RHOB are based in physics-based from rock lab-derived Modulus (Mavko 1988; Dvorkin & Nur 1996), V_p/V_s vs I_p (Avseth & Ødegaard, 2004; Avseth et al. 2009 & 2010) or from Lambda-Rho/Mu-Rho relationships (Goodway 1997 & 2001, Hoffe et al. 2008).

Estimating V_p & V_s assuming an isotropic rock can be derived from the bulk modulus (K), shear modulus (μ) and density (ρ) using:

$$V_p = \sqrt{\frac{k+4/3\mu}{\rho}} \quad \text{and} \quad V_s = \sqrt{\frac{\mu}{\rho}} \quad (4)$$

To estimate V_p & V_s , the first one needs to determine the mineral values for K and ρ (see appendix) and then estimate the variation in values with porosity. This variation with porosity generally requires estimating using a combination of a soft (Reuss lower) or hard (Voigt upper) estimation.

Reuss (soft lower bound) is the minimum possible moduli for a material by averaging parameters using the denominator:

$$\frac{1}{M_R} = \sum_1^N f_i \frac{1}{M_i} \quad (5)$$

and the Voigt or the upper bound (hard) represents a straight numerical averaging of moduli (maximum possible).

$$M_V = \sum_i^N f_i M_i \quad (6)$$

where in equations 5 & 6, f_i is the volume fraction of the i^{th} component with a modulus M_i .

Density (ρ) generally follows a Voigt averaging, while bulk (K) & shear moduli (μ) generally lie between the bounds. Various averaging techniques such as Voigt-Reuss-Hill, Hashin-Shtrikman, or Kuster-Toksoz, along with the more popular Gassmann's formulation, have been proposed. Nur proposed (Mavko, 1998) a more physical basis for associating a rock's mineralogy to its log properties by assuming that the bulk and shear modulus are equivalent to the mineral grains at extremely low porosities and the Reuss bound when the rock loses cohesion (Φ_c critical porosity). This report is not intended to investigate the relative merits of these techniques but to use the derived machine learning mineralogy to determine the associated petrophysical properties and, thus, potential match to rock lab-derived properties.

The isotropic formulas for deriving moduli's from V_p , V_s and density (ρ) are:

$$K = \rho(V_p^2 - \frac{4}{3} V_s^2) \quad \mu = \rho V_s^2 \quad (7)$$

Likewise, the Poisson's ratio can be derived using V_p and V_s by:

$$\sigma = \frac{V_p^2 - 2V_s^2}{2(V_p^2 - V_s^2)} \quad (8)$$

Another rock-physic model involves cross-plotting of V_p/V_s versus I_p , which is relatively straightforward using petrophysical logs. The final common cross-plot we use for evaluating V_p , V_s and density is the Lambda-Rho ($\lambda\rho$) Mu-Rho ($\mu\rho$) variation in lithology and fluids (Goodway 1997 & 2001, Hoffe et al. 2008). The formulas used for estimating Lambda-Rho (λ) and Mu-Rho (μ) are:

$$\lambda = \rho(V_p^2 - 2V_s^2) \quad \mu = \rho V_s^2 \quad (9)$$

As mentioned earlier, most of these empirical relationships, along with the rock physic formulation, are for siliciclastic rocks at reservoir depths, temperatures, and pressure. As the petrophysical log used for the Machine Learning solution in this report vary from near the water bottom to over 5000 meters below the sea floor, some variation is expected. Fortunately, the input mineralogy is dominantly siliciclastic (88%), with shale making up 72% of the overall, thus providing a reasonable background trend that we can use to normalize data between wells and derive mineralogical relationships.

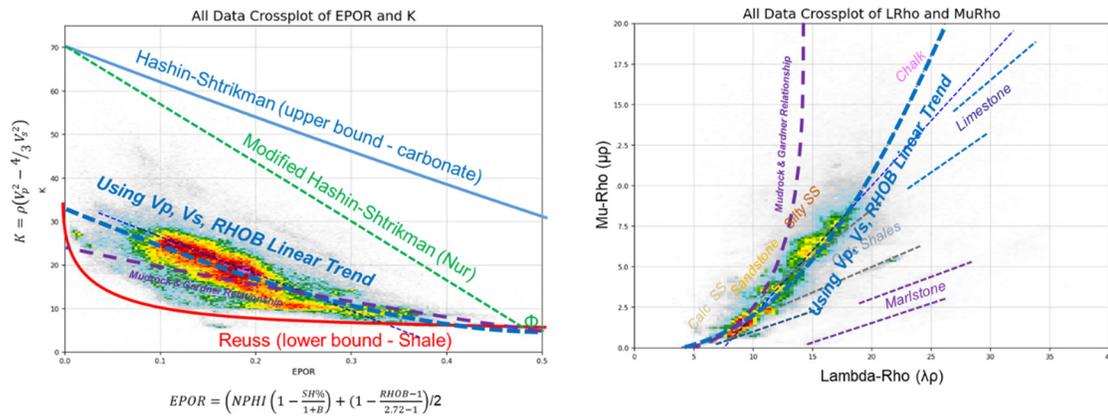


FIG. 2. Bulk Modulus (left) and Lambda-Rho/Mu-Rho (right) cross plots of input data.

The cross plot (**Error! Reference source not found.**) shows the input data using the derived bulk modulus (K) and the more common Lambda-Rho ($\lambda\rho$) Mu-Rho ($\mu\rho$); the colouring is by data density and is dominated by the shale. The right-hand panel also contains labels for some of the trace mineralogy. Please see the appendix if you are interested in cross-plots of the various mineralogy.

MACHINE LEARNING 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, producing a dichotomous outcome (yes/no). We use the implementation known as XGBoost (Chen & Guestrin, 2016), which conveniently adds to the handling of missing data and increases speed by parallel processing. In our test, this library outperformed other tools.

As XGBoost builds multiple trees sequentially, each new tree corrects some errors made by a previous tree; the model can become computationally 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, Guarido, & Trad, 2022). For example, a geological framework has been shown to dominate lithofacies and mineralogy classification (Emery D. J., Guarido, Trad, & Innanen, 2021; Emery, Guarido, & Trad, 2022). We have removed all location and formation interpretations to make the XGB regression more general for estimating missing log values.

FEATURE ENGINEERING

For this analysis, as training data, we used 118 North Sea wells 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 & Steward, 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), calliper (CAL), and borehole size (BS). As the training data was 88% siliciclastic and dominantly shale, and we aimed to create a general solution, we chose to estimate a single global trend and analyze the residuals (FIG 3). While a more detailed porosity relationship (Ehrenberg, Nadeau, & Steen, 2009) could be appropriate, for simplicity, we used a linear depth trend estimated for each log type. A local trend was used to create the residual for SP, where mud conditions dominate the response.

	Trend Removed			Residual Statistics			
	intercept	slope	r^2	Coverage	Mode	Mean	STD
GR	63.16	-0.0028	0.0060	100%	0.0	0.157	25.891
RDEP	0.027	-0.000165	0.1530	99%	0.0	0.064	0.372
RMED	0.027	-0.000165		97%	0.0	0.054	0.316
NPHI	0.506	0.0000741	0.3190	67%	0.0	-0.005	0.092
RHOB	1.85	-0.00019	0.5480	87%	0.0	-0.015	0.155
DPHI	0.5	0.00011	0.5480	87%	0.0	-0.009	0.091
Vp	1550.3	-0.6252	0.5340	93%	0.0	68.717	488.260
Vs	797	-0.3202	0.3030	20%	0.0	20.678	313.334
PEF	4.13	-1.23E-05	0.0000	61%	0.0	0.160	0.753
SP	nan	nan	nan	68%	0.0	0.026	2.892
NPHI-DPHI	nan	nan	nan	67%	-0.11	-0.100	0.102

FIG. 3. Trends estimated from input logs and statistics for the residuals.

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, which was more a carbonate system than siliciclastic, where we needed 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 ($V_p \times RHOB$). To guarantee input independence for the XGBoost solution, testing was done using either the slowness (DT, DTS), the velocity (V_p , V_s), or the V_p & V_p/V_s ratio.

The Lithoclasses provided with the 118 wells from the FORCE 2020 competition were divided into Mineralogy (Emery, 2022). While each class had significant overlap, FIG 4 contains a summary of the mean for each residual attribute.

	Cal SS	Sandstone	Silty-SS	Silty-shale	Shale	Organic-SH	Claystone	Marlstone	Calc-shale	Arg-Limestone	Limestone	P-Limestone	Dolomite	Anhydrite	Halite	Tuff	Coal	Basement
GR	-30	-28	-3.4	2.7	11.2	49	34	-15	-11	-14	-45	-42	-12	-38	-30	-23	-1	-9.1
RDEP	0.36	0.00	0.12	0.04	-0.07	0.21	-0.14	0.09	0.00	0.19	0.43	0.13	0.16	1.60	2.18	0.00	0.56	0.08
RMED	0.42	-0.01	0.13	0.04	-0.05	0.23	-0.13	0.10	0.04	0.32	0.45	0.15	0.15	1.56	0.73	0.00	0.65	0.08
NPHI	-0.17	-0.07	-0.03	0.02	0.03	0.08	0.10	-0.06	-0.05	-0.10	-0.19	-0.14	-0.05	-0.27	-0.20	0.07	0.15	-0.03
RHOB	0.01	-0.10	-0.03	-0.02	-0.01	-0.09	-0.11	0.12	0.12	0.15	0.20	0.13	0.14	0.40	-0.43	-0.02	-0.57	0.00
DPHI	0.01	-0.06	-0.02	-0.01	-0.01	-0.05	-0.06	0.07	0.07	0.09	0.12	0.08	0.08	0.23	-0.25	-0.01	-0.34	0.00
Vp	649.0	235.5	212.0	-5.4	-177.4	-226.2	-392.5	393.3	186.7	898.0	1396.6	1104.8	459.0	1781.8	446.6	100.0	-483.2	205.2
Vs	418.0	143.8	108.8	56.1	-98.1	-129.5	-316.8	-18.0	132.5	495.6	639.9	825.0	6.2	1065.1	nan	98.6	-337.2	83.3
PEF	0.13	-0.01	0.06	0.21	0.13	0.26	0.34	0.41	0.21	0.27	0.38	0.36	0.34	0.69	2.52	0.14	-0.36	0.13
SP	-1.00	-0.47	0.09	0.13	0.05	0.41	0.10	0.17	0.05	-0.01	-0.25	0.20	-0.24	-0.72	0.06	0.03	0.41	0.18
NPHI-DPHI	0.04	0.01	-0.08	-0.14	-0.14	-0.17	-0.14	-0.10	-0.13	-0.08	-0.04	-0.05	-0.14	-0.07	0.31	-0.15	0.02	-0.09

FIG. 4. Mean values by mineralogy for residual following trend removal.

RESULTS

As DTS is the most under-sampled petrophysical log, we will cover its results first (FIG 5). XGBoost models were built using the FORCE2020 dataset and evaluated against the Jeanne d’Arc and NW Australia data. A feature importance analysis (**Error! Reference source not found.** 6) for Vs estimation indicates the dominance of Vp followed by NPHI and RDEP. A surprising outcome was the relatively low importance of GR and RHOB.

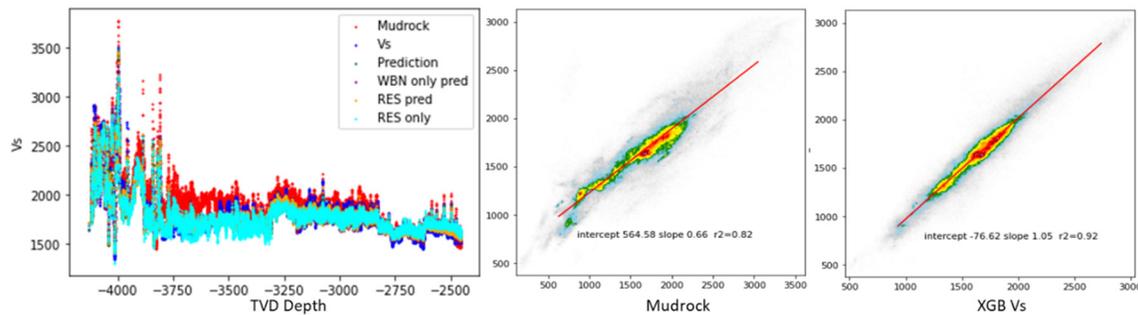


FIG. 5. Vs prediction versus Mudrock & measure log (left), Mudrock vs real value (center), XGB prediction vs real (right).

Evaluation of RHOB indicates high importance (**Error! Reference source not found.** 6, 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 balanced feature importance, with Vp still being the most significant but a more balanced RDEP, NPHI, and GR. Minor importance was found in the PEF and RESdiff logs, and the lowest importance was found in 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 (**Error! Reference source not found.**, upper right) showed a low correlation, and the result from XGBoost (**Error! Reference source not found.**, lower right) showed a much higher correlation. The feature importance indicates that Vp can be estimated from NPHI and DRES (FIG 6) and to a lesser GR and RESdiff.

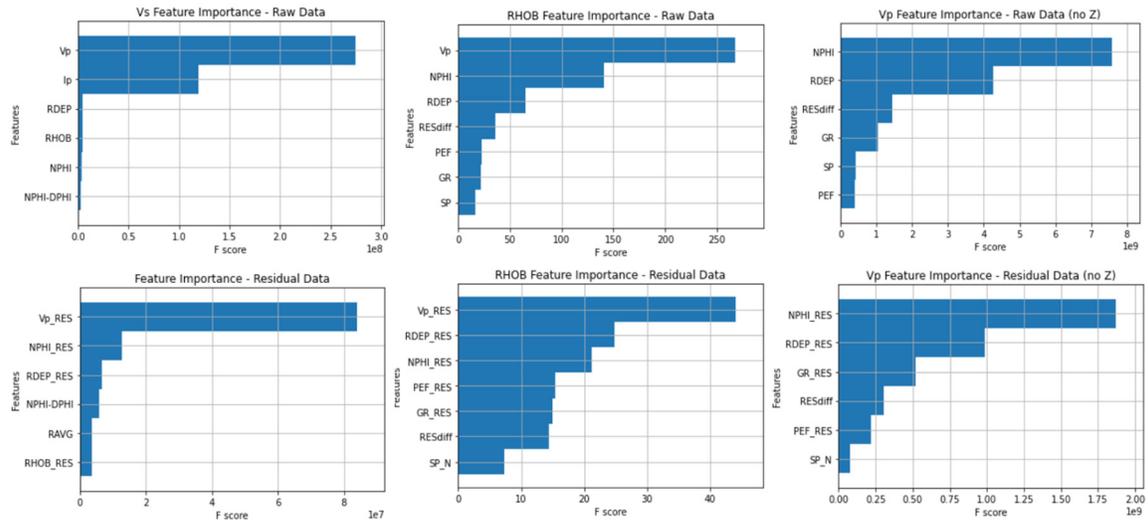


FIG. 6. Feature Importance (normal – top, residual – bottom) for Vs (left), RHOB (center), Vp (right).

OBSERVATIONS

The evaluation against the wells (FIG 7) 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 ($R^2 = 0.05$), while machine learning shows a significant correlation ($R^2 = 0.76$). The Vp estimation using only RDEP and GR reduces R^2 (0.71), but the regional trend dominates this. I am estimating changes from the residuals to show more of the difference in R^2 (0.57 and 0.40 LWD) with an improved RMSE.

Predicting RHOB with the addition of RDEP & NPHI again shows an improvement compared to Gardner’s (R^2 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 (R^2 0.61 from 0.53, RMSE 53329 compared to 64406).

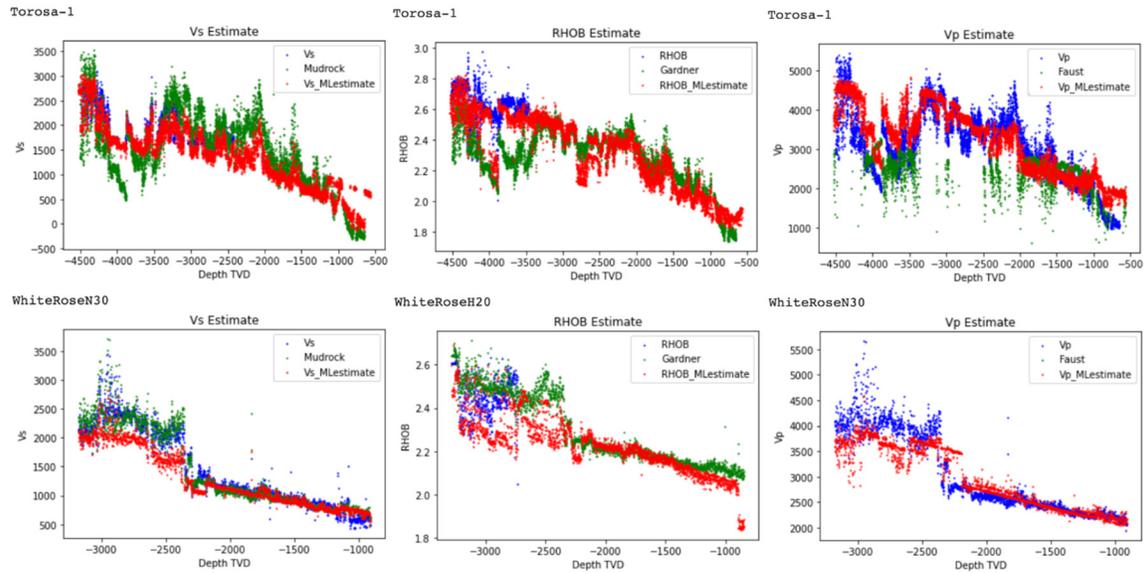


FIG. 7. Example blind estimation, blue real, green empirical relationship & red result from machine learning.

The results are more striking when comparing the performance of estimating mineralogy using logs, empirical relationships, or machine learning estimation (FIG 8). The Machine Learning logs quickly improve the empirical relationship but still fall short of matching the results from using the actual logs. Upon investigation, three significant problems were observed:

- As expected, the Empirical Relationship performed the worst and tended not to add much additional information for the estimation of mineralogy as they predominately represent a rescaling of existing information. Likewise, the observed shift between the empirical relationship and the input data is believed to be the result of the various geological location (North Sea vs Gulf of Mexico) and the broad variation in burial depths (0-5500m).
- The machine learning estimated logs were best when using the residual attributes but still tended to favour the larger classes and performed inadequately in areas with poor log coverage (FIG 9).
- The solution for the Poseidon wells, which were in a predominately carbonate environment, had a reasonable log prediction for shape but tended to have a poor machine learning estimation of Mineralogy. Likely the result of the under-sampling of carbonates in the learning data and errors in the log normalization.

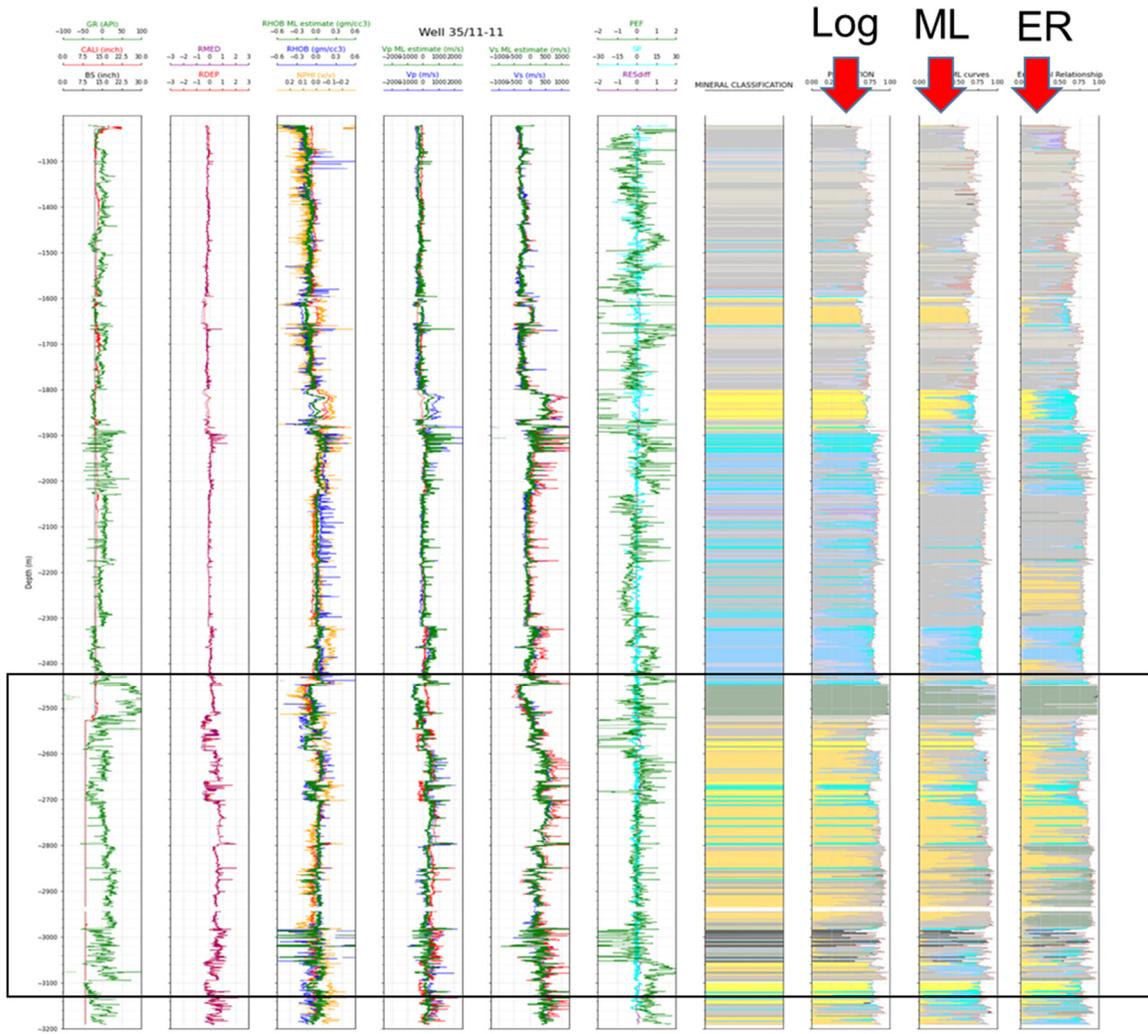


FIG. 8. Mineralogy estimation using petrophysical logs (Log), machine learn estimated logs (ML) and empirical relationship (ER).

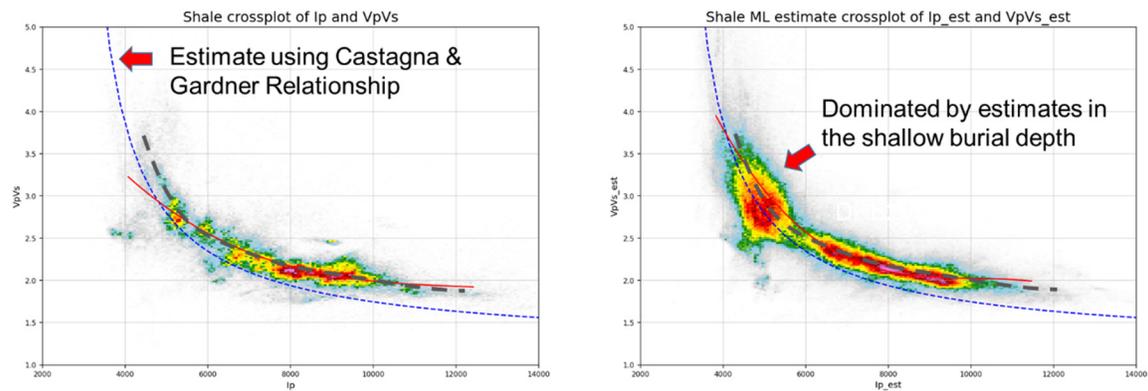


FIG. 9. Comparison of Empirical Relationship with input shale mineralogy (left) and XGBoost estimation (right). The large scatter centred around $I_p=5000$ results from the shortage of shallow input data (under 2000m burial depth) coupled with the uncertainty in mineralogy.

FUTURE WORK - CROSS PLOT OBSERVATIONS

We were fortunate to provide a petrophysical analysis for a learning dataset. However, the small sample size for the carbonates and trace mineralogy does cause significant difficulty in mineralogy classification and estimation of missing logs required for AVO. The ideal solution would be estimating logs using a more physic-informed approach, and contained in the appendix is the cross plot of the normalized input data, which we hope to use for future work in 2023. A preview of this analysis is contained in FIG 10, indicating the mineral within the data set.

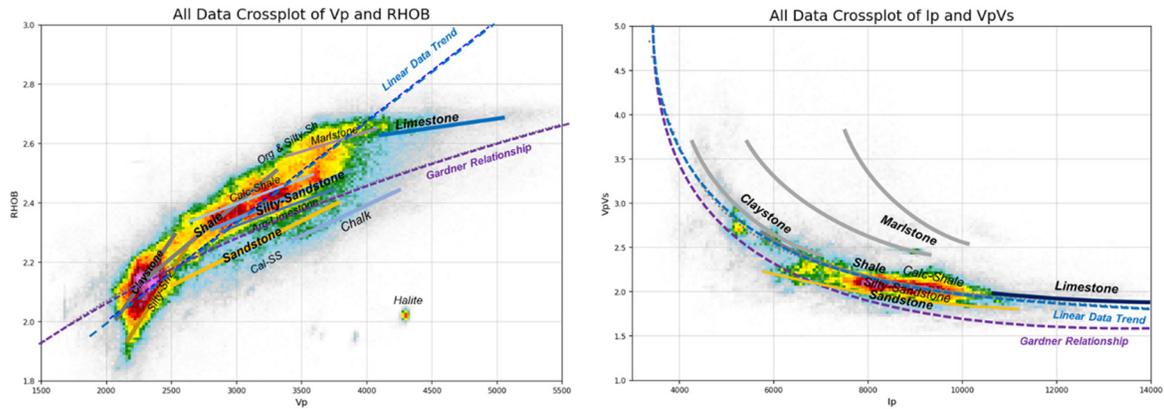


FIG. 10. Vp vs RHOB with Gardner line (left) and Vp/Vs vs Ip (right). Lines represent the most common profile for selective mineralogizes.

CONCLUSIONS

Machine Learning analysis for the 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.

ACKNOWLEDGMENTS

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

- Abdurrachman, M. F., & Muhsin, M. D. (2021). Reducing Uncertainties in Shear Wave Petrophysical Log Prediction by using Deep Neural Network and Machine Learning Methods. *Indonesian Petroleum Association*.
- Aderiran, A. E., & Aizeneokhai, A. P. (2019). Compressional-Shear Velocity Model of Toki Field using Support Vector Regression, Offshore Niger Delta. *J. Phys.*
- Aki, K., & Richards, P. G. (1980). *Quantitative Seismology*. W. H. Freeman.

- Anemangely, M., Ramezanzadeh, 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., & Odegaard, E. (2004). Well Log and Seismic Data Analysis using Rock Physics Templates. *First Break*, 22, 37-43.
- Avseth, P., Jørstad, A., van Wijngaarden, A. J., & Mavko, G. (2009). Rock physics estimation of cement volume, sorting, and net-to-gross in North Sea sandstones. *The Leading Edge*, 28(1), 98-108.
- Avseth, P., Mukerji, T., Mavko, G., & Dvorkin, J. (2010). Rock-physics diagnostics of depositional texture, diagenetic alterations, and reservoir heterogeneity in high-porosity siliciclastic sediments and rocks—A review of selected models and suggested workflows. *Geophysics*, 75(5), 75A31-75A4.
- Azadpour, M., Saber, M. R., Javaherian, A., & Shabani, M. (n.d.). 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.
- Brown, A. R. (1999). Interpretation of 3-dimensional seismic data. *AAPG Memoir*, 42, 341.
- Castagna, J. P., Batzle, M. L., & Eastwood, R. L. (1985). Relationship Between Compressional-Wave and Shear-Wave Velocities in Clastic Silicate Rocks. *Geophysics*, 50, 571-581.
- Castagna, J. P., Batzle, M. L., & Kan, T. K. (1993). Rock physics – The link between rock properties and AVO response: in Offset-Dependent Reflectivity – Theory and Practice of AVO Analysis. In J. Castagna, & M. Backus, *Investigations in Geophysics* (pp. 135-171). Tulsa, Oklahoma: Society of Exploration Geophysicists.
- Castagna, J. P., Swan, H. W., & Foster, D. J. (1998). Framework for AVO gradient and intercept interpretation. *Geophysics*, 63, 948-956.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Connolly, P. (1998). Elastic Impedance. *The Leading Edge*, 18(4), 438.
- Downton, J. E., Collet, O., P, H. D., & Colwell, T. (2020). Theory-Guided Data Science-Based Reservoir Prediction of a North Sea Oil Field. *The Leading Edge*, 742-749.
- Dvorkin, J., & Nur, A. (1996). Elasticity of high-porosity sandstones: Theory for two North Sea data sets. *Geophysics*, 61(5), 1363-1370.
- Eastwood, R. L., & Castagna, J. P. (1983). Basis for Interpretation of Vs/Vp Ratios in Complex Lithologies. *Proceedings of 24th SPWLA Annual Logging Symposium*.
- Ehrenberg, S. N., Nadeau, P. H., & Steen, O. (2009). Petroleum reservoir porosity versus depth: Influence of geological age. *AAPG Bulletin*, 93(10), 1281-1296.
- Emery, D. J., & Steward, R. R. (2006). Using VP/VS to explore for sandstone reservoirs: well log and synthetic seismograms from the Jeanne d'Arc basin, offshore Newfoundland. *CREWES Research Report*, 18.
- Emery, D. J., Guarido, M., & 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*.
- Emery, D. J., Guarido, M., Trad, D. O., & Innanen, K. A. (2021). Lessons Learned, Pitfalls and Feature Engineering for FORCE 2020: Log Facies Classification using Machine Learning. *Geoconvention 2021*.
- Emery, D., Guarido, M., Russell, B., & Trad, D. (2022). Machine learnings and lessons learned on improvements to Castagna's mudrock, Gardner's density, and Faust's velocity estimation. *International Meeting for Applied Geoscience & Energy* (p. 4). Houston: SEG/AAPG.
- Faust, L. Y. (1951). Seismic Velocity as a Function of Depth and Geologic Time. *Geophysics*, 16, 192-206.
- Faust, L. Y. (1953). A Velocity Function including Lithologic Variation. *Geophysics*, 18, 271-288.
- FORCE. (2022, February). *Results of the FORCE 2020 lithology machine learning competition*. Retrieved from <https://www.npd.no/en/force/Previous-events/results-of-the-FORCE-2020-lithology-competition/>
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189-1232.
- Gardner, G. H., Garner, L. W., & Gregory, 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 S. H. Danborn, & S. Domenico, *Shear-wave exploration* (pp. 122-136). Society of Exploration Geophysicists.
- Gassmann, F. (1951). Elastic waves through a packing of spheres. *Geophysics*, 16, 271-288.

- Goodway, B. (2001). AVO and Lamé constants for rock parameterization and fluid detection. *CSEG Recorder*, 26(6), 39-60.
- Goodway, W., Chen, T., & Downton, J. (1997). Improved AVO fluid Detection and Lithology Discrimination Using Lamé Petrophysical Parameters. *SEG 67th Annual International Meeting, Denver*.
- Greenberg, M. L., & Castagna, J. P. (1992). Shear-wave velocity estimation in porous rocks: theoretical formulation, preliminary verification and applications. *Geophysical Prospecting*, 40, 195-209.
- Guarido, M. (2019). Machine Learning Strategies to Perform Facies Classification. *Geoconvention 2019*.
- Guarido, M., Emery, D. J., Macquet, M., T. D., & Innanen, K. A. (2020). The Pitfalls and Insights of Log Facies Classification for a Machine Learning Contest. *CREWES Research Report*, 32, 18.
- Hacikoylu, P., Dyorkin, J., & Mavko, G. (2006). Resistivity-velocity transforms revisited. *The Leading Edge*, 1006-1009.
- Hall, B. (2016). Facies classification using machine learning. *The Leading Edge*, 35(10), 906-909.
- Hall, M., & Hall, B. (2017). Distributed collaborative prediction: Results of the machine learning contest. *The Leading Edge*, 36(3), 267-269.
- Hampson, D. P., Russell, B. H., & Bankhead, B. (2005). Simultaneous inversion of pre-stack seismic data. *SEG Technical Program 2005*.
- Hampson, D., Schuelke, J. S., & Quirein, J. A. (2001). Use of multi-attribute transforms to predict log properties from seismic data. *Geophysics*, 66, 220-231.
- Hoffe, B., Perez, M., & Goodway, W. (2008). AVO interpretation in LMR space: A primer. *Back to Exploration*, 31-34.
- Iwuoha, S. C., Pedersen, P. K., Clarkson, C. R., & Gates, I. D. (2019). A working method for estimating dynamic shear velocity in the Montney formation. *Methods X*, 6, 1876-1993.
- Lee, M. W. (2006). A simple method of predicting S-wave velocity. *Geophysics*, 71(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 2021*.
- Mavko, G., Mukerji, T., & Dvorkin, J. (1998). *The Rock Physics Handbook: Tools for seismic analysis in porous media*. Cambridge University Press.
- Mitchum, R. M., Vail, P. R., & Thompson, S. (1977). The depositional sequence as a basic unit for stratigraphic analysis. In C. E. Payton, *Seismic Stratigraphy - Applications to hydrocarbon exploration* (Vol. 26, pp. 53-62). AAPG Memoir.
- Naeine, E. Z., & Prindle, K. (2018). Machine learning and learning from machines. *The Leading Edge*, 886-893.
- Nur, A., Mavko, G., Dvorkin, J., & Gal, D. (1995). Critical porosity: the key to relating physical properties to porosity in rocks. *65th Annual International Meeting* (p. 878). Society of Exploration Geophysics.
- Ostrander, W. J. (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., & Hampson, D. (1991). Comparison of post-stack seismic inversion methods. *SEG Technical Program Expanded Abstracts 1991*, 876-878.
- Russell, B., & Lines, L. (2013). A Gassmann-consistent rock physics template. *CSEG Recorder*.
- Russell, B., Hampson, D., Schuelke, J., & Quirein, J. (1997). Multiattribute seismic analysis. *The Leading Edge*, 16, 1439-1443.
- Shuey, R. T. (1985). A simplification of the Zoeppritz equations. *Geophysics*, 50(4), 609-614.
- Tanner, M. T., & Sheriff, R. E. (1977). Application of amplitude frequency and other attributes to stratigraphic and hydrocarbon determination. In C. E. Payton, *Seismic stratigraphy - applications to hydrocarbon exploration* (Vol. 26, pp. 301-327). Tulsa: AAPG Memoir.
- Tatham, R. H. (1982). Vp/Vs and Lithology. *Geophysics*, 47, 336-344.
- Tosaya, C., & Nur, A. (1982). Effects of diagenesis and clays on compressional velocities in rocks. *Geophysical Research Letters*, 9(1), 5-8.
- Walls, J. D., Taner, M. T., Guidish, T., Taylor, G., Dumas, D., & Derzhi, N. (1999). North Sea reservoir characterization using rock physics, seismic attributes, and neural networks; a case history. *SEG Technical Program Expanded Abstracts 1999*, 1572-1575.

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.

APPENDIX

Petrophysical rock properties provide for your reference. The two tables are an amalgamation of published values from various sources compiled over the last 15 years. No weight has been applied, and a straight average is displayed where more than one value is available. Values in red are estimates with questionable validity.

The cross-plots after the tables are from the normalized 118 wells provided as part of the Norwegian Petroleum Directorate for the FORCE 2020 competition. The mineralogy is from lithoclasses redistribution to mineralogy, as discussed in Emery (2021).

Petrophysical Log Properties

Information compiled from Schlumberger, 1989, 1990, 2003; Baker, 2002; Castagna, 1985, 1993; Mavko 1998, Crain 2000)

Minerals	Density gm/cm ³	V _{rms} m/sec	V _{sma} m/sec	DT (logs us/ft)	DT (logs us/m)	DTS (logs us/ft)	DTS (logs us/m)	Vp/Vs	Ip	GR	P _e	Resistivity	φ _{critical}	Chemical Formula
Quartz	2.65	5779	3930	52.7	173	78	254	1.47	15.30	-	1.81	300-5x 10 ³	0.40	SiO ₂
Calcite	2.71	6500	3372	47	154	90	297	1.93	17.62	-	5.08	5000- 10 ⁷	0.70	CaCO ₃
Dolomite	2.87	7080	3909	43	141	78	256	1.81	20.28	-	3.14	500- 10 ⁵		CaCO ₃ MgCO ₃
Anhydrite	2.97	5640	3110	54	177	98	322	1.81	16.76	-	5.07	1000- 10 ⁴		CaSO ₄
Halite	2.10	4620	2710	66	216	118	387	1.70	9.70	-	4.67	10 ² - 10 ⁷		NaCl
Gypsum	2.33	5880	2770	52	170	110	361	2.12	13.69	-	3.80	10-1.2x 10 ²		CaSO ₄ (H ₂ O) ₂
Feldspars - Alkali	2.55	4420	2638	69	226	116	379	1.68	11.28	~220	2.9	10 ³		KAlSi ₃ O ₈
Feldspars - Average	2.62	4690	2390	65	213	128	418	1.96	12.29					
F - Plagioclase (Albite)	2.60	6740	4110	45	148	85	279	1.64	17.51	~200	1.7	10 ³		NaAlSi ₃ O ₈ (Albite)
F - Plagioclase (Anorthite)	2.76	7690	4925	40	130	62	203	1.56	21.19	~200	3.1	10 ³		CaAl ₂ Si ₂ O ₈ (Anorthite)
Kaolinite	2.30	3020	1410	101	331	216	709	2.14	6.93	80-130	1.6	10 ¹	0.60	Al ₂ Si ₄ O ₁₀ (OH) ₂
Illite	2.67	4320	2540	71	231	120	394	1.70	11.52	250-300	3.0	10 ¹	0.60	K _{1-1.5} Al ₄ (Si _{7-6.5} Al _{1-1.5})O ₂₀ (OH) ₄
Montmorillonite/Smectite	2.32	4330	2310	70	231	132	433	1.87	10.05	150-200	1.9	10 ¹	0.60	(Ca,Na) ₇ (Al,Mg,Fe) ₄ (Si,Al) ₈ O ₂₀ (OH) ₄ (H ₂ C
Chlorite	2.83	3400	1600	90	294	191	625	2.13	9.61	180-250	5.8	10 ¹	0.60	(Mg,Fe,Al) ₆ (Si,Al) ₄ O ₁₀ (OH) ₈
Anthracite Coal	1.51	2900	1357	105	345	225	737	2.14	4.39		0.2	10- 10 ³		CH ₃₅₈ N ₀₀₉ O ₂₂₂
Bituminous Coal	1.30	2540	1184	120	394	257	845	2.15	3.29		0.2			CH ₇₉₃ N ₀₁₅ O ₀₇₈
Lignite	1.16	2038	942	150	491	324	1062	2.16	2.36		0.2	9-2x 10 ²		CH ₃₄₈ N ₀₁₅ O ₂₁₁
Basalt	2.75	5950	2600	51	168	117	385	2.29	16.39	60.0	4.0	300- 10 ³		
Chalk	2.10	2600	1300	117	385	234	769	2.00	5.46	-	5.1	50- 10 ²		CaCO ₃
Granite	2.56	6000	3300	51	167	92	303	1.82	15.36	220.0	3.0	1000- 10 ⁶		SiO ₂ (60-70%) a little of Al ₂ O ₃ , CaO, MgO z
Mixed Rock Lithology														
Clay (Average)	2.64	3400	1600	90	294	191	625	2.13	8.97	~150	3.4	20- 10 ²		
Shale	1.8-2.7	2616		63-170							3.5	50- 10 ³	0.60	
Feldspars - Average	2.62	4690	2390	65	213	128	418	1.96	12.29	200	2.5	10 ²		
Unconsolidated Sandstone	2.65	5180	3730	58.8	193			13.73				10 ²		
Clean Sandstone (p10%)	2.57	5750	3730	53.0	174	82	268	1.54	14.78	15-30	1.7	50- 10 ²	0.40	
Dirty Sandstone (p10%)	2.55	5950	3870	51.2	168	79	258	1.54	15.15		2.7	10 ³		
Conglomerate	2.57	5950	3870	51	168	79	258	1.5	15.2	25.0	2.7	10 ⁴	0.40	
Limestone	2.63	6420	3400	47.5	156	90	294	1.89	16.85	10-20	5.0	5000- 10 ⁵	0.70	
Dolomite (P>5.5%)	2.77	7000	4000	43.5	143	76	250	1.75	19.36	8-15	3.14	500- 10 ³		
Dolomite (P<5.5%)	2.85	7000	4000	43.5	143	76	250	1.75	19.95		3.05	1000- 10 ⁴		
Fluids														
Water	0.998	1450	0	210	690			1.45			0.36	10 ¹		H ₂ O
Sea Water (100,000 ppm)	1.025	1460	0	209	685			1.50			0.73	10 ⁻²		
Sea Water (200,000 ppm)	1.146	1490	0	205	671			1.71			1.12	10 ⁻³		
Brine	1.0686	1470	0	207	680			1.57				10 ⁻²		
Oil	0.85	1450	0	210	690			1.23			0.12	10 ⁴		
Methane	0.0086	630	0	484	1587			0.01			0.09	10 ⁸		

Rock Physic Properties			Bulk modulus K (Gpa)	Poisson Ratio σ^*	Shear modulus μ (Gpa)	Lamb constant μ (Vp,Vs)	Vp from K, μ , & ρ	Vs from μ , & ρ	Vp/Vs From Poisson Ratio	Vp/Vs from Vel.	Bulk modulus K from Vp & Vs	Poisson Ratio σ From Vp & Vs	Lamb constant μ from Vp & Vs	Lamb constant μ from K & σ	
Minerals	VPma m/sec	VSma m/sec	Density kg/m3												
Quartz	5779	3930	2647	37.0	0.08	44.4	6.6	6028	4097	1.48	1.47	33.9	0.07	40.9	43.8
Calcite	6500	3372	2710	71.8	0.31	30.6	52.9	6447	3361	1.92	1.93	73.4	0.32	30.8	30.5
Dolomite	7080	3909	2865	86.5	0.24	48.8	56.1	7272	4126	1.72	1.81	85.3	0.28	43.8	53.6
Anhydrite	5640	3110	2973	57.1	0.28	30.4	37.0	5730	3198	1.80	1.81	56.2	0.28	28.8	30.2
Halite	4620	2710	2100	24.8	0.25	14.9	14.0	4610	2664	1.73	1.70	24.3	0.24	15.4	14.9
Gypsum	5880	2770	2328	40.0			44.8				2.12	56.7	0.36	17.9	
Feldspars - Alkali	4420	2638	2551		0.25		14.3				1.68	26.2	0.22	17.8	
Feldspars - Average	4690	2390	2620	37.5	0.32	15.0	27.7	4685	2393	1.94	1.96	37.7	0.32	15.0	15.3
F - Plagioclase (Albite)	6740	4110	2598	66.0	0.28	26.3	30.2	6237	3182	1.80	1.64	59.5	0.20	43.9	34.9
F - Plagioclase (Anorthite)	7690	4925	2755				29.3			1.41	1.56	73.8		66.8	
Kaolinite	3020	1410	2295	14.8	0.28	5.0	11.8	3059	1481	1.81	2.14	14.8	0.36	4.6	7.6
Illite	4320	2540	2666	32.4	0.24	14.0	15.4	4377	2292	1.71	1.70	26.8	0.24	17.2	20.4
Montmorillonite/Smectite	4330	2310	2320	29.0	0.21	7.9	18.7	4128	1845	1.65	1.87	27.0	0.30	12.4	20.9
Chlorite	3400	1600	2828	22.5	0.36	7.1	18.2	3362	1585	2.14	2.13	23.0	0.36	7.2	6.9
Anthracite Coal	2900	1357	1513	9.4	0.38	4.4	7.2	3177	1705	2.27	2.14	9.0	0.36	2.8	
Bituminous Coal	2540	1184	1295	4.7	0.35	1.8	4.7	2337	1179	2.08	2.15	5.9	0.36	1.8	
Lignite	2038	942	1160	2.5	0.38	0.9	2.8	1786	881	2.27	2.16	3.4	0.36	1.0	
Basalt	5950	2600	2754	58.5	0.20	4.0	60.3	4814	1205	1.63	2.29	72.7	0.38	18.6	43.9
Chalk	2600	1300	2100	9.0	0.35	3.2	7.1	2513	1234	2.08	2.00	9.5	0.33	3.5	3.0
Granite	6000	3300	2560	91.0	0.31	28.0		7080	3307	1.91	1.82	55.0	0.28	27.9	39.6
Mixed Rock Lithology															
Clay (Average)	3400	1600	2639	21.6	0.35	6.9	17.0	3418	1620	2.08	2.13	21.5	0.36	6.8	7.2
Shale															
Unconsolidated Sandstone	5180	3730	2650	38.5	0.14	36.9	-2.6	5753	3732	1.55	1.39	21.9	-0.04	36.9	
Clean Sandstone (p10%)	5750	3730	2570	38.5	0.14	36.9	13.5	5842	3789	1.55	1.54	37.3	0.14	35.8	
Dirty Sandstone (p10%)	5950	3870	2547	40.9	0.13	39.7	13.9	6070	3948	1.53	1.54	39.3	0.13	38.1	
Limestone	6420	3400	2625	70.0	0.31	31.3	47.5	6524	3453	1.91	1.89	67.7	0.31	30.3	
Dolomite (P>5.5%)	7000	4000	2765	79.0	0.26	45.6	47.0	7111	4061	1.76	1.75	76.5	0.26	44.2	
Dolomite (P<5.5%)	7000	4000	2850	79.0	0.26	45.6	48.5	7004	4000	1.76	1.75	78.9	0.26	45.6	
Fluids															
Water	1450		998	2.3195	0.5	0.0		1525				2.098	0.5		
Sea Water (100,000 ppm)	1460		1025	2.401	0.5	0.0		1531				2.185	0.5		
Sea Water (200,000 ppm)	1490		1146		0.5							2.544	0.5		
Brine			1069	2.752	0.5	0.0		1605							
Oil	1450		850	0.862	0.5	0.0		1007				1.787	0.5		
Methane	630		9	0.001325	0.5	0.0		393				0.003	0.5		

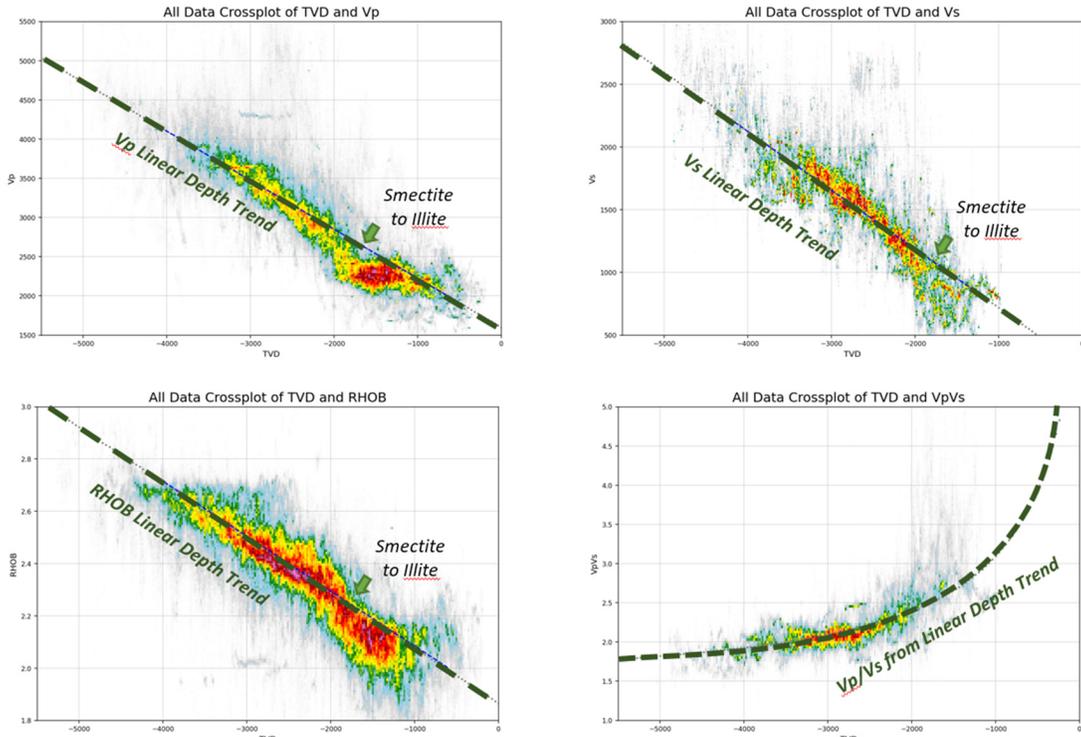


FIG. A. Input data plotted against depth, strong discontinuity at -1800 m is assumed to be the location of smectite clay conversion into illite.

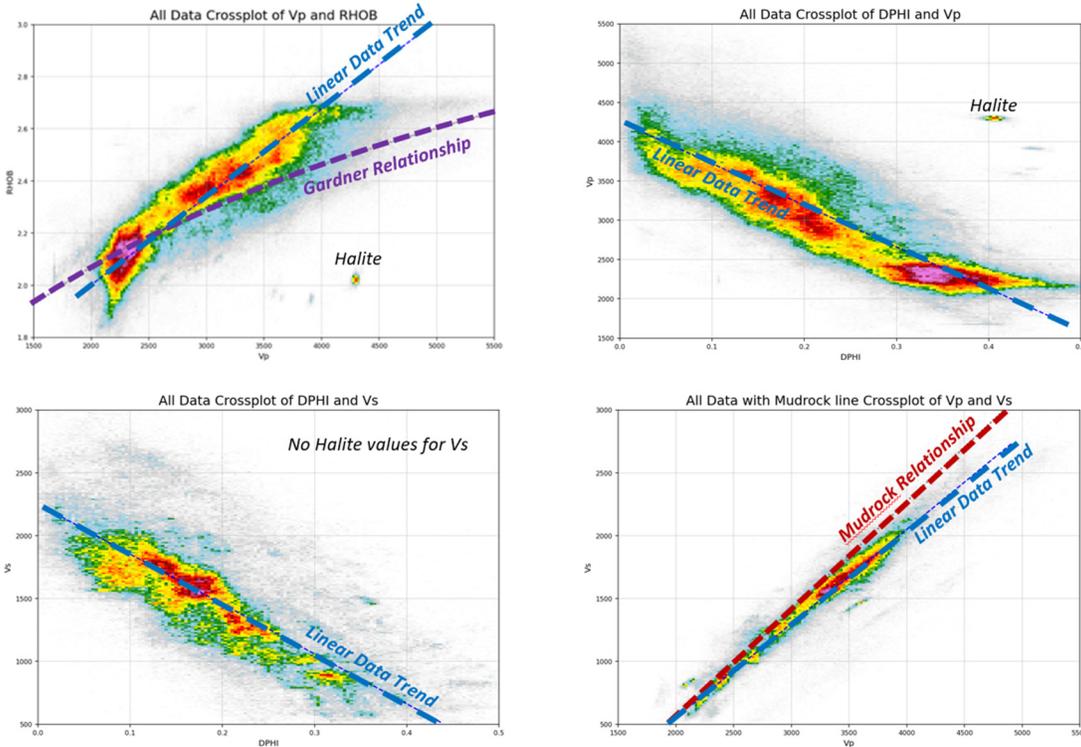


FIG. B. Input data V_p , V_s , RHOB, and V_p/V_s plotted against each other. DPHI is the porosity estimated using limestone density (2.72).

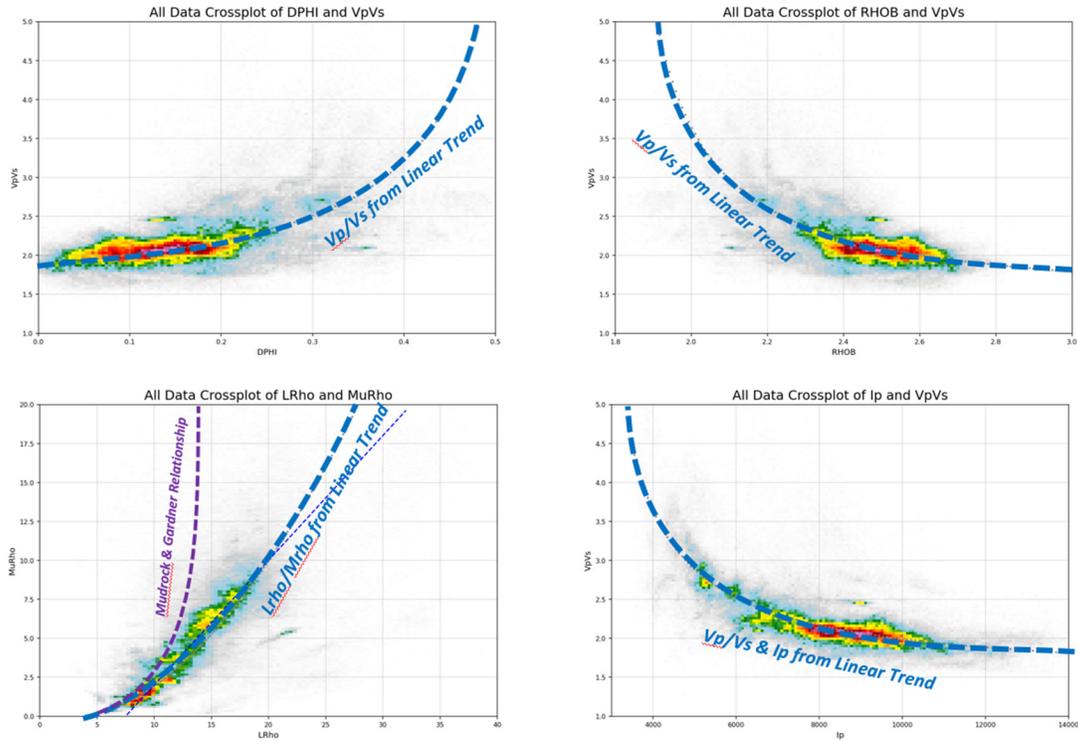
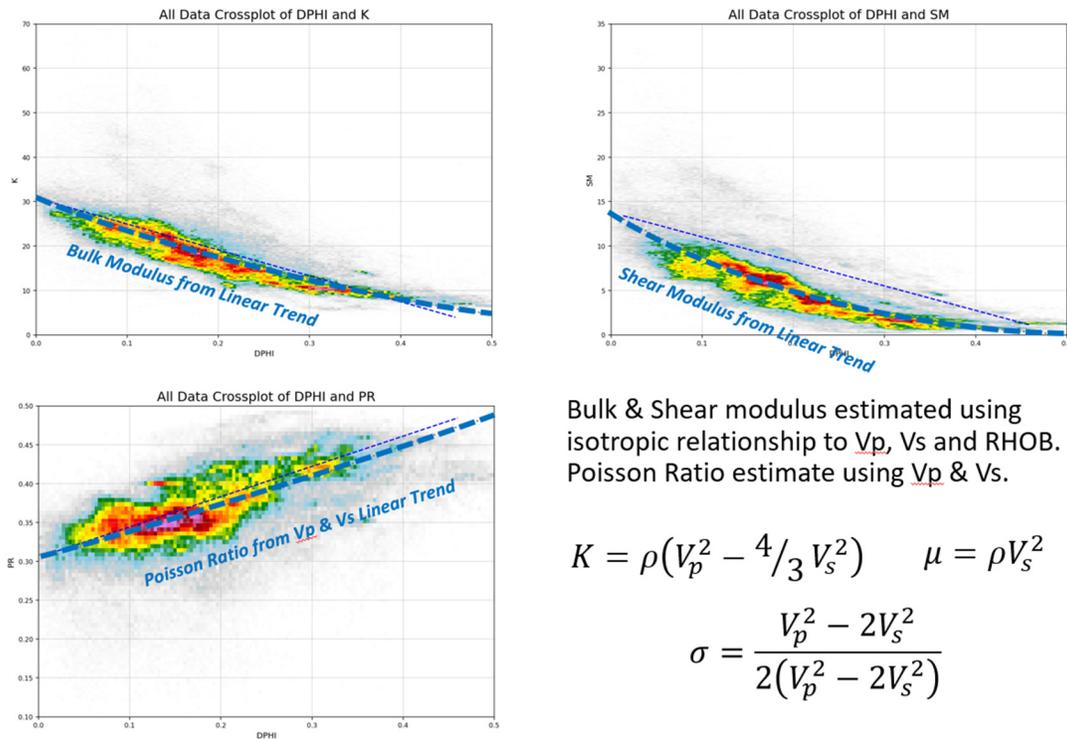


FIG. C. Vp/Vs and Lambda-Rho/Mu-Rho plot using all the input data



Bulk & Shear modulus estimated using isotropic relationship to Vp, Vs and RHOB. Poisson Ratio estimate using Vp & Vs.

$$K = \rho(V_p^2 - \frac{4}{3} V_s^2) \quad \mu = \rho V_s^2$$

$$\sigma = \frac{V_p^2 - 2V_s^2}{2(V_p^2 - 2V_s^2)}$$

FIG. D. Rock-Physics Parameters of all input data (estimated from Vp, Vs and RHOB) plotted against DPHI estimate using Limestone density (2.72 gm/cc³).

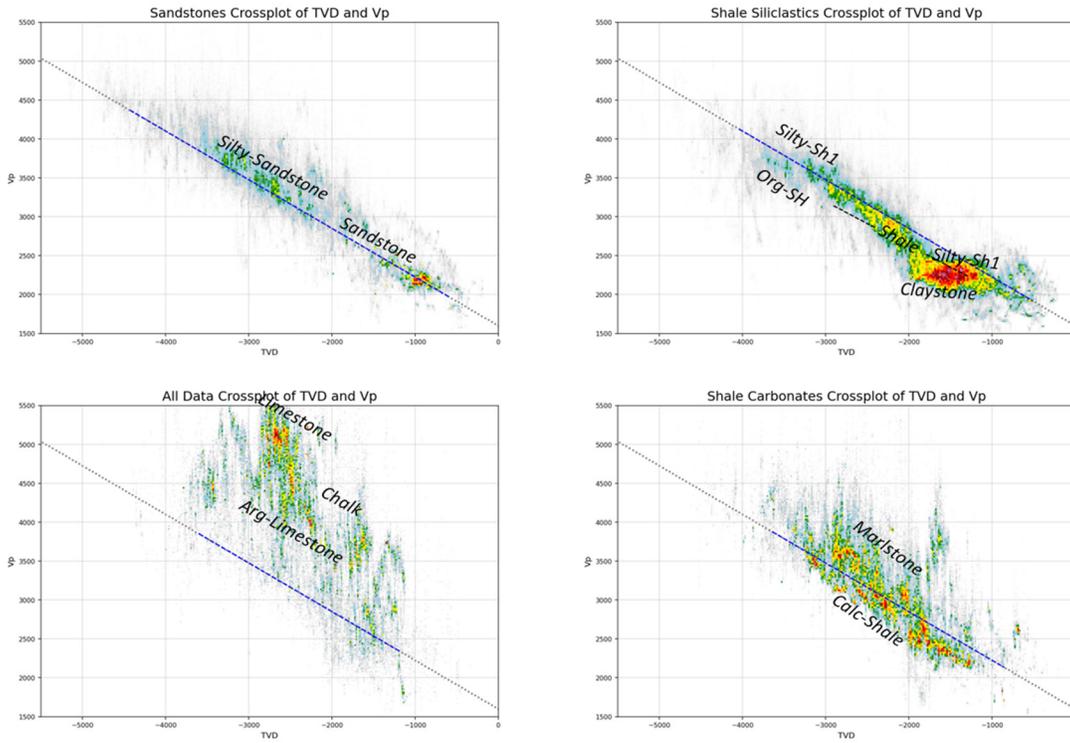


FIG. E. Vp versus Depth for sandstone (upper-left), siliciclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The line is from the 100% data linear fit.

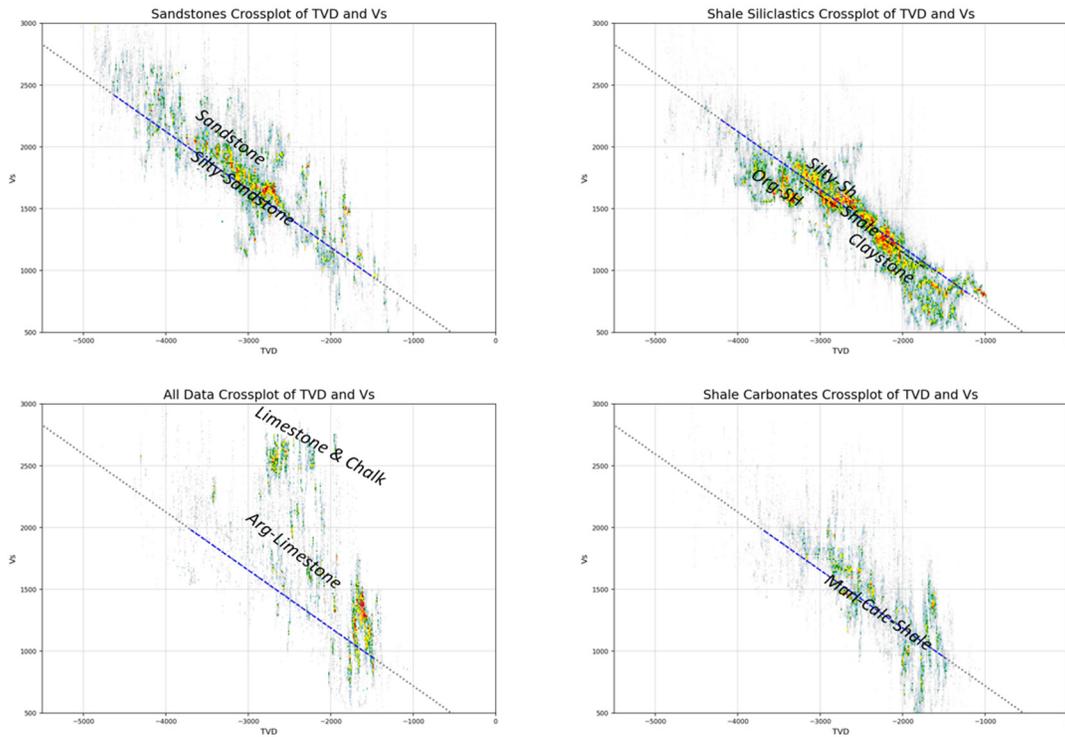


FIG. F. Vs versus Depth for sandstone (upper-left), siliciclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The line is from the 100% data linear fit.

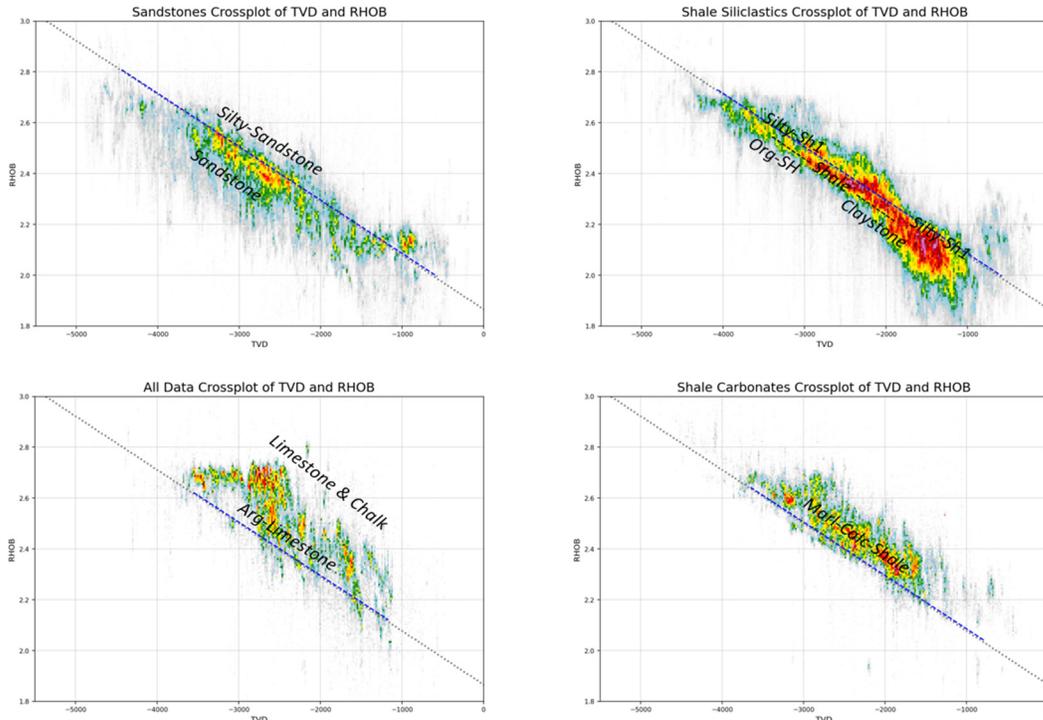


FIG. G. RHOB versus Depth for sandstone (upper-left), siliclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The line is from the 100% data linear fit.

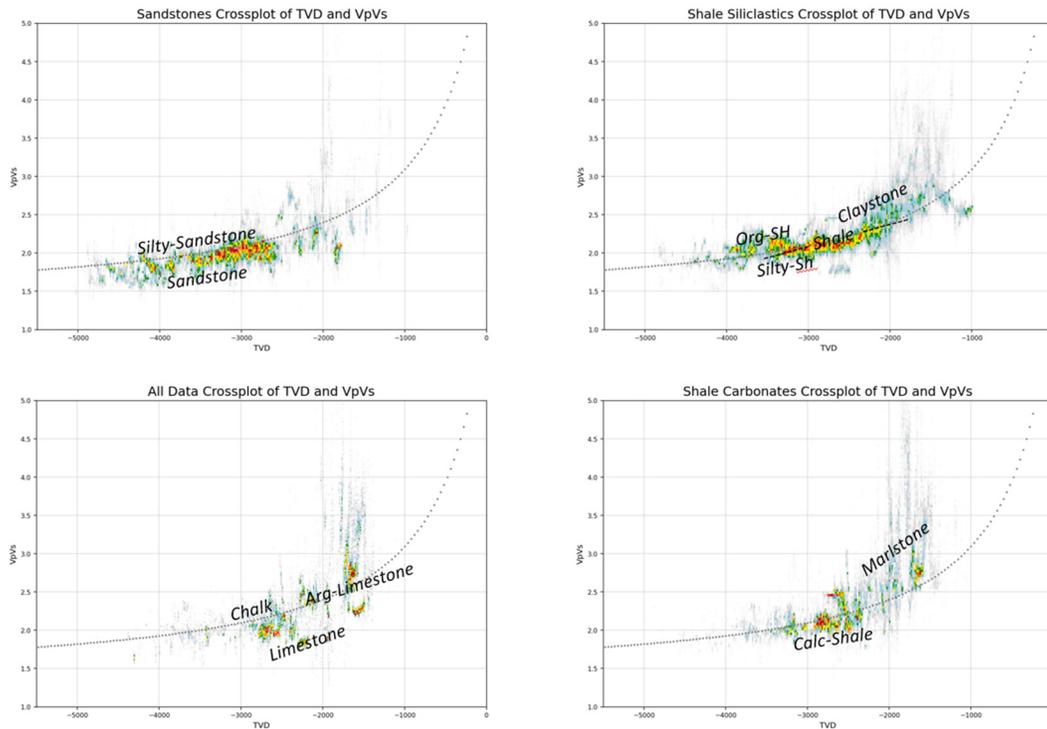


FIG. H. Vp/Vs versus Depth for sandstone (upper-left), siliclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The line is from the 100% data linear fit.

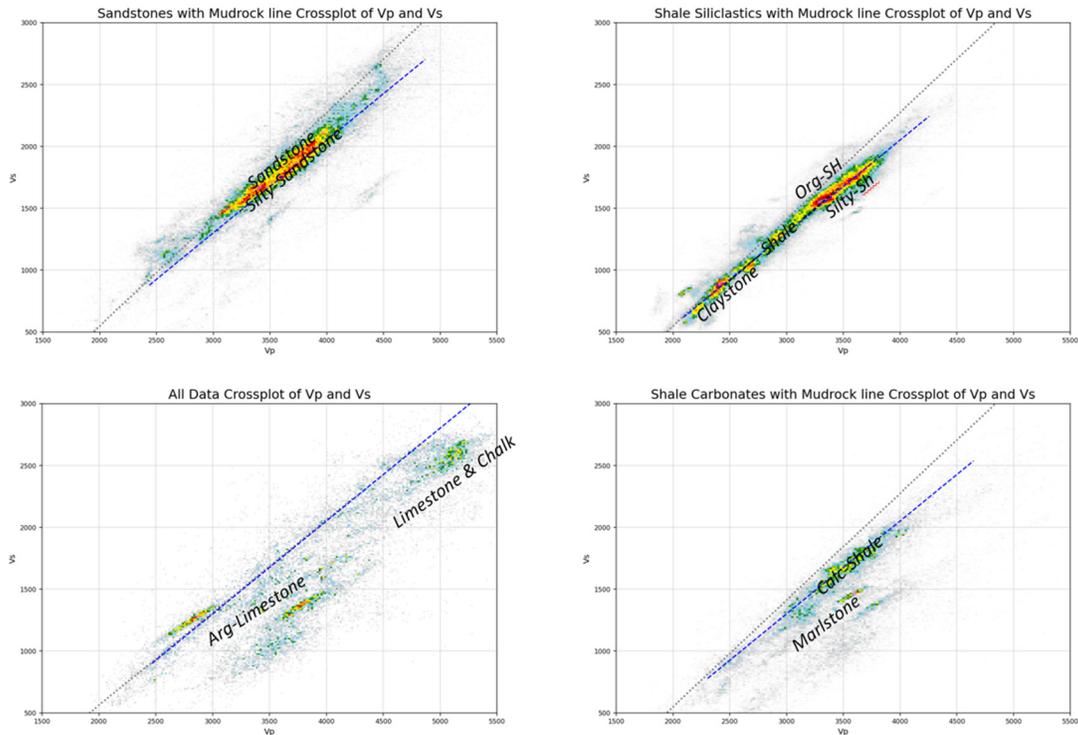


FIG. I. Vs versus Vp for sandstone (upper-left), siliciclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The Grey curved line is from the Mudrock relationship, while the blue line is the fit to the 100% data.

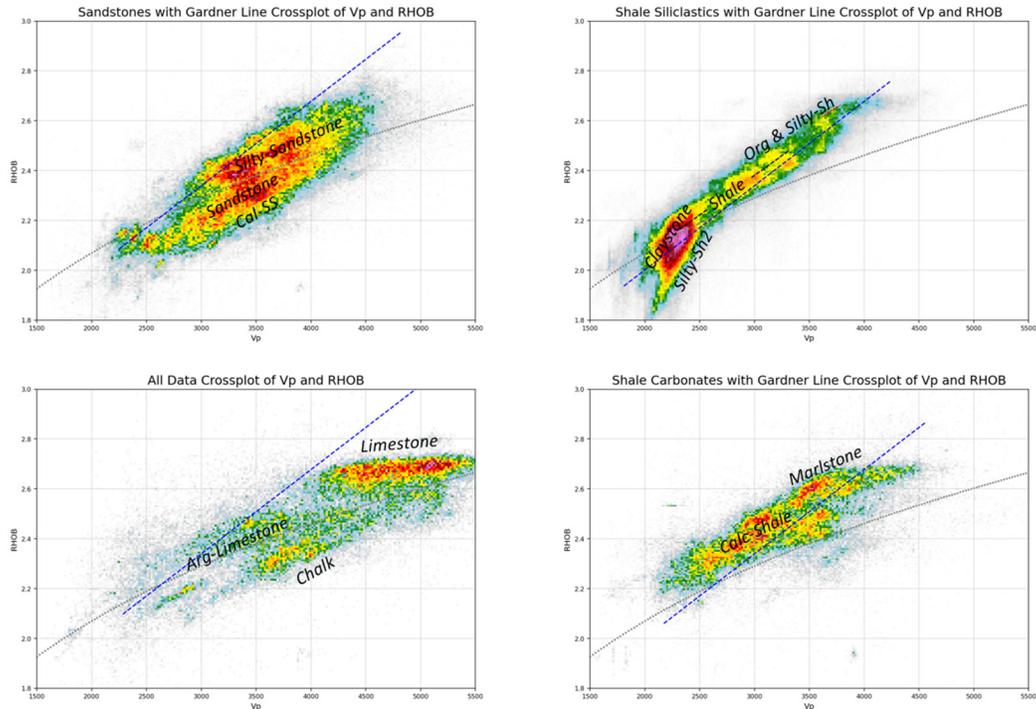


FIG. J. Vp versus RHOB for sandstone (upper-left), siliciclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The grey curved line is from the Gardner relationship, while the blue line is the fit to the 100% data.

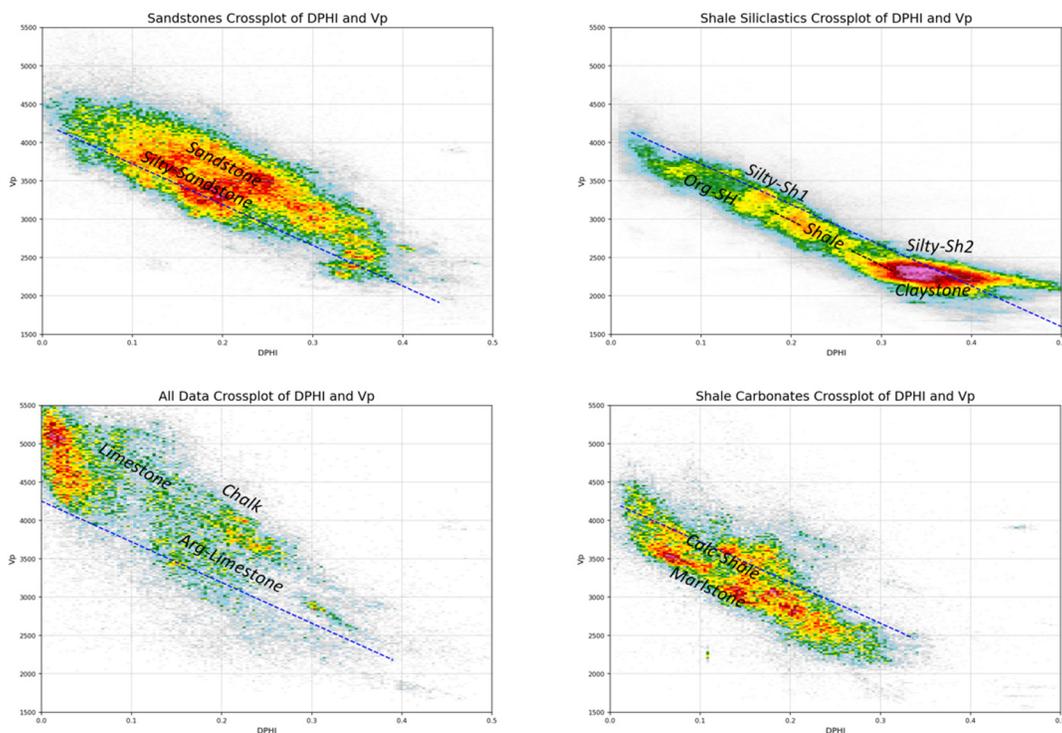


FIG. K. V_p versus DPHI for sandstone (upper-left), siliciclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The blue line fits 100% data.

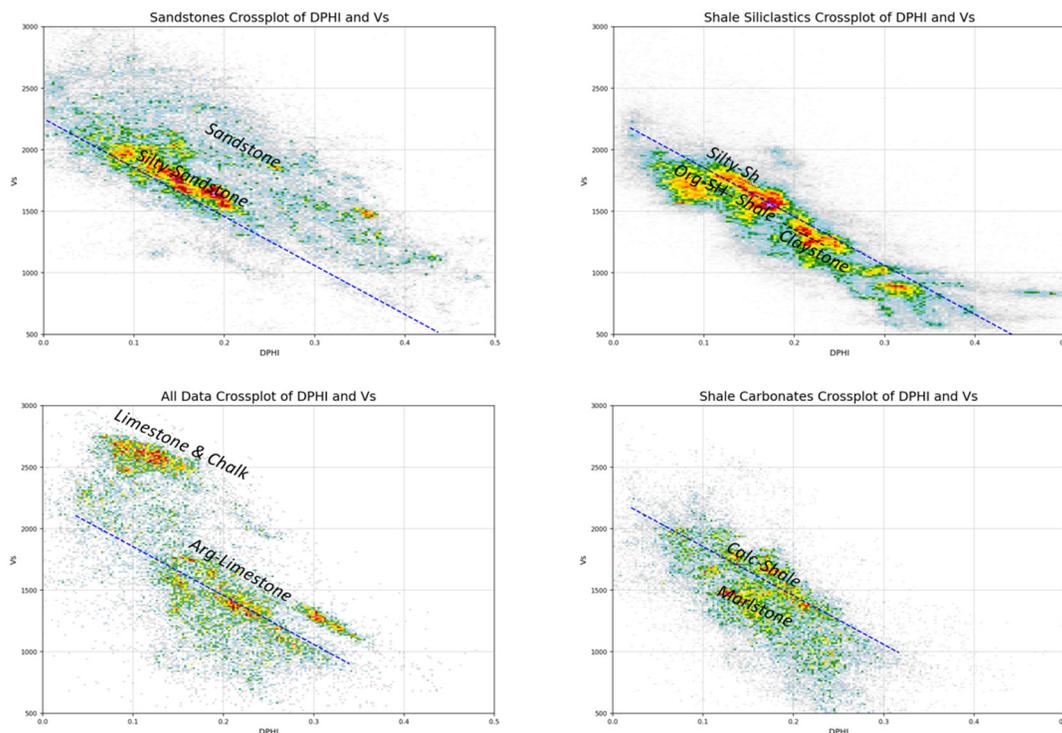


FIG. L. V_s versus DPHI for sandstone (upper-left), siliciclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The blue line fits 100% data.

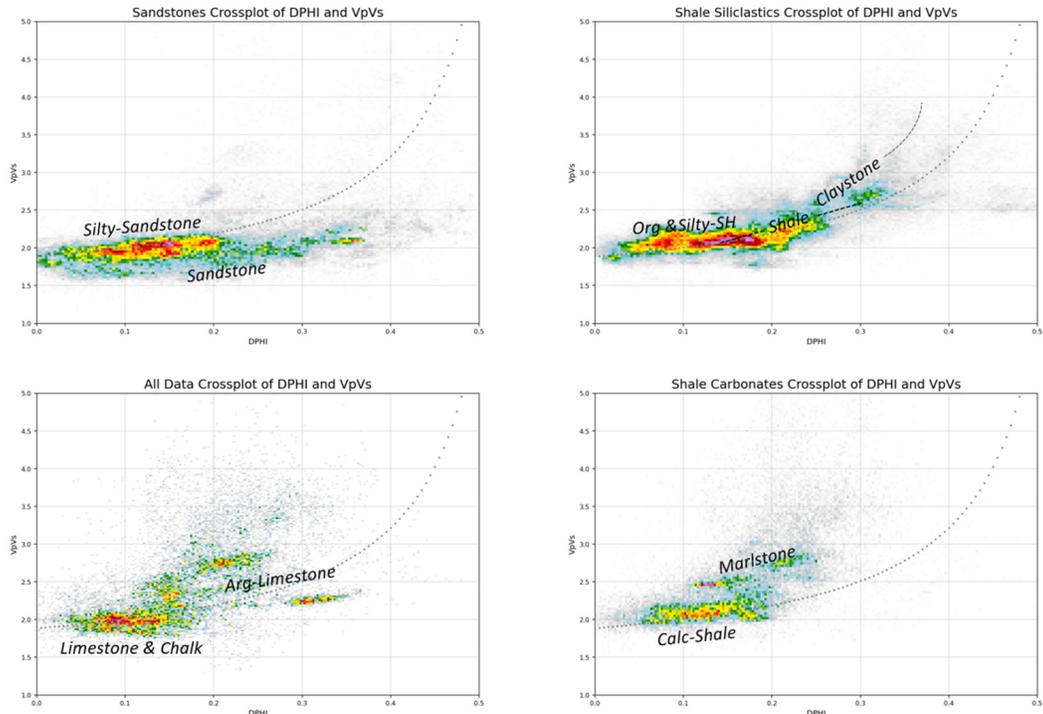


FIG. M. Vp/Vs versus DPHI for sandstone (upper-left), siliclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. Line fit to the 100% data.

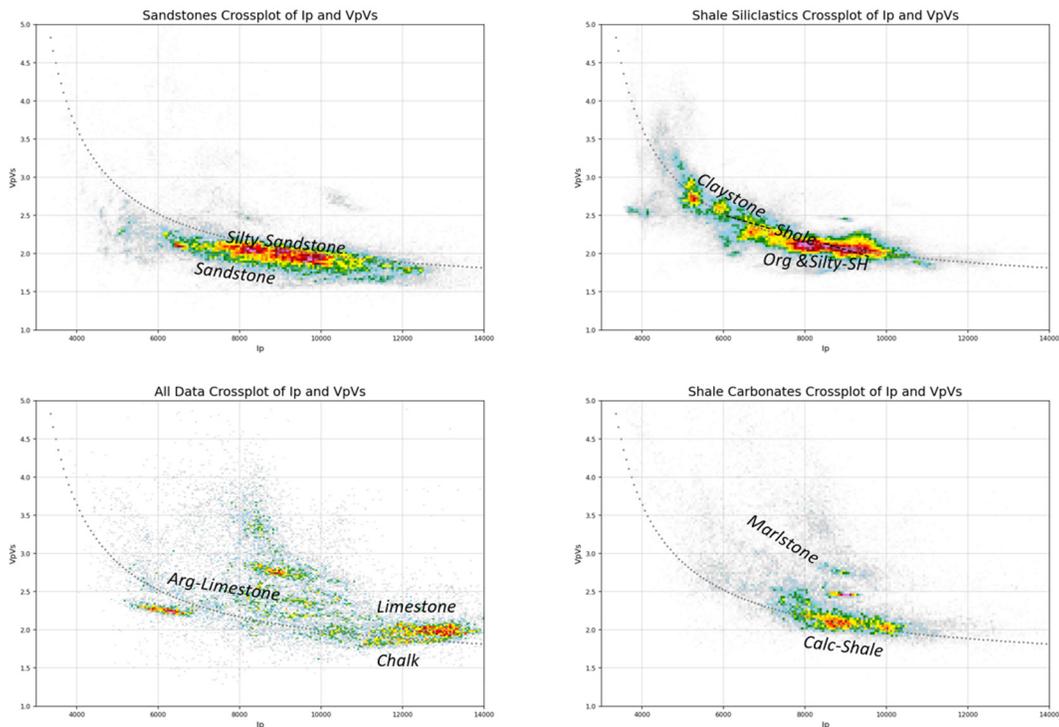


FIG. N. Vp/Vs versus Ip for sandstone (upper-left), siliclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. Line fit to the 100% data.

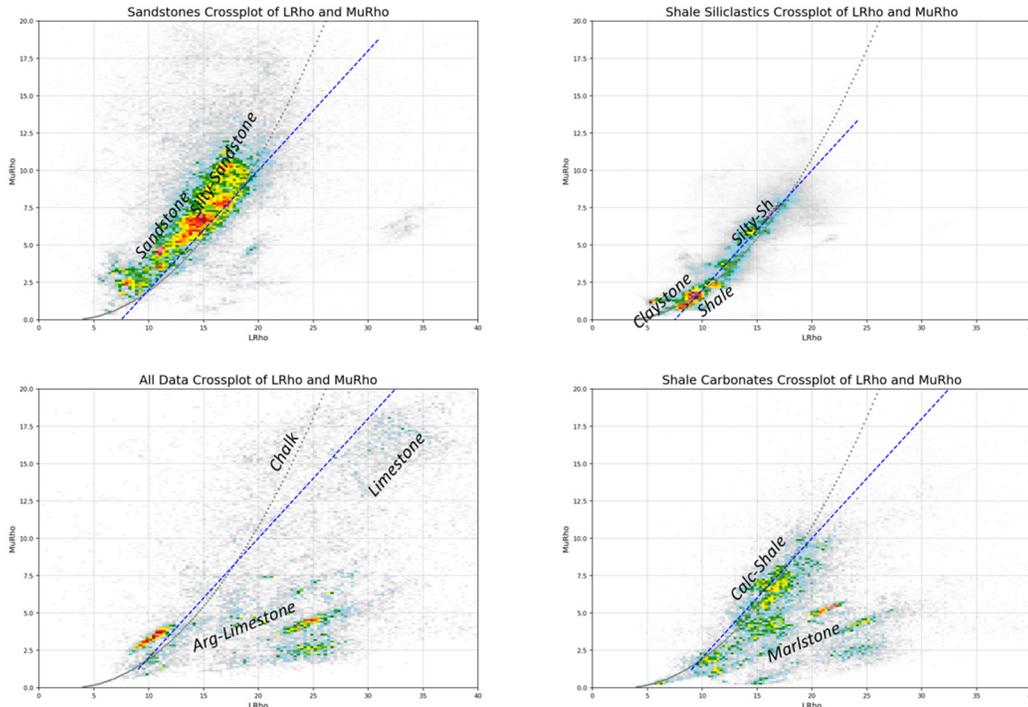


FIG. O. Lambda-Rho vs Mu-Rho for sandstone (upper-left), siliciclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. The Grey line from linear Vp, Vs & RHOB fits the 100% data and the blue line linear fit the 100% data.

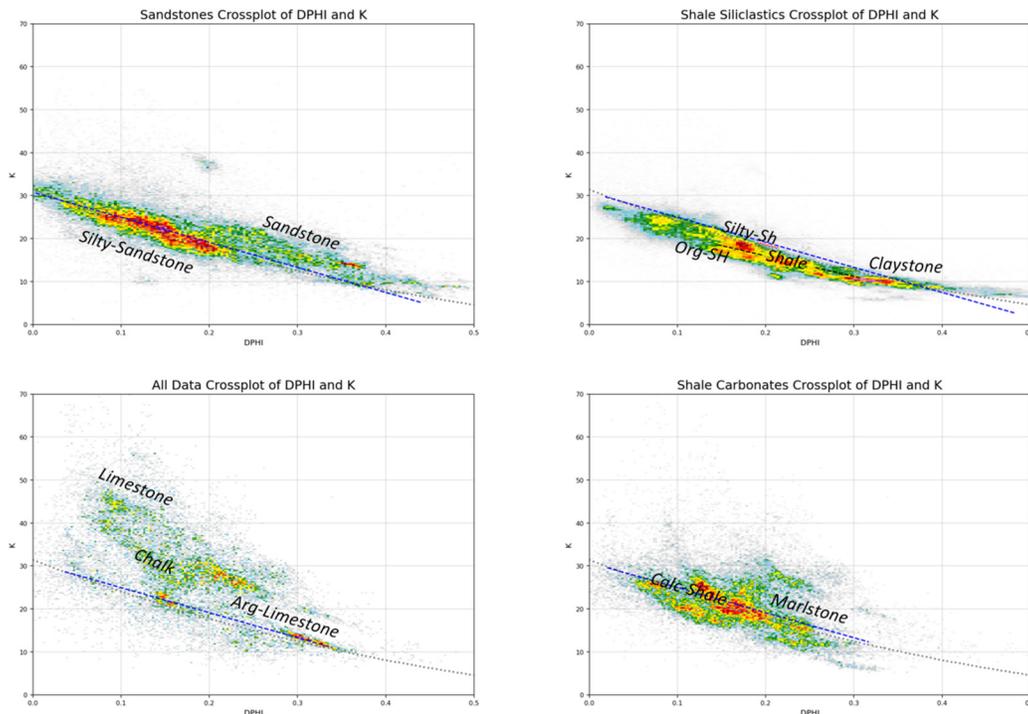


FIG. P. Bulk Modulus (K) vs DPHI (limestone) for sandstone (upper-left), siliciclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. Line linear fit to the 100% data.

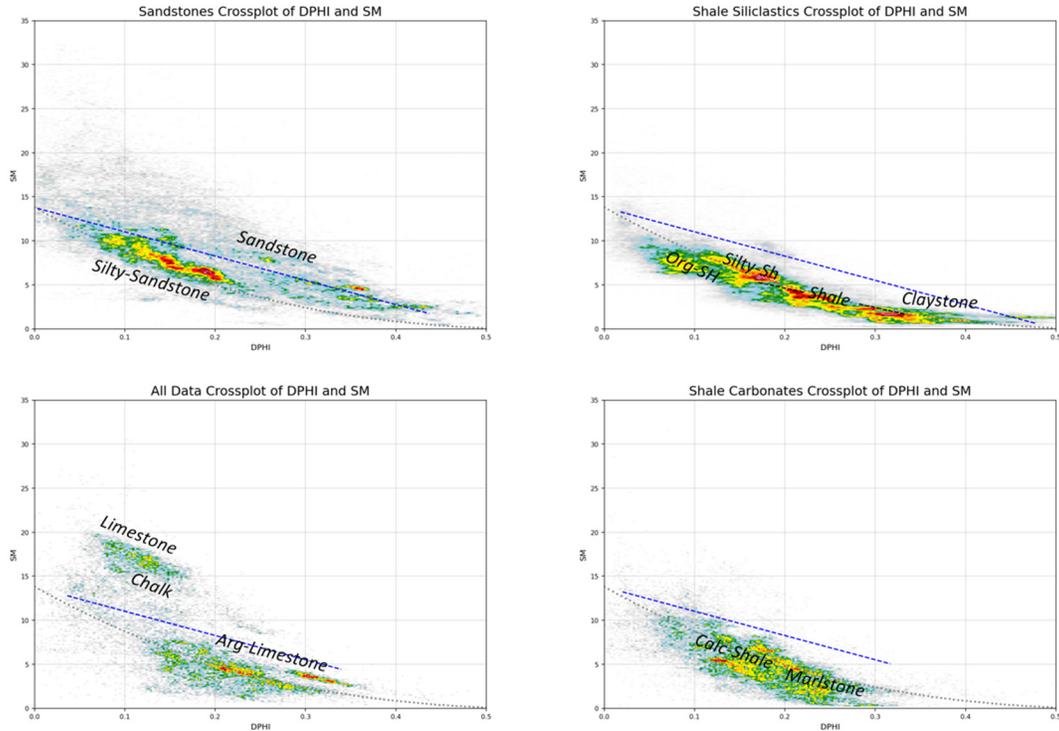


FIG. Q. Shear Modulus (μ) vs DPHI (limestone) for sandstone (upper-left), siliclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. Line linear fit to the 100% data.

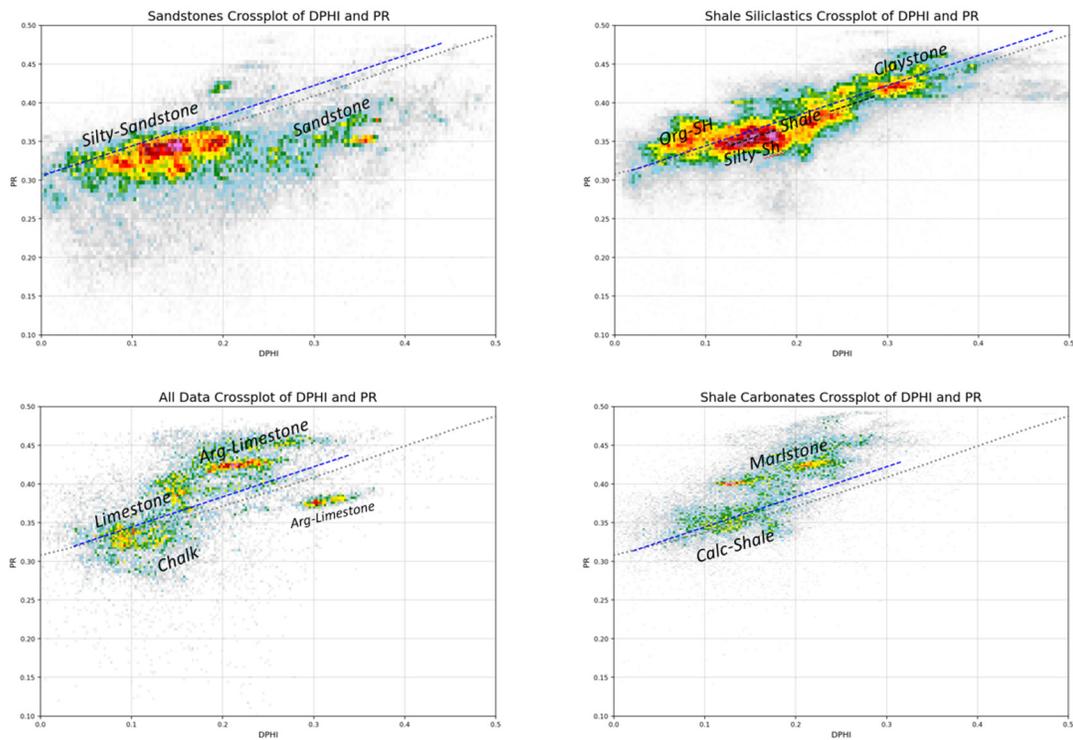


FIG. R. Poisson Ratio (ν) vs DPHI (limestone) for sandstone (upper-left), siliclastic shales (upper-right), calc-shales (lower-right) and limestone (lower-left) mineralogy. Line linear fit to the 100% data.