SPWLA 64th Annual Logging Symposium, June 10-14, 2023
DOI: 10.30632/SPWLA-2023-0084

BEST PRACTICES IN AUTOMATIC PERMEABILITY ESTIMATION: MACHINE-LEARNING METHODS VS. CONVENTIONAL PETROPHYSICAL MODELS

Oriyomi Raheem, Wen Pan, Carlos Torres-Verdín, and Misael M. Morales, The University of Texas at Austin

This paper was prepared for presentation at the SPWLA 64th Annual Logging Symposium held in Lake Conroe, TX, USA, June 10-14, 2023.

ABSTRACT

Estimation of flow-related petrophysical properties from borehole geophysical measurements is challenging in the presence of spatially complex rocks. This paper develops best practices for permeability estimation by comparing results obtained with machine learning methods and conventional petrophysical models.

We preprocess core data acquired in key wells that incorporate expert knowledge, depth-matched core porosity with log-calculated porosity, generated two support systems (core and log resolution), and trained with the predicted porosity and smoothed permeability.

Feature engineering is implemented on the resulting data suite to select the well logs that best correlate with permeability. Dimensionality reduction techniques are used to generate latent-space well logs, from which models are trained to estimate permeability. From the latent space models, we perform regression using random forests, k-nearest neighbors, artificial neural network, and Timur-Coates's model to estimate the logarithm of permeability. Finally, the uncertainty of the estimated permeability is calculated based on the validation variance function for the test set. Results are compared based on the relative standard error of permeability estimations. The methods are tested on data sets from both conventional and unconventional rocks.

Results indicate that random forest and neural networks best estimate permeability from triple-combo well logs across a wide range of variation with an average of 16% relative standard error when using the original well logs. Estimations improve when using latent-space well logs with discrete wavelet transforms. Machine learning algorithms reduced the error to less than 13%, while implementing a fully-connected autoencoder resulted in less than 10% error using the core-resolved support system. We obtained a 5% average permeability estimations error in well-log resolution, a 50% further decrease compared to the core support. The Timur-Coates model/approach is the most reliable for data sets with *a priori* information about irreducible water saturation, yielding less than 22% relative standard error, yet it requires prior data classification to improve estimation accuracy.

The new procedure is computationally efficient, with estimations obtained in less than 2 minutes of CPU time. Uncertainty estimates show that permeability calculations are accurate, as their distributions border the true values within ± 5 mD. However, it is important to note that training wells must cover the widest possible range of measurements and petrophysical properties, and core data are smoothed to well-log resolution to improve the estimation of permeability in test wells. Data normalization does not always improve machine-learning estimations, especially across very low or high permeability ranges, resulting in a 25% increase in permeability estimation error compared to non-normalized data.

INTRODUCTION

Permeability prediction is an active research area in the oil and gas industry. Previous efforts on machine learning techniques have been published to solve the challenge of uncertainty in permeability prediction (Negara et al., 2016; Belozerov et al., 2018; Bennis et al., 2019; Eriavbe et al., 2019; Arigbe, 2020; Singh et al. 2020; Bennis et al., 2023). The complexity of using different approaches to solving the interpretation problems is associated with the need to obtain an accurate and balanced assessment of reservoir properties (Belozerov et al., 2018; Mohamed et al., 2023). Anand et al., (2010) and Guimarães, C., (2021) incorporated well logs and well-testing data in permeability prediction. Interpretation challenges complicate machine-learning algorithms by relying on the accurate assessment of petrophysical properties and a minimum well-log suite for formation description (Negara et al., 2016). Furthermore, drilling mud types, device quality and condition, salinity changes, and field noise in recordings tend to require expert knowledge to interpret each well individually. There is a need for an expedient

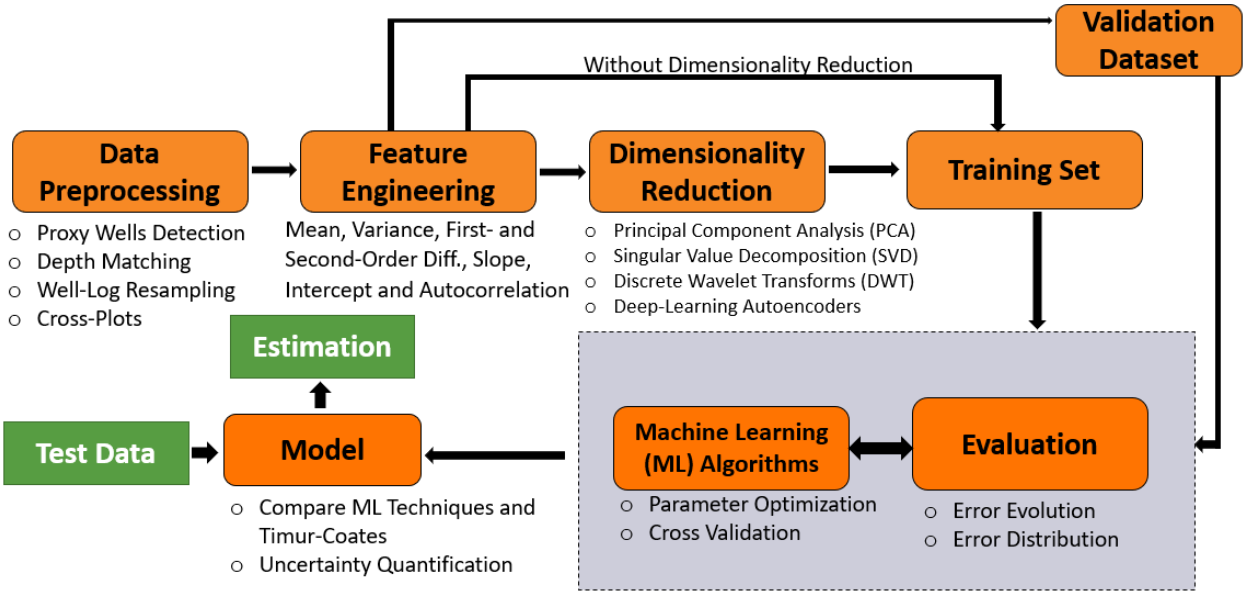

Fig. 1—Workflow for automatic estimation of permeability from core porosity and well logs. The workflow starts with data preprocessing, feature engineering, and dimensionality reduction. Furthermore, multiple machine-learning models are compared to Timur-Coates's model and the uncertainty of permeability estimations is quantified as part of the process.

interpretation method of well logs that (a) saves time and can be applied to a field case with many wells, (b) implements a quantitative formula between rock permeability and conventional well logs, and (c) is accompanied with uncertainty quantification.

This paper develops a fast estimation workflow (**Fig. 1**) that can be applied to many wells. Interpretation is on a case-by-case basis. Field cases considered for testing and benchmarking are described below.

Field A is a downthrown from Pemberton to the east and upthrown from Sculptor block to the north. The gas field covers an area of approximately 15 square kilometers at the gas-down-to (GDT) of 3038 meters true vertical depth. Field B is a vast gas field covering Kansas, Oklahoma, and Texas. It includes rock permeabilities as low as less than 0.5 md in the Council Grove. Field C is a tight-gas sandstone providing 72% (342 Tcf) of the projected unconventional resource (474 Tcf) for the United States. It covers Wyoming, Colorado, and Utah. Field D is located at the center of the North Sea. It mainly produces oil from Middle Jurassic age sandstones, covering 2703-3200 meters depth with uncertain cross-communication across heavy faults at the western part. We directly map well logs to interpreted petrophysical properties and train mapping using the properties in key wells that incorporate expert knowledge and core data. We also implement machine learning algorithms and Timur Coates's model to estimate the logarithm of permeability from well logs. Relative standard errors in the prediction are compared to core permeability for proper assessment.

METHOD

Data Preprocessing

We develop the first-pass input vs. output correlations using the r-squared and Seaborn methods (Géron, 2019) as part of the data-preparation stage including proxy-wells-detection, core-log depth matching, well-log resampling, and cross plots. Two support systems are constructed as part of this method: (a) core resolution, i.e., interpolated triple-combo well logs to core depth, and (b) well-log resolution-calculated core bulk density from core grain density using **Eq. 1** given by

$$\rho_b = \phi_t(\rho_f - \rho_m) + \rho_m, \quad (1)$$

where ρ_b is bulk density, ϕ_t is total porosity, ρ_f is fluid density and ρ_m is matrix density and smoothed to well-log resolution to match the density log, predicted porosity from the resulting log, apply the same smoothing function to permeability, and trained with the predicted porosity and smoothed permeability.

Feature Engineering

Feature engineering was used for data augmentation before feeding the training dataset into the models (**Fig. 2**).

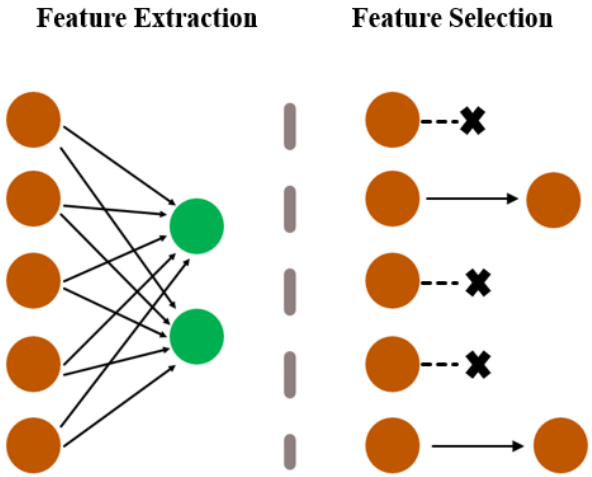

Fig. 2—Graphical illustration of feature extraction and selection.

From the well logs provided, we generated more features by calculating their moving-window average of mean, variance, autocorrelation, first- and second-order gradient, slope, and intercept. The best features were selected by aggregating features with mean ranks.

Dimensionality Reduction

Dimensionality reduction techniques (**Fig. 3**) were implemented such as principal component analysis (PCA), singular value decomposition (SVD), discrete wavelet transforms (DWT), and deep learning-based auto-encoders (**Fig. 4**) to generate latent-space well logs, from which models were trained to estimate permeability. The central idea of PCA (**Eq. 2**) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables while retaining as much as possible the variation present in the data set. It can be described as follows:

PCA: Covariance(X) → Eigen-decomposition → Z, assumes the square matrix, X, such that

$$X = Q\lambda Q^t, \tag{2}$$

where Q is an orthogonal matrix, Q^t is the transpose of orthogonal matrix and λ is the diagonal basis of singular values.

PCA uses eigen-decomposition of the normalized covariance matrix which can also be performed by the singular value decomposition (SVD) (**Fig. 11**) of the data matrix S given by

$$\textbf{SVD: } S = U\Sigma V^t, \tag{3}$$

where U is left-singular orthonormal matrix of eigenvectors of a matrix SS^t, Σ is the diagonal matrix of singular values, (the square root of the eigenvalues of SS^t), and V^t is the orthonormal matrix of the right-singular vectors.

The most common approach to the multilevel discrete wavelet transforms (DWT) involves further decomposition of only the approximation sub-band at each subsequent level, providing a multi-resolution decomposition of the data matrix. In two dimensions, the discrete wavelet transform produces a set of coefficients corresponding to the possible combinations of the wavelet decomposition filters over the two separate axes (note that for n dimensions there are 2^n sets of coefficients).

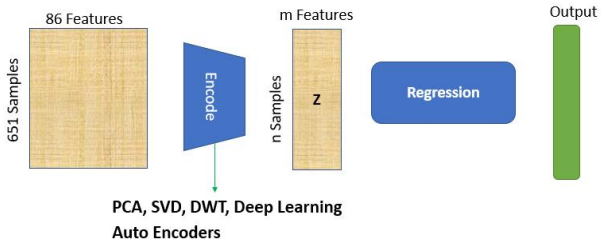

Fig. 3—Dimensionality reduction techniques were implemented such as principal component analysis (PCA), singular value decomposition (SVD), discrete wavelet transforms (DWT), and deep learning-based auto-encoders to generate latent-space well logs from which models were trained to estimate permeability.

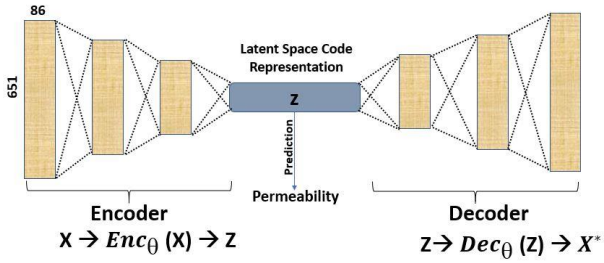

Fig. 4—Deep Learning auto encoders.

DWT: The wavelets are discretely sampled. It captures frequency and location and provides coefficients from a series of multi-resolution high -and low-pass filters.

Autoencoder: reduces the input data (X) to a lower dimensional space (Z) and then reconstructs the full data X^* (**Fig. 4**), through a set of learned weights and biases. It can be fully connected or convolutional.

Machine Learning (ML) Algorithms

From the latent space models, we performed regression using random forest (RF), kernelized support vector regression (SVR), ridge, lasso, k-NN, artificial neural network (ANN), and Timur-Coates's methods to estimate the logarithm of permeability from core porosity and well logs (gamma ray, bulk density, neutron porosity, and photoelectric factor).

Support Vector Regression (SVR): It is a supervised machine-learning method for segmentation. It forms optimal separated hyperplanes when the classes overlap through the concept of margin and support vectors (training data in the margin that impact model fit). In this way, training data well within the decision boundary has no influence on the results. SVR achieves nonlinearity by basis expansion feature space (typically to a high-dimensional space) and performs optimization by maximizing margins and minimizing the regularized error jointly.

Random Forest (RF): It is a decision tree-based algorithm that performs optimization within an ensemble. It improves the accuracy of the trees by randomizing selection such that for each split, only a subset of the available predictors are candidates for splits. It prevents a single strong predictor from dominating the entire set of trees by forcing diversity among the trees, thereby decorrelating them.

Artificial Neural Network (ANN): A sequential Keras-based deep learning model was used in this study. Essentially, data flow from one dense layer to another layer in the given order until the data finally reach the output layer. Weight and bias parameters of each layer are optimized to reconstruct the prediction X^* from X.

Parameter optimization and cross-validation were performed to determine the best parameters for prediction in each method. To ensure no data leakage, we employed group k-fold cross-validation (**Table 1**). The process was complemented with error evolution, histograms, and quantification. When the distribution of the target variable is mostly Gaussian, for example, large

Table 1—Group k-fold cross-validation (CV).

	Well GU2HI	Well GU3HI	Well KDRK23_1
1ST CV	Test	Train	Train
2ND CV	Train	Test	Train
3RD CV	Train	Train	Test

or small values far from the mean value, the mean absolute value, MAE (**Eq. 4**), is used as an estimator. It is calculated as the average of the absolute difference between the actual and predicted values, i.e.,

$$MAE = \frac{\sum_{i=1}^{n_d} |y_i - x_i|}{n_d}, \quad (4)$$

where n_d = total number of data points, x_i = true value, and y_i = prediction; it is also robust to outliers. To avoid overfitting, approximately 30% of the training data were set aside for validation.

Evaluation of Prediction Models

The ML models were tested with a dataset from another well. Results were compared to conventional Timur-Coates's permeability prediction model (**Eq. 5**) (Timur, 1968). Absolute permeability in terms of porosity and irreducible water saturations is written as

$$k = A\frac{\phi^B}{{s_{wirr}}^C}, \quad (5)$$

where k is permeability, ϕ is porosity, s_{wirr} is irreducible water saturation, and A, B, and C are fitting parameters.

From **Eq. 5** we derive a simplified version of the relationship, given by

$$\log k = B \log \phi + \log A - C \log s_{wirr}, \quad (6)$$

which is employed in machine learning training to estimate the free coefficients (A, B, and C).

For datasets without initial water saturation, we estimated Archie's exponents from well logs using Pickett's plot, calculated the water saturation using Archie's equation, i.e.,

$$S_w = \left(\frac{R_w}{\phi^m R_t}\right)^{\frac{1}{n}}, \quad (7)$$

where S_w is water saturation, R_w is water resistivity, R_t is measured resistivity, ϕ is porosity, m is the porosity

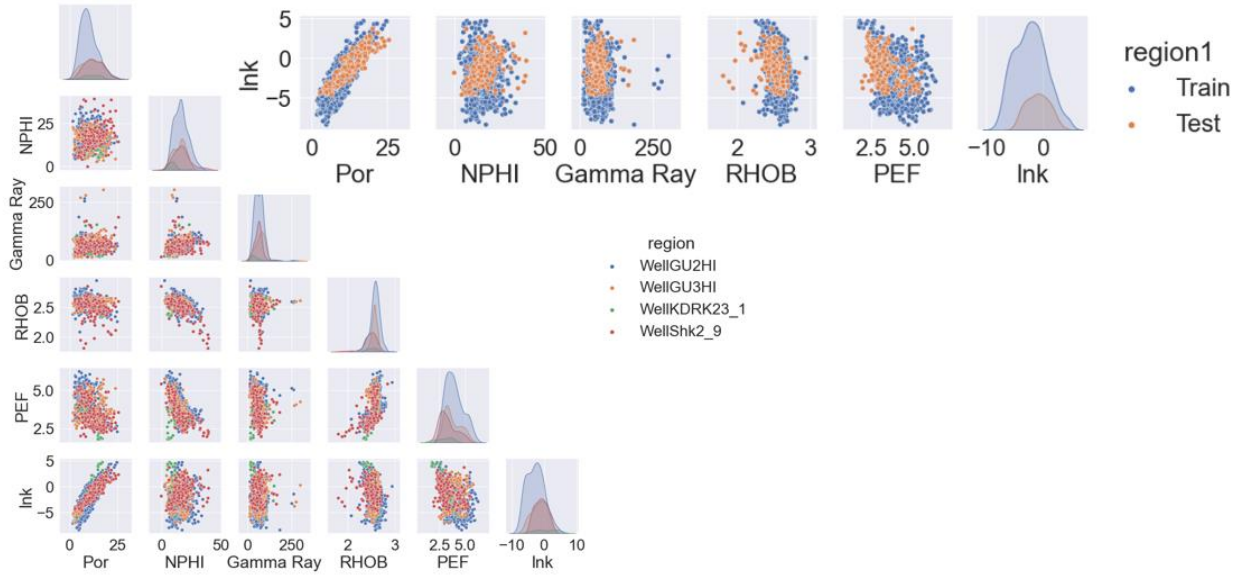

Fig. 5—Cross plots of resampled well logs to core depth. Upper right cross plots show the train test division (Field Case B).

exponent, and n is the saturation exponent. The estimated water saturation values are resampled to core depths.

To evaluate the model's performance, we compared results using the relative standard error (**Eq. 8**) given by

$$RSE = \frac{\sigma(error)}{\sigma(samples)}\ , \qquad (7)$$

where RSE is the relative standard error and σ is the standard deviation.

To further verify the width of prediction, we calculated the variance function based on the validation set and applied the model and error function to estimate uncertainty in the test set. We tested this procedure on fields covering a wide permeability range (**Table 2**).

Table 2—Description of the Field Cases Considered in this Paper.

Fields	Permeability Ranges
A	Mid to high
B	Low to mid-high
C	Very low
D	Mostly high

DISCUSSION OF RESULTS

Exploratory Data Analysis

Figure 5 shows that the logarithm of permeability correlates well with porosity while is highly non-correlated with other input features at core-data depths. It also indicates that training data cover the widest possible range of the test data. Core and calculated well-log porosities were used for depth matching as shown in track 4 of **Fig. 6**. To further improve the resolution of input features, we compared core and well-log resolution support systems. The calculated core bulk density and permeability were smoothed by the same function to well-log resolution (**Fig. 7**).

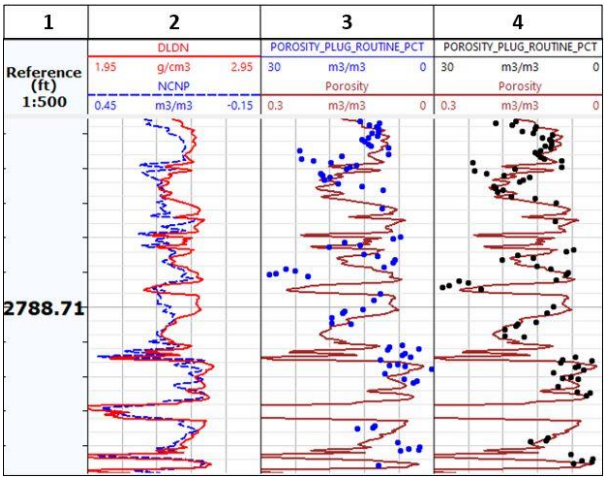

Fig. 6—Well 2-9 depth matched with core and calculated well-log porosity. Track 1: Reference Depth. Track 2: Bulk Density and Neutron Porosity. Track 3: Mismatch between core and calculated porosity from well logs. Track 4: Matched core and calculated porosity from well logs.

Having generated more features, we adopted multiple ML feature augmentation techniques including linear regression, Lasso, Ridge, random forest, and recursive feature elimination to correlate the logarithm of permeability with all input features arranged by mean of ranks. By aggregating the rankings from these methods, we selected the best features. In order to avoid feature cross-correlation (**Figs. 8** and **9**), we generated latent space well logs using PCA, SVD, DWT, and deep-learning autoencoders. With a 95% variance explained, 30 principal components will retain the variation present in the data (**Fig. 10**).

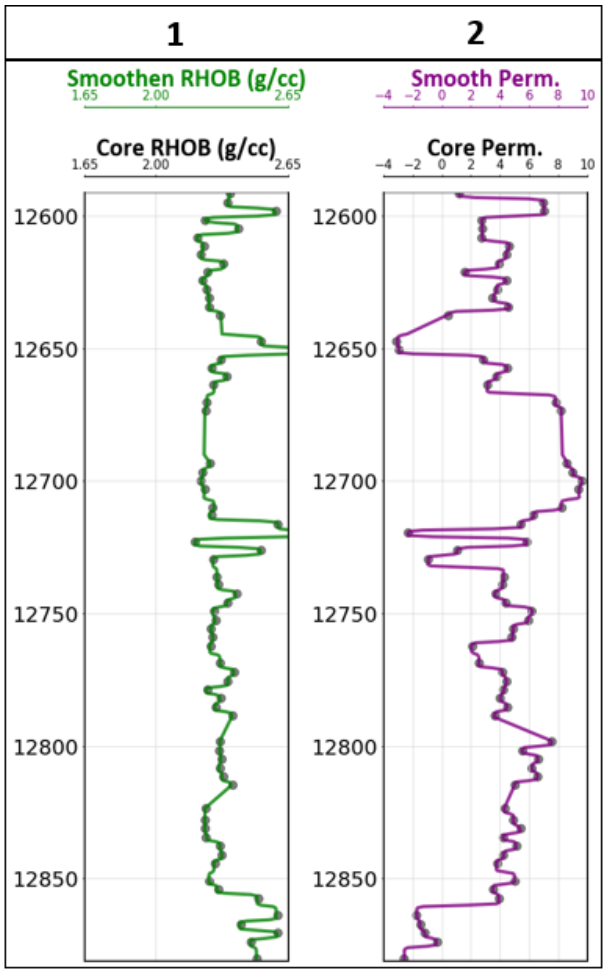

Fig. 7—Core-to-well-log resolution. Track 1: Core vs. smoothed bulk density. Track 2: Core vs. smoothed permeability – Field D.

Machine Learning (ML) Algorithms and Evaluation

We performed hyperparameter tuning (**Fig. 12**) to select the optimal parameters for each ML method. The Scikit-learn techniques (Géron, 2019) have fewer parameters to optimize than deep learning-based techniques. We selected the corresponding loss functions (**Fig. 14**) based on the distribution of the target variable (**Fig. 13**) and volume of outliers. For example, MAE (**Eq. 4**) is used for Gaussian variables.

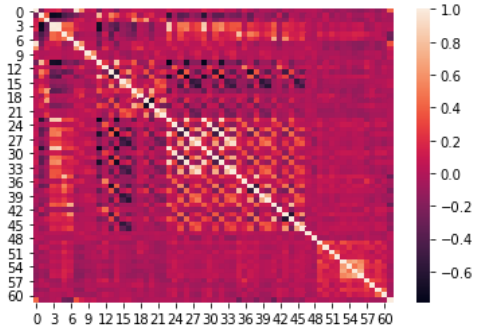

Fig. 8—Before PCA – Correlation exists between input features. Input features are porosity, neutron porosity, gamma ray, bulk density, photoelectric factor, and their corresponding moving-window average of mean, variance, autocorrelation, first- and second-order gradient, slope, and intercept.

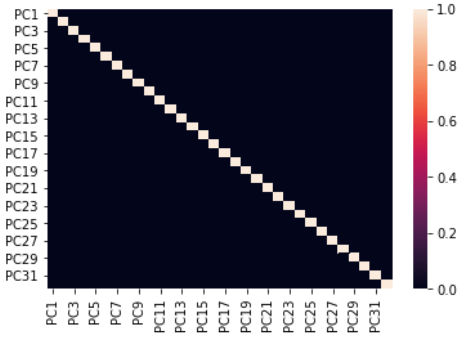

Fig. 9—After PCA – No Correlation exists between transformed input features, referred to as principal components.

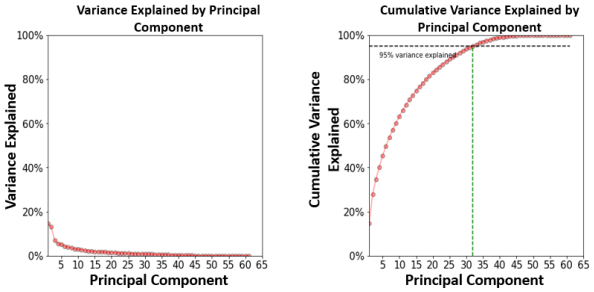

Fig. 10—Screen plot to support the number of dimensions to retain. It considers the variance contributions from each PC.

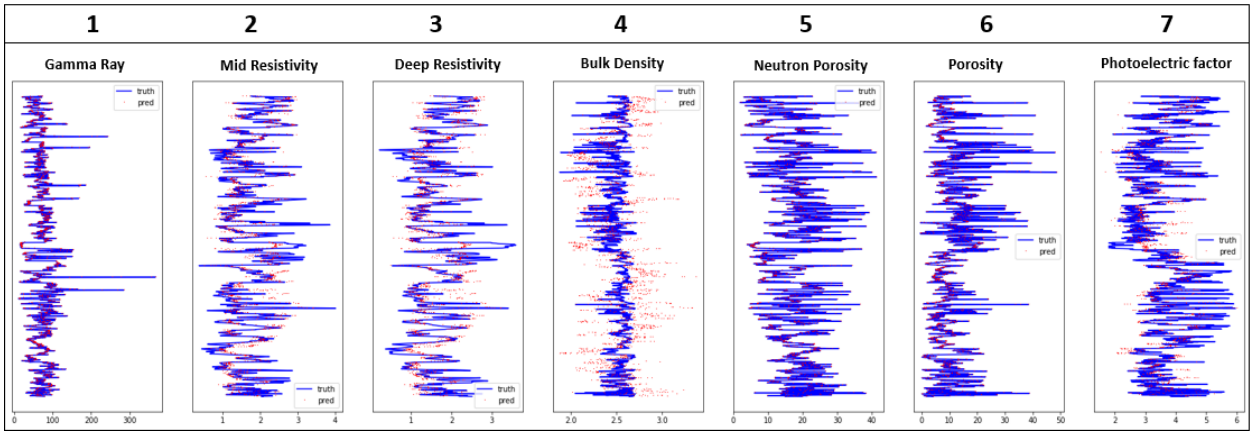

Fig. 11—Latent Space Well Logs. Singular vectors that contain the most energy are: Track 1: Gamma Ray; Track 5: Neutron Porosity, and Track 6: Porosity. Track 2: Mid Resistivity; Track 3: Deep Resistivity; Track 4: Bulk Density and Track 7: Photoelectric factor have the least energy.

distribution of the target variable (**Fig. 13**) and volume of outliers. For example, MAE (**Eq. 4**) is used for Gaussian variables.

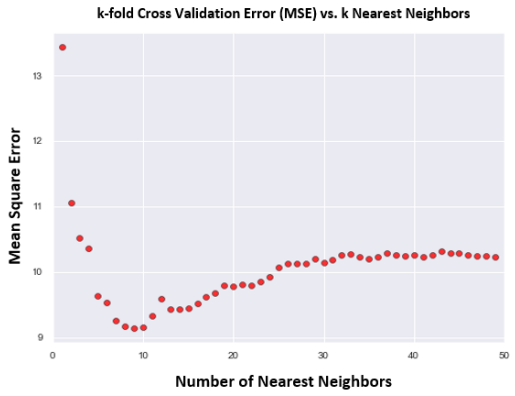

Fig. 12—k-fold cross-validation error vs. k-NN. Parameter optimization used to select the optimal nearest neighbors.

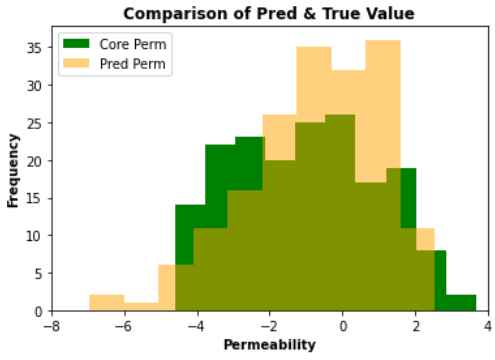

Fig. 13—Comparison of predicted and true-value histograms for the logarithm of permeability.

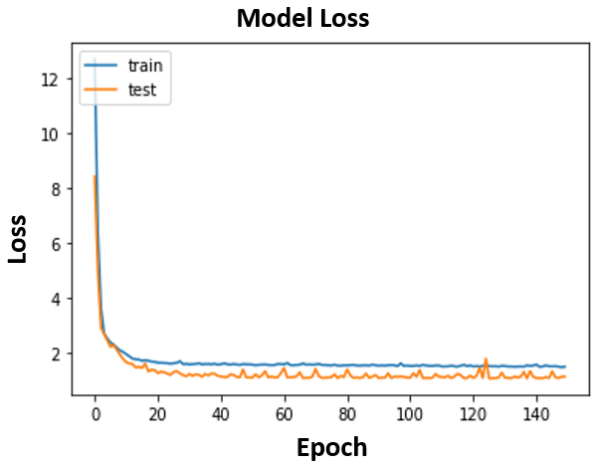

Fig. 14—Error evolution in neural network. Model loss vs. the number of epochs.

Model Performance

RF and ANN best estimate permeability (**Fig. 15** and **16**) from triple-combo well logs across a wide range of variation (0.001 – 2000 mD). An average of 16% relative standard error was calculated from field A in **Table 3** and 13% relative standard error from field B in **Table 4** when using the original well logs (**Figs. 17** and **18**). We observed an improvement while using latent-space well logs with DWT. Calculations from machine learning algorithms reduced the estimation error to less than 13%, while implementing a fully-connected autoencoder resulted in less than 10% error (**Tables 5** and **6**) using the core resolved support system (**Fig. 19**). We obtained a 5% average permeability estimation error (**Tables 5** and **6**) in well-log resolution, a 50% further decrease (**Figs. 20** and **21**) compared to the core support.

Table 3— Percent Error Averages: Testing error – Field A.

	Without PCA			Timur Coates		
	RMSE	RSE	MAE	RMSE	RSE	MAE
Timur Coates's Class (*phi/Swirr*)				2.7	31.7	2.28
Ridge	2.51	30.83	1.89			
Lasso	2.45	30.51	1.91			
kNN	2.84	37.19	2.27			
Random Forest	1.07	15.84	1.70			
SVR	2.51	36.58	1.87			
Neural Network	1.15	16.73	1.79			

Table 4— Percent Error Averages: Testing error - Field B.

		With PCA			Without PCA			Timur Coates		
		RMSE	RSE	MAE	RMSE	RSE	MAE	RMSE	RSE	MAE
Timur Coates	**Without Class**							0.88	28.82	0.69
	Class (*phi/Swirr*)							0.86	26.77	0.687
Ridge					0.92	29.95	0.71			
Lasso					0.94	30.58	0.72			
kNN		2.88	50.12	2.39	1.35	42.39	1.07			
Random Forest		2.99	71.16	2.43	0.02	13.15	0.79			
SVR		3.12	67.55	2.50	1.25	35.66	1.04			
Neural Network	**MSE**				1.21	36.47	0.99			
	MAE				0.09	12.58	0.88			

Table 5— Percent Error Averages: Testing error core vs. well-log comparison - Field B.

	SVD			DWT			FULLY CONNECTED AUTOENCODER		
	Core RSE	**Log RSE**	**% Diff.**	**Core RSE**	**Log RSE**	**% Diff.**	**Core RSE**	**Log RSE**	**% Diff.**
Ridge	16.06	19.48	-21%	16.63	18.90	-14%	23.79	27.61	-16%
Lasso	17.19	19.07	-11%	17.18	19.01	-11%	19.35	24.37	-26%
kNN	39.39	7.67	81%	38.35	17.96	53%	36.63	18.15	50%
Random Forest	17.60	5.26	70%	15.84	10.27	35%	33.74	15.40	54%
SVR	37.63	14.66	61%	15.57	40.97	-163%	21.23	31.12	-47%
Neural Network	15.94	14.98	6%	13.48	26.14	-94%	10.56	18.85	-79%

Table 6— Percent Error Averages: Testing error core vs. well-log comparison - Field D.

	SVD			DWT		FULLY CONNECTED AUTOENCODER	
	Core RSE	**Log RSE**	**% Diff.**	**Log RSE**	**% Diff.**	**Log RSE**	**% Diff.**
Ridge	24.46	21.77	11%	22.35	9%	16.44	33%
Lasso	25.13	22.01	12%	21.74	13%	16.85	33%
kNN	24.30	14.14	42%	11.49	53%	21.53	11%
Random Forest	21.03	7.65	64%	5.86	72%	29.35	-40%
SVR	25.05	20.03	20%	22.82	9%	21.16	16%
Neural Network	22.79	15.12	34%	13.02	43%	15.99	30%

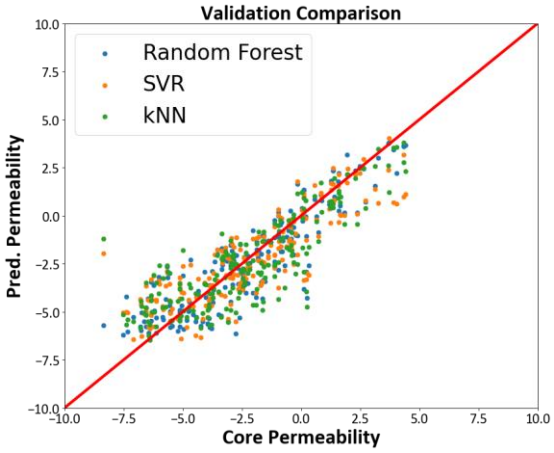

Fig. 15—Validation: Without PCA - Field B.

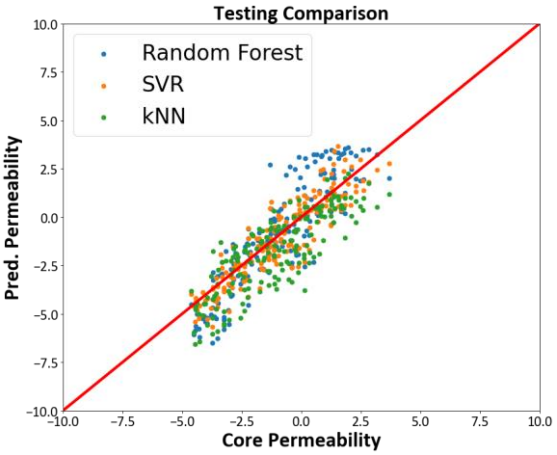

Fig. 16—Prediction: Without PCA - Field B.

Conventional Petrophysical Model

Using Pickett's plot (**Fig. 22**), Archie's exponents (**Table 7**) were calculated from well logs, and the resulting water saturation was estimated using Archie's **Eq. 7**, which was further resampled to core depths. The Timur-Coates model is the most reliable for data sets with *a priori* information about irreducible water saturation, yielding less than 22% relative standard error (**Tables 3** and **4**), yet it requires prior data classification to improve estimation accuracy (**Figs. 23 – 25**). **Figures 26** and **27** show good fitting and prediction using Timur-Coates coefficients (**Table 8**).

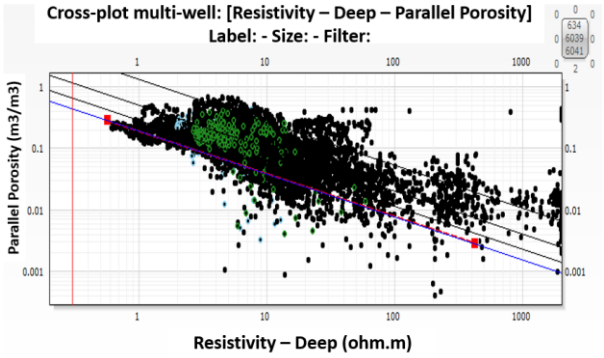

Fig. 22—Pickett Plot. Used for the estimation of Archie's parameters from which initial water saturation is calculated.

Table 7—Archie's Parameters Obtained from Pickett's Plot.

Parameters	Values
a	1.00
m	1.43
n	2.00
R_w	0.09

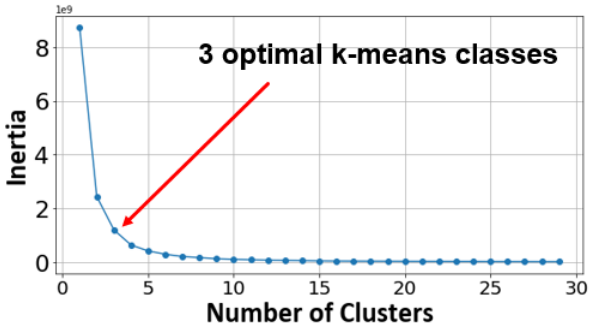

Fig. 23—k-means classification.

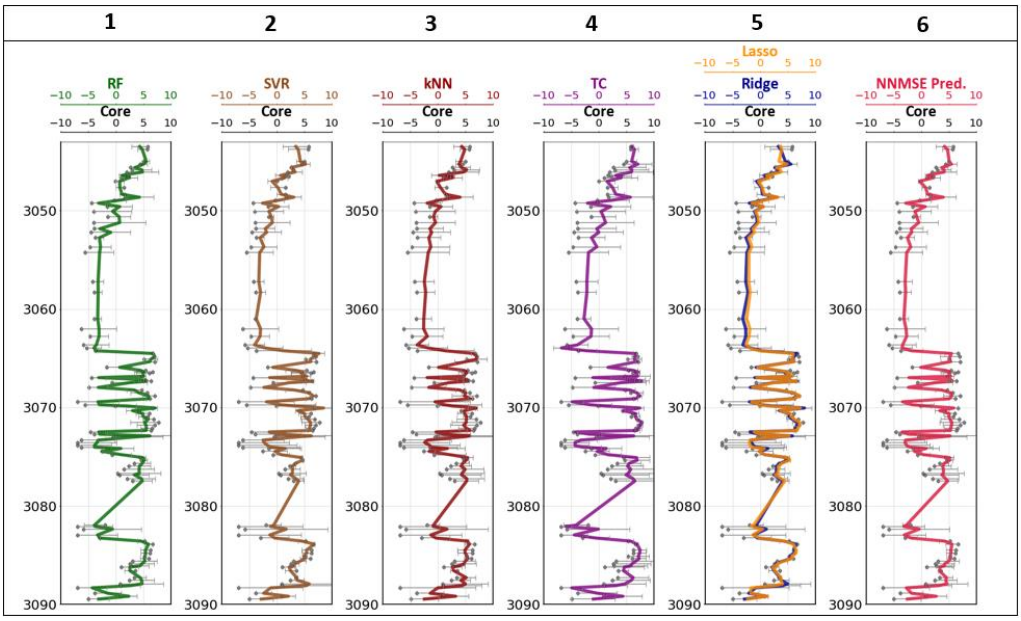

Fig. 17—Comparison of permeability prediction. Most predictions are fairly consistent with core permeability; they differ in calculated errors - **Field A**. Track 1: Core permeability (dots), random forest prediction and error bars. Track 2: Core permeability (dots), support vector regression prediction and error bars. Track 3: Core permeability (dots), k-nearest neighbor prediction and error bars. Track 4: Core permeability (dots), Timur Coates's prediction and error bars. Track 5: Core permeability (dots), Lasso and Ridge predictions and error bars. Track 6: Core permeability (dots), neural network prediction and error bars.

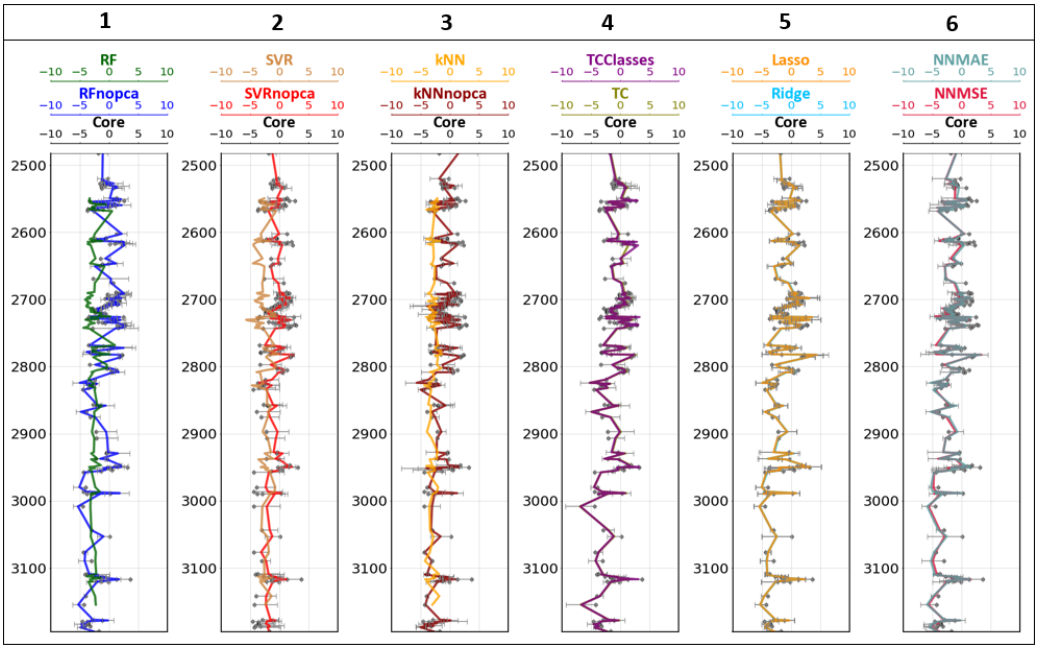

Fig. 18— Comparison of permeability prediction **with and without principal component analysis**. Most predictions are fairly consistent with core permeability without PCA - **Field B**. Track 1: Core permeability (dots), random forest predictions and error bars. Track 2: Core permeability (dots), support vector regression predictions and error bars. Track 3: Core permeability (dots), k-nearest neighbor predictions and error bars. Track 4: Core permeability (dots), Timur Coates's predictions and error bars. Track 5: Core permeability (dots), Lasso and Ridge predictions and error bars. Track 6: Core permeability (dots), neural network predictions and error bars.

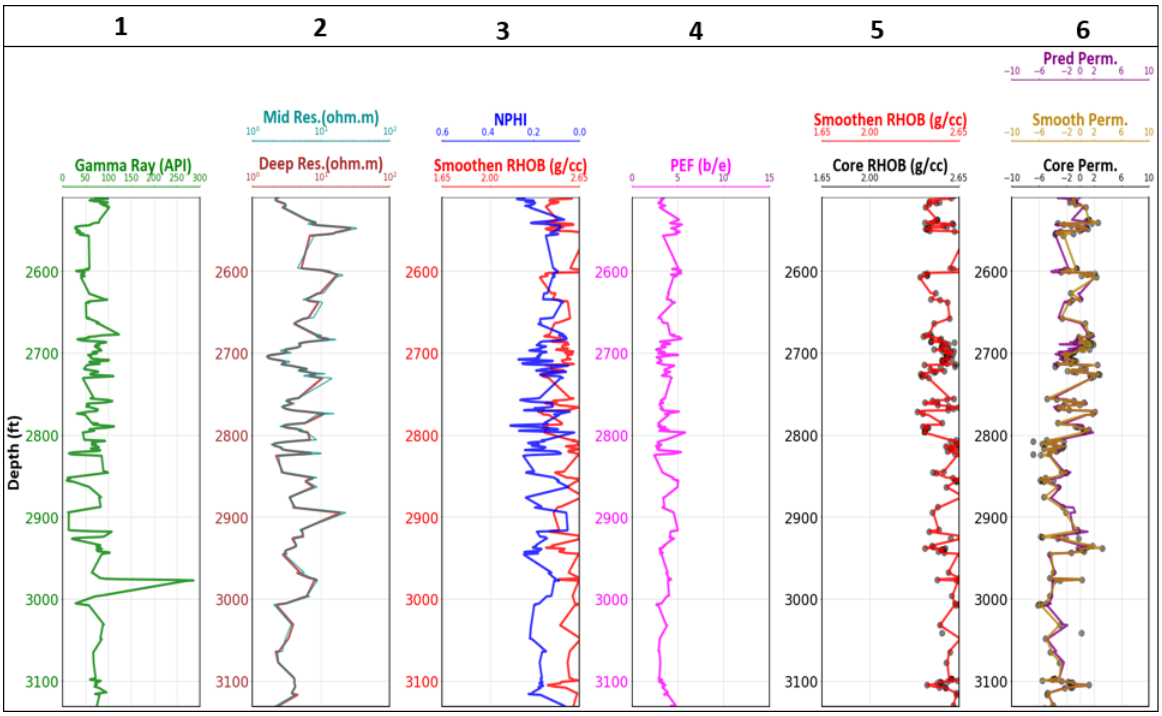

Fig. 19—Smooth permeability prediction using **core resolution** – Well 3HI- **Field B**. Track 1: Gamma ray within the zone of prediction. Track 2: Mid and Deep resistivities. Track 3: Neutron porosity and smooth bulk density. Track 4: Photoelectric factor. Track 5: Core bulk and smooth bulk density. Track 6: Comparison of core and smoothed core permeability and predicted permeability.

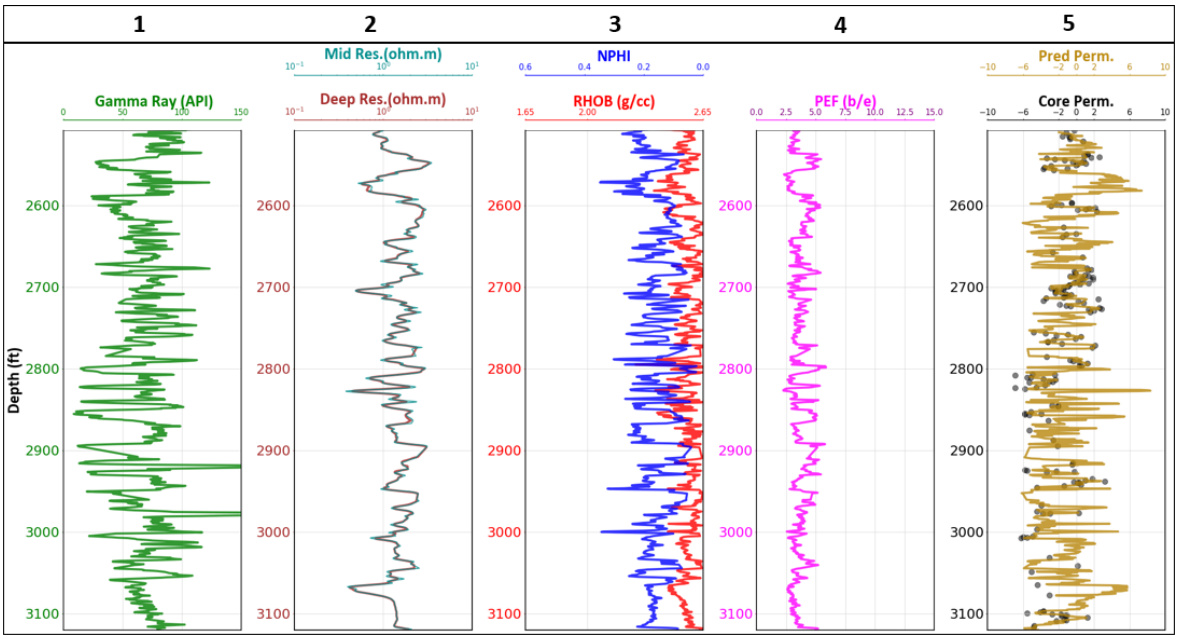

Fig. 20—Smooth permeability prediction using **log resolution** – Well 3HI- **Field B**. Track 1: Gamma ray within the zone of prediction. Track 2: Mid and Deep resistivities. Track 3: Neutron porosity and bulk density. Track 4: Photoelectric factor. Track 5: Comparison of core permeability and smooth predicted permeability.

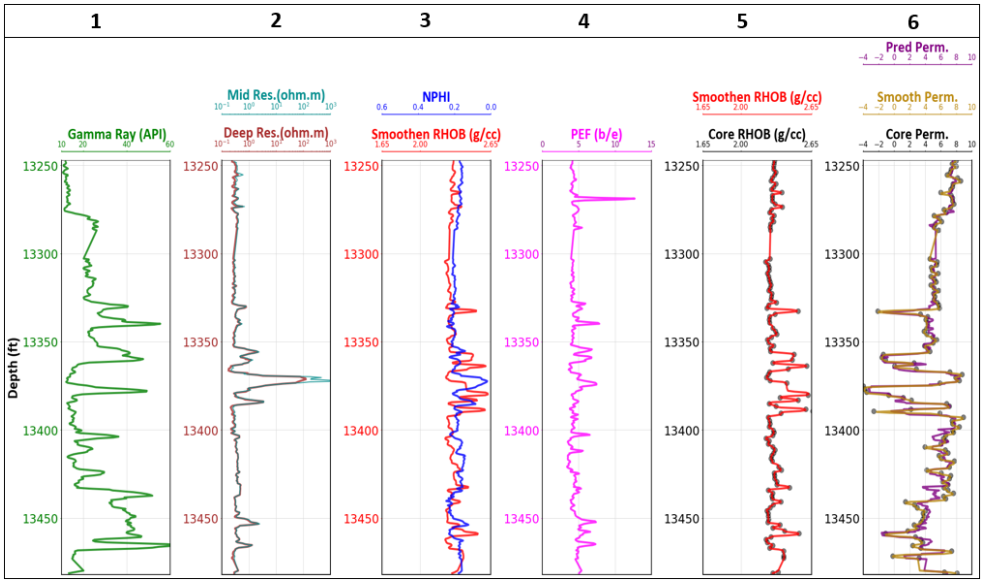

Fig. 21—Smooth permeability prediction using **log resolution** – Well BBT- **Field D**. Track 1: Gamma ray within the zone of prediction. Track 2: Mid and Deep resistivities. Track 3: Neutron porosity and smooth bulk density. Track 4: Photoelectric factor. Track 5: Core bulk and smooth bulk density. Track 6: Comparison of core and smoothened core permeability and predicted permeability.

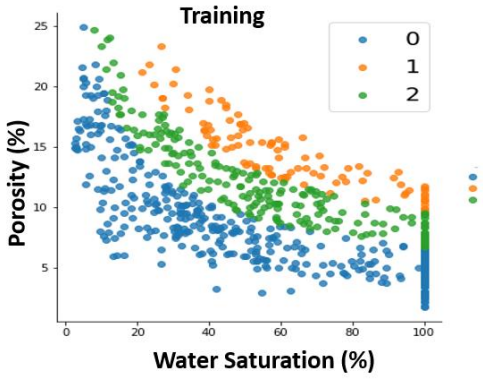

Fig. 24—Classifying and training Timur Coates' coefficients - Field B.

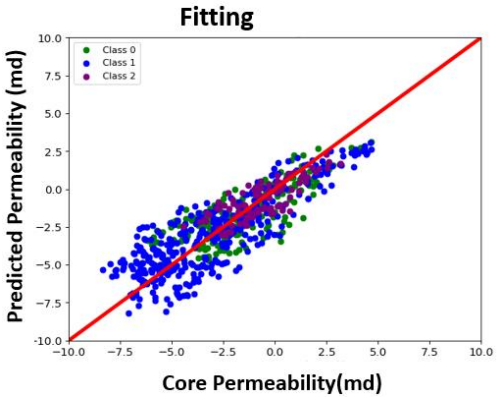

Fig. 26—Timur Coates' fitting – Field B.

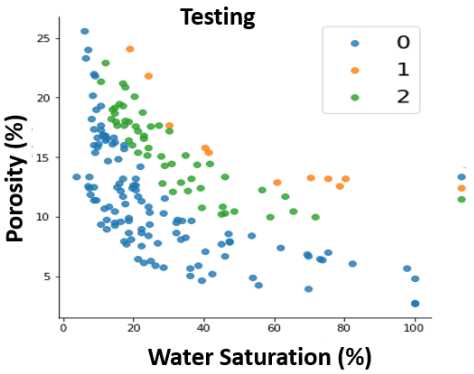

Fig. 25—Predicting Classes on Test Data - Field B.

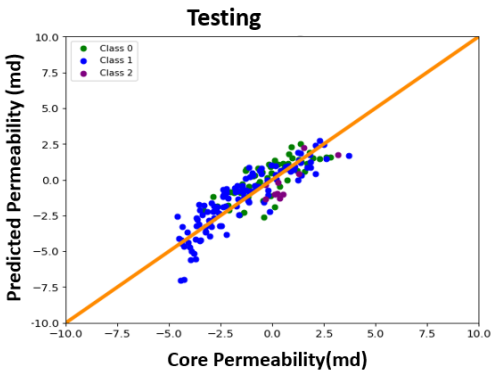

Fig. 27—Timur Coates' Prediction - Field B.

Table 8—Timur Coates's Constants.

	A	B	C
Class 0	0.12962	2.32847	1.57996
Class 1	1.7e-08	5.99745	-0.32180
Class 2	1.6e-03	3.74990	1.14829

Uncertainty Prediction: SVR – Field B

Uncertainty estimates show that permeability calculations are accurate, as their distributions border the true values within ± 5 mD. Regression is more accurate within 1 standard deviation (**Fig. 28**) than double standard deviation (**Fig. 29**).

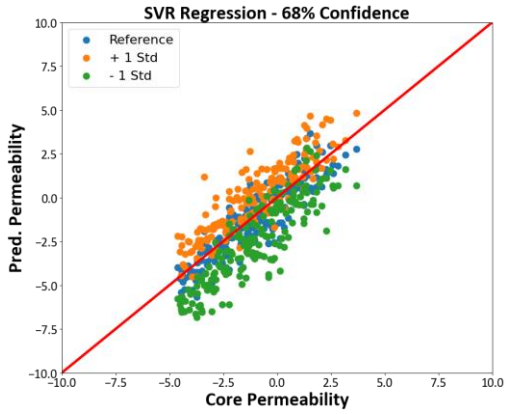

Fig. 28—Uncertainty prediction within 1 standard deviation using SVR - Field B.

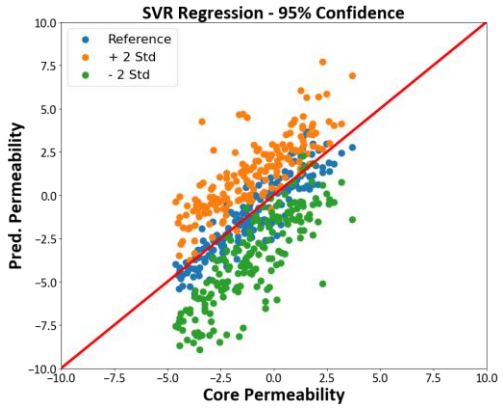

Fig. 29—Uncertainty prediction within 2 standard deviations using SVR – Field B.

The proposed workflows have proven to be widely applicable. Therefore, they can be used to estimate permeability in both conventional and unconventional reservoirs.

CONCLUSIONS

We introduced a comprehensive machine learning-based workflow for automatic permeability estimation. The procedure can be applied to data sets from a variety of carbonate and clastic (shaly and clean/shale-free) rocks, both conventional and low-permeability formations. It is computationally efficient and accurate. Permeability calculations using the workflow introduced in this paper are accurate within ± 5 mD.

Performing feature engineering, dimensionality reduction techniques, and training on latent well logs can expose the physics of input datasets to decrease permeability prediction errors. Random Forest and Artificial Neural Networks best estimate permeability with a 5% average permeability estimations error in well-log resolution. Machine Learning methods outperform the Timur-Coates method for all the field examples examined in this paper.

For best practices and results, training wells must cover the widest range of measurements and petrophysical/fluid properties and core data must be smoothed to well-log resolution. In some cases, it might be necessary to perform rock classification as part of the training/prediction process. Data normalization does not always improve machine learning prediction results, especially in very low and high permeability formations.

Field examples examined in this paper included cases of carbonate and clastic rocks, permeability being closely related to porosity, neutron porosity, bulk density, gamma ray, and photoelectric factor. Prior to this study, conventional interpretation methods that assumed permeability was only related to porosity and irreducible water saturation failed to incorporate other rock properties (for instance, rock type) thereby limiting its application across diverse hydrocarbon fields. The application of the method introduced in this paper confirms that it is possible to estimate flow-related petrophysical properties from well logs using machine-learning methods.

ACKNOWLEDGEMENTS

This project was funded by the University of Texas at Austin's Research Consortium on Formation Evaluation, jointly sponsored by AkerBP, Baker Hughes, BP, Chevron, CNOOC International, ConocoPhillips, ENI, Equinor, ExxonMobil, Fieldwood Energy E&P Mexico, Halliburton, Inpex Corporation, Oxy, Petrobras, Repsol, Schlumberger, Total Energies, Wintershall, and Woodside. Carlos Torres-Verdín is thankful for the financial support provided by the Brian James Jennings Memorial Endowed Chair in Petroleum and Geosystems Engineering.

Abbreviations

ANN = Artificial Neural Network
CV = Cross Validation
DWT = Discrete Wavelets Transform
k-NN = k Nearest Neighbors
lnk = Natural Logarithm of Permeability
MAE = Mean Absolute Error
NN = Neural Networks
NPHI = Neutron Porosity
PCA = Principal Component Analysis
PEF = Photoelectric Factor
Perm = Permeability
Por = Porosity
RHOB = Bulk Density
RSE = Relative Standard Error
RMSE = Root Mean Square Error
SVD = Singular Values Decomposition
SVR = Support Vector Regression

Symbols

σ = standard deviation
n_d = total number of data points
x_i = true value
y_i = prediction
k = permeability
s_{wirr} = irreducible water
s_{wi} = initial water saturation
ϕ, *Por* = Porosity
A, B, C = fitting parameters
a = Winsauer factor
m = porosity exponent
n = saturation exponent
R_w = formation water resistivity

REFERENCES

Anand, V., Freedman, R., Crary, S. et al., 2010, Predicting Effective Permeability to Oil in Sandstone and Carbonate Reservoirs from Well Logging Data, Paper SPE-134011-PA. SPE Reservoir Evaluation and Engineering 14 (06). DOI: 10.2118/134011-PA.

Arigbe, O.D., 2020, *Uncertainty Reduction in Reservoir Parameters Prediction from Multiscale Data Using Machine Learning in Deep Offshore Reservoirs*, Ph.D. dissertation, Robert Gordon University, Scotland (June 2020).

Belozerov, B., Bukhanov, N., Egorov, D. et al., 2018, Automatic Well Log Analysis Across Priobskoe Field Using Machine Learning Methods, Paper SPE-191604 presented at the SPE Russian Petroleum Technology Conference, Moscow, Russia, October 15-17. DOI: 10.2118/191604-18RPTC-MS.

Bennis, M., and Torres-Verdín, C., 2019, Estimation of Dynamic Petrophysical Properties from Multiple Well Logs Using Machine Learning and Unsupervised Rock Classification, Paper SPWLA-2019-KKKK, *SPWLA 60th Annual Logging Symposium*, The Woodlands, Texas, USA, June 15-19. DOI: 10.2118/134011-MS.

Bennis, M., and C. Torres-Verdín, 2023, Automatic Multiwell Assessment of Flow-Related Petrophysical Properties of Tight Gas Sandstones Based on The Physics of Mud-Filtrate Invasion. Paper SPE-214668-PA. SPE Reservoir Evaluation and Engineering, 1–22. DOI:10.2118/214668-PA.

Eriavbe, F. E., and Okene, U. O., 2019, Machine Learning Application to Permeability Prediction Using Log & Core Measurements: A Realistic Workflow Application for Reservoir Characterization, Paper SPE-198874 presented at the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, August 5-7. DOI: 10.2118/198874-MS.

Géron, A., 2019, *Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, Inc., Sebastopol, CA 95472. ISBN: 978-1-492-03264-9.

Guimarães, C., 2021, *Permeability Predictions Using Borehole Logs and Well Testing Data: A Machine*

Learning Approach, MS thesis, PUC-Rio, Rio De Janeiro (March 2021).

Mohamed, T. S., Bennis, M., Torres-Verdín, C., Merletti, G., and Gelvez, C., 2023, Using Formation-Tester Measurements to Estimate Depth of Invasion and Water Saturation in Deeply Invaded Tight-Gas Sandstones, SPWLA 64th Annual Logging Symposium, Conroe, Texas, June 10-14.

Negara, A., Jin, G., and Agrawal, G., 2016, Enhancing Rock Property Prediction from Conventional Well Logs Using Machine Learning Technique-Case Studies of Conventional and Unconventional Reservoirs, Paper SPE-183106 presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 7-10. DOI: 10.2118/183106-MS.

Singh, M., Makarychev, G., Mustapha, H., et al., 2020, Machine Learning Assisted Petrophysical Logs Quality Control, Editing, and Reconstruction, Paper SPE-202977 presented at the Abu Dhabi International Petroleum Exhibition & Conference, Abu Dhabi, UAE, November 9-12. DOI: 10.2118/202977-MS.

Timur, A., 1968, An Investigation of Permeability, Porosity, And Residual Water Saturation Relationship for Sandstone Reservoirs, Paper SPWLA-1968-vIXn4a2. SPWLA Journal of Formation Evaluation and Reservoir Description Petrophysics 9 (04).

ABOUT THE AUTHORS

Oriyomi Raheem is a Ph.D. student in the Hildebrand Department of Petroleum and Geosystems and a Graduate Research Assistant in the Formation Evaluation Consortium Research Program at The University of Texas at Austin. He received his MSc degree in Petroleum Engineering from the Khalifa University of Science Technology. He is currently an officer of The SPWLA Student Chapter at The University of Texas at Austin.

Wen Pan received a Ph.D. in Petroleum Engineering from the University of Texas at Austin in 2022. His research interests include spatial statistics, data analytics, machine learning, optimization, geomodelling, petrophysics, seismic processing, and reservoir simulation. He currently works with Shell as an Artificial Intelligence Resident.

Misael M. Morales is a Ph.D. student in the Hildebrand Department of Petroleum and Geosystems Engineering and a Graduate Research Assistant in the Digital Reservoir Characterization Technology and Formation Evaluation Consortia at the University of Texas at Austin. He obtained his BSc in Petroleum Engineering and Applied Mathematics, and an MSc in Mathematics, from the University of Tulsa. His research focuses on machine learning applications for subsurface energy resource modeling, characterization, and simulation, specifically in dimensionality reduction techniques for inverse modeling, data assimilation, uncertainty quantification and optimization.

Carlos Torres-Verdín received a Ph.D. in Engineering Geoscience from the University of California at Berkeley in 1991. During 1991-1997, he held the position of Research Scientist with Schlumberger-Doll Research. From 1997-1999, he was Reservoir Specialist and Technology Champion with YPF (Buenos Aires, Argentina). Since 1999, Dr. Torres-Verdín has been affiliated with the Department of Petroleum and Geosystems Engineering of the University of Texas at Austin, where he is currently a Full Professor, and holds the Brian James Jennings Memorial Endowed Chair in Petroleum and Geosystems Engineering. Dr. Torres-Verdín is the founder and director of the Research Consortium on Formation Evaluation at the University of Texas at Austin. He is the recipient of the Cockrell School of Engineering's 2016-2017 Lockheed Martin Aeronautics Company Award for Excellence in Engineering Teaching, the 2014 Gold Medal for Technical Achievement from the SPWLA, the 2008 Formation Evaluation Award from the SPE, and the 2006 Distinguished Technical Achievement Award from the SPWLA. He is a Distinguished Member of the Society of Petroleum Engineers (SPE), an Honorary Member of the Society of Exploration Geophysicists (SEG), and a receiver of the Conrad Schlumberger Award from the European Association of Geoscientists and Engineers (EAGE).