

Estimating Compressional Velocity and Bulk Density Logs in Marine Gas Hydrates using Machine Learning

Fawz Naim^a, Ann E. Cook^a and Joachim Moortgat^a

^aSchool of Earth Sciences, The Ohio State University, Columbus, OH, USA

Corresponding author: Fawz Naim (naim.9@buckeyemail.osu.edu)

Abstract: Compressional velocity (V_p) and bulk density (ρ_b) logs are essential for characterizing gas hydrates and near-seafloor sediments, however it is sometimes difficult to acquire these logs due to poor borehole conditions, safety concerns or cost related issues. We present a machine learning approach to predict either compressional V_p or ρ_b logs with a high accuracy and low error in near seafloor sediments within water saturated intervals, in intervals where hydrate is filling fractures and intervals where hydrate occupies the primary pore space. We use scientific quality logging-while drilling well logs, gamma ray, ρ_b , V_p , and resistivity, to train the machine learning model to predict V_p or ρ_b logs. Of the six machine learning algorithms (Multilinear Regression, Polynomial Regression, Polynomial Regression with Ridge Regularization, K Nearest Neighbors, Random Forest and Multilayer Perceptron) tested, we find that the Random Forest and K Nearest Neighbors algorithms are best suited to predict V_p and ρ_b logs based on coefficients of determination (R^2) greater than 70% and mean absolute percentage errors less than 4%. Due to the high accuracy and low error results for V_p and ρ_b prediction in both hydrate and water saturated sediments, we argue our model can be applied in most LWD wells to predict V_p or ρ_b logs in near seafloor siliciclastic sediments on continental slopes irrespective of the presence or absence of gas hydrate.

Keywords: Gas hydrate; well logs; compressional velocity; bulk density; random forest; k nearest neighbors

1. Introduction

Natural gas hydrate occurs in near seafloor sediments worldwide; detecting and quantifying gas hydrate is a challenge, but important for understanding the amount and contribution of gas hydrate in the global carbon cycle as well as for assessing gas hydrate as a prospective energy resource [1,2]. Out of the different methods for interpreting hydrate, downhole logging measurements are the most accurate way to identify the amount of gas hydrate in the subsurface.

The most common downhole logs used for interpreting gas hydrate are compressional velocity (V_p), resistivity and bulk density (ρ_b) [3]. The measurement response for V_p in hydrate bearing sediments depends on whether hydrate occurs in the primary pore space or as fill in veins or fractures. In coarse grained sand or silt hydrate nucleates in the primary pore space [4,5]. When hydrate saturation exceeds ~40% hydrate begins forming a rigid framework ; at that saturation there is a distinct increase in formation moduli that increases V_p relative to water saturated sediments [6]. Hydrate in marine muds and clays is usually observed in fractures, and those fractures likely grow in place due to the formation of hydrate and methane supplied via microbial methanogenesis [7]. V_p , however, does not usually increase significantly in hydrate filled fractures, as these accumulations usually have lower hydrate saturation than sand or silt layers [8].

Gas hydrate increases the electrical resistivity as it is an electrical insulator [3]. When hydrate is in the primary pore space, resistivity increases with increasing hydrate

saturation [9,10]. However, when hydrate is in near-vertical fractures the increase in resistivity is not only related to the amount of hydrate but also depends on fracture orientation [11].

Bulk density in near seafloor sediments provides the most accurate measurement of porosity (e.g., [3]). Porosity is essential for calculating hydrate saturation using both resistivity and V_p [12,13]. Therefore, bulk density is linked to the interpretation of hydrate. However, the small difference in the bulk density of hydrate (0.92 g/cm^3 [12]) and pore water (1.02 g/cm^3) makes hydrate effectively undetectable from the bulk density log.

Hydrate interpretation relies on good quality V_p and q_b logs, however these logs are sometimes poor quality or are not acquired in near seafloor sediments. For example, there are ~70 LWD scientific ocean drilling holes with missing V_p or q_b logs on the Lamont Doherty Earth Observatory database.

Machine learning is an effective tool that can be used for building both linear and non-linear correlations to predict or fill-in missing data [14,15]. Supervised learning is a type of machine learning that trains a model using input and output features from a labeled input dataset and predicts on a new or novice dataset to test the accuracy of the model [16]. For supervised learning, available data is often split into training and validation data sets [16]. The training dataset is used so that the computer model can learn; in addition, a small proportion of the training dataset is used to validate the model [16].

Supervised machine learning models have been applied to marine geology [17–22] and geophysics [23,24], geochemistry [25,26] and gas hydrate [26–29] datasets to predict different physical properties. For example, Graw et al. [30] used the Random Forest algorithm to predict global seafloor sediment bulk density using core measurements acquired by scientific ocean drilling programs. Sain and Kumar [31] used artificial neural networks to interpret subsurface geological features with a combination of seismic attributes. Similarly, Farfour & Mesbah, Ismail et al., Ramya et al. [17,18,19] used artificial neural networks for interpreting subsurface features such as gas chimneys, channels, hydrocarbon saturated rocks using marine seismic data. Dumke and Berndt [32] used V_p logs, local geological information (such as water depth and distance to basement) and the Random Forest algorithm to predict subseafloor V_p trends worldwide. In a more related study in an Arctic permafrost region, Singh et al. [27] used a variety of different machine learning algorithms and well log combinations to predict gas hydrate saturation.

In this work, we use a machine learning model to predict V_p and q_b logs in near seafloor sediments, which includes both water saturated and hydrate bearing sediments. This includes predicting V_p and q_b logs and their variation with different depths and different hydrate morphologies including hydrate in pores and hydrate in fractures. Our model results have broader relevance and are not only applicable to marine hydrate systems but may also be useful for researchers working to identify shallow natural hazards such as overpressure intervals or landslides in near seafloor marine sediments [33,34]. In these cases, V_p and q_b are essential inputs for computing overburden stress and pore pressure [33,35]. In addition, our model results will be useful for well to seismic ties, since V_p and q_b logs are essential inputs to link seismic data (measured in time) to well logs (measured in depth) (e.g. [36]).

2. Data

For our machine learning model, we use data only acquired by logging-while-drilling (LWD) tools as they collect the highest quality well logging datasets in a borehole. This is because LWD tools are placed directly behind the drill bit and acquire data before sediments have time to erode [3]. This ensures that the machine learning model is trained on quality data and can make predictions with high accuracy.

We downloaded all the available LWD data from 22 holes from three primary locations on continental slopes from the Lamont-Doherty Earth Observatory database to train, validate and test each machine learning model (Figure 1): seven holes from the Gulf of Mexico collected by the Gas Hydrate Joint Industry Project (JIP) Leg II [37], three holes

from Cascadia Margin collected during Integrated Ocean Drilling Program (IODP) Expedition 311 [38] and 12 holes from Bay of Bengal collected during the Indian National Gas Hydrate Program (NGHP) Expedition 01 [39]. All these holes host a range of siliciclastic sediment types and some of these holes contain natural gas hydrates.

2.1. Training Holes

We use LWD datasets from 20 holes from the Gulf of Mexico, Cascadia Margin and Bay of Bengal to train the machine learning model (Figure 1).

The training holes from the northern Gulf of Mexico were drilled by JIP Leg II and are in Green Canyon (Figure 1) and Alaminos Canyon (Figure 1) [37]. The three holes in Green Canyon in Block 955 (GC955) in ~2 km of water and with sediments sourced from turbidite channel-levee complexes as well as hemipelagic marine muds [40,41]. Hole GC955-H has high quality LWD data drilled to 590 mbsf that includes 412 m of water saturated sediments, 144 m of near-vertical gas hydrate filled fractures in clay sediments with low hydrate saturations, and a 34 m of hydrate in the primary pore space of a coarse silt reservoir with saturation ranging from 30–80% [42]. Holes GC955-Q and GC955-I also have high quality LWD data to 461 and 671 mbsf in mostly water saturated sediments [37]. Alaminos Canyon Block 21 (AC21) lies in the northwestern Gulf of Mexico at a water depth of ~1.5 km. Holes AC21-A and AC21-B are drilled to a depth of 536 and 340 mbsf. Sediments in both holes are primarily water saturated marine muds, with one ~60 m water saturated sand interval that is part of a large submarine fan system [43,44].

IODP Expedition 311 drilled and logged turbidite sequences on the Cascadia subduction zone (Figure 1). The training Holes U1325A, U1327A and U1328A from Cascadia Margin (the yellow dots in Figure 1) are mostly water saturated but also have gas hydrate accumulations. The average gas hydrate saturation ranges from 4–10%, with local maximums up to 80% [10]. In Hole U1325A, drilled to a depth of 350 mbsf, most of the hydrate is present in thin sands (< 23 cm) [10]. Hole U1327A, drilled to a depth of 300 mbsf, is water saturated except for an 18 m thick high resistivity interval composed of hydrate-saturated turbidite lenses [45]. Hole U1328A is drilled to a depth of 300 mbsf; in this hole, gas hydrate filled fractures were identified on the resistivity image logs from the seafloor to 46 mbsf, while the remaining 254 m are water saturated marine muds [45].

The training holes drilled and logged offshore India as a part of NGHP-01 have high quality LWD data. Holes 2A, 2B, 3A, 4A, 5A, 5B, 6A, 7A, 10A and 11A lie in the Krishna-Godavari Basin and the Holes 8A and 9A are located in the more northern Mahanadi Basin (Figure 1; Table 1). Both locations have clay-rich sediments that are primarily water saturated; almost all gas hydrate encountered during the NGHP-01 Expedition occurs in marine muds in near-vertical fractures [39].

We use all the available LWD logging data from NGHP-01 holes except some data from Hole 10A. Hole 10A is located at a paleo-vent site in the Krishna-Godavari Basin and consist of a webby network of veins and fractures [46]. The propagation resistivity logs in Hole 10A exceed the accuracy range in the interval 43–90 mbsf and are not valid measurements [11]. Therefore, we do not use the data in the interval 43–90 mbsf from Hole 10A for training the model and use the data below 90 mbsf.

Hole	Location	Drilling Project	Water Depth (m)	Total Depth Drilled (mbsf)	Water Saturated Intervals (m)	Hydrate in fractures (m)	Hydrate in pores (m)
GC955-H	Gulf of Mexico	JIP Leg II	2033	590	412	144	34
GC955-I			2064	671	666	0	~5
GC955-Q			1985	461	437	0	~24
AC21-A			1490	536	436	79	21
AC21-B			1488	340	301	0	39
WR313-G			2000	1043	<753	>246	44
WR313-H			1966	1000	626	325	49
U1325A	Cascadia Margin	IODP Expedition 311	2192	350	>349	0	<0.23
U1327A			1305	300	282	0	18
U1328A			1267	300	254	46	0
NGHP-01-02A	Bay of Bengal	NGHP Expedition 01	1058	50	50	0	0
NGHP-01-02B			1058	250	250	0	0
NGHP-01-03A			1076	300	91	209	0
NGHP-01-04A			1081	300	280	20	0
NGHP-01-05A			945	200	161	39	0
NGHP-01-05B			945	200	163	37	0
NGHP-01-06A			1160	350	339	11	0
NGHP-01-07A			1285	260	220	40	0
NGHP-01-10A			1038	205	82	123	0
NGHP-01-11A			1007	200	180	20	0
NGHP-01-08A			1689	350	313	37	0
NGHP-01-09A			1935	330	230	100	0

Table 1. Training, validation and test datasets for the machine learning model. Holes WR313-G and WR313-H are used to test and all the other holes are used to train and validate the model by splitting into a 70% (train) and 30% (validation) ratio.

142

143

144

145

146

147

2.2. Test Holes

We use two Walker Ridge LWD holes, Holes WR313-G and WR313-H for testing, which assesses the predictability of the model (the white dot in Gulf of Mexico, Figure 1). We selected these two holes for testing as they host the three key intervals that we are focusing on for our machine learning model: water saturated sediments, hydrate in the primary pore space and hydrate in near-vertical fractures.

These holes were drilled in the Terrebonne mini-basin in the Gulf of Mexico with a water depth of about 2 km [47]. A total of ~1220 m in Holes WR313-G and WR313-H are water saturated with a low background resistivity that ranges from 1–2 Ω m. Hydrate with hydrate saturation of 50–90% occurs in the primary pore space of sand or silt layers a total of 50 m between both holes [47].

Hydrate also occurs in near-vertical fractures in marine mud over a total thickness of ~520 m between both holes [48]. Free gas is also present in an interval ~2 m in the Hole WR313-G just below the gas hydrate stability zone [48]. However, we did not include free gas in our machine learning model as this was the only hole with any free gas intervals. The lack of data in free gas intervals is not surprising; in general, free gas intervals are carefully avoided during scientific ocean drilling because they present a potential drilling hazard.

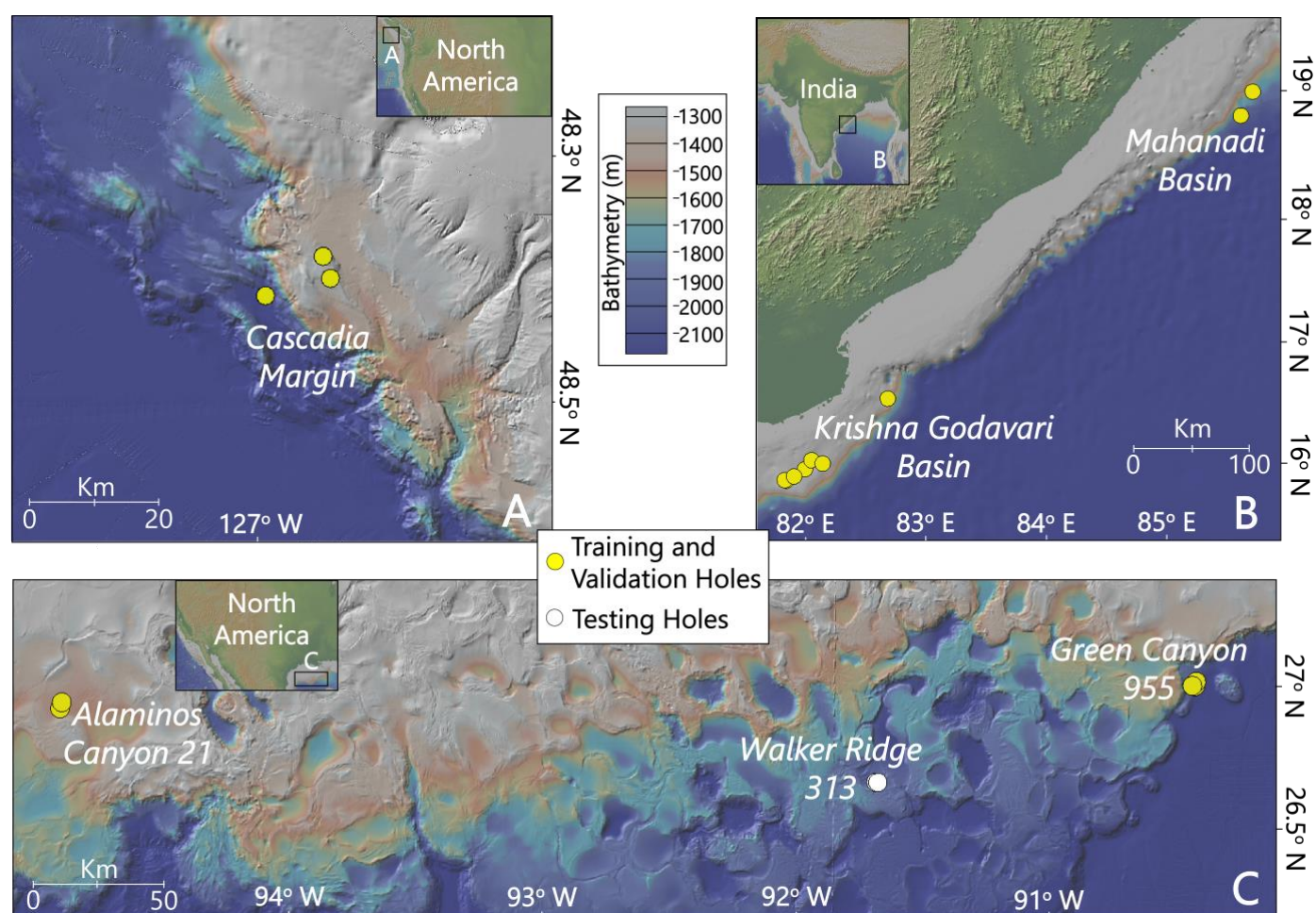


Figure 1. Maps showing the training holes (yellow dots) and testing holes (white dots) (A) Holes located at Cascadia Margin. (B) Holes located offshore India. (C) Holes located in the Gulf of Mexico.

Holes	Location	Gas Hydrate Occurrence	LWD Tools	Logs Used
GC955-H GC955-I GC955-Q AC21-A AC21-B	Gulf of Mexico	Gas hydrate occurs in all the holes.	EcoScope geoVISION sonicVISION	Density Caliper (DCAV), Bulk Density (RHOB), Calibrated and Filtered Gamma Ray (GRMA_FILT), RING Resistivity, Propagation Resistivity (A16L, A40L, P16H, P28H, P40H), V_p
U1327A U1328A U1325A	Cascadia Margin	Gas hydrate occurs in holes U1327A and U1328A.	adnVISION EcoScope geoVISION sonicVISION	Density Caliper (DCAV), Bulk Density (RHOB), Calibrated and Filtered Gamma Ray (GRMA_FILT), RING Resistivity, Propagation Resistivity (A16L, A40L, P16H, P28H, P40H), V_p
NGHP-01-02A NGHP-01-02B NGHP-01-03A NGHP-01-04A NGHP-01-05A NGHP-01-05B NGHP-01-06A NGHP-01-07A NGHP-01-08A NGHP-01-09A NGHP-01-10A NGHP-01-11A	Bay of Bengal	Gas hydrate occurs in all the holes except 02A and 02B	EcoScope geoVISION sonicVISION	Density Caliper (DCAV), Bulk Density (RHOB), Calibrated and Filtered Gamma Ray (GRMA_FILT), RING Resistivity, Propagation Resistivity (A16L, A40L, P16H, P28H, P40H), V_p

Table 2. Holes used for training the machine learning model. V_p is computed from the compressional slowness log, DTCO.

3. Methods

3.1. Machine Learning Algorithms

We predict V_p logs using gamma ray, q_b and resistivity as inputs and q_b logs using gamma ray, V_p and resistivity as inputs using the 20 training holes (Table 2) and test the model in the two Walker Ridge holes (Holes WR313-G and WR313-H). We use all these logs as inputs because they are important for interpreting sediment type, the morphology of hydrate and hydrate saturation. For example, gamma ray differentiates between sand and clay rich sediments. Bulk density measures the electron density of matrix and pore fluids. Resistivity is used to identify gas hydrate at low and high saturations and V_p is used to identify gas hydrate at high saturation [49]. We focus on predicting V_p and q_b logs because they are often poor quality in near seafloor sediments. We do not predict resistivity logs because there are often many resistivity channels collected and, in general, deeper penetrating resistivity logs are often the highest quality measurements in near seafloor sediments.

We use six supervised machine learning algorithms and compare the accuracy and error for each algorithm using R^2 and mean absolute percentage error (MAPE). We selected these algorithms as they have been used previously in geoscience applications

[27,30,31,32]. Some machine learning algorithms have hyperparameters that can be tuned to predict output with the highest accuracy and least error. We use gridsearchcv technique to select the best set of hyperparameters for predicting V_p and Q_b . Gridsearchcv is a cross validation method that splits the training data into different parts and validates the model on each part iteratively while training the model on the remaining set of data points, searching for the optimum set of hyperparameters using all the possible user defined hyperparameter combinations [50]. We split our training data into five folds and perform hyperparameter tuning with gridsearchcv using the process as described by [51,52] (Figure 2). The spreadsheets generated after gridsearchcv with all the possible combination of hyperparameters for each algorithms are provided with the Supplementary Material. We perform k-fold cross validation for all the algorithms using the 20 holes for predicting V_p and Q_b logs on different parts of the dataset (statistics appear in the Supplementary Material). We use a k-fold of five that divides the training data into five parts and validates the machine learning model on each part (Figure 2). This helps identify which algorithms are more consistent in predicting V_p and Q_b that are not biased for a specific set of data points.

A brief description of each algorithm Is given below:

a. Multilinear Regression: Multilinear regression develops a correlation between the provided inputs and output on a labeled training dataset using a linear relationship and the resulting linear model is used to predict values for a new dataset [53]. This algorithm does not require hyperparameter tuning.

b. Polynomial Regression: This algorithm defines a relationship between the input and output parameters based on a n^{th} degree polynomial. The user defines the degree of the polynomial and then the algorithm transforms the input data into a polynomial equation [54]. For a supervised learning model, the same equation is then used to predict outputs on a novice dataset. Herein, we tested polynomial equations from orders two to six and chose a 4th order polynomial equation after hyperparameter tuning.

c. Polynomial Regression with Ridge Regularization (L2): L2 regularization reduces overfitting by adding a penalty term that can be used to reduce the magnitude of large coefficients in the equation [55]. Here, we combine a 4th order polynomial equation with a ridge regression fit on the training data. We use a regularization of 0.001 and 0.01 to predict V_p and Q_b respectively.

D. K Nearest Neighbors: This algorithm uses feature similarity between input and output points in a space to make predictions [56]. Whenever a new dataset is input into the model, the Euclidean distance from the training data points is calculated for all the new data points and then the nearest neighboring values are selected based on the k value, which defines the search criteria and selects k number of nearest neighbors from the input (e.g. [56]). Another parameter, the weight attribute, weighs different points in the neighborhood corresponding to their respective Euclidean distances. The closeness that is calculated as the Euclidean distance from training points is then used to predict an output based on the class of the nearest neighbors [57]. We select $k=7$ and 'distance' as the weight attribute as they fit the model best for predicting V_p and Q_b .

e. Random Forest: As described in [58], and by other research works in geosciences such as Bressan et al., Hou et al., Shalaby et al., [20,22,25] Random Forest uses a bootstrap aggregating method that uses a combination of decision trees and takes the mean out of all the decision trees to generate the final output. Decision trees mimic the structure of a tree and consist of several nodes that terminate on a leaf node [59]. Leaf nodes are representative of class labels, and all other nodes signify feature attributes. Each branch of the tree used in Random Forest is subdivided into nodes based on the conditions that the algorithm tries to construct with reference to the input data provided [59]. This structure of Random Forest reduces variance and avoids overfitting. Herein, we use Random Forest by constructing a forest with '400' trees, 'sqrt' as the max_features that

defines the size of the features to be considered while splitting a node, '1' as min_samples_leaf that refers to the minimum number of samples at the leaf node, '15' as max_depth that refers to the maximum depth of the tree from the root node to the leaf node and '2' as min_samples_split that refers to the minimum number of samples required to split a node.

f. Multilayer Perceptron: A Multilayer Perceptron is an artificial neural network that uses artificial neurons with an input layer, a hidden layer and an output layer to make non-linear predictions based on the inputs provided to it [60]. It is inspired by the structure of biological neurons that receive signals from other neurons via interconnections [61,62]. It has been frequently applied in the geosciences [17-23,25,27,29,31]. An important part of a multilayer perceptron is the choice of activation function, which defines the output from a neuron. We use the 'relu' activation function, which is a piecewise linear function [63], along with four and five hidden layers to predict V_p and q_b respectively as it provides the best fit.

In order to implement the machine learning algorithms, we use only well log data sampled at 0.5 ft (0.1524 m) depth intervals. We also normalize the inputs to range from 0 to 1 [64]. This ensures that each variable is contributing equally to the model. Normalization is particularly important for algorithms that use distance-based attributes to improve accuracy and reduce error [65]. We normalize the inputs when using all the above algorithms except for Random Forest because it does not depend on distance-based attributes.

3.2. Prediction of q_b and V_p

We predict q_b and V_p for Holes WR313-G and WR313-H using the six machine learning algorithms by creating a training dataset from the 20 holes with the available LWD logs from the Gulf of Mexico, Cascadia Margin and Bay of Bengal (Table 2). As a part of the well log quality control for the training dataset we eliminate washout zones > 5 m thick where borehole diameter ≥ 5 cm more than the bit size to remove intervals with poor data. We keep thinner washout intervals because the machine learning model needs to be trained on some poor-quality data along with good quality data to avoid overfitting.

For all algorithms, the training dataset consists of 34,341 sets of data at discrete depths with 30,478 data points corresponding to water saturated intervals, 2938 data points corresponding to intervals with gas hydrate in near-vertical fractures and 925 data points corresponding to intervals with gas hydrate in the primary pore space. Each well log in the training dataset has 34,341 sets of data at discrete depths, or 34341 values of gamma ray, q_b , ring resistivity, propagation resistivity and V_p . We split the training dataset and use 70% for training the model and 30% for validation (Figure 2). The validation dataset is kept separate from the training dataset to observe if the model is consistent enough in making predictions. We also perform feature selection analysis to select the best combination of input well logs to predict V_p and q_b using both the workflows (the statistics are shown in Section 5 of Supplementary Material).

We predict V_p and q_b using different sets of well logs as inputs and describe each of these sets as a Case. We use two different workflows to predict V_p ; Case 1 and 2, where we use bulk density and gamma ray logs, but different resistivity logs. For Case 1, ring resistivity is the only resistivity dataset used as an input. For Case 2, we use propagation resistivities (A16L, A40L, P16H, P28H, P40H) instead of ring resistivity along with gamma ray and bulk density. To predict q_b Case 1, we use gamma ray, ring resistivity and V_p as input well logs and for q_b Case 2 we use gamma ray, propagation resistivities and V_p as input well logs.

3.3. Downsampling the predicted results

V_p and q_b logs have a lower vertical resolution than the other logs. For example, V_p has a vertical resolution of ~61 cm [66] and q_b has a vertical resolution of ~30 cm [67] while ring resistivity has a resolution of ~5–7 cm [68] and gamma ray has a vertical resolution of ~31 cm [67]. The vertical resolution of the propagation resistivity logs ranges from ~21 cm to ~121 cm [67]. Therefore, we downsample the predicted outputs using a moving average filter while estimating V_p for Case 1 and 2 and q_b for Case 1 only.

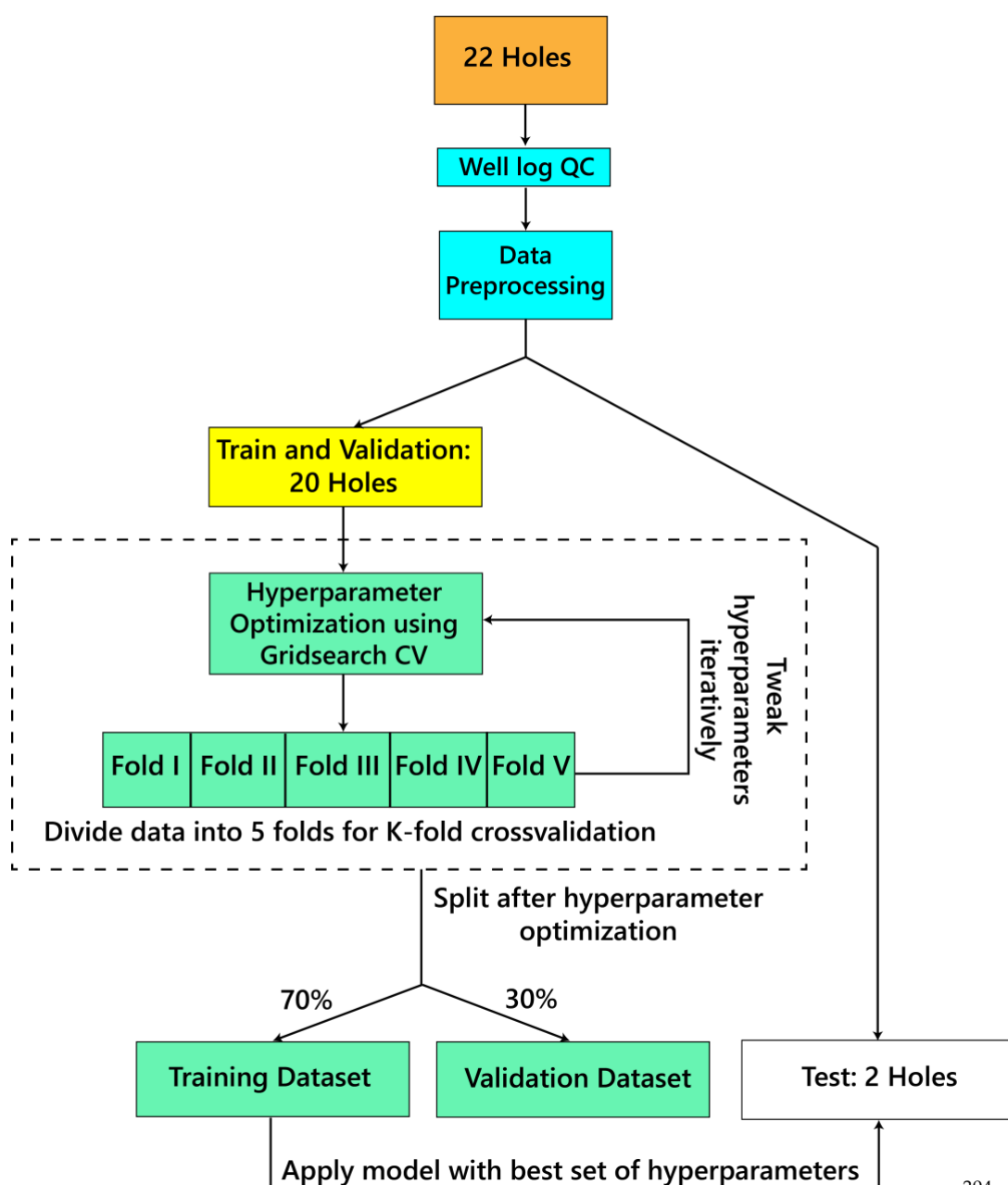


Figure 2. The workflow used for the data and the machine learning models in this study.

4. Results and Discussion

Our study is the first to use centimeter scale resolution LWD data to predict V_p and q_b logs in near seafloor sediments. Out of the six algorithms, we find that Random Forest and K Nearest Neighbors are more robust and can predict V_p and q_b logs with a high accuracy (R^2) greater than 70% and low error (MAPE) less than 4% on training, validation, test data (Figure 3 and Table 3). In addition, Random Forest and K Nearest Neighbors have consistently high accuracy for k-fold cross validation across different folds (Supplementary Material). Random Forest has been used across the geosciences to tackle a variety of different problems [30, 69, 70], however our study shows that K Nearest Neighbors is a strong machine learning method and may be viable for other geoscience applications..

Multilinear Regression and Multilayer Perceptron have also been used in geoscience studies [27, 31, 71] but do not perform as well herein as Random Forest and K Nearest Neighbors to predict V_p and q_b logs. Multilinear Regression has an accuracy of only ~30–60% and a higher error of 4–6% for training, validation and test data (Figure 3 and Table 3). This low accuracy signifies that the relationship among different well logs is not linear; this is an important point, because missing log data is commonly approximated using linear equations. Similarly, Multilayer Perceptron has overall low accuracy varying from 55–59% on training, validation, and test data.

Polynomial Regression and Polynomial Regression with ridge regularization have extremely poor accuracy in the main hydrate bearing sands in WR313-G and WR313-H (Figure 3 and Table 3). Moreover, Polynomial Regression and Polynomial Regression with ridge regularization perform poorly on different folds while performing k-fold cross validation (Supplementary Material).

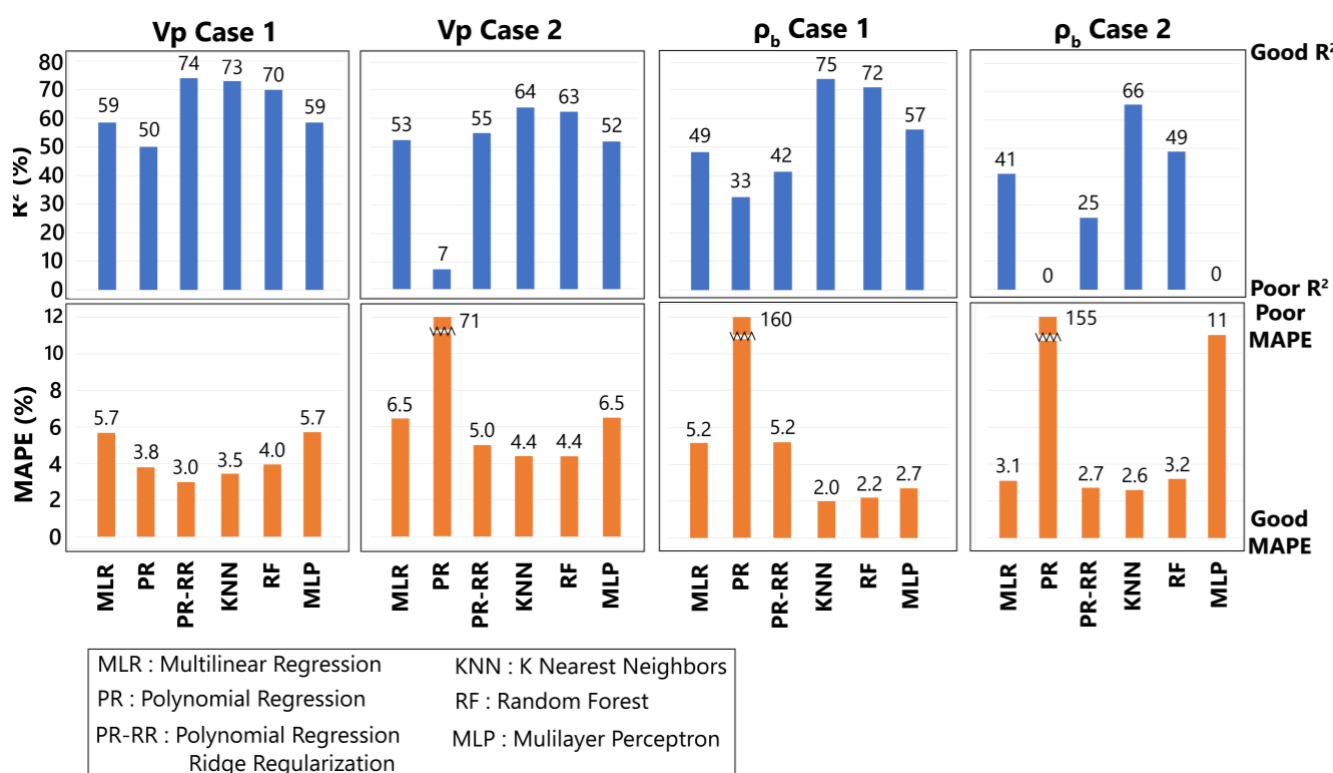


Figure 3. R^2 accuracy and mean absolute percentage error (MAPE) for V_p and q_b prediction. Averaged over the two Walker Ridge Holes WR313-G and WR313-H.

	Training R ² (%)	Training MAPE (%)	Validation R ² (%)	Validation MAPE (%)	Test R ² (%)	Test MAPE (%)
V _p Case 1	56	4.18	55	4.18	59	5.69
V _p Case 2	62	3.6	64	3.46	53	6.45
ρ _b Case 1	31	5.18	31	5.12	49	5.17
ρ _b Case 2	46	4.13	48	4.13	41	3.14
Polynomial Regression (4 th order)						
	Training R ² (%)	Training MAPE (%)	Validation R ² (%)	Validation MAPE (%)	Test R ² (%)	Test MAPE (%)
V _p Case 1	91	2.46	90	2.48	50	3.79
V _p Case 2	88	2.02	0.015	4.83	7.0	71.4
ρ _b Case 1	62	3.54	60	3.49	33	160
ρ _b Case 2	82	2.33	0	11	0	155
Polynomial Regression (4 th order) with Ridge Regularization						
	Training R ² (%)	Training MAPE (%)	Validation R ² (%)	Validation MAPE (%)	Test R ² (%)	Test MAPE (%)
V _p Case 1	85	2.69	83	2.7	74	2.99
V _p Case 2	82	2.42	81	2.34	55	4.99
ρ _b Case 1	57	3.86	57	3.8	42	5.16
ρ _b Case 2	75	2.78	75	2.81	25	2.70
K Nearest Neighbors						
	Training R ² (%)	Training MAPE (%)	Validation R ² (%)	Validation MAPE (%)	Test R ² (%)	Test MAPE (%)
V _p Case 1	100	0	94	1.98	73	3.45
V _p Case 2	100	0	86	1.79	64	4.4
ρ _b Case 1	100	0	76	2.55	75	2.00
ρ _b Case 2	100	0	85	1.86	66	2.65
Random Forest						
	Training R ² (%)	Training MAPE (%)	Validation R ² (%)	Validation MAPE (%)	Test R ² (%)	Test MAPE (%)
V _p Case 1	99	1.07	96	1.60	70	3.96
V _p Case 2	97	1.05	91	1.60	63	4.40
ρ _b Case 1	93	1.51	81	2.30	72	2.19
ρ _b Case 2	95	1.16	89	1.71	49	3.18
Multilayer Perceptron						
	Training R ² (%)	Training MAPE (%)	Validation R ² (%)	Validation MAPE (%)	Test R ² (%)	Test MAPE (%)
V _p Case 1	56	4.20	55	4.19	59	5.66
V _p Case 2	62	3.59	63	3.45	52	6.53
ρ _b Case 1	46	4.47	45	4.40	57	2.70
ρ _b Case 2	0	6.37	0	6.38	0.1	11

Table 3. Train/Validation accuracy and error metrics computed over the 20 training holes with a 70:30 split over training data. Test metrics is computed for the two Walker Ridge Holes WR313-G and WR313-H (taking average R^2 and MAPE over the two holes).

4.1. Formation V_p Prediction

Random Forest and K Nearest Neighbors have high R^2 , low MAPE and are more consistent than the other algorithms, therefore we compare these two algorithms and focus on how these results vary in water saturated intervals, hydrate in the primary pore space and hydrate in fractures (Table 4). A unique aspect of our study is that we consider hydrate in different morphologies and the effect on machine learning results.

a. Water Saturated Intervals

In Figures 4 and 5 water saturated intervals are primarily identified by low resistivity and are signified by a white background. In these water saturated intervals the predicted V_p closely matches the measured V_p with a low percent error (Figure 6) for both algorithms using V_p Case 1 ($R^2 \approx 75\%$). However, the R^2 for predicted V_p for V_p Case 2 is 66% (MAPE 4.7%) for Random Forest and 70% (MAPE 4.0%) for K Nearest Neighbors. This indicates that either Random Forest or K Nearest Neighbors can be used for estimating V_p in water saturated intervals with ring resistivity as one of the inputs in the training model. However, the propagation resistivity can also be used to predict V_p in water saturated sediments if ring resistivity is not available (Case 2). The high accuracy and low percent error for these results may suggest that these models could be applied in datasets in near-seafloor water saturated sediments to predict accurate V_p where high quality input logs are available.

b. Hydrate in Fractures

We compare the predicted V_p results with the measured V_p for WR313-G and WR313-H in the intervals where hydrate is identified in near-vertical fractures. Intervals where hydrate occurs in near-vertical fractures are highlighted in yellow in Figures 3 and 4. Propagation resistivity measurements are the most sensitive to resistivity anisotropy caused by near-vertical hydrate filled fractures; near-vertical resistivity fractures cause a characteristic curve separation in propagation resistivity curves that depends on the fracture angle, hydrate resistivity, the measurement type and the spacing of the measurement sondes [11]. In general, no significant increase in V_p is observed in near-vertical fracture intervals, which is likely due to the low concentration of hydrate in the bulk sediment [8]. The Random Forest V_p prediction results have a low accuracy and higher percent error (Figure 6) using Case 1 (using ring resistivity) but a higher accuracy and lower percent error (Figure 6) with Case 2 (using propagation resistivity: A16L, A40L, P16H, P28H, P40H) (Table 4); this is consistent with the observation that a set of propagation resistivity logs are sensitive to near vertical fractures while a single resistivity measurement (in this case, ring resistivity) cannot be used to identify near-vertical gas hydrate filled fractures. However, the accuracy of the K Nearest Neighbors algorithm is lower for Case 2 ($R^2=48\%$ and MAPE=4.2%) as compared to Case 1 ($R^2=73\%$ and MAPE=2.4%). These contradictory results may be due to the fact that gas hydrate filled fractures form complex 3D networks [46] with a variety of fracture angles [8] and the anisotropy caused by these networks may result in data which is difficult to fit by a machine learning model.

This suggests that some caution is required while predicting V_p when hydrates occur in near-vertical fractures. Thus, in order to predict V_p for hydrates in near-vertical fractures, the Random Forest algorithm with propagation resistivities (Case 2) or the K Nearest Neighbors algorithm with ring resistivity (Case 1) are the best algorithms and datasets.

c. Hydrate in Pores

Hydrate bearing sands are highlighted in blue on Figures 4 and 5. These intervals have a significant increase in the measured V_p log and a corresponding increase in the resistivity logs. In hydrate bearing sands (Figures 4 and 5) the Random Forest algorithm closely replicates the measured V_p log using Case 1 ($R^2 = 81\%$ and $MAPE = 6.5\%$), and we recommend this algorithm over K Nearest Neighbors ($R^2 = 71\%$ and $MAPE = 10\%$) in locations with high saturation gas hydrate. This is because Random Forest predictions better match the measured V_p log both in thick sand accumulations and thin sands (< 5 m in thickness) as compared to K Nearest Neighbors. In addition, a higher accuracy is observed when the ring resistivity log (Case 1) is used over a suite of propagation resistivity logs (Case 2). This is likely due to a better vertical resolution of ring resistivity (5-7 cm) from geoVISION* tool [68] as compared to the propagation resistivities (~21-121 cm resolution) from EcoScope* tool [67]. Therefore, the ring resistivity measurement is able to resolve thinner beds and improves the accuracy of V_p prediction using Case 1.

Of course, high saturation gas hydrate is not a common occurrence. Even so, data in these intervals may still benefit from prediction algorithms. For example, [72] observed that the presence of high-saturation hydrate in pores can cause loss of signal while acquiring V_p logs in boreholes. This may make it difficult to interpret formation V_p logs due to poor data quality. Our prediction results for hydrate in pores may improve the interpretation for V_p logs in such cases where V_p data is compromised due to loss of signal.

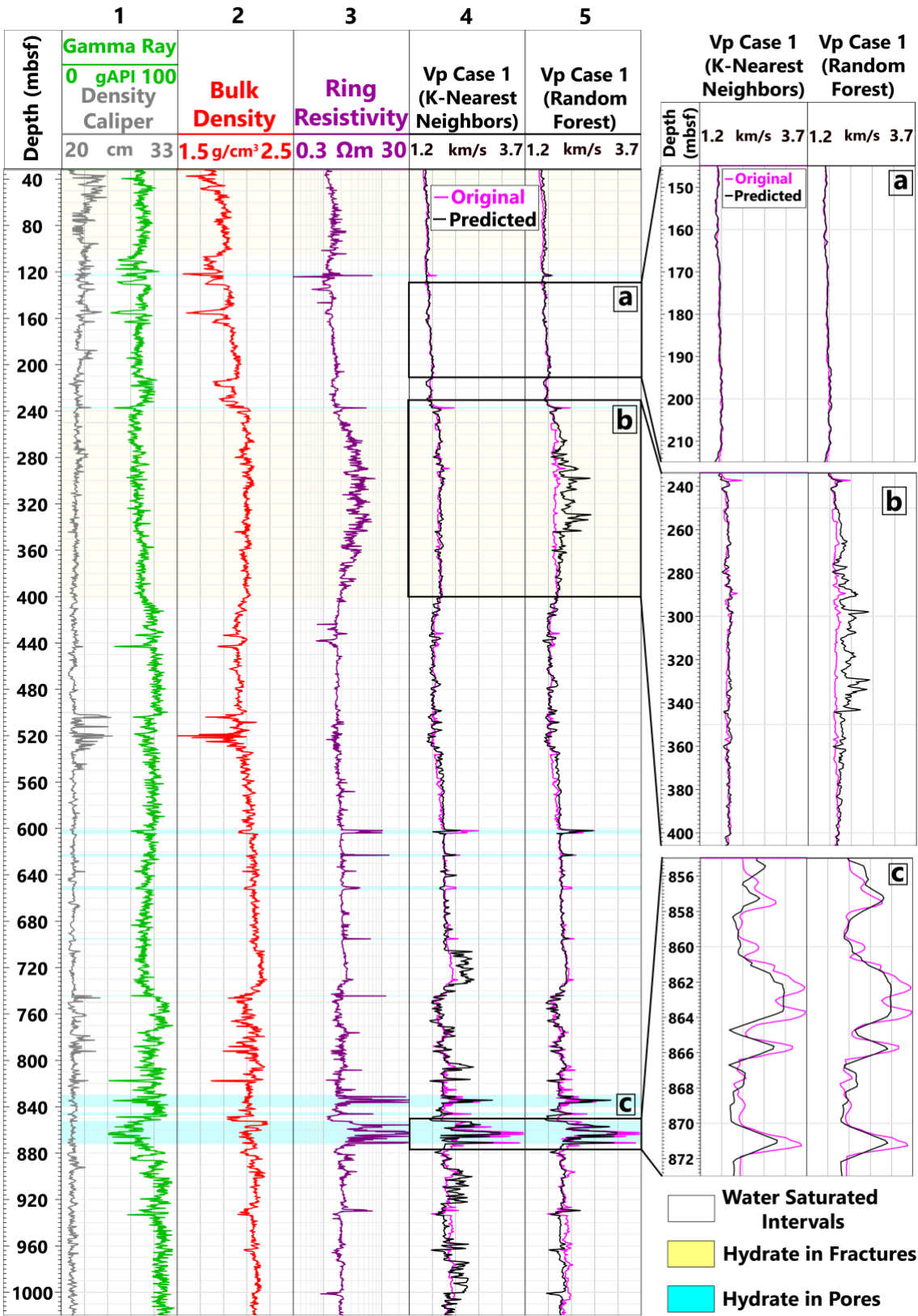


Figure 4. LWD data from 31 – 1043 mbsf (m below sea floor) in Hole WR313-G showing the original and predicted results from K Nearest Neighbors (Track 4) and Random Forest (Track 5) for V_p Case 1. Insets show (a) water saturated interval (b) interval with hydrates in fractures and (c) interval with hydrate in pore space.

409
410
411
412
413

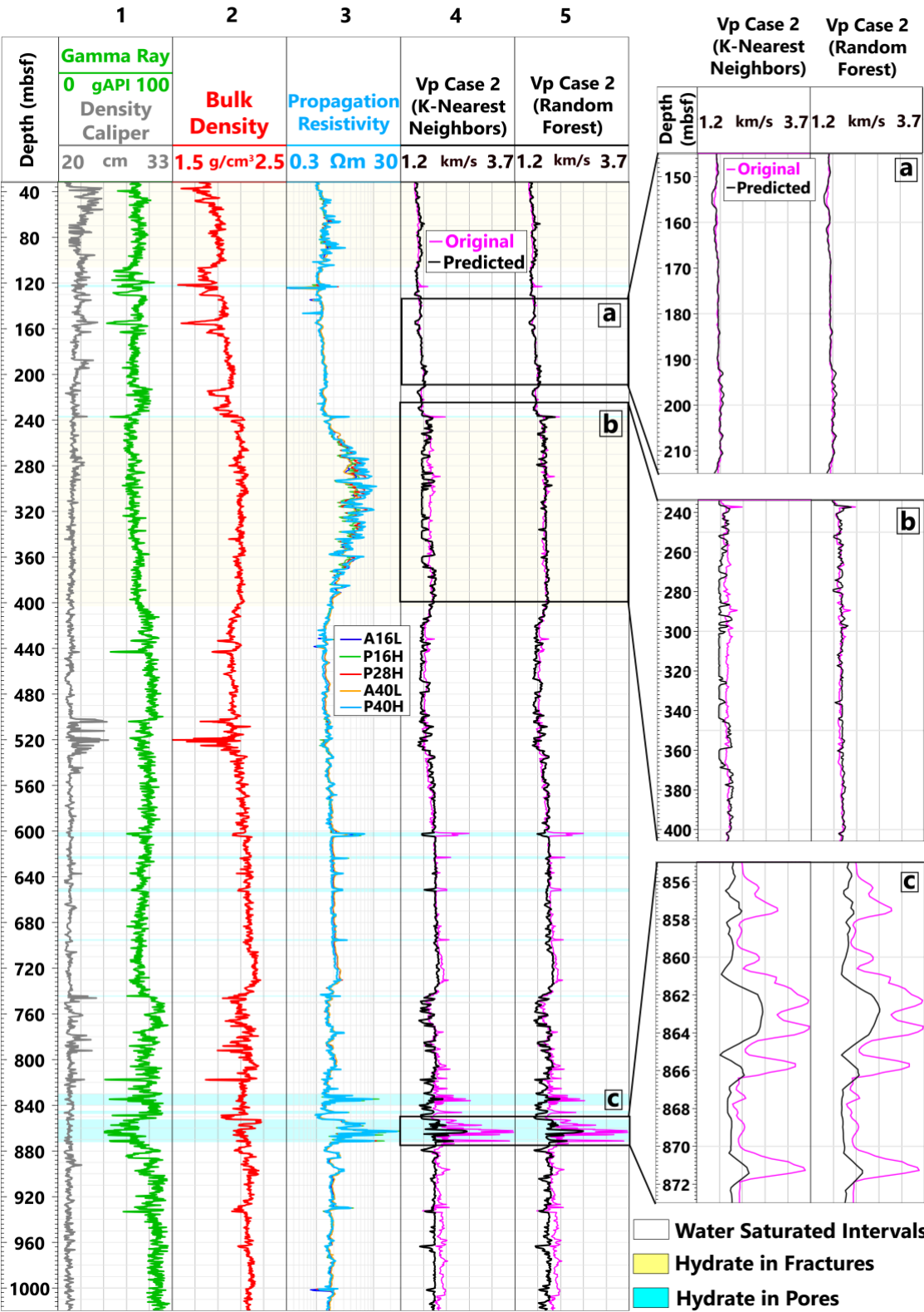


Figure 5. LWD data from 31 – 1043 mbsf (m below sea floor) in Hole WR313-G showing the original and predicted results from K Nearest Neighbors (Track 4) and Random Forest (Track 5) for V_p Case 2. Insets show (a) water saturated interval (b) interval with hydrates in fractures and (c) interval with hydrate in pore space.

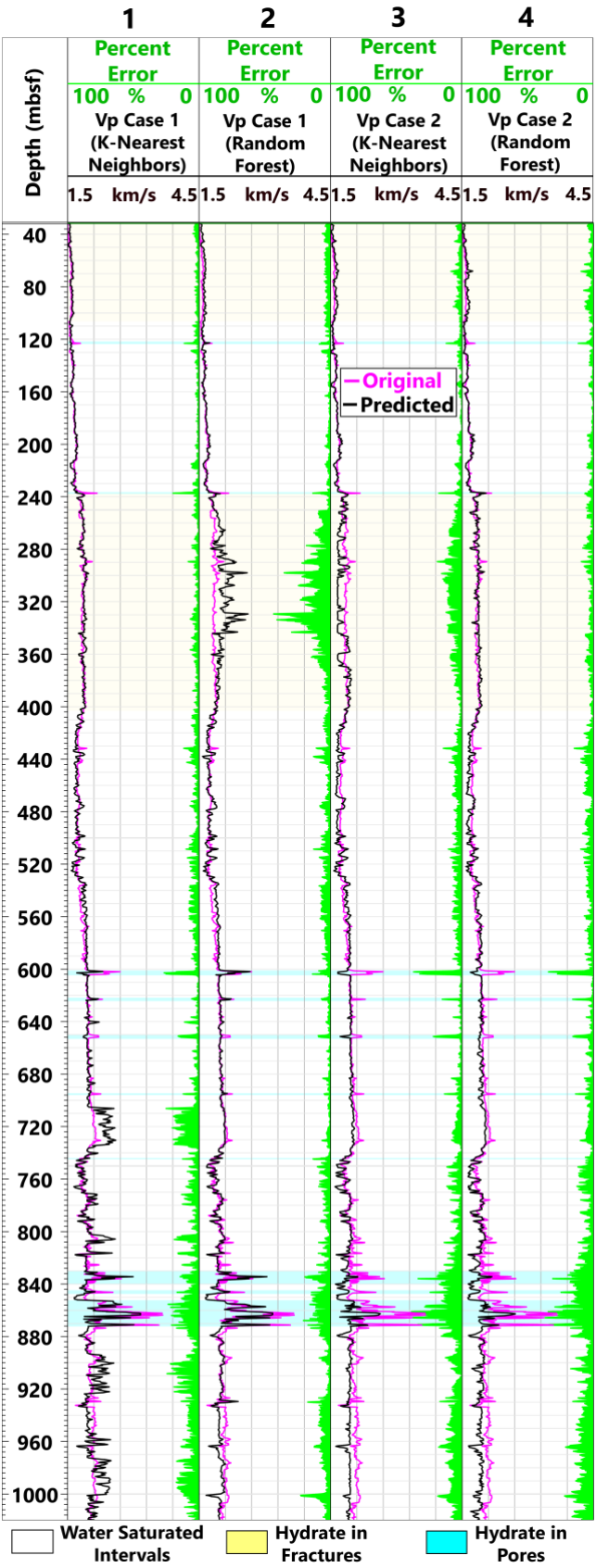


Figure 6. LWD data from 31 – 1043 mbsf (m below sea floor) in Hole WR313-G showing the original and predicted results for V_p Case 1 and 2 using K Nearest Neighbors and Random Forest along with the percentage error for different depth intervals.

		Random Forest		K Nearest Neighbors	
		R ²	MAPE	R ²	MAPE
V _p Case 1 (Input logs: Gamma Ray, Bulk Density, Ring Resistivity)	Complete log interval	70%	3.9%	73%	3.4%
	Water Saturated	74%	3.0%	75%	3.6%
	Hydrate in Fractures	54%	5.9%	73%	2.4%
	Hydrate in Pores	81%	6.5%	71%	10%
V _p Case 2 (Input logs: Gamma Ray, Bulk Density, Propagation Resistivity)	Complete log interval	63%	4.4%	64%	4.4%
	Water Saturated	66%	4.7%	70%	4.0%
	Hydrate in Fractures	68%	2.8%	48%	4.2%
	Hydrate in Pores	69%	14%	63%	15%
ρ _b Case 1	Complete log interval	72%	2.2%	75%	2.0%

Table 4. Statistical analysis for V_p and ρ_b prediction averaged over Holes WR313-G and WR313-H for water saturated sediments, gas hydrate in near vertical fractures and gas hydrates in the primary pore space. (MAPE = mean absolute percentage error).

4.2. Bulk Density Prediction

We predict ρ_b log with high accuracy and low error using Random Forest and K Nearest Neighbors algorithms. We choose Case 1 (with gamma ray, ring resistivity and V_p as inputs) for ρ_b prediction over Case 2 (with gamma ray, propagation resistivities and

V_p as inputs), since q_b Case 2 overfits the model. This is because q_b Case 2 predicts q_b with a high accuracy and low error on training and validation datasets but the prediction becomes poor for the test holes (Table 3). Unlike V_p prediction, we do not assess the different hydrate morphologies for q_b prediction as Case 1 fits all the intervals (Figure 7).

The bulk density measurement is important for hydrate interpretation as it provides the most accurate measurement of porosity in near seafloor sediments. Porosity is used to compute hydrate saturation along with resistivity and V_p . Our bulk density model (Case 1), therefore will be valuable to estimate the bulk density measurement in the near seafloor sediments in locations where bulk density is not collected, such as the Nankai Trough [73] and the Hikurangi Margin [34, 74].

4.3. Prediction at Deeper Depths

We observe that the accuracy decreases and error increases for V_p and q_b prediction at deeper depths (>600 mbsf) in the test dataset (Figures 6 and 7). This is likely because the total drilled depth for the training holes ranges from ~200-600 mbsf, however the total drilled depth for the validation holes is ~1000 mbsf. Both V_p and q_b are a function of depth, i.e., both increase with increasing depth. Therefore, the model can predict V_p and q_b with a higher accuracy and low error for depths where training data is available (<600 mbsf).

4.4 Further Data Limitations

One work is limited by the availability of scientific ocean drilling LWD data. We use all the publicly available data (22 holes) from the Lamont Doherty Earth Observatory database to train, validate and test the model. If more data becomes publicly available in the future, further data can be incorporated to improve the model.

If a user wants to apply our models to new data, the V_p model requires gamma ray, resistivity, q_b logs and the q_b model requires the gamma ray, resistivity and V_p logs, otherwise the model cannot be accurately applied. Moreover, our model is only applicable to siliciclastic near seafloor sediments in marine settings. It cannot be used for permafrost environments or in lithified rock.

The well log data that we use for this project is few tens to few hundreds of megabytes in size and the machine learning algorithms take 30 seconds to 2 minutes for execution. However, the computation time increases to 3-4 hours while performing hyperparameter optimization that compares several hundreds of combinations of different hyperparameters for different algorithms. For Random Forest, the computation time for hyperparameter optimization is higher and takes about 10 hours.

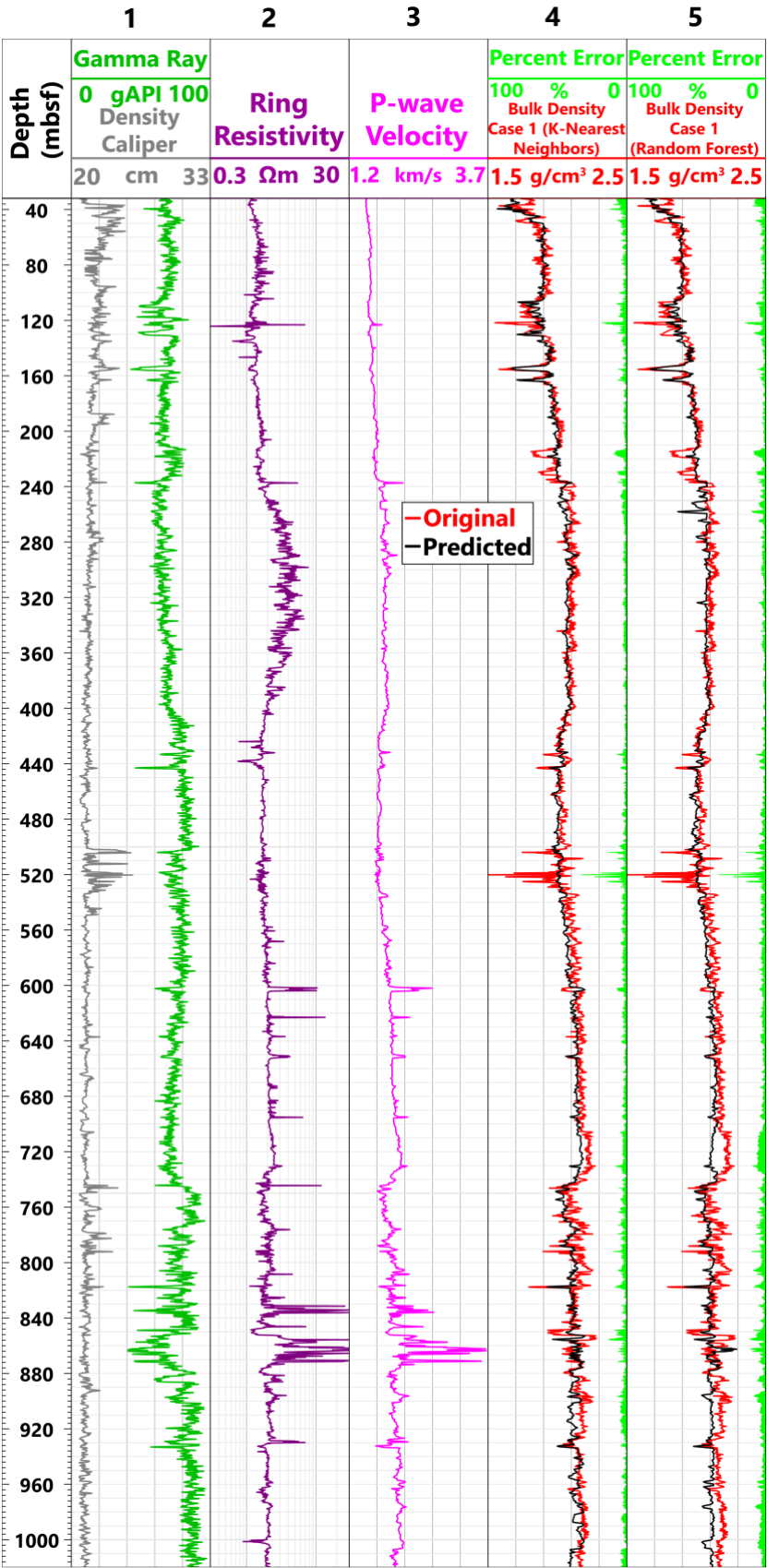


Figure 7. LWD logs showing bulk density Case 1 for Hole WR313-G comparing the results and percentage error associated with different depth intervals for K Nearest Neighbors and Random Forest algorithms.

472
473
474
475

4.5. Neutron Porosity

The neutron porosity log measures the hydrogen concentration in the formation which is related to the porosity of the formation [49]. In clay rich environments, however, the apparent neutron porosity can be larger and noisy due to the presence of hydroxyl ions (OH⁻) associated with clay minerals [49]. Therefore, bulk density is the preferred log in near-seafloor sediments to interpret porosity because it most closely replicates in situ porosity [3]. For this reason, we used bulk density as an input log in Case 1 and 2.

We tested the neutron porosity log as an input for our machine learning model to predict V_p (Figure 8). When applying neutron porosity to the two Walker Ridge Holes (WR313-G and WR313-H), we found that the predicted V_p in clayey zones does not correlate as well as when the bulk density was used (Case 1). In contrast, [27] shows that both neutron porosity and porosity derived from bulk density can be used interchangeably as an input in a machine learning model used to compute hydrate saturation in a permafrost location in Canada (Figure 8). The neutron porosity works in the model of [27] because the lithology is primarily sand whereas we apply our machine learning model to both sand and clay rich intervals. Caution should always be exercised if using neutron porosity in mud or clay rich environments.

4.6. Model application in non-hydrate sites

Even though we train our machine learning model using borehole data from hydrate drilling expeditions in the Gulf of Mexico, Cascadia Margin and Offshore India, we argue our model can still be applied in boreholes missing data not only in hydrate systems but in siliciclastic near seafloor sediments on the continental slopes. While this paper is focused on hydrate systems, most of the data used in the model (89%) is from water saturated marine sediments; in these systems our model can predict V_p and q_b with a high accuracy and a low percent error (Figure 6 and 7).

One factor that might affect the machine learning model is porewater salinity. This is because resistivity is a function of porewater salinity in high porosity sediments. In general, an increase in porewater salinity would reduce resistivity. This would reduce predicted V_p and q_b . Conversely, a decrease in porewater salinity would increase resistivity and also the predicted V_p and q_b . For example, porewater salinity variations can be due to the formation or dissociation of hydrate [10,75]. Porewater salinity can also vary in places with shallow salt diapirs [76]. While these situations where pore water salinity varies are not very common and porewater salinity is normally standard seawater, caution should be taken in any location where there may be a significant change in porewater salinity. There are many holes where this model can be applied in marine sediments on continental margins. For example, our model could be used to predict V_p and q_b logs for the ~70 LWD scientific ocean drilling holes on Lamont Doherty Earth Observatory database with missing V_p or q_b logs. Even more holes have missing or damaged V_p or q_b logs where our model could be applied.

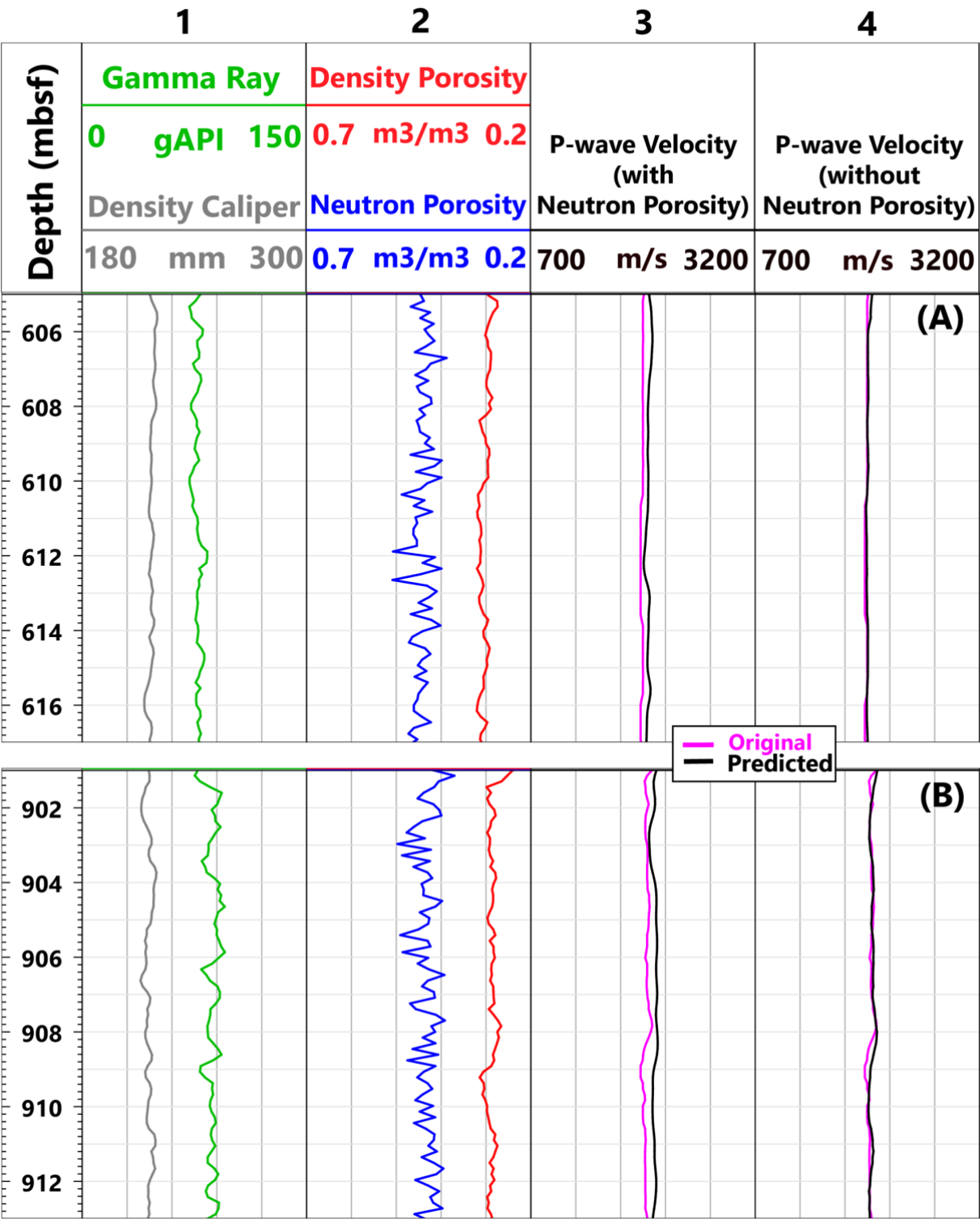


Figure 8. LWD logs for WR313-G showing two different clay rich intervals A and B with V_p prediction results using Random Forest algorithm before and after eliminating Neutron Porosity (Case 1) from the training model.

5. Conclusion

In this work, we present a novel machine learning approach to predict V_p and q_b logs in marine gas hydrates and their variation with different depth intervals and different hydrate morphologies. We predict V_p log using gamma ray, bulk density and resistivity as inputs and q_b log using gamma ray, V_p and resistivity logs as inputs. To identify the best algorithms, we use six machine learning algorithms and compare the results. We find that Random Forest and K Nearest Neighbors algorithms can be used to predict V_p and q_b logs with a high degree of accuracy and low error in near seafloor sediments with water saturated intervals, intervals where hydrate is filling fractures and intervals where hydrate is in the primary pore space. Due to a good match between the measured and predicted logs both hydrate bearing and water saturated intervals, our model can be applied to siliciclastic near seafloor sediments where either V_p or q_b logs are missing. Our model for V_p or q_b prediction is applicable not only to hydrate systems but also useful for researchers working to identify shallow natural hazards such as submarine landslides as well as conducting studies by integrating well and seismic data.

Acknowledgements

The authors would like to thank Debashis Konwar for helping with interpreting borehole sonic data and Schlumberger for providing Techlog software at The Ohio State University. The LWD log data in this paper was downloaded from Lamont-Doherty Earth Observatory (<https://mlp.ldeo.columbia.edu/logdb/>). All machine learning codes are provided on the link: https://colab.rsearch.google.com/drive/10oUw7hIh5aBD3q56Ix8iu0a3mUJTyRWF#scrollTo=mZ_q2FHcxv4K. All the machine learning codes along with the training and testing data files (csv format) are attached with the supplementary material. This work was supported by the US Department of Energy DE-FE0023919; and National Science Foundation 1752882.

We also thank the anonymous reviewers.

Author Contributions

F.N. conceived the main idea for the manuscript and designed the figures. F.N. and A.C. wrote the manuscript. A.C. secured the funding. J.M. reviewed and improved the codes.

Data Availability Statement

Detailed information related to WR313-H log figures, k-fold cross validation, hyperparameter tuning for all the algorithms using gridsearchcv (available as spreadsheets) and distribution of observed data points is available in the Supplementary Material.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof nor any of their employees make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or an agency thereof.

References

1. Collett, Timothy S, Johnson, A. H., Knapp, C. C., & Boswell, R. (2009). Natural gas hydrates: A review. *AAPG Memoir*.
2. Kvenvolden, K. A., & Lorenson, T. D. (2001). The global occurrence of natural gas hydrate. In *Geophysical Monograph Series* (Vol. 124). <https://doi.org/10.1029/GM124p0003>
3. Goldberg, D., Kleinberg, R. L., Weinberger, J. L., Malinverno, A., McLellan, P. J., & Collett, T. S. (2010). 16. Evaluation of Natural Gas-Hydrate Systems Using Borehole Logs. In *Geophysical Characterization of Gas Hydrates*. <https://doi.org/10.1190/1.9781560802197.ch16>
4. Kerkar, P. B., Horvat, K., Mahajan, D., & Jones, K. W. (2013). Formation and dissociation of methane hydrates from seawater in consolidated sand: Mimicking methane hydrate dynamics beneath the seafloor. *Energies*, 6(12). <https://doi.org/10.3390/en6126225>
5. Li, Z. dong, Tian, X., Li, Z., Xu, J. ze, Zhang, H. xiang, & Wang, D. ju. (2020). Experimental study on growth characteristics of pore-scale methane hydrate. *Energy Reports*, 6. <https://doi.org/10.1016/j.egy.2020.04.017>
6. Yun, T. S., Francisca, F. M., Santamarina, J. C., & Ruppel, C. (2005). Compressional and shear wave velocities in uncemented sediment containing gas hydrate. *Geophysical Research Letters*. <https://doi.org/10.1029/2005GL022607>
7. Oti, E. A., Cook, A. E., Phillips, S. C., & Holland, M. E. (2022). Using x-ray computed tomography to estimate hydrate saturation in sediment cores from Green Canyon 955, northern Gulf of Mexico. *AAPG Bulletin*, 106(5). <https://doi.org/10.1306/05272120051>

8. Cook, A. E., Goldberg, D. S., & Malinverno, A. (2014). Natural gas hydrates occupying fractures: A focus on non-vent sites on the Indian continental margin and the northern Gulf of Mexico. In *Marine and Petroleum Geology* (Vol. 58, Issue PA). <https://doi.org/10.1016/j.marpetgeo.2014.04.013>
9. Collett, T. S., & Ladd, J. (2000). Detection of gas hydrate with downhole logs and assessment of gas hydrate concentrations (saturation) and gas volumes on the Blake Ridge with electrical resistivity log data. *Proceedings of the Ocean Drilling Program: Scientific Results*.
<https://doi.org/10.2973/odp.proc.sr.164.219.2000>
10. Malinverno, A., Kastner, M., Torres, M. E., & Wortmann, U. G. (2008). Gas hydrate occurrence from pore water chlorinity and downhole logs in a transect across the northern Cascadia margin (Integrated Ocean Drilling Program Expedition 311). *Journal of Geophysical Research: Solid Earth*, 113(8). <https://doi.org/10.1029/2008JB005702>
11. Cook, A. E., Anderson, B. I., Malinverno, A., Mrozewski, S., & Goldberg, D. S. (2010). Electrical anisotropy due to gas hydrate-filled fractures. *Geophysics*, 75(6). <https://doi.org/10.1190/1.3506530>
12. Helgerud, M. B., Dvorkin, J., Nur, A., Sakai, A., & Collett, T. (1999). Elastic-wave velocity in marine sediments with gas hydrates: Effective medium modeling. *Geophysical Research Letters*.
<https://doi.org/10.1029/1999GL900421>
13. Lee, M. W., & Collett, T. S. (2011). In-situ gas hydrate saturation estimated from various well logs at the Mount Elbert Gas Hydrate Stratigraphic Test Well, Alaska North Slope. *Marine and Petroleum Geology*. <https://doi.org/10.1016/j.marpetgeo.2009.06.007>
14. Nelwamondo, F. V., Golding, D., & Marwala, T. (2013). A dynamic programming approach to missing data estimation using neural networks. *Information Sciences*, 237.

<https://doi.org/10.1016/j.ins.2009.10.008>

15. Pelckmans, K., De Brabanter, J., Suykens, J. A. K., & De Moor, B. (2005). Handling missing values in support vector machine classifiers. *Neural Networks*, 18(5–6).

<https://doi.org/10.1016/j.neunet.2005.06.025>

16. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>

17. Farfour, M., & Mesbah, M. (2020). Machine intelligence vs. human intelligence in geological interpretation of seismic data. *2020 International Conference on Decision Aid Sciences and Application, DASA 2020*. <https://doi.org/10.1109/DASA51403.2020.9317032>

18. Ismail, A., Ewida, H. F., Nazeri, S., Al-Ibiary, M. G., & Zollo, A. (2022). Gas channels and chimneys prediction using artificial neural networks and multi-seismic attributes, offshore West Nile Delta, Egypt. *Journal of Petroleum Science and Engineering*, 208.

<https://doi.org/10.1016/j.petrol.2021.109349>

19. Ramya, J., Somasundareswari, D., & Vijayalakshmi, P. (2020). Gas chimney and hydrocarbon detection using combined BBO and artificial neural network with hybrid seismic attributes. *Soft Computing*, 24(3). <https://doi.org/10.1007/s00500-019-04064-6>

20. Bressan, T. S., Kehl de Souza, M., Girelli, T. J., & Junior, F. C. (2020). Evaluation of machine learning methods for lithology classification using geophysical data. *Computers and Geosciences*, 139. <https://doi.org/10.1016/j.cageo.2020.104475>

21. Ismail, A., Radwan, A. A., Leila, M., Abdelmaksoud, A., & Ali, M. (2023). Unsupervised machine learning and multi-seismic attributes for fault and fracture network interpretation in the Kerry Field,

- Taranaki Basin, New Zealand. *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, 9(1). <https://doi.org/10.1007/s40948-023-00646-9>
22. Hou, M., Xiao, Y., Lei, Z., Yang, Z., Lou, Y., & Liu, Y. (2023). Machine Learning Algorithms for Lithofacies Classification of the Gulong Shale from the Songliao Basin, China. *Energies*, 16(6). <https://doi.org/10.3390/en16062581>
23. Lou, Y., Li, S., Liu, N., & Liu, R. (2022). Seismic volumetric dip estimation via a supervised deep learning model by integrating realistic synthetic data sets. *Journal of Petroleum Science and Engineering*, 218. <https://doi.org/10.1016/j.petrol.2022.111021>
24. Yang, L., & Sun, S. Z. (2020). Seismic horizon tracking using a deep convolutional neural network. *Journal of Petroleum Science and Engineering*, 187. <https://doi.org/10.1016/j.petrol.2019.106709>
25. Shalaby, M. R., Jumat, N., Lai, D., & Malik, O. (2019). Integrated TOC prediction and source rock characterization using machine learning, well logs and geochemical analysis: Case study from the Jurassic source rocks in Shams Field, NW Desert, Egypt. *Journal of Petroleum Science and Engineering*, 176. <https://doi.org/10.1016/j.petrol.2019.01.055>
26. Gjelsvik, E. L., Fossen, M., & Tøndel, K. (2023). Current overview and way forward for the use of machine learning in the field of petroleum gas hydrates. *Fuel*, 334(P2), 126696. <https://doi.org/10.1016/j.fuel.2022.126696>
27. Singh, H., Seol, Y., & Myshakin, E. M. (2021). Prediction of gas hydrate saturation using machine learning and optimal set of well-logs. *Computational Geosciences*. <https://doi.org/10.1007/s10596-020-10004-3>

28. Yu, Z., & Tian, H. (2022). Application of Machine Learning in Predicting Formation Condition of Multi-Gas Hydrate. *Energies*, 15(13). <https://doi.org/10.3390/en15134719>
29. Rebai, N., Hadjadj, A., Benmounah, A., Berrouk, A. S., & Boualleg, S. M. (2019). Prediction of natural gas hydrates formation using a combination of thermodynamic and neural network modeling. *Journal of Petroleum Science and Engineering*, 182. <https://doi.org/10.1016/j.petrol.2019.106270>
30. Graw, J. H., Wood, W. T., & Phrampus, B. J. (2021). Predicting Global Marine Sediment Density Using the Random Forest Regressor Machine Learning Algorithm. *Journal of Geophysical Research: Solid Earth*. <https://doi.org/10.1029/2020JB020135>
31. Sain, K., & Kumar, P. C. (2022). *Meta-attributes and Artificial Networking: A New Tool for Seismic Interpretation*. AGU-John Wiley & Sons. <https://doi.org/10.1002/9781119481874>
32. Dumke, I., & Berndt, C. (2019). Prediction of seismic p-wave velocity using machine learning. *Solid Earth*, 10(6). <https://doi.org/10.5194/se-10-1989-2019>
33. Flemings, P. B., Behrmann, J. H., John, C. M., & Scientists, and the E. 308. (2006). Expedition 308 summary. *Proceedings of the 27ntegrated Ocean Drilling Program*, 308. <https://doi.org/10.2204/iodp.proc.308.101.2006>
34. Pecher, I. A., Barnes, P. M., LeVay, L. J., & and the Expedition 372 Scientists. (2019). Creeping Gas Hydrate Slides. *Proceedings of the International Ocean Discovery Program*, 372. <https://doi.org/10.14379/iodp.proc.372A.2019>
35. Zoback, M. D. (2007). Reservoir Geomechanics. In *Reservoir Geomechanics*. <https://doi.org/10.1017/CBO9780511586477>

36. Liner, C. L. (Ed.). (2016). Synthetic Seismogram, Tuning, and Resolution. In *Elements of 3D Seismology* (Vol. 19, p. 0). Society of Exploration Geophysicists.
<https://doi.org/https://doi.org/10.1190/1.9781560803386>
37. Collett, Timothy S., Lee, M. W., Zyrianova, M. V., Mrozewski, S. A., Guerin, G., Cook, A. E., & Goldberg, D. S. (2012). Gulf of Mexico Gas Hydrate Joint Industry Project Leg II logging-while-drilling data acquisition and analysis. *Marine and Petroleum Geology*.
<https://doi.org/10.1016/j.marpetgeo.2011.08.003>
38. Riedel, M., Collett, T. S., Malone, M. J., Mitchell, M., Guérin, G., Akiba, F., Blanc-Valleron, M. M., Ellis, M., Hashimoto, Y., Heuer, V., Higashi, Y., Holland, M., Jackson, P. D., Kaneko, M., Kastner, M., Kim, J. H., Kitajima, H., Long, P. E., Malinverno, A., ... Yoshioka, H. (2009). Gas hydrate drilling transect across northern Cascadia margin – IODP Expedition 311. *Geological Society Special Publication*, 319. <https://doi.org/10.1144/SP319.2>
39. Collett, T.S., Boswell, R., Cochran, J. R., Kumar, P., Lall, M., Mazumdar, A., Ramana, M. V., Ramprasad, T., Riedel, M., Sain, K., Sathe, A. V., & Vishwanath, K. (2014). Geologic implications of gas hydrates in the offshore of India: Results of the National Gas Hydrate Program Expedition 01. *Marine and Petroleum Geology*, 58(PA). <https://doi.org/10.1016/j.marpetgeo.2014.07.021>
40. Kevin Meazell, P., Flemings, P. B., Santra, M., & Johnson, J. E. (2020). Sedimentology and stratigraphy of a deep-water gas hydrate reservoir in the northern Gulf of Mexico. *AAPG Bulletin*, 104(9). <https://doi.org/10.1306/05212019027>

41. Santra, M., Flemings, P. B., Scott, E., & Kevin Meazell, P. (2020). Evolution of gas hydrate-bearing deep-water channel-levee system in abyssal Gulf of Mexico: Levee growth and deformation. *AAPG Bulletin*, 104(9). <https://doi.org/10.1306/04251918177>
42. Flemings, P. B., Phillips, S. C., Boswell, R., Collett, T. S., Cook, A. E., Dong, T., Frye, M., Goldberg, D. S., Guerin, G., Holland, M. E., Jang, J., Meazell, K., Morrison, J., O'Connell, J. I., Petrou, E. G., Pettigrew, T., Polito, P. J., Portnov, A., Santra, M., ... You, K. (2020). Pressure coring a Gulf of Mexico deep-water turbidite gas hydrate reservoir: Initial results from the University of Texas-Gulf of Mexico 2-1 (UT-GOM2-1) Hydrate Pressure Coring Expedition. *AAPG Bulletin*, 104(9). <https://doi.org/10.1306/05212019052>
43. Cook, A. E., & Tost, B. C. (2014). Geophysical signatures for low porosity can mimic natural gas hydrate: An example from Alaminos Canyon, Gulf of Mexico. *Journal of Geophysical Research: Solid Earth*, 119(10). <https://doi.org/10.1002/2014JB011342>
44. Frye, M., Shedd, W. W., Godfriaux, P. D., Dufrene, R. S., Collett, T. S., Lee, M. W., Boswell, R., Jones, E., McConnell, D. R., Mrozewski, S., Guerin, G., & Cook, A. E. (2010). Gulf of Mexico gas hydrate joint industry project leg II: Results from the Alaminos Canyon 21 Site. *Proceedings Of The Annual Offshore Technology Conference*, 2.
45. Expedition 311 summary. (2006). In *Proceedings of the IODP, 311*. <https://doi.org/10.2204/iodp.proc.311.101.2006>

46. Rees, E. V. L., Priest, J. A., & Clayton, C. R. I. (2011). The structure of methane gas hydrate bearing
sediments from the Krishna-Godavari Basin as seen from Micro-CT scanning. *Marine and
Petroleum Geology*, 28(7). <https://doi.org/10.1016/j.marpetgeo.2011.03.015>
47. Frye, M., Shedd, W., & Boswell, R. (2012). Gas hydrate resource potential in the Terrebonne Basin,
Northern Gulf of Mexico. *Marine and Petroleum Geology*, 34(1).
<https://doi.org/10.1016/j.marpetgeo.2011.08.001>
48. Hillman, J. I. T., Cook, A. E., Daigle, H., Nole, M., Malinverno, A., Meazell, K., & Flemings, P. B.
(2017). Gas hydrate reservoirs and gas migration mechanisms in the Terrebonne Basin, Gulf of
Mexico. *Marine and Petroleum Geology*, 86. <https://doi.org/10.1016/j.marpetgeo.2017.07.029>
49. Ellis, D. V., & Singer, J. M. (2007). Well Logging for Earth Scientists. In *Well Logging for Earth
Scientists*. <https://doi.org/10.1007/978-1-4020-4602-5>
50. Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter
optimization for machine learning models based on Bayesian optimization. *Journal of Electronic
Science and Technology*, 17(1), 26–40. <https://doi.org/10.11989/JEST.1674-862X.80904120>
51. Kutty, A. A., Wakjira, T. G., Kucukvar, M., Abdella, G. M., & Onat, N. C. (2022). Urban resilience
and livability performance of European smart cities: A novel machine learning approach. *Journal of
Cleaner Production*, 378(November 2021), 134203. <https://doi.org/10.1016/j.jclepro.2022.134203>
52. Wakjira, T. G., Ibrahim, M., Ebead, U., & Alam, M. S. (2022). Explainable machine learning model
and reliability analysis for flexural capacity prediction of RC beams strengthened in flexure with

- FRCM. *Engineering Structures*, 255(December 2021), 113903. 741
<https://doi.org/10.1016/j.engstruct.2022.113903> 742
53. Su, Y., Gao, X., Li, X., & Tao, D. (2012). Multivariate multilinear regression. *IEEE Transactions on* 743
Systems, Man, and Cybernetics, Part B: Cybernetics, 42(6). 744
<https://doi.org/10.1109/TSMCB.2012.2195171> 745
54. Ostertagová, E. (2012). Modelling using polynomial regression. *Procedia Engineering*, 48. 746
<https://doi.org/10.1016/j.proeng.2012.09.545> 747
55. Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning Second 748
Edition. *Math. Intell.*, 27(2), 83–85. [https://doi.org/link.springer.com/book/10.1007/978-0-387-](https://doi.org/link.springer.com/book/10.1007/978-0-387-84858-7) 749
[84858-7](https://doi.org/link.springer.com/book/10.1007/978-0-387-84858-7) 750
56. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. In *Data Mining:* 751
Concepts and Techniques. <https://doi.org/10.1016/C2009-0-61819-5> 752
57. Wu, X., Kumar, V., Ross, Q. J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., 753
Yu, P. S., Zhou, Z. H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data 754
mining. *Knowledge and Information Systems*, 14(1). <https://doi.org/10.1007/s10115-007-0114-2> 755
58. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. 756
<https://doi.org/10.1023/A:1010933404324> 757
59. Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, 13, 758
1063–1095. <https://doi.org/10.5555/2188385.2343682> 759
60. Liu, Z., & Liu, J. (1998). Seismic-controlled nonlinear extrapolation of well parameters using neural 760
networks. *Geophysics*, 63(6). <https://doi.org/10.1190/1.1444496> 761

61. McCormack, M. D. (1991). Neural computing in geophysics. *The Leading Edge*, 10(1). 762
<https://doi.org/10.1190/1.1436771> 763
62. McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. 764
The Bulletin of Mathematical Biophysics, 5(4). <https://doi.org/10.1007/BF02478259> 765
63. Nielsen, M. A. (2014). Neural Networks and Deep Learning. *Determination Press*. 766
64. García, S., Luengo, J., & Herrera, F. (2015). Data Preprocessing in Data Mining. *Intelligent Systems* 767
Reference Library, 72. <https://doi.org/10.1007/978-3-319-10247-4> 768
65. Aksoy, S., & Haralick, R. M. (2001). Feature normalization and likelihood-based similarity measures 769
for image retrieval. *Pattern Recognition Letters*, 22(5). [https://doi.org/10.1016/S0167-](https://doi.org/10.1016/S0167-8655(00)00112-4) 770
[8655\(00\)00112-4](https://doi.org/10.1016/S0167-8655(00)00112-4) 771
66. Schlumberger. (2010). *sonicVISION: Real-time LWD sonic for advanced drilling optimization and* 772
formation evaluation. <http://www.slb.com> 773
67. Schlumberger. (2015). *EcoScope Log Quality Control Reference Manual*. <http://www.slb.com> 774
68. Schlumberger. (2007). *geoVISION Brochure: Resistivity imaging for productive drilling*. 775
<http://www.slb.com> 776
69. Nguyen, H., Savary-Sismondin, B., Patacz, V., Jenssen, A., Kifle, R., & Bertrand, A. (2022). 777
Application of random forest algorithm to predict lithofacies from well and seismic data in Balder 778
field, Norwegian North Sea. *AAPG Bulletin*, 106(11). <https://doi.org/10.1306/08142221032> 779
70. Zou, C., Zhao, L., Xu, M., Chen, Y., & Geng, J. (2021). Porosity Prediction With Uncertainty 780
Quantification From Multiple Seismic Attributes Using Random Forest. *Journal of Geophysical* 781
Research: Solid Earth, 126(7). <https://doi.org/10.1029/2021JB021826> 782

71. Lorenzen, R. (2018). Multivariate linear regression of sonic logs on petrophysical logs for detailed reservoir characterization in producing fields. *Interpretation*, 6(3). <https://doi.org/10.1190/INT-2018-0030.1>
72. Guerin, G., & Goldberg, D. (2002). Sonic waveform attenuation in gas hydrate-bearing sediments from the Mallik 2L-38 research well, Mackenzie Delta, Canada. *Journal of Geophysical Research*. <https://doi.org/10.1029/2001jb000556>
73. Tobin, H., Hirose, T., Ikari, M., Kanagawa, K., Kimura, G., Kinoshita, M., Kitajima, H., Saffer, D., Yamaguchi, A., Eguchi, N., Maeda, L., Toczko, S., Bedford, J., Chiyonobu, S., Colson, T. A., Conin, M., Cornard, P. H., Dielforder, A., Doan, M. L., ... Saito, S. (2020). *Expedition 358 summary*. <https://doi.org/10.14379/iodp.proc.358.101.2020>
74. Saffer, D. M., Wallace, L. M., Barnes, P. M., Pecher, I. A., Petronotis, K. E., LeVay, L. J., Bell, R. E., Crundwell, M. P., Engelmann de Oliveira, C. H., Fagereng, A., Fulton, P. M., Greve, A., Harris, R. N., Hashimoto, Y., Hüpers, A., Ikari, M. J., Ito, Y., Kitajima, H., Kutterolf, S., ... Wu, H.-Y. (2019). *Expedition 372B/375 summary*. <https://doi.org/10.14379/iodp.proc.372b375.101.2019>
75. Almenningen, S., Iden, E., Fernø, M. A., & Ersland, G. (2018). Salinity Effects on Pore-Scale Methane Gas Hydrate Dissociation. *Journal of Geophysical Research: Solid Earth*, 123(7). <https://doi.org/10.1029/2017JB015345>
76. Hanor, J. S., & Mercer, J. A. (2010). Spatial variations in the salinity of pore waters in northern deep water Gulf of Mexico sediments: Implications for pathways and mechanisms of solute transport. *Geofluids*, 10(1–2). <https://doi.org/10.1111/j.1468-8123.2009.00271.x>