



Value of Geologically Derived Features in Machine Learning Facies Classification

Julie Halotel^{1,2}  · Vasily Demyanov¹ ·
Andy Gardiner¹

Received: 12 November 2018 / Accepted: 1 November 2019 / Published online: 16 November 2019
© The Author(s) 2019

Abstract The aim of this work is to demonstrate how geologically interpretative features can improve machine learning facies classification with uncertainty assessment. Manual interpretation of lithofacies from wireline log data is traditionally performed by an expert, can be subject to biases, and is substantially laborious and time consuming for very large datasets. Characterizing the interpretational uncertainty in facies classification is quite difficult, but it can be very important for reservoir development decisions. Thus, automation of the facies classification process using machine learning is a potentially intuitive and efficient way to facilitate facies interpretation based on large-volume data. It can also enable more adequate quantification of the uncertainty in facies classification by ensuring that possible alternative lithological scenarios are not overlooked. An improvement of the performance of purely data-driven classifiers by integrating geological features and expert knowledge as additional inputs is proposed herein, with the aim of equipping the classifier with more geological insight and gaining interpretability by making it more explanatory. Support vector machine and random forest classifiers are compared to demonstrate the superiority of the latter. This study contrasts facies classification using only conventional wireline log inputs and using additional geological features. In the first experiment, geological rule-based constraints were implemented as an additional derived and constructed input. These account for key geological features that a petrophysics or geological expert would attribute to typical and identifiable wireline log responses. In the second experiment, geological interpretative features (i.e., grain size, pore size, and argillaceous content) were used as additional independent inputs to enhance the prediction accuracy and

✉ Julie Halotel
julie.halotel@gmail.com

¹ Heriot-Watt University School of Energy Geoscience Infrastructure and Society, Edinburgh, UK

² Present Address: Equinor, Oslo, Norway

geological consistency of the classification outcomes. Input and output noise injection experiments demonstrated the robustness of the results towards systematic and random noise in the data. The aspiration of this study is to establish geological characteristics and knowledge to be considered as decisive data when used in machine learning facies classification.

Keywords Machine learning · Lithofacies · Random forest · Support vector machine

1 Introduction

Lithofacies classification based on wireline logs is often subject to interpretation bias as it is traditionally performed by an expert, who assigns a lithology to a particular set of wireline log responses then cross-references the interpretation to the lithology observed in cores (when available). The challenges in manual facies interpretation are often associated with lack of data, poor data quality, interpretational uncertainty, variation in terminological conventions, and the increased workload for the expert when using large datasets. Biases are highly likely to be incorporated into the final decision about the vertical succession when the expert's prior knowledge regarding the lithology expected within the field influences the interpretation. Compiling a series of well logs from heavily explored basins or giant oil/gas fields is an intensively time-consuming and strenuous task to perform. The uncertainty associated with well log interpretation must be accounted for, as it may have a significant impact on development decisions. This interpretational uncertainty is usually difficult to determine using a traditional, deterministic approach.

Automation of lithofacies classification using machine learning tools appears to a promising solution for the challenges faced during traditional manual interpretation. A machine learning facies classification contest was organized by the Society of Economic Geologists (SEG) in 2016 (Hall 2016), aiming to achieve the greatest accuracy in facies predictions for blind wells. Out of a large number of teams using various learning-based techniques, the highest accuracy achieved was 65%. Tree-based algorithms, such as boosted trees and random forest, coupled with augmented features performed best (Hall and Hall 2017). This level of accuracy, although impressive for automatic systems, still leaves much room for improvement, and the ability to quantify the uncertainty in such automatic facies interpretation is also lacking. The top three teams developed an automated classification relying on decision-tree algorithms coupled with either engineered features (i.e., automatically derived from wireline data) or spatial gradients. Engineered features can be derived in various ways, e.g., by numerical processing (averaging, deriving variance, gradients, etc.) or by taking account of domain knowledge about the importance of certain particular aspects (e.g., high/low values, joint increase/decrease, joint opposite change, etc.). The algorithms developed on engineering-derived features followed standard computer science practice, which may lack an adequate representation of the geological context through the integration of petrophysical characteristics of the classified rocks themselves. Thus, the solutions obtained in such a “mechanistic” fashion, although undoubtedly of quality, depart

from the variable lithology in the field and are limited in their ability to account for domain knowledge and any other relevant information during the classification.

Facies interpretation relies on the geological context described by geologically interpretative features and petrophysical rules. A lithofacies is primarily defined based on the texture, mineralogy, and grain size, viz. geologically derived characteristics that are often shared amongst similar lithofacies when an individual characteristic is considered (e.g., two facies might share a medium grain size). However, when associating those characteristics (e.g., considering grain size, pore size, and argillaceous content together), unique combinations can be defined and thus might describe a unique lithology, albeit subject to some uncertainty which must be accounted for. The present study considers whether features derived from the lithology of rocks (i.e., texture, grain, and pore size) could improve the accuracy of automated facies classification.

Two experiments integrating geological knowledge are presented herein. The first experiment investigates the value of an additional input geological-rule-based constraint with respect to the test accuracy and reproducibility of both heterogeneity and spatial variation. The second experiment examines the impact of additional geologically independent features, through evaluation of the test prediction accuracy and input feature importance. Then, the impact of random and systematic noise on the robustness of the classifier is investigated. Finally, the classification uncertainty is assessed using different combinations of training/validation sets.

A random forest (RF) classifier was chosen for this study. A brief overview of the method is followed by a description of the setup, the hyperparameter tuning process through training, and the model selection. The random forest classification results are then benchmarked against another widely used classifier, i.e., the support vector machine (SVM). The paper is organized as follows: Sect. 3 details a brief overview of the classifiers used, the experimental setup, especially the generation of the geologically derived additional input features as wells, and a description of the implementation of the noise injection; Sect. 4 introduces the results of classification using the supplementary features and noise; Sect. 5 is devoted to discussing the results, then Sect. 6 draws conclusions.

2 Brief Geological Overview

2.1 Local Geology: the Panoma Gas Field and Lower Permian Council Grove Group

This paper considers the facies classification problem as presented in Hall (2016). The case study considers the lithologies of the Council Grove Group from the Panoma Gas Field located in southwest Kansas, USA. The field is part of the Giant Hugoton Field, whose reservoir consists of cycles of shelf carbonates grading laterally into nonmarine siliciclastics at the basin margins (Heyer 1999). The vertical facies succession exhibits fourth-order eustatic cycle sequences (Olson et al. 1997; Heyer 1999; Dubois et al. 2006). This cyclicity produced a predictable vertical succession of eight major lithofacies (Dubois et al. 2003), reflecting a shoaling-upward pattern as a result of rapid sea level fluctuations occurring in a shelf or ramp environment, further characterized

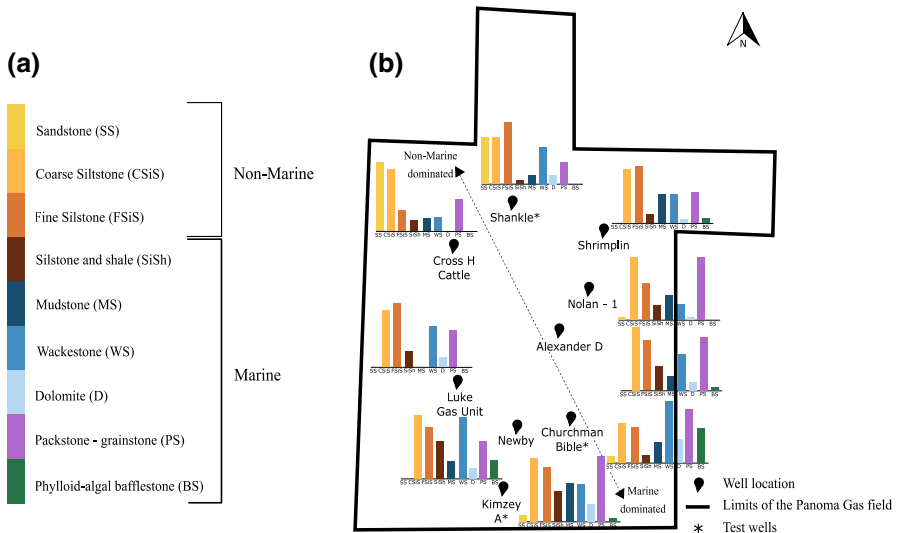


Fig. 1 Lithofacies in the Council Grove Group (a) and their distribution at the studied wells (b)

by a low-relief, laterally extensive, shallow ramp. The ramp dipped gently southward and migrated in response to cyclic sea level fluctuations (Siemers and Ahr 1990).

2.2 Facies

The cyclical vertical succession of the Council Grove Group exhibits an expected pattern of eight main facies (Fig. 1a). A simplified set of 8 (+ 1) lithofacies were derived from Puckette et al. (1995) for this study after Hall (2016) (Fig. 1a). It is important to note that, although eight principal lithofacies are defined in earlier literature (Puckette et al. 1995; Heyer 1999), nine facies are considered herein, specifically: sandstone (1), coarse siltstone (2), fine siltstone (3), siltstone and shale (4), mudstone (5), wackestone (6), dolomite (7), packstone–grainstone (8), and phylloid algal bafflestone (9). Facies 1–3 are non-marine-dominated facies, while facies 4–9 are marine-dominated facies. The phylloid algal bafflestone facies (9) is distinguished from the packstone–grainstone facies (8) for the sole purpose of classification using machine learning, as it displays distinct wireline log properties (Dubois et al. 2003; Hall 2016).

The transitional nature of the depositional environment coupled with a cyclical process of deposition produced a lateral trend of the facies, which is observable in the field (Fig. 1b). The northwest side of the Panama Field is dominated by nonmarine shale and siltstone, while the southeast is mainly marine dominated (Dubois et al. 2003). This lateral trend presents an additional challenge for automated classification of the Council Grove Group facies. Another difficulty in classifying the facies arises from their similarities, a problem which is also commonly encountered during manual interpretation. As an attempt to take these similarities into account, adjacent facies were defined by Dubois et al. (2007) based on shared petrophysical properties (Table 1).

Table 1 Facies and their respective adjacent facies

Facies	Description	Adjacent facies ^a	Adjacent facies for noise injection
1	Sandstone (SS)	2	2
2	Coarse siltstone (CSiS)	1, 3	3
3	Fine siltstone (FSiS)	2	1
4	Siltstone and shale (SiSh)	5	5
5	Mudstone (MS)	4, 6	6
6	Wackestone (WS)	5, 7, 8	4
7	Dolomite (D)	6, 8	8
8	Packstone–grainstone (PS)	6, 7, 9	9
9	Phylloid algal bafflestone (BS)	7, 8	7

^aAs described by Dubois et al. (2007)

3 Experimental Setup

Supervised machine learning classification implies setting up a system of inputs (i.e., explanatory variables), whose combination results in a target variable (i.e., facies labels). Defining the correct combination of inputs is essential for the classification outcome and plays a major role in the performance of the classifier.

The output data in this supervised classification study are interpreted lithofacies, which are subject to human bias and uncertainty. Unlike classic categorical variable classification problems, geological facies classification may have a nonlinear fuzzy nature, subject to interpretational bias. In practice, lithofacies classes can exhibit gradational or nonlinear boundaries, depending on their environment of deposition and the abruptness of environmental changes.

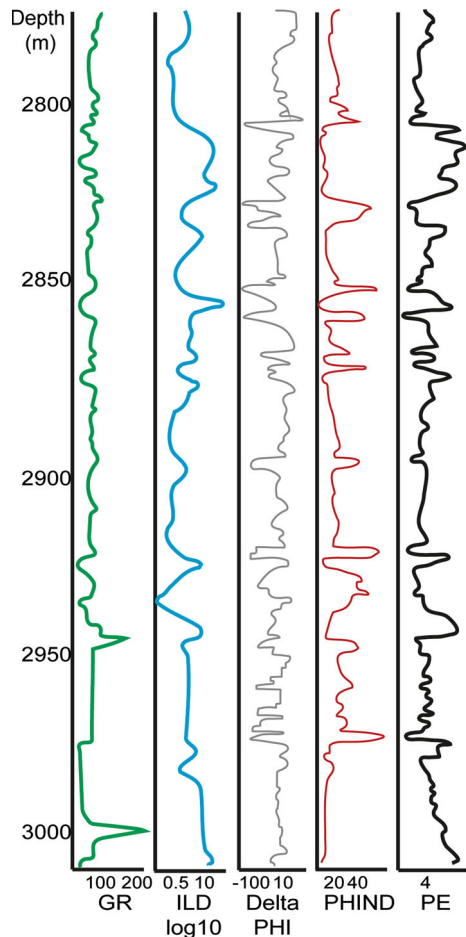
3.1 From Data to Features

3.1.1 Conventional Wireline Log Input Features

As the foundation of the classifier, the input features are key elements in label prediction using ML. Here, the data were compiled from the open-source dataset provided by Hall (2016) and Dubois et al. (2006); it includes input features from nine wells and one pseudowell (Recruit F9). The pseudowell amalgamates data to better represent the phylloid algal bafflestone facies (9), which is underrepresented in the real wells.

As mentioned in the “[Introduction](#),” traditional manual lithofacies interpretation is performed by assigning wireline log responses to a particular lithology. For this study, three conventional wireline logs were available: gamma ray (GR), the logarithm of the deep induction resistivity (ILD_log10), and the photoelectric factor (PE). It is also common, when identifying lithofacies from wireline logs, to investigate the

Fig. 2 Schematic examples of wireline logs: gamma ray (GR), logarithm of the deep induction resistivity (ILD_log10), average of the neutron and density log porosity (Δ PHI), difference between the neutron log and the density porosity (PHIND), and photoelectric factor (PE) for the Shankle well



behavior of the logs in parallel. Here, the average of the neutron and density log porosity (Δ PHI), and the difference between the neutron log and the density porosity (PHIND) were available (Fig. 2). The absolute values of the two log curves plus their degree of separation provide important information on both the porosity and the density of the fluids (e.g., water versus hydrocarbon or presence of gas in the formation) (Hartmann 1999). Finally, the open-source data also include two geologically derived input features: a depositional indicator (NM_M) and a stratigraphic cycle relative position (RelPos). The first geological variable, NM_M, allows differentiation between lithofacies with similar petrophysical characteristics but deposited in different environments (NM: Nonmarine; M: Marine). The second variable, RelPos, is an indicator of the relative position within a stratigraphic cycle on a linear scale between 0 and 1, where 0 is the bottom of each half-cycle interval and 1 is the top of the half-cycle (Fig. 3).

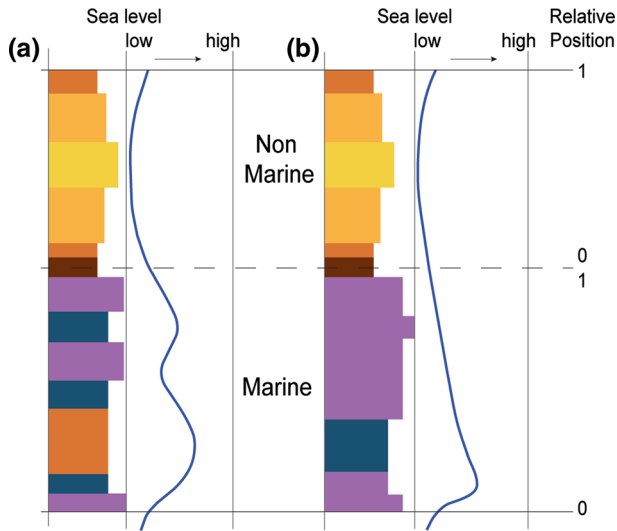


Fig. 3 Idealized Council Grove Group cycles replotted from Dubois et al. (2006). Two cycle types are inferred from the stacking pattern of the lithofacies and the relative sea level curve: **a** symmetric with complexity, and **b** asymmetric. The relative position, RelPos, is the input reflecting the position within each half-cycle, being 0 at the bottom and 1 at the top

Table 2 Summary of petrophysical rules used to construct the additional geologically constrained input

Facies	Geology	Petrophysical GR rule
3	Hot bed marker	> 200 API
4	Limestone matrix	40–120 API
7	Increasing amount of allochems with decreasing amount of limestone matrix	35–55 API
8	Anhydrite cement	< 37 API

3.1.2 Design of Petrophysically Constrained Wireline Log Input

A petrophysically constrained input was created by assigning a facies number to a wireline log pattern identified as characteristic and unique for that facies (Table 2), mimicking the method an expert would use for manual facies interpretation. The packstone lithofacies (7) is associated with gamma-ray values ranging from 35 to 55 API (Heyer 1999) from an increasing amount of allochems (skeletal grains) and a decreasing amount of limestone matrix (Puckette et al. 1995). The wackestone facies (4) is composed of a limestone matrix which is associated with gamma ray readings within 40–120 API (Heyer 1999). Similarly, another facies has unique wireline log responses corresponding to its composition: the packstone–grainstone lithofacies (8) is cemented by anhydrite (Puckette et al. 1995) and is, therefore, believed to exhibit GR responses always lower than 37 API (Heyer 1999).

Finally, due to its depositional environment, the fine siltstone facies (3) occurs above a hardground surface of skeletal phosphate corresponding to the highstand system tract. Puckette et al. (1995) defined this zone as a “hot” marker bed in a gamma-ray log due to its high glauconite component. Such a hot marker bed yields a high gamma ray signature, usually greater than 200 API.

These distinct gamma readings were found in each well, and their respective facies numbers were assigned to the readings to form an additional input. The geological constraining input was purposely created and used to more accurately define the petrophysical responses of the facies of the Council Grove Group, with the goal of producing a better performing classifier using a more thorough learning technique.

3.1.3 Design of Geologically Interpretative Input Features

The aim of this study is to assess the value of interpretative geological features, such as the geological texture of the rocks. This can be converted to input features for integration in the machine learning classification. These features, although derived from an independent set of measurements, may not be considered completely independent because facies identification and interpretation are largely based on sedimentological features observed in cores, which are related to rock texture characteristics.

The geologically interpretative input features were constructed and derived from geologically interpretative characteristics observed in rock: grain size, pore size, and content of clay minerals. The extensive petrophysical study of the lithofacies performed by Dubois et al. (2003) and their digital rock classification system provided essential tools for the derivation of these additional inputs. Grain size (GS) is one geological characteristic that is easily assessed and described, being most often described qualitatively (e.g., very coarse sand, medium sand, very fine sand, etc.). Quantitative ranges (e.g., 1000–2000 μm , 250–500 μm , and 62–125 μm , respectively) are likewise implemented as tools to describe the corresponding qualitative ranges, being exploited to derive and build the additional input features. Each of the eight lithofacies (facies 9 being incorporated with facies 8) is classified within a defined grain size range from the lithology observed in the core (Dubois et al. 2003).

The input feature for each individual well was derived by assigning a random grain size value within the appropriate range to each depth instance. Such assignment of a random rather than constant (e.g., average) value reflects the natural grain size variability within rocks belonging to the same lithofacies. Use of a random value implies no subscale correlation (i.e., nuggets) below the resolution of the facies unit with the cycle. It is assumed here that the subsurface variability considers features within the facies log resolution, whereas features below this resolution (i.e., burrows, ripple cross lamination, thin beds and laminae, etc.) are not considered. Randomization of the additional inputs (e.g., grain size) also helps to avoid overfitting, when one of the continuous input variables imposes a one-to-one value correspondence with the categorical output.

The random assignment was then performed using the argillaceous content (AC; the percentage of clay mineral contained within the rocks) of each lithofacies in a similar way. Two more inputs were constructed using the same random assignment technique

on both the principal pore sizes (P_Pore) and subsidiary pore sizes (S_pore), which are linked to lithology.

Overall, four additional geological interpretative features were constructed: GS, AC, P_pore, and S_pore. A unique combination of these four inputs defines a specific facies; For example, the packstone–grainstone facies (8) is defined as exhibiting a grain size of medium sand (i.e., average grain sizes between 250 and 500 μm), a clean argillaceous content of less than 1%, a fine principal pore size (i.e., lying between 125 and 250 μm), and a microporous size of subsidiary pores (i.e., < 31 μm) (Dubois et al. 2006).

3.2 Classifier Algorithms

3.2.1 Random Forest (RF)

Classification of an instance label using a random forest (RF) algorithm is based upon the label (i.e., facies label) predicted most frequently by an ensemble of decision trees (Breiman 2001; James et al. 2013). A decision tree makes a decision after following a classification from the tree's root to a leaf. The decision path includes nodes, branches, and leaf nodes. Each node represents a feature, each branch describes a decision, and each leaf depicts a qualitative outcome. A random forest is an ensemble, or forest, of such decision trees that aims to predict the leaf label with the least error. One way to minimize the error is to determine the best number of features to be used at the root and at the nodes. The best number of features is the feature combination with the highest information gain. RF implements entropy as a measure of impurity to characterize the information gain.

As indicated by the name itself, RF is a machine learning tool incorporating random components into its core processes. Each tree is grown and fit through a bagging process based on random selection (with replacement) of samples in the training set, which improves the stability and accuracy of the machine learning classification (Breiman 2001; James et al. 2013). A decision tree is subsequently grown on each of those training subsets; all the grown trees are then amalgamated into an ensemble of decision trees, or forest. The random selection of inputs or combination of inputs as split candidates at each tree node is a further component acting towards the integration of randomness into the algorithm (Breiman 2001; James et al. 2013). The random subset of input features considered at a node is the result of the selection of only a small group from the full set of features to split on. The bootstrap procedure and the random feature selection reduce the correlation between the trees by decreasing the strength of individual trees (which reduces the risk of overfitting in the presence of a strong, dominant predictor; James et al. 2013; Pelletier et al. 2017).

As in any data-driven model, RF has two hyperparameters to be tuned: the number of trees in the forest and the number of features to be randomly selected for splitting at each node (Cracknell and Reading 2014). The optimization of these two hyperparameters was performed through cross-validation. The best combination of both the number of trees and the number of features was selected as the combination with the lowest validation error (i.e., prediction error on the validation set). The number of

trees in the forest was set to range from 100 to 800, with an increment of 100 based on numbers commonly used in literature (around 500 trees). The number of features ranged from 1 to 7, or 1 to 11 depending on the experiment and the number of input features used. It should be noted that, if the maximum number of features is selected as the optimized hyperparameter (i.e., 7 or 11), the experiment is no longer a random forest classifier but rather a bagging classifier (James et al. 2013). The prediction errors were computed for the predictions resulting from each unique combination of the two parameters, and the combination with the lowest error was selected as the optimized hyperparameters.

Considered to be a powerful classifier, RF is believed to be relatively robust to outliers and noise, while being easily parameterized (Breiman 2001). On the other hand, this algorithm could be subject to severe overfitting, due to the reliance of the core of the algorithm on classification and regression trees (CARTs; Evans et al. 2011).

3.2.2 Support Vector Machine (SVM)

SVM classification is computed from the idea of a linear boundary exhibiting a maximum width separating two classes. This linear boundary is defined by a subset of data points—support vectors which define the widest decision margin (Cortes and Vapnik 1995). A nonlinear separability problem is then solved by mapping the boundary in a multidimensional feature space using a kernel transformation of the support vectors (Vapnik 1995). The optimal hyperplane in the feature space has the maximum distance to the support vectors, which are the closest data instances to this hyperplane (Pelletier et al. 2017). SVM is a supervised model that is trained to find support vectors best defining the hyperplane based on the training data. Such training is routinely done using a standard SVM functionality.

Two hyperparameters of the SVM classifier require optimization: the regularization factor C and the γ parameter. C controls the trade-off between the model complexity and its capability to reproduce outliers in the data. It imposes a bound on the number of support vectors which subsequently allow for variation of the roughness of the hyperplane. The γ parameter is proportional to the inverse of the kernel width and controls the continuity of the classification boundary. Optimal choice of these hyperparameter combinations is important due to their impact on the classification results. A poor choice of hyperparameters may lead to either over- or underfitting (Kanevski et al. 2009). The parameter optimization is achieved when the cross-validation errors depending on C and γ are minimal (Kanevski and Maignan 2004; Friedman et al. 2001). C and γ reach a minimal cross-validation error at 10 and 1, respectively (Fig. 4).

3.3 Experimental Dataset Design for Classifier Training

Following the choice of algorithm, it is important to tune the hyperparameters of the classifier and select the right model. These decisions are made based on splitting the data into training, validation, and test sets. The approach used to design these three sets varies in literature. The training set is the one onto which the model is fit; it is the set from which the classifier learns the rules to link the inputs with the output

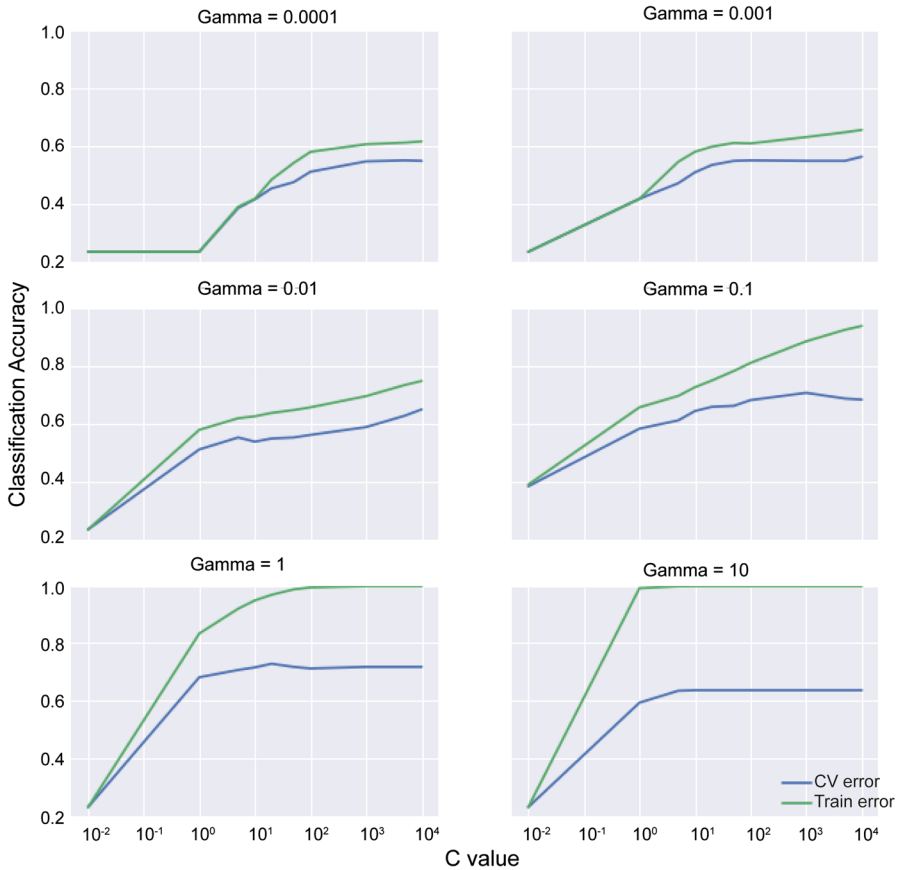


Fig. 4 Plots of training and cross-validation errors against the complexity—both C and γ . The optimized parameters for which the cross-validation error reaches a minimum are 10 and 1, respectively. Replotted from Hall (2016)

labeled targets. A training set must therefore be representative of the patterns to be classified to produce adequate estimates (Duda et al. 2012). The training set is used to fit the model for any combination of hyperparameters, while the validation set is used to select the optimal combination of hyperparameters (tuning). Finally, the test set is used to assess the performance of the final optimized model (Kanevski et al. 2009).

Cross-validation (k -fold) is often used for designing sets for training–validation data. This approach consists of dividing the data into several k folds (Duda et al. 2012). The first fold is treated as the validation set, while the remaining $k - 1$ folds are used as the training set to fit the model (James et al. 2013). A similar approach is used here, but instead of using random folds, the full set of instances for an entire individual well are used as equivalent to the folds.

The test well must be set aside initially to assess the performance of the final classifier. Three different test wells were selected for this experiment: Shankle, Kimzey A, and Churchman Bible (Fig. 1b). The selection of three different test wells aims to

assess the classification performance with respect to the lateral trend of facies observed in the field. Each of the test wells is located in a different region, having a different relative distribution of facies. The Shankle well is dominated by nonmarine facies, while the Churchman Bible well is marine dominated. The facies at the Kimzey A well are present in relatively equal proportions. Varying the test well location aims to investigate the accuracy and performance of the classifier under varying facies distributions.

To design the training and validation sets, two wells in any one case were set aside for validation, while the remaining seven wells were used for training. Together, the two validation wells and seven training wells represent one possible training/validation case. As a result of the lateral facies trends, the instances within the training and validation sets may vary considerably. To accommodate this, the well sets for validation and training were reshuffled until all possible combinations were used. This was motivated by the lack of a priori means to select the combination of wells able to predict with the highest accuracy. Shuffling of the training and validation sets produces predictions which may vary from one data combination to another. Here, the uncertainty range is assessed across the 36 possible predictions based on different combinations of training and validation sets for each of the three blind test well experiments.

A slightly different experimental setup was implemented to investigate the value of the additional geologically constrained input. The training data were set to seven wells: Luke G U, Kimzey A, Cross H Cattle, Nolan, Recruit F9, Newby, and Churchman Bible, while the validation set was composed of two wells: Shrimplin and Alexander D (Fig. 1b). The test set contained the remaining well (Shankle). The choice of training and validation sets was made such that at least one sample from each facies was contained in the training set, increasing the likelihood of correct facies assignment and therefore higher classification accuracy. Additionally, for comparison purposes with regards to the performance of the RF classifier, another combination of training, validation, and test sets was examined, where the Shankle well was replaced with the Churchman Bible well as the test well (i.e., Shankle well is now in the training set), while the validation set remained unchanged.

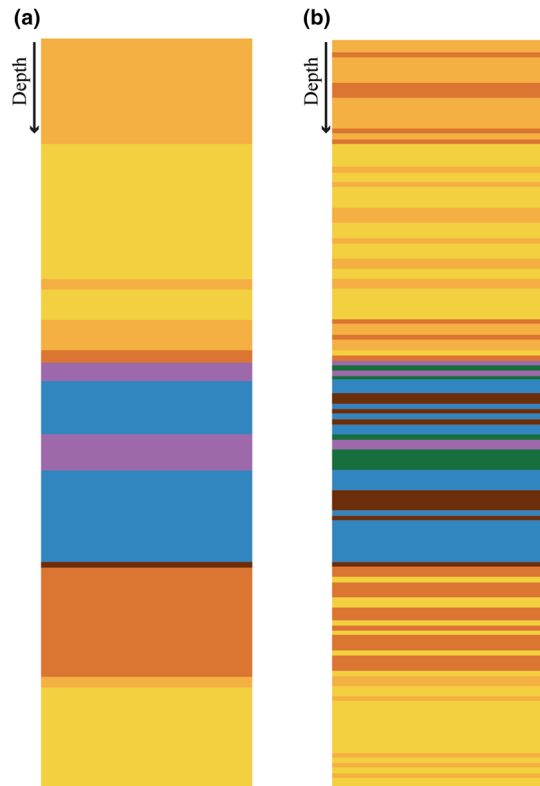
3.4 Noise Injection

Noise injection procedures were performed to investigate the robustness of the classifier to noise and the impact in terms of the degree of misinterpretation. Two types of noise injection were applied to the training and validation sets: nonsystematic random noise and systematic noise (Pelletier et al. 2017). Noise injection is also sometimes used as a training boost and control of model complexity (Kanevski et al. 2009).

3.4.1 Nonsystematic Noise

A nonsystematic or random noise injection procedure was applied to the input variables. An additional geologically derived input feature, here the grain size, was shuffled with respect to the depth reading to make it 100% irrelevant to the target output. The depth shuffling of the feature instances was performed for 100% of the depth instances

Fig. 5 Extract of logs of vertical facies succession with depth in the Shankle well for true facies (a) and after 30% systematic noise injection (b)



and in a completely random manner. This approach renders the noise injection non-systematic (i.e., no pattern or direct control on the label flipping). Such noise injection is performed to evaluate the robustness of the model to random/measurement error or to an irrelevant input.

3.4.2 Systematic Noise

Systematic noise injection displays a pattern of disruption of the output data to account for inevitable misinterpretation, by systematically swapping the target labels between a pair of adjacent facies (Pelletier et al. 2017). Here, 30% of each interpreted facies was systematically swapped with one of its adjacent facies, as defined by Dubois et al. (2006) (Table 2). Adjacent facies account for misinterpretation as they represent pairs of facies which display a transition zone where the sediments of the two facies are “mixed” due to a gradual change in the depositional environment (Fig. 5). Consequently, the systematic noise injection procedure allows evaluation of the robustness of the model to these errors in interpretation and judgment.

Table 3 Experiment 2: impact of additional geologically derived input

Test accuracy	Shankle (%)	Churchman Bible (%)
Seven inputs, ISPL team ^a classifier	59.1	60.1
Eight inputs, ISPL team ^a classifier	60.1	62.4

^aThe RF script used here was based on the script developed by the ISPL team of the 2016 machine learning SEG contest

4 Analysis of Results

4.1 Experiment 1: SVM versus RF

4.1.1 Test Prediction Accuracy

The support vector machine classifier was trained using the code completed by Hall (2016). The prediction of the Shankle facies succession achieved a total test prediction accuracy of 41%. Compared with the SVM algorithm, the algorithm for the RF was scripted by modifying the available code from the ISPL team of the 2016 ML contest (Hall 2016). The RF predicted the lithofacies of the Shankle well with an accuracy of 59.1% (Table 3). Churchman Bible was subsequently used as the test well to study the classifier performance when changing the nature of the test set: marine-dominated well versus non-marine-dominated (i.e., Shankle) well. The prediction of the Churchman Bible non-marine-dominated facies sequences using a random forest algorithm reached a test prediction accuracy of 60.1% (Table 3).

A geologically consistent classifier should not only reach the lowest test error but should also provide predictions with the most geological character, with spatial correlation, heterogeneity, and distribution being reproduced. Therefore, other methods were used on top of the test accuracy to estimate the performance of each prediction. Statistical analysis tools were used to assess the heterogeneity and the spatial correlation and variability. These include the coefficient of variation, the distribution of the thickness of reservoir facies (facies 3), and variograms.

4.1.2 Heterogeneity

Heterogeneity is a term related to the level of intrinsic variability in any geological characteristic (Fitch et al. 2015). A pattern in variability might be observed at multiple scales within a single reservoir. It is therefore very important to define the scale at which heterogeneity is studied, as the variability could differ across all observational scales (Fitch et al. 2015). Here, the heterogeneity is considered at a stratigraphic level (i.e., the variability between facies). A commonly used technique to easily investigate the heterogeneity is the coefficient of variation (CV; Fitch et al. 2015). A coefficient lower than 0.5 suggests that the facies succession is homogeneous, while a coefficient lying between 0.5 and 1 highlights a heterogeneous succession, with increasing het-

erogeneity towards 1. This variability assessment technique is usually performed to investigate the permeability variability (Jensen et al. 2000) but is applied here to the thickness of facies intervals. The CV evaluates bulk heterogeneity, disregarding its spatial character; the calculation technique investigates the variability within the data distribution while removing the influence of the original scale of measurement (Fitch et al. 2015). The coefficient of variation is calculated using Eq. (1).

$$CV = \frac{\sigma}{\text{mean}}, \quad (1)$$

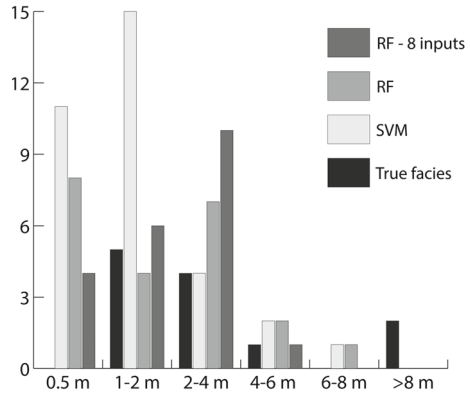
where σ is the standard deviation and “mean” is the arithmetic average (Brown 2012).

Focus has been devoted to the study of heterogeneity at a stratigraphic level because of its importance towards flow simulation modeling. Facies predictions using ML tools should therefore reproduce the degree of intrinsic variability within the true facies to produce a more realistic flow simulation model. Additionally, it should be noted that other techniques to evaluate the heterogeneity of geological characteristics have been used in literature. Fitch et al. (2015) also provide examples of two other commonly used statistical characteristics of heterogeneity: the Lorenz coefficient and the Dykstra–Parson coefficient.

The CV for the SVM predictor was found to be 0.601, indicating that the vertical distribution is heterogeneous. The RF predicted the vertical facies succession for the Shankle well to be heterogeneous (CV=0.64), while that for the Churchman Bible is considered to be homogeneous (CV=0.42). However, it must be noted that the latter difference in heterogeneity is also exhibited in the true facies succession of both wells (0.65 and 0.43, respectively). The CV in this study represents a relative level of heterogeneity with regards to the facies code used for this study. The RF predictions are closer to the true facies than the SVM predictions in terms of layer variability, therefore evidencing the greater capability of this classifier to reproduce the heterogeneity level of the true vertical stratigraphic sequence.

The heterogeneity of the vertical succession of the facies predicted by the classifier is also assessed by the distribution of facies interval thicknesses. Indeed, the greater the number of thin layers within the succession, the more prone the succession is to be heterogeneous; the facies intervals would alternate considerably more frequently than when thick intervals compose the sequence. Facies 3 was selected for thickness distribution based on its being the dominant reservoir lithofacies in the Shankle well and also allowing for sampling of layer thickness. From the distribution in Fig. 6, it is apparent that the vertical succession predicted by SVM is composed of a higher number of thin layers than the correct facies succession, notably for those less than 1 m and between 1 and 2 m. The further assessment of the heterogeneity through the distribution of the reservoir thickness of facies 3 for the Shankle well (Fig. 6) demonstrates predictions of thinner layers in comparison with the true thicknesses, with overestimation of the number of intervals under 1 m thickness and between 2 and 4 m. Likewise, the interval thickness distribution of the Churchman Bible facies 3 indicates predictions of thicker layers with a greater number of thicknesses lying between 4 and 6 m, coupled with a fewer layers of thickness between 1 and 4 m. The

Fig. 6 Distribution of reservoir thickness, fine siltstone facies (3), for the Shankle well



RF prediction is also superior to the SVM one with regards to the thickness distribution of the reservoir facies.

4.1.3 Spatial Continuity

The spatial aspect of the vertical facies heterogeneity was further analyzed using variograms, a conventional geostatistical exploratory correlation tool (Matheron 1963). Variograms quantify spatial correlation to help explore spatial patterns, trends, cyclicity, and randomness based on data. Spatial correlation analysis with variograms is important to make inferences about the scale of the correlation and relevance of a stationarity assumption of some sort. Strong vertical trends and periodic behavior are expected in the case considered herein due to the alternating features in the marine and nonmarine depositional environments. It is important to demonstrate how machine learning classifiers are able to identify and reproduce nonstationary spatial correlation. Amongst the true facies variograms, some display distinct vertical correlation patterns: facies 1 displays a periodic pattern, facies 3 shows a strong correlation pattern, while facies 4 presents a trend (Fig. 7). The variograms for the SVM predictions highlight a lack of reproducibility of both the trend of facies 4 and the periodicity of facies 1. The spatial variation amongst the data of the RF predictions investigated with variograms demonstrates the capability of the classifier to reproduce both the periodicity observed in facies 1 and the general trend in facies 4. Yet, both classifiers are unable to reproduce the well-correlated pattern of the true facies 3.

This demonstrates that variograms can be an important measure of how well a classifier is able to reproduce spatial heterogeneity, and not just how closely the data are matched according to some averaged error measure. Spatial heterogeneity measures can be used as additional criteria in learning-based classifiers to help enhance the interpretability of the outcomes.

4.1.4 SVM versus RF: Comparative Discussion

The overall higher accuracy and performance of the RF classifier are believed to be associated with the algorithm itself. From its core processes of tree building,

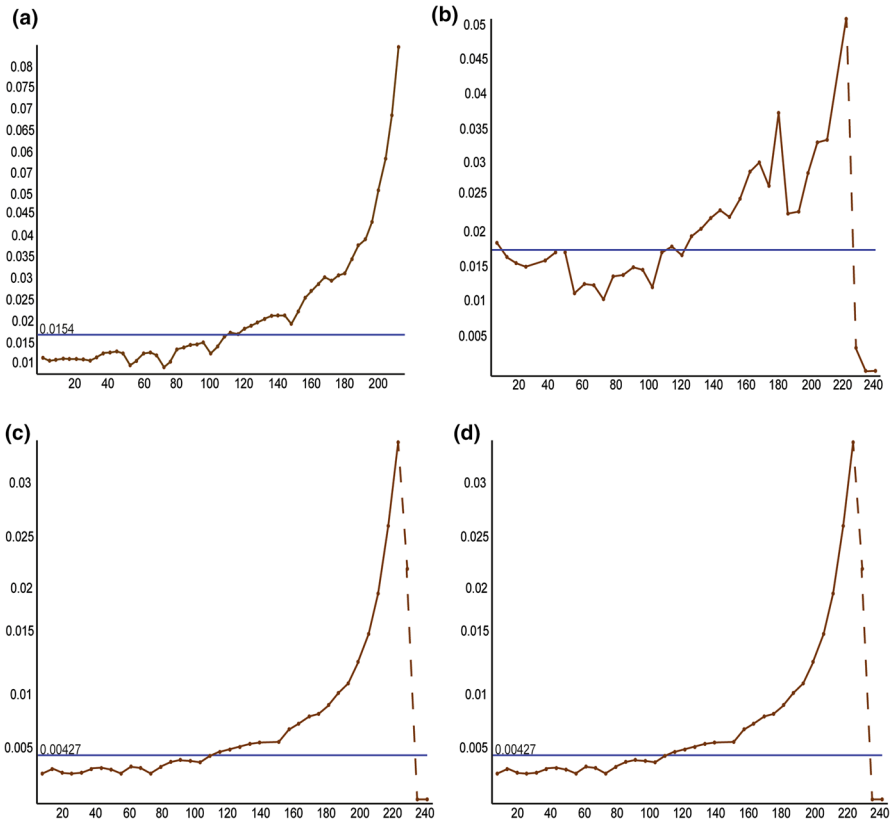


Fig. 7 Variograms for facies 4 in the Shankle well, exhibiting a trend for true facies (a), no pattern for SVM predictions (b), a trend for RF seven-input predictions (c), and a trend for RF eight-input predictions (d)

the RF algorithm performs a classification which decorrelates each individual tree (Sect. 3.2.1). Any strong, dominant correlator inside the data is only selected in a few bootstrapped training subsets, allowing the less strong predictors to be more frequently picked at a tree node. This increased likelihood of a RF predictor independent of its strength results in a more accurate representation than those obtained from the SVM.

4.2 Experiment 2: RF with Additional Geological Constraining Input

The classifier is set up as in the previous random forest experiment, except for the additional geologically constrained input integrated within the input features; a total of eight input features were used for this experiment. It was trained on the training dataset, validated on the validation set for parameter optimization, then exposed to either the Shankle or Churchman Bible well to compute the test prediction accuracy (Table 3).

The level of heterogeneity for the predicted sequences of facies is described by the coefficient of variation. As described previously, the correct facies succession for the

Table 4 Experiment 3: impact of additional interpretative geological inputs

Average test accuracy	Shankle (%)	Kimzey A (%)	Churchman Bible (%)
Eleven-input classifier	99	98	97
Seven-input classifier	54	52	53

Shankle well has a CV value suggesting a heterogeneous vertical sequence, while the Churchman Bible succession has a coefficient suggesting homogeneous layering. The random forest with additional geological constraints scored a CV of 0.64 and 0.42 for Shankle and Churchman Bible, respectively.

The heterogeneity with regards to the distribution of thickness of facies 3 resulted in predictions exhibiting a relatively similar distribution of thickness to that of the Shankle well true facies (Fig. 6). The significant difference is the lack of layers with thickness greater than 8 m, and a slight increase in layers less than 4 m thick. A similar conclusion was drawn for Churchman Bible, where the only difference is a greater number of layers less than 4 m thick.

The variograms for the RF with additional geological constraints correctly reproduced the periodicity and trend exhibited by the Shankle well facies 1 and facies 4 respectively (Fig. 7). Identically to the previous random forest experiment, the predicted facies did not reproduce the well-correlated pattern observed in facies 3.

4.3 Experiment 3: RF with Additional Geological Independent Features

4.3.1 Test Prediction

The value of the additional geological interpretative inputs was first investigated by computing the total test prediction accuracy (Eq. 2) from the predicted facies labels for each combination of training and validation sets. The total prediction accuracy is a measure of the classifier performance with regards to the predicted labels in a test set.

$$\text{Total Accuracy} = \frac{1}{N} \sum_{i=1}^N (y_i^* = y_i), \quad (2)$$

i.e., the ratio of the correctly predicted labels y_i^* to the total number of labels N .

From the combinations of training and validation sets introduced in Sect. 3.3, a total of 36 test prediction accuracies were obtained for each of the three test wells. These 36 accuracy values were subsequently averaged to obtain a single accuracy value per test well. The results introduced and discussed here are the arithmetic means of the accuracy values (Table 4).

A considerable increase in the accuracy for the three test wells was observed for the 11-input classifier (i.e., the classifier with the additional geological features) (Table 4). Where the seven-input classifier achieves an average accuracy of only 54%, the 11-

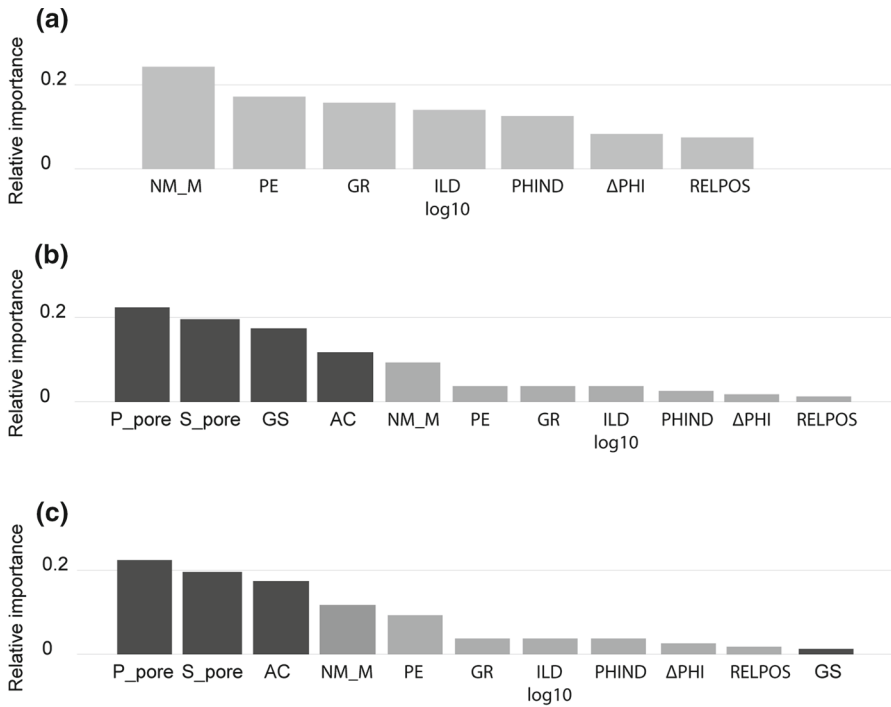


Fig. 8 Relative feature importance for the RF 7-input classifier (a), RF 11-input classifier (b), and RF 11-input and random GS feature classifier (c)

input classifier reaches an average accuracy as high as 99%. Incorporating additional geological inputs into the classification clearly results in a substantial boost in the test prediction accuracy.

4.3.2 Feature Importance

Another approach to assess the value of additional geological features is to examine the importance of each of the features when considered at a node for the tree to reach a final decision. NM_M, GR, and PE are the three most informative features when the trees built on the seven inputs come to a decision (Fig. 8a). Adding the four geologically interpretative features, the weights of the original features are reduced (see the shift to the right in Fig. 8b). This suggests that the four additional geological inputs are the most valuable and meaningful for the trees to decide. A novel approach to feature attribution for tree ensembles has been proposed to resolve inconsistency in existing methods (Lundberg et al. 2018), but further examination of this lies beyond the scope of this study.

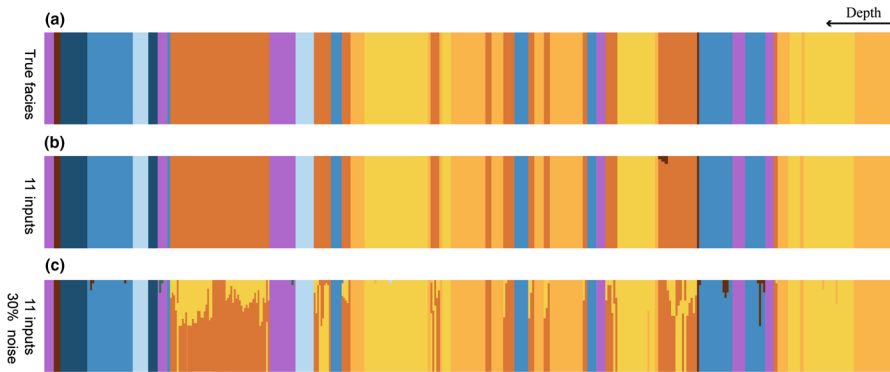


Fig. 9 Logs of vertical facies succession with depth in the Shankle well for true facies (a), the proportion (amalgamation of the predictions from the 36 combinations of training and validation sets) of predicted facies for the RF-11 input classifier (b), and the proportion of predicted facies for the RF-11 input with 30% systematic noise classifier (c)

4.3.3 Noise Injection

Injecting random noise (i.e., 100% depth-shuffled grain size inputs) results in the classifier disregarding this feature. This is emphasized by the change of rank of the grain size input with regards to the feature importance (from third most important to least important, Fig. 8c), while the average test accuracy prediction remains as high as 99%.

Systematic noise was injected into the training and validation sets for the Shankle well as the test well. The average test prediction accuracy of the 36 combinations of training/validation wells is 92% (Fig. 9).

5 Discussion

5.1 Value of Additional Geological Inputs

5.1.1 Geologically Constrained Input

The addition of a geologically constrained input (i.e., eight-input classifier) resulted in only a marginal improvement in test prediction accuracy, from 59.1% to 60.1% for the Shankle test well and from 60.1% to 62.4% for the Churchman Bible well (Table 3). The improvement in accuracy for the Churchman Bible is twice as large as that for the Shankle. This greater increase in accuracy is believed to be related to the facies dominating each particular well. The Churchman Bible well is marine dominated and hence mostly exhibits carbonate facies. The additional geologically constrained input was constructed based on petrophysical rules that mostly apply to marine facies (Table 2). Consequently, the constraining input results in a more accurate classification of such marine facies. Thus, when the marine facies compose the majority of the vertical succession, the total test prediction accuracy is higher.

Considering only the test prediction accuracy, one would lean towards the conclusion that the addition of the geological constraining input increases the accuracy of the facies prediction, albeit only slightly. Both classifiers predict facies succession with the same heterogeneity level (i.e., same CV value). Both classifiers therefore adequately represent the inherent heterogeneity of the vertical stratigraphic facies succession of the Council Grove Group in both the Shankle and Churchman Bible wells. Additionally, the distribution of layer thickness of facies 3 does not suggest that one classifier outperforms the other; the frequencies of thickness are comparable (Fig. 8). Attention should nonetheless be drawn to the fact that both classifier predictions might impact reservoir flow simulation modeling. A greater number of thin layers could potentially play a major role in controlling the flow trajectory, requiring a change in the reservoir simulation model.

The test prediction accuracy suggests that the classifier with the additional input constrained from geological rules exhibits marginally superior and more accurate performance. Yet, with respect to the heterogeneity and distribution, the superiority of the eight-input classifier may not be significant.

5.1.2 Geological Interpretative Features

Addition of the geologically interpretative features leads to near-perfect classification with accuracy close to 100%. This is related to the fact that the additional input data, although from an independent source, are related to the output labels derived by an interpreter based on the geological characteristics observed in the core. Randomization of the additional input values over the overlapping intervals may further mitigate this effect and reflect a continuous character of lithology changes at a smaller scale in Nature. The following noise injection experiments were performed to test whether the results are subject to overfitting.

Robustness to Noise and Misinterpretation Two noise experiments were performed. One procedure introduced random noise to one of the most important features: grain size. The other noise procedure consisted in introducing systematic noise to the facies labels. Note that the noise injection procedure was not performed on the wireline logs.

The depth shuffling of the grain size input reduced its weighted importance to almost null (Fig. 8c), while the average test prediction accuracy remained at 99%. The latter suggests that the classifier can filter the most informative features necessary to make the most accurate decision. This implies that the model detects and disregards any input feature that might be irrelevant or might have been affected by random or measurement errors.

Injecting systematic noise to a level as high as 30% is equivalent to systematically misinterpreting three instances for their adjacent labels (those facies that can be often confused with each other) every ten instances in the training set. Although, this frequency might appear relatively high, the results of the noise injection procedure suggest that the classifier remained robust. The average test prediction accuracy of 92% highlights the robustness of the model towards this misinterpretation error (Fig. 9). Some pairs of adjacent facies appear to be more sensitive to systematic noise injection than others, exhibiting a higher variability in instance prediction. This is especially

observed between the fine siltstone facies (3) and sandstone facies (1). The greater sensitivity of these facies to noise is further exposed when computing receiver operator characteristic (ROC) cumulative curves. ROC curves illustrate the trade-off between the true-positive rate (i.e., the positives correctly classified divided by the total positive instances) and false-positive rate (i.e., the negatives incorrectly classified divided by the total negative labels; Fawcett 2006). The area under the curve (AUC) is a common tool used to investigate the performance of classifiers (where a perfect classifier has an AUC of 1; Fawcett 2006). The AUC for facies with higher sensitivity to noise such as facies 3 and facies 1 is marginally lower than for the other facies with little to no sensitivity to noise. The greater variability in prediction of facies 3 in the presence of systematic noise is attributed to the four additional independent features sharing no similarity between these two facies. Therefore, the introduction of noise creates additional confusion with regards to the interpretation of facies 3.

The RF classifier is believed to have an innate capacity to counteract misleading and diverging rules arising from instances being mislabeled, as long as the correct labeling of facies is in a greater and major proportion. The robustness of the classifier towards systematic noise and the presence of mislabeled training data is thought to be related to the good generalization ability of the algorithm (Pelletier et al. 2017). This ability might be regarded as a valuable tool for classification including facies of similar lithologies which are easily confused for each other.

Uncertainty Quantification Uncertainty in facies classification is quantified by the range of outcomes across the multiple classifiers computed based on the unique training/validation well combinations. The 36 predictions are combined into a single log of vertical succession of facies, which displays the proportion of the predicted facies at each depth record (Fig. 10). The uncertainty across the range of models (i.e., predictions from different combinations of training and validation wells) encompasses the “true” facies interpretation; the latter is subjective rather than objective and becomes increasingly uncertain with sparse and incomplete data. As each combination’s prediction is unique, combining them allows the identification of the different lithofacies predicted at each depth. When predicting on a test set, the conditioning data are the key to the accuracy achieved; the combination of training and validation data that will predict with the greatest accuracy is not identifiable a priori. The technique of predicting with all available combinations of sets resolves the latter issue, thus the interpretational uncertainty is quantified and monitored. The assumption that the 36 combinations were equally probable was made. This, however, may not be the case, and a probabilistic approach could be applied with Bayesian inference based on posterior analysis of the classified facies (Lin 2002), although further examination of this lies beyond the scope of this study.

Comparing the 7-input log with the 11-input log (Fig. 10b, c), the predictions without additional geological inputs appear to have a higher level of uncertainty associated with each depth. The proportion of predicted facies falls within two and four facies for each depth. This suggests that different combinations of training and validation sets can predict different lithofacies for the same depth. In that case, the correct choice of the combinations used would be vital. The use of fewer features (i.e., without the additional independent geological inputs) results in an overestimated heterogeneity. One element

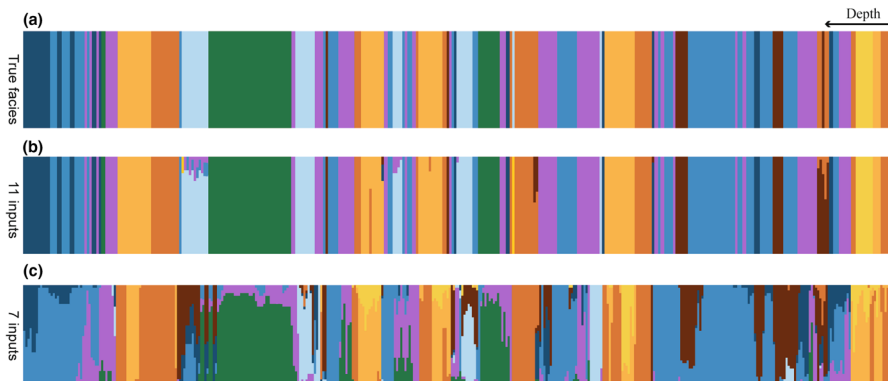


Fig. 10 Logs of vertical facies succession with depth in the Churchman Bible well for true facies (a), the proportion of predicted facies for the RF 11-input classifier (b), and the proportion of predicted facies for the RF 7-input classifier (c)

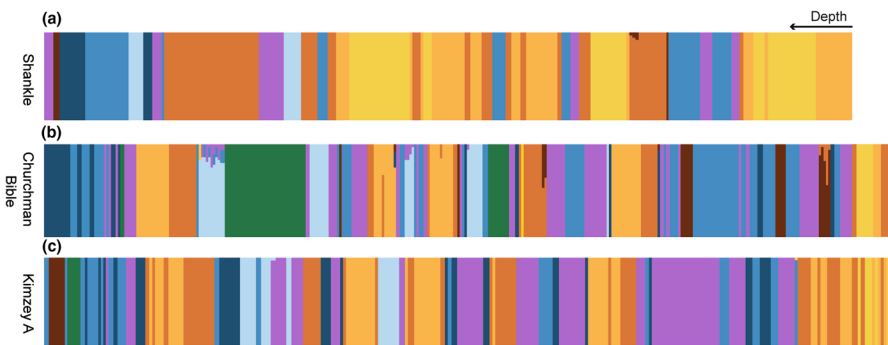


Fig. 11 Logs of vertical facies succession with depth for the proportion of predicted facies from the RF 11-input classifier for the Shankle (a), Churchman Bible (b), and Kimzey A (c) wells

to note is the consistency in the incorrectly predicted facies (i.e., false positives) being predicted for their adjacent facies for both the 7-input and 11-input classifiers; non-marine-dominated lithofacies were predicted as non-marine-dominated facies, and the identical scenario applies to marine-dominated facies. The value of the additional geologically derived inputs is thus appreciated in their capability to predict facies with less uncertainty and variability for a wide range of training and validation sets.

As previously introduced (Sect. 3.3), three test experiments were performed here to investigate the performance of the 11-input classifier under differing conditions of facies distribution. The uncertainty appears marginally higher in the Churchman Bible well (Fig. 11), which is in the marine-dominated zone in the southeast region of the field. Consequently, the vertical succession of the well is largely dominated by carbonate rocks. Identifying carbonate rocks based solely on petrophysical characteristics is sometimes a difficult task. Grain size might be difficult to distinguish between adjacent ranges (e.g., the distinction between very fine-grain carbonate and fine-grain carbonate might be difficult to quantify). Similarly, pore sizes within carbonates vary substantially, sometimes within a single lithofacies. These complexities in identifying

carbonate facies from observed geological interpretative features is presumably the explanation for a higher uncertainty in more complex wells (i.e., carbonate-dominated wells) such as Churchman Bible.

6 Conclusions

Machine learning classification incorporating geological knowledge and petrophysical properties of the rocks predicts facies in test wells with higher accuracy. It is believed that adding geological interpretative features (i.e., grain size, pore size, and argillaceous content) to the automated classification leads to a huge boost in test accuracy, which was proven to be robust to random noise and misinterpretation through noise injection procedures. Addition of a geologically constraining input was established to increase the test accuracy only marginally, while still reproducing the heterogeneity and variability inherent to the field. With regards to the heterogeneity and variability of lithofacies, the RF algorithm proved to be superior to the SVM algorithm. The former additionally predicts instances with greater test accuracy than the former. Lithofacies interpretation, whether it be automatic or manual, is associated with a large degree of uncertainty. One way to encompass and quantify the uncertainty in prediction was demonstrated here using different combinations of training and validation sets.

A few challenges remain with regards to this novel proposed automated classification approach for wells for which cores are not available. Ideas to build the additional feature inputs using geological inherent constraints (i.e., field facies distribution, NM_M, and well location) are being investigated. Furthermore, it is recognized that, if available, additional data should be considered in order to enhance the classification performance. Such data could be additional wireline logs (e.g., PEF, RHOB, SP) or additional geological characteristics (i.e., the mineralogy of each facies). We anticipate that the use of additional geological interpretative features, combined with conventional wireline log features, will lead to considerable improvement in the facies classification of cored wells.

Acknowledgements The authors would like to thank Matt and Brendon Hall for sharing the Council Grove Group dataset and the support vector machine code, and the ISPL team from the 2016 contest for their code, on which some of the scripts used in this study were based. The authors are grateful to the Institute of Petroleum Engineering Research Facilitation Budget for support. Valuable feedback from the IAMG 2018 conference presentation was greatly appreciated. Professional proofreading was provided by Malcolm Hodgskiss, and the authors would like to thank him for his contribution.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Brown CE (2012) Applied multivariate statistics in geohydrology and related sciences. Springer, Berlin
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297

- Cracknell MJ, Reading AM (2014) Geological mapping using remote sensing data: a comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Comput Geosci* 63:22–33
- Dubois MK, Byrnes AP, Bohling GC, Seals SC, Doveton JH (2003) Statistically-based lithofacies predictions for 3-D reservoir modeling: examples from the Panoma (Council Grove) Field, Hugoton Embayment, Southwest Kansas. In: *Proceedings of the American Association of Petroleum Geologists annual convention*, vol 12, p A44
- Dubois MK, Byrnes AP, Bohling GC, Doveton JH (2006) Multiscale geologic and petrophysical modeling of the giant Hugoton gas field (Permian), Kansas and Oklahoma, USA
- Dubois MK, Bohling GC, Chakrabarti S (2007) Comparison of four approaches to a rock facies classification problem. *Comput Geosci* 33(5):599–617
- Duda RO, Hart PE, Stork DG (2012) *Pattern classification*. Wiley, Hoboken
- Evans JS, Murphy MA, Holden ZA, Cushman SA, Drew CA, Wiersma YF, Huettmann F (2011) Modeling species distribution and change using random forest. In: Drew C, Wiersma Y, Huettmann F (eds) *Predictive species and habitat modeling in landscape ecology*. Springer, New York, pp 139–159
- Fawcett T (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27(8):861–874
- Fitch PJ, Lovell MA, Davies SJ, Pritchard T, Harvey PK (2015) An integrated and quantitative approach to petrophysical heterogeneity. *Mar Pet Geol* 63:82–96
- Friedman J, Hastie T, Tibshirani R (2001) *The elements of statistical learning*, vol 1(10). Springer series in statistics. Springer, New York
- Hall B (2016) Facies classification using machine learning. *Lead Edge* 35(10):906–909
- Hall M, Hall B (2017) Distributed collaborative prediction: results of the machine learning contest. *Lead Edge* 36(3):267–269
- Hartmann DJ (1999) Predicting reservoir system quality and performance. In: *Exploring for oil and gas traps*, vol 9, pp 1–154
- Heyer JF (1999) Reservoir characterization of the Council Grove Group, Texas County, Oklahoma
- James G, Witten D, Hastie T, Tibshirani R (2013) *An introduction to statistical learning*, vol 112. Springer, New York
- Jensen J, Lake LW, Corbett PW, Goggin D (2000) *Statistics for petroleum engineers and geoscientists*, vol 2. Gulf Professional, Houston
- Kanevski M, Maignan M (2004) *Analysis and modelling of spatial environmental data*, vol 6501. EPFL Press, Lausanne
- Kanevski M, Timonin V, Pozdnukhov A (2009) *Machine learning for spatial environmental data: theory, applications, and software*. EPFL Press, Lausanne
- Lin Y (2002) Support vector machines and the Bayes rule in classification. *Data Min Knowl Discov* 6(3):259–275
- Lundberg SM, Erion GG, Lee SI (2018) Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*
- Matheron G (1963) Principles of geostatistics. *Econ Geol* 58(8):1246–1266
- Olson TM, Babcock JA, Prasad KVK, Boughton SD, Wagner PD, Franklin MH, Thompson KA (1997) Reservoir characterization of the giant Hugoton gas field, Kansas. *AAPG Bull* 81(11):1785–1803
- Pelletier C, Valero S, Inglada J, Champion N, Marais Sicre C, Dedieu G (2017) Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sens* 9(2):173
- Puckette J, Boardman DR II, Al-Shaieb Z (1995) Evidence for sea-level fluctuation and stratigraphic sequences in the Council Grove Group (Lower Permian), Hugoton Embayment, Southern Mid-Continent
- Siemers WT, Ahr WM (1990) Reservoir facies, pore characteristics, and flow units: Lower Permian Chase Group, Guymon-Hugoton Field, Oklahoma. In: *SPE annual technical conference and exhibition*. Society of Petroleum Engineers
- Vapnik V (1995) *The nature of statistical learning theory*. Springer, Berlin