Real-Time Prediction of Equivalent Circulation Density for Horizontal Wells Using Intelligent Machines

Ahmed Alsaihati, Salaheldin Elkatatny,* and Abdulazeez Abdulraheem

ABSTRACT: Equivalent circulation density (ECD) is an important part of drilling fluid calculations. Analytical equations based on the conservation of mass and momentum are used to determine the ECD at various depths in the wellbore. However, these equations do not incorporate important factors that have a direct impact on the ECD, such as bottom-hole temperature, pipe rotation and eccentricity, and wellbore roughness. This work introduced different intelligent machines that could provide a real-time accurate estimation of the ECD for horizontal wells, namely, the support vector machine (SVM), random forests (RF), and a functional network (FN). Also, this study sheds light on how principal component analysis (PCA) can be used to reduce the dimensionality of a data set without loss of any important information. Actual field data of Well-1, including drilling surface parameters and ECD measurements, were collected from a 5–7/8 in. horizontal section to develop the models. The performance of the models was assessed in terms of root-mean-square error (RMSE) and coefficient of determination ($R^2$). Then, the best model was validated using unseen data points of 1152 collected from Well-2. The results showed that the RF model outperformed the FN and SVM in predicting the ECD with an RMSE of 0.23 and $R^2$ of 0.99 in the training set and with an RMSE of 0.42 and $R^2$ of 0.99 in the testing set. Furthermore, the RF predicted the ECD in Well-2 with an RMSE of 0.35 and $R^2$ of 0.95. The developed models will help the drilling crew to have a comprehensive view of the ECD while drilling high-pressure high-temperature wells and detect downhole operational issues such as poor hole cleaning, kicks, and formation losses in a timely manner. Furthermore, it will promote safer operation and improve the crew response time limit to prevent undesired events.

1. INTRODUCTION

Drilling fluid hydraulics has played an important role in well design when drilling a vertical or extended reach well. Therefore, accurate model and optimized drilling fluid hydraulics are crucial to allow engineers to properly design a well profile, thereby improving drilling efficiency, reducing risks, and decreasing nonproductive time (NPT). Additionally, it empowers the drilling crew to fully examine the downhole conditions and identify potential problems such as drill string washout, plugged nozzles, and the presence of a gas kick in the well.

Equivalent circulation density (ECD) is an important aspect of drilling fluid hydraulics which helps in avoiding kicks and drilling fluid losses, particularly in deep high-pressure high-temperature (HPHT) wells, where the temperature is significant and the margin between pore pressure and fracture pressure is minute. Drilled cuttings can increase the effective drilling fluid density and reduce fluid flow area, which in turn increases the value of the ECD. Thus, ECD calculations can be used as a baseline to monitor hole cleaning in real time while drilling.

Hydraulics programs have been widely used to evaluate the detailed hydraulic calculations that are required in planning, execution, and post-well review phases of drilling a well. The user has the option to select which rheological model (Bingham plastic, Power-law, Herschel–Bulkley) the hydraulic calculations are based on. Each rheological model, however, requires different data inputs, which can be obtained from a full lab test. These parameters are then fed into the software to compute the drilling fluid hydraulics and its associated parameters such as ECD. It has been observed that there is a discrepancy between computed data and the data recorded in the field.2

One factor contributing to the inconsistency between calculated and actual drilling hydraulics is the assumption
that the rheological properties of the drilling fluid are independent of pressure and temperature.\textsuperscript{1} This can be valid in shallow wells, where temperature changes are insignificant, and hence, the rheological variations are small. Furthermore, the mathematical equations used to calculate the drilling hydraulics entail a set of assumptions such as: (i) concentric annular and circular sections, (ii) laminar and turbulent flow, where plug flow is considered as laminar flow, while transition flow is neglected, and (iii) steady-state flow, where fluid properties at any single point in the system do not change. Such potentially unrealistic assumptions cannot be fulfilled in all drilling conditions.\textsuperscript{5}

In HPHT wells, evaluations and analyses of the effect of temperature and pressure on drilling hydraulics and kick probability are necessary.\textsuperscript{1,6} Rommeveit and Bjorkevoll\textsuperscript{1} developed two models, a static ECD model and a dynamic ECD model, which incorporate the temperature profile along the wellbore and allow the mud properties to be dependent on pressure and temperature. They concluded that the models made it possible to make realistic and reliable evaluations of any operational concerns during drilling especially when, for example, pit gain and standpipe pressure (SPP) deviate from the predictable value. Scheid et al.\textsuperscript{7} performed an experiment to evaluate the frictional pressure loss through pipes, annuli, and accessories such as pipe tool joint and stabilizers to estimate the drilling hydraulic calculations and the corresponding ECD with four types of drilling fluid. They stated that the results obtained in the study can be used in the drilling industry for accurate drilling hydraulic calculations.

Dokhani et al.\textsuperscript{3} concluded that eccentricity ($\epsilon$), which describes how off-center a pipe is in an open hole or casing, affects frictional pressure loss and hence the ECD for Herschel–Bulkley fluid. Eccentricity reduces overall frictional pressure loss if $\epsilon$ is $>0.1$ and it will be neglected if $\epsilon$ is $0.8$.\textsuperscript{8–10} Moreover, Dokhani et al.\textsuperscript{3} evaluated the effect of pipe rotations on frictional pressure loss. They observed that a drilling fluid’s apparent viscosity decreases as pipe rotation increases; hence, overall frictional loss decreases. This phenomenon is a result of the shear-thinning properties of Herschel–Bulkley fluid.\textsuperscript{11} Dokhani et al.\textsuperscript{3} also recommended considering the effect of roughness of the wellbore wall to have a more accurate estimation of drilling hydraulic calculations such as ECD.

A more accurate and reliable way to evaluate the ECD is to use a downhole pressure sensor, which consists of high-accuracy pressure gauges to measure the annular pressure. The sensors provide important real-time downhole pressure information which will allow the drilling crew to make a faster and better decision.\textsuperscript{12} However, such sensors are extremely expensive.\textsuperscript{13}

The literature review shows that major discrepancies exist between actual drilling hydraulic values and those predicted by previously accepted mathematical equations. In addition, several studies have suggested consideration of other factors, including pipe eccentricity, wellbore roughness, pressure and temperature, and pipe rotation speed to improve the accuracy of drilling hydraulic calculations. This, however, would mandate intensive mathematical intervention.

An alternative approach to estimating the ECD with higher accuracy is by using artificial intelligence (AI) and machine learning (ML).\textsuperscript{14,15} which map the inputs and outputs based on a defined algorithm.\textsuperscript{15} AI is efficient enough to minimize human efforts in various areas and does not necessitate fundamental knowledge of physics and science. AI has become an important subject in the drilling industry,\textsuperscript{15} where streaming data are continuously structured from the surface, downhole sensors and logging. AI and ML allow for new methods to learn from these data to mitigate drilling challenges, describe trends in real time, and automate drilling.\textsuperscript{16}

This study introduces three intelligent techniques to predict the ECD which can help in reducing the overall cost of drilling operation by not using downhole pressure sensors for future offset wells. The developed models would give the drilling crew the capability to monitor hole cleaning, maintain the wellbore pressures in horizontal extended-reach wells in real-time, and reduce the risk of formation fracture and collapse. Furthermore, they can improve the warning time to detect kicks to maintain well control.

1.1. Application of AI and ML in the Drilling Industry.

Drilling is a major aspect of the oil and gas industry which is considered to be highly expensive and risky. Therefore, tools that can improve the drilling operation at a low cost are essential. There are some applications related to AI and ML in drilling operations such as well planning, rate of penetration (ROP) optimization, well integrity, detecting problems, procedure decision making, and pattern recognition.

The large number of publications on the application of AI and ML in the drilling industry indicate that AI and ML can potentially reduce drilling costs and promote safety at the rig site.\textsuperscript{16,17}

Abdelgawad et al.\textsuperscript{13} used SPP, ROP, and mud weight (MW) as input parameters to predict the ECD using an artificial neural network (ANN) and an adaptive neuro-fuzzy inference system (ANFIS) in an 8–1/2 in. vertical hole section. The two models predicted the ECD with a correlation coefficient ($R$) of 0.99 and AAPE of 0.22%. Alkinani et al.\textsuperscript{14} collected data from more than 2000 wells located around the world and used an ANN to build a model to predict the ECD. The input parameters for the model were flow rate, MW, nozzles total flow area, plastic velocity, revolutions per minute, weight on bit (WOB), and yield point. Bayesian Regularization (BR) was used as a training algorithm because it had the highest $R^2$ (0.982).

Alsabaa et al.\textsuperscript{18} used ANFIS to predict the rheological properties of invert emulsion mud in real time. Al-Yami et al.\textsuperscript{19} used Bayesian Belief Network (BBN) to establish an intelligent drilling system based on different fluid and reservoir properties. This tool can be utilized to train young engineers in different drilling perspectives such as well control, underbalanced drilling, drilling fluid, and cementing best practices. Ahmadi\textsuperscript{20} simulated the performance of various types of drilling fluid’s rheology under different conditions using a support vector machine (SVM) with good agreement between lab results and prediction. Elkatatny et al.\textsuperscript{21} used an ANN, which incorporates drilling fluid and drilling mechanical properties, to predict ROP with high accuracy. Elkatatny et al.\textsuperscript{22} used actual field data to build an ANN to predict the top-depth of four geological formations in real time while drilling. The best results were obtained when scaled conjugate gradient back-propagation (TRAINSCG) with 20 neurons was used. The input parameters used in the study to develop the model were the mechanical drilling parameters including ROP, Q, RS, SPP, torque, and WOB.

1.2. Support Vector Machines. SVMs are supervised ML models that analyze data for classification or regression problems\textsuperscript{17} and can lead to a high performance in particular
samples from the original data set are not included. These samples are made from the data set, about one-third of the observations. It splits nodes in each tree considering a limited number of variables.30

The feature vectors from input space to kernel space for linearly nonseparable data sets, the kernel matrix computation requires computational resources. The popular kernel functions are27 linear kernel function, polynomial kernel function, Gaussian RBF, and randomized blocks analysis of variance (ANOVA RB) kernel.

The selection of kernel functions is essentially dependent on the nature of the data set. The linear kernel ranks behind the polynomial kernel, and it is useful in large sparse data vectors. On the other hand, the polynomial kernel is commonly used in image processing. While the ANOVA RB kernel is generally used for regression tasks, the Gaussian RBF is mostly applied if the user lacks prior knowledge.27

1.3. Random Forests. The RF is an ensemble learning technique that can be used for classification and regression problems.28 The RF combines hundreds or thousands of decision trees and trains each one on a slightly different set of observations. It splits nodes in each tree considering a limited number of features, in a process called bootstrapping. The final prediction of the RF is made by averaging the predictions of each tree in a process called aggregation. When a bootstrap sample is made from the data set, about one-third of the samples from the original data set are not included.30 These samples are called out-of-bag data and are used to measure internally the accuracy of the RF. The RF outperforms a single decision tree because of its ability to limit overfitting without substantially increasing the margin of error.29

1.4. Functional Network. A functional network (FN) is a generalization of an ANN, which can be accomplished by using multiple arguments and learnable functions, that is, in a FN, the activation functions associated with neurons are not fixed but learned from data.31 In an ANN, the weights associated with the neurons must be learned, while they are suppressed in a FN.25 Another characteristic of a FN is that the specification of the initial topology could be based on the features of the problem. Therefore, knowledge about the problem can help in developing the network structure.

1.5. Principal Component Analysis. PCA is a useful statistical technique that has many applications in many fields and is a common technique for finding patterns in a data set with high dimensions.32 The objective of PCA is to reduce the dimensionality of a large data set of variables or features into a smaller one without the loss of any important information.33 PCA finds a lower dimensional space (W) to transform the data (X = [x1, x2, ..., xN]) from a higher dimensional space (R^N) to a lower dimensional space (R^M), where N represents the total number of observations (rows in a data set) and xi represents the ith observation. Each observation is represented by M features or variables (columns in a dataset), that is, each observation is represented as a point in M-dimensional space.34

A PCA space for a data set that contains a number of features K has K principal components. These K principal components are uncorrelated, orthonormal, and represent the direction of the maximum variance.35 The first component, (PC1 ∈ R^M), always represents the maximum variation of the data; (PC2 ∈ R^M) represents the second-largest variation of the data, while (PCn ∈ R^M) represents the least variation of the data.32,33 The principal components that represent more than 90% of the variation in a data set are often considered.

The following sequence shows how to calculate PCs using the Covariance Matrix Method for a given data set (X = {x1, x2, ..., xN}):

- Compute the mean of all variables
- Subtract the mean of each variable from all observations corresponding to that variable
- Compute the covariance matrix for the centered data set
- Compute the eigenvectors and eigenvalues of the covariance matrix
- Sort eigenvectors according to their corresponding eigenvalues
- Select the eigenvectors that have the largest eigenvalues. The selected eigenvectors represent the projection space of PCA
- Project all observations on the lower dimensional space of PCA

2. METHODOLOGY

2.1. Data Collections. Two types of actual field data of Well-1 were collected from a 5–7/8 in. horizontal section: (i) drilling surface parameters and (ii) ECDs. The drilling surface parameters including flow rate (Q), hook-load (HL), ROP, rotary speed (RS), SPP, WOB, and surface drilling torque (T) were obtained from surface real-time transmitter sensors, while ECDs were obtained from a pressure-while-drilling (PWD) sensor. A total of 3567 data points of ECD were obtained at the same depth of the drilling surface parameters. Table 1 shows the statistical parameters of the whole data set. Q ranges from 250 to 296.5 GPM; HL ranges from 256 to 286 klf; ROP ranges from 3.5 to 59.6 ft/h; RS ranges from 59 to 141 RPM; SPP ranges from 59 to 2354.8 psi; WOB ranges from 5.1 to 2354.8 klf; T ranges from 2.9 to 83.4 klf.lbf; ECD ranges from 0.96 to 83.4.

| statistical parameters | Q (gal/min) | HL (klf) | ROP (ft/h) | RS (RPM) | SPP (psi) | WOB (klf) | T (klf.lbf) | ECD (Pcf) |
|------------------------|-------------|----------|------------|-----------|-----------|-----------|------------|-----------|
| minimum                | 250.0       | 256.0    | 3.5        | 59.0      | 2354.8    | 5.1       | 2.9        | 83.4      |
| maximum                | 296.5       | 286.0    | 59.6       | 141.0     | 3656.5    | 20.1      | 10.0       | 95.5      |
| mean                   | 276.7       | 267.4    | 23.0       | 119.7     | 3031.7    | 15.2      | 6.9        | 90.4      |
| standard deviation     | 10.3        | 5.5      | 6.2        | 17.1      | 257.7     | 3.0       | 1.2        | 3.2       |
| skewness               | -1.66       | 0.62     | 0.18       | -0.95     | -0.15     | -0.96     | -0.04      | -0.37     |
| kurtosis               | 1.09        | 0.16     | 1.76       | 1.34      | -0.11     | 0.10      | -0.85      | -0.90     |
RPM; SPP ranges from 2354.8 to 3656.5 psi; WOB ranges from 5.1 to 20.1 kbf; T ranges from 2.9 to 10 kft.lbf; and the ECD ranges from 83.4 to 95.5 pcf.

Figure 1 compares the linear relationship of the input parameters used to train the model with the ECD. Figure 1 shows that HL, SPP, and T have a strong relationship with the ECD, that is, 0.71, 0.87 and 0.85, respectively. Q has a moderate linear relationship with an ECD of −0.38. On the other hand, ROP, RPM, and WOB have relatively a low relationship of −0.13, 0.1, and 0.11, respectively.

2.2. Data Splitting. In ML, it is necessary to build a model that makes accurate predictions for future data. Thus, the data set is divided into two portions, training and testing sets. The training set is used to ensure that the machine recognizes patterns in the data set, while the testing set is used to evaluate how well the machine can predict unseen data based on its training.

In this analysis, seven surface drilling parameters were used as independent variables (inputs): Q, HL, ROP, RS, SPP, WOB, and T, while ECD was used as a dependent variable (output). The data set was randomly split, with a ratio of 80:20. Eighty percent of the data was selected for training to ensure that the models capture most of the ECD variation while drilling with various surface drilling parameters. The training set has 2742 data points, while the testing set has 827 data points. Tables 2 and 3 show the statistical parameters of the training and testing sets, respectively.

2.3. Model Building. Python library’s Scikit-Learn was used to build the SVM model. The SVM parameters, known as hyperparameters (e.g., regularization parameter C, gamma, and kernel type), were tuned using a built-in function in Scikit-Learn known as GridSearchCV to evaluate the improvement and performance of the SVM. Two types of kernel, RBF and linear were tried while varying the value of C from 0.001 to 1000 and the type of gamma (i.e. auto and scale). Likewise, Scikit-Learn was used to develop the RF model. The model parameters, including the maximum depth of the tree “max_depth”, the maximum features to be considered when splitting the node in each tree “max_features”, and the number of the trees in the forest “n_estimators”, were tuned using GridSearchCV. Different values of max_depth from 3 to 21 and three types of max_features (i.e., auto, sqrt, and log_) were tried while varying the n_estimator from 3 to 150. The rest of the model parameters, including minimum_samples_split, minimum_sample_leaf, maximum_leaf_nodes, and minimum_impurity, were kept as default values. MATLAB code was used to build the FN model. Two methods, FN Forward-Backward Method (FNFBM) and FN Exhaustive-Backward Method (FNEBM), were studied with two types of relationship: linear and nonlinear.

2.4. Using PCA for Dimensionality Reduction. Standardizing a data set refers to shifting the distribution of each variable to have a unit scale, that is, a mean of zero and a standard deviation of one, which is a necessary step for the PCA algorithm. The values of the input parameters were standardized using the following equation

\[ Y = \frac{X - \bar{X}}{\sigma} \]  

where Y is the normalized input parameter, X is the input parameter to be normalized, and σ is the standard deviation of the input variable.

Then, Python library’s Scikit-Learn was used to apply the PCA algorithm to the input variables of the data set of Well-1. The transformation to PCA space was completed in three steps: (i) instantiate the PCA by passing the number of principal components to the constructor, (ii) call the fit, which will find the covariance matrix, the eigenvectors, and eigenvalues of the covariance matrix, and (iii) transform the data set into the PCA space. Then, the transformed data set was fed into another RF model for training and testing. The PCA-based RF model improvement and performance while using different numbers of PCs were evaluated.

3. RESULTS AND DISCUSSION

3.1. Model Assessment. The SVM model performance with different parameter type was presented in Table 4. The optimal parameters for each kernel type were obtained using GridSearchCV. Table 4 shows that the SVM model with the RBF kernel had the highest \( R^2 \) and the lowest root-mean-square error (RMSE) compared to the linear kernel. The SVM with the RBF kernel predicted the ECD with an \( R^2 \) of 0.97 and an RMSE of 0.54 in the training set, while the \( R^2 \) and RMSE were 0.97 and 0.58, respectively, in the testing set. On the other hand, the SVM with the linear kernel predicted the ECD with an \( R^2 \) of 0.95 and an RMSE of 0.71 in the training set and with an \( R^2 \) and RMSE of 0.95 and

Table 2. Statistical Parameters of the Training Set (2742 Data Points)

| statistical parameters | Q (gal/min) | HL (kbf) | ROP (ft/h) | RS (RPM) | SPP (psi) | WOB (kbf) | T (kft.lbf) | ECD (Pcf) |
|------------------------|-----------|---------|-------------|---------|---------|---------|-----------|-----------|
| minimum                | 249.4     | 256.1   | 3.5         | 59.0    | 2379.7  | 5.5     | 3.7       | 83.4      |
| maximum                | 296.6     | 285.2   | 59.6        | 141.3   | 3632.1  | 20.0    | 10.0      | 95.5      |
| mean                   | 276.7     | 267.4   | 23.0        | 119.8   | 3035.3  | 15.2    | 6.9       | 90.4      |
| standard deviation     | 10.3      | 5.5     | 6.2         | 16.9    | 258.0   | 3.0     | 1.2       | 3.2       |
| skewness               | −1.67     | 0.61    | 0.22        | −0.93   | −0.15   | −0.96   | −0.05     | −0.39     |
| kurtosis               | 1.11      | 0.12    | 1.88        | 1.28    | −0.14   | 0.08    | −0.87     | −0.89     |

https://dx.doi.org/10.1021/acsomega.0b05570
ACS Omega 2021, 6, 934−942

937
The RF model predicted the ECD with an RMSE of 0.23 and $R^2$ of 0.99 in the training set and with an RMSE of 0.42 and $R^2$ of 0.99 in the testing set. The optimum model parameters are shown in Table 5. Figure 3a,b are cross-plots of the actual and predicted ECD of the training and testing sets, respectively.

The FN model performance with different methods and relationship types is presented in Table 6. Table 6 shows that the best results were obtained when FNxEBM was used with a nonlinear relationship type. The model predicted the ECD with an RMSE of 0.44 and $R^2$ of 0.99 in the training set and with an RMSE of 0.45 and $R^2$ of 0.99 in the testing set. Figure 4a,b denotes cross-plots of the actual and predicted ECD of the training and testing sets, respectively, when FNxEBM with a nonlinear relationship was used.

The optimum results of each model are summarized in Table 7. Table 7 shows that the RF is the most accurate model in estimating the ECD with a low RMSE and high $R^2$ in the training and testing sets. The FN was the second most accurate model, followed by the SVM as the least accurate model.

3.2. Validation of the Developed RF. The best model, RF, was selected to be validated using a total of 1152 unseen data points collected from Well-2. The compatibility of the input parameters in Well-2 was checked to ensure they are in the same range as the data set, which was used to train the RF model. Table 8 shows the statistical parameters of the data set of Well-2. The RF predicted the ECD in Well-2 with an RMSE of 0.35 and $R^2$ of 0.95. Figure 5 shows the actual and predicted ECD as a function of the depth index of Well-2. The depth index was used to hide the actual depth of the drilled section.

3.3. PCA for Dimensionality Reduction. The input variables of the original data set collected from Well-1 were standardized using eq 1. Then, the data set was transformed to PCA space using Python library’s Scikit-Learn. Table 9 shows a sample of the training set after transformation to PCA space. It is important to study the variation that each PC accounts for in the data set to perform dimensionality reduction. Figure 6 is a screen plot, which is a graphical representation of the percentages of the variation that each PC accounts for in the whole data set. The first thing to notice is that the first principal component PC1 accounts for 36.42% of the variation in the data set; the second principal component PC2 accounts for 26.68%; the third principal component PC3 accounts for 17.30%; the fourth principal component PC4 accounts for 13.01%; the fifth principal component PC5 accounts for 4.47%; the sixth principal component PC6 accounts for 1.29%, and the seventh principal component PC7 accounts for 0.75%. This...
means that PC1, PC2, PC3, and PC4 directions collectively explain 93.41% of the total variation in the data set, while PC5, PC6, and PC7 combined explain only 6.51% of the total variation in the data set. Thus, even though the points in the data set form a cloud in a dimensional space ($R^7$), PCA shows that these points cluster near a four-dimensional plane ($R^4$) spanned by PC1, PC2, PC3, and PC4.

The contribution of each variable (Q, HL, ROP, RS, SPP, WOB, or T) in each principal component is presented in Table 10. How all of this needs to be interpreted? For example, in studying PC1, the second entry "0.58", HL, is the largest, which means a change in one unit of HL tends to effect the ECD more than a change of one unit of Q, ROP, RS, SPP, WOB, or T. The first entry "0.52", which corresponds to Q, is the next most important factor in determining the ECD. On the other hand, the second last entry "0.14", WOB, is the least important.
in determining the ECD. Similarly, other PCs can be interpreted.

The PCA-based RF model performance in predicting the ECD in the same testing set used in Section 3.1, while using different numbers of PCs, was evaluated and compared with the SVM and FN, which were trained and tested using 100% variation of the data set (i.e., seven dimensions) as discussed in Section 3.1. Table 11 shows that as the number of PCs for the data set increases from one principal component to four principal components, the $R^2$ and RMSE improve. However, when more than four principal components were considered, RMSE increased. The PCA-based RF with only four principal components outperformed the SVM, with an RMSE of 0.54 and $R^2$ of 0.97. In addition, the PCA-based RF with only four principal components performed almost similar to the FN that had an RMSE of 0.45 and $R^2$ of 0.99. Furthermore, even if only two dimensions PC1 and PC2 (63.10% of variation) were considered, the PCA-based RF would perform almost similar to the SVM with an RMSE of 0.63 and $R^2$ of 0.96. In other words, the PCA-based RF model with four inputs (i.e., PC1, PC2, PC3, and PC4) performed better than the SVM and almost similar to the FN, which were trained with seven inputs (i.e., Q, HL, ROP, RS, SPP, WOB, and T). This showed how the PCA technique is powerful enough to transform the data set from a higher dimensional space to a lower dimensional space without loss of any important information, which in turn increases the speed of the training process for a model.

![Figure 5. Actual and predicted ECD as a function of the depth index of Well-2.](image)

![Figure 6. Percentages of the variation that each PC accounts for in the whole data set.](image)

**Table 9. Sample of the Training Set after Transformation to PCA Space**

| sample | PC1   | PC2   | PC3   | PC4   | PC5   | PC6   | PC7   | ECD (Pcf) |
|--------|-------|-------|-------|-------|-------|-------|-------|-----------|
| 1      | 0.0157| 5.6083| −0.9000| −1.1646| 0.5203| −0.5557| −0.5946| 83.57     |
| 2      | −0.1004| 5.2729| 0.1541| −1.5667| 1.3405| −0.0313| −0.1090| 83.45     |
| 3      | 0.2006| 4.9792| −0.1247| −1.5161| 1.4103| 0.5679| 0.1762| 83.42     |
| 4      | 0.4135| 4.8381| −0.1609| −1.3059| 1.4539| 0.7850| 0.2212| 83.44     |
| 5      | −0.3242| 5.0952| 0.0931| −1.2110| 1.4730| −0.0422| 0.0816| 83.47     |

**Table 10. Contribution of Variables in Each Principal Component**

| variable | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
|----------|-----|-----|-----|-----|-----|-----|-----|
| Q        | 0.52| 0.27| 0.10| 0.27| 0.48| 0.03| 0.58|
| HL       | 0.58| 0.03| 0.27| 0.00| 0.12| 0.47| 0.59|
| ROP      | 0.19| 0.30| 0.08| 0.89| 0.15| 0.23| 0.01|
| RS       | 0.26| 0.53| 0.31| 0.11| 0.73| 0.10| 0.01|
| SPP      | 0.25| 0.61| 0.02| 0.30| 0.35| 0.34| 0.48|
| WOB      | 0.14| 0.14| 0.85| 0.05| 0.26| 0.37| 0.17|
| T        | 0.45| 0.41| 0.29| 0.15| 0.09| 0.68| 0.23|

**Table 11. Comparison of PCA-Based RF Performance with FNN and SVM Using Different PCs**

| no. PC | variation % | PCA-RF testing | FN testing | SVM testing |
|--------|-------------|----------------|------------|-------------|
| PC1    | 36.42       | 1.78 0.70      | 0.45 0.99  | 0.58 0.97   |
| PC1 and PC2 | 63.10       | 0.63 0.96      | 0.45 0.99  | 0.58 0.97   |
| PC1 to PC3 | 80.40       | 0.62 0.96      | 0.45 0.99  | 0.58 0.97   |
| PC1 to PC4 | 93.41       | 0.54 0.97      | 0.45 0.99  | 0.58 0.97   |
| PC1 to PC5 | 97.88       | 0.57 0.97      | 0.45 0.99  | 0.58 0.97   |
| PC1 to PC6 | 99.17       | 0.59 0.97      | 0.45 0.99  | 0.58 0.97   |
| PC1 to PC7 | 100         | 0.60 0.97      | 0.45 0.99  | 0.58 0.97   |
4. CONCLUSIONS

Nowadays, modern drilling involves multiple interconnecting activities. Therefore, obtaining real-time information about ongoing operations is crucial for safe and efficient drilling operations. Three intelligent machines SVM, RF, and FN were developed to predict the ECD while drilling horizontal wells. PCA was applied to reduce the data set dimensions and compare the result of the PCA-based RF with the SVM and FN. Based on the results, the following can be concluded:

- The RF achieved the lowest RMSE and the highest $R^2$ compared to the FN and SVM. The RF predicted the ECD with an RMSE of 0.23 and $R^2$ of 0.99 in the training set and with an RMSE of 0.42 and $R^2$ of 0.98 in the testing set.
- The FN achieved the second-lowest RMSE of 0.44 and 0.45 in the training and testing sets, respectively. The $R^2$ was as high as the RF of 0.99 in the training and testing sets.
- The SVM with the kernel type had the highest RMSE and the lowest $R^2$ when compared with that of the RF and FN models. The SVM with the optimum parameters predicted the ECD with an RMSE of 0.54 and $R^2$ of 0.97 in the training set, while the RMSE and $R^2$ were 0.58 and 0.97, respectively, in the testing set.
- The RF predicted the ECD in Well-2 with an RMSE of 0.35 and $R^2$ of 0.95.
- The PCA-based RF model with only four principal components outperformed the SVM with an RMSE of 0.54 and $R^2$ of 0.97.

■ AUTHOR INFORMATION

Corresponding Author

Salaheldin Elkatatny — Department of Petroleum Engineering, College of Petroleum & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; orcid.org/0000-0002-7289-3715; Email: elkatatnyy@kfupm.edu.sa

Authors

Ahmed Alsaihati — Department of Petroleum Engineering, College of Petroleum & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

Abdulazeez Abdulraheem — Department of Petroleum Engineering, College of Petroleum & Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.0c05570

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This research received no external funding.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to acknowledge King Fahd University of Petroleum & Minerals (KFUPM) for utilizing the research facilities in conducting this research work. In memoriam of Dr. Hisham Nasr-El-Din, my Ph.D. advisor who taught me how to write strong technical papers.

■ REFERENCES

(1) Rommeveit, R.; Bjorkevoll, K. S. Temperature and Pressure Effects on Drilling Fluid Rheology and ECD in Very Deep Wells. Presented at the 1997 SPE/IADC Middle East Drilling Technology Conference, Manama, Bahrain; Society of Petroleum Engineers, November, 1997. SPE/IADC 39282. pp 1–10.
(2) Maglione, R.; Robotti, G. A Computer Program to Predict Stand Pipe Pressure While Drilling Using the Drilling Well as Viscometer. Presented at the Eleventh Petroleum Computer Conference, Datta, Texas, USA; Society of Petroleum Engineers, June, 1996. SPE 35994. pp 1–11.
(3) Dokhani, V.; Pordel, S. M.; Karimi, M.; Salehi, S. Evaluation of Annular Pressure Losses While Casing drilling. Presented at the SPE Annual Technical Conference and Exhibition, New Orleans, USA; Society of Petroleum Engineers, September, 2013. SPE 166103. pp 1–15.
(4) Ergie, O.; Vajargah, A. K.; Ozbayoglu, M. E.; Van Oort, E. Improved ECD Prediction and Management in Horizontal and Extended Reach Wells with Eccentric Drill strings. Presented at the SPE/IADC Drilling Conference and Exhibition, Fort Worth, Texas, USA; Society of Petroleum Engineers, 1-3 March, 2016. IADC/SPE-178785-Ms. pp 1–21.
(5) Raba, H. Well Engineering and Construction; Entrac Petroleum: London, 2001; pp 1–650.
(6) Isambourg, P.; Anfinsen, B. T.; Marken, C. Volumetric Behavior of Drilling Muds at High Pressure and High Temperature. Presented at the 1996 SPE European Petroleum Conference, Milan, Italy; Society of Petroleum Engineers, 22-24 October, 1996. SPE 36830. pp 1–9.
(7) Scheid, C. M.; Cacada, L. M.; Rocha, D. P.; Araha, P. E.; Arago, A. F., Martins, A. L. Prediction of Pressure Losses in Drilling Fluid Flows in Circular and Annular Pipes and Accessories. Presented at the Latin American and Caribbean Petroleum Engineering, Conference, Cartagena, Colombia; Society of Petroleum Engineers, May, 2009. SPE 122072. pp 1–13.
(8) Cartalos, U.; Dupuis, D. An Analysis Accounting for The Combined Effect of Drill string Rotation and Eccentricity on Pressure Losses in Slim Hole Drilling. Presented at the SPE/IADC Drilling Conference, Amsterdam, Netherlands; Society of Petroleum Engineers, February, 1993. SPE 25769. pp 1–11.
(9) Escudier, M. P.; Gouldson, I. W.; Oliveira, P. J.; Pinho, F. T. Effects of inner cylinder rotation on laminar flow of a Newtonian fluid through an eccentric annulus. Int. J. Heat Fluid Flow 2000, 21, 92–103.
(10) Mokhtari, M.; Ermila, M. A.; Tutuncu, A. N.; Karimi, M. Computational Modeling of Drilling Fluids Dynamics in Casing Drilling. Presented at the SPE Eastern Regional Meeting, Lexington, Kentucky, USA; Society of Petroleum Engineers, October, 2012. SPE 161301. pp 1–13.
(11) Hemphill, T.; Ravi, K. Calculation of Drill Pipe Rotation Effects on Axial Flow: An Engineering Approach. Presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas; Society of Petroleum Engineers, October, 2005. SPE 97158. pp 1–4.
(12) Halliburton.com. 2020, https://www.halliburton.com/content/dam/ps/public/ss/contents/Data_Sheets/web/Pressure-While-Drilling-Sensor-Datasheet.pdf?nav=en-US_sperry_public [accessed 10 Oct 2020].
(13) Abdelgawad, K. Z.; Elzenary, M.; Elkatatny, S.; Mahmoud, M.; Abdulraheem, A.; Patil, S. New Approach to Evaluate the Equivalent Circulating Density (ECD) Using Artificial Intelligence Techniques. J. Pet. Explor. Prod. Technol. 2018, 9, 1569–1578.
(14) Alkini, H. H.; Al-Hameedi, A.; Dunn-Noman, S.; Al-Alwani, M. A.; Mutar, R. A.; Al-Bazza, W. H. Data-Driven Neural Network Model to Predict Equivalent Circulation Density ECD. Presented at the SPE Gas and Oil Technology Showcase and Conference, Dubai, UAE;
Society of Petroleum Engineers, 21-23 October, 2019. SPE-198612. pp 1–9.

(15) Balaji, K.; Rabiei, K.; Suicmez, V.; Canbaz, C.; Agharzeyva, Z.; Tek, S.; Bulut, U.; Temizel, C. Status of Data-Driven Methods and Their Applications in Oil and Gas Industry. In *Proceeding of the SPE Europe featured at 80th EAGE Conference and Exhibition, Copenhagen, Denmark*; Society of Petroleum Engineers, 11-14 June, 2018. SPE-190812-MS.

(16) Alsaihati, A.; Elkatatny, S.; Mahmoud, A. A.; Abdulraheem, A. Use of Machine Learning and Data Analytics to Detect Downhole Abnormalities While Drilling Horizontal Wells, With Real Case Study. *J. Energy Resour. Technol.* 2020, 143, 043201.

(17) Bello, O.; Holzmann, J.; Yaqoob, T.; Teodoriu, C. Application of Artificial Intelligence Methods in Drilling System Design and Operations: A Review of The State of The Art. *J. Artif. Intell. Soft Comput. Res.* 2015, 5, 121–139.

(18) Alsabaa, A.; Gamal, H.; Elkatatny, S.; Abdulraheem, A. Real-Time Prediction of Rheological Properties of Invert Emulsion Mud Using Adaptive Neuro-Fuzzy Inference System. *Sensors* 2020, 20, 1669.

(19) Al-Yami, A. S.; Al-Shaarawi, A.; Al-Bahrani, H.; Wagle, V. B.; Al-Gharbi, S.; Al-Khudairi, M. B. Using Bayesian Network to Develop Drilling Expert Systems. *Presented at the SPE Heavy Oil Conference and Exhibition, Kuwait City, Kuwait*; Society of Petroleum Engineers, 6-8 December, 2016. SPE-184168. pp 1–21.

(20) Ahmedi, M. A. Toward Reliable Model for Prediction Drilling Fluid Density at Wellbore Conditions: A LSSVM model. *Neurocomputing* 2016, 211, 143–149.

(21) Elkatatny, S. M.; Tariq, Z.; Mahmoud, M. A.; Al-AbdulJabbar, A. Optimization of Rate of Penetration Using Artificial Intelligence Technique. *Presented at the 51st US Rock Mechanics/Geomechanics Symposium, San Francisco, California, USA*, 25-28 June, 2017. ARMA-17-771. pp 1–8.

(22) Elkatatny, S.; Al-AbdulJabbar, A.; Al-AbdulJabbar, A.; Mahmoud, A. A. New Robust Model to Estimate Formation Tops in Real Time Using Artificial Neural Networks (ANN). *Petrophysics* 2019, 60, 825–837.

(23) Hearst, M. A.; Dumais, S. T.; Osuna, E.; Platt, J.; Scholkopf, B. Support Vector Machines. *IEEE Intell. Syst. Their Appl.* 1998, 13, 18–28.

(24) Vapnik, V. *The Nature of Statistical Learning Theory*, 2nd ed; Springer: New York, pp 1–314, 1995.

(25) Anifowose, F.; Abdulraheem, A. Fuzzy logic-driven and SVM-driven hybrid computational intelligence models applied to oil and gas reservoir characterization. *J. Nat. Gas Sci. Eng.* 2011, 3, 505–517.

(26) Durka, B. A Classification Algorithm Using Mahalanobis Distance Clustering of Data with Applications on Biomedical Data Sets. Msc Dissertation; Middle East Technical University: Ankara, Turkey, 2011.

(27) Awad, M.; Khanna, R. Support Vector Machines for Classification. In: *Efficient Learning Machines*; Apress: Berkeley, CA, 2015.

(28) Zhang, C.; Ma, Y. *Ensemble Machine Learning*; Springer US: Boston, MA, 2012.

(29) Hegde, C.; Gray, K. E. Use of Machine Learning and Data Analytics to Increase Drilling Efficiency for Nearby Wells. *J. Nat. Gas Sci. Eng.* 2017, 40, 327–335.

(30) Breiman, L. Random Forests. *Mach. Learn.* 2001, 45, 5–32.

(31) Iglesias, A.; Gámez, A. Applying Functional Networks to Fit Data Points from B-Spline Surfaces. *Proceedings Computer Graphics International (CGI’01)*, 2001, pp 329–332.

(32) Smith, L. I. A Tutorial on Principal Component Analysis. 2002, http://www.cs.otago.ac.nz/cosc453/studenttutorials/principalcomponents.pdf (accessed 2020-07-25).

(33) Tharwat, A. Principal component analysis - a tutorial. *Int. J. Appl. Pattern Recognit.* 2016, 3, 197.

(34) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemometr. Intell. Lab. Syst.* 1987, 2, 37–52.