Advances in Geo-Energy Research

Vol. 9, No. 3, p. 172-184, 2023

Original article

Variable interaction empirical relationships and machine learning provide complementary insight to experimental horizontal wellbore cleaning results

David A. Wood*

DWA Energy Limited, Lincoln LN59JP, United Kingdom

Keywords:
Hole-cleaning factors
cuttings carrying performance
cuttings transport feature importance
optimized empirical relationships
cuttings-bed concentrations

Cited as:
Wood, D. A. Variable interaction empirical relationships and machine learning provide complementary insight to experimental horizontal wellbore cleaning results. Advances in Geo-Energy Research, 2023, 9(3): 172-184.
https://doi.org/10.46690/ager.2023.09.05

Abstract:
Long horizontal wellbore sections are now a key requirement of oil and gas drilling, particularly for tight reservoirs. However, such sections pose a unique set of borehole-cleaning challenges which are quite distinct from those associated with less inclined wellbores. Experimental studies provide essential insight into the downhole variables that influence borehole cleaning in horizontal sections, typically expressing their results in multivariate empirical relationships with dimensionless cuttings bed thickness/concentration ($H\%$). This study demonstrates how complementary empirical $H\%$ relationships focused on pairs of influential variables can be obtained from published experimental data using interpolated trends and optimizers. It also applies five machine learning algorithms to a compiled multivariate (10-variable) interpolated dataset to illustrate how reliable $H\%$ predictions can be derived based on such information. Seven optimizer-derived empirical relationships are derived using pairs of influential variables which are capable of predicting $H\%$ with root mean squared errors of less than 1.8%. The extreme gradient boosting model provides the lowest $H\%$ prediction errors from the 10-variable dataset. The results suggest that in drilling situations where sufficient, locally-specific, information for multiple influential variables is available, machine learning methods are likely to be more effective and reliable at predicting $H\%$ than empirical relationships. On the other hand, in drilling conditions where information is only available for a limited number of influential variables, empirical relationships involving pairs of influential variables can provide valuable information to assist with drilling decisions.

1. Introduction

Long horizontal wellbore sections are now recognized as the most effective way to develop tight reservoirs, and the length of the horizontal sections drilled is increasing particularly in certain shale formations to optimize resource recovery. Hole cleaning and drilling cuttings transport is a challenge for wellbores of all configurations and inclinations due to the large number of variables that influence it (Li and Walker, 2001). For horizontal wellbore sections, it is a particular challenge because cuttings beds tend to form more readily on the lower side of the wellbore due to gravitational forces and often slow down cuttings transport (Sun et al., 2013). Such cuttings beds, if they are allowed to remain in the horizontal sections exert high frictional and torque forces on the drill pipe leading to inefficiency and slow rates of penetration (ROP) (Mahmoud et al., 2020a).

As the length of horizontal sections increases so do the impacts of the cuttings beds on borehole cleaning efficiency and the ability to move logging tools and completion equipment in and out of a wellbore (Nazari et al., 2010). Inefficient borehole cleaning leading to thick cuttings beds accumulating in horizontal sections is a major cause of drilling problems and non-productive time (Power et al., 2000). Potential problems include the drill string becoming stuck and "pack-offs" blocking the circulation of the drilling fluid, which can damage subsurface equipment. High drill-string torque/friction

Yandy Scientific Press

*Corresponding author.
E-mail address: dw@dwasolutions.com (D. A. Wood).

Received August 10, 2023; revised August 25, 2023; accepted September 10, 2023; available online September 13, 2023.

associated with cuttings-bed buildups is responsible for drill-string wear that can ultimately lead to its failure (Busch and Johansen, 2020).

Borehole cleaning is influenced by multiple factors including the fluid flow rate and flow regime (turbulent, laminar, etc.) in the borehole annulus, drill-pipe rotation speed, drilling fluid density, wellbore inclination, drill-pipe eccentricity in the borehole, drilling fluid rheology, rate of penetration, flow regime, annulus diameter relative to borehole diameter, drilling cuttings size, shape, and density (Moroni et al., 2009; Wang et al., 2013; Zico et al., 2023). Models dating back to the 1950s have been proposed to explain the cuttings-carrying capacity of drilling fluids (Williams and Bruce, 1951) incorporating the key factors that influence borehole cleaning (Hopkin, 1967).

The empirical models used most widely in drilling operations to provide real-time indications of hole cleaning performance, as reviewed by Mahmoud et al. (2020b) and Al-Rubaii et al. (2023) incorporate many of the influential factors mentioned. However, such models struggle to incorporate all the influential factors relevant to specific locations, and most do not specifically focus on horizontal wellbore sections. Formation lithology, borehole rugosity, depth, temperature, and pressure are additional factors that most multi-variable empirical models do not consider. Wellbore stability plays a role in cuttings transport efficiency and depends largely on the physicochemical characteristics of the rock formations penetrated in relation to those of the drilling fluid (Sun et al., 2023).

The height of accumulated cuttings beds tends to decrease as the inclination of the borehole decreases, making them a greater challenge in horizontal wellbores (Zhu et al., 2021). The type of drilling fluid used (oil-based, synthetic, or water-based) does not in itself influence hole cleaning performance, rather it is the drilling fluid density, rheology, and fluid flow regime that influence cuttings-transport efficiency (Hemphill and Larsen, 1996). Turbulent fluid-flow regimes tend to be more effective for borehole cleaning at most annulus fluid velocities and wellbore inclinations (Nazari et al., 2010; Piroozian et al., 2012). While drill cuttings particles of smaller size tend to be carried more easily in the circulating drilling fluids (Akhshik et al., 2016), drill cuttings of higher density and flatter shapes form thinner cuttings beds at the same flow rates than less dense more rounded particles (Tomren et al., 1986). This is a relevant factor in horizontal sections as denser flatter cuttings particles will take longer to reduce the flowing area of the annulus (Pandya et al., 2019). Nevertheless, denser cutting particles are more difficult to dislodge and move from cuttings beds, making it important to focus on the cuttings-lifting abilities of the drill fluids and flow regimes applied (Jimmy et al., 2022).

Periodic back reaming during drilling can help to disrupt and break up cuttings beds and avoid the need to trip the drill string from the wellbore as cuttings beds accumulate (Zhu et al., 2023). Alternatively, drilling a provisional pilot narrow-diameter borehole followed by reaming can substantially reduce the volume of the cuttings in the wellbore annulus and lower the equivalent circulating density of the drilling fluid (Lin et al., 2016), thereby improving hole cleaning. Additionally, various downhole cuttings-removal tools positioned in the drill string are commonly deployed to create vortexes when drilling long-horizontal sections to improve hole cleaning, particularly in wells drilling shale formations (Chen et al., 2022). Nevertheless, it remains important to adopt drilling designs that minimize the build-up of cuttings beds as much as possible. Indeed, efforts are being made to develop reliable transient hole-cleaning models validated with real wellbore results and deploying them as digital twins to monitor borehole cleaning in real-time (Arévalo et al., 2022) and/or automated-advisory systems (Forshaw et al., 2022).

Simulation studies, particularly those involving computational fluid dynamics (CFD), applied to horizontal wellbores and supported by experimental validation have proved to be effective in providing insight to cuttings-transport processes (Zakerian et al., 2018; Yeo et al., 2021). CFD models provide a means of better understanding cuttings particle behavior in complex fluid-flow regimes. Hajipour (2020) developed a CFD model to simulate turbulent two-phase flow regimes in horizontal wellbore annuli by applying the Herschel-Bulkley model. Vaziri et al. (2020) applied a CFD model to evaluate the hole-cleaning influences on foam-based drilling fluids in highly inclined wellbores. Awad et al. (2021) applied CFD to compare cuttings particle settling velocities in Newtonian versus non-Newtonian fluid-flow regimes. Epelle et al. (2019) evaluated the behavior of polydisperse spherical-particulate systems by applying both Eulerian-Eulerian and Lagrangian-Eulerian CFD methods. Zhu et al. (2023) simulated cuttings particle buoyancy and waveform distributions in cuttings beds applying a layered-mesh CFD model. The Lagrangian-Eulerian CFD technique is more complex and time-consuming but tends to provide more realistic representations of individual particle dynamics in various flow regimes (Epelle and Gerogiorgis, 2019).

Laboratory-based experimental studies play a key role in providing insight into the inter-relationships between the various factors influencing borehole cleaning in horizontal wellbores (Loureiro et al., 2010; Li and Luft, 2014). The test equipment used typically consists of a flow loop representing a drill-string/annulus configuration through which various fluids and cuttings-like materials can be circulated under a range of conditions including inclinations (Ma et al., 2023). However, most systems used to date can only operate at ambient pressure and temperature conditions. The experimental flow loops reported use both opaque (Piroozian et al., 2012) and transparent flow-loop systems at reduced (Leporini et al., 2019) and full-scale facilities (Xu et al., 2013). By carefully measuring the injection rates, weights, and volumes of cuttings materials, typically sand of different particle sizes, it is possible to establish the equivalent ROP of the system configuration of such flow loops. Transparent experimental equipment setups enable slow-motion visualizations of the circulating systems. The results of such experiments are often expressed in dimensionless terms as cuttings bed concentrations (%) rather than absolute cuttings-bed thickness measures (Song et al., 2017; Han et al., 2022). This facilitates the scaling of the results to various operational systems (Busch et al., 2020).

Machine learning (ML) is now extensively used to pro-

vide accurate predictions of experimentally derived cuttings bed data from a few (typically up to about five) influential variables. Ozbayoglu et al. (2002) were among the first to develop an artificial neural network (ANN) model to predict cuttings bed thickness in inclined wellbores. Their model considered a range of distinct fluid-flow regimes making it of more practical value than empirical relationships that tend to be focused on specific flow regimes. ANN, support vector machines (SVM), and multilinear regression (MLR) models have been widely applied for more than two decades to estimate drilling operational data including cuttings transport performance (Olukoga and Feng, 2021). However, as described in Section 2.3 several other ML algorithms have also been successfully applied to borehole cleaning datasets.

As the data available from borehole cleaning experimental studies and horizontal wellbores drilled in specific geological locations, ML will likely become more widely used than multivariate empirical relationships which tend to be more restricted to specific geological conditions and drilling fluid flow regimes. This study interpolates the data trends from a published experimental study to demonstrate how novel, and potentially operationally useful, empirical relationships with cuttings concentration in horizontal wellbores between pairs of influential variables can be derived using optimizers to complement a multivariate ML borehole cleaning dataset.

2. Materials and method

2.1 Cutting concentration experimental dataset

The results of a recently published hole-cleaning experimental dataset (Ma et al., 2023) evaluating highly inclined and horizontal wellbore conditions in turbulent fluid-flow regimes, derived from an experimental setup developed by the China University of Petroleum, are used in this study. Those results are configured and interpolated in this study to illustrate the benefits of applying customized empirical formulas and multivariate ML to derive additional practical insights from that dataset to those provided by multivariate empirical relationships. The results as presented by Ma et al. (2023) are in the form of seven graphical relationships, with each graph displaying the impact of two influential wellbore variables on dimensionless cuttings concentration in the annulus ($H\%$).

Collectively, the influential variables displayed in those graphs are annulus velocity (Va), cuttings generation rate (Cg), drill-pipe rotation speed (N), drill-pipe eccentricity (ε), drilling fluid density (Pf), effective viscosity (μe), cuttings particle (sand) size (Ds), wellbore inclination (θ; bimodal 80° or 90°). In addition, the power-law fluid index (n) and consistency factor (k) were provided for each experimental setup evaluated. These ten variables provide a useful multivariate dataset with which to evaluate horizontal and near-horizontal wellbore conditions.

Each graph presented (Ma et al., 2023) involves between eight and twenty experimentally recorded data points with a series of curved lines connecting the points. This study interpolates between those data points along the presented curved trends to sample those trends with between seventy-five and one hundred and eleven data points. The greater density of interpolated data points enables the optimizers and ML models to establish more statistically viable relationships from which predictions can be made.

2.2 Customizing empirical formula relationships using optimizers

A common approach is to apply non-linear regression methods to the experimental dataset to derive a single multivariate empirical formula for predicting $H\%$. However, such an approach is mathematically quite restrictive in that it typically derives single exponents for each variable or ratios of variables. Moreover, such relationships often do not fit the individual variables with high precision. In this study, more flexible formulas are fitted to pairs of influencing variables to predict $H\%$ from the interpolated graphical information. This is achieved using the customized-formula optimization method described by Wood (2022) which uses optimizers to rapidly determine the coefficients of combined non-linear and linear terms to establish empirical relationships that generate low prediction errors. For this study, two independent variables are considered in each relationship to determine dependent variable $H\%$. By using the generalized formula displayed in Eq. (1) for each relationship, $H\%$ can be predicted from the interpolated dataset with very low prediction errors:

$$H\% = ax_1^b + cx_1 + dx_2^e + fx_2 + g \tag{1}$$

where x_1 and x_2 are the two independent variables, and a to g are the unknown coefficients determined by the optimizers by minimizing the objective error functions root mean squared error (RMSE) and mean absolute error (MAE) when the formula is applied to the interpolated experimental dataset.

In this study, the optimal values of the coefficients are derived in Excel spreadsheets using the built-in "Solver" optimizers (Frontline Solvers, 2023)). These solvers are the Generalized Reduced Gradient (GRG) method (the GRG2 extension of the Simplex algorithm) (Lasdon et al., 1978), and the non-smooth, non-deterministic "Solver Evolutionary Algorithm", a genetic algorithm customized with features derived from other evolutionary algorithms. The two optimizers are configured and coded in Visual Basic for Applications to generate multiple runs using RMSE and MAE separately as their objective function. It would be straightforward to use other optimizers coded in other systems to conduct such optimization. However, it is convenient to do this in an Excel spreadsheet as the results from each optimizer run can be readily compiled and statistically assessed to determine the optimum solution from multiple optimizer runs.

2.3 Machine learning for multivariate cuttings-bed dataset analysis

Several ML model studies addressing hole-cleaning experimental datasets have been published in the past decade. For example, Rooki et al. (2014) obtain superior cuttings-concentration predictions from foam drilling experiments based on drilling parameters from an ANN compared to a multi-linear regression model. Ulker and Sorgun (2016) ob-

tained lower cuttings bed thickness prediction errors with their ANN model compared to MLR, SVM, and K-nearest neighbor (KNN) models applied to an experimentally-derived dataset for wellbore with inclinations from 60° to 90°. Rooki and Rakhshkhorshid (2017) applied a radial basis function neural network to model cuttings concentration during underbalanced drilling from drilling variables. Agwu et al. (2020) developed an ANN model to predict drilling cuttings settling velocity. Han et al. (2022) found that an ANN model provided more reliable predictions of cuttings bed height from flow-loop experimental data for horizontal wellbore configurations than support vector regression (SVR), recurrent neural network, and long short-term memory models. These models have, for the most part, been applied to relatively small experimental datasets (less than about 200 data points).

Other researchers have achieved lower cuttings-bed prediction errors with models other than ANN. For example, (Al-Azani et al., 2018, 2019) found that their SVM model slightly outperformed their ANN models in the prediction of cuttings concentration from mud density and mud rheological variables derived from experimental tests on deviated wellbores. Recent studies have also shown ensemble-tree models to be very effective for predicting cuttings-bed experimental datasets. For example, Awojinrin (2022) applied random forest (RF), gradient boosting (GB), and adaptive boosting (Adaboost) models, and an ensemble of those three models, to predict drilling cuttings concentration from drilling-fluid rheology and operational drilling variables from published datasets. Alsaihati and Elkatatny (2023) applied RF, GB, and extreme gradient boosting (XGB) models to estimate drill cuttings size from a synthetic dataset.

In this study, MLR, KNN, RF, SVR, and XGB supervised ML models are developed to predict $H\%$ from a ten-variable (N, θ, ε, Pf, Cg, Va, n, k, μe, and Ds), 641-point dataset compiled from all the interpolated data points extracted from the experimental results presented by Ma et al. (2023). They are evaluated and compared by applying the multi-K-fold cross-validation technique (Wood, 2023) to determine their $H\%$ prediction performance and to establish the model that generates the lowest prediction errors. The relative influences of each input feature on the prediction solutions of the best-performing model are also evaluated.

Ridge is an MLR algorithm that applies L2 regularization (Hastie, 2020) by adding a squared penalty to its error function to reduce solution complexity. The learning rate (λ) is the key control value with an optimum value of 0.01 for the dataset studied. Regardless of the regularization method applied, a key limitation of MLR models is that they assume parametric relationships between the variables, which is not the case for the compiled dataset (Harrell, 2015).

KNN is non-parametric, regression-free, and works by establishing the similarity between data records. The algorithm was developed by Fix and Hodges (1951) and is widely used for ML due to its simplicity and rapidity. KNN makes no parametric assumptions about the datasets it evaluates. It is widely applied for the classification of datasets with categorical dependent variables and for regression of datasets with dependent variables characterized by continuous distributions. Its key control variables are the number of nearest neighbors to be used in its assessment (optimized to 5 in this study) and the Minkowski distance metric (Shahid et al., 2009) to be used ($p = 1$ or Manhattan distance).

RF is a decision-tree-ensemble algorithm that varies its "bagging" and "out-of-bag" selections for the multiple trees it constructs and takes the mean solution of the collective predictions generated by the decision trees evaluated (Ho, 1998). For this study, the RF model was optimized with 750 estimators (individual decision trees), with each tree having a maximum depth limit of 20 levels, a mean squared error splitting criterion, and tree pruning to remove nodes making low contributions controlled by the ccp-alpha cost complexity parameter.

SVR is a widely used non-parametric algorithm able to apply different kernel functions to define hyperplanes in multi-dimensional (variable) space that separate the data points for prediction purposes. It tends to generate fewer prediction errors with highly non-linear exist between the variables when a radial-basis-function kernel is employed (Chang et al., 2010). SVR was initially developed by Cortes and Vapnik (1995) and its performance is highly sensitive to the control parameter values selected. The key control values applied in this study are the error-regularization factor (C = 2,000,000), radial-basis-function depth of influence or gamma (γ is set to 'scale'), and the error-tolerance margin (ε = 0.001) establishing an error limit beyond which penalties are applied.

XGB is a tree-ensemble algorithm, which employs a parallel gradient-boosting process incorporating L1 and L2 regularization terms, applied to multiple decision trees (Chen and Guestrin, 2016). It tends to be more flexible than other tree-ensemble algorithms because it involves more control variables, which require careful tuning. The XGB control variable values applied in this study are number of estimators = 750, maximum depth limit for each tree = 10, learning rate (eta) = 0.05, subsample = 0.7, and column sample per tree = 0.8.

These five models were selected because they apply distinctive algorithms including both parametric and non-parametric assumptions. Collectively they are, therefore, able to assess and predict the non-linear datasets effectively. These ML models were coded in Python and executed using Scikit-Learn functions (Scikit Learn, 2023a) including the randomized search function to identify the optimum control- parameter values (Scikit Learn, 2023b), MinMaxScaler to normalize all data variables to the ranges of -1 to 1 (Scikit Learn, 2023c), and the Cross-validation function (Scikit Learn, 2023d) customized in this study, to deliver a multi-K-fold analysis of each algorithm.

2.4 Prediction performance assessment metrics

Multiple statistical prediction performance metrics are calculated to assess the results of the empirical formula optimization methods and the ML models evaluated. The formulas used to calculate these commonly employed metrics are described in Fig. 1.

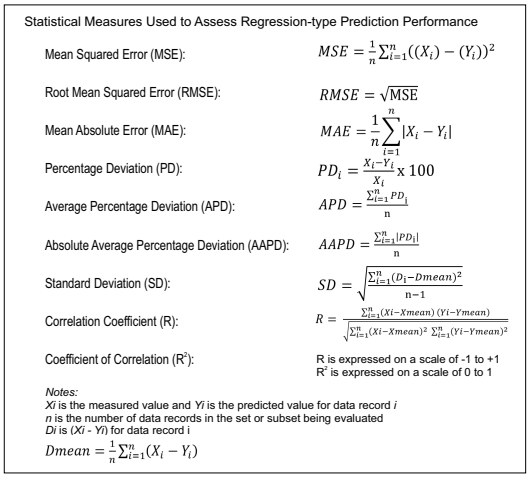

Fig. 1. Statistical metrics assessing prediction errors for regression-type analysis used in this study to evaluate optimized empirical formula and ML model predictions.

3. Results

3.1 Interpolated experimental data trends of variables influencing $H\%$

The interpolated data distributions of $H\%$ experimental results derived from the graphical presentations of Ma et al. (2023) are displayed in Figs. 2 and 3. From the distribution of data points in Figs. 2 and 3 it is apparent that $H\%$ is most sensitive to variables Va, N, and Cg (reflecting ROP), findings that are to be expected based on historical empirical models (Al-Rubaii et al., 2023). However, the data trends reveal distinctive inflection points in the associated Va, Cg, and N variations considered (Figs. 2(a) and 2(b)). Expressing the trends displayed in Figs. 2 and 3 with non-linear empirical formulas involving two independent variables with $H\%$ as the dependent variable can provide insights into these relationships with practical benefits for drilling operators.

3.2 Empirical relationships between groups of variables

The extrapolated experimental data makes it possible to establish empirical non-linear relationships between groups of the variables considered that are of practical value to drillers/drilling engineers providing guidance during the drilling of near-horizontal wellbore sections. Applying the customized formulaic optimization approach (Wood, 2022) determines empirical formulas that fit the interpolated data displayed in Figs. 2 and 3 with minimum errors.

3.2.1 $H\% = f(Va, Cg)$

Eq. (2) displays the optimized formulaic relationship between $H\%$, Va and Cg (Fig. 2(a)):

$$H\% = 437.61Va^{-0.2837} + 73.67Va - 22.11Cg^{-0.1711} + 0.73Cg - 499.96 \quad (2)$$

The relatively low errors that result from applying Eq. (2) are illustrated in Fig. 4(a) and detailed in Table 1. These results relate to experiments conducted on a horizontal wellbore configuration in which drill-pipe rotation speed (N) was 60 rpm, drill-pipe eccentricity (ε) was 0, and drilling fluid specific gravity (Pf) was 1. Variable Cg is directly correlated with the implied ROP from the experiments varying between 10 and 50 m/h to achieve a range of Cg from 7.8 to 26 kg/min.

3.2.2 $H\% = f(N, Va)$

Eq. (3) displays the optimized formulaic relationship between $H\%$, Va and N (Fig. 2(b)):

$$H\% = -4.31N^{0.8825} + 2.26N - 72.01Va^{0.0646} - 37.64Va - 0.6869 \quad (3)$$

The relatively low errors that result from applying Eq. (3) are illustrated in Fig. 4(b) and detailed in Table 1. These results relate to experiments conducted on a horizontal wellbore configuration in which ROP was 10 m/h, ε was 0.5, Pf was 1.1, power-law fluid behavior index or exponent (n) was 0.51 and consistency factor (k) was 0.41. The experimental results reveal (Fig. 2(b)) that the relationship between N and $H\%$ is highly non-linear with inflection points occurring in each Va curve between N = 30 and 40 rpm. For N values below the inflection point, the $H\%$ value increases more steeply as N declines.

3.2.3 $H\% = f(\varepsilon, Va)$

Eq. (4) displays the optimized formulaic relationship between $H\%$, Va and ε (Fig. 2(c)):

$$H\% = 85.51\varepsilon^{1.1681} - 57.03\varepsilon - 16.86Va^{1.6091} - 39.96Va + 73.23 \quad (4)$$

The relatively low errors that result from applying Eq. (4) are illustrated in Fig. 4(c) and detailed in Table 1. These results relate to experiments conducted on a horizontal wellbore configuration in which N was 60 rpm, ROP was 10 m/h, Pf was 1.1, n was 0.585, and k was 0.1371. The experimental results reveal (Fig. 2(c)) that the relationship between C and $H\%$ is non-linear, with $H\%$ increasing more rapidly at all Va speeds considered once ε increases above 0.25.

3.2.4 $H\% = f(pf, Va)$

Eq. (5) displays the optimized formulaic relationship between $H\%$, Va and Pf (Fig. 2(d)):

$$H\% = 5.02Va^{4.9152} - 63.92Va + 12.25Pf^{-1.8777} - 2.30Pf + 63.21 \quad (5)$$

The relatively low errors that result from applying Eq. (5) are illustrated in Fig. 4(d) and detailed in Table 1. These results relate to experiments conducted on a horizontal wellbore configuration in which N was 60 rpm, ε was 0, Cg was 7.8 kg/min, n was 0.585 and k was 0.1371. The experimental results reveal (Fig. 2(d)) that the relationship between Pf and $H\%$ is moderately non-linear with $H\%$ decreasing by relatively small increments as Pf increases across a wide range of Va. The impact of Pf on $H\%$ is higher at low Va rates than at high Va rates.

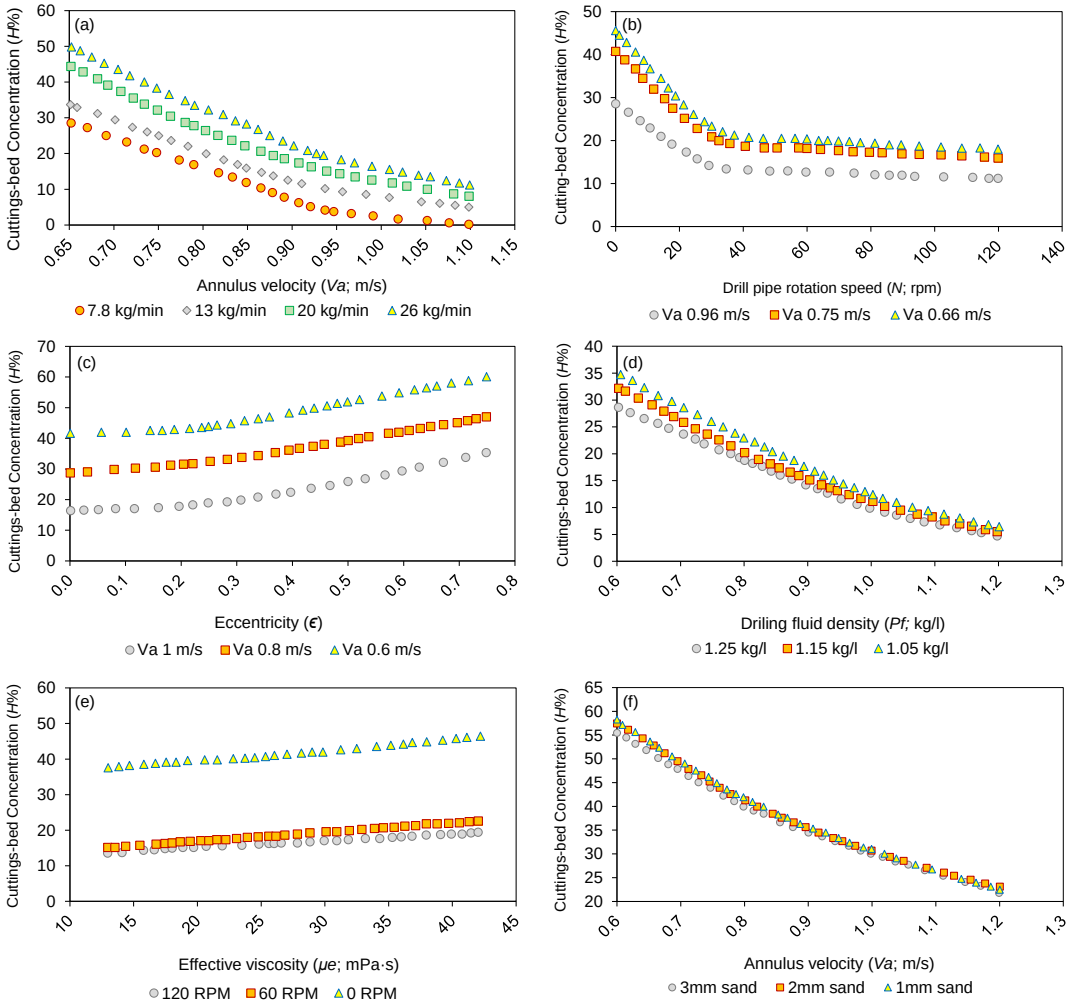

Fig. 2. Impacts of influencing variables on cuttings-bed concentration ($H\%$) in horizontal wellbore sections with data points interpolated from the experimental results of Ma et al. (2023). Impact of: (a) cuttings generation rate (Cg), (b) annulus velocity (Va), (c) annulus velocity (Va), (d) drilling fluid density (Pf), (e) drill-pipe rotation speed (N), (f) sand cutting size (Ds).

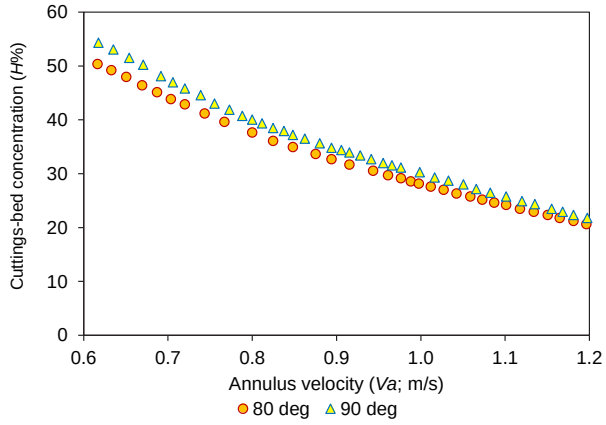

Fig. 3. Impacts of wellbore inclination angle on cuttings-bed concentration ($H\%$) in near-horizontal wellbore sections with data points interpolated from the experimental results of Ma et al. (2023).

3.2.5 $H\% = f(\mu e, Va)$

Eq. (6) displays the optimized formulaic relationship between $H\%$, μe and N (Fig. 2(e)):

$$H\% = 6.9\mu e^{-0.1389} + 0.27\mu e + 10.96N^{-0.1330} - 0.029N + 2.38 \quad (6)$$

The relatively low errors that result from applying Eq. (6) are illustrated in Fig. 4(e) and detailed in Table 1. These results relate to experiments conducted on a horizontal wellbore configuration in which ε was 0, Pf was 1.1, and Cg was 7.8 kg/min. The experimental results reveal (Fig. 2(e)) that the relationship between μe and $H\%$ is approximately linear with $H\%$ increasing slightly as μe increases. The impact of μe on $H\%$ becomes more relevant at low values of N.

3.2.6 $H\% = f(Ds, Va)$

Eq. (7) displays the optimized formulaic relationship between $H\%$, Ds and Va (Fig. 2(f)):

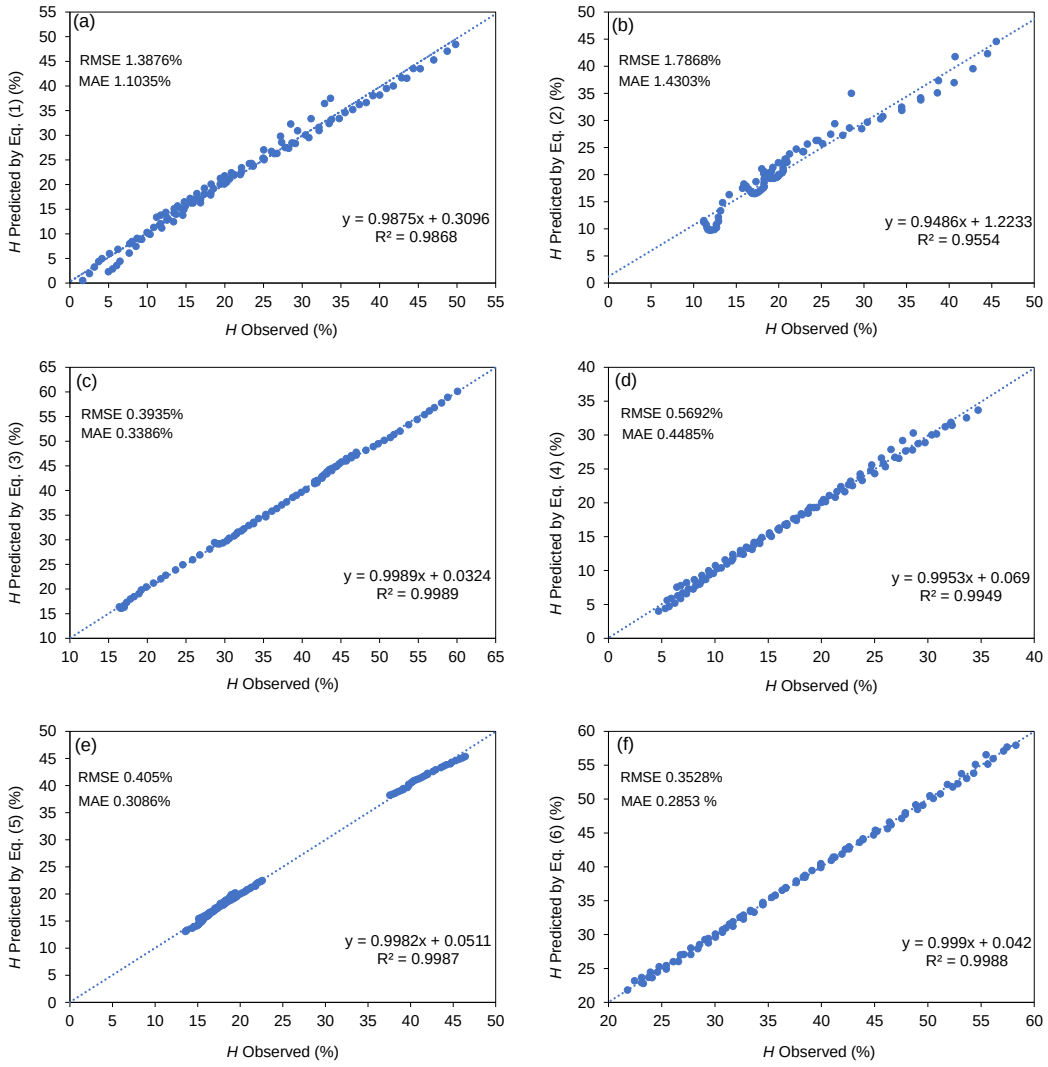

Fig. 4. Predicted versus observed $H\%$ values and prediction errors associated with applying Eq. (2)-(7) to the horizontal wellbore sections with data points interpolated from the experimental results of Ma et al. (2023).

$$H\% = 133.44\mu e^{-0.4014} + 8.38Va - 0.867Ds^{-64.1021} - 1.14Ds - 108.87 \quad (7)$$

The relatively low errors that result from applying Eq. (7) are illustrated in Fig. 4(f) and detailed in Table 1. These results relate to experiments conducted on a horizontal wellbore configuration in which N was 0, ε was 0, Pf was 1.05, Cg was 7.8 kg/min, n was 0.4792 and k was 0.7959. The experimental results reveal (Fig. 2(f)) that the impact of Ds, in the sand particle-size range studied (1 to 3 mm) on $H\%$ was small and almost indiscernible for Va values > 1. For $Va < 1$ the smaller sand particles generated lower $H\%$ values than the larger sand particles.

3.2.7 $H\% = f(Ds,Va)$

Eq. (8) displays the optimized formulaic relationship between $H\%$, θ and Va (Fig. 3):

$$H\% = 28.96Va^{-1.0030} - 14.17Va + 65.71\theta^{-97.4082} + 0.23\theta - 5.56 \quad (8)$$

The relatively low errors that result from applying Eq. (8) are illustrated in Fig. 5 and detailed in Table 1. These results relate to experiments conducted on a highly deviated wellbore with θ varying from 80° to 90° with a configuration in which N was 0, ε was 0.75, Pf was 1.05, Cg was 7.8 kg/min, n was 0.5526 and k was 0.2284. The experimental results reveal (Fig. 2) that the impact of θ on $H\%$, for the high-angle wellbore deviation range evaluated, was small but more discernible at the lowest Va values evaluated. The $H\%$ value was lower at θ = 80° than at θ = 90° (Fig. 3).

3.3 Machine learning models to predict $H\%$ from multiple influential variables

By compiling the interpolated experimental data (Figs. 2 and 3) into a dataset involving ten variables (N, θ, ε, Pf, Cg,

Table 1. Prediction errors associated with formulaic optimized solutions.

Eq.	Influencing vanables	RMSE (%)	MAE (%)	R^2	APD (%)	AAPD (%)	SD (%)
(2)	*Va*, *Cg*	1.3876	1.1035	0.9868	22.1694	29 .3386	1.930
(3)	*N*, *Va*	1.7868	1.4303	0.9554	-0.5630	7.0783	1.7940
(4)	ε, *Va*	0.3935	0.3386	0.9989	0.0284	1.0127	0.3957
(5)	*Pf*, *Va*	0.5692	0.4485	0.9949	0.3273	3.5349	0.5720
(6)	μe, *Cg*	0.4050	0.3086	0.9987	0.0168	1.4045	0.4071
(7)	*Ds*, *Va*	0.3528	0.2853	0.9988	-0.0085	0.7890	0.3547
(8)	θ, *Va*	0.4986	0.3727	0.9977	0.6019	1.1337	0.4628

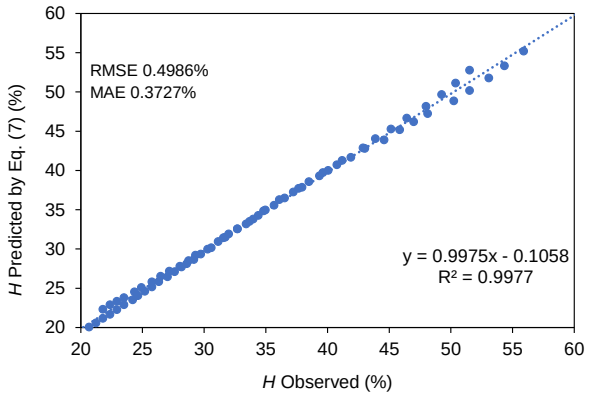

Fig. 5. Predicted versus observed *H*% values and prediction errors associated with applying Eq. (8) to the near-horizontal wellbore sections with data points interpolated from the experimental results of Ma et al. (2023).

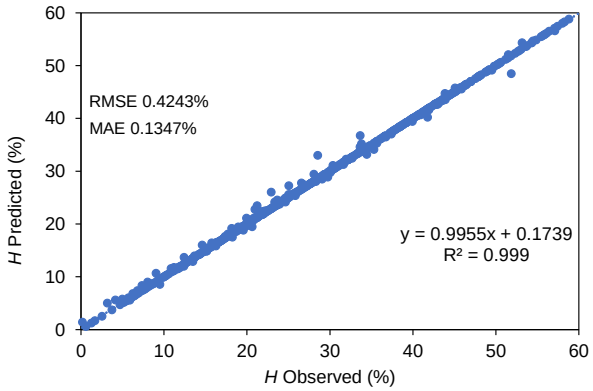

Fig. 6. Predicted versus observed *H*% values and prediction errors associated with applying the XGB model to all data records of the 10-input variable dataset for horizontal wellbore sections with 641 data points interpolated from the experimental results of Ma et al. (2023).

Va, *n*, *k*, μe, and *Ds*) supervised machine learning models can be used to predict *H*%. The machine learning method applied to the 10-variable dataset adds the ability to consider situations where a more comprehensive set of input variables are available rather than rely on the empirical relationships each based on just two variables. The specific advantage offered by a machine learning approach is that it does not rely on a single multivariate non-linear equation. That means that ML methods can be: adapted to suit different subsurface and borehole conditions; periodically updated as the number of available experimental data points increases; and, readily modified to include or exclude certain influential variables, depending on the data points available.

The results of the 10-fold cross-validation analysis (repeated three times) applying the Ridge (multi-linear regression), KNN, SVR, RF, and XGB models to the compiled 10-input-variable compiled dataset (641 data records) to predict *H*% are shown in Table 2. From those results, it is apparent that the XGB model generates the lowest *H*% prediction errors of the models considered. As the relationships between several of the input variables and *H*% are substantially non-linear, it is not surprising that the Ridge model provides *H*% predictions with high errors than the other ML models which are better able to deal with non-linearity. The mean plus one standard deviation (SD) metric (Table 2) is considered the most useful in comparing the prediction performance of the models. The performance of the ML models is ranked XGB (best) > RF > SVR > KNN > Ridge (worst) in terms of RMSE mean plus one SD. However, using the MAE mean plus one SD, the ML models are ranked XGB (best) > RF > KNN > SVR > Ridge (worst). The ensemble decision tree models (XGB and RF) provide predictions with substantially fewer errors than the other models evaluated based on that metric for both MAE and RMSE.

Table 3 provides the results of the multi-K-fold cross-validation analysis for the best-performing XGB model. This demonstrates that the 5-fold (repeated 6 times), 10-fold (repeated 3 times) and 15-fold (repeated twice) cross-validation analyses generate lower prediction errors compared to the 3-fold (repeated 10 times) and 4-fold (repeated 8 times) cross-validation analyses. For the RMSE metric, the 15-fold analysis generates the lowest mean values but due to higher standard deviations for the 10-fold and 15-fold analysis, it is the 5-fold analysis that generates the lowest mean plus one SD RMSE (Table 3). For the MAE metric, the 15-fold analysis generates the lowest mean and mean plus one SD value (Table 3). These results suggest that assigning 80% or more of the data records for model training (20% or less to validation) generates the most repeatable *H*% prediction results for the

Table 2. 10-fold cross validation *H*% prediction error statistics based on thirty cases applying different ML algorithms to compiled dataset with ten input variables (641 Data Records; 3 Repeat 10-fold Cross-validation Runs).

ML model	RMSE (%)			MAE (%)		
	Mean	SD	Mean + 1 SD	Mean	SD	Mean + 1 SD
XGB	0.8217	0.2116	1.0333	0.5077	0.0792	0.5870
RF	1.2241	0.2019	1.4260	0.8280	0.1000	0.9280
SVR	2.1788	0.2071	2.3859	1.6550	0.1552	1.8102
KNN	2.0924	0.4588	2.5512	1.0494	0.1964	1.2459
Ridge	6.9101	0.6639	7.5740	5.3410	0.4994	5.8404

Table 3. Multi-fold cross validation *H*% prediction error statistics for the XGB model applied to the compiled dataset with ten input variables (641 Data Records; Multi-fold Cross-validation Runs).

ML model	RMSE (%)			MAE (%)		
	Mean	SD	Mean + 1 SD	Mean	SD	Mean + 1 SD
3-fold (10 repeats)	1.0422	0.1440	1.1862	0.6550	0.0655	0.7205
4-fold (8 repeats)	0.9311	0.1556	1.0867	0.5752	0.0772	0.6525
5-fold (6 repeats)	0.8908	0.1132	1.0040	0.5445	0.0572	0.6016
10-fold (3 repeats)	0.8217	0.2116	1.0333	0.5077	0.0792	0.5870
15-fold (2 repeats)	0.7613	0.2825	1.0438	0.4723	0.1069	0.5792

compiled dataset.

Fig. 6 displays the prediction performance of the XGB model trained randomly with 90% of the data records to the entire compiled dataset.

3.4 Influence of the input variables on the *H*% predictions of the XGB model

The ensemble decision tree models (XGB and RF) provide information on the relative importance of the input variables to their model solutions. This feature-importance information is displayed in Fig. 7.

Although the XGB and RF models both provide *H*% prediction for the compiled 10-variable borehole cleaning dataset (641 data records), they do so by assigning different levels of relative importance to each of the ten input variables available. The RF model relies heavily on just two input variables (Fig. 7), *Va* and *N*, with weightings of 0.52 and 0.34, respectively. The RF model assigns importance weighting of ≤ 0.05 to the other eight variables.

On the other hand, the XGB model assigns meaningful weightings to seven of the variables, *n*, *N*, *k*, *Va*, *Pf*, ε, and *Cg*, with weighting arranged in descending value order of ~0.25, ~0.21, ~0.17, ~0.12, ~0.09, ~0.07, and ~0.06, respectively. Neither of the models assigns weighting above ~0.01 to the variables θ, μe, and *Ds*. The broader spread of input-variable weightings employed by the XGB model enables it to outperform the RF model in terms of the accuracy of its *H*% predictions for the compiled dataset (Fig. 7).

There is a tendency in hole cleaning and cuttings transport studies to attempt to develop multivariate empirical equations

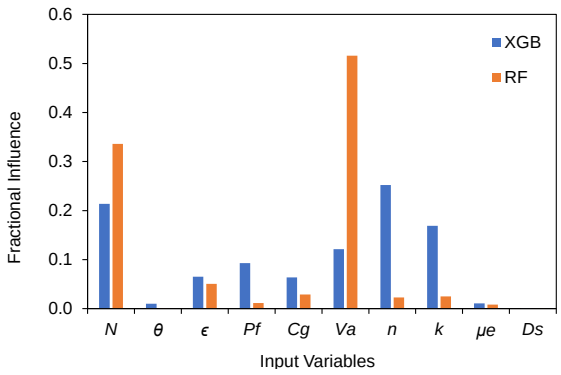

Fig. 7. Feature importance comparisons for the XGB and RF model solutions applied to the 10-input variable compiled dataset to predict *H*.

to predict cuttings bed thickness/concentrations or hole-cleaning indices from combinations of experimental, simulation, and actual wellbore data. However, such empirical formulas tend to be limited to a specific set of subsurface geological and borehole conditions, wellbore trajectories, and drilling techniques, which makes them difficult to rely upon for more generalized applications. Also, as so many variables influence borehole cleaning performance, with greater or lesser impacts in different areas, the multivariate empirical formulas proposed to date do not take into account all influential factors.

4. Discussion

Rather than attempt to develop cumbersome multivariate formulas with ten or more variables, a case can be made to focus on the relative impacts of pairs of influential variables on hole cleaning based on specific sets of conditions. This can provide simpler and more easily applied and tested empirical equations that can reliably reproduce the relationships between variable pairs established by experimental results. Experiments typically only record a limited set of data points across the ranges of interest for each influential variable. It is often, therefore, helpful, to interpolate between the limited experimentally recorded data points to downscale the available dataset and provide more (albeit approximated) data points at smaller scale intervals to improve the derivation of feasible empirical equations for the range of conditions considered.

A large number of interpolated data points used to develop multiple pair-wise relationships for specific borehole configurations can also be compiled to form multivariate datasets for evaluation by ML models to predict hole-cleaning metrics such as cuttings bed thickness/concentration. The advantage of this approach, compared to cumbersome multivariate empirical relationships to predict hole-cleaning indices is that the ML dataset can be periodically updated/expanded with additional data points. It is much easier to retrain and/or adjust ML models in terms of variable feature selection to suit the conditions pertinent to specific wellbores than to customize or redefine multivariate empirical equations to suit specific wellbore conditions. Additionally, over time, it should be possible to compile larger ML multivariate datasets, incorporating both experimental and historical wellbore recordings covering a wide range of subsurface conditions. Such ML datasets should eventually become more easily generalized for applications to suit a wide range of wellbore configurations in a specific area.

The results presented in this study illustrate the benefits of the proposed approach applied to recently published experimental results about a specific set of highly deviated/horizontal wellbore conditions to determine influences on dimensionless cuttings-bed thickness/concentration (Ma et al., 2023). That study established a multivariate empirical equation based on the observed experimental results that were evaluated using historical data from an actual horizontal wellbore section of 1,230 m length drilled at greater than 2,500 m true vertical depth. Their evaluation involved the comparison of wellbore pump pressure predicted by their developed cuttings-carrying relationships, combined with a multi-phase fluid-flow model with measured data collected from the historically drilled horizontal section. The predicted versus measured pump pressures agreed within 5%, verifying the veracity of the relationships established from the experimental data. The pair-wise empirical relationships established in this study (Eq. (2)-(8)) provide additional, complementary relationships that could be used in real-time to inform drillers on how best to adjust drilling conditions to minimize cuttings-bed thickness/concentration in horizontal wellbore sections, as drilling progresses, based on the limited data available in real-time. Moreover, the multivariate interpolated experimental dataset could be used together with a previously trained XGB model to rapidly predict with reliable accuracy the likely dimensionless cuttings-bed concentration that would likely result from various alternative combinations of drilling conditions.

To further develop the proposed approach and expand the conditions over which it could be reliably applied, further experimental data combined with actual wellbore recordings are required. This is particularly the case for relationships between $H\%$ and variables n and k, which exert a substantial influence on the XGB model applied to the dataset studied (Fig. 7). The specific conditions evaluated experimentally in the dataset studied are restricted in several ways. The specific restrictions are (1) the cuttings lithology evaluated is limited to sand with particle size varying from 1 to 3 mm; (2) the borehole inclinations considered are limited to 80° and 90°; and, (3) drilling fluid specific gravities considered are limited to 1.05 to 1.25. Many long, near-horizontal well bore sections are now being drilled into shale reservoirs with various lithological and physical properties. Further experiments are therefore required that evaluate cuttings of larger sizes, shapes and different shale compositions/densities. As the inclinations of many near-horizontal wellbores varies intermittently along their lengths, further experiments are required that consider multiple inclinations in the range of 70° to 90° to determine the impacts of wellbore inclination in more detail. Likewise, experiments addressing a wider range of drilling fluid specific gravities and rheological properties are required. The information gained from such additional experiments would substantially improve the generalizability of the multivariate ML database currently available.

In addition to wider-ranging experimental conditions being evaluated, observations from real wellbores are required to establish the impacts of various subsurface conditions on key wellbore variables influencing hole cleaning performance. For instance, further information is required regarding the impacts on pair-wise variable relationships related to borehole cleaning of (1) a wide range of subsurface pressures and temperatures, (2) borehole and annular diameters, (3) wellbore pump pressures, and (4) ROPs in different lithologies. With the addition of test results from a broader range of experimental conditions combined with observations from a diverse range of real horizontal wellbore sections, the pair-wise empirical relationships established in this study could be further refined. Moreover, such information would strengthen the multivariate dataset available for ML models, making their predictions more reliable and generalizable across a wider range of sub-surface conditions. Additionally, for larger datasets deep learning algorithms may be more effective than the XGB algorithm in generating even lower prediction errors. Future studies should therefore also consider a wider range of both ML and deep learning models in attempts to further reduce prediction errors with larger multi-variate datasets.

5. Conclusions

Experimental testing of the factors influencing cuttings-bed formation in horizontal wellbore sections provides useful guidance regarding conditions to avoid and strive for during drilling operations. However, as there are so many factors that

influence cuttings-bed concentration most multivariate empirical relationships proposed as potentially useful operational indicators of horizontal wellbore cleaning performance work only for a limited set of downhole conditions. This study uses recently publish experimental results to demonstrate that by interpolating between test data points it is possible to derive, with the aid of optimizers a series of empirical relationships that capture the influences of pairs of variables on *H%*. Seven empirical equations, each involving two independent variables, are derived using standard spreadsheet Solver optimizers. These equations fit the interpolated data with RMSE for *H%* predictions of less than 1.8% for cuttings concentration values between ~2% and 60%.

The interpolated experimental data is also compiled into a multivariate dataset involving ten independent variables with *H%* as the dependent variable and involving 641 data records. Five machine learning algorithms (multi-linear regression, K-nearest neighbor, random forest, support vector regression, and XGB) are optimized to predict *H%* using this dataset. Based on multi-K-fold cross-validation analysis, the XGB model predicts *H%* for the compiled dataset with the lowest errors: mean RMSE = 0.76%; mean MAE = 0.47% for the 15-fold cross-validation case. The XGB model, through its feature importance function, reveals that it utilizes information primarily from seven of the ten influential variables available in the compiled dataset when making its *H%* predictions.

As more experimental and real horizontal wellbore data become available in specific plays, it is considered likely that ML models applied to multivariate datasets (10 + influential variables) will become more reliable for hole cleaning prediction than multivariate empirical relationships proposed for more general use. However, simpler non-linear optimized empirical relationships (with two or three influential variables), such as those derived in this study, are likely to complement the ML models for assisting with real-time drilling decisions.

Conflict of interest

The author declares no competing interest.

References

Agwu, O. E., Akpabio, J. U., Dosunmu, A. Artificial neural network model for predicting drill cuttings settling velocity. Petroleum, 2020, 6(4): 340-352.

Akhshik, S., Behzad, M., Rajabi, M. CFD-DEM simulation of the hole cleaning process in a deviated well drilling: The effects of particle shape. Particuology, 2016, 25: 72-82.

Al-Azani, K. H., Elkatatny, S., Abdulraheem, A., et al. Prediction of cutting concentration in horizontal and deviated wells using support vector machine. Paper SPE 192193 Presented at SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition, Dammam, Saudi Arabia, 23-26 April, 2018.

Al-Azani, K. H., Elkatatny, S., Ali, A., et al. Cutting concentration prediction in horizontal and deviated wells using artificial intelligence techniques. Journal of Petroleum Exploration and Production Technology, 2019, 9: 2769-2779.

Al-Rubaii, M., Al-Shargabi, M., Al-Shehri, D., et al. A novel efficient borehole cleaning model for optimizing drilling performance in real time. Applied Sciences, 2023, 13: 7751.

Alsaihati, A., Elkatatny, S. A new method for drill cuttings size estimation based on machine learning technique. Arabian JournaI for Science and Engineering, 2023, in press, https://doi.org/10.1007/s13369-023-08007-0.

Arévalo, P. J., Forshaw, M., Starostin, A., et al. Monitoring hole-cleaning during drilling operations: Case studies with a real-time transient model. Paper SPE 210244 Presented at SPE Annual Technical Conference and Exhibition, Houston, Texas, 3-5 October, 2022.

Awad, A. M., Hussein, I. A., Nasser, M. S., et al. CFD modeling of particle settling in drilling fluids: Impact of fluid rheology and particle characteristics. Journal of Petroleum Science and Engineering, 2021, 199: 108326.

Awojinrin, G. T. Machine learning workflow for the determination of hole cleaning conditions. Paper SPE 212381 Presented at Annual Technical Conference and Exhibition, Houston, Texas, 3-5 October, 2022.

Busch, A., Johansen, S. T. Cuttings transport: On the effect of drill pipe rotation and lateral motion on the cuttings bed. Journal of Petroleum Science and Engineering, 2020, 191: 107136.

Busch, A., Werner, B., Johansen, S. T. Cuttings transport modeling-part 2: dimensional analysis and scaling. SPE Drilling & Completion, 2020, 35(1): 69-87.

Chang, Y., Hsieh, C. J., Chang, K., et al. Training and testing low-degree polynomial data mappings via linear SVM. Journal of Machine Learning Research, 2010, 11(4): 1471-1490.

Chen, T., Guestrin, C. XGBoost: A scalable tree boosting system. Paper Presented at Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, USA, 13-17 August, 2016.

Chen, F., Liu, Z., Huo, Y., et al. Mechanical mechanism and removal effect of efficient vortexing cuttings removal tool. Advances in Mechanical Engineenng, 2022, 14(12): 1-14.

Cortes, C., Vapnik, V. Support-vector networks. Machine Learning, 1995, 120 (3): 273-297.

Epelle, E. I., Gerogiorgis, D. I. Drill cuttings transport and deposition in complex annular geometries of deviated oil and gas wells: A multiphase flow analysis of positional variability. Chemical Engineering Research and Design, 2019, 151: 214-230.

Epelle, E. I., Obande, W., Okolie, J. A., et al. CFD modelling and simulation of drill cuttings transport efficiency in annular bends: Effect of particle size polydispersity. Journal of Petroleum Science and Engineering, 2022, 208: 109795.

Fix, E., Hodges, J. L. Discriminatory analysis, nonparametric

discrimination: Consistency properties. Technical Report, USAF School of Aviation Medicine, 1951.

Forshaw, M. J., Qahtani, Y. S., Aramco, S., et al. Validation of full transient hole cleaning model, with at-scale datasets, implementation into an automated advisory system. Paper SPE 32081 Presented at SPE Offshore Technology Conference, Houston, Texas, 2-5 May, 2022.

Frontline Solvers. Excel Solver–non-linear optimization methods 2023.

Hajipour, M. CFD simulation of turbulent flow of drill cuttings and parametric studies in a horizontal annulus. SN Applied Sciences, 2020, 2: 1146.

Han, Y., Zhang, X., Xu, Z., et al. Cuttings bed height prediction in microhole horizontal wells with artificial intelligence models. Energies, 2022, 15: 8389.

Harrell, F. E. Regression Modeling Strategies (second edition). Cham, Switzerland, Springer, 2015.

Hastie, T. Ridge regularization: An essential concept in data science. Technometrics, 2020, 62(4): 426-433.

Hemphill, T., Larsen, T. I. Hole-cleaning capabilities of water- and oil-based drilling fluids: a comparative experimental study. SPE Drilling & Completion, 1996, 11(4): 201-207.

Ho, T. K. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(8): 832-844.

Hopkin, E. A. Factors affecting cuttings removal during rotary Drilling. Journal of Petroleum Technology, 1967, 19(6): 807-814.

Jimmy, D., Wami, E., Ogba, M. I. Cuttings lifting coefficient model: A criteria for cuttings lifting and hole cleaning quality of mud in drilling optimization. Paper SPE 212004 Presented at SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 1-3 August, 2022.

Lasdon, L. S., Waren, A. D., Jain, A., et al. Design and testing of a generalized reduced gradient code for nonlinear programming. ACM Transactions on Mathematical Software, 1978, 4 (1): 34-50.

Leporini, M., Marchetti, B., Corvaro, F., et al. Sand transport in multiphase flow mixtures in a horizontal pipeline: An experimental investigation. Petroleum, 2019, 5: 161-170.

Li, J., Luft, B. Overview of solids transport studies and applications in oil and gas industry-experimental work. Paper SPE 171285 Presented at SPE Russian Oil and Gas Exploration & Production Technical Conference and Exhibition, Moscow, Russia, 14-16 October, 2014.

Li, J., Walker, S. Sensitivity analysis of hole cleaning parameters in directional wells. SPE Journal, 2001, 6(4): 356-363.

Lin, T., Wei, C., Zhang, Q., et al. Calculation of equivalent circulating density and solids concentration in the annular space when reaming the hole in deepwater drilling. Chemistry and Technology of Fuels and Oils, 2016, 52: 70-75.

Loureiro, B. V., Paula, R. S., Serafim, M., et al. Experimental evaluation of the effect of drill string rotation in the suspension of a cuttings bed. Paper SPE 122071 Presented at SPE Latin American and Caribbean Petroleum Engineering Conference, Lima, Peru, 1-3 December, 2010.

Ma, Y., Yang, C., Liu, X. On hole cleaning evaluation method in highly deviated/horizontal well sections. Journal of Physics: Conference Series, 2023, 2442: 012037.

Mahmoud, A. A., Elzenary, M., Elkatatny, S. New hybrid hole cleaning model for vertical and deviated wells. JournaI of Energy Resources TechnoIogy, 2020a, 142(3): 034501.

Mahmoud, H., Hamza, A., Nasser, M. S., et al. Hole cleaning and drilling fluid sweeps in horizontal and deviated wells: Comprehensive review. Journal of Petroleum Science and Engineering, 2020b, 186: 106748.

Moroni, N., Ravi, K., Hemphill, T., et al. Pipe rotation improves hole cleaning and cement-slurry placement: mathematical modeling and field validation. Paper SPE 124726 Presented at SPE Offshore Europe Oil and Gas Conference and Exhibition, Aberdeen, UK, 8-11 September, 2009.

Nazari, T., Hareland, G., Azar, J. J. Review of cuttings transport in directional well drilling: Systematic approach. Paper SPE 132372 Presented at SPE Western Regional Meeting, Anaheim, California, 27-29 May, 2010.

Olukoga, T. A., Feng, Y. Practical machine-learning applications in well-drilling operations. SPE Drilling & Completion, 2021, 36(4): 849-867.

Ozbayoglu, E. M., Miska, S. Z., Reed, T., et al. Analysis of bed height in horizontal and highly-inclined wellbores by using artificial neural networks. Paper SPE 78939 Presented at SPE International Thermal Operations and Heavy Oil Symposium and International Horizontal Well Technology Conference, Calgary, Alberta, 4-7 November, 2002.

Pandya, S., Ahmed, R., Shah, S. Effects of particle density on hole cleanout operation in horizontal and inclined wellbores. Paper SPE 194240 Presented at SPE/ICoTA Well Intervention Conference and Exhibition, The Woodlands, Texas, 26-27 March, 2019.

Piroozian, A., Ismail, I., Yaacob, Z., et al. Impact of drilling fluid viscosity, velocity and hole inclination on cuttings transport in horizontal and highly deviated wells. Journal of Petroleum Exploration and Production Technology, 2012, 2: 149-156.

Power, D. J., Hight, C., Weisinger, D., et al. Drilling practices and sweep selection for efficient hole cleaning in deviated wellbores. Paper SPE 627794 Presented at IADC/SPE Asia Pacific Drilling Technology, Kuala Lumpur, Malaysia, 11-13 September, 2000.

Rooki, R., Ardejani, F. D., Moradzadeh, A. Hole cleaning prediction in foam drilling using artificial neural network and multiple linear regression. Geomaterials, 2014, 4(1): 47-53.

Rooki, R., Rakhshkhorshid, M. Cuttings transport modeling in underbalanced oil drilling operation using radial basis neural network. Egyptian Journal of Petroleum, 2017, 26(2): 541-546.

Scikit Learn. Supervised and unsupervised machine learning models in Python, 2023a.

Scikit Learn. RandomizedSearchCV function for control parameters, 2023b.

Scikit Learn. MinMaxScaler function for data pre-processing, 2023c.

Scikit Learn. Cross-validation: evaluating estimator performance, 2023d.

Shahid, R., Bertazzon, S., Knudtson, M. L., et al. Comparison of distance measures in spatial analytical modeling for health service planning. BMC Health Services Research, 2009, 9: 200.

Song, X., Xu, Z., Wang, M., et al. Experimental study on the wellbore-cleaning efficiency of microhole-horizontal-well drilling. SPE Journal, 2017, 22(4): 1189-1200.

Sun, X., Tao, L., Zhao, Y., et al. Numerical simulation of hole cleaning of a horizontal wellbore model with breakout enlargement section. Mathematics, 2023, 11(14): 3070.

Sun, X., Wang, K., Yan, T., et al. Review of hole cleaning in complex structural wells. The Open Petroleum Engineering Journal, 2013, 6: 25-32.

Tomren, P. H., Iyoho, A. W., Azar, J. J. Experimental study of cuttings transport in directional wells. SPE Drilling Engineering, 1986, 1(1): 43-56.

Ulker, E., Sorgun, M. Comparison of computational intelligence models for cuttings transport in horizontal and deviated wells. Journal of Petroleum Science and Engineering, 2016, 146: 832-837.

Vaziri, E., Simjoo, M., Chahardowli, M. Application of foam as drilling fluid for cuttings transport in horizontal and inclined wells: A numerical study using computational fluid dynamics. Journal of Petroleum Science and Engineering, 2020, 194: 107325.

Wang, K., Yan, T., Sun, X., et al. Review and analysis of cuttings transport in complex structural wells. The Open Fuels & Energy Science Journal, 2013, 6: 9-17.

Williams, C. E., Bruce, G. H. Carrying capacity of drilling muds. Journal of Petroleum Technology, 1951, 3(4): 111-120.

Wood, D. A. Dataset insight and variable influences established using correlations, regressions, and transparent customized formula optimization, in Sustainable Geoscience for Natural Gas Sub-surface Systems, edited by D. A. Wood and J. Cai, Elsevier (Gulf Professional Publishing), Amsterdam, pp. 383-408, 2022.

Wood, D. A. Predicting total organic carbon from few well logs aided by well-log attributes. Petroleum, 2023, 9: 166-182.

Xu, J., Ozbayoglu, E., Miska, S. Z., et al. Cuttings transport with foam in highly inclined wells at simulated downhole conditions. Archives of Mining Sciences, 2013, 58: 481-494.

Yeo, L., Feng, Y., Seibi, A., et al. Optimization of hole cleaning in horizontal and inclined wellbores: A study with computational fluid dynamics. Journal of Petroleum Science and Engineering, 2021, 205: 108993.

Zakerian, A., Sarafraz, S., Tabzar, A., et al. Numerical modeling and simulation of drilling cutting transport in horizontal wells. Journal of Petroleum Exploration and Production Technology, 2018, 8: 455-474.

Zhu, N., Ding, S., Shi, X., et al. Cuttings transport: Back reaming analysis based on a coupled layering-sliding mesh method via CFD. Petroleum Science, 2023, in press, https://doi.org/10.1016/J.PETSCI.2023.06.009.

Zhu, N., Huang, W., Gao, D. Dynamic wavy distribution of cuttings bed in extended reach drilling. Journal of Petroleum Science and Engineering, 2021, 198: 108171.

Zico, M. N. A., Rahman, M. A., Yusuf, H. B., et al. CFD modeling of drill cuttings transport efficiency in annular bends: Effect of hole eccentricity and rotation. Geoenergy Science and Engineering, 2023, 221: 211380.