

Original Research Paper

Real-Time Simulation Method Using Monte Carlo and Clustering Algorithms: A Case Study Covering Drill Bit Wear

Rodolfo Wilvert Reitz and Paulo José de Freitas Filho

*PerformanceLab, Departamento de Informática e Estatística,
Universidade Federal de Santa Catarina, Florianópolis, SC, Brazil*

Article history

Received: 01-09-2017

Revised: 22-01-2018

Accepted: 17-02-2018

Corresponding Author:

Rodolfo Wilvert Reitz,
Universidade Federal de Santa
Catarina, P.O. Box 476, 88040-
900, Florianópolis, SC, Brazil
Email:
rodolfo.reitz@posgrad.ufsc.br

Abstract: Minimization of the cost of drilling an oil well is primarily achieved by reducing the operation completion time, which in turn can be achieved by increasing the Rate of Penetration (ROP). The ROP is the result of a combination of factors, such as lithological formation, operational parameters and bit wear. This paper addresses bit wear during drilling, using a method that combines a physical equation, techniques for risk analysis and data mining to estimate the behavior of bit wear per meter drilled. Experiments were conducted with real-world data to test the method's validity and accuracy and the results demonstrated the relevancy of this approach.

Keywords: Simulation, Monte Carlo, Risk Analysis, Drill Bit Wear, Clustering Algorithms

Introduction

Machine prognostics is an area of engineering dedicated to the study of methods for estimating the Remaining Useful Life (RUL) of a system, machine, or component (Widodo and Yang, 2011). There are two main approaches to estimate RUL (Chen *et al.*, 2012). The first is physical and more complex and difficult to perform, but offers more precise results. The second option is to construct predictive models based on data. An example of this approach is models based on neural networks.

When drilling wells for extraction of petroleum, the behavior of bit wear, or the bit's RUL, is a crucial factor in the prediction of the final cost of a drilling operation. However, it is not currently possible to measure bit wear during drilling (Lin and Ting, 1996). Because of this limitation, there are a number of models for estimating bit wear during the drilling process available in the literature, the majority of which are analytical.

Analytical models and simulation techniques can be combined into a hybrid approach to risk analysis for estimating the RUL of the bit, to provide support for decision-making (Mostafavi *et al.*, 2011). An analytical model can be employed to estimate bit wear and the Monte Carlo Method (MCM) can be used to calculate the risk (probabilistic) associated with the predicted degree of wear.

While MCM may be the most widely-used method for generating scenarios for risk assessment, it is important that certain precautions are taken when using this method. If the relationships between variables are not well studied and dealt with inadequately, it is possible that the method will produce scenarios that are incompatible with the real-world system. For instance, it may produce incompatible values for two or more input variables.

Determination of the relationships between variables can be a complex task. Certain techniques for separating data into clusters, so that each cluster contains observations with similarities to each other, can minimize the problem of handling relationships between variables.

One important issue related to analytical models for predicting bit wear in real time is that the majority of them require inclusion of certain coefficients to achieve model fit. In general, these coefficients are based on the behavior of the variable for bit wear and they are identified using historical data. Therefore, if data from well P , drilled in the past, are employed as the basis for setting the coefficients of a model, this will result in good predictions if the well to be drilled has similar characteristics to those of well P .

Using a prognostic method to estimate the degree of bit wear has relevance for optimization of the drilling process. If the decision to withdraw a bit that has become

completely worn is delayed or is taken prematurely, the cost of drilling can increase significantly.

The main objective of this article is to present a way for the prediction of bit wear in real time. The process employs historical data and its risk analysis estimated degree of wear is updated according to data from the drilling run in progress. The method proposed is a hybrid based on an analytical model for predicting bit wear and the MCM.

This paper is organized as follows: Section 2 a brief overview of clustering, random sampling, the MCM and risk management; Section 3 discusses related work; Section 4 describes the proposed method; Section 5 the experiments conducted with the method; and, finally, Section 6 contains a discussion of the results as well as a conclusion of the study.

Background

Clustering

Clustering is a process for grouping a set of observations into several clusters according to a similarity function (like Euclidean distance), so that similar observations are in the same cluster (Sumathi and Sivanandam, 2006).

Due to its simplicity, *k*-means is the most popular algorithm for clustering. Figure 1 illustrates how this algorithm works. It is basically a loop that attributes all observations to their nearest cluster and finds the centroid of each cluster. The loop only ends when no more observations change clusters. There are several ways to select *k* observations to be the initial centroids one of which is to select the ones that are most distant from each other.

Random Sampling

The Probability Density Function (PDF) for a random continuous variable *X* can be represented by a non-negative function *f(x)*, with the area between the *x*-axis and the curve equal to 1, which describes the probability *P* that the random variable will take a given value of *x*. The probability of a given interval [*a*, *b*] occurring is given by the integral of *f(x)* in the required interval $\int_a^b f(x)dx$.

Considering the PDF of variable *X*, a cumulative distribution function *F(x)*, which represents the probability that a value less than *x* will occur, i.e., $F(x) = P(X \leq x)$. To generate a random value *x*, a random number *r* in the range 0 to 1 is drawn according to a uniform distribution, that means $P(X \leq x) = r$. The random value *x* is obtained using the inverse function of *F(x)* written as $G(F(x)) = G(p) = x$. Figure 2 illustrates the graphical relationship between *F(x)* and

$G(F(x))$. The example shows that a random number $r = 0.5$ corresponds to $x \approx 270.9$ (Costa e Lima *et al.*, 2015). This concept is widely employed in many methods for generating samples, such as the MCM.

The PDF of variable *X* can be obtained from a specialist or (more commonly) by using some type of fitting method, such as Chi-Square or K-S method.

Monte Carlo Method

The MCM is used to generate values for a simulation model. A complete simulation involves hundreds or even thousands of scenarios, in each of which a sample of each of the model's random input variables is obtained. Using this sampling scheme, for each scenario, a value for each output variables is calculated and stored. In the end, the distribution of the values stored reflects the probability of the outputs that are possible in the system being modeled. Figure 3 illustrates the MCM schematically.

One of the major advantages of using the MCM is the ability to use computational tools to test a wide range of scenarios and their respective results.

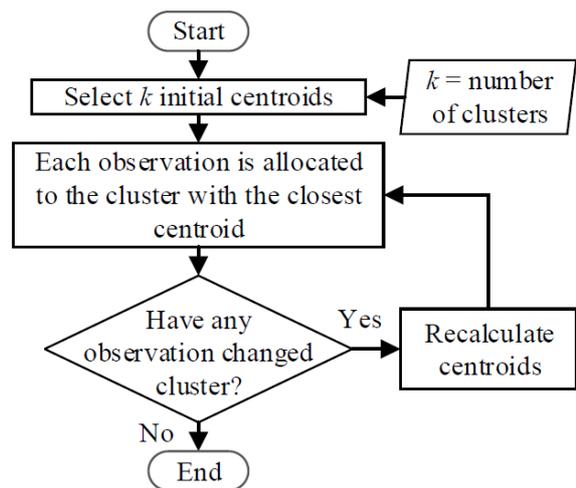


Fig. 1: *k*-means flow diagram

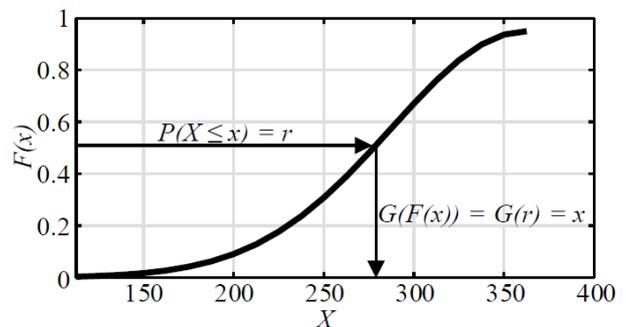


Fig. 2: Illustration of the relationship between *F(x)* and $G(F(x))$

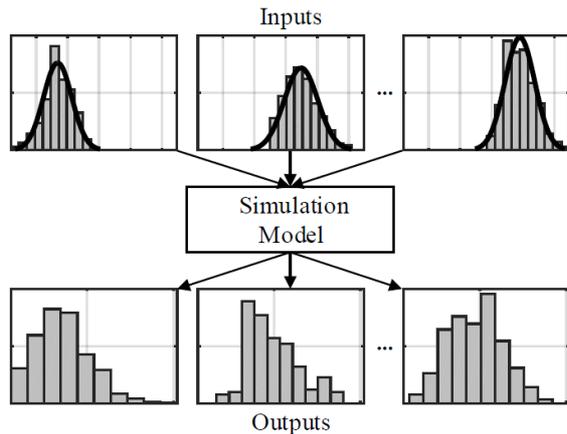


Fig. 3: Schematic illustration of uncertainty propagation by MCM

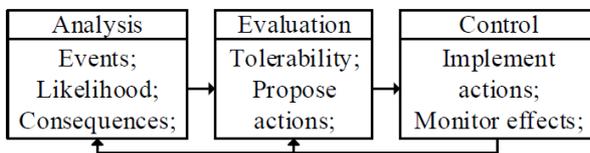


Fig. 4: Risk analysis, evaluation and management

Risk Management

Risk management is a continuous process with the objective to identify, analyze and assess relevant events (critical or beneficial) in a system or related to an activity, to decide whether or not the risk of an event occurring is tolerable and to identify and introduce risk control actions. The elements of risk management are illustrated in Fig. 4.

A commonly used definition for Risk Analysis (RA) is “Systematic use of available information to identify hazards and to estimate the risk to individuals, property and the environment” (IEC 60300-3-9, 1995). RA is used to identify the causes of relevant events, to determine the possible consequences, to identify actions to prevent or favor and to form a basis for deciding whether or not the risk related to a system is tolerable. RA is carried out to provide answers to the following three main questions (Kaplan and Garrick, 1981):

- *Q1. What can go wrong?* To identify the possible relevant events to the system and the uncertainties that can trigger them
- *Q2. What is the likelihood of that happening?* To determine the likelihood of uncertainties identified in *Q1* that may lead to the relevant event
- *Q3. What are the consequences?* To identify the consequences of events that were described in *Q1* to the system and action to prevent or mitigate them

In the Risk Evaluation (RE), judgments are made on the tolerability of the risk on the basis of a RA taking into account the consequences and/or some other risk acceptance or rejection criteria. Another objective of RE is to propose actions to be used in risk control, which may involve adjustments in the RA process.

In Risk Control (RC), the decision to stop or continue a particular activity at the current risk level must be made. Furthermore, RC can involve modification of existing actions or implementation of new ones and monitor the effects of these changes.

Many researchers use MCM to study the uncertainty propagation in mathematical models, similar as shown in Fig. 3, which consists of choosing probability distribution functions that best reflects the uncertainty behaviors (for each of the input variables). Afterwards, a repeated sampling of values from distributions to use as model input to calculate the output values is performed. Finally, a large number of output values is obtained, that may be plotted as a histogram to identify the output distribution shape.

Related Work

Drilling an oil well can involve high costs. Maintaining all drilling parameters in order to minimize this cost is a complex task.

Many researchers have been working on this problem, developing equations and mathematical models to represent the drilling process, enabling drilling teams to make decisions based on the results produced by these models. For example, optimizing Rate of Penetration (ROP), i.e., achieving the highest possible drilling speed, resulting in a lower cost, may be one of the objectives of these models.

Models can also help decide the correct time to remove the drill bit due to its wear. The costs involved in performing a drill change are great. On the other hand, keeping a worn drill can result in even greater costs than its replacement.

One of the most accepted models for estimating ROP is the Bourgoyne and Young (BYM) (Bourgoyne Jr. and Young Jr., 1974; Bourgoyne *et al.*, 1986). The model considers ROP to be the product of eight exponential functions ($ROP = f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5 \cdot f_6 \cdot f_7 \cdot f_8$), each of them has its own exponent, denoted by a_1 through a_8 , which are found through multiple regression from historical data. Each exponential represents a factor that affects ROP. The seventh function models the drill wear ($f_7 = e^{-a_7 h}$), where h is the fractional bit tooth wear. However, by observing the term (f_7) and simulating the drilling of a well with the same h , the wear (f_7) will have the same effect on ROP throughout the drilling, which is not in line with reality, since ROP tends to decrease as drill wear increases.

Warren (1987) published a model (Equation 1) for predicting ROP, which also does not consider that drill wear affects the ROP. However, Hareland and Hoberock (1993) later addressed this limitation by adding a function for bit wear W_f (Equation 2) to the Warren model:

$$ROP = \left[\frac{f_c \times \left(\frac{a \times CCS^2 \times d_b^2}{RPM \times WOB^2} + \frac{b}{RPM \times d_b} \right)}{+ \frac{c \times d_b \times \gamma_f \times \mu}{F_{jm}}} \right]^{-1} \quad (1)$$

Where:

- ROP = Rate of penetration (m/h)
- a, b, c = Bit coefficients (dimensionless)
- CCS = Rock confined compressive strength (psi)
- d_b = Bit diameter (in)
- RPM = Bit revolutions (rev/min)
- WOB = Weight on bit ($klbf$)
- γ_f = Drilling fluid density (Ib/gal)
- μ = Drilling fluid plastic viscosity (c_p)
- f_c = Chip-hold down function (dimensionless)
- F_{jm} = Modified jet impact force function ($klbf$)

$$ROP = W_f \times (\text{Warren ROP Model})$$

$$W_f = 1 - \frac{W_c \times \sum_{i=1}^m WOB_i \times RPM_i \times Abr_i \times CCS_i}{8} \quad (2)$$

Where:

- W_f = Wear function (dimensionless)
- W_c = Wear coefficient (dimensionless)
- m = Total depth ($meter$)
- Abr = Abrasiveness (dimensionless)

The function W_f (Equation 2) is formed by the linear combination of four parameters that are obtained along the drilling and a W_c coefficient. When drilling a well with similar characteristics to the well used for fitting the model, the same value can be applied to the W_c coefficient. When correctly fitted, W_f (Equation 2) will produce a value between 0 and 1, where 0 indicates a new, unused bit and 1, a completely worn bit. This value can be converted and vice-versa, to the IADC Dull Bit Grading scale with values from 0 (new) to 8 (worn). The W_f (Equation 2) function seems to be more suitable for estimating drill wear in real time as it considers all drilling data to the desired depth.

Other researchers have also addressed the wear of the drill. Liu *et al.* (2014) provide for the loss of the cut material from the drill cutters. Rashidi *et al.* (2008) presented a method to estimate the wear of the drill bit from the combination of BYM and Mechanical Specific Energy (MSE), which later used CCS instead of BYM (Rashidi *et al.*, 2010).

Lin and Ting (1996) described an approach that employed a neural network, in which the inputs to the network are the average thrust force, torque, feedrate, diameter and revolutions per minute. The output from the network is the average wear (final wear divided by the total number of meters drilled).

Many previous studies in the drilling area of the petroleum industry have employed a similar approach to risk analysis, which consists in generating the inputs for an analytical model using MCM:

- One of the experiments conducted by Udegbumam (2015) employed an analytical model to estimate permissible limits for bottomhole pressure
- Mostafavi *et al.* (2011) and Peterson *et al.* (1993) used an analytical model to estimate ROP and duration of drilling, respectively
- Cunha (2004) dealt with the case of bit breakage at the well bottom, for which there are two possible solutions, the first is to fish for the bit, which will take an unknown length of time and the second it to drill a sidetrack around the bit; both solutions increase the drilling costs. The authors used risk analysis in combination with an analytical model of costs in order to answer the question of which option is more cost-effective

These three studies present their results as probability distributions, which makes it possible to estimate the likelihood of a given output value falling within a given interval.

Proposed Method

The proposed method combines an analytical model, simulation techniques (MCM) and data mining (clustering) to evaluate the bit wear during the drilling of an oil well. In contrast with analytical models, there is only one output for a given input. This proposal will produce many outputs for a single input, along with their degrees of uncertainty.

To estimate the drill wear, the W_f (Equation 2) function of Hareland and Hoberock (1993) was employed for its simplicity and adequacy to the problem since it considers the entire length drilled by the drill. Since the available database did not contain information on the Abr variable, it was removed from Equation 2. The CCS variable was substituted by UCS (Unconfined Compressive Strength). We believe that UCS is an excellent substitute for the two variables that were excluded. Due to these limitations, (Equation 3) was adopted in the method:

$$W_f = 1 - \frac{W_c \times \sum_{i=1}^m WOB_i \times RPM_i \times UCS_i}{8} \quad (3)$$

where, UCS = unconfined compressive strength (psi); All other symbols defined in Equation 2.

The method is composed of three steps, which are described below, whose purpose is to identify Probability Density Functions (PDFs) associated to the input variables (WOB , RPM and UCS) per cluster and the PDFs related to the W_c coefficient per set of historical data.

Step 1

In this first step, based on historical data files, two procedures are performed: (a) Clustering, separating all data into k groups, using the k -means technique and (b) fitting, where the likely probability distributions (PDFs) are identified for each input variable (WOB , RPM and UCS) in each group formed. For example: With the data obtained in 3 perforations $d = 3$ sets of historical data are obtained. At the end of this step, $k \times 3$ distributions (one for each variable) are obtained. Figure 5 illustrates this first step.

Step 2

The second step is dedicated to the generation of s sampled datasets for each historical dataset. The greater the s quantity of generated datasets, the better the fitting of the distributions that will represent W_c . These are generated from a given historical h dataset in such a manner that for each i record, a window of $[i-w, \dots, i]$ records is utilized (where w is the size of the window) to seek the C cluster that best represents, i.e., the cluster in which the sum of the distances between $[i-w, i]$ records and the centroid is the lowest. This window was used to smooth the selection of clusters when traversing a historical dataset. The

larger the window size, the smaller the variability of the selected clusters.

After identifying the best C cluster, their PDFs are used to generate s new records, each of which is associated to the datasets 1, 2, ..., s being generated. Figure 6 illustrates this step. The example refers to the record $i = 3$ of a historical h dataset, considering $w = 2$. After performing this procedure for all records in h , a total of s data sets were generated, as shown in Fig. 7.

This step is performed for each historical dataset, so there will be generated s datasets for each of them.

Step 3

The third and final stage is to find the distributions that will represent the W_c coefficient of the analytical model (Equation 3) in each historical h dataset. For each h (Fig. 8), a PDF is obtained using the s values for W_c obtained by regression from their generated datasets.

The distributions associated with the inputs (per cluster) and the distributions for the W_c coefficient, by historical dataset, can be extended to simulate the behavior of drill wear during drilling in real time. This application is described below.

Real-time Application

The real-time simulation consists of using the PDFs of C cluster and the historical D dataset to generate s new record samples for the WOB , RPM and UCS variables, with the respective values that the W_c coefficient must assume throughout the drilling. The C cluster and the D dataset are the ones that best represent (the shorter distance) the last records observed in real time.

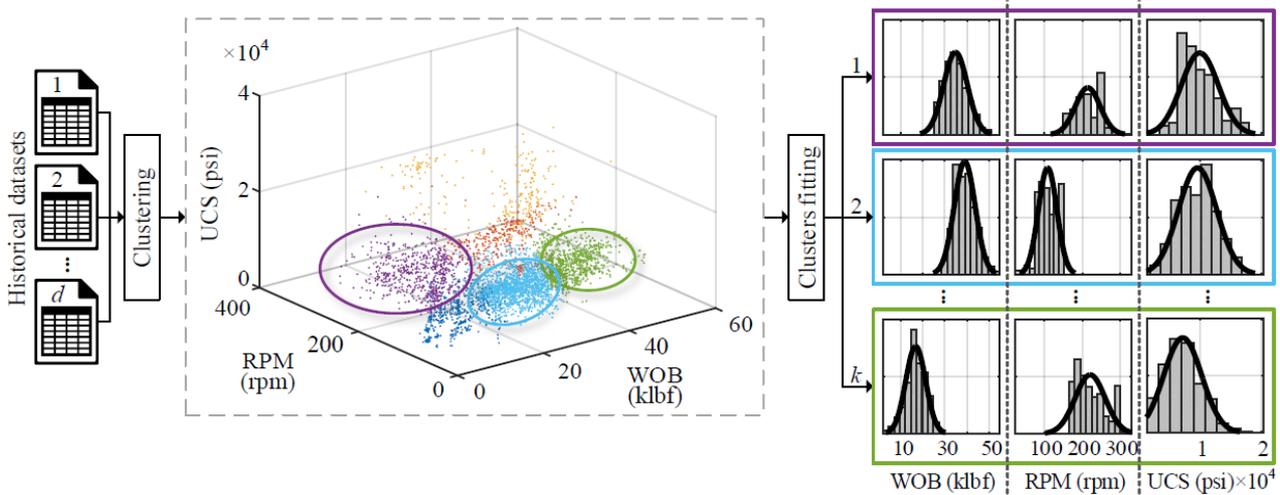


Fig. 5: First step of the method: Data clustering and fitting of the PDFs for each variable in each cluster

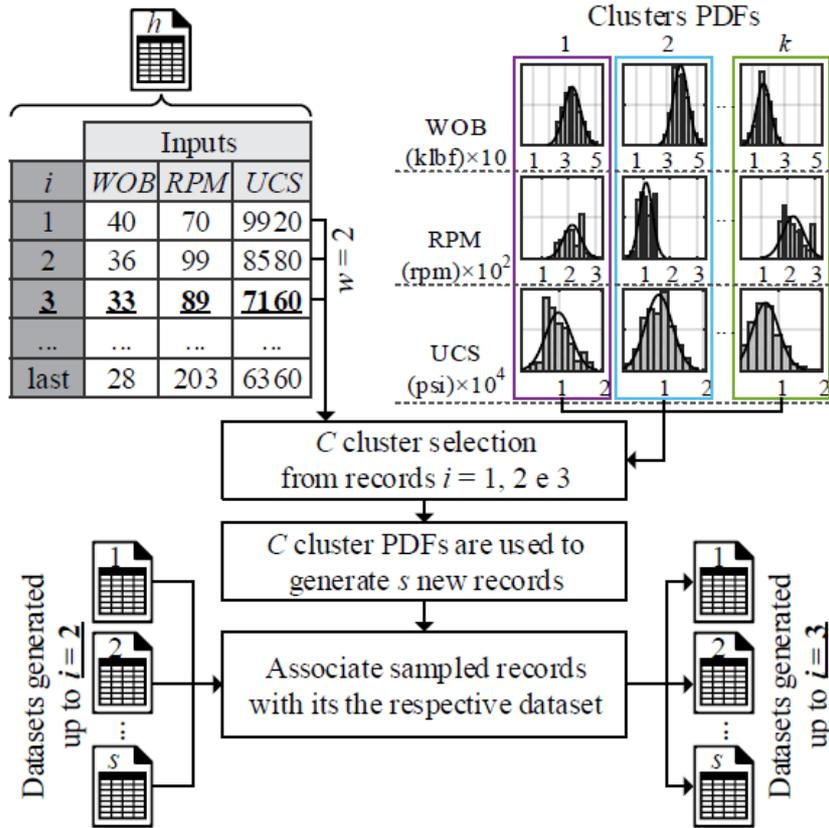


Fig. 6: Update of the s generated datasets for a record $i = 3$ in a historical h dataset

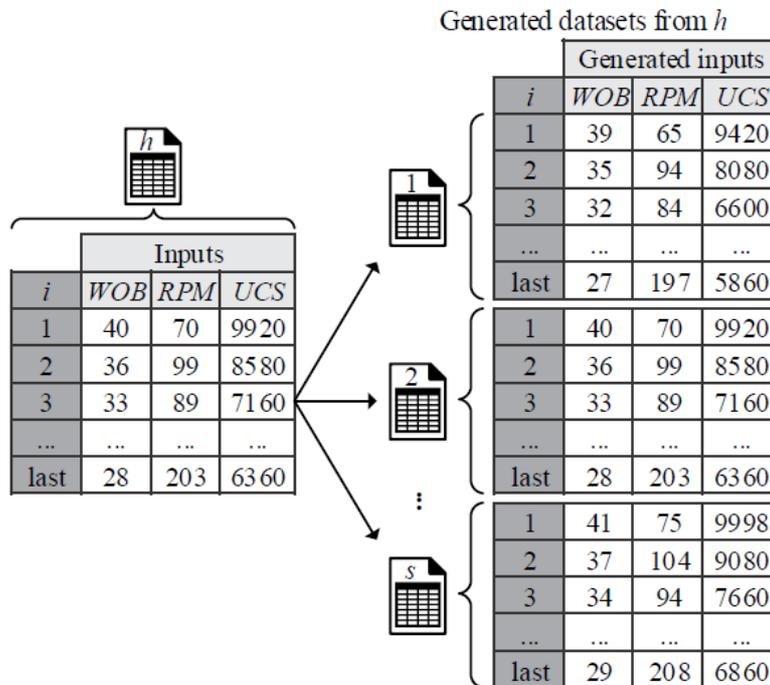


Fig. 7: Generated datasets from a historical h dataset

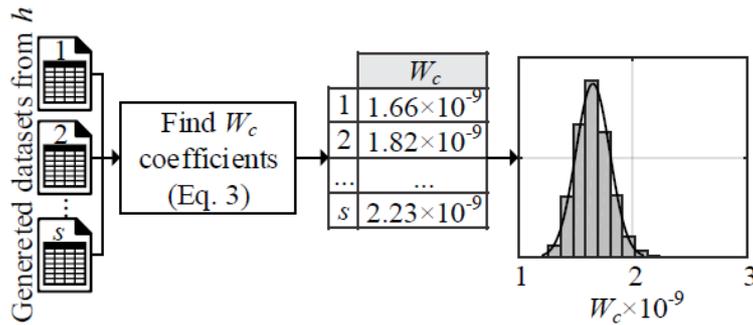


Fig. 8: The PDF identification step for the W_c coefficient for a given historical h dataset

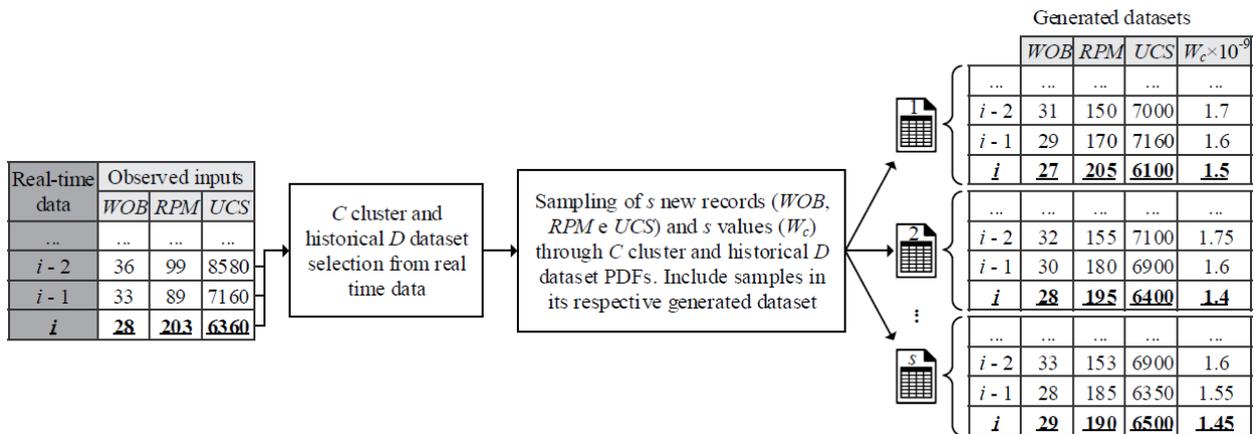


Fig. 9: Update of generated datasets in the given i meter

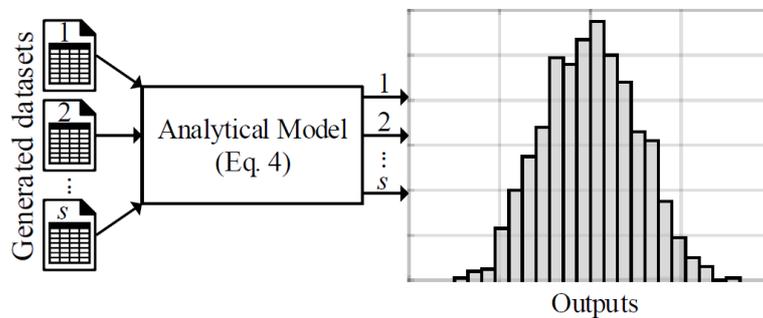


Fig. 10: Application of the generated datasets in the analytical model

For the use in real time simulation, a minor alteration was made to Equation 3, so that the W_c coefficient could take different values for each drilled meter, as seen in Equation 4:

$$W_f = 1 - \frac{\sum_{i=1}^m W_{ci} \times WOB_i \times RPM_i \times UCS_i}{8} \quad (4)$$

where, W_{ci} is wear coefficient at i meter (dimensionless); all other symbols defined in Equation 3.

As shown in Fig. 9, for each i meter drilled, the observed information of WOB , RPM and UCS in meters

$[i-w, i]$ are used to select the most appropriate C cluster, so its PDFs are used in the sampling of s new records. Each sampled record is included in its respective generated dataset. Similarly, the historical D dataset is selected and its PDF, which represent the W_c coefficient, is used in the sampling of s values for the W_c coefficient.

At this point, the s datasets are up to date for the current i meter with new records and values for the W_c coefficient. Then, as shown in Fig. 10, each of them is used as analytical model input (Equation 4) to obtain the respective wear. Thus, the s outputs can be presented through a histogram, allowing visualization of the probable values of the bit wear and their probabilities.

Experiments

This section details the use of the method addressing the problem of drill bit wear during drilling of an oil well.

One issue to be taken into consideration is that the IADC Dull Bit Grading scale is a classification code that can take any integer from within a set of 9 classes: 0, 1, 2, 3, 4, 5, 6, 7 and 8. The value of W_f can take infinite values within the interval $[0, 1]$. When it is transformed into an IADC Dull Bit Grading scale score, values may occur that are not within the set of IADC classes, making it necessary to perform adjustments for an acceptable classification. When Equation 3 is correctly fitted, W_f takes values from 0 to 1. However, differences between the well used to fit the W_c coefficient and the well in the process of being drilled can cause outputs beyond the interval expected for W_f , creating difficulties for conversion to the IADC Dull Bit Grading scale. A result value greater than 8 indicates that the bit is completely worn out.

The method proposed was tested using two databases. The first one was used in the experiments *I*, *II* and *III*, in which contains drilling data of 35 Polycrystalline Diamond Compact (PDC) type drill bits with a diameter of $8\frac{1}{2}$, ranging from 30 to 446 drilled meters and the wear is 1 to 8 on the IADC Dull Grading scale. The second one was employed in experiment *IV*, it has data from 21 Impregnated type drill bits with diameters of $8\frac{1}{2}$ and $12\frac{1}{4}$ with drilled meters and final bit wear varying respectively from 28 to 467 m and from 1 to 8 on the IADC Dull Grading scale.

The drilling data for each drill bit contains: The final bit wear, the operational parameters (*WOB* and *RPM*) and the lithological profiles (*UCS*) per drilled meter.

Experiment I

Of the 35 runs in the dataset, 28 were used to fit the PDFs per cluster (inputs) and per historical dataset (W_c) and the other 7 were used to test the method.

The *k*-means algorithm was used in step 1 to identify clusters. The number of clusters was set at $k = 6$. Figure 11 shows the clusters obtained and Fig. 12 demonstrates the PDFs fitted in three of the clusters.

From the histograms of clusters 2 and 6 in Fig. 12, it is possible to observe that *RPM* has a behavior inversely proportional to *WOB*. This situation was detected due to the use of clusters. This implies that the data generated by the PDFs of these clusters will not contain scenarios in which *WOB* and *RPM* are incompatible, i.e., *WOB* and *RPM* elevated at the same time.

In step 2, a total of $s = 3000$ sampled datasets (or simulated runs) were generated using the MMC. The window size of $w = 5$ m was used to identify the most representative cluster and historical run.

The W_c histograms and their respective fitted distribution for some of the runs (step 3) are showed in Fig. 13. The observed differences in W_c results are due to the observed variations in input variables such as: Length drilled by the bit and final wear.

The objective of experiment 1 is to compare the real final wear with the one predicted by the proposed method, using the remaining 7 runs that have not yet been used and were reserved for this test. The median of the simulated results for the last drilled meter is taken as an estimate of final bit wear. The RMSE (Root-Mean-Square Error) was used as an indicator of the accuracy, measuring the mean difference on the same scale between the real result and the predicted one; lower RMSE scores indicate better accuracy.

The real and estimated final wear for each test run are shown in Fig. 14. The RMSE obtained for the 7 test runs was 1.03 in IADC Dull Bit Grading scale. This value is equivalent to approximately 12.5% of the wear scale used.

However, the obtained error disregards the response of the method as not punctual, but presented in the form of a histogram, with probabilities associated with possible values in an interval. Thus, even if the point estimate is not precise, the value of the real final wear may be, with a greater or lesser probability, among the possible wear values predicted by the histogram.

Figure 15 illustrates the behavior of the median values for wear per drilled meter. The test run selected had a drilling depth of 254 m and the final bit wear was class 5 in IADC Dull Bit Grading scale.

Figure 16 illustrates histograms from simulations for meters 64, 128, 192 and 254. The final wear score estimated for this run was 4.82 in IADC Dull Bit Grading scale, whereas the real wear score was 5. The histogram for the last meter in Fig. 16 (254 m) shows that there was a probability of a wear score of 5. We know that the estimate does not provide a fixed value and even though 4.82 was the most common wear score, in some of the simulations wear was closer to 5.

With the 28 runs used to fit the PDFs per cluster (inputs) and by historical dataset (W_c), the performance was slightly smaller when compared to the test runs, since the RMSE obtained was 1.64 on the IADC Dull Grading scale. The real and estimated final wear for each run used to fit the PDFs are shown in Fig. 17.

Experiment II

The previous experiment evaluated the method with 28 runs selected for fitting the PDFs and another 7 for testing. In this experiment, the three steps of the method were performed 1000 times. Each time, 28 runs (80%) were randomly selected for fitting and the remaining 7 (20%) were used for testing. The RMSE obtained by each execution were stored.

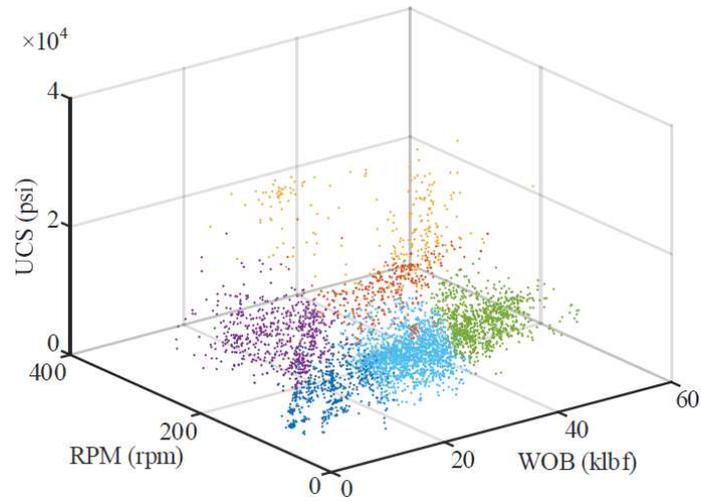


Fig. 11: Clusters obtained in step 1

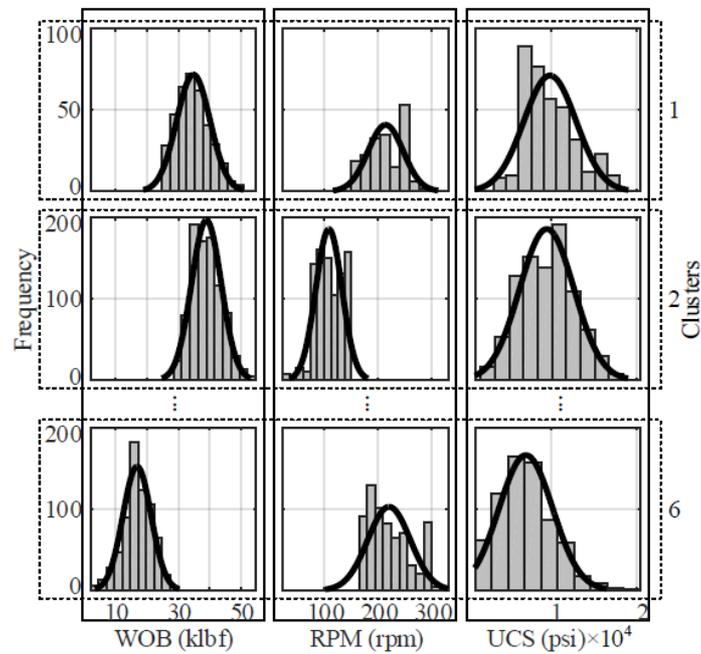


Fig. 12: Fitted PDFs to three clusters

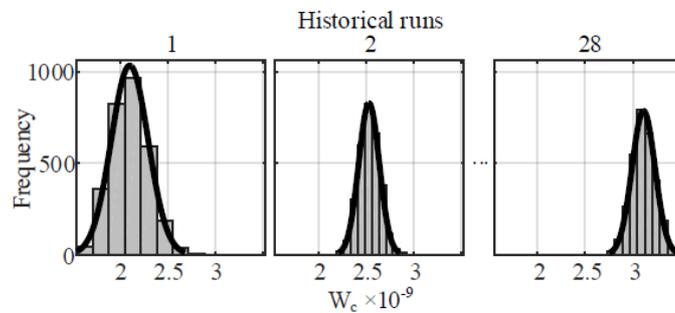


Fig. 13: PDFs identified for the W_c coefficient in three historical runs

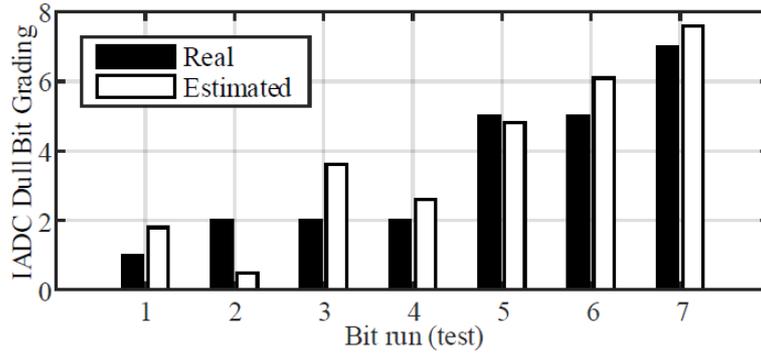


Fig. 14: Comparisons between real wear and predicted wear

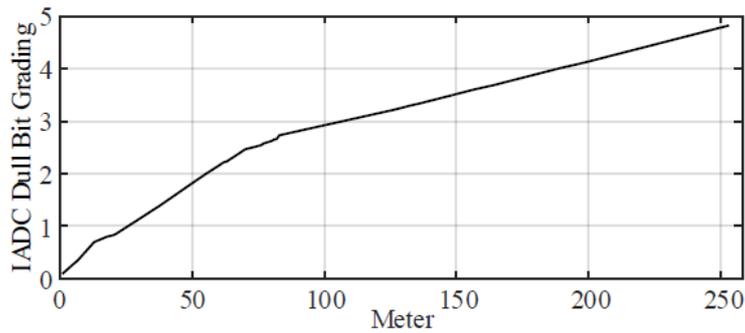


Fig. 15: Median value of the simulated wear per drilled meter

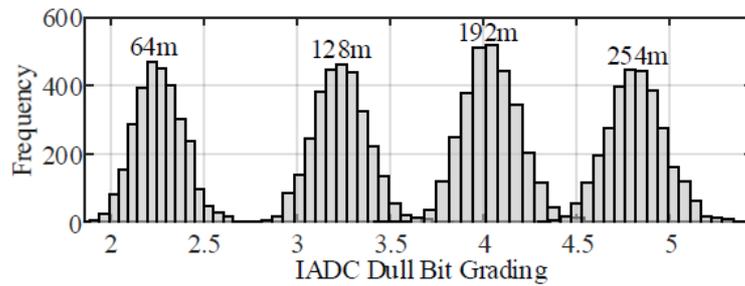


Fig. 16: Histograms for simulations of 64, 128, 192 and 254 m

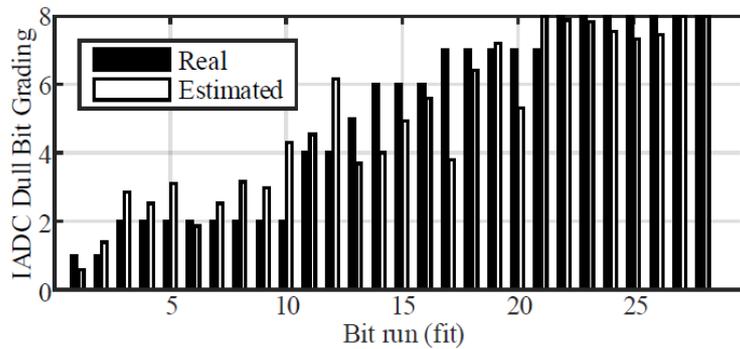


Fig. 17: Comparisons between real wear and predicted wear for runs used to fit the PDFs

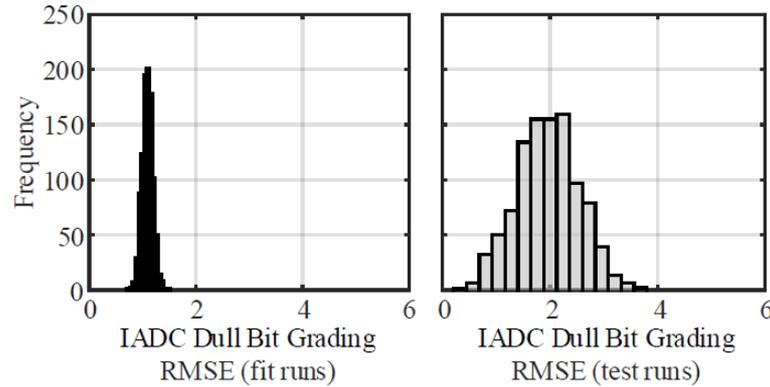


Fig. 18: RMSE obtained in 1000 executions of the proposed method with different runs for fitting and testing

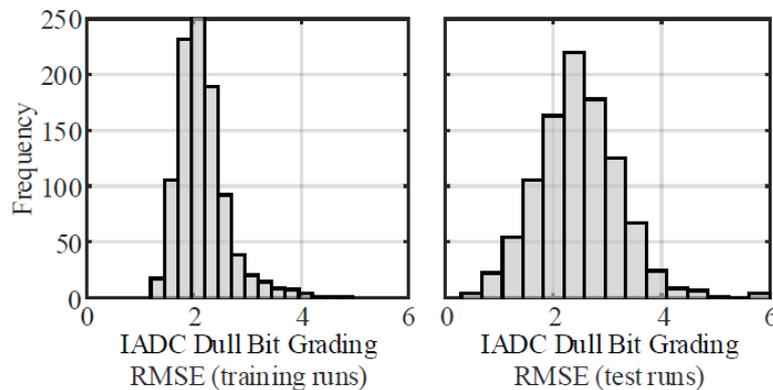


Fig. 19: RMSE obtained in 1000 neural networks with different training and test runs

Through sensitivity analysis, the total k clusters to be formed in step 1 and the size of the window of record w used in step 2 where the best results obtained were $k = 8$ and $w = 6$, respectively. In order to reduce the computational time, in each execution $s = 400$ runs are generated in step 2 to carry out the fitting process of the PDF that represents W_c coefficient in step 3.

For the simulations of the fitting and testing runs, the same parameters $w = 6$ and $s = 400$ were used. Figure 18 shows the two histograms for the sample of 1000 RMSE obtained. The mean RMSE and standard deviation obtained for the fitting runs were 1.09 and 0.1, respectively (IADC Dull Bit Grading). Whereas for the testing runs, it reached an average RMSE and standard deviation of 1.95 and 0.57, respectively (IADC Dull Bit Grading).

Experiment III

A comparative experiment was carried out using an approach similar to that of Lin and Ting (1996), which uses the mean values of the run variables as input and the average wear output per meter.

In this experiment, similar to the previous one, 1000 neural networks were trained and tested with 80% and 20% of the randomly selected runs at each execution for training and testing, respectively.

The neural network used was Cascade-forward with 15 neurons in the hidden layer and the training function was Levenberg-Marquardt. The average runs of the same WOB , RPM and UCS variables of the previous experiments were used as input to the network.

In the 1000 executions, the final wear predicted by simulating the training and testing runs per meter were stored and compared to the real wear through the RMSE, which are shown in Fig. 19.

On average, the RMSE is 2.16 and 2.47 with standard deviation 0.47 and 0.76 on the IADC Dull Bit Grading scale for the training and testing runs, respectively.

Even the neural network for the training runs (80%) had difficulties in estimating wear, since the proposed method had a lower RMSE average and lower standard deviation for both, runs used for fitting the PDFs (80%) and for the testing runs (20%).

Experiment IV

Similar to experiment II, but using data from Impregnated type drill bits. The method was executed 500 times. Each time, 16 ($\approx 76\%$) and 5 ($\approx 24\%$) runs are randomly selected to fit the PDFs and test the method respectively.

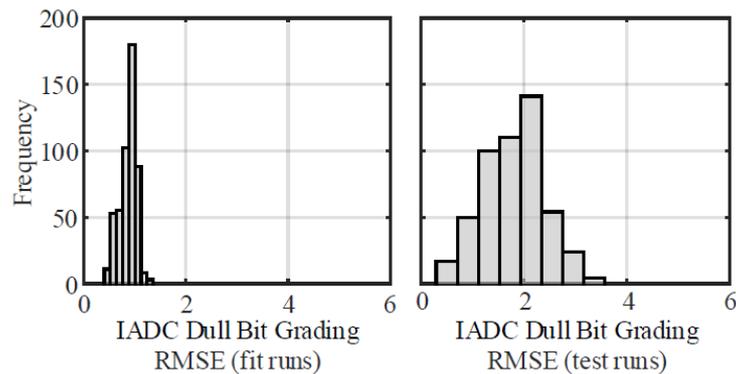


Fig. 20: RMSE obtained in 500 executions of the proposed method with different runs for fitting and testing, using data from Impregnated type drill bit

The parameters used for step 1 was $k = 4$ and in step 2 were $w = 8$ and $s = 500$. The same values for the parameters w and s were used when simulating fitting runs and testing runs. Figure 20 shows the two histograms for the sample of 500 RMSE obtained.

The standard deviations obtained in this experiment were close to those obtained in experiment II. However, the mean values for RMSE for the fitting and testing runs were slightly better, reaching 0.86 and 1.81 respectively.

Conclusion

The results of the conducted experiments show that the proposed simulation technique was capable of producing estimates of bit wear that were close to the real final wear. Since the method's output is a distribution of wear, it is possible that the true value falls within the distribution at a certain level of probability.

One of the objectives in the drilling process is to maximize ROP, so that the cost of drilling is reduced. Bit wear has a direct influence on ROP and the method described here provides support for decision-making in situations in which the decision-maker needs to know whether or not the bit should be pulled because of wear.

One of the advantages of the method proposed here is that it does not offer the decision-maker a single value for estimated wear at each meter drilled. Rather, it provides a histogram illustrating the risk of accepting or rejecting the predicted bit wear value for that depth.

Another advantage of the proposed method is that there is no need to provide the coefficients for the analytical model during its use in real time. The historical data provide these coefficients, selected on the basis of similarity. Furthermore, the proposed method here is not dependent on the analytical model of wear employed. The engine of the simulation model in this study were Equations 3 and 4, for bit wear, but this could easily be substituted with a different wear equation. It should also be pointed out that combining the MCM with clusterization reduces the likelihood that scenarios incompatible with the drilling process will occur.

The criterion used to select the historical run in the real-time simulation during drilling, which is then used to draw values for the W_c coefficient creates considerable difficulty and has a very significant effect on the results of the experiments in the presented method.

This study estimates the behavior of bit wear in real time during the drilling process by combining analytical models, risk analysis techniques and data mining. Analysis of the results of the simulations conducted using real drilling data to test the method supports the conclusion that in the majority of cases, the results are compatible with the results in the real-world data, as shown by the RMSE achieved in the experiments.

Acknowledgement

To Professor Paulo José de Freitas Filho for his support, time, revisions and incentives. To the members of the laboratory that were part of this research directly or indirectly: Mariana Dehon Costa e Lima, Silvia Modesto Nassar, Mauro Roisenberg and Tiago Mazzutti.

Funding Information

To the Coordination of Improvement of Higher Education Personnel through the Social Demand Program (CAPES/DS) that fostered this research associated with the Post-Graduation Program in Computer Science of the Federal University of Santa Catarina.

Author's Contributions

Both authors contributed equally to this research.

Ethics

This study is self-contained and includes unpublished material. The authors confirm that they have read and approved this document and there is no ethical issue involved.

References

- Bourgoyne Jr, A.T. and F.S. Young Jr., 1974. A multiple regression approach to optimal drilling and abnormal pressure detection. *Society Petrol. Eng. J.*, 14: 371-384. DOI: 10.2118/4238-PA
- Bourgoyne, A.T., K.K. Millheim, M.E. Chenevert and F.S. Young. 1986. *Applied drilling engineering*. SPE Textbook Series Vol. 2. ISBN: 9781555630010. pp: 502.
- Chen, C., B. Zhang and G. Vachtsevanos, 2012. Prediction of machine health condition using neuro-fuzzy and Bayesian algorithms. *IEEE Trans. Instrument. Measur.*, 61: 297-306. DOI: 10.1109/TIM.2011.2169182
- Costa e Lima, M.D., S.M. Nassar and P.J. de Freitas Filho, 2015. Simulation of oil drilling time series using Monte Carlo and Bayesian Networks. *Proceedings of the Winter Simulation Conference*, Dec. 6-9, IEEE Xplore Press, Huntington Beach, CA, USA, pp: 1195-1205. DOI: 10.1109/WSC.2015.7408245
- Cunha, J.C., 2004. Risk analysis application for drilling operations. *Proceedings of the Canadian International Petroleum Conference*, Jun. 8-10, Petroleum Society of Canada. DOI: 10.2118/2004-148
- Hareland, G. and L.L. Hoberock, 1993. Use of drilling parameters to predict in-situ stress bounds. *Proceedings of the SPE/IADC Drilling Conference*, Feb. 22-25, Society of Petroleum Engineers. DOI: 10.2118/25727-MS
- IEC 60300-3-9, 1995. *Dependability Management - Application Guide: Risk Analysis of Technological Systems*. International Electrotechnical Commission, Geneva. <https://webstore.iec.ch/publication/14642>
- Kaplan, S. and B.J. Garrick, 1981. On the quantitative definition of risk. *Risk Anal.*, 1: 11-27. DOI: 10.1111/j.1539-6924.1981.tb01350.x
- Lin, S.C. and C.J. Ting, 1996. Drill wear monitoring using neural networks. *Int. J. Machine Tools Manuf.*, 36: 465-475. DOI: 10.1016/0890-6955(95)00059-3
- Liu, Z., C.N. Marland, D. Li and R. Samuel, 2014. An analytical model coupled with data analytics to estimate PDC bit wear. *Proceedings of the SPE Latin America and Caribbean Petroleum Engineering Conference*, May 21-23, Society of Petroleum Engineers. DOI: 10.2118/169451-MS
- Mostafavi, V., G. Hareland and B.S. Aadnoy, 2011. Assessment And Mitigation of Drilling Risks Using ROP Model Coupled With Monte-Carlo Simulation. *Proceedings of the 45th US Rock Mechanics/Geomechanics Symposium*, Jun. 26-29, American Rock Mechanics Association, San Francisco, California. <https://www.onepetro.org/conference-paper/ARMA-11-128>
- Peterson, S.K., J.A. Murtha and F.F. Schneider, 1993. Risk analysis and Monte Carlo simulation applied to the generation of drilling AFE estimates. *Proceedings of the SPE Annual Technical Conference and Exhibition*, Oct. 3-6, Society of Petroleum Engineers, Houston, Texas. DOI: 10.2118/26339-MS
- Rashidi, B., G. Hareland and R. Nygaard, 2008. Real-time drill bit wear prediction by combining rock energy and drilling strength concepts. *Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference*, Nov. 3-6, Society of Petroleum Engineers, Abu Dhabi, UAE. DOI: 10.2118/117109-MS
- Rashidi, B., G. Hareland and A. Wu, 2010. New approach in real-time bit wear prediction. *Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference*, Nov. 1-4, Society of Petroleum Engineers, Abu Dhabi, UAE. DOI: 10.2118/136008-MS
- Sumathi, S. and S.N. Sivanandam, 2006. *Introduction to Data Mining and its Applications*. 1st Edn., Springer, Berlin, ISBN: 3540343512, pp: 828.
- Udegbunam, J.E., 2015. *Improved Well Design with Risk and Uncertainty Analysis*. Unpublished dissertation in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Stavanger, Stavanger, Norway.
- Warren, T.M., 1987. Penetration rate performance of roller cone bits. *SPE Drill. Eng.*, 2 9-18. DOI: 10.2118/13259-PA
- Widodo, A. and B.S. Yang, 2011. Machine health prognostics using survival probability and support vector machine. *Expert Syst. Applic.*, 38: 8430-8437. DOI: 10.1016/j.eswa.2011.01.038