IADC/SPE-212445-MS

Real-Time Digital Log Generation from Drilling Parameters of Offset Wells Using Physics Informed Machine Learning

Prasham Sheth, Sai Shravani Sistla, Indranil Roychoudhury, Mengdi Gao, Crispin Chatar, Jose Celaya, and Priya Mishra, SLB

This paper was prepared for presentation at the IADC/SPE International Drilling Conference and Exhibition, Stavanger, Norway, 7 – 9 March 2023.

Abstract

By 2026, USD 5.05 billion will be spent per year on logging while drilling (LWD) according to the market report from **Fortune Business Insights (2020).** Logging tools and wireline tools are costly services for operators to pay for, and the companies providing the services also have a high cost of service delivery. They are, however, an essential service for drilling wells efficiently. The ability to computationally generate logs in real time using known relationships between the rock formations and drilling parameters provides an alternative method to generate formation evaluation information. This paper describes an approach to creating a digital formation evaluation log generator using a novel physics-informed machine learning (PIML) approach that combines physics-based approaches with machine learning (ML) algorithms.

The designed PIML approach learns the relationships between drilling parameters and the gamma ray (GR) logs using the data from the offset wells. The decomposition of the model into multiple stages enables the model to learn the relationship between drilling parameters data and formation evaluation data. It makes it easier for the model to generate GR measurements consistent with the rock formations of the subject well being drilled. Since the computationally generated GR by the model is not just dependent on the relationships between drilling parameters and GR logs, this model is also generalizable and capable of being deployed into the application with only retraining on the offset wells and no change in the model structure or complexity. For this paper, the drilling of the horizontal section will not be discussed as this was done as a separate body of work.

Historically collected data from the US Land Permian Basin wells is the primary dataset for this work. Results from the experiments based on the data collected from five different wells have been presented. Leave-one-out validation for each of the wells was performed. In the leave-one-out validation process, four of the wells represent the set of offset wells and the remaining one becomes the subject well. The same process is repeated for each of the wells as they are in turn defined as a subject well. Results show that the framework can infer and generate logs such as GR logs in real time.

Introduction

GR is a primary LWD measurement. The GR logs help drillers as well as geoscientists to infer the distribution of the rock formations. The process of collecting such GR logs requires a tool to be deployed within the bottomhole assembly (BHA). The tool uses GR particles ejected from different elements to estimate the GR logs. These tools are not only prone to failures, but also their deployment increases the cost of drilling. This paper presents a methodology to computationally generate GR logs for the subject well based on the collected real-time information. This method could lower the cost-of-service delivery for any LWD measurement along with the added robustness against tool failures and will facilitate data-driven decision making.

In previous studies discussed in the subsequent section, researchers used statistical algorithms and ML methods such as convolutional neural networks (CNN) and long short-term memory networks (LSTM) to predict some wireline logs using drilling parameters and other available wireline logs. However, this study presents a PIML approach that combines ML algorithms and physics-inspired models to computationally generate GR logs using drilling parameters. It should be noted that this research assumes a one-to-one relationship between GR and rock formations.

The designed PIML approach learns the relationships between the drilling parameters and the formation evaluation logs using the data from the offset wells. The explicit decomposition of the model into multiple stages compels the PIML approach to learn this relationship so the distribution of the rock formation is taken into consideration. This decomposition of the model ensures that the computationally generated log is consistent with the rock formations. This approach is generalizable because the final computationally generated GR is dependent on the drilling parameters and extracted intermediate rock formation information. Thus, it can be claimed that this approach has the capability of being deployed into the application with just retraining on the offset wells and no change in the model structure or complexity.

The organization of the paper is as follows. First, the related work is described, followed by the formulation of the problem. Then the details of the designed framework are presented, followed by a presentation of the experimental setup and the results. Finally, the paper is concluded with a summary and discussion of future work.

Related Work

Recently with the development of data-driven algorithms, there have been a lot of applications designed to use such techniques in the oil and gas industry. The applications range from analyzing the data for inferring certain properties to using the data to forecast properties. Generating wireline logs based on recorded data is one of such highly explored areas of research. Different techniques have been explored for synthetic log generation. Each of these techniques differs based on the methodology used, the computationally generated log as well as the features used for designing the data-driven technique.

In Yu et al. (2021), the authors worked on generating compressional sonic travel time (DTC) and shear sonic travel time logs (DTS) using deep learning approaches such as artificial neural networks (ANN) and recurrent neural networks (RNN) based on other available wireline logs. The authors demonstrate the criticality of different preprocessing techniques including focused data cleaning and clustering to improve the quality of data. In Chaikine and Gates (2020), the authors focus on generating DTS logs using convolutional RNN. An inception-based CNN is presented in Kanfar et al. (2020) that combines a temporal convolutional network (TCN) as the deep learning model to generate density, porosity, and sonic logs using the drilling parameters along with the mechanical specific energy (MSE) as the inputs. Alzate et al. (2014) provides a way to generate the transit-time curves for primary or compressional waves (DTP) and secondary or shear waves, based on the measurements of GR, neutron-porosity, and density logs using an ANN with multiple-input single-output (MISO) autoregressive moving average with exogenous input (ARMAX). Gowida et al. (2020) presents a way to generate the formation bulk density based on the recorded drilling parameters using an ANN combined with an adaptive neuro-fuzzy inference system (ANFIS). In Zhong et al. (2020), the authors devise a way to generate synthetic density logs using XGBoost as the model. The drilling parameters along with LWD logs were used as the input to the model.

Ibrahim and Elkatatny (2022), Osarogiagbon et al. (2020), and Chen and Zhang (2020) present different data-driven ways to generate GR logs. The difference between these methods is in the input features and the specifications of the algorithms used. Osarogiagbon et al. (2020) utilizes drilling parameters obtained from mud logging and measurement while drilling (MWD) for real-time prediction of GR logs that is used as a lithology identifier. In this work, techniques such as RNN, long short-term memory recurrent neural network (LSTM-RNN), TCN, gated recurrent unit (GRU) network, nonlinear autoregressive network with exogenous inputs (NARX), and simple artificial neural network (ANN) were tested. On the other hand, Ibrahim and Elkatatny (2022) use support vector machines (SVM) and random forest (RF) based models to generate the GR logs using the drilling parameters. Chen and Zhang (2020) have designed an ensemble LSTM model based network to generate GR logs using different drilling parameters along with certain other LWD and MWD logs.

The novelty of the work presented in this paper is that formation evaluation logs can be computationally generated in real time using only the drilling parameters, and no other LWD or MWD logs. The framework designed also integrates the physics-based model to provide better generalizability and wider scale applicability. The structure of the framework is generic to be deployed after retraining without any hyperparameter tuning, which is not necessarily the case in the other works.

Problem Setup

If multiple wells are being drilled within a specific region, the drilling cost could be reduced if we have a method to computationally generate formation evaluation logs. In this specific scenario, the computationally generated substitute can be used in place of the formation evaluation logs measured using traditional tools. The digital logs would be computationally generated using the data collected from the other nearby wells, i.e., offset wells. These offset wells are assumed to be analogous to the subject well in terms of the log readings, i.e., the distribution of the formations within the surface can be considered as similar as is the case for nearby wells. For proving the concept, this work will focus on GR logs as the primary formation evaluation measurement. GR logs were prevalent measurements in the data collected.

Fig. 1 shows the formulation of the problem in a graphical representation. In order to generate GR logs (GR_s) for the subject well (denoted as W_s), this approach takes input from the set of offset wells (W_q) and the subject well (W_s), (1) drilling parameters for offset wells (X_O), (2) GR logs for offset wells (GR_O), and (3) subject well drilling parameters (X_S).

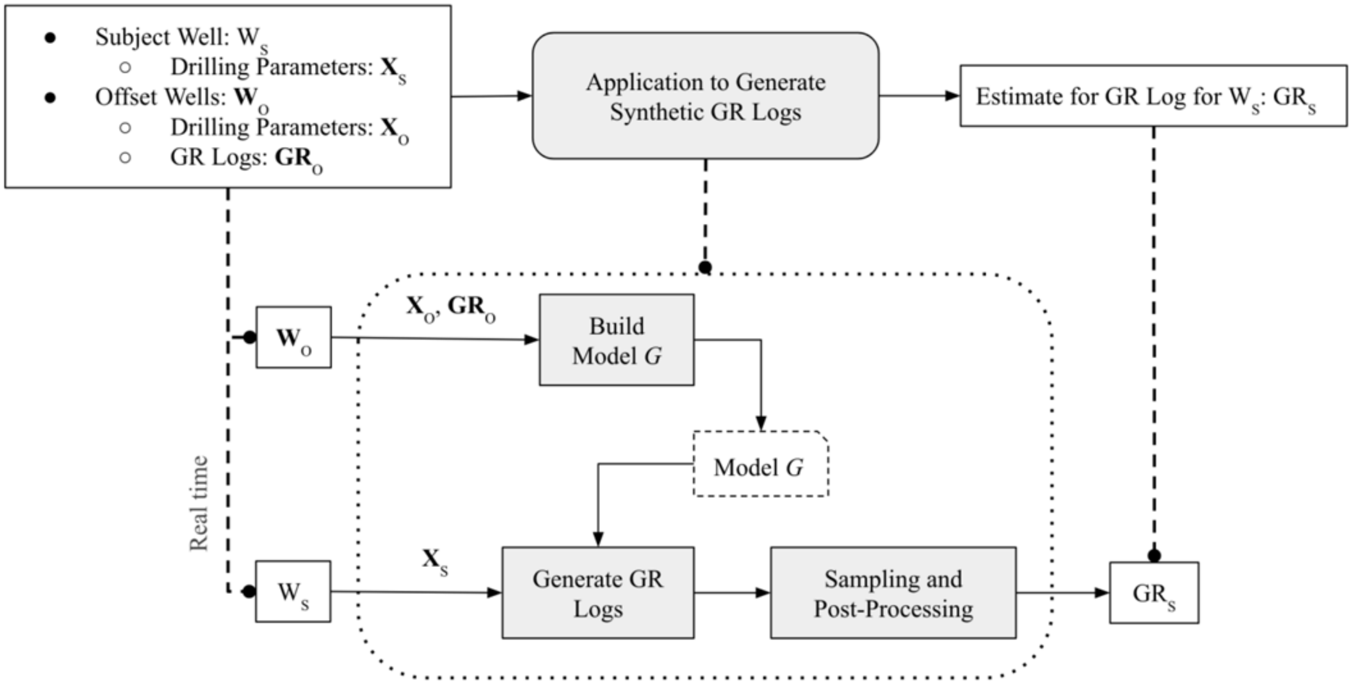

Figure 1—Flowchart representing the operation process.

The work will focus on building the model G that would be used for generating GR logs (GR_s) for the subject well (W_S) as the drilling parameters (X_s) are recorded in real time. The performance of Model G heavily depends on the data from the offset well. It is thus crucial to select offset wells that are analogous to the subject well.

Challenges

From the literature review, it was established that there are methods such as the one presented in Osarogiagbon et al. (2020) that use other wireline logs like high-resolution acoustic logs, sonic logs, and density for generating GR logs. These wireline logs use expensive sensors and are not immediately available in real time as we are drilling. Thus, the model required for generating the GR measurements while drilling would need to use only the real-time available surface drilling parameters. The relationship between the surface drilling parameters and GR logs is extremely complicated because GR is closely related to formations. It would be naive to believe otherwise. Thus, it would require a complex model to gauge the existing relationships.

The relationship between GR and drilling measurements changes with different formations and different regions. The final model parameters for different sets of offset wells would vary based on the distribution of the data from the offset wells. To make the solution generalizable, a single architecture needs to be retrained based on the offset wells specified by the users, making the modeling efforts more challenging.

Our Approach

It is established from the context of the problem that the relationship between rock formations and GR logs can be exploited to generate GR logs. One way of solving this problem would be directly learning the relationship between drilling measurements and the GR logs obtained from the offset wells and then using the drilling measurements from the subject well to generate the corresponding GR logs. The task of learning the relationship between drilling measurements and the formations using the data is implicit in such a solution. However, one of the shortcomings of this approach is that the structure and complexity of the model needs to be changed and adapted to reflect the different relationships varying with different sets of offset wells. As a result, such an approach cannot be used directly in a general application that requires retraining the model with changing sets for the offset wells.

Therefore, we adopt a solution approach where the model is broken down into multiple stages so as to exploit the relationships between rock formations and GR logs. The former stages focus on extracting the information to represent the underlying rock formations and the latter ones use the extracted information in conjunction with the observed surface drilling parameters to generate the GR logs. This explicit decomposition of the model allows the model to adhere to the physical properties that define the GR logs. The model can be visualized as shown in Fig. 2 which shows the former stages as submodel G_1 and the latter stages as submodel G_2.

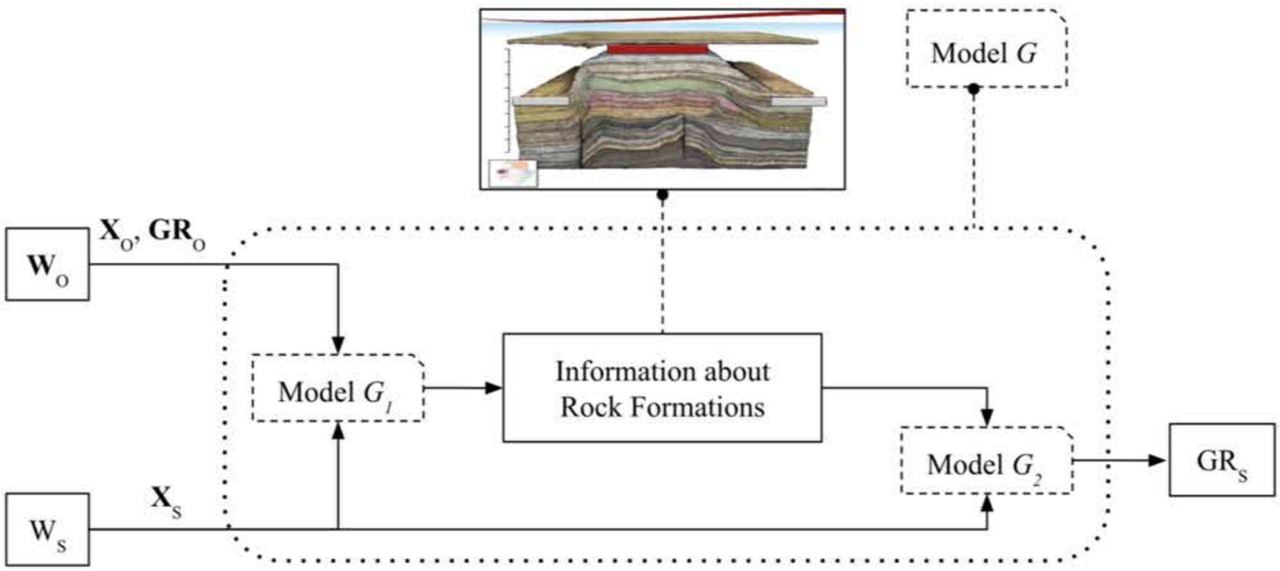

Figure 2—Workflow for multi-staged decomposed framework. Model G_1 (representing the former stages) extracts the latent information for the formation using the drilling parameters, and Model G_2 (representing the latter stages) uses the extracted rock formation information obtained through Model along with the drilling measurements to generate the GR logs.

The explicit decomposition of the model G into multiple stages addresses the challenge of generating GR logs for various sets of offset wells. Firstly, the model is compelled to learn the relationship between drilling parameters and rock formations, and between the rock formations and GR logs, separately. This makes it easier for the model to generate GR logs consistent with the rock formations of the subject well being drilled. Secondly, the model is generalizable because the final GR logs computationally generated by the model are not just dependent on the relationships between surface drilling measurements and GR logs learned by the model.

The Physics Informed Machine Learning (PIML) Framework

The physical rules that represent a real-world scenario are typically based on the underlying physics that guide the particular phenomenon. Accurate modeling of the phenomenon involves designing appropriate mathematical equations to represent these physical rules. Typically, such equations derived with the help of physics require the estimation of multiple constants and can be computationally expensive and time-consuming.

The alternate method to estimating the constants is to learn a data-driven model based on recorded data that represents the working of the phenomenon. The performance of such data-driven models relies heavily on the data used for training the model. These models are not always generalizable to data points outside the distribution of the training datasets. Moreover, the predictions from these models might not abide by all the physics laws underlying the phenomenon.

Physics models can be combined with data-driven approaches to mitigate the extensive process of estimating the constants while simultaneously adhering to the physics laws, to get the best of both worlds. As shown in Karniadakis et al. (2021), this combining physics models with data-driven approaches leads to PIML. Moreover, the work done in Jeong et al. (2020) and Sheth et al. (2022) establishes that the PIML approaches show significant improvement in modeling physical phenomena compared to modeling them with physics or data-driven approaches separately.

The overall approach consists of different modules that are discussed in the following sections. Fig. 3 and Fig. 4 visualize the training and inference process for the designed approach. In the training workflow, 'Calculate MSE' block calculates the *MSE* value. 'Build Physics Model' block computes the physics model estimate for GR represented by GR_{PHY}. Details of the physics model are not divulged because of intellectual property considerations. 'Extract Formation Information' block computes the automated formation information represented by F_O. 'Build Empirical Model' block computes the empirical model estimate for GR represented by GR_{emp}. 'Train Formation Classification Model' block trains the classification model represented by G_{FCL}. 'Train Formation-Based Regression Models' block trains the different regression models represented by G_{FR}. For inference block represented in Fig. 4., the learned regression models G_{FR}, along with formation classification model G_{FCL} are used to computationally generate GR logs of the subject well.

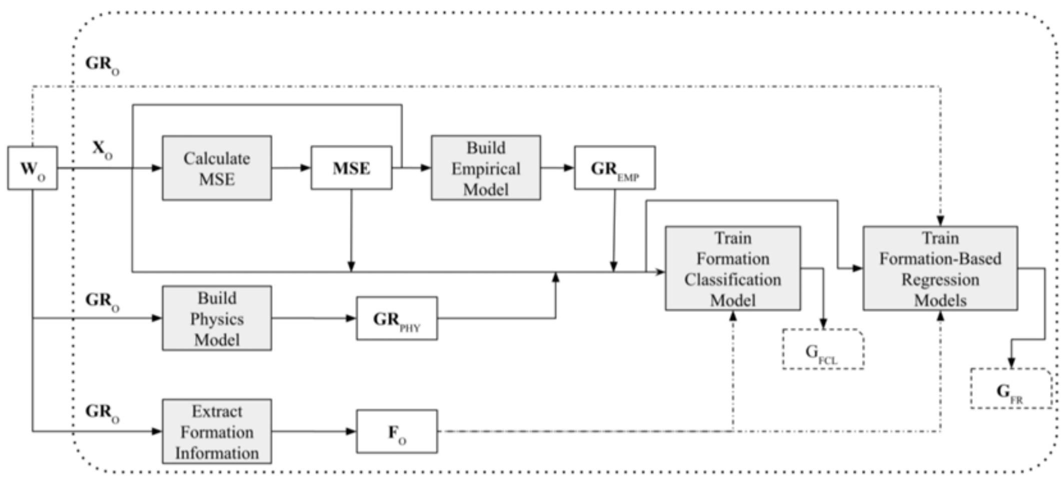

Figure 3—Flowchart representing the PIML framework for the training process.

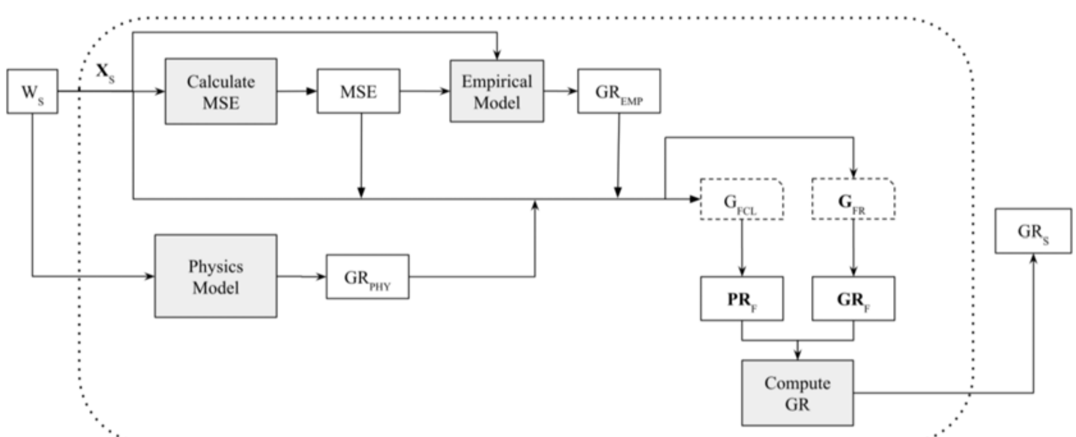

Figure 4—Flowchart representing the PIML framework for the inference process.

Mechanical Specific Energy. *MSE* is the energy required to remove a unit volume of rock. Surface weight on bit (*SWOB*) is responsible for indenting the rock, and torque (*TQX*) is responsible for breaking it. These two forces act independently. The axial work done is determined by *SWOB* and the axial distance per time is determined by the Rate of Penetration (*ROP*). The rotational work done is calculated using *TQX* and revolutions per minute (*RPM*). This total work done is then divided by the volume of the rock to calculate *MSE*. Thus, it can be controlled using drilling parameters - *TQX, RPM, ROP,* and *SWOB*. It follows a physical formula as presented in Teale (1965) and Dupriest et al. (2022):

$$MSE = \left(\frac{120 \cdot TQX \cdot RPM}{ROP} + \frac{SWOB}{\pi} \cdot \frac{4}{D^2}\right),$$

where D is the diameter of the drill bit.

Hence, it does not 'learn' anything from the data, instead just 'uses' the given data to produce the *MSE* value. *MSE* calculation is free from learning anything from the data, can adapt to the input data easily and does not need any retraining. *MSE* allows for standardizing the reactions to the drilling activity performed through a certain rock formation. These reactions are recorded in terms of the drilling parameters. *MSE* is calculated for all the wells and using it as a feature in the following modules aids the task of understanding the relationship between rock formations and drilling parameters. As described in the related work section, different works have also used *MSE* as an additional feature for predicting different wireline logs.

Automated Formation Extraction. Formation tops are an integral part of the decision making in any drilling process. GR logs provide a signature to the rock formations. Analyzing and understanding the patterns in these logs can help identify the changes in formations and hence the formation tops.

Time series clustering is a method that allows us to group similarly shaped time series elements. This method of clustering is analogous to the traditional method of analyzing the patterns of wireline logs. In this work, an automated way of finding the formation information has been designed based on the similarity between these methods. The designed method allows the users to control the granularity for the analysis. A single GR log is broken down into multiple segments based on the specifications. Clustering is applied to group these parts. Each of the clusters is uniquely represented by the central element. These parts of GR logs are assigned to an individual group using these centers. A set of labels for each point in depth is obtained using these alignments. Majority-based voting is carried out on overlapping segments to obtain a final label for a specific point.

Further, a methodology to include the prior information based on the rock formations of the region of the wells is developed. This is useful to increase the robustness of the approach. Probabilistic labels are obtained to signify the uncertainty involved in the algorithmic predictions and specified prior information. Analyzing the labels obtained for each point would provide enough information to mark the formation tops. This provides a way to use the real-time information coming in the form of the drilling measurements and use them in conjunction with the extracted rock formation information.

Physics Model for GR Log Generation. The trajectory-based average of the GR values across all the offset wells could serve as a physical estimate for the GR logs of the subject well. The average representation could be indexed based on both measured depth (*MD*) and true vertical depth (*TVD*). This average could be precomputed and stored based on the data from the offset wells. This average representation can be mapped to different observed data points of the subject well in real time to get a physical estimate of the GR log. The mappings are unique as it is dependent both on the *TVD* and *MD* value. By considering the trajectories of the wells, this model captures the similarities in the rock formations that different wells are traversing through in turn learning the relationship between the rock formations and the GR logs. In this way, the physics model could be considered as a type of three-dimensional mapping where the subject well could be sampled based on the upcoming well plan. The advantage of this method is that any simple model could suffice.

Empirical Model for GR Log Generation. The empirical model uses the drilling parameters to get the empirical estimate of GR logs. This model computes the estimate by performing computations based on the statistics of the data. Particularly, it considers the distance between different data points based on the feature space defined by drilling parameters to determine the empirical estimate of the GR logs. This estimate is dependent on the weights assigned based on the distance in the depth for different points. Even though the empirical model is developed using data, it is more of a statistical model and not an ML model and hence, does not 'learn' anything from the data, instead just 'uses' the given data to produce the estimates. The physics model differs from the empirical model in the features used for estimating the GR logs. The former uses the trajectory information while the latter uses the observed drilling parameters. Fig. 5 shows the schematic representation of the empirical model.

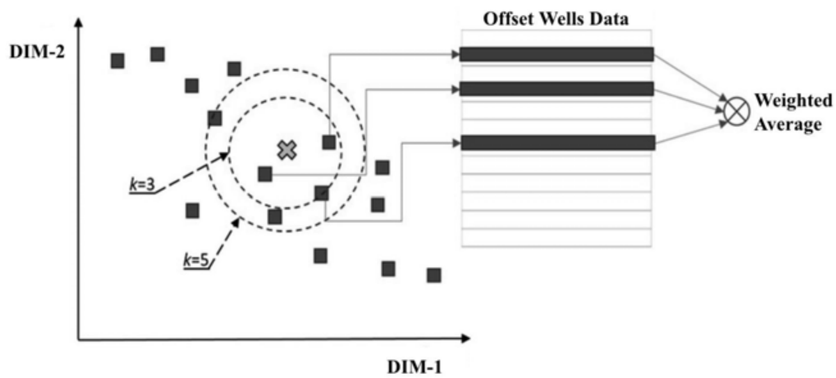

Figure 5—Schematic representation for the process of generating the empirical physics model using the data from offset wells. The plot on the left shows the 'reduced' 2-dimensional feature space represented by Dimension-1 (DIM-1) and Dimension-2 (DIM-2).

Formation Classification and Formation-Based Regression. Given the drilling parameters, calculated *MSE,* formation information extracted from the automated formation extraction model, physical estimate of the GR logs from the physics model, and empirical estimate of the GR logs from the empirical model, a classification model was trained for predicting the formation classes for each depth. The number of formation classes (N) is determined based on the number of formations observed from the formation extraction part. At every depth, the classification model gives us the predicted probabilities for the belongingness to each formation class. For a data point, these probabilities from the classification model can be denoted as $PR_{F_i}, i \in [1, N]$. The formation class with the highest probability is denoted by c'. A regression model was trained for each of the corresponding formation classes to extract the formation-based GR logs. For a data point, the estimate of the GR value for $Class_i$ is $GR_{F_i}, i \in [1, N]$.

The final predicted GR logs (GR_s) can be calculated using two ways namely unweighted estimate and weighted estimate. The unweighted GR log estimation is GR_{Fc*} (i.e.) the GR log corresponding to the formation class with maximum probability (i.e.) c'. The weighted GR log estimation is $\sum_{i=1}^{N} PR_{F_i} \cdot GR_{F_i}$, where PR_{F_i} is the probability of $Class_i$ obtained from the classification model, and GR_{F_i} is the GR estimated from the corresponding regression models.

Experiments

Dataset. Data collected while drilling can be indexed based on time and depth. The time-indexed data represents all the data throughout the time while the well is being drilled as well as where no drilling activity is conducted. The depth-indexed data is recorded at regular intervals of depth and has no data corresponding to the stagnant part while drilling. In this work for generating GR logs, the GR logs change with depth, and hence, depth-indexed data is used.

For the results discussed in this paper, data from the US Land Permian Basin wells is used as the primary dataset since the subject wells tested in real time are from this location. Clustering of the wells based on their geographical location is performed and the wells are recursively divided into subgroups. The final subgroup of five wells was obtained after further analysis of the similarity of the nature of the GR logs. The data collected from these five different wells have been used to validate the designed approach. This is to mimic the selection of the offset wells in the real-time workflow expected.

The data corresponding to each well has multiple fields, both surface drilling measurements, and non-surface drilling measurements. The data represents the surface features indexed on measured depth and is recorded at appropriate intervals. Each data point is denoted by depth d, feature vector X_d, and GR value GR_d. Using this representation, a single data point in the dataset corresponding to depth d can be represented as (d, X_d, GR_d). If there are N-data points in the data corresponding to a well, the entire well's data can be represented by $\left(d_i, X_{d_i}, GR_{d_i}\right), \quad i \in [1, N]$

The relevant data fields corresponding only to the surface drilling measurements have been selected with the help of drilling experts. The shortlisted surface drilling measurements in the feature vector *MD, SWOB,* hook load (*HKLD*), *RPM, ROP, TQX,* standpipe pressure (*SPPA*).

Data Preprocessing. A set of quality checks were performed on the data to ensure consistency across all the wells. The outliers in the GR logs were removed based on the typical range for GR logs. [0-200] API was used as the range for this body of work. It is also important to make sure the units of all the recorded fields are consistent. Moreover, the unwanted noise captured in the data due to different reasons in the field needs to be removed. Along with the noise, missing values could occur for some features in the data. Data imputation techniques were deployed to tackle this problem.

Imputation is a technique where the missing data values are filled by statistical methods to maintain a consistent distribution with the overall data. The imputation of GR logs should not be affected by any of the features as it is an independent variable. Thus, the data imputation of features and GR logs was performed separately. Some of the imputation methods used for feature imputation include backward fill, forward fill, interpolation, exponential weighted moving average (EWMA), random forest (RF) method-based imputation, and multiple imputations by chained equations (MICE). The methods used for GR log imputation are backward fill, forward fill, and interpolation.

Feature imputation techniques are applied to each feature individually. To determine the best imputation method for a particular feature, the feature is imputed using all the different methods mentioned above, and the best method is selected based on an error metric. For each of the methods and the resulting imputed values, the root mean squared error (*RMSE*) is calculated over the known data that were set to be missing as follows:

$$RMSE(F, M) = \sqrt{\left(\frac{\sum_{i=1}^{N}\left(F_i - F_i'\right)^2}{N}\right)}$$

where method M is the imputation method in consideration, data attribute F is the attribute being imputed. N is the number of known-data points. F_i represents the data point's known value. F'_i represents the imputed value for the same data point. The data attribute is imputed with the selected method, and the missing values are filled. The above process is repeated over all the attributes for all the wells. The method yielding the lowest RMSE value is selected.

Training and Testing. Different models were trained using the data from the selected offset wells. The modeling work has been accomplished by using open-source ML libraries, and training was done in our local setup. Leave-one-out validation has been performed for each of the wells, where four of them become the offset wells and the remaining one becomes the subject well.

Results and Analysis

In this section different plots showing the results from the leave-one-out validation are presented. Fig. 6 shows the results from the experimental scenario where Well: 1 was the subject well and the others i.e. Well: 2, Well: 3, Well: 4, Well: 5 become the offset wells. Detailed plots for all the wells separated out are summarized in Appendix: Fig. 1. The plots of the computationally generated GR logs (weighted and unweighted) from the defined PIML framework on the specified set of five wells are shown in Fig. 7 and Fig. 8 respectively. As the regression models used in the PIML framework are Bayesian in nature, the mean value of the different predictions from the model is the final computationally generated GR log. This is represented by the blue line in the plots. The confidence interval represented by the gray shaded region in the plots is covering two standard deviations for the computationally generated GR logs. Moreover, the extracted confidence interval captures the uncertainty in the framework that could benefit the applications that can be built upon this framework.

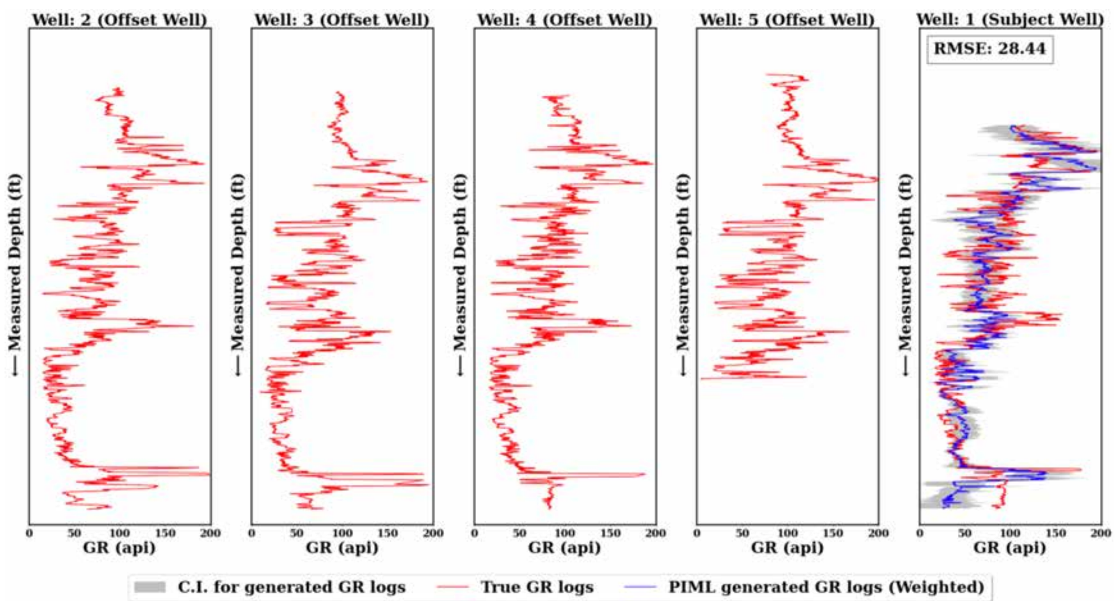

Figure 6—Results from the experimental scenario where Well: 1 was the subject well and the others i.e. Well: 2, Well: 3, Well: 4, Well: 5 become the offset wells. The red solid line represents the ground truth recorded GR logs from the wells. The blue solid line represents the computationally generated GR logs (weighted). The gray shaded region represents the confidence interval for the computationally generated GR logs.

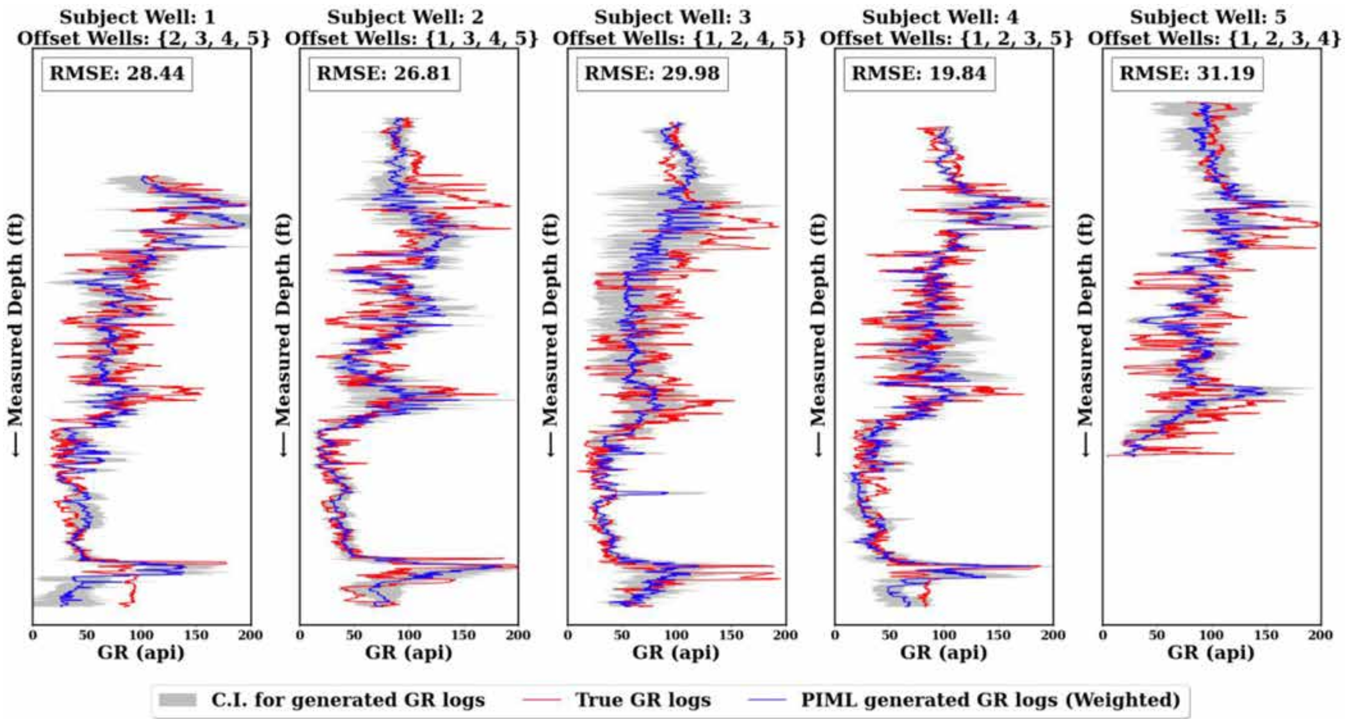

Figure 7—Leave-one-out validation results for different wells. The red solid line represents the ground truth recorded GR logs from the wells. The blue solid line represents the computationally generated GR logs (weighted). The gray shaded region represents the confidence interval for the computationally generated GR logs.

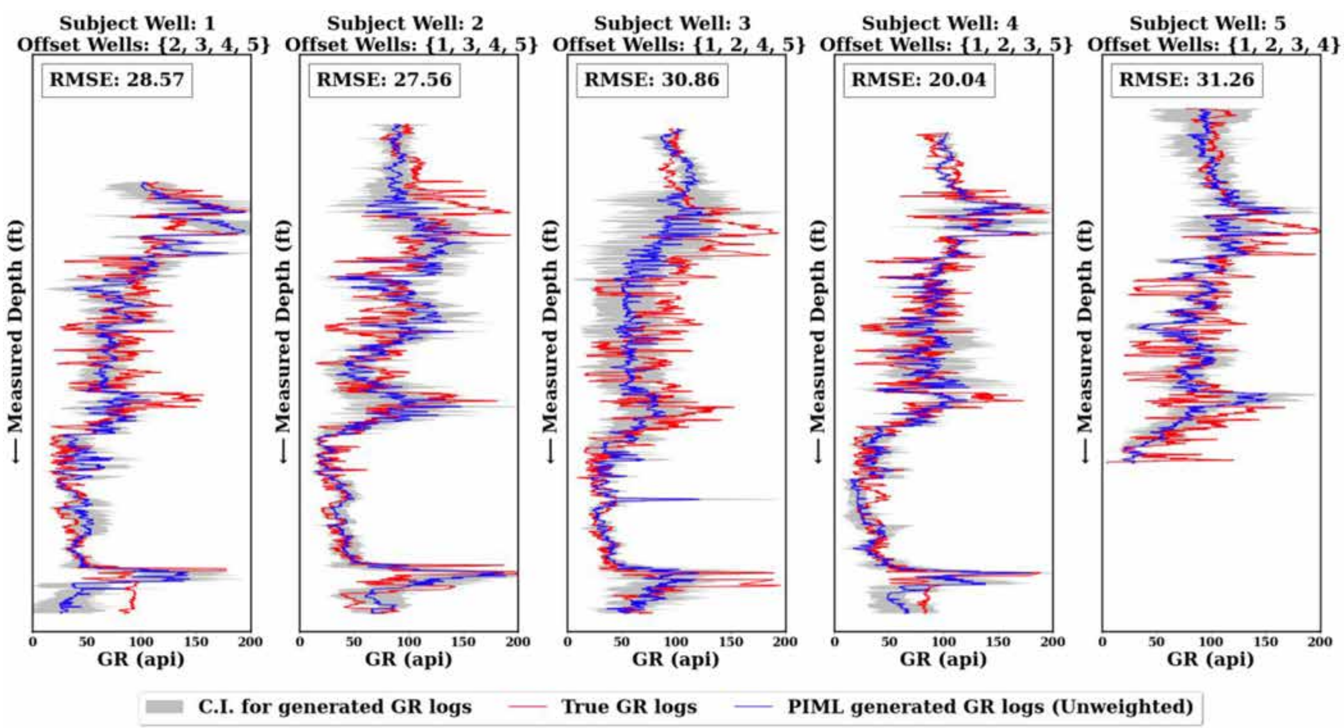

Figure 8—Leave-one-out validation results for different wells. The red solid line represents the ground truth recorded GR logs from the well. The blue solid line represents the computationally generated GR (unweighted) logs. The gray shaded region represents the confidence interval for the computationally generated GR logs.

Based on the qualitative analysis of the results, it can be noted that the computationally generated GR logs using the developed framework closely follow the trends in the ground truth recorded GR logs. As the computationally generated GR logs approximate each crest and trough, they could be helpful to the drillers and geoscientists to do a preliminary analysis of different aspects of drilling such as the drill bit position, the rock formation type, and eventually could serve as a guide for directional drilling. RMSE values (calculated based on the mean representation shown in the plots) along with the QQ-plots (shown in Fig. 11) have been used for performing the qualitative analysis of the results. QQ-plot aids in comparing the distributions of the recorded logs and the computationally generated logs. The dotted line in the QQ-plot that occurs as a 45° straight line ($x = y$) represents the ideal scenario where recorded logs exactly align with the computationally generated logs. The closer the computationally generated GR logs are to the ground truth recorded logs, the closer the line for this computationally generated GR log would be to this ideal line. Based on these plots from Fig. 11, it can be noted that the distributions of computationally generated GR logs (weighted and unweighted) highly resemble the distributions of the ground truth recorded logs. Although the plots represented in Fig. 6 and Fig. 8 show that the computationally generated GR logs (weighted and unweighted) are similar; it has been established from further detailed experiments that the uncertainty from the unweighted generation is much higher than the weighted generation. This is attributed to the fact that the weighted generation considers the probability of the datapoint belonging to all the rock-formation classes and thus, has a better chance of mitigating the errors coming from the classification block.

Fig. 9 and Fig. 10 show the intermediate GR logs computationally generated from the physics model and empirical model, respectively. They follow the same notation as described above except that the confidence interval of these computationally generated GR logs is not available.

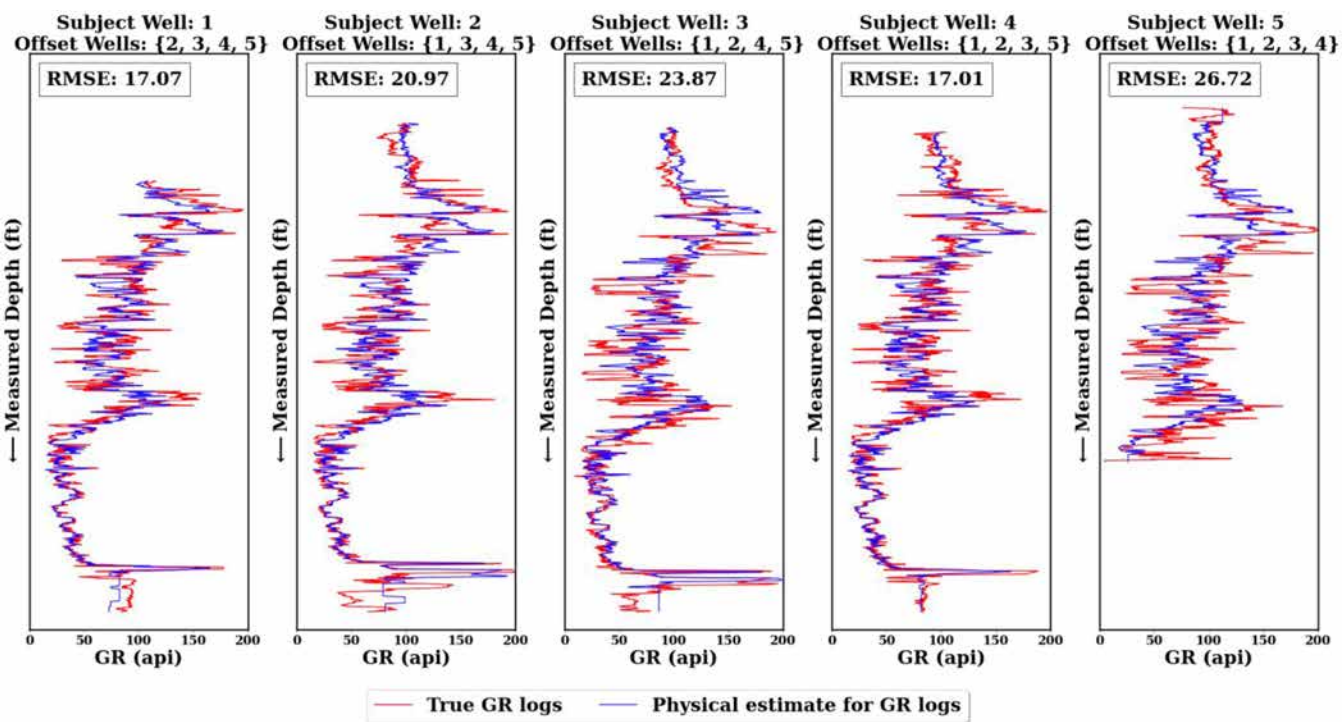

Figure 9—Leave-one-out validation results for different wells. The red solid line represents the ground truth recorded GR logs from the well. The blue solid line represents the computationally generated GR logs from the physics model.

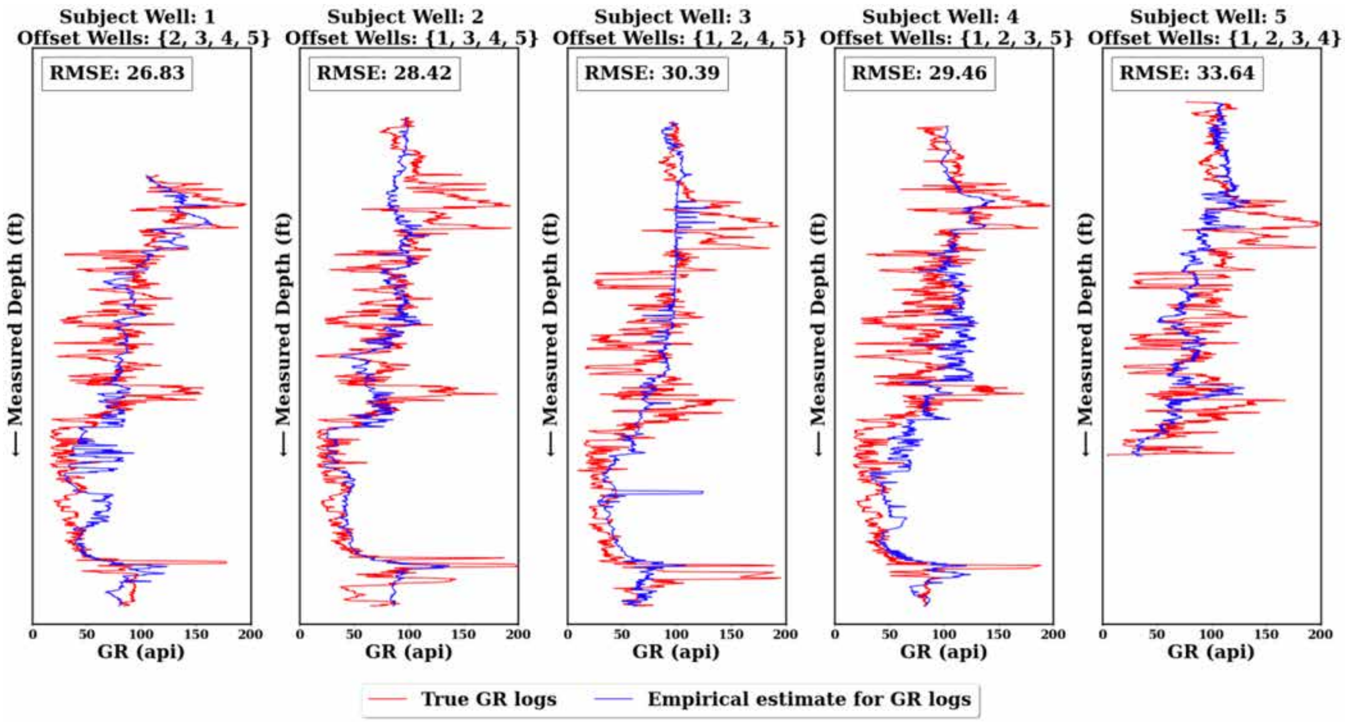

Figure 10—Leave-one-out validation results for different wells. The red solid line represents the ground truth recorded GR logs from the well. The blue solid line represents the computationally generated GR logs from the empirical model.

From the plots (Fig. 9 and Fig. 10) of the computationally generated GR logs from the physics model and the empirical model, it can be seen that the physics model succeeds well in capturing the patterns of the ground truth recorded GR logs. This trend can be attributed to the high correlation between the GR logs of the specified set of offset wells. However, it can fail to capture the exact trends in the cases of low correlation between the offset wells. Further detailed experiments performed on a few other different sets of offset wells with lower correlation showed that the performance of the physics model is highly dependent on the chosen set of offset wells. It was also established that the faults occurring through the different rock-type formations could hamper the performance of the physics model. Also, the model in this work is designed for low-angle paths. The performance of the physics model could deteriorate with the model being exposed to high-angle noisy well path trajectories. Hence, the hybrid approach of physics and ML is more valuable in terms of generalizability and applicability. The plots from Fig. 10 show that the distribution of the computationally generated GR logs differs significantly from the ground truth recorded GR logs. Thus, the empirical model is not sufficient as a standalone model for this application. However, it is helpful in capturing the relationship between the observed drilling parameters. Supplementing the physics model with the empirical model would allow for combining the information from the drilling parameters from the former with the trajectory information from the latter. This provides the model with grounds for learning from all the available information, thereby improving the performance of the overall framework.

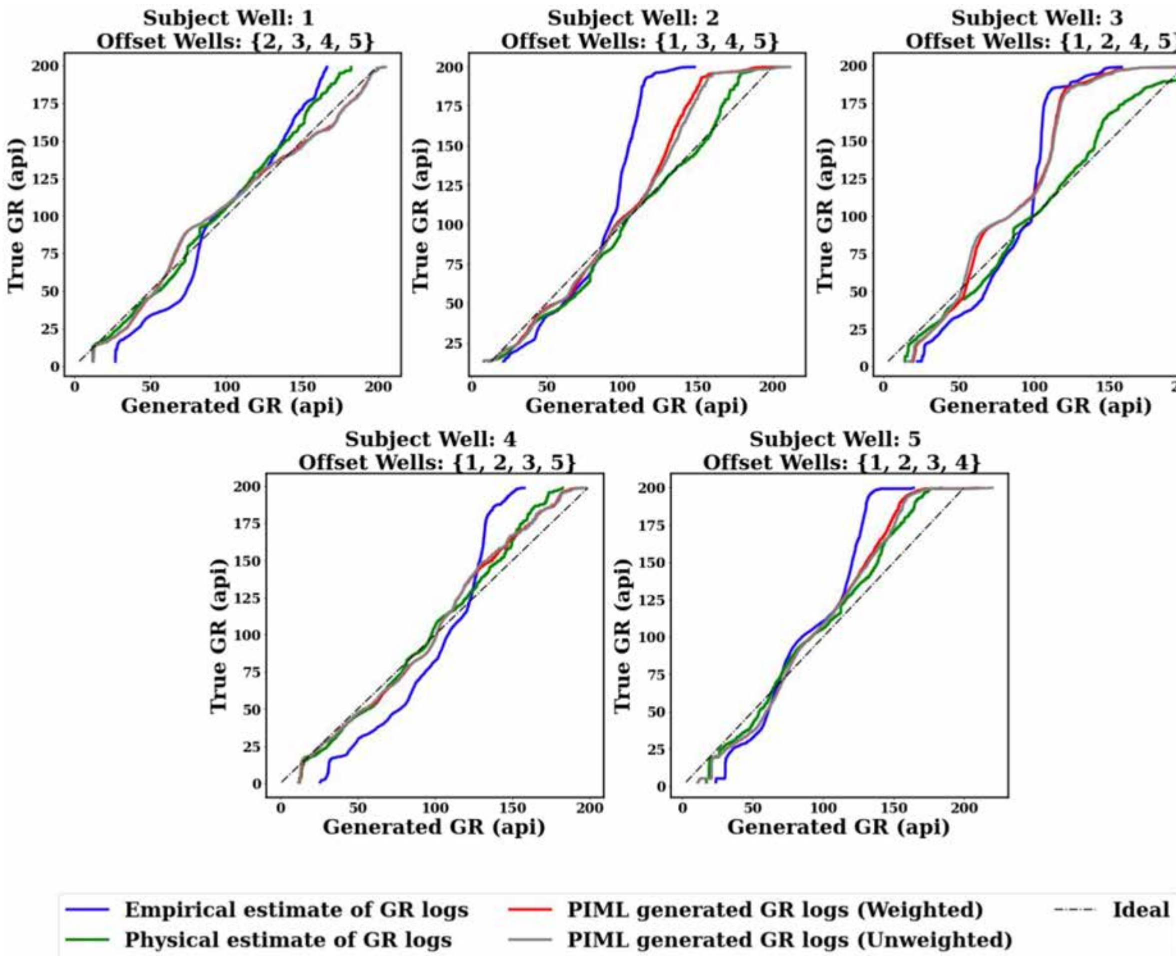

Figure 11—QQ-plots comparing the performance of different models.

Conclusion

This body of work tests the idea that a digital log could be computationally generated if the learned relationships between drilling parameters data and formation evaluation logs can be identified and applied to drilling parameters data in real time. The cost-of-service delivery is increasing across the industry for service companies, and the presented novel approach to this problem will allow reduced delivery costs using the computationally generated log.

The problem was addressed by breaking down the idea into a simple substitute for a physics model for the logs learned from the offset data. This was then paired with the real-time application of the data-driven generator for the logs. While it is possible to define a completely standalone physics-based model that may be able to predict the shape of a log, the hybrid approach holds the key to generating more accurate logs in cases where the physics model is not robust enough for standalone usage. An example of this is where the offsets are not highly correlated. In our work so far, we have addressed low angle application as with higher angle well paths, additional requirements need to be considered and applied and fall beyond the scope of what is presented here. Recall that the hybrid framework designed provides better generalizability and wider scale applicability. The structure of the framework is generic to be deployed after retraining without any hyper-parameter tuning, which is not necessarily the case in the other works.

Multiple wells from various locations were used and the respective GR logs have been computationally generated from the learned relationships with real-time drilling parameters data. The results were validated with recorded mode logs. Results thus far have proven to be successful in the development of such an application. As stated previously, formation evaluation logs can be computationally generated in real time using only the drilling parameters, and not any other LWD or MWD logs as shown by the results. Although our work described here focuses on GR logs without any loss of generality, this could be applied to any of the additional formation evaluation logs.

The method presented here with some minor modifications could be deployed and used in the field. It is possible to replace the physics model described with any one of the industry's leading applications. Data-driven methods may also be tweaked to allow consideration of local knowledge of formations. An example of this may provide a unique model for handling special formation cases such as salt.

As the generated GR logs could approximate each crest and trough, it could help the drillers and geoscientists to do a preliminary analysis of where the drill bit is, what is the formation type and eventually guide for directional drilling. In some cases, applying the methods described may lead to a complete replacement for LWD tools or at least a proxy accurate enough to be used for real-time decision making at the well site.

References

Alzate, G.A., Arbeláez-Londoño, M., Agudelo, A.N. et al: "Generating Synthetic Well Logs by Artificial Neural Networks (ANN) Using MISO-ARMAX Model in Cupiagua Field," SPE Latin America and Caribbean Petroleum Engineering Conference. OnePetro, (May 2014). SPE-169388-MS https://doi.org/10.2118/169388-MS

Chaikine, I.A. and Gates, I.D.: "A New Machine Learning Procedure to Generate Highly Accurate Synthetic Shear Sonic Logs in Unconventional Reservoirs," SPE Annual Technical Conference and Exhibition. OnePetro, (October 2020) https://doi.org/10.2118/201453-MS

Chen, Y. and Zhang, D.: "Well Log Generation Via Ensemble Long Short-Term Memory (EnLSTM) Network," *Geophysical Research Letters* **47**, Issue 23 (November 2020) e2020GL087685 https://doi.org/10.1029/2020GL087685

Dupriest, F., Pastusek, P., Lai, S. et al: "Standardization of Mechanical Specific Energy Equations and Nomenclature," SPE/IADC Drilling Conference and Exhibition, OnePetro, (March 2022) SPE-208777-MS https://doi.org/10.2118/208777-MS

The global logging while drilling (LWD) market size was USD 2.82 billion in 2018 and is projected to reach USD 5.05 Billion by 2026, exhibiting a CAGR of 7.6% during the forecast period. 2020. *Fortune Business Insights*, https://www.fortunebusinessinsights.com/industry-reports/logging-while-drilling-lwd-market-101919 (accessed 20 October 2021)

Gowida, A., Salaheldin, E., Abdulrahim, A. et al: "Synthetic Well-Log Generation: New Approach to Predict Formation Bulk Density While Drilling Using Neural Networks and Fuzzy Logic," International Petroleum Technology Conference. OnePetro, (January 2020) https://doi.org/10.2523/IPTC-19787-MS

Ibrahim, A.F., and Elkatatny, S.: "Real-Time GR Logs Estimation While Drilling Using Surface Drilling Data; AI Application," *Arabian Journal for Science and Engineering* **47**, No. 9 (2022) 11187–11196 https://doi.org/10.1007/s13369-021-05854-7

Jeong, C., Yu, Y., Mansour, D., Vesslinov, V., and Richard M.: "A Physics Model Embedded Hybrid Deep Neural Network for Drillstring Washout Detection," IADC/SPE International Drilling Conference and Exhibition, OnePetro, (March 2020) SPE-199629-MS https://doi.org/10.2118/199629-MS

Kanfar, R., Shaikh, O., Yousefzadeh, M. et al: "Real-Time Well Log Prediction from Drilling Data Using Deep Learning," International Petroleum Technology Conference, OnePetro, (January 2020) https://doi.org/10.48550/arXiv.2001.10156

Karniadakis, G.E., Kevrekidis L.G., Lu, L. et al: "Physics-Informed Machine Learning," *Nature Reviews Physics* **3**, pp. 422–440 (May 2021) https://doi.org/10.1038/s42254-021-00314-5

Osarogiagbon, A.U., Oloruntobi, O., Khan, F. et al: "Gamma Ray Log Generation from Drilling Parameters Using Deep Learning," *Journal of Petroleum Science and Engineering* **195** (December 2020) 107906 https://doi.org/10.1016/j.petrol.2020.107906

Sheth, P., Roychoudhury I., Chatar C. et al: "A Hybrid Physics-Based and Machine-Learning Approach for Stick/Slip Prediction," IADC/SPE International Drilling Conference and Exhibition, OnePetro, (March 2022) SPE-208760-MS https://doi.org/10.2118/208760-MS

Teale, R.: "The Concept of Specific Energy in Rock Drilling," *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, Vol. **2**, Issue 1, pp. 57–73. Pergamon, (March 1965) https://doi.org/10.1016/0148-9062(65)90022-7

Yu, Y., Xu, C., Misra, S. et al: "Synthetic Sonic Log Generation with Machine Learning: A Contest Summary from Five Methods," *The SPWLA Journal of Formation Evaluation and Reservoir, Description* **62**, No. 04 (2021): 393–406 https://doi.org/10.30632/PJV62N4-2021a4

Zhong, R., Johnson Jr, R., and Chen, Z.: "Generating Pseudo Density Log from Drilling and Logging-While-Drilling Data Using Extreme Gradient Boosting (XGBoost)," *International Journal of Coal Geology*, Vol. **220** (March 2020): 103416 https://doi.org/10.1016/j.coal.2020.103416

Appendix

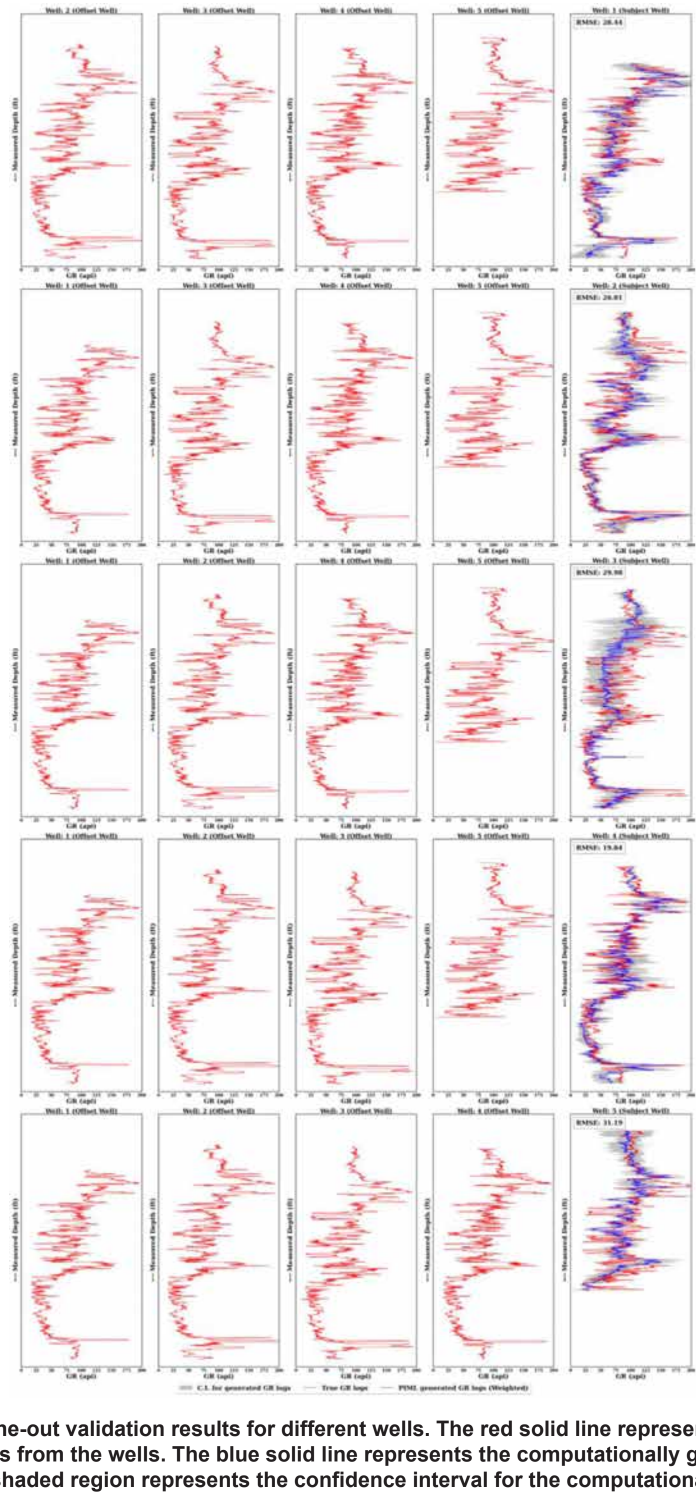

Figure 1—Leave-one-out validation results for different wells. The red solid line represents the ground truth recorded GR logs from the wells. The blue solid line represents the computationally generated GR logs (weighted). The gray shaded region represents the confidence interval for the computationally generated GR logs.