PROCEEDINGS, 46th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 15-17, 2021
SGP-TR-218

Real-Time Model for Thermal Conductivity Prediction in Geothermal Wells Using Surface Drilling Data: A Machine Learning Approach

Cesar Vivas and Saeed Salehi

Well Construction Technology Center (WCTC), The University of Oklahoma, Norman, OK, USA

cesar.vivas@ou.edu

Keywords: Machine Learning, Geothermal, Thermal Conductivity, Geothermal Well, Real-Time Drilling Data, Python, Supervised Learning

ABSTRACT

The thermal conductivity of rocks is a strategic component to evaluate the potential of a geothermal resource. This rock property represents the rate of the amount of heat transferred by conduction through a cross-sectional area, characterizing the rock's ability to transmit heat. Thermal conductivity is measured directly in the laboratory on core samples or empirical correlations from electric logs. Although these methods can provide thermal conductivity information, they are expensive and sometimes difficult to implement. During drilling geothermal wells, surface sensors are collecting drilling parameters in real-time. That data has been mainly used to provide information about drilling conditions, prevent or determine potential risks, and help drilling optimization. This study focuses on surface drilling data in a machine-learning workflow to predict the thermal conductivity at the bit while drilling. In this study, we have used a public data set from Utah FORGE geothermal wells project. In this workflow, the main objective is to predict the thermal conductivity using various surface-drilling data variables, such as the penetration rate (ROP), weight on bit (WOB), torque, flow rate, and others that are usually real-time monitored. We used actual thermal conductivity values directly obtained from samples to train and test the data set. We analyzed four different supervised regression algorithms to link real-time drilling data to thermal conductivity. The algorithms predict the thermal conductivity in the test well with precision above 80%. Although we generated this methodology based on information from the Utah FORGE project, it can potentially be extended to other fields. This study's results are a small step to demonstrate the data analytics capabilities to generate valuable information at a relatively low cost.

1. INTRODUCTION

Geothermal energy benefits from the energy stored in the subsurface in the form of heat. Due to its capability to produce constant energy independently of seasonal factors, geothermal energy is a very attractive renewable energy source. Drilling geothermal wells is a complicated and expensive operation, especially considering the high temperatures, hard and abrasive rock, and corrosive environment (Vivas et al. 2020). These conditions make collecting data from geothermal wells a challenging task. Downhole sensors used in wireline logs or logging while drilling tools are affected by the high temperatures and corrosive conditions.

In drilling operations, surface sensors are placed around the drill rig. These sensors continuously record data 24/7 and transmit that data in real-time to operator offices, where that data can be analyzed immediately (Sletcha et al. 2020). This data is used to analyze drilling parameters, evaluate drilling conditions, look for optimization, and prevent potential problems. With the development of different data analytic solutions, machine learning offers an opening perspective of the data generated continuously in geothermal drilling operations. Diverse studies have focused on the analysis of drilling data to solve problems and optimize processes in geothermal wells. Kiran and Salehi (2020) presented a study where supervised and unsupervised algorithms were used to identify lost circulation zones using drilling data from a geothermal well. Bayan and Zulkarnain (2020) presented a machine learning approach to predict stuck pipe events in geothermal operations using drilling parameters. Yuswandari et al. (2019) proposed a methodology to predict the ROP in a geothermal well using the drilling parameter of offset wells.

The concept of using machine learning algorithms to predict petrophysical properties using drilling data has also been investigated. Gupta et al. (2020) presented a machine learning approach workflow to predict lithology facies in real-time. Kanfar et al. (2020) proposed using drilling parameters to predict petrophysical logs, such as density, porosity, and sonic logs, in real-time using drilling parameters. Chen et al. (2020) presented a methodology for lithology classification using drilling data. Al-AbdulJabbar et al. (2020) presented a machine learning approach analyzing drilling parameters using Artificial Neural Networks to predict porosity.

Identification of the reservoir rocks' thermal conductivity is one of the critical variables of geothermal modeling. Thermal conductivity is a physical property of materials to conduct their molecule's kinetic energy to other molecules in a different energy state. It directly controls the steady-state temperature in the reservoir (Di Sipio et al. 2013). Core samples and drilling cuttings are frequently used for thermal conductivity measurement (Fuchs and Foster 2013). However, collecting rock samples, mostly from coring, is very expensive and challenging.

Additionally, coring intervals are controlled to a limited range, so analyzing thermal conductivity for long intervals is not feasible. Due to these difficulties, thermal conductivity predictions using petrophysical log response data are gaining popularity. Several studies have investigated statistical approaches to generate thermal conductivity. Predictor variables generated by data regressions based on petrophysical log responses were applied to predict thermal conductivity (Brigaud et al. 1990, Hartmann et al. 2005, Fuchs and Foster

2013, Fuchs et al. 2015). Since statistical solutions are applied to predict thermal conductivity, machine learning algorithms' usage to thermal conductivity prediction is reasonable.

This study generates a machine learning approach to predict thermal conductivity by using drilling data. Drilling parameters are recorded from the surface. Surface sensors are not exposed to the same harsh conditions than wireline or logging while drilling tools sensors. The advantage of using drilling parameters is that it does not require any additional infrastructure. The reason is that recording those parameters is a standard procedure in drilling operations.

1.1 FORGE Geothermal Field

The source of the data used to generate this study is from the Utah FORGE project. This is a geothermal project for enhanced geothermal system (EGS) research. All the information processed in this study belongs to the well 58-32. The well 58-32 is located in Milford, UT, 217 miles south of Salt Lake City, as shown in Figure 1 (Utah FORGE, 2018). The well 58-32 reached a depth of 7,536 ft. drilling more than 4,500 ft. of granite. Once the drilling process was finished, a temperature log was run. The bottom hole temperature recorded was 386ºF at the total depth (TD) of the well. Most of the information generated by this project is public and can be accessed at Geothermal Data Repository (GDR) webpage (https://gdr.openei.org).

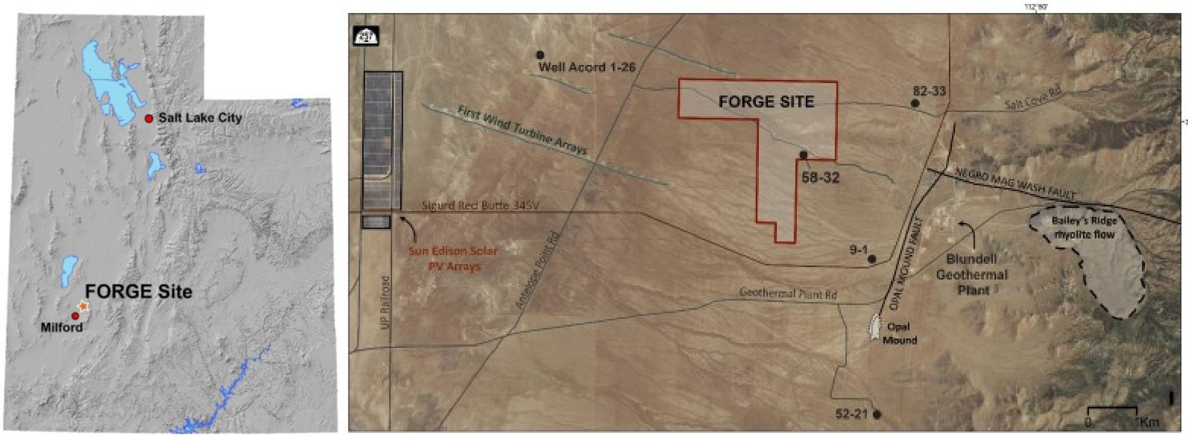

Figure 1: Location of FORGE site and the Well 58-32 (Utah FORGE, 2018)

2. METHODOLOGY

2.1 Existing Data

Information from logs, drilling reports, mud logs, and recorded parameters was analyzed to generate the model. The drilling parameters were collected from sensors around the rig. Standpipe pressure (SPP), rotary speed (RPM), rate of penetration (ROP), weight on bit (WOB), flow rate (FLOW_IN), and torque were the parameters selected for the analysis. These parameters were considered because they are the most frequently recorded, making this data easily available. Besides, these parameters are measured from the surface, and the sensors that are collecting those parameters are not affected by the high temperatures of geothermal operations. Table 1 summarizes the information considered. In Figure 2, the drilling parameters considered for this study are presented by the depth.

Table 1 Summary of information used to build the model*

Report or Log	Relevant Information
Daily Drilling Reports	BHA information, Bit size, and type, and nozzles MW, PV, YP
Pason drilling parameters per foot	SPP (psi), Torque (psi), ROP (ft/hr), WOB (klbs), FLOW_IN (gpm), RPM (rev/min).
Directional Survey	Azimuth Inclination
Core Data and Drilling Cuttings	Matrix Thermal Conductivity

*Source of information: Energy and Geoscience Institute at the University of Utah. (EGI 2018).

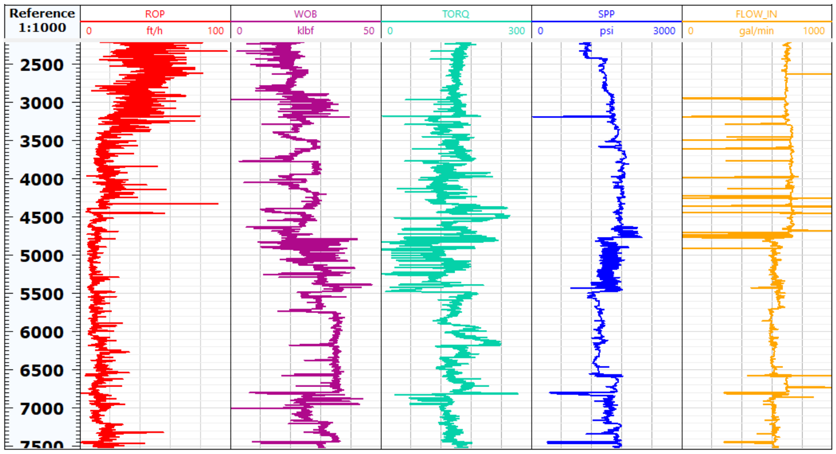

Figure 2: Drilling data for FORGE the Well 58-32

This study aims to build a model that predicts thermal conductivity in the presence of different unrelated drilling variables. The adaptive algorithms used in this study identify patterns in the data. Then, the model learns from the observations. As more observations the model is exposed, the better is this predictive performance (Rokach and Maimon 2005). Figure 3 presents the workflow used in this study to analyze the drilling parameters and generate the thermal conductivity predictors.

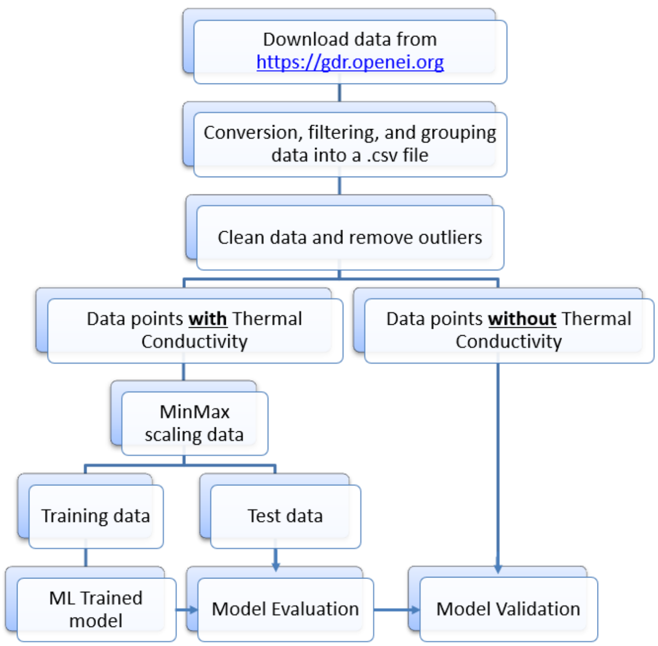

Figure 3: Workflow used in this study.

2.2 Data Preprocessing

Before using the data in a data analytics algorithm, they need to be preprocessed. Incomplete or inconsistent data may lead to unhelpful pattern extraction. In this stage, the raw data is cleaned and organized for the data analytics processing. During preparation, the raw data is checked for errors. This step aims to eliminate erroneous data (redundant, incomplete, or incorrect data).

For removing the outliers, boxplots provide a useful tool for their identification. As observed in Figure 2, some flow rate points are erroneous. For instance, there are values of flow rate equals zero that is probably incorrect. This is in consideration that the drilling operation requires mud circulation to remove cuttings. Even though if drilling operation is performed in total mud losses, a situation that is not uncommon in geothermal drilling, you still need to pump drilling fluid. The initial flow rate values were around 700 gpm up to 4,700ft of depth. Then, for the rest of the well, the flow rate was about 600 gpm. The mud pumps have different configurations that can be adjusted for certain flow intervals. Considering this, values of 1000 gpm and above or values below 500 gpm, are probably outliers. Figure 4 shows that the boxplot has a good representation of the data points, where outliers are easily identified. The same consideration is applied for standpipe pressure. Considering the flow rate window, values below 1000 psi are probably outliers.

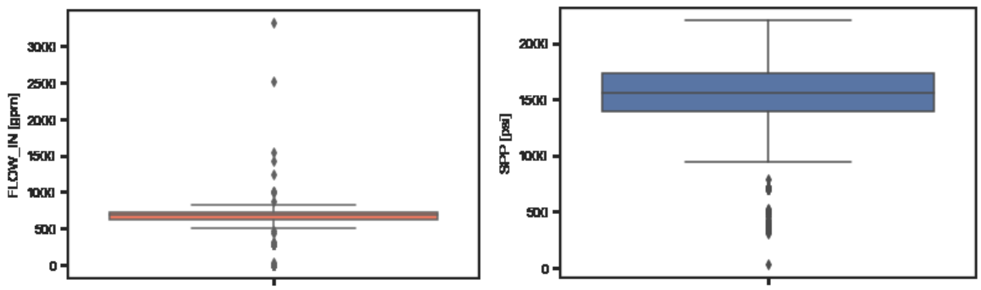

Figure 4: Boxplots for outliers' detection in flow rate (gpm) [left], and standpipe pressure (psi) [right].

A scatter plot with the drilling parameters is plotted with outliers removed, identifying relationships between the drilling parameters selected. As observed in Figure 5, there are values with positive correlation, such as torque and rotary speed, flow rate, and standpipe pressure. Those relations are normal since those variables are directly related. The ROP shows a moderate negative correlation with the weight on the bit and the standpipe pressure. This correlation is typical since standpipe pressure is expected to increase with the depth, and ROP is expected to decreases with depth. The increase in WOB with depth is probably a consequence of the reduction in the ROP. In general, the parameters follow a regular pattern conventionally seen in drilling operations.

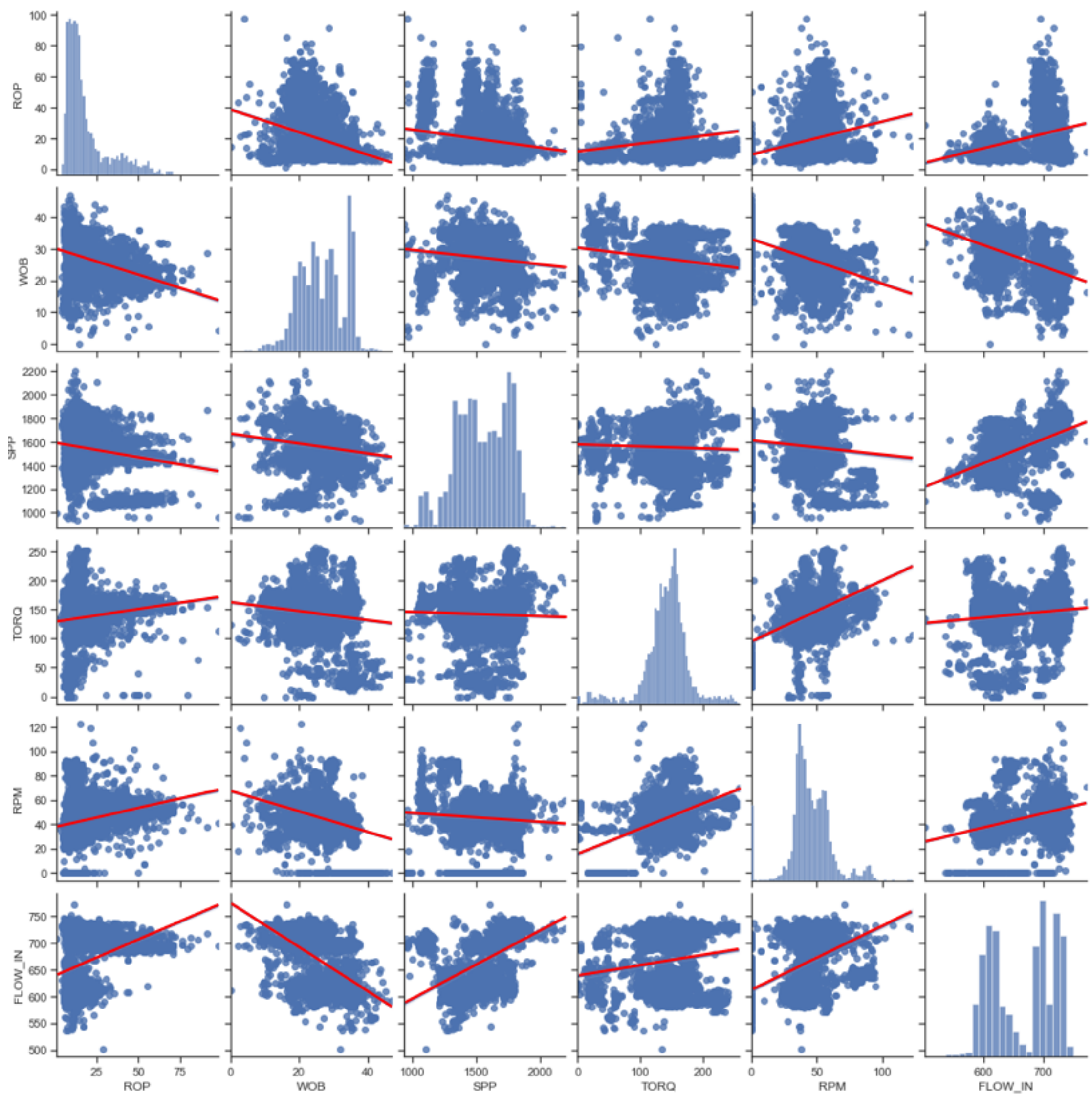

Figure 5: Scatter plot comparing the correlation between the standpipe pressure (SPP [psi]), rotary speed (RPM [rev/min]), rate of penetration (ROP [ft/hr]), weight on bit (WOB [klbs]), flow rate (FLOW_IN [gpm]), and torque.

Once the outliers were removed, the data was cleaned and scaled. Normalizing data to be able to analyze it optimally is essential for the data analysis.

2.2.1 Matrix thermal conductivity (MTC)

The data of MTC of the well 58-32 was generated from drilling cuttings. Fifty-four data intervals of 10 ft. were identified. Those data points are in the interval of thedrilling parameters recorded. Then, the data intervals that contain MTC values were separated from the data that does not have MTC values. In Figure 6, it is presented the relationships between drilling parameters and the thermal conductivity values. There are no strong correlations of the drilling parameters.

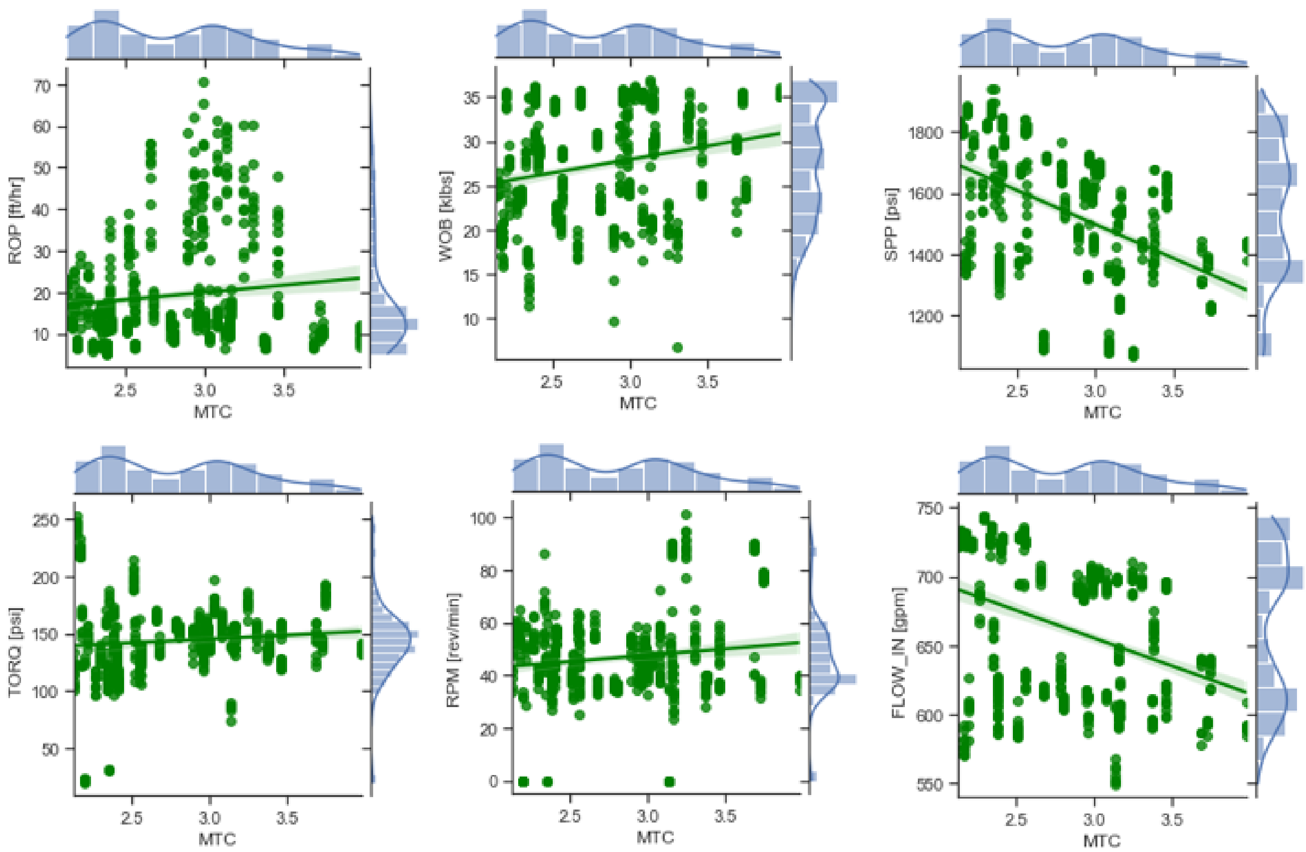

Figure 6: Scatter plot comparing the correlation between the thermal conductivity values (MTC [W/(m.K)]) vs. standpipe pressure (SPP [psi]), rotary speed (RPM [rev/min]), rate of penetration (ROP [ft/hr]), weight on bit (WOB [klbs]), flow rate (FLOW_IN [gpm), and torque (TORQ [psi]).

The data points that contain MTCs values (in this case, a data point has one value of each ROP, WOB, FLOW_IN, Torque, RPM, SPP, and MTC collected at the same depth) were separated, normalized, and split into data for training (85%) and data for testing (15%). The rest of the data that does not contain MTC values was used to evaluate the predictor's functionality in new data.

3. MACHINE LEARNING ALGORITHMS

Machine learning algorithms are widely used for the regression of patterns embedded in the datasets. This study has used two types of supervised learning-based approaches to assess each algorithm's accuracy that can quantify the matrix thermal conductivity from FORGE 58-32 well data. These models include Decision Tree, Random Forest, *k*-Nearest Neighbors, and Deep Neural Multilayer Perceptron Regressors. These are supervised learning algorithms. The reduction of a function is managed to predict any process's input value based on the repetition of a series of previous examples. In our case, the output of this process is a thermal conductivity numerical value.

3.1 Decision Tree algorithm

Tree-based learning algorithms have the advantage of comprehensibility, easy to understand and interpret (Rokach and Maimon 2005). The Decision Tree algorithm does not require extensive data preparation and has a relatively low computing cost (Ling et al. 2004). One noticeable downside of this algorithm is that the model's tree depth con generate overfitting. This overfitting will consequently reduce the accuracy of the model to predict thermal conductivity values.

The decision tree depth was adjusted in this analysis starting from 3, where the model underpredicts (training R_2=54%, test R_2=49.6%). In this case, the depth of the tree was increased to improve the prediction of MTC values. In this dataset, the best ratio between thermal conductivity actual values vs. predicted values has an R_2 (how the prediction fits with the regression) in the training data of 86.6% and the test data of 77.6% (Figure 7). The depth of the tree for this scenario is 8. The risk of increasing the depth of the tree disproportionately is to overfit the model. In this case, the decision tree moreover learns from the data noise and performs inaccurately with the new data.

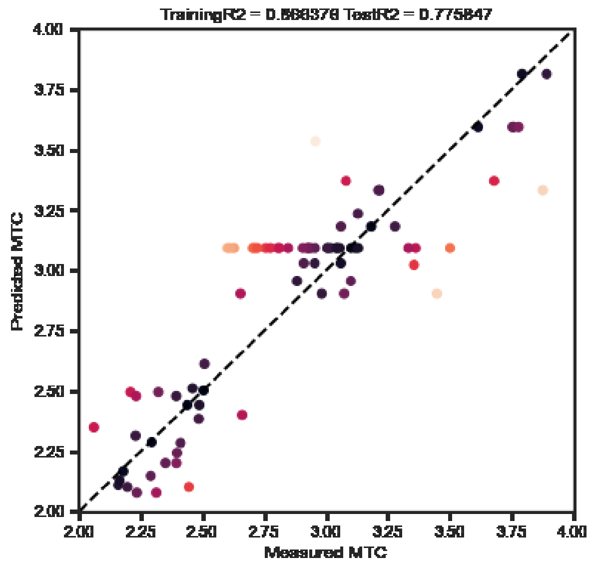

Figure 7: Thermal conductivity (MTC [W/(m.K)]) prediction accuracy using Decision Tree algorithm.

3.2 Random Forest algorithm

Random forest is another tree-based learning algorithm introduced by Breiman (2001). This consists of tree-based classifiers divided randomly. In this approach, each decision tree created the rule based on the answer and used it to find the answer that matched the rules. The multiple trees generate solutions that are voted for the prediction. In the end, the answer with the highest votes is the selected prediction. The randomness aims to create a variance that compensates the individual decision trees' tendency to overpredict (Breiman 2001).

In this dataset, the best ratio between thermal conductivity actual values vs. predicted values has an R_2 in the training data of 84.9% and in the test data of 78.4% (Figure 8). This result was obtained with the depth of the tree of 8, the same used in the decision tree regressor. In contrast to the decision tree regressor, the risk of overfitting is reduced if there are enough "trees" in the forest.

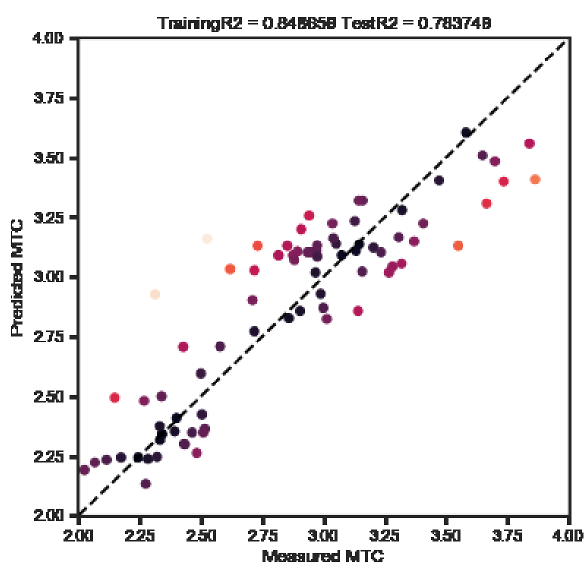

Figure 8: Thermal conductivity (MTC [W/(m.K)]) prediction accuracy using Random Forest algorithm.

3.3 Deep Neural Multilayer Perceptron Regressor

The Multilayer Perceptron is a deep neural algorithm introduced by Rosenblatt (1958). This technique analyzes which input features are or are not important. The output derivative expression will be extracted from the input; then, it will be seen how this can calculate a qualified multilayer perceptron sensitivity to each input function (Ruck et al. 1990).

When this neural network supervised algorithm was applied to this dataset, the ratio between thermal conductivity actual values vs. predicted values has an R_2 in the training data of 36.9% and the test data of 48.6% (Figure 9). In this case, we observed that the prediction of thermal conductivity values with this algorithm is lesser than the results obtained with the tree-based algorithms.

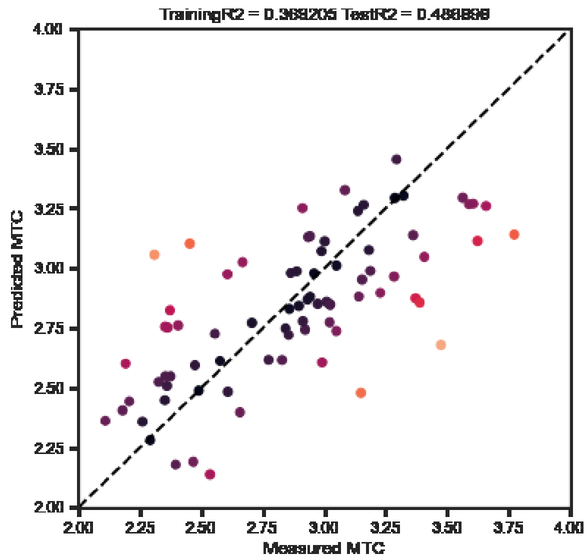

Figure 9: Thermal conductivity (MTC [W/(m.K)]) prediction accuracy using Decision Tree algorithm.

3.4 *k*-Nearest Neighbor (kNN) algorithm

k-Nearest Neighbor is a machine learning algorithm presented by Fix and Hodges (1951). This algorithm searches the closest observations to the response variable tried to predict. Then, the algorithm classifies the variable based on most of the surrounding data. This algorithm computes the distance between the variable of interest to be classified and the rest of the training dataset parameters. Then select the closest "k" elements. Finally, carry out a "poll" among the k points: those of a ruling class will decide its final classification.

When this nonparametric supervised algorithm was applied to this dataset, the ratio between thermal conductivity actual values vs. predicted values has an R_2 in the training data of 88% and the test data of 83.6% (Figure 10). The k-neighbors used were 6. This model has the highest prediction fit with the regression compared with the other models used.

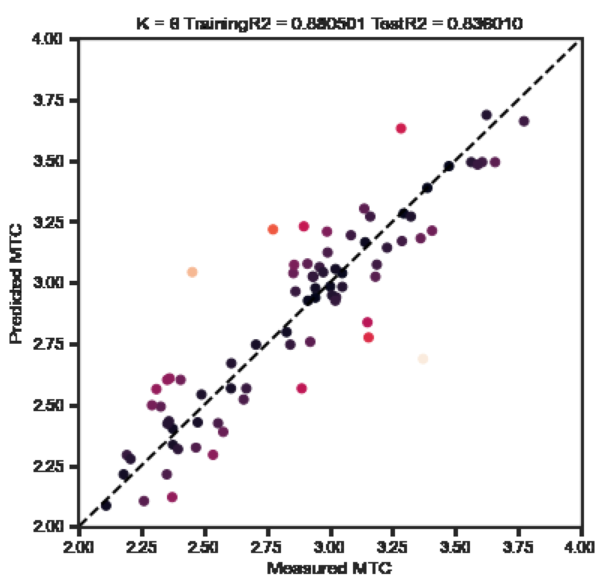

Figure 10: Thermal conductivity (MTC [W/(m.K)]) prediction accuracy using the k-Nearest Neighbor algorithm.

3.5 Testing the Real-Time thermal conductivity predictor

To evaluate how the model can work with the new data, the drilling data that does not contain MTC values were analyzed. In this case, the values of ROP, WOB, RPM, FLOW_IN, SPP, and Torque were computed with the k-Nearest Neighbor regressor to compute thermal conductivity. Figure 11 shows the predictor variables (ROP [track 1], WOB [track 2], RPM [track 3], FLOW_IN [track 4], SPP [track 5], and Torque [track 6]), and the response variable (MTC [track 7]). The data used to produce the predicted MTC is independent of the data used to build the model. To picture the prediction's performance, the real MTC values directly measured from rock samples were also plotted (blue dots in track 7).

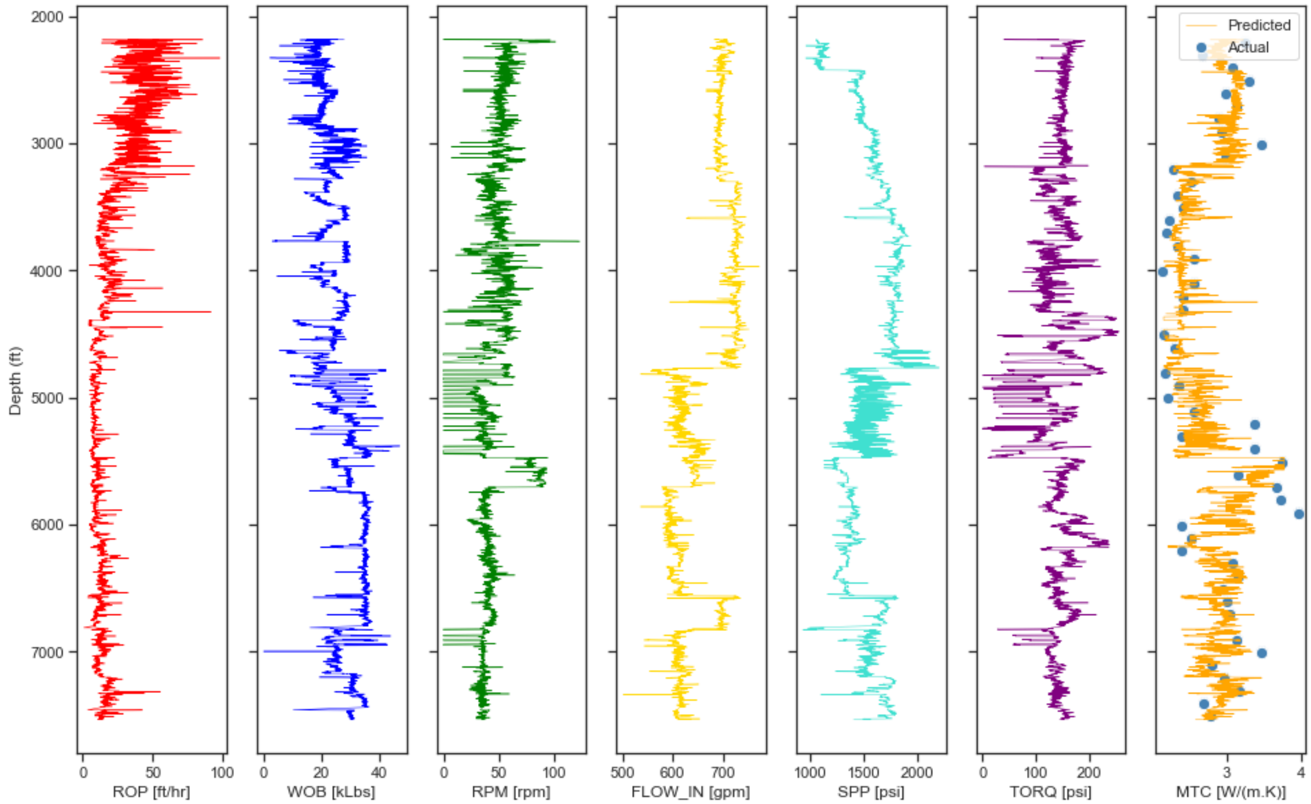

Figure 11: kNN algorithm thermal conductivity prediction accuracy. The predictor variables are ROP [track 1], WOB [track 2], RPM [track 3], FLOW_IN [track 4], SPP [track 5], and Torque [track 6], and the response variable is MTC [track 7].

It is observed that MTC values predicted with drilling data have a decent matching with MTC values measured from drilling cuttings. Since those drilling parameters are measured at the surface, they are less prone to fail to the high temperatures and corrosive environment found in geothermal wells than downhole sensors. Since the penetration rate, weight on bit, rotary speed, flow rate, torque, and standpipe pressure are standard parameters recorded in drilling operations, no additional instrumentation is required.

As the drilling variables selected are recorded in real-time, the model can predict MTC values in real-time. This application's potential is to make decisions in real-time, decide the best point to stop the drilling or deepen the well if necessary. This makes the usage of drilling data to predict petrophysical parameters a desirable avenue to obtain relevant data in real-time at a low cost.

4. CONCLUSION

In this study, we tested the concept of using machine learning algorithms to predict thermal conductivity using drilling parameters. Four different algorithms, Decision Tree, Random Forest, k-Nearest Neighbors, and Deep Neural Multilayer Perceptron Regressors, were evaluated. The ratio between thermal conductivity actual values vs. predicted values with kNN has an R_2 in the training data of 88% and the test data of 83.6%. The k-neighbors used was 6. This model has the highest prediction fit with the regression compared with the other models used. The predictor generated with kNN was then applied to new data to see the predictor's capability to predict MTC values. We found that the kNN regressor trained with drilling parameters was capable to predicted MTC values.

The impact of this study is the application of drilling parameters to predict thermal conductivity. Drilling parameters are easier to collect, and they are not significantly affected by high temperatures. The parameters selected are parameters conventionally collected in drilling operations, so this will not represent additional effort or resources. As the drilling data is generated in real-time, this tool can predict thermal conductivity in real-time. The ability to obtain MTC in real-time can allow instant decision-making or make changes or adjustments to the well plan if necessary. This solution can generate relevant information to estimate the potential of geothermal wells cost-effectively.

ACKNOWLEDGMENTS

The authors of this paper would like to thank the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Geothermal Program Office Award Number DE-EE0008602 for providing support to attend 46th Workshop on Geothermal Reservoir and present this work.

REFERENCES

Al-AbdulJabbar, A., Al-Azani, K., and Elkatatny, S. "Estimation of Reservoir Porosity From Drilling Parameters Using Artificial Neural Networks." Petrophysics **61** (2020): 318–330.

Bayan, M., and Zulkarnain.: Stuck Pipe Prediction in Geothermal Well Drilling at Darajat Using Statistical and Machine Learning Application. Proceedings of the 3rd Asia Pacific Conference on Research in Industrial and Systems Engineering. Association for Computing Machinery, New York, NY, USA, (2020), 100–104.

Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001).

Chen G, Chen M, Hong G, Lu Y, Zhou B, Gao Y. A New Method of Lithology Classification Based on Convolutional Neural Network Algorithm by Utilizing Drilling String Vibration Data. Energies. 2020; 13(4):888.

Di Sipio, E., S. Chiesa, and E. Destro.: Rock thermal conductivity as key parameter for geothermal numerical models. Energy Procedia **40**, (2013), 87–94.

Energy and Geoscience Institute at the University of Utah-EGI: Utah FORGE Logs and Data from Deep Well 58-32 (MU-ESW1) [data set] (2018). Retrieved from http://gdr.openei.org/submissions/1006.

Fix, E. and Hodges, J.L.: Discriminatory analysis: nonparametric discrimination, consistency properties. Book. USAF School of Aviation Medicine, Randolph Field, Texas. (1951).

Fuchs, S, Balling, N., and Forster, A.: Calculation of thermal conductivity, thermal diffusivity and specific heat capacity of sedimentary rocks using petrophysical well logs, Geophysical Journal International, **203**, (2015), 1977–2000.

Fuchs, S, and Forster, A.: Well-log-based prediction of thermal conductivity of sedimentary successions: a case study from the North German Basin, Geophysical Journal International, **196**, (2014), 291–311.

Gupta, I., Tran, N., Devegowda, D., Jayaram, V., Rai, C., Sondergeld, C., and Karami, H.: Looking Ahead of the Bit Using Surface Drilling and Petrophysical Data: Machine-Learning-Based Real-Time Geosteering in Volve Field, SPE J. **25** (2020): 990–1006.

Hartmann, A., Rath V., and Clauser C.: Thermal conductivity from core and well log data. International Journal of Rock Mechanics and Mining Sciences, **42**, (2005), 1042-1055.

Kanfar, R., Shaikh, O., Yousefzadeh, M. and Mukerji, T.: Real-Time Well Log Prediction from Drilling Data Using Deep Learning (2020)

Kiran, R, and Salehi, S.: Assessing the Relation between Petrophysical and Operational Parameters in Geothermal Wells: A Machine Learning Approach, *Proceedings,* 45th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2020).

Ling, C.X. Yang, Q., Wang, J., and Zhang, S.: Decision trees with minimal costs. In Proceedings of the twenty-first international conference on Machine learning (ICML '04). Association for Computing Machinery, New York, NY, 69, (2004).

Molaei, F., Kemeny, J., and Mehdi R.: Investigation of Thermal Conductivity in Rock Using Molecular Dynamics Simulation. Paper presented at the 54th U.S. Rock Mechanics/Geomechanics Symposium. (2020).

Rokach, L. and Maimon, O.: Clustering Methods. In Data Mining and Knowledge Discovery Handbook, ed. O. Maimon and L. Rokach, Springer Science+Business Media (2005), 321–352. New York City, New York, USA.

Rosenblatt, F.: The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review, **65**(6), 386–408. (1958).

Ruck, D.W., Rogers, S.K., and Kabrisky, M.: Feature Selection Using a Multilayer Perceptron. Journal of Neural Network Computing, Volume 2, Number 2, (1990), pp 40-48.

Sletcha, B., Vivas, C., K. Saleh, F., Ghalambor, A., and Salehi, S.: Digital Oilfield: Review of Real-time Data-flow Architecture for Upstream Oil and Gas Rigs. Paper presented at the SPE International Conference and Exhibition on Formation Damage Control, Lafayette, Louisiana, USA, (2020).

Utah FORGE: "Phase 2B Final Topical Report" Frontier Observatory for Research in Geothermal Energy – University of Utah (DOE Project DE-EE0007080) – Milford Site, Utah (2018).

Vivas, C., Salehi, S., Tuttle, J., and Rickard, B.: Challenges and Opportunities of Geothermal Drilling for Renewable Energy Generation. GRC Transactions, Vol. 44, (2020), 904-918.

Yuswandari, A., Prayoga, and Purba, D.: Rate of Penetration (ROP) Prediction Using Artificial Neural Network to Predict ROP for Nearby Well in a Geothermal Field. *Proceedings,* 44th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2019).