PROCEEDINGS, 44th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 11-13, 2019
SGP-TR-214

Estimation of Bottom Hole and Formation Temperature by Drilling Fluid Data: A Machine Learning Approach

Sercan Gul[1], Volkan Aslanoglu[1], Mahmut Kaan Tuzen[2], Erdinc Senturk[2]
[1] Middle East Technical University, Petroleum & Natural Gas Engineering Department, Ankara, Turkey
[2] Zorlu Energy, Reservoir Management, Istanbul, Turkey
sercan.gul@gmail.com, aslanoglu.volkan@metu.edu.tr, mahmut.tuzen@zorlu.com, erdinc.senturk@zorlu.com

Keywords: machine learning, bottom hole temperature, formation temperature, drilling fluids

ABSTRACT

Two of the most critical measurements during and after drilling a geothermal well are the bottom hole circulating temperature (BHCT) and the static formation temperature (SFT). BHCT is critical for bottom hole assembly (BHA), drilling fluid and cement slurry designs. SFT is significantly essential since it directly correlates with the amount of renewable energy power that can be produced from the well. Currently, these data are obtained from various equipment such as measurement while drilling (MWD), logging while drilling (LWD) and temperature logs. The measurements are used for calculations related to geothermal power plant construction as well as the drilling and completion designs of subsequent wells in the field. However, data from MWD is not always available. On the other hand, the process of taking a temperature log is time-consuming and expensive.

In this paper, we are proposing a machine learning approach to predict the BHCT and SFT in real-time using drilling fluid data (mud weight, rheological properties, flow in and out temperatures and circulation time). For the analyses, data from various wells were obtained from a large independent operator. Data available includes casing and drill string design, daily mud reports, flow rate, bottom hole temperature readings from MWD and the temperature log data obtained after drilling and testing the wells. This data was used for training and testing the machine learning algorithms. Two different models (random forest and XGBoost) were trained. 80% of data was used for training while 20% of data was used for testing the performance of the algorithm.

A perfect match with the trained models and testing dataset was observed with mean absolute percentage errors (MAPE) of less than 1% for both algorithms. The trained models can provide both BHCT and SFT with extremely high accuracies using the drilling fluid data which can be recorded on the surface in real-time. This paper presents a novel approach to estimate geothermal well temperatures and is believed to be very beneficial for practicing engineers to save a significant amount of time and cost in geothermal development projects.

1. INTRODUCTION AND BACKGROUND

During drilling and completion of geothermal wells, it is essential to predict the downhole temperature of drilling fluid and cement systems as well as the temperature of downhole formations for adequate design and composition of drilling fluids, cement slurries, and the bottom hole assemblies. Previous studies show that increased fluid temperature decreases the density and rheological properties of the fluid (Johnson, et al. 2018). Therefore, accurate knowledge of BHCT is essential to properly design both mud weight and rheology according to downhole conditions (Gul, Kuru and Parlaktuna 2017). Thickening time of cement slurries is also directly correlated with temperature. Optimized cement slurry design by accurate prediction of both SFT and BHCT will decrease non-productive time (NPT) and contribute further in cost savings. Similarly, sensors used in bottom hole assemblies have critical temperature limits over which signal quality issues can be observed. There are various studies on predicting BHCT or SFT using physics-based models, analytical models or artificial neural networks.

An early study in the subject (Cao, Hermanrud and Lerche 1988) estimates the SFT by inversion of borehole temperature. This study uses a numerical model called Formation Temperature Estimation (FTE) model. The model also estimates mud temperature at the time circulation stops, thermal invasion distance into the formation, and formation thermal conductivity. The model requires the use of lithology data. However, it was developed and tested for only one well and worked only in static conditions. Moreover, the model does not consider the effect of drilling fluid parameters (e.g., rheology or mud weight) in its calculations.

A thermal simulator developed by researchers (Garcia, et al. 1998) estimates the temperature distribution in the drill pipe and annulus while circulating. The simulator works with inputs of well geometry, the fluid and flow characteristics and the initial formation temperature. However, the use of this model requires the source code and expensive solutions of numerical models which is time-consuming.

An artificial neural networks (ANN) approach was presented (Bassam, et al. 2010) for the calculation of SFT in geothermal wells. The ANN model was trained using the features of BHCT, shut-in times, and the temperature gradients. Even though the model has high accuracy (>95%), it is not able to measure BHCT. Similar to the previous model, this approach also does not take into consideration the drilling fluid parameters (e.g., rheology or mud weight).

Another study (Feng 2011) discussed the prediction of temperatures in deep-water vertical oil and gas wells. The study presented two analytical methods. The proposed methods used finite difference discretization to predict the circulating drill pipe and annulus temperatures in steady-state. However, these models were only tested in offshore wells but not in geothermal wells, which have larger temperature gradients compared to conventional oil and gas wells (Gul and Aslanoglu 2018).

A study using GTEMP simulation (Tekin and Akin 2011) estimates the formation and bit temperatures using the features of mud in and out temperatures. The study was conducted in two different cases; the first case concerning the cooling tower effect and the second case not concerning this. Formation temperatures were estimated for five different wells and compared with reservoir temperature data obtained from the Horner Plot method. Estimations deviated within 3.6% to 25.2% in Case 1 and 1.5% to 24.5% in Case 2.

It was observed that some studies require formation temperature for bottom hole temperature estimations and some others require the reverse. There also some models which use the well depth or formation temperature gradient as features for accurate predictions. The models which use the same feature as in the proposed model are not as accurate. In this study, we are presenting a novel machine learning approach which can predict the BHCT and SFT with accuracies higher than 99% using the features of mud weight, rheological properties, flow in and out temperatures and bottoms up time of the fluid.

2. MACHINE LEARNING

Machine learning regression algorithms were used here to make more reliable predictions for the complicated system of temperature distribution in geothermal wells. Commonly used machine learning regression models include:

1. Linear/polynomial regressions;
2. Neural networks (ex: artificial, multi-layer perceptron, convolution);
3. Decision-tree-based models (ex: random forest, XGBoost).

Linear/polynomial regression is computationally fast because it uses only small amounts of data (Gul, Johnson, et al. 2019). However, a regression model can be very challenging to design for non-linear data. As a result, when it comes to modeling complicated feature-output relationships, such models are not very useful. Neural networks can model complex non-linear relationships given a reasonable amount of training data. Bassam et al. (2010) proposed a prediction of SFT using an ANN approach for geothermal wells. Other studies (Haklidir and Haklidir 2019) proposed a long-short-term memory (LSTM) approach to relating the input variable (hydro geochemistry data) and output variable (reservoir temperature) associated with the problem.

However, a neural network approach is generally outperformed by other machine learning approaches when the number of training samples is small. Tree-based approaches gather many trees to learn complicated and non-linear feature-output relationships. Bagging (random forest) and boosting (gradient boosting) are standard methods used. Trained tree-based models are particularly easy to interpret and understand because the trained decision boundaries are practical and intuitive (Gul, Johnson, et al. 2019). For these reasons, we used random forest and XGBoost regressions to predict SFT and BHCT in this study.

3. DATA DETAILS

The obtained data was in a low frequency, especially in drilling fluid rheology values (3-4 full mud checks for a day). During drilling, the mud in and mud out temperatures were available for every thirty minutes (48 data for a day). The bottom hole temperature data was obtained in every connection from MWD. SFT data was available from temperature logs which were taken after drilling and completing the wells. Since all these data were obtained in various depths, linear interpolation on all the features was performed to match the depth index. After interpolations, a total of 11006 lines of data became available for two wells.

For the data mining analysis in this study, random-forest and XGBoost regression machine learning algorithms from Scikit-Learn library (Pedregosa, et al. 2011) were used. Predictions were made by predicting both the SFT and BHCT. The optimum tuning parameters in random-forest and XGBoost regression models were evaluated by cross-validation (GridSearchCV) and minimizing the MAPE. The minimum and maximum values of the independent variables in the dataset are provided in **Table 1**. The trained models are valid only in the ranges shown in **Table 1**, because they may give erroneous results when extrapolated. For the analyses, 80% of the data from the dataset was used to train the model, while the rest of the data (20%) was used for testing model accuracy and analyzing MAPE. Only the results from the random forest regression approach are presented here because this approach yielded a superior performance.

Table 1. Details of the parameters in the training dataset.

Description	Bottoms up time (min)	Temperature difference (°C)	Mud weight (ppg)	m	τ_y	K	Bottom hole temperature (°C)	Formation temperature (°C)
count	11006	11006	11006	11006	11006	11006	11006	11006
mean	40.10	9.97	9.36	0.64	2.08	0.31	99.87	171.40
std	8.89	5.12	0.29	0.05	0.88	0.11	26.72	46.51
min	12.44	0.90	8.50	0.47	0.38	0.03	33.30	50.60
25%	34.11	6.10	9.30	0.61	1.68	0.24	76.40	147.22
50%	42.78	10.32	9.46	0.64	2.00	0.29	98.20	183.29
75%	46.95	12.71	9.52	0.68	2.28	0.37	128.90	213.86
max	55.69	38.60	10.00	0.90	9.77	0.96	143.30	219.55

As mentioned before, the data for this study was obtained from two different vertical geothermal wells which were designed and constructed similarly to each other. Well schematics are shown in **Fig. 1**. Both wells were drilled in 26", 17.5", 12.25" and 8.5" sections and cased by 20", 13.375", 9.625" casings and 7" standard and slotted liners. The final depth of Well A was 3508m while the final depth of Well B was 3025m.

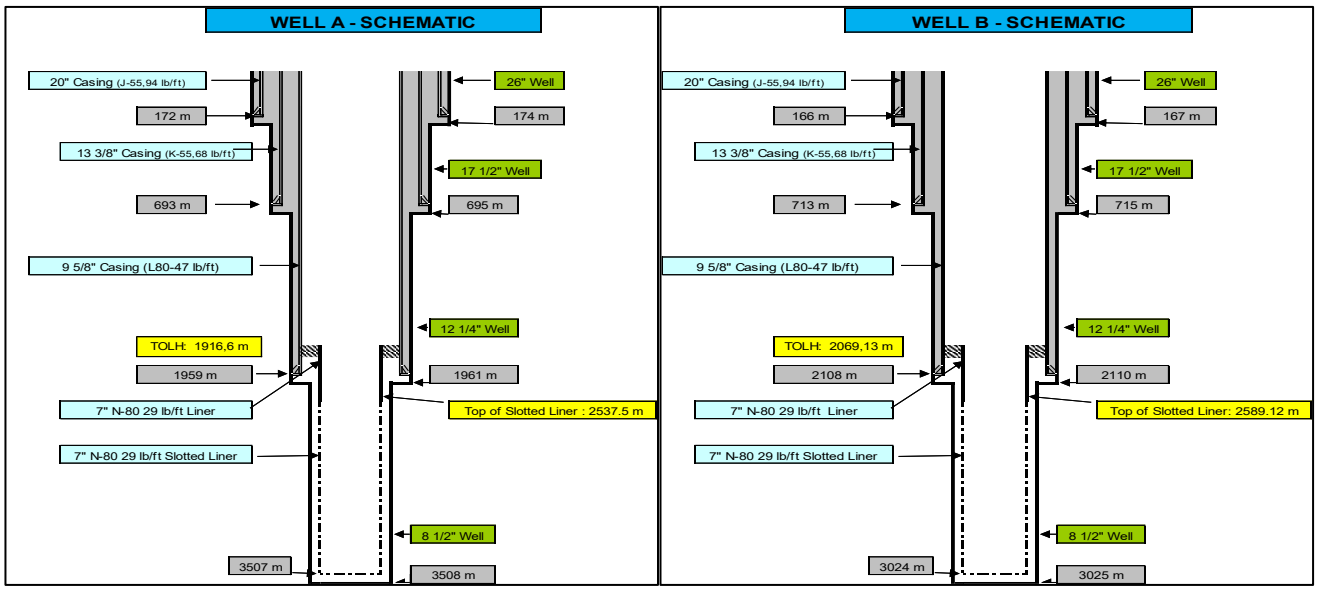

Figure 1. Design schematics of Well A (left) and Well B (right).

Even though the designs of both wells were very similar, the temperature distributions observed were slightly different. Mud in, out, BHCT and SFT for both wells are illustrated in **Fig. 2.** Both wells were drilled in different formations where the formation gradients were also slightly different. Cooling towers were used more often in Well B, which was the reason for sudden decreases in mud in, out and BHCT temperatures. Well A observed losses at 3500m at which new mud was pumped into the well which decreased the mud temperatures. Well B observed similar losses at 3000m, and the same response on temperatures was observed due to pumping new and cold mud to the formation. A direct relationship of an increase in mud in and out temperatures with depth can be seen on both Well A and Well B, as expected.

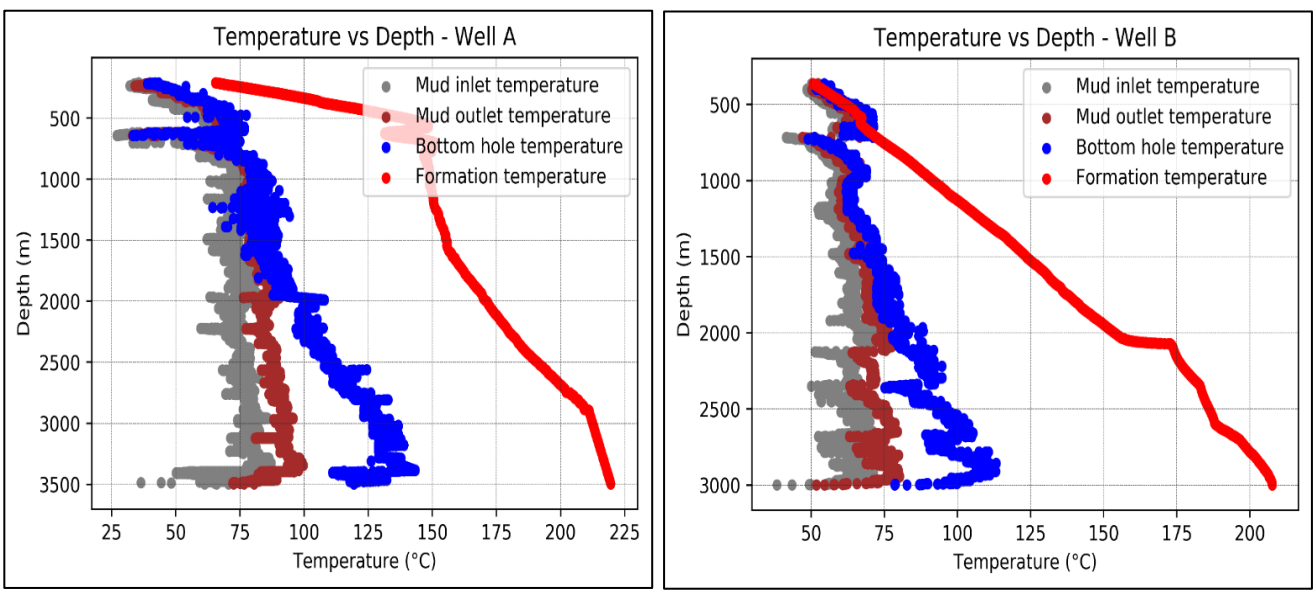

Figure 2. Temperature distribution data plots for Well A (left) and Well B (right).

Accurate predictions of temperatures in geothermal wells require the time drilling fluid spends in the wellbore during drilling or circulation. To calculate this feature, drillstring, and casing design information for each section was used. The pipe and annulus capacities were calculated using **Eqs. 1 and 2.**

$$C_A = (D_{bit}^2 - OD_{DP}^2) / 1029.4 \quad (1)$$

$$C_P = (ID_{DP}^2) / 1029.4 \quad (2)$$

where C_A is annulus capacity (bbl/ft), C_P is pipe capacity (bbl/ft), D_{bit} is bit diameter in the section (in), OD_{DP} is drill pipe outside diameter (in) and ID_{DP} is drill pipe inside diameter (in).

Once the capacities of annulus and drill pipe are known, well capacity was calculated using **Eq. 3**.

$$C_W = C_A + C_P \quad (3)$$

where C_W is the well capacity (bbl/ft).

Once the well capacity was obtained, the volume at each depth for the same section was calculated using **Eq. 4**.

$$V = C_w * L \quad (4)$$

where V is total circulating volume (bbl), C_w is well capacity (bbl/ft), and L is the length of the section (ft).

This calculation was performed for each section separately assuming each casing setting depth is zero and using the section lengths instead of depths. Afterward, the volume of the previous section was recalculated by changing the bit diameter with casing inside diameter and adding this to the calculated circulating volume. Once the volume was known, the bottoms up time were calculated using **Eq. 5**.

$$t = V * 42 / Q \quad (5)$$

where V is total circulating volume (bbl), Q is flow rate (gpm), and t is bottoms up circulating time (min).

The bottoms up time vs. depth graph for each well is shown in **Fig. 3**. The intermittent areas correspond to each casing setting point where the bottoms up time significantly decreased due to the decreased diameters by casings. As expected, the bottoms up times mostly increase as the well gets deeper. Even though Well A and Well B were similar in their casing and drill string design, the flow rates during the operations were different which resulted in different patterns on bottoms up time vs. depth graphs.

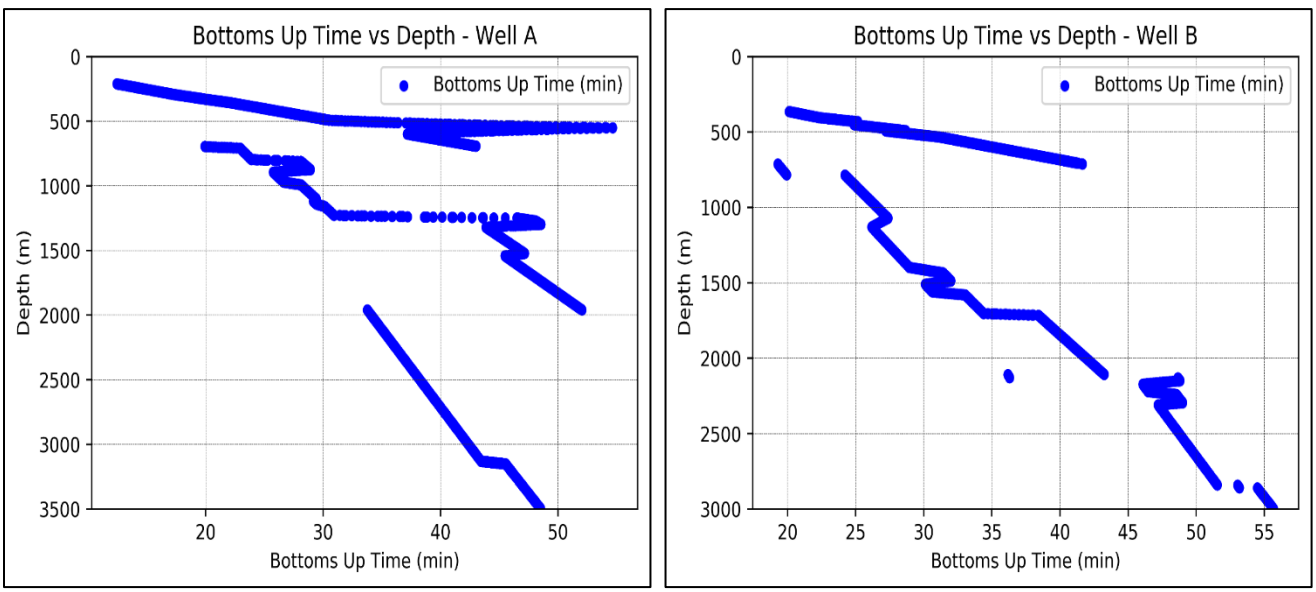

Figure 3. Calculated bottoms up time vs depth for Well A (left) and Well B (right).

Maximum bottoms up times observed were 58 min for Well A and 56 min for Well B. More flow rate fluctuations were observed in Well A, while a more stable flow rate was observed in Well B during drilling each section. The well design and selected flow rates can change from one well to another, however, the time drilling fluid spends in the well is the primary factor affecting the heating up of the fluids. Since the proposed model work with the surface measurements of mud in and out temperature, the times for heating up was also crucial to properly train the models and therefore to obtain accurate predictions for both BHCT and SFT.

The coefficient of determination (R^2), mean absolute error (MAE) and mean absolute percentage error (MAPE) were used to test the model accuracies. **Eqs. 6 and 7** explain the calculations of MAE and MAPE.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(y_{i-calculated} - y_{i-measured})| \quad (6)$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|(y_{i-calculated} - y_{i-measured})|}{y_{measured}} \quad (7)$$

where $y_{i\text{-calculated}}$ is each calculated data point, $y_{i\text{-measured}}$ is each measured data point, and n is the total number of data points.

4. REGRESSION RESULTS

Machine learning regressions for BHCT were developed using the data from both Well A and Well B. The minimum BHCT in the dataset was 33.3°C while the maximum was 143.3°C. Mud in and out temperatures, bottoms up time, rheological properties characterizing drilling fluid rheology (m, K and τ_y) and mud weight were used as features for the predictions of BHCT in random forest regressions. **Fig. 4** illustrates the measured and predicted values of BHCT using random forest regression with 10% confidence interval. The regression provided a fit with a MAE of 0.63°C and MAPE of 0.71%. The coefficient of determination (R^2) of this regression is 0.997. There are only three outliers out of 2200 data points in the test dataset which suggests that these points are outliers due to slight measurement errors in raw data (maximum absolute error 18.44 °C, maximum absolute percentage error 22.43%).

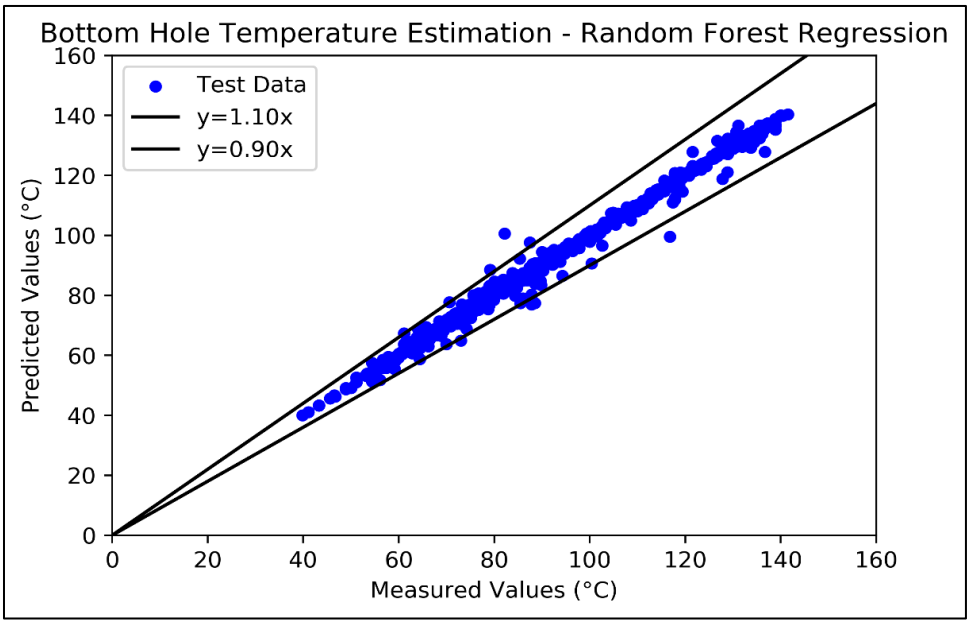

Figure 1. Bottom hole circulating temperature estimation accuracy with 10% confidence interval using the features of fluid rheological properties (τ_y, K and m), mud weight, mud in and out temperatures and bottoms up time.

The regressions for SFT were developed using the same data and features in the previous case. **Fig. 5** illustrates the measured and predicted values of SFT using random forest regression with 10% confidence interval. The regression provided a fit with an MAE of 0.07°C and MAPE of 0.06%. R^2 of this regression is 0.999. There was only one outlier out of all test data with a maximum absolute error 11.09 °C and maximum absolute percentage error 10.42%.

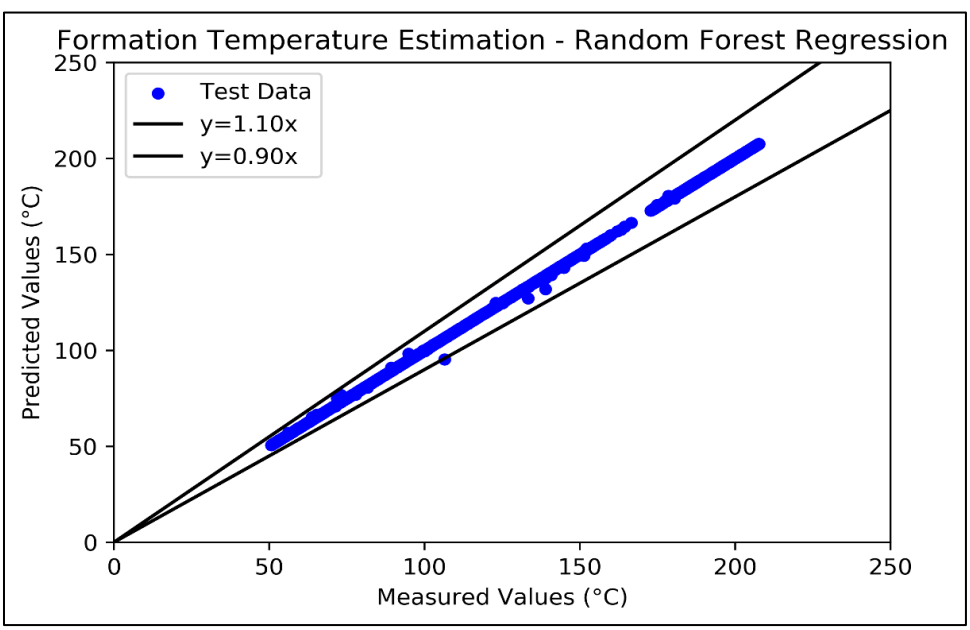

Figure 2. Formation temperature estimation accuracy with 10% confidence interval using the features of fluid rheological properties (τ_y, K and m), mud weight, mud in and out temperatures and bottoms up time.

As mentioned before, random forest regression showed superior performance compared to XGBoost regression to estimate BHCT and SFT using fluid data. For this reason, only random forest regressions were illustrated in the previous figures. **Table 2** summarizes the errors in estimations of BHCT and SFT by random forest and XGBoost machine learning regressions with their average and maximum errors and coefficient of determination results.

Table 2. Summary of errors and coefficient of determination in different regression models for estimation of the bottom hole and formation temperatures.

Calculated Parameter	Regression Method	Maximum Absolute Error (°C)	Mean Absolute Error (°C)	Max Absolute Percentage Error (%)	Mean Absolute Percentage Error (%)	R^2
BHCT	Random Forest Regression	18.44	0.63	22.43	0.71	0.997
BHCT	XGBoost Regression	42.53	2.93	52.44	3.26	0.973
SFT	Random Forest Regression	11.09	0.07	10.42	0.06	0.999
SFT	XGBoost Regression	21.31	2.01	22.9	1.98	0.970

5. DISCUSSION AND CONCLUSIONS

- For the wells investigated, BHCT and SFT can be determined with excellent accuracy using machine learning techniques.
- The machine learning relationships can be different for different geothermal wells drilled in different fields or formations. But, once a relationship is characterized using sufficient field data, it is evident that both BHCT and SFT can be estimated with high accuracies.
- This approach offers a novel way to obtain fluid temperature profiles in real-time and can be improved when used together with automated drilling fluid measurement methods which were explained with further detail on other studies (Gul, Johnson, et al. 2019).
- It was evident that the bottom hole temperature measurements from MWD and temperature log data after drilling the wells are required for optimum performance and analysis after the project. This method, however, is not to replace but to improve them:
    - Verifying the quality of MWD tools using the regressions is beneficial. A better-predicted bottom hole temperature will elevate the drilling fluid quality which improves the rate of penetrations and results in further cost savings.
    - Using this method, there is no need to wait for the temperature logs which are taken after the well is finished. After each section, proper estimations of formation temperatures can be done, and cement slurries can be designed accordingly. This will decrease the cement thickening times (which is sometimes up to 48 hours for geothermal wells (Gul and Aslanoglu 2018)) and contribute significantly to decreasing well costs.

NOMENCLATURE

C_A = annulus capacity, L^2, bbl/ft
C_P = pipe capacity, L^2, bbl/ft
C_W = well capacity, L^2, bbl/ft
D_{bit} = bit diameter, L, m
ID_{DP} = inside diameter, L, in
OD_{DP} = outside diameter, L, in
R^2 = coefficient of determination
τ_y = yield stress, m/Lt^2, Pa
l = length, L, m
m = fluid behavior index
n = number of data points
t = time, t, min
K = consistency index, mt^{m-2}/L, Pa.s^m
Q = volumetric flow rate, L^3/t, gpm
V = volume, L^3, gallons

GLOSSARY

ANN = Artificial Neural Networks
BHA = Bottom Hole Assembly
BHCT = Bottom Hole Circulating Temperature
FTE = Formation Temperature Estimation
LSTM = Long-Short Term Memory
LWD = Logging While Drilling
MAE = Mean Absolute Error
MAPE = Mean Absolute Percentage Error
MWD = Measurement While Drilling
NPT = Non-Productive Time
SFT = Static Formation Temperature

ACKNOWLEDGMENTS

The authors like to thank Zorlu Energy Group for providing the datasets for the analyses as well as their support and encouragement throughout the study.

REFERENCES

Bassam, A, E Santoyo, J Andaverde, and OM Espinoza-Ojeda. 2010. "Estimation of static formation temperatures in geothermal wells by using an artificial neural network approach." *Computers and Geosciences* (Elsevier) 36 (9): 1191-1199. Accessed January 18, 2019. doi:doi.org/10.1016/j.cageo.2010.01.006.

Cao, S, C Hermanrud, and I Lerche. 1988. "Estimation of formation temperature from bottom-hole temperature measurements: COST #1 well, Norton Sound, Alaska." *Geophysics* (Society of Exploration Geophysicists) 53 (12): 1619-1621. Accessed January 18, 2019. doi:10.1190/1.1442447.

Duan, Chenyang, Xiuhua Zheng, Bairu Xia, and Zhiqing Wang. 2016. "Temperature distribution modeling in geothermal wellbore and formation during the well test in Yangyi geothermal field, Tibet." *41st Workshop on Geothermal Reservoir Engineering.* Stanford: Stanford University. Accessed January 18, 2019.

Feng, Ming. 2011. *The temperature prediction in deepwater drilling of vertical well.* College Station, Texas: Texas A&M University. Accessed January 18, 2019. https://core.ac.uk/download/pdf/9068886.pdf.

Garcia, A, I Hernandez, G Espinosa, and E Santoyo. 1998. "Temlopi: A thermal simulator for estimation of drilling mud and formation temperatures during drilling of geothermal wells." *Computers and Geosciences* (Elsevier) 24 (5): 465-477. Accessed January 18, 2018. doi:10.1016/S0098-3004(98)00025-9.

Gul, Sercan, and Volkan Aslanoglu. 2018. "Drilling and well completion cost analysis of geothermal wells in Turkey." *43rd Workshop on Geothermal Reservoir Engineering.* Stanford: Stanford University. Accessed January 18, 2019. https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2018/Gul.pdf.

Gul, Sercan, Ergun Kuru, and Mahmut Parlaktuna. 2017. "Experimental investigation of cuttings transport in horizontal wells using aerated drilling fluids." *Abu Dhabi International Petroleum Exhibition and Conference.* Abu Dhabi: Society of Petroleum Engineers. Accessed January 18, 2019. doi:10.2118/188901-MS.

Gul, Sercan, Mitchell David Johnson, Ali Karimi Vajargah, Zheren Ma, Besmir Buranaj Hoxha, and Eric van Oort. 2019. "A data driven approach to predict frictional pressure losses in polymer-based fluids." *SPE/IADC Drilling Conference and Exhibition.* The Hauge: Society of Petroleum Engineers. Accessed January 18, 2019.

Haklidir, Fusan Tut, and Mehmet Haklidir. 2019. "Geothermal reservoir temperatures prediction with a deep learning model: a case study western Anatolia." *Stanford Geothermal Workshop.* Stanford: Stanford University. Accessed January 18, 2019.

Johnson, Mitchell, Sercan Gul, Ali Karimi, Besmir Buranaj Hoxha, and Eric van Oort. 2018. "Real-time friction factor monitoring: characterization of drag reduction in polymer-based fluids." *AADE Fluids Conference and Exhibition.* Houston: American Association of Drilling Engineers. Accessed January 18, 2019. https://www.researchgate.net/publication/323104728_Real-time_Friction_Factor_Monitoring_Drag_Reduction_Characterization_of_Polymer-Based_Fluids.

Pedregosa, Fabian, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-learn: machine learning in python." *Journal of Machine Learning Research* (Journal of Machine Learning Research) 12 (October): 2825-2830. Accessed January 18, 2019.

Priyangga, Heneka Yoma, Auzan Abirama Soedarmo, Konark Ogra, and Aliya Zhabagina. 2017. "Numerical approach for temperature estimation in geothermal drilling operation." *42nd Workshop on Geothermal Reservoir Engineering.* Stanford: Stanford University. Accessed January 18, 2019. https://pangea.stanford.edu/ERE/db/IGAstandard/record_detail.php?id=27970.

Tekin, Sema, and Serhat Akin. 2011. "Estimation of the formation temperature from the inlet and outlet mud temperatures while drilling geothermal formations." *Thirty-Sixth Workshop on Geothermal Reservoir Engineering.* Stanford: Stanford University. https://pangea.stanford.edu/ERE/pdf/IGAstandard/SGW/2011/tekin.pdf.