PROCEEDINGS, 48th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 6-8, 2023
SGP-TR-224

Drilling the Perfect Geothermal Well: An International Research Coordination Network for Geothermal Drilling Optimization Supported by Deep ML and Cloud Based Data Aggregation

Adam Schultz[1], Pradeepkumar Ashok[2], Alain Bonneville[1], Daniel Bour[3], Rolando Carbonari[4], Trenton Cladouhos[5], Geoffrey Garrison[6], Roland Horne[7], Gunnar Kaldal[8], Susan Petty[6], Robert Rallo[9], Carsten Sorlie[10], Dang Ton[7], Matt Uddenburg[11], Eric van Oort[2], Leandra Weydt[12]

[1] College of Earth, Ocean and Atmospheric Sciences, Oregon State University, Corvallis OR, USA; [2] Department of Petroleum & Geosystems Engineering, University of Texas at Austin, Austin TX, USA; [3] Bour Consulting, Granite Falls, WA, USA; [4]Institute of Earth Sciences, Hebrew University of Jerusalem, Israel; [5] 4010 Stone Way North Suite 400, Seattle WA, USA; [6]AltaRock Energy, Inc. Seattle WA, USA; [7]Precourt Institute for Energy, Stanford University, Stanford, CA, USA; [8]Iceland Geosurvey (ISOR), Kópavogur, Iceland; [9]Pacific Northwest National Laboratory, Richland, WA, USA; [10]Equinor, Trondheim, Norway; [11]Stravan Consulting, Philadelphia, PA, USA; [12]Technische Universitat, Darmstadt, Germany

Adam.Schultz@oregonstate.edu, pradeepkumar@mail.utexas.edu, bonnevia@oregonstate.edu, daniel_bour@outlook.com, rolandocarbonari@hotmail.com, ttcladouhos@gmail.com, ggarrison@altarockenergy.com, horne@stanford.edu, Gunnar.Skulason.Kaldal@isor.is, spetty@hotrockenergy.org, robert.rallo@pnnl.gov, cso@equinor.com, tonthataidang@gmail.com, muddenberg@stravan.co, vanoort@austin.utexas.edu, leandra.weydt@tu-darmstadt.de

Keywords: geothermal well, machine learning, expert system, well optimization

ABSTRACT

Motivated by the potential of recent developments in data science to minimize the cost of drilling geothermal wells and to reduce well failures, the *EDGE: Drilling the Perfect Geothermal Well Consortium* (EDGE) developed a database of geothermal drilling data from 113 wells in various geologic settings. After extensive exploratory data analysis and data quality control checks, the database was used to develop an expert system using machine learning tools designed to guide geothermal drillers to realize substantial reductions in the cost of geothermal wells. Given the experience gained in the assimilation of well data from multiple sources into the EDGE repository, a series of best practices were developed for ingestion of heterogenous data, mitigating the impacts of missing and erroneous data on subsequent analyses. The data repository was designed, built, and stored securely by Pacific Northwest National Laboratory (PNNL) so that proprietary data provided to the consortium remains proprietary to the provider, while at the same time enabling those data as well as open access data to form the training set for the expert system. Teams at Oregon State University (OSU), Stanford University (Stanford), and the University of Texas at Austin (UTA) used modern machine learning (ML) techniques to develop methods for predicting cost, Rate of Penetration (ROP), bit wear, and other important parameters for a variety of environments. The EDGE consortium expert system will be available for users through a tiered subscription model designed to encourage organic growth of the well database, and continuous improvements of the expert system through expansion of the well data training set.

1. INTRODUCTION

Carbonari, et al (2021) reviewed the impetus for geothermal well drilling optimization, noting that drilling operations alone constitute 30%-70% of the total cost of geothermal projects (Saleh, et al., 2020; Dumas et al., 2013). While a variety of factors are known to be causative for increasing drilling and completion costs, efforts to provide a more systemic overview and ranking of these factors has been hampered by the limitations of existing well data sets. For instance, efforts have been made to reduce non-productive time (NPT), well-casing failures, the possibility and impact of lost circulation, and premature drill bit failure (Salehi et al., 2013; Karimi et al., 2011; Saleh et al., 2020; Imaizumi et al.; 2019; Miyazaki et al., 2019; Raymond, et al., 2012). Both physics-based and data-driven (ML) models have been applied to the problem of optimizing drilling parameters by focusing in on increasing the rate of penetration (ROP) and the mechanical specific energy (MSE), i.e. (Alali et al., 2020; Sabah, et al., 2019; Hedge, et al., 2017, 2018; Basarir et al., 2014). The limitations of existing data sets often reduced the applicability of some of these previous studies to specific well fields and geological contexts and well configurations, and they may have relied on detailed logs that are not always available from drilling operators. Furthermore, while improving ROP will likely have a positive impact on drilling costs, a systemic overview of all factors that drive such costs up is warranted.

In part to address some of these limitations, the US Department of Energy Geothermal Technologies Office supported EDGE Consortium developed a sustainable, organically expandable database of complex, heterogeneous geothermal drilling data from 113 wells in various geologic settings in the US and Iceland (see Figure 1 below). An expert system was trained on that database using machine learning tools to guide geothermal drillers to substantially reduce the cost of geothermal wells. Best practices were developed by the EDGE Consortium for ingestion of heterogenous geothermal data containing missing and erroneous data, complementing previous efforts at establishing best practices for geothermal drilling (Finger and Blankenship, 2010). The resulting cloud-hosted FAIR data repository based on CKAN (Winn, 2013), compatible and interoperable with existing geothermal data repositories was designed, built, and stored by PNNL while assuring that data provided remains proprietary to the provider.

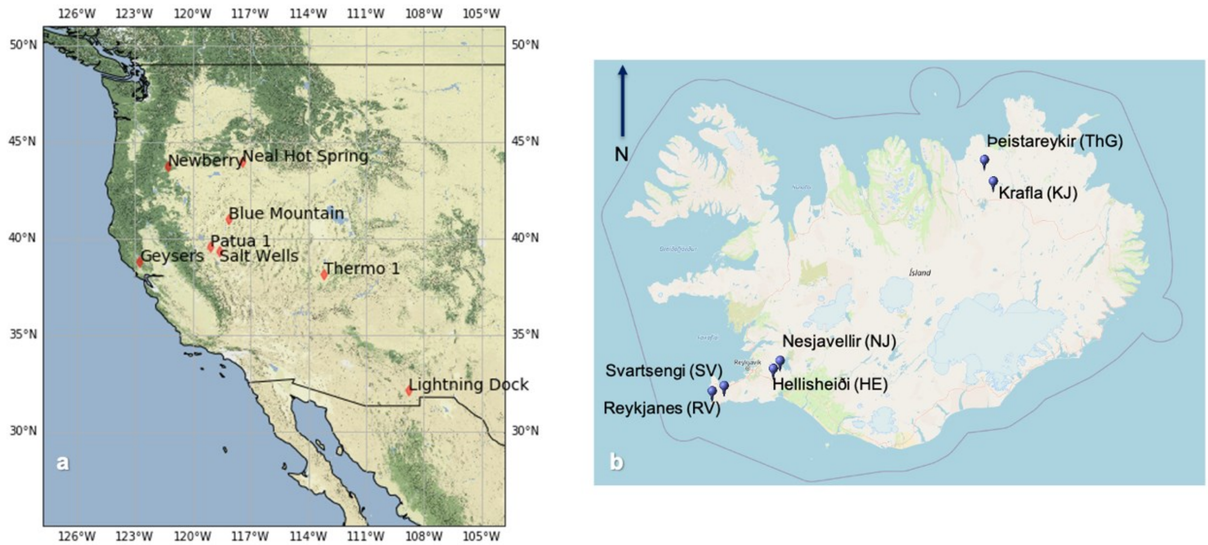

Figure 1: (left) Location of geothermal fields in US and (right) in Iceland that provided well data for the EDGE database and analysis. (Source: Carbonari, et al., 2021).

EDGE Consortium teams at OSU, Stanford University, and UTA carried out exploratory data analysis, devised methods to detect and correct erroneous and/or missing data, and developed ML models to predict: (a) Rate of Penetration (ROP), (b) Non-productive Time (NPT), (c) the occurrence of 'Problems', and (d) drilling costs. Data analytics and modeling approaches used/developed included: (a) Self-Organizing Maps (SOM), (b) Random Forest Classifier (RFC), (c) Natural Language Processing (NLP) of driller's comments, (d) Probabilistic models, and (e) Process mining.

2. DATA MANAGEMENT

A FAIR data repository has been implemented that includes both raw data and curated data to support all data analytics and modeling efforts. A web application was created, based on CKAN (Winn, 2013) with identified standards for model persistence and sharing (ONNX, PMML, and PFA) and deployed as a virtual machine instance within PNNL's public cloud with access control to preserve proprietary information. To ensure compatibility and interoperability with existing geothermal data repositories each dataset includes basic metadata descriptors that provide information on the content and purpose of the files, enabling sufficient public-facing information for data discovery that both preserves confidential information but provides those interested in gaining access to those data information on the data provider.

This is an important and, perhaps, easily misunderstood feature of the database. While the principal of completely open access data promotes development of new analysis and interpretation methods by different groups unhampered by data access restrictions, and independent validation of results by third parties, the reality of the geothermal industry is that much of the available well data is proprietary and considered commercially sensitive. This condition has hampered the release of data for analyses such as described here.

By assuring that the proprietary nature of much of the geothermal well data set is respected, it becomes possible to use those data as well as open access data as part of a much larger training set without revealing commercially sensitive information to end-users of the expert system. Upon weighing the benefits and costs of restricting our analysis to open data sets vs. potentially a much larger training set, the EDGE Perfect Geothermal Well Consortium determined that the advantages of allowing for inclusion of proprietary well data outweighed the other considerations. Furthermore, the database was designed at the outset to be expandable organically and supported by a toolset to ease the process of ingesting heterogeneous and data sets containing errors and omissions, thus allowing for future expansion with both open access and proprietary well data. The validation of the expert system by geothermal drilling operators therefore rests on real-world cost reductions they might achieve.

The process of data assimilation and curating was described previously by Carbonari, et al., 2021), but it is worth emphasizing the extensive preprocessing, exploratory data analysis, curation and data imputation is required when heterogeneous data sets from a variety of drilling operators, geothermal fields and countries are to be readied for analysis and curated for re-use. Considering the heterogeneity of existing data sets and to encourage harmonization of geothermal drilling data, a set of best practices for geothermal data has been developed by the EDGE Consortium. The relationship of the data management components of the project to the other components (process analysis and visualization, geothermal drilling models and well failure analysis, and drilling optimization is seen in Figure 2 below.

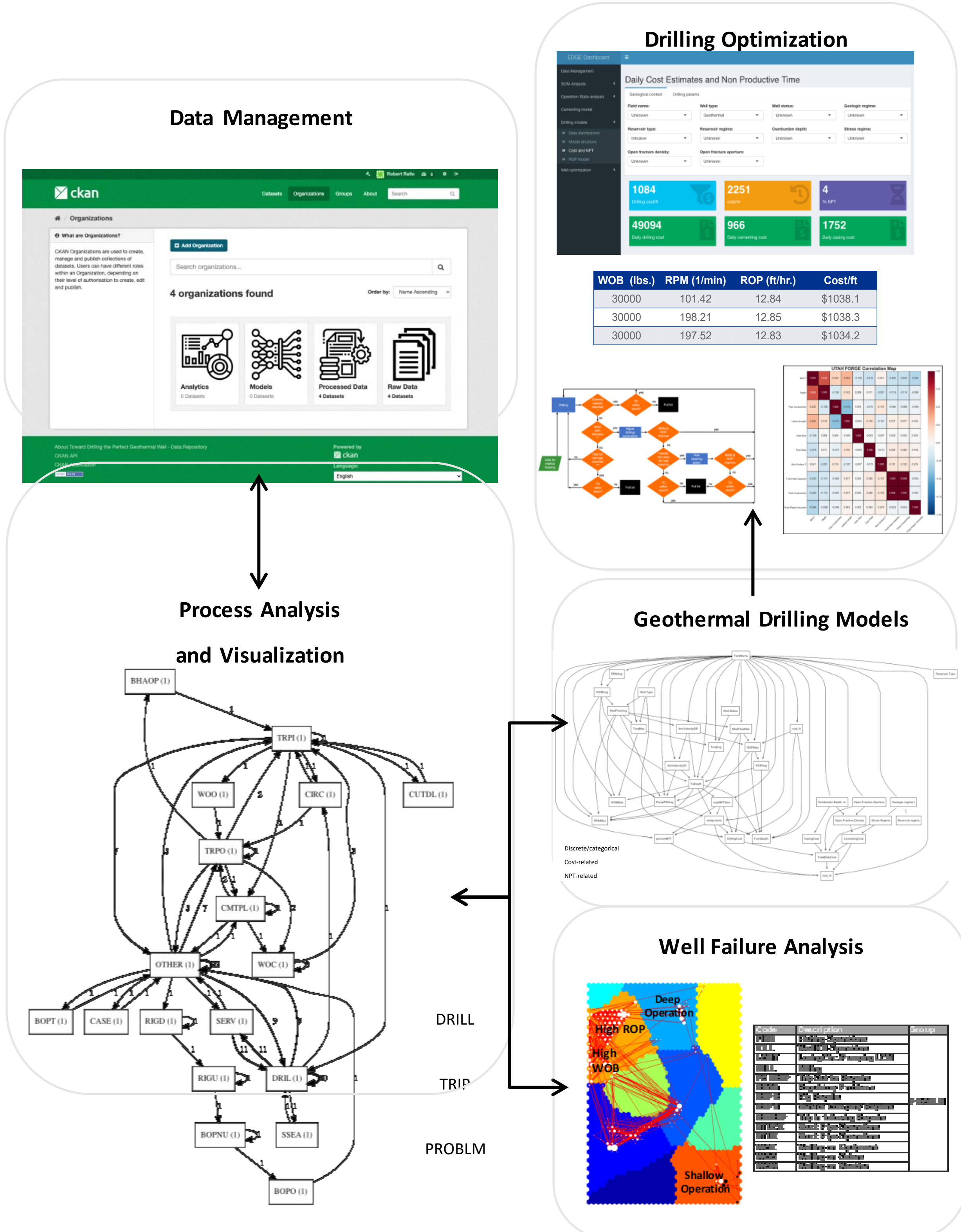

Figure 2: The elements of the EDGE Perfect Geothermal Well Research Coordination Network data flow process, with feedback loops between data ingestion/management, process analysis and visualization, training of geothermal drilling models and well failure analysis. The process allows for detection of erroneous and/or missing data, correction or deletion of such data, iterative refinement of the ML based models and well failure analysis (all within the hybrid open and secure proprietary data space of the well data), and then the use of the resulting tools for drilling optimization.

At present the EDGE well data repository contains contributions from AltaRock Energy, Cyrq Energy, ISOR and open access data from the Geothermal Data Repository (GDR). These data have been assembled, preprocessed, and cleaned, resulting in usable data from 113 wells located in the US and Iceland. A lexicon for metadata to characterize geological context and reservoir characteristics was defined as part of this process. A high-level summary appears in Table I below.

Table I Geological Context and Reservoir Characteristics Vocabulary

Geologic regime	Reservoir Type	Overburden Depth	Stress Regime	Reservoir regime	Open Fracture Density	Open Fracture Aperture
shield	metavolcanic	shallow	compressional	hot water	high	high
platform	metasedimentary	medium	extensional	wet steam	medium	medium
orogen	intrusive	deep	strike-slip	dry steam	low	low
basin	subvolcanic			superheated		
LIP	sedimentary					
extended crust						

3. DATA ANALYTICS

Data models were developed to use the data records in the EDGE well data repository to predict: (a) Rate of Penetration (ROP), (b) Non-productive Time (NPT), (c) the occurrence of 'Problems', and (d) drilling costs. The definition of problems and failures was based on the collective experience and discussions of the drilling experts from both geothermal and oil & gas industries who are part of the project team, as well as other well-known experts in the geothermal industry. Data analytics and modeling approaches used/developed included: (a) Self-Organizing Maps (SOM), (b) Random Forest Classifier (RFC), (c) Natural Language Processing (NLP) of driller's English language comments, (d) Probabilistic models, and (e) Process mining. Examples are shown in Figure 4 below.

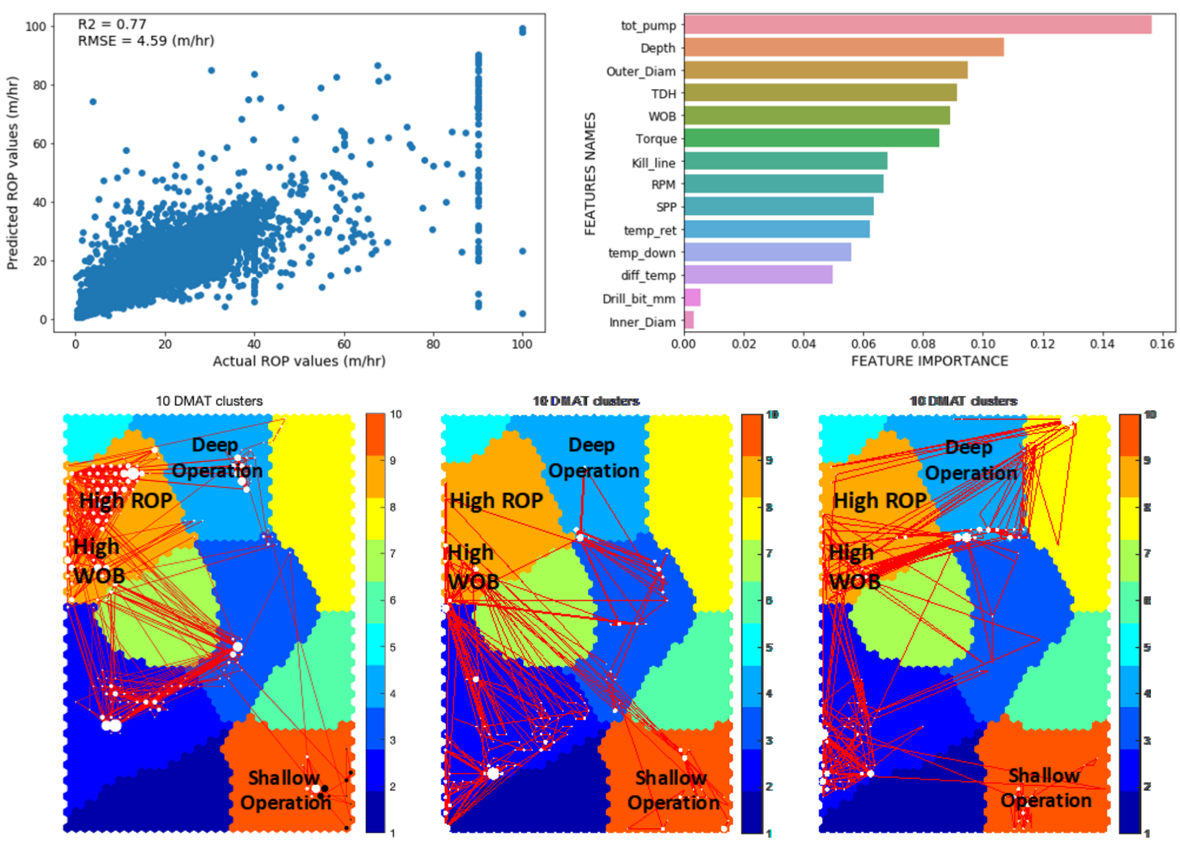

Figure 3: Output of machine learning model vs. observations in the database for Rate of Penetration. (Top, left) ROP prediction with Random Forest cross plot of predicted vs. actual ROP values. (Top, right) Feature importance ranking plot. (Bottom) Operation trajectories of three wells projected on top of the SOM clustering.

Challenges were identified in assimilating and analyzing these data. It is necessary to embed uncertainty in the models, and relevant features in the data for drilling cost optimization appear as free-form text comments in the data that are often ad hoc and not clearly following a consistent lexicon. This required extensive data extraction and categorization. Despite these challenges, conclusions can be drawn. Perhaps most obvious among these is the daily rate of the drill rigs is a key cost driver – it is important to minimize time on rig. Time on rig is influenced by multiple factors including ROP, but also NPT, premature bit wear or trip outs caused by inadequately tracked bit wear history, delays caused by lack of replacement equipment, personnel issues, etc. Figure 4 shows the quasi-linear relationship between rig days on location and well costs.

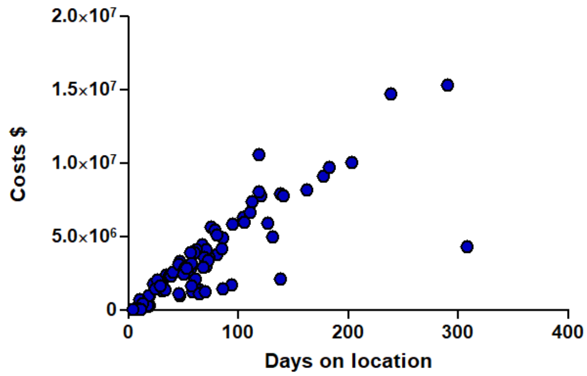

FIGURE 4: Well cost vs. rig days on location showing quasi linear relationships across all wells analyzed in the EDGE repository.

The drilling optimization system has learned to predict Rate of Penetration (ROP) by optimal choice of drilling parameters; to reduce Nonproductive time (NPT) based on a root-cause analysis and optimal choice of drilling parameters, and to predict the occurrence of upcoming adverse drilling events based on recent data in the well, including plain English comments from the rig crew. The optimization will be deployed as a collection of program tools, and provides the capability for geothermal community members to add their own drilling data.

While ROP was not found to be the primary driver of well costs (although it is a significant contributor to time on rig), the most significant variables controlling the ROP were found to be the pumping rate, depth, and the diameter of the hole. Using these and other parameters, a model with a population correlation coefficient with respect to the well data of R^2=0.77 was achieved. These results were obtained analyzing wells primarily drilled in igneous basement rock. While a promising result, more work is needed to determine how to best incorporate lithological complexity seen in sedimentary basins. The EDGE Consortium particularly seeks to expand the database to incorporate a larger population of wells from basin settings.

The primary cause of well failures during geothermal drilling was identified to be downhole tool failure due to hard rock drilling and high temperatures. Tracking both bit wear and tooth wear metrics in real-time can potentially help avoid both premature and overdue trip outs. Its notable that no real-time bit pull criteria are currently in use. The temporal sampling rate of the well data set proved a limitation, since the highest resolution was one sample every ten seconds and higher resolution is desired (Ashok, et al., 2022a). A novel bit pull criterion was developed and the algorithm was demonstrated on historical data sets. With regard to downhole heat management, the state-of-the-art does not include real-time tools to estimate Bottom Hole Circulating Temperature (BHCT) as a function of depth, to the required accuracy. We have evaluated the impact of different drilling parameters on BHCT, exploring the complex relationship among them, and developed a fast-running ML model for BHCT estimation (Ashok, et al., 2022b).

4. COST MODEL

A robust Bayesian Network cost model has been designed that is applicable to a variety of geological settings and well designs using a framework suitable for a wide range of scenarios. High level analysis using this model shows that the primary driver of cost is rig time, followed by casing, drilling mud, and cement. Using the cost model, it is possible to quickly analyze the cost impact of changes in drilling operations, allowing for optimization drilling procedure. Data on well failure is also included in the database, which enables users to optimize well design and operations for reduced risk of failure in addition to cost. Figure 5 shows the total well costs originally planned vs. that achieved showing that cost overruns can be quite significant, and that the relative importance of factors driving up well costs varies with the depth of well.

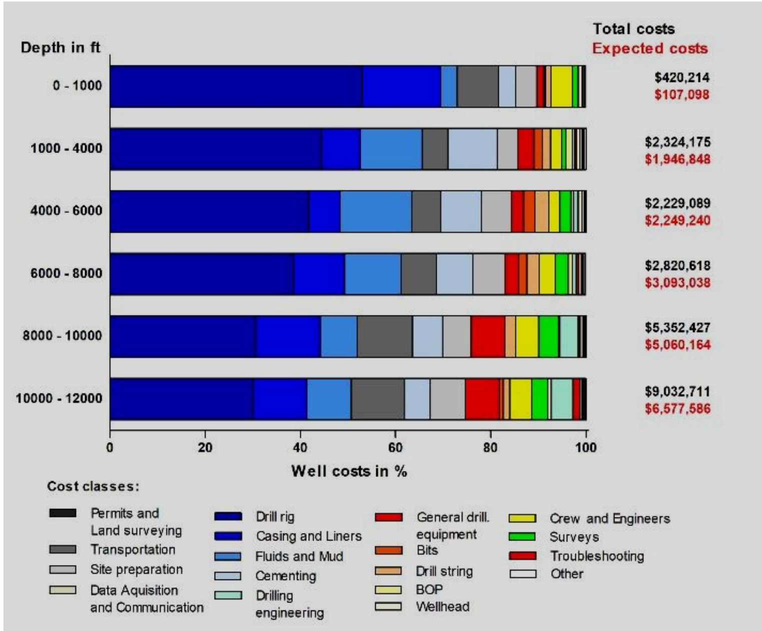

Figure 5: Overview of well costs by category sorted by total well depth. Parameters shown are the result of grouping over 600 cost codes into more general classes. The dominance of time-on-rig decreases with well depth as other factors such as transportation, drill equipment and drill string issues become more prominent.

5. SUSTAINABILITY THROUGH ORGANIC GROWTH AND CONTINUOUS TRAINING

Additional refinement of the existing models, and development of related analytical tools is a logical next step, a process that would be aided by expansion of the well database to cover an even wider range of geological settings and well configurations. We invite geothermal developers to become EDGE expert system users, and for those interested in tailoring the system to their specific requirements, by providing additional drilling and contextual data additional training of the expert system will improve the system organically. Such tools will help developers to streamline well design and operations with the aim or reducing costs significantly. A subscription system is being launched with a tiered structure to encourage users to become data providers at minimal cost.

6. CONCLUSIONS

There is significant potential for drilling cost reduction if the appropriate data are collected at the appropriate sample collection rate, with metadata including driller's notes (ideally using a harmonized lexicon) and if the appropriate data analyses (both in real-time and post-well) are conducted. The data required for drilling cost reduction analysis comes from a variety of sources. At present there is no reference standard available for data that is most useful to potentially reduce geothermal well construction costs. The EDGE Consortium proposes adoption of such standards, and a report that documents the key data for geothermal data drilling analysis is being finalized.

Through the EDGE Consortium efforts, a drilling optimization system has learned from its database of 113 wells to predict ROP by optimal choice of drilling parameters, to reduce NPT based on root-cause analysis and optimal choice of drilling parameters, and to predict the occurrent of adverse drilling events based on recent data in the well, including plain English language comments from the rig crew. The drilling optimization system integrates a collection of program tools that provide analysis, modeling, and optimization capabilities to the geothermal community and allows for the incorporation of their own drilling data in the training set to better optimize the system for their needs.

7. ACKNOWLEDGEMENT

This research has been conducted in the framework of the EDGE project supported by the Department of Energy under Award Number DE-EE0008793 – 0000. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for United States Government purposes.

REFERENCES

Alali, A. M., Abughaban, M. F., and Aman, B. M.: Hybrid Data Driven Drilling and Rate of Penetration Optimization, Journal of Petroleum Science and Engineering, 108075, (2020).

Ashok, P. et al.: Optimizing Rate of Penetration and Tripping Decision-Making using Real-Time Bit Wear Monitoring While Drilling Geothermal Wells, *Geothermal Rising Conference*, Reno, NV (2022a).

Ashok, P., et al.: Predicting Bottomhole Circulating Temperature while Drilling High Enthalpy Geothermal Wells Using Machine Learning Models, *Geothermal Rising Conference*, Reno, NV (2022b).

Basarir, H., Tutluoglu, L., and Karpuz, C.: Penetration rate prediction for diamond bit drilling by adaptive neuro-fuzzy inference system and multiple regressions, *Engineering Geology*, 173, 1-9, (2014).

Carbonari, R., D. Ton, A. Bonneville, D. Bour, T. Cladouhos, G. Garrison, R. Horne, S. Petty, R. Rallo, A. Schultz, C. Sørlie, Ingolfur Orn Thorbjornnson, M. Uddenburg, L., Weydt: First year report of EDGE project: An International Research Coordination Network for Geothermal Drilling Optimization Supported by Deep Machine Learning and Cloud Based Data Aggregation, *PROCEEDINGS*, *46th Workshop on Geothermal Reservoir Engineering*, Stanford University, Stanford, California, February 15-17, 2021

Dumas, P., Antics, M., and Ungemach, P.: Report on geothermal drilling. *Co-Funded by the Intelligent Energy Europe Programme of the European Union*, (2013).

Finger, J., and Blankenship, D.: Handbook of best practices for geothermal drilling, *Sandia National Laboratories*, Albuquerque, NM (2010).

Imaizumi, H., Ohno, T., Karasawa, H., Miyazaki, K., Akhmadi, E., Yano, M., Miyashita, Y., Yamada, N., Miyamoto, T., Tsuzuki, M., Kubo, S., and Hishi, Y.: Drilling Performance of PDC bits for Geothermal Well Development in Field Experiments, *Proceedings*, 44th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2019).

Karimi, M., Moellendick, T. E., and Holt, C.: Plastering effect of casing drilling; a qualitative analysis of pipe size contribution, *SPE Annual Technical Conference and Exhibition*, Society of Petroleum Engineers, (2011).

Miyazaki, K., Ohno, T., Karasawa, H., and Imaizumi, H.: Performance of polycrystalline diamond compact bit based on laboratory tests assuming geothermal well drilling, *Geothermics*, 80, 185-194., (2019).

Raymond, D. W., Knudsen, S. D., Blankenship, D. A., Bjomstad, S., Barbour, J., Drilling, B., Schen, A., and Downhole, N. O. V.: PDC Bits Outperform Conventional Bit in Geothermal Drilling Project (No. SAND2012-7841C), *Sandia National Lab(SNL-NM)*, Albuquerque, NM (2012).

Sabah, M., Talebkeikhah, M., Wood, D. A., Khosravanian, R., Anemangely, M., and Younesi, A.: A machine learning approach to predict drilling rate using petrophysical and mud logging data, *Earth Science Informatics*, 12(3), 319-339, (2019).

Saleh, F. K., Teodoriu, C., Ezeakacha, C. P., and Salehi, S.: Geothermal Drilling: A Review of Drilling Challenges with Mud Design and Lost Circulation Problem, *Proceedings*, 45th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2020).

Salehi, S., Mgboji, J., Aladasani, A., and Wang, S.: Numerical and Analytical Investigation of Smear Effect in Casing Drilling Technology: Implications for Enhancing Wellbore Integrity and Hole Cleaning, *SPE/IADC Drilling Conference*, Society of Petroleum Engineers, (2013).

Winn, J.: Open data and the academy: an evaluation of CKAN for research data management, (2013).