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

A Progress Report on GeoThermalCloud Framework: An Open-Source Machine Learning Based Tool for Discovery, Exploration, and Development of Hidden Geothermal Resources

Bulbul Ahmmed[1], Maruti K. Mudunuru[2], Luke Frash[1], and Velimir V. Vesselenov[3]

[1]Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545. [2]Watershed and Ecosystem Science, Pacific Northwest National Laboratory, Richland, WA 99352. [3]EnviTrace LLC. Santa Fe, NM 87501

Email: ahmmedb@lanl.gov

Keywords: *Play fairway analysis, unsupervised machine learning, NMFk, prospectivity analysis*

ABSTRACT

GeoThermalCloud is a Department of Energy's Geothermal Technologies Office funded project to develop an open-source tool (https://github.com/SmartTensors/GeoThermalCloud.jl) to discover hidden geothermal resources using machine learning and prospecting enhanced geothermal systems (EGS). We named the geothermal resources exploration component ***GeoThermalCloud-RE*** while the EGS prospecting component ***GeoThermalCloud-EGS***. ***GeoThermalCloud-RE*** utilizes unsupervised machine learning (ML) to automate data analyses and interpretations by extracting hidden signatures to elucidate geothermal prospects. Also, it enables the identification of critical measurements needed to identify geothermal resource signatures. ***GeoThermalCloud-RE*** can be applied to (1) analyze large sparse field datasets, (2) assimilate model simulations, (3) perform transfer learning (between sites with different exploratory levels), (4) label geothermal data types, resources, and processes, (5) identify high-value data acquisition targets, and (6) guide geothermal exploration and production by selecting optimal exploration, production, and drilling strategies. ***GeoThermalCloud-EGS*** is a machine learning-based alternative to GeoDT, a fast, simplified multi-physics solver to evaluate EGS designs in uncertain geologic systems. This paper will briefly update our progress from the project's onset to the present.

1. INTRODUCTION

The project is motivated by the challenges, risks, and costs associated with geothermal exploration and production (V. V. Vesselinov et al., 2022). Many processes and parameters impacting geothermal conditions are poorly understood. Diverse datasets are available to help characterize subsurface geothermal conditions (public and proprietary; satellite, airborne surveys, vegetation/water sampling, geological, geophysical, etc.). Yet, it is unclear how to properly leverage these datasets for geothermal exploration due to an incomplete understanding of how physical processes impacting subsurface geothermal conditions are represented in these observations. Recent advancements in machine learning (ML) promise to resolve these issues (V. V Vesselinov et al., 2022).
The tremendous challenges and risks of geothermal exploration and production bring the demand for novel ML methods and tools that can (1) analyze large field datasets, (2) assimilate model simulations (large inputs and outputs), (3) process sparse datasets, (4) perform transfer learning (between sites with different exploratory levels), (5) extract hidden geothermal signatures in the field and simulation data, (6) label geothermal resources and processes, (7) identify high-value data acquisition targets, and (8) guide geothermal exploration and production by selecting optimal exploration, production, and drilling strategies (Mudunuru et al., 2022). Our goals and work under Phases 1 and 2 (as proposed) of this project address all these needs.

Under Phase I&II, we have developed ***GeoThermalCloud-RE*** and ***GeoThermalCloud-EGS***. ***GeoThermalCloud-RE*** is an unsupervised ML-based tool to discover and extract new (unknown/hidden) geothermal signatures in existing site, synthetic, and regional datasets. Our ML analyses also identified high-value data acquisition strategies that can reduce geothermal exploration/production costs and risks. Moreover, ***GeoThermalCloud-RE*** categorized geothermal data, which is applied to generate geothermal data labels (e.g., geothermal resource types). ***GeoThermalCloud-RE*** allows for the treatment of both public and proprietary datasets. This is an essential feature considering the high sensitivities associated with using proprietary data. Moreover, the ***GeoThermalCloud-RE*** framework includes a series of advanced pre-processing, post-processing, and visualization tools, which tremendously simplify its application for real-world problems. These tools make the ML results understandable and visible even for non-experts. Thus subject-matter expertise is not a critical requirement to use the ***GeoThermalCloud-RE*** framework.

GeoThermalCloud-EGS is an enhanced geothermal system (EGS) prospecting tool. It is an ML version of GeoDT (Frash et al., 2023; Frash, 2022, 2021; Mudunuru et al., 2023). GeoDT is very fast modeling tool to run thousands of realization tweaking reservoir, drilling, and geothermal plant parameters. The main mechanism is to use ***GeoThermalCloud-RE*** for geothermal resources exploration to find favorable geothermal locations and then use ***GeoThermalCloud-EGS*** for exploring EGS prospectivity. Figures 1 and 2 demonstrate the schematics of ***GeoThermalCloud-RE*** and ***GeoThermalCloud-EGS***, respectively.

We have used ***GeoThermalCloud-RE*** on ten geothermal datasets. Eight datasets include site/real data, including a large and sparse dataset of the Great Basin, and two datasets are synthetic data. The analyses found critical information that could not be found using supervised

ML or exploratory statistical analyses. Most of the data and analyses are available on GitHub as well. Obtained results can be reproduced and further expanded by adding additional data. Practitioners and researchers are welcome to utilize ***GeoThermalCloud-RE*** to solve other geothermal problems. ***GeoThermalCloud-EGS*** can be used for studying FORGE EGS prospectivity.

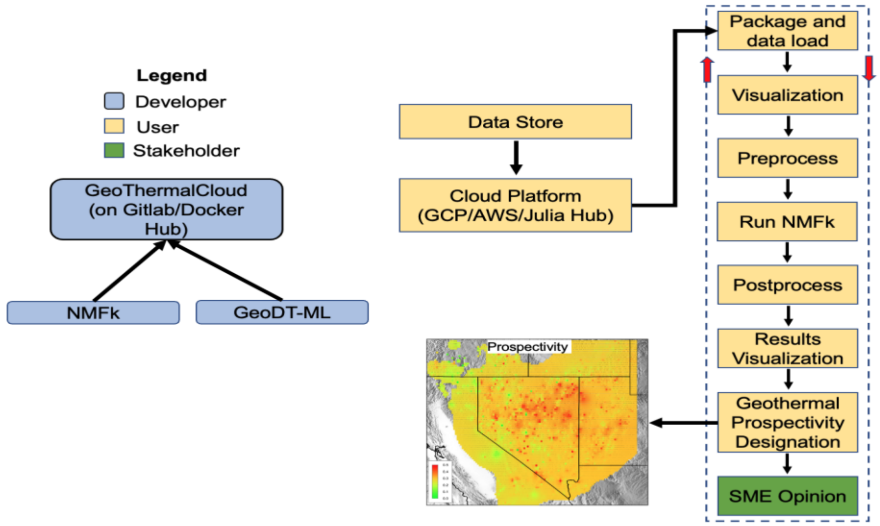

Figure 1: Schematic of *GeoThermalCloud-RE* for geothermal resources exploration.

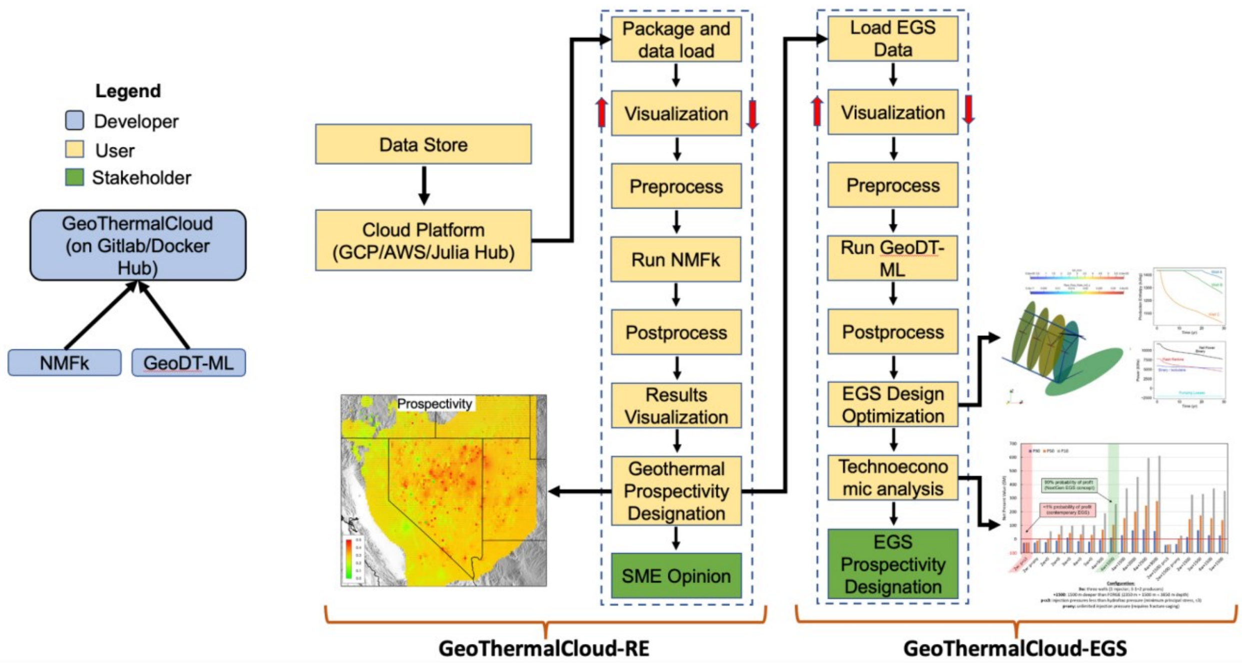

Figure 2: Schematic of *GeoThermalCloud-RE* for geothermal resources exploration and *GeoThermalCloud-EGS* for EGS design and prospectivity analysis.

2. GEOTHERMALCLOUD CAPABILITIES

2.1 GeoThermalCloud-RE

GeoThermalCloud capabilities include (1) analyzing large field datasets, (2) assimilating model simulations (large inputs and outputs), (3) processing sparse datasets, (4) performing transfer learning (between sites with different exploratory levels), (5) extracting hidden geothermal signatures in the field and simulation data, (6) labeling geothermal resources and processes, (7) identifying high-value data acquisition targets, and (8) guiding geothermal exploration and production by selecting optimal exploration, production, and drilling strategies. The GeoThermalCloud is an open-source tool available at https://github.com/SmartTensors/GeoThermalCloud.jl (a part of our SmartTensors framework; http://tensors.lanl.gov, https://github.com/SmartTensors) (Mudunuru et al., 2022; V. V Vesselinov et al., 2022).

2.2 GeoThermalCloud-EGS

GeoThermalCloud-EGS is an ML-based version of GeoDT, which is a fast, simplified multi-physics solver to evaluate EGS designs in uncertain geologic systems (Frash et al., 2023; Frash, 2022, 2021; Mudunuru et al., 2023). It is numerically efficient enough to model thousands of realizations in a few hours using a desktop computer. The underlying assumptions of this model are empirically based on laboratory and field data to partially account for complex coupled processes obviating running expensive numerical simulations (Frash, 2021). The intent of this model is to run it with full uncertainty, as informed by a broad spectrum of relevant prior laboratory and field measurements, and to reduce the uncertainty only when suitable information is available. When a promising EGS design is identified, it can be investigated in greater detail and at higher fidelity using other more powerful, but more expensive, numerical modeling codes. The primary features of ***GeoThermalCloud-EGS*** include (Figure 3):

1. Pressure and flow rate prediction for 3D networks of intersecting wells and fractures modeled as pipes and nodes.
2. Hydraulic stimulation prediction with shear and tensile mechanisms where fracture apertures depend on effective stress.
3. Transient heat production predictions that depend on fluid enthalpy, rock conductivity, and stored energy change over time.
4. Electrical power generation using the combined single-flash Rankine and isobutane binary cycle.
5. Net present value prediction based on geothermal cost estimation tools, electricity sales, and a simple earthquake cost model.

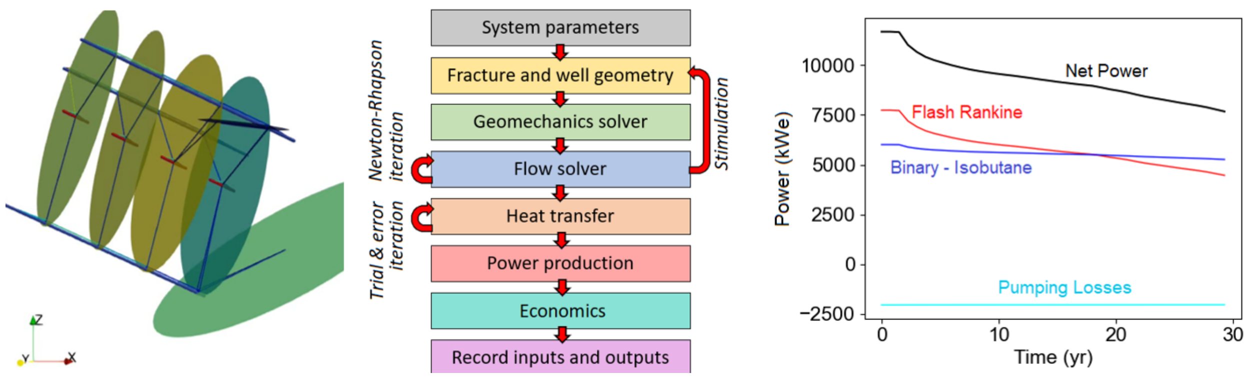

Figure 3: GeoDT or *GeoThermalCloud-EGS* stochastically predicts reservoir parameters, flow networks, hydraulic stimulation, heat production, power production, injection-induced seismicity potential, and ultimately net present value by fast and simplified methods. Most models complete in ~15 seconds using a common desktop computer with a single processor thread.

3. GEOTHERMALCLOUD ANALYSES

3.1 GeoThermalCloud-RE

ML methods embedded in the **GeoThermalCloud** have been extensively tested and validated against various datasets (Figure 4) (Mudunuru et al., 2022; V. V Vesselinov et al., 2022). Outputs of these applications have been published in presentations, conference papers, and peer-reviewed papers. The analyzed ML applications are

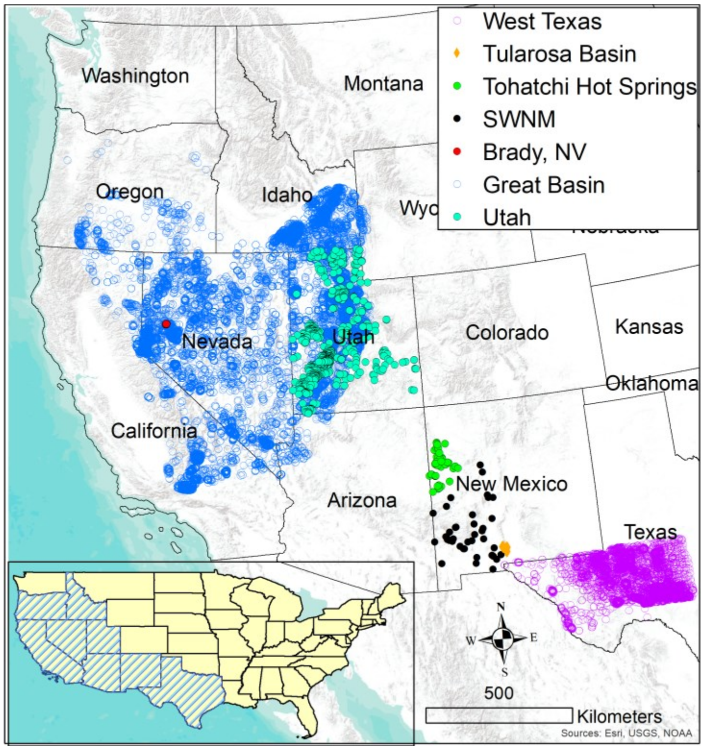

Figure 4: Locations of seven out of 8 analyzed site datasets by the GeoThermalCloud framework. The other site dataset is in Hawaii, not shown here.

1. **Great Basin**: In this dataset, we analyzed 18 shallow water chemistry attributes at 14,342 locations. This work extracted hidden geothermal signatures associated with low-, medium-, and high-temperature hydrothermal systems, their dominant characterization attributes, and spatial distribution within the study area (Ahmmed and Vesselinov, 2022). The analyses are based on the public data available on the Nevada Bureau of Mines and Geology website.
2. **Southwest New Mexico (SWNM)**: Here, we analyzed 18 attributes at 44 locations and identified low- and medium-temperature hydrothermal systems; found dominant attributes and spatial distribution of extracted hidden hydrothermal signatures; demonstrated blind predictions of the regional physiographic provinces (Vesselinov et al., 2020; V. V. Vesselinov et al., 2022).
3. **Brady site, Nevada**: We identified key geologic factors controlling geothermal production in the Brady geothermal field (Siler et al., 2021).
4. **Tularosa Basin, New Mexico**: Analyzed 21 Play Fairway Analysis (PFA) attributes at 120 locations (Ahmmed et al., 2022); data comes from past PFA work in this region (Bennett and Nash, 2017). ML analyses identified geothermal signatures associated with low-, medium-, and high-temperature hydrothermal systems. Dominant attributes and spatial distribution of the geothermal signatures were also defined.
5. **Tohatchi Springs, New Mexico**: Explored 19 geothermal attributes at 43 locations in Tohatchi Springs, New Mexico (Ahmmed et al., 2020b). Successfully defined geothermal signatures associated with low- and medium-temperature hydrothermal systems. Also, we found their dominant attributes and spatial distribution.
6. **Hawaii**: Analyzed four islands' data separately and jointly; ML identified low-, medium-, and high-temperature hydrothermal systems and their dominant characterization attributes (Ahmmed et al., 2020a).
7. **Utah FORGE**: Performed prospectivity analysis to identify future drilling locations using geological, geochemical, and geophysical attributes (Ahmmed and Vesselinov, 2021). Maps of temperature at depth and heat flow are constructed based on the available data. Processed data includes satellite (InSAR), geophysical (gravity, seismic), geochemical, and geothermal attributes. Prospectivity maps were generated, and drilling locations were proposed for future geothermal field exploration.
8. **EGS Collab:** Field experiment data processed to extract dominant temporal patterns observed in 49 data streams; erroneous measurement attributes and periods automatically identified; interrelated data streams automatically identified. This work has not been published yet.

Future Case Study: In the coming months, we will add recently released geothermal exploration data by Great Basin Center for Geothermal Energy (Ayling et al., 2022) to the geochemistry data set we analyzed in Ahmmed and Vesselinov (Ahmmed and Vesselinov, 2022; Nevada Bureau of Mines and Geology, 2012).

3.2 GeoThermalCloud-EGS

Multiple datasets have been generated using GeoDT for EGS Collab and Utah FORGE site. One was used for the PIVOT 2022 Datathon to simulate the whole geothermal development cycle from the initial well design to the end of production. This dataset includes the Utah FORGE site characteristics and its measured uncertainties. The database includes 44,492 unique realizations, each with at least 30 years of production. Based on site characteristics, fractures are stochastically created (Figure 5). Next, simulations are performed to compute power outputs for each situation.

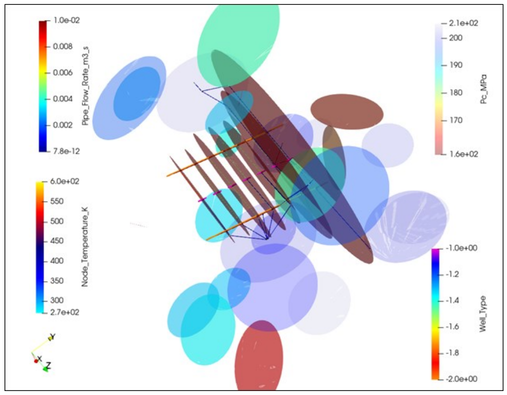

Figure 5: Example stochastically generated fracture and well scenario with injection into one well across seven isolated intervals and production from two bounding wells. The parallel hydraulic fractures propagated from each injection interval are shown in red, the color indicating that these fractures require relatively low pressure for activation (*Pc*). Note that most, but not all, of the scattered natural fractures require significantly higher pressures to activate.

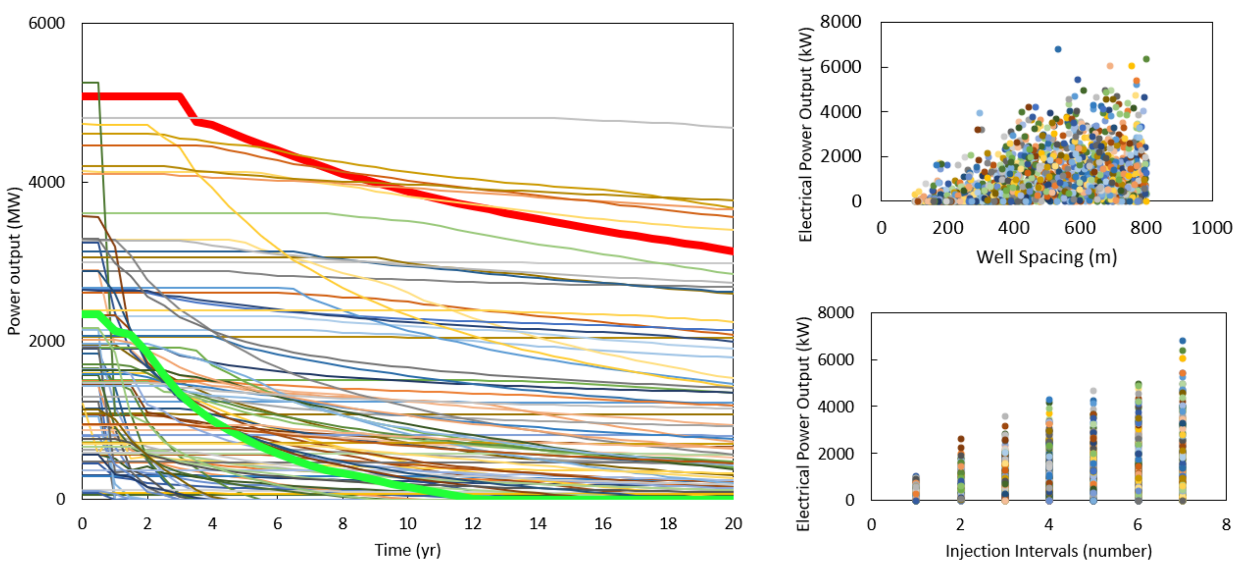

Figure 6: Geothermal power production simulations based on the parameters described in Table 3.8.1 in (Vesselinov, n.d.). In the time series plot, a high-performing case is highlighted in red, and a poor performer is highlighted in green. There is also a clear link between the well spacing and power output in addition to the number of injection intervals (i.e., isolated zones) and power output (plots on the right).

4. HOW TO USE IT?

GeoThermalCloud can be used in three ways (i) on Julia, (ii) on Python, and (iii) on a Cloud platform (e.g. JuliaHub, Google Cloud Platform, Amazon Web Services, Azure Cloud Services through Docker). Julia's installation is explained at https://github.com/SmartTensors/GeoThermalCloud.jl and also given below.

```
import                                                              Pkg
Pkg.add("GeoThermalCloud")
import                                                  GeoThermalCloud

GeoThermalCloud.SWNM()  #  performs  analyses  of  the  Southwest  New  Mexico  region
GeoThermalCloud.GreatBasin()  #  performs  analyses  of  the  Great  Basin  region
GeoThermalCloud.Brady() # performs analyses of the Brady site, Nevada.
```

The Python installation process is described below:

```
$                                                               python3
import                                                            julia
julia.install()
from                  julia                 import                 Base
from                  julia                 import                 Main
Main.eval("import Pkg; Pkg.build(\"GeoThermalCloud\")")
```

Docker container development is still in progress. We will provide an update on how to use GeoThermalCloud when it is ready for use.

5. CONCLUSIONS

GeoThermalCloud is an open-source cloud-based ML framework for geothermal exploration that can simultaneously handle both public and proprietary datasets. Also, it consists of a series of advanced pre-processing, post-processing, and visualization tools that tremendously simplify its application for real-world problems. These tools make the ML results understandable and visible even for non-experts; therefore, ML and subject-matter expertise are not critical requirements to use our ML framework. **GeoThermalCloud** utilizes a series of novel LANL-developed patented ML tools called *SmartTensors* (https://github.com/SmartTensors). *SmartTensors* has already been applied to solve a wide range of real-world problems, from COVID-19 to wildfires (http://tenosrs.lanl.gov), and it has won two 2021 R&D 100 awards, including a bronze award for market disruptor tools. Now, it has two components (i) GeoThermalCloud-RE and (ii) GeoThermalCloud-EGS.

GeoThermalCloud-RE is developed to process and analyze diverse small and large datasets. Also, it can handle sparse datasets with missing values. It analyzes and finds actionable information to enable decision-makers to make sound decisions for geothermal exploration, development, and production. It finds such actionable information by finding mapping functions between all input parameters. We analyzed eight diverse site datasets and found critical information that would not be possible by visual inspection or any other statistical tools. Overall, **GeoThermalCloud-RE** can (1) analyze large field datasets, (2) assimilate model simulations (large inputs and outputs), (3) process sparse datasets, (4) perform transfer learning (between sites with different exploratory levels), (5) extract hidden geothermal signatures in the field and simulation data, (6) label geothermal resources and processes, (7) identify high-value data acquisition targets, and (8) guide geothermal exploration and production by selecting optimal exploration, production, and drilling strategies.

GeoThermalCloud-EGS is an ML-based version of GeoDT, a fast, simplified multi-physics solver to evaluate EGS designs in uncertain geologic systems. It is numerically efficient enough to model thousands of realizations in a few hours using a desktop computer. It is designed to find prospective enhanced geothermal systems in hot, dry rocks.

ACKNOWLEDGMENT

This research is based upon work supported by the U.S. Department of Energy's (DOE) Office of Energy Efficiency and Renewable Energy (EERE) under the Geothermal Technology Office (GTO) Machine Learning (ML) for Geothermal Energy funding opportunity, Award Number DE-EE-3.1.8.1. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (Contract No. 89233218CNA000001). Additional information regarding the datasets and codes can be obtained from Bulbul Ahmmed (ahmmedb@lanl.gov).

This paper was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

REFERENCES

Ahmmed, B., Lautze, N., Vesselinov, V.V., Dores, D., Mudunuru, M.K. Unsupervised machine learning to extract dominant geothermal attributes in Hawaii Island Play Fairway data. Geothermal Resources Council, Online, (2020a).

Ahmmed, B., Vesselinov, V.V. Machine learning and shallow groundwater chemistry to identify geothermal prospects in the Great Basin, USA. Renew. Energy 197, 1034–1048. https://doi.org/10.1016/j.renene.2022.08.024, (2022).

Ahmmed, B., Vesselinov, V.V. Prospectivity Analyses of the Utah FORGE Site using Unsupervised Machine Learning, in: Geothermal Rising Conference, San Diego, CA, (2021).

Ahmmed, B., Vesselinov, V.V., Middleton, R.S. Geothermal resource analysis at Tohatchi hot springs, New Mexico (No. LA-UR-21-23827). Los Alamos National Laboratory, Online, (2020b).

Ahmmed, B., Vesselinov, V.V., Rau, E., Mudunuru, M., Karra, S. Machine learning and a process model to better characterize hidden geothermal resources, (2022).

Ayling, B., Faulds, J.E., Rivera, A.M. INGENIOUS - Great Basin Regional Dataset Compilation (Data No. 1881483).

Bennett, C.R., Nash, G.D., 2017. The Convergence of Heat, Groundwater & Fracture Permeability: Innovative Play Fairway Modelling Applied to the Tularosa Basin. Ruby Mountain Inc. and Energy & Geoscience Institute, Salt Lake City, UT, (2022).

Frash, L., Carey, J.W., Ahmmed, B., Meng, M., Sweeney, M., K.C., B., Iyare, U. A Proposal for Profitable Enhanced Geothermal Systems in Hot Dry Rock, in: 48th Stanford Geothermal Workshop. Presented at the Stanford Geothermal Workshop, Stanford, CA, (2023).

Frash, L.P. Optimized Enhanced Geothermal Development Strategies with GeoDT and Fracture Caging, in: Proceedings of 47th Workshop on Geothermal Reservoir Engineering. Presented at the 47th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, (2022).

Frash, L.P. Geothermal Design Tool (GeoDT): Proceedings of the 46th Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, February 15-17, (2021).

Mudunuru, M.K., Ahmmed, B., Frash, L.P. Deep Learning for Modeling Enhanced Geothermal Systems, in: PROCEEDINGS of 48th Workshop on Geothermal Reservoir Engineering. Presented at the Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, (2023).

Mudunuru, M.K., Vesselinov, V.V., Ahmmed, B. GEOTHERMALCLOUD: MACHINE LEARNING FOR GEOTHERMAL RESOURCE EXPLORATION. J. Mach. Learn. Model. Comput. 3. https://doi.org/10.1615/JMachLearnModelComput.2022046445, Nevada Bureau of Mines and Geology, 2012. Great Basin Groundwater Geochemical Database, (2012).

Siler, D.L., Pepin, J.D., Vesselinov, V.V., Mudunuru, M.K., Ahmmed, B. Machine learning to identify geologic factors associated with production in geothermal fields: A case study using 3D geologic data, Brady geothermal field, Nevada. Geotherm. Energy, (2021).

Vesselinov, V.V., GeoThermalCloud: Cloud Fusion of Big Data and Multi-Physics Models using Machine Learning for Discovery, Exploration and Development of Hidden Geothermal Resources (No. LA-UR-20-30264). Los Alamos National Laboratory, (2021).

Vesselinov, V. V, Ahmmed, B., Frash, L.P., Mudunuru, M.K. GeoThermalCloud: Cloud Fusion of Big Data and Multi-Physics Models using Machine Learning for Discovery, Exploration and Development of Hidden Geothermal Resources, in: Proceedings of 47th Stanford Geothermal Workshop. Presented at the 47th Stanford Geothermal workshop, 47th Stanford Geothermal workshop, p. 11, (2021).

Vesselinov, V. V., Ahmmed, B., Mudunuru, M.K., Pepin, J.D., Burns, E.R., Siler, D.L., Karra, S., Middleton, R.S., 2022. Discovering hidden geothermal signatures using non-negative matrix factorization with customized k-means clustering. Geothermics 106, 102576. https://doi.org/10.1016/j.geothermics.2022.102576, (2022).

Vesselinov, V.V., Mudunuru, M.K., Ahmmed, B., S, K., Middleton, R.S., 2020. Discovering signatures of hidden geothermal resources based on unsupervised learning, in: Proceedings of the 45th Annual Stanford Geothermal Workshop, (2020).