GeoThermalCloud for EGS – An Open-source, User-friendly, Scalable AI Workflow for Modeling Enhanced Geothermal Systems

Maruti K. Mudunuru, Bulbul Ahmmed, and Luke Frash

[1]Environmental Subsurface Science Group, Pacific Northwest National Laboratory, Richland, WA 99352

[2]Earth & Environmental Science Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Keywords

Enhanced Geothermal Systems, Geothermal Design Toolkit, Deep learning, Sensitivity Analysis, Techno-Economics

ABSTRACT

Enhanced Geothermal Systems (EGS) offer vast potential to expand the use of geothermal energy. Heat is extracted from engineered systems by injecting relatively cold water into subsurface fractures, which are in contact with hot dry rock, and brought back to surface through production wells. Creating an EGS requires improving the natural permeability of hot crystalline rocks. In this short conference paper, we present a reproducible workflow for modeling EGS. Our workflow, called the GeoThermalCloud (GTC) for EGS, leverages recent advances in machine learning, deep learning, and high-performance computing. This GTC framework is currently being made open-source, user-friendly, and reproducible through Python scripts as well as Google Colab/Jupyter Notebooks. The GTC for EGS modeling scripts are made available at https://github.com/SmartTensors/GeoThermalCloud.jl/tree/master/EGS and will be updated to serve the geothermal community. Current GTC framework provides scripts to train deep learning (DL) models for techno-economics and data worth analysis. The Geothermal Design Tool (https://github.com/GeoDesignTool/GeoDT.git), a fast and simplified multi-physics solver, is used to develop a database for training DL models. This paper provides details on the scripts to curate, process, and train DL models. The scripts can easily be modified to train on databases generated by other popular open-source simulators such as PFLOTRAN, STOMP, TOUGH, and GEOSX or commercial software such as ResFrac and COMSOL.

1. Introduction

Enhanced Geothermal Systems (EGS) are engineered geothermal systems, which offer potential for dramatically expanding the use of geothermal energy (Brown et al., 2012). In this engineered system, cold water is injected into hot dry rock and is allowed to flow through a fracture network. The resulting hot fluid is extracted from production wells to generate electricity. The U.S.

Department of Energy's GeoVision (2019) estimates that more than 100GWe of economically viable power capacity could be extracted from the southwestern basin and range (GeoVison, 2019 DOE-MYPP 2022, EarthShot Initiative, 2022). However, high upfront costs and long development timelines generally characterize geothermal resource development projects (Hamm et al., 2021). This can lead to lengthy investment payback periods relative to many other utility-scale power generation projects (e.g., wind, solar). Moreover, projects employing new EGS designs and stimulation technologies to harness this renewable resource and produce usable power can have higher risks (Becker et al., 2018). To overcome this challenge of reducing costs and improving economics for geothermal projects, we need to understand feasible and non-feasible EGS designs better. Specifically, a detailed techno-economic analysis is required to successfully expand and accelerate EGS deployment in the western U.S. (DOE-MYPP, 2022; Sec-2). A workflow that combines geothermal data, multi-physics process models, and economics to assess good and bad EGS design parameters will allow us to overcome such a challenge (Sec-2.4 and Sec-2.5 in DOE-GTO MYPP, 2022). Recent deep learning (DL) advances have shown promise in developing such a workflow (Okoroafor et al., 2022). Here, we provide a scalable methodology (laptop to high-performance computing resources) to curate and analyze EGS datasets. An initial development of this scalable methodology, GeoThermalCloud (GTC), is available at https://github.com/SmartTensors/GeoThermalCloud.jl/tree/master/EGS.

2. GeoThermalCloud (GTC) for EGS workflow

In this section, we describe the workflow scripts for GTC for EGS techno-economic analysis. First, GTC for resource exploration is performed to estimate geothermal energy potential. Then, GTC for EGS allows us to evaluate and rank the prospectivity of a site and perform techno-economics for resource development. Fig. 1 describes the entire GTC workflow for exploration and EGS development. The GTC for exploration can be found in previous publications (e.g., Mudunuru, M.K et al., 2022). The GTC techno-economic analysis for EGS is the novelty of this work. The Python scripts for the workflow development are available at https://github.com/SmartTensors/GeoThermalCloud.jl/tree/master/EGS/GeoDT_ML_v1/Python_Scripts. Equivalent Jupyter Notebooks and Google Colab notebooks will be made available in future at this GTC GitHub location.

2.1 Data processing and curation

The GeoDT code (https://github.com/GeoDesignTool/GeoDT.git**)** is used to generate the training database (Frash, 2021; Frash, 2022; Frash et al., 2022). The data for DL modeling is available at https://github.com/SmartTensors/GeoThermalCloud.jl/tree/master/EGS/GeoDT_ML_v1/Data. In our study, a total of 4078 realizations are generated. The Python scripts *get_inp_out.py* and *get_preprocessed_data.py* are used to process the raw data and curate it with various pre-processing methods such as StandardScaler, MinMaxScaler, MaxAbsScaler, RobustScaler, PowerTransformer (Yeo-Johnson), QuantileTransformer (uniform output), and QuantileTransformer (Gaussian output). The Python script *get_train_val_test_splits.py* allows us to split the curated data into 80% training, 10% validation, and 10% testing. When the DL model identifies a promising EGS design, it can then be investigated in greater detail. For example, we can use high-fidelity process models and simulation codes such as PFLOTRAN (Lichtner et al., 2015) to explore promising EGS scenarios. This current study does not include the use of high-fidelity codes, but these Python scripts can be leveraged and modified to perform such DL analysis with minimal effort.

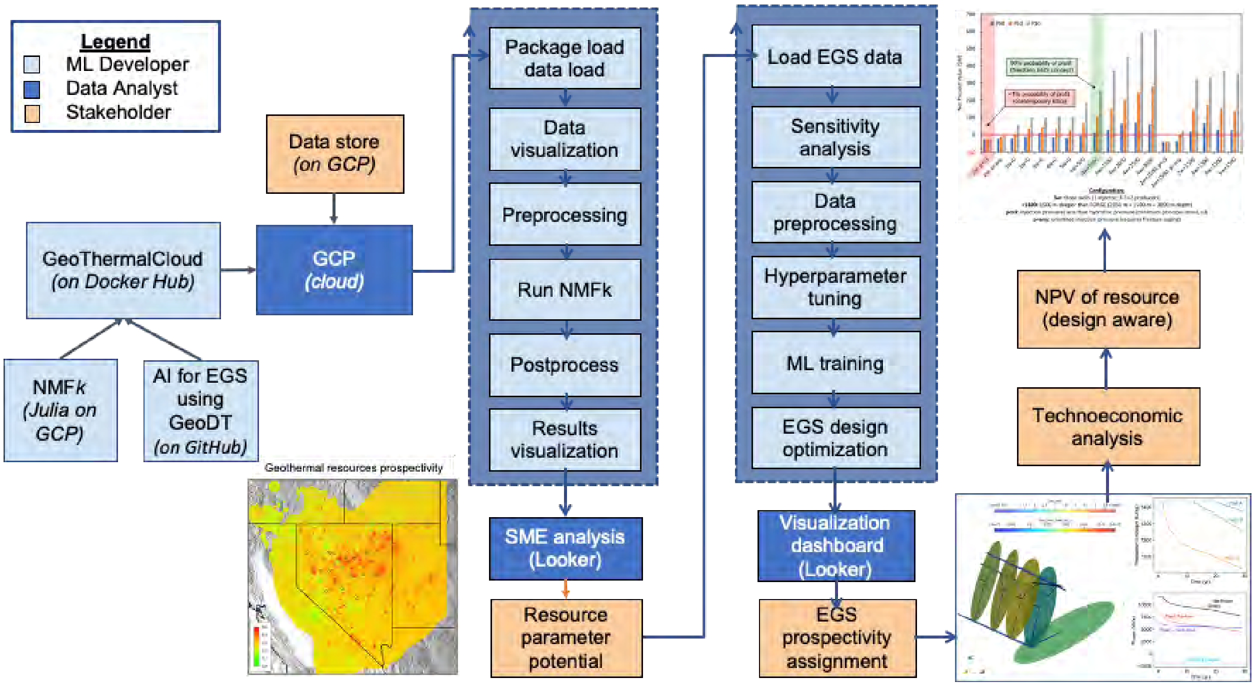

Figure 1: This figure describes the GTC framework and its two components – exploration and EGS development. The exploration component analyzes and curates play fairway analysis datasets to find the resource potential within a region. The resource component builds on these potential maps and assesses the EGS prospectivity to find and rank the most promising sites for further analysis.

2.2 Local and global sensitivity analysis

The data worth analysis is performed using the *get_ftest_mi_npv.py* and *get_ftest_mi_npv_others.py* scripts. These Python scripts allow us to perform local and global data worth analysis. Sensitivity analysis is performed using two different approaches, F-test and mutual information (MI). F-test is a univariate linear regression tests returning F-statistic and p-values. It provides insights on the linear dependency of a given EGS design parameter with respect to economics (e.g., undiscounted cashflow), thereby allowing us to identify potentially predictive design parameters for DL model training for undiscounted cashflow. On the other hand, mutual information provides insights on non-linear dependency between EGS design parameters and undiscounted cashflow. The MI between an EGS design parameter and undiscounted cashflow is a non-negative value and is equal to zero if and only if two variables are independent, and higher values mean higher non-linear dependency.

2.3 DL model training and hyperparameter tuning

This curated data is given as input to deep neural networks, which are trained on multiple cores available on high-performance computing machines (HPC). This AI training at scale is performed in parallel, allowing us to train and tune various deep neural networks in minimal time. We combine Python and AI modules such as mpi4py, multiprocessing, parallel hdf5, and TensorFlow to achieve this training at scale. The performance of the trained DL models is compared using the validation loss, and a tuned model is then selected. This hyperparameter tuning is computationally intensive and requires a lot of HPC resources. Python scripts such as *get_dir_hp_dnn_*.py* and *get_dnn_results_*.py* are available to achieve this. They provide specifics on how to run on MacOSX, Linuc, and HPC resources. In our case, we trained these models on a HPC resource at PNNL using 20,000 CPU cores. Fig. 2 shows a plot of one such DL model training and inference.

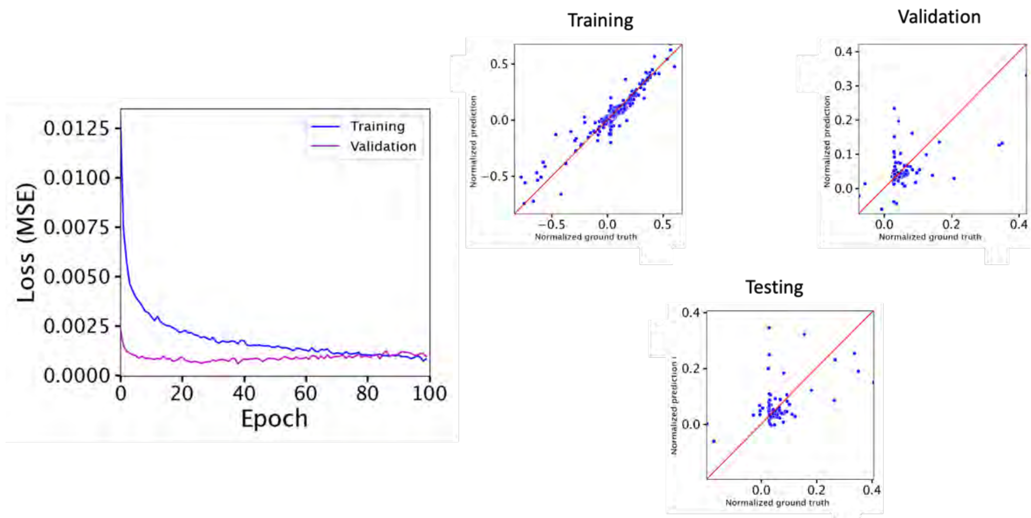

Figure 2: This figure provides a preliminary DL model training loss and one-to-one plots for training, validation, and test datasets. More than 20,000 DL models are training on HPC resources to estimate the EGS economics. This trained deep neural network model has three hidden layers, with neurons = [1000, 500, 250] in each of these layers. Leaky ReLU is used as an activation function with alpha value = 0.1. The dropout value, which allows for minimizing over-fitting during the training process, is assigned a value of 0.1. The total number of epochs for training is equal to 100. Batch size, which is the number of training samples that a DL model sees for each iteration in an epoch is equal to 64. The resulting DNN has approximately 750K trainable weights.

3. Conclusions and next steps

In this study, we developed and provided preliminary DL workflow scripts to estimate EGS economics from design parameters. The database for DL model training is developed using GeoDT, a multi-physics solver. Sensitivity analysis using F-test and mutual information is performed on this database to gain insights into the GeoDT parameters. The DL model training requires HPC resources as training and hyperparameter tuning is computational expensive. To overcome this challenge, we will also provide notebooks and pre-trained ML models in the GitHub for the geothermal community. Advanced hyperparameter tuning scripts using open-source softwares such as DeepHyper and Keras-Tuner will also be made available at https://github.com/SmartTensors/GeoThermalCloud.jl/tree/master/EGS.

Acknowledgement and Disclaimer

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).

Disclaimer

The training database using GeoDT employs very-high estimates of seismic risk and an unproven new muti-physics model. The actual FORGE project uses lower injection rates, lower injection volumes, closer well spacing, shallower depths, and seismicity mitigation measures that our model needs to include. In short, this work is not a prediction for the Utah FORGE project. 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. and Vesselinov, V.V., 2022. Machine learning and shallow groundwater chemistry to identify geothermal prospects in the Great Basin, USA. Renewable Energy, 197, pp.1034-1048.

Beckers, K.J. and McCabe, K., 2018. Introducing GEOPHIRES v2. 0: updated geothermal techno-economic simulation tool (No. NREL/CP-5500-70856). National Renewable Energy Lab.(NREL), Golden, CO (United States).

Brown, D.W., Duchane, D.V., Heiken, G. and Hriscu, V.T., 2012. Mining the earth's heat: hot dry rock geothermal energy. Springer Science & Business Media.

Energy EarthShots, DOE Office of Policy, https://www.energy.gov/policy/energy-earthshots-initiative, 2022

Enhanced Geothermal Shot, An Energy EarthShot Initiative by DOE-GTO, https://www.energy.gov/eere/geothermal/enhanced-geothermal-shot

GTO Multi-Year Program Plan (FY 2022-2026), https://www.energy.gov/sites/default/files/2022-02/GTO%20Multi-Year%20Program%20Plan%20FY%202022-2026.pdf, 2022.

GeoVision Report, Harnessing the Heat Beneath Our Feet, DOE-GTO, 2019.

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

Frash, L.P.: Geothermal Design Tool (GeoDT), Proceedings, 46th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2021).

Frash, L.P., Fu, P., Morris, J., Gutierrez, M., Neupane, G., Hampton, J., Welch, N., Carey, J.W., Kneafsey, T.: Fracture Caging to Limit Induced Seismicity, Geophysical Research Letters, 48(1), e2020GL090648 (2020).

Frash, L.P., Li, W., Meng, M., Carey, J.W., Sweeney, M.: Enhanced Geothermal System Design Using GeoDT and Fracture Caging — EGS Collab Stimulation Prediction Study, Proceedings, 56th US Rock Mechanics/Geomechanics Symposium, Santa Fe, NM (2022).

Frash, L.P., Carey, J. W., Ahmmed, B., Meng, M., Sweeney, M., Bijay, K. C., Iyare, U., Madenova, Y., Let's Make Enhanced Geothermal Systems Work, Proceedings, 48th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2023).

Hamm, S.G., Anderson, A., Blankenship, D., Boyd, L.W., Brown, E.A., Frone, Z., Hamos, I., Hughes, H.J., Kalmuk, M., Marble, A. and McKittrick, A.M., 2021. Geothermal Energy R&D: An Overview of the US Department of Energy's Geothermal Technologies Office. Journal of Energy Resources Technology, 143(10), p.100801.

Lichtner, P.C., Hammond, G.E., Lu, C., Karra, S., Bisht, G., Andre, B., Mills, R. and Kumar, J., 2015. PFLOTRAN user manual: A massively parallel reactive flow and transport model for describing surface and subsurface processes (No. LA-UR-15-20403). Los Alamos National Lab.(LANL), Los Alamos, NM (United States); Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States); Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States); OFM Research, Redmond, WA (United States).

Mudunuru, M.K., Vesselinov, V.V. and Ahmmed, B., 2022. GeoThermalCloud: Machine Learning for Geothermal Resource Exploration. Journal of Machine Learning for Modeling and Computing, 3(4).

Samek, W., Montavon, G., Lapuschkin, S., Anders, C.J. and Müller, K.R., 2021. Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), pp.247-278.

Okoroafor, E.R., Smith, C.M., Ochie, K.I., Nwosu, C.J., Gudmundsdottir, H. and Aljubran, M.J., 2022. Machine learning in subsurface geothermal energy: Two decades in review. Geothermics, 102, p.102401.

Vesselinov, V.V., Frash, L., Ahmmed, B. and Mudunuru, M.K., 2021, June. Machine Learning to Characterize the State of Stress and its Influence on Geothermal Production. In Geothermal Rising Conference, San Diego, CA.

Mudunuru, M.K., Vesselinov, V.V. and Ahmmed, B., 2022. GeoThermalCloud: Machine Learning for Geothermal Resource Exploration. *Journal of Machine Learning for Modeling and Computing*, *3*(4).