

GEOHERMALCLOUD: MACHINE LEARNING FOR GEOTHERMAL RESOURCE EXPLORATION

Maruti K. Mudunuru,^{1,*} Velimir V. Vesselinov,² & Bulbul Ahmmed³

¹Earth System Measurement & Data Group, Atmospheric Sciences & Global Change Division, Pacific Northwest National Laboratory, Richland, Washington 99352, USA

²EnviTrace LLC. Santa Fe, New Mexico 87501, USA

³Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

*Address all correspondence to: Maruti K. Mudunuru, Earth System Measurement & Data Group, Atmospheric Sciences & Global Change Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA, E-mail: maruti@pnnl.gov

Original Manuscript Submitted: 10/9/2022; Final Draft Received: 12/6/2022

Geothermal is a renewable energy source that can provide reliable and flexible electricity generation for the world. In the past decade, play fairway analysis (PFA) studies identified that geothermal resources without surface expression (e.g., blind/hidden hydrothermal systems) have vast potential. However, a comprehensive search for these blind systems can be time-consuming, expensive, and resource-intensive, with a low probability of success. Accelerated discovery of these blind resources is needed with growing energy needs and higher chances of exploration success. Recent advances in machine learning (ML) have shown promise in shortening the timeline for this discovery. This paper presents a novel ML-based methodology for geothermal exploration towards PFA applications. Our methodology is provided through our open-source ML framework, GeoThermalCloud <https://github.com/SmartTensors/GeoThermalCloud.jl>. The GeoThermalCloud uses a series of unsupervised, supervised, and physics-informed ML methods available in SmartTensors AI platform <https://github.com/SmartTensors>. Through GeoThermalCloud, we can identify hidden patterns in the geothermal field data needed to discover blind systems efficiently. Crucial geothermal signatures often overlooked in traditional PFA are extracted using the GeoThermalCloud and analyzed by the subject matter experts to provide ML-enhanced PFA (ePFA), which is informative for efficient exploration. We applied our ML methodology to various open-source geothermal datasets within the U.S. (some of these are collected by past PFA work). The results provide valuable insights into resource types within those regions. This ML-enhanced workflow makes the GeoThermalCloud attractive for the geothermal community to improve existing datasets and extract valuable information often unnoticed during geothermal exploration.

KEY WORDS: geothermal energy exploration, unsupervised machine learning, play fairway analysis, SmartTensors artificial intelligence platform, hidden signatures

1. INTRODUCTION

Geothermal energy is a nearly inexhaustible renewable resource that harnesses Earth's heat (Brown et al., 2012). Broad deployment of geothermal energy can provide essential benefits to the U.S. domestic energy needs, including electricity production, thermal energy storage, mining of critical materials (e.g., lithium), and efficient heating and cooling solutions (Cheng, 2022; Wicks, 2022). As of 2021, the U.S. installed geothermal powerplant capacity is more than 3.8 gigawatts (GW) (Hamm et al., 2021). In the last decade, the U.S. Department of Energy (DOE) Geothermal Technologies Office (GTO) recently conducted a study called GeoVision to understand and assess the geothermal potential in our country (Geo, 2019). Based on the state-of-the-art geothermal technology (e.g., recent advances in drilling and data analytics), the 2019 GeoVision analysis report suggests that more than 30 GW of new and undiscovered hydrothermal resources is recoverable. Additionally, we can access 100+ GW of new geothermal energy by developing enhanced geothermal systems (EGS). However, various technical and nontechnical barriers exist to successful exploration, development, and widespread geothermal energy deployment (Imolauer et al., 2010; Levine and Young, 2017; Young et al., 2017). A significant barrier is efficient discovery and accelerated exploration of hidden or blind geothermal resources (Dobson, 2016).

Blind systems are shown to have three times more resource potential within the same region than existing and identified hydrothermal systems (Dobson, 2016). However, these systems lack surface thermal expressions at the ground surface, making it more challenging to discover them compared to traditional, surface-manifested hydrothermal resources. As a result, the probability of successfully locating and developing these resources without a comprehensive and expensive exploration and research is very low (Gehring and Loksha, 2012). To overcome these challenges of discovering blind systems, the U.S. DOE-GTO has funded multiple projects through three phases to develop a systematic approach to quantify geothermal prospects while reducing overall exploration costs (Hamm et al., 2021). This approach to improving geothermal development projects' success rate is called play fairway analysis (PFA). The PFA concept borrowed from the oil and gas industries allows us to identify potential locations of blind hydrothermal systems and quantify geothermal potential in the regions of interest (Holmes and Fournier, 2022). Traditional PFA was able to identify various regions of interest for geothermal exploration (Faulds et al., 2016). However, the main limitation of this traditional analysis is effectively combining sparse and multisource data streams with missing values to better explore geothermal resources (Moraga et al., 2022; Mudunuru et al., 2017; Siler et al., 2021; Smith, 2021; Vesselinov et al., 2020).

In this study, we present GeoThermalCloud, a novel machine learning (ML) methodology to enhance and accelerate PFA by overcoming these limitations. Unsupervised (self-supervised) ML algorithms based on non-negative matrix factorization (NMF) and custom k -means clustering (NMF k) provide means to overcome this challenge (Alexandrov and Vesselinov, 2014; Vesselinov et al., 2019). The NMF k algorithm is one of the tools available in the SmartTensors artificial intelligence (AI) platform <https://github.com/SmartTensors> to preprocess sparse datasets with missing values. The GeoThermalCloud is developed based on this powerful AI platform for extracting hidden geothermal signatures from field data. Specifically, the GeoThermalCloud facilitates curation and analysis of geothermal datasets <https://github.com/SmartTensors/GeoThermalCloud.jl>. The uniqueness of the GeoThermalCloud is that it is flexible, open-source, and provides site/regional data, scripts, examples, figures, and developed geothermal prospectivity maps for the geothermal community. It has built-in preprocessing, postprocessing, and

state-of-the-art visualization tools for nonexperts. Therefore, experts and nonexperts can equally utilize this capability without going through a steep learning curve. Based on subject matter expertise (SME), the discovered signatures using the GeoThermalCloud are assessed and correlated to geothermal resource type (e.g., low, medium, and high-temperature resources) (Siler et al., 2021). We can use the extracted signatures upon SME assessment to build geothermal prospectivity, develop exploration risk maps, and guide future data collection strategies (Vesselinov et al., 2022b).

The paper is organized as follows: Section 1 discusses state-of-the-art for PFA and its limitations. Section 2 describes the proposed methodology for ML-enhanced PFA (ePFA). Section 3 describes the PFA data curated from open-source geothermal repositories for NMF k analysis. Section 4 presents the results and discussion of ML-analyzed geothermal data for popular PFAs in the USA. Finally, Section 5 presents our conclusions.

2. PROPOSED METHODOLOGY

This section describes the proposed ML methodology used to build the GeoThermalCloud. GeoThermalCloud capabilities and workflow 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. In traditional PFA analysis, first, the SMEs review and analyze the data to estimate the geothermal potential in a region (Faulds et al., 2016; Ito et al., 2017; Lautze et al., 2017; Lindsey et al., 2021; Siler et al., 2017). Second, dataset attributes and associated evidence of geothermal system occurrence are combined using weighted products or weighted sums. Finally, geothermal PFA models are generated to estimate the site- or regional-scale prospectivity in the form of maps or other data visualization tools. Our ePFA workflow has four steps combining the strengths of SMEs and ML to develop maps of geothermal potential (Ahmed and Vesselinov, 2022b). The site- or regional-scale data are curated and processed using ML to discover hidden features/signatures in the first step. SMEs interpret these signatures to associate them with a resource type (e.g., low/medium/high-temperature geothermal systems). In the second step, ML analyses are performed at other sites/regions to transfer knowledge (e.g., reusing discovered geothermal patterns, signatures, and relationships between geothermal data attributes) to the analyzed site/regional datasets. In the third step, physics insights gained from process models and simulations are applied to enrich the analyzed dataset (e.g., fill data gaps). Finally, the information from the above three individual steps is combined using risk analysis and uncertainty quantification methods to develop site-specific prospectivity maps and future data acquisition strategies.

Through the ePFA, we can incorporate physics into the open-source GeoThermalCloud framework in two ways. First, we can generate synthetic data conditioned on site-specific knowledge (e.g., geology) using multiphysics codes such as PFLOTRAN (Hammond et al., 2014; Lichtner et al., 2020), GeoDT (Frash, 2021). These data are ingested into the SmartTensors AI platform to develop ML models for the geothermal quantities of interest (e.g., thermal power production, temperature, chemical concentration). The second approach, which has gained traction recently, is through physics-informed neural networks (Cuomo et al., 2022; Raissi et al., 2019) and differentiable programming (Innes et al., 2019; Rackauckas et al., 2021). We can

incorporate governing laws of physics (e.g., mass, energy, reactive transport) as loss functions in neural networks when developing ML models. We can also differentiate simulators such as PFLOTRAN and GeoDT using differentiable programming methods. Both these capabilities are available to GeoThermalCloud by interfacing with the SmartTensors AI platform.

For this ePFA workflow development, we use an unsupervised ML algorithm called NMF k (Alexandrov and Vesselinov, 2014; Vesselinov et al., 2022a) to discover hidden patterns in geothermal datasets. The NMF k is a ML method that combines NMF with customized k -means clustering. NMF learns a parts-based representation (Cichocki et al., 2009; Lee and Seung, 1999) of the analyzed geothermal data matrix. The k -means clustering groups similar NMF learned pieces together and discovers unique underlying patterns (Wagstaff et al., 2001). Customization of the k -means clustering provides insights on the similarity of learned parts within a given cluster compared to other identified groups (Iliev et al., 2018; Vesselinov et al., 2019). This study builds a data matrix X of size (n, m) , after curating the raw geothermal datasets. This geothermal data matrix is developed by combining geophysical, geological, hydrological, and thermal data sources. For instance, the DOE's Geothermal Data Repository (<https://gdr.openei.org/>) provides a venue to download such datasets for ML analysis. The preprocessor normalizes the raw data such that the elements of X are non-negative. If needed, it also log transforms individual attributes to reduce the effects of outliers and non-normal distributions. Here, n is the number of geothermal data attributes, and m is the number of spatial locations of geothermal measurements. The NMF k is part of the SmartTensors AI platform and can be found at <https://github.com/SmartTensors>.

NMF of the NMF k algorithm decomposes $X_{n \times m}$ into two matrices, $W_{n \times k}$ (basis/mixing matrix) and $H_{k \times m}$ (coefficient/attribute matrix) as

$$X = W \times H + \epsilon(k), \quad (1)$$

where $\epsilon(k)$ is an error matrix representing the discrepancy between X and its estimate $W \times H$ for the specified k , while k is the unknown/hidden number of signatures present in the data and is always smaller than n and m . The W matrix represents how the measurement locations are related to the hidden signatures. The H matrix depicts the relationship between attributes and hidden signatures. In this methodology, the number of hidden signatures (k) is unknown and is identified by performing a series of NMF analyses for $k = 2, 3, \dots, d$, where $d \leq \min\{m, n\}$. The NMF process minimizes the following objective function, \mathcal{L} , based on the Frobenius norm for a specified k , such that the entries of the resulting W and H matrices are non-negative:

$$\mathcal{L} = \|X - W \times H\|_F \quad \text{such that} \quad W, H \geq 0 \quad \forall \quad n, m, k. \quad (2)$$

In the minimization, the following rules are used to update the W and H matrices (Lee and Seung, 1999):

$$W_{ij} \leftarrow W_{ij} \frac{(X H^T)_{ij}}{(W H H^T)_{ij}}, \quad (3)$$

and

$$H_{ij} \leftarrow H_{ij} \frac{(W^T X)_{ij}}{(W^T W H)_{ij}}. \quad (4)$$

For each k , NMF is solved for a series of random initial guesses (typically, 1000 or more) for W and H matrices. The random initial guesses for the matrices are generated uniformly within the bounds of the input data ranges. The least value of \mathcal{L} for a given k is assumed as the best value for

the reconstruction error. After completing the NMF process, the columns of the 1000 estimated H matrices are clustered into k clusters using a customized k -means clustering. Alternatively, we cluster the rows of the 1000 estimated W matrices. Typically, we cluster the smaller matrix. However, k is also unknown in the k -means clustering. The algorithm consecutively examines a specified k by obtaining 1000 H matrices for each feature. The custom k -means clustering keeps the number of solutions in each cluster equal to the number of NMF runs. For example, for the case with $k = 2$, after the execution of 100 NMF runs (performed with different random initial guesses for the elements of W and H matrices), each of the two clusters will contain 100 solutions each. We enforce the condition that the two clusters have an equal number of solutions (Alexandrov and Vesselinov, 2014; Vesselinov et al., 2018). This is because each NMF simulation contributes an equal number of solutions for each k .

During clustering, the similarity between two clusters is assessed using the silhouette width/value (Rousseeuw, 1987), which is essentially the cosine norm:

$$\rho(p, q) = 1 - \frac{\sum_{i=1}^n p_i q_i}{\sum_{i=1}^n p_i^2 \sum_{i=1}^n q_i^2}, \quad (5)$$

where p_i and q_i are components of vectors \mathbf{p} and \mathbf{q} . The silhouette value quantifies how similar an object is to its own cluster compared to other clusters and varies from -1 to $+1$; high values indicate that the object is well matched to its own cluster and poorly matched to neighboring clusters. The combination of reconstruction error [$\mathcal{L}(k)$] and the silhouette value are used to determine the optimal number of hidden signatures. The reconstruction error $\mathcal{L}(k)$ decreases with increasing k . However, after a certain k , the silhouette value $S(k)$ sharply falls to a negative value even though $\mathcal{L}(k)$ decreases. This is the point at which k_{opt} is chosen. If k is low, the silhouette value will be high, and $\mathcal{L}(k)$ can be high. This may result in underfitting. For high k , the silhouette value will be low and the solution may be overfit, the best estimate for k is a number that optimizes both $\mathcal{L}(k)$ and the silhouette value.

3. DATA

In this section, we describe the open-source geothermal datasets analyzed using GeoThermalCloud. The GeoThermalCloud uses NMF k not only for the execution of ML analyses but also applies the postprocessing and visualization tools provided by NMF k . Datasets include hydrological, geophysical, geomechanical, geochemical, and geological attributes. The geothermal dataset also covers regional conditions and corresponding resource types (e.g., low, medium, and high temperature) in Nevada, Utah, California, Oregon, Idaho, New Mexico, Texas, and the Hawaiian Islands in the USA. Figure 1 shows the spatial locations of the analyzed data. All the data attributes are rescaled within the range of 0.0–1.0 using a unit range transformation. If needed, some of the attributes are log-transformed. In the following paragraphs, we briefly describe the processed data for ML analysis. Additional information on these datasets is described in Ahmmed (2020a,b) and Vesselinov et al. (2021), which were curated from the Geothermal Data Repository (Weers and Anderson, 2016; Weers et al., 2022).

The Nevada geothermal PFA Phases I, II, and III identified Great Basin as the largest region with hidden geothermal resources. The study area comprises 14,341 spatial locations at which geochemical data are sampled. A total of 18 different geochemical attributes are available for ML analysis. Within the Frontier Observatory for Research in Geothermal Energy, Utah

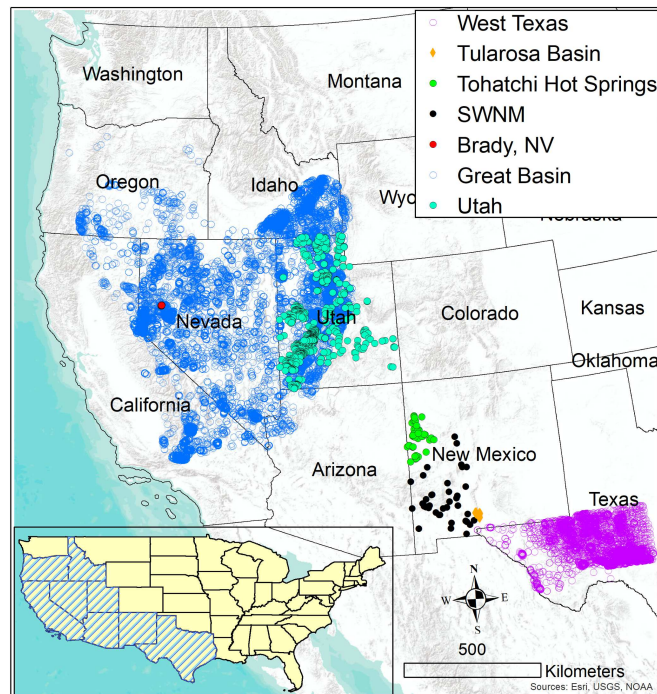


FIG. 1: Geothermal field data: Field locations at which geothermal data are available for ePFA analysis. These field data include thermal, geophysical, geomechanical, geochemical, and geological attributes. These datasets are collected and curated at various locations in NV, UT, CA, OR, ID, NM, and HI.

(UtahFORGE) region, 22 data attributes, including satellite interferometric synthetic-aperture radar (InSAR), geophysical (e.g., gravity, seismic), geochemical, and geothermal attributes, are collected at 102 locations. Twenty-one data attributes collected during Phases I and II at 120 different locations are analyzed in the Tularosa Basin. In the Tohatchi Hot Springs area in northern New Mexico, we investigated 19 data attributes observed at 41 wells. A total of 15 data attributes collected at 247 locations during Phases I and II of Hawaiian geothermal PFA are analyzed using NMF k . The data attributes employed in the current ML study are related to geothermal processes. For instance, they can be used as a proxy or indirect measure of rock-water interactions to understand reservoir temperatures, geothermal conditions, reservoir boundaries, and heat source type (e.g., meteoric, magmatic, mixed) (Dobson, 2016; Fournier, 1977; Fridriksson and Ármannsson, 2007; Klein, 2007).

4. RESULTS AND DISCUSSION

In this section, we provide the results of our ML methodology on the PFA datasets described in Section 3.

4.1 Great Basin

Great Basin includes multiple known and hidden geothermal resources, which have been studied in past decades (Faulds et al., 2016). GeoThermalCloud is used to analyze geochemical

data, which can indirectly infer geothermal conditions (e.g., reservoir temperatures, conditions, boundaries, and heat source type). Geochemistry provides insight into water-rock interactions and fluid circulation at various depths (e.g., shallow, deep). ML analysis is performed on a curated data matrix consisting of 18 data attributes (e.g., temperature, pH, total dissolved solids (TDS), anions, and cation concentrations) at 14,321 locations. Our NMF k algorithm identified three optimal signatures based on the loss function, \mathcal{L} , and the silhouette width value.

Figure 2 shows the dominant attributes in the discovered signatures, the spatial distribution of these identified signatures, and their association with a resource type (e.g., low, medium, high). Signatures A, B, and C define modestly, highly, and moderately prospective hydrothermal systems. Signature A represents modestly hydrothermal systems because of the low contribution of groundwater temperature in this extracted signature. The dominant attributes of this signature are TDS, Br⁻, B⁺, and δO^{18} . Signature B represents highly prospective hydrothermal systems due to the high contribution of temperature. The dominant attributes of this signature are pH, Al³⁺, Be²⁺, and quartz and chalcedony geothermometers. Signature C defines moderately prospective hydrothermal systems because of the medium contribution of temperature. The dominant attributes of the signature are Mg²⁺ and Ca²⁺. Signatures B and C distribution suggest that the significant portions of the Great Basin region have prospective geothermal resources. Areas with

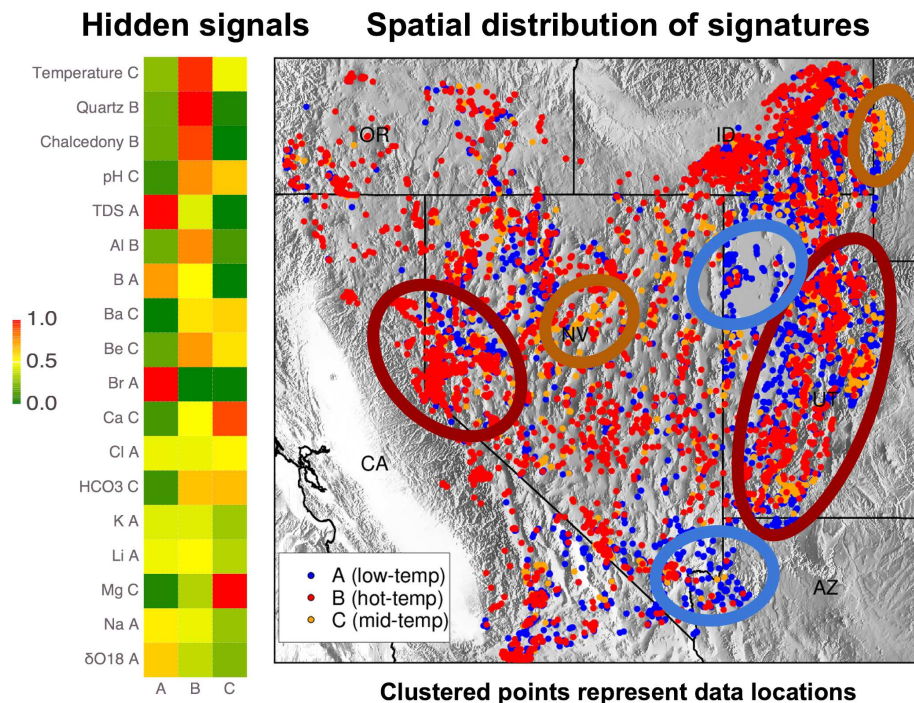


FIG. 2: ML analysis of Great Basin data: ML results for geothermal data within the Great Basin region in Nevada. The left figure shows the optimal hidden geothermal signatures discovered by NMF k , which is three. The right figure provides the spatial distribution of these signatures mapped onto the geothermal data collection locations based on resource type (e.g., modestly, highly, and moderately prospective). The ellipses mark the regions with a high density of similar resource-type signature locations. The H -matrix contains the locations and associated spatial distribution of signatures at these locations. This color-coded H -matrix is mapped back to the spatial distribution map.

a high density of B and C locations are labeled with ellipses in Fig. 2. Some of these locations align with existing geothermal resources and sites such as Dixie Valley and Brady geothermal areas in Nevada (Faulds et al., 2016).

4.2 UtahFORGE

UtahFORGE is a field laboratory located in Milford, Utah, to study EGS (Allis et al., 2016; Moore et al., 2019). This field laboratory provides a unique opportunity to develop and test new technologies for characterizing and creating sustainable geothermal systems in a controlled environment. During two phases of EGS site development in UtahFORGE, geothermal data at 102 locations with 22 attributes were collected. These include satellite interferometric synthetic-aperture radar (InSAR), geophysical (gravity, seismic), geochemical, and geothermal attributes. The NMF k method was applied to this dataset and discovered four optimal hidden geothermal signatures. Figure 3 shows the discovered signatures, their relationship to data attributes, and the associated mapping of these signatures to the sampled locations. Signatures A and B are related to favorable geothermal conditions, which have different dominant attributes. Signature A's key attributes are gravity, seismic, InSAR, and specific geochemical species. The high contribution from the InSAR features indicates that locations covered by Signature A are influenced by EGS activity. Signature B's dominant attributes are temperature, heat flow, pH, and ions such as K⁺

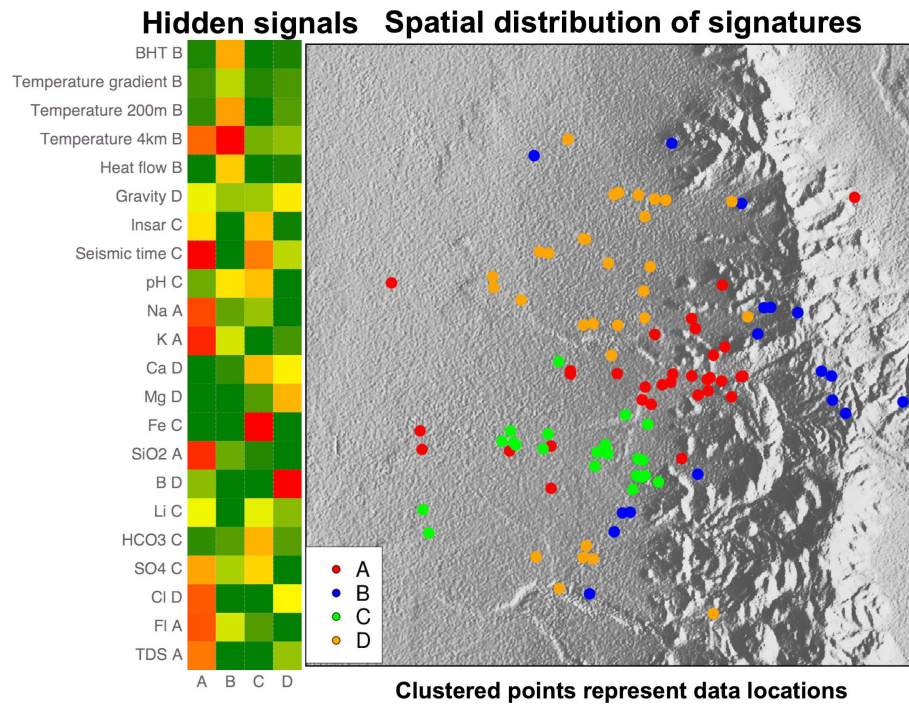


FIG. 3: ML analysis of UtahFORGE data: ML results for UtahFORGE geothermal data. The left figure shows the four hidden geothermal signatures discovered by NMF k . The right figure provides the spatial distribution of these signatures based on low-, medium-, and high-temperature resource types. According to Williams et al. (2008), low-, medium-, and high-temperature resources correspond to $< 90^{\circ}\text{C}$, $90\text{--}150^{\circ}\text{C}$, and $> 150^{\circ}\text{C}$, respectively.

and FI^- . These heat, pH, and temperature attributes are good indicators of potential resources that may be hidden within this region. In a nutshell, NMFk analysis solidifies the idea that signature B locations can be a source to discover blind systems in Utah.

4.3 Tularosa Basin

The Tularosa Basin in the southern Rio Grande Rift within New Mexico consists of promising geothermal plays based on Phase I and II PFA studies (Brandt, 2015; Nash, 2017; Nash et al., 2017). The presence of known geothermal systems, high-temperature slim hole wells, quaternary faults, and relatively high heat flow sheds light on possible blind geothermal systems that may be present in the study area. During the traditional Tularosa Basin PFA, various datasets representing the heat of the Earth, such as water chemistry, temperature gradients, 2 m surface temperature, and heat flow, are collected to assess the geothermal potential. We investigate 21 attributes collected during this PFA, including field geology (geological reconnaissance and mapping), gravity surveys, shallow temperature surveys, well water sampling and geothermometry, temperature logging, and magnetotelluric surveys. Figure 4 shows the results of NMFk applied to this Tularosa Basin PFA dataset. Signature C defines the potential for hidden hydrothermal resources due to favorable characteristics associated with geological structures for geothermal exploration. Critical attributes in this signature include heat flow, SiO_2 , silica geothermometer, Critical attributes in this signature include heat flow, SiO_2 , silica geothermometer,

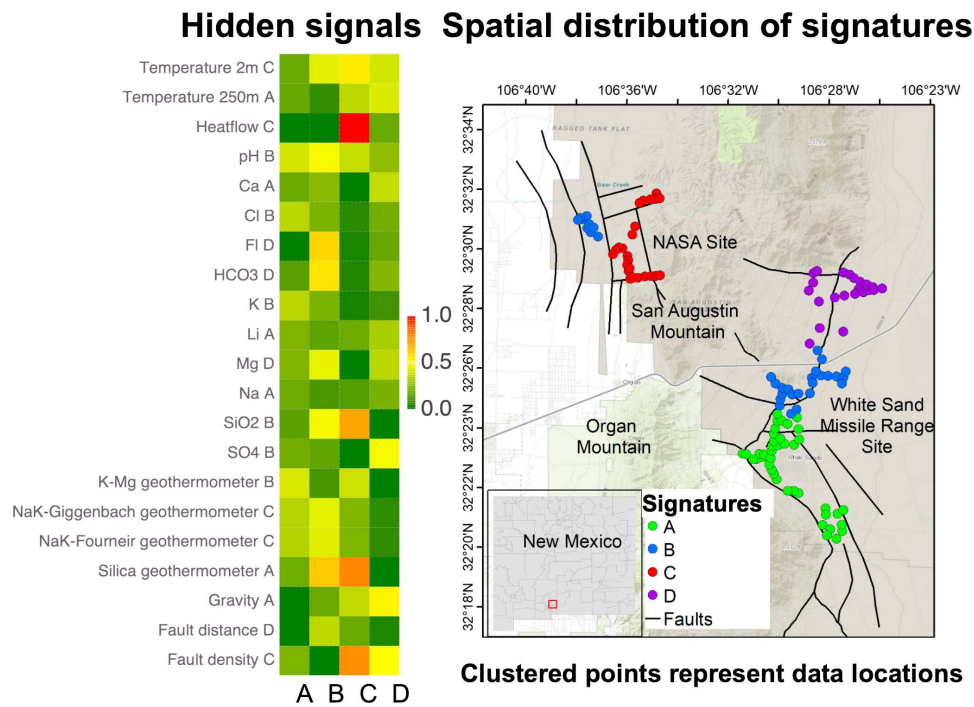


FIG. 4: ML analysis of Tularosa Basin data: ML results for geothermal data from the Tularosa Basin in New Mexico. The left figure shows the four hidden geothermal signatures discovered by NMFk. The right figure provides the spatial distribution of these signatures. Signature C has dominant attributes (e.g., heat flow, SiO_2 , silica geothermometer, fault density) that correspond to the location of hidden geothermal resources in this region.

and fault density. The spatial distribution of this signature suggests that resources are close to NASA and other military installations, making geothermal energy cost-effective for power needs for several facilities in this region.

4.4 Tohatchi Hot Springs

Tohatchi Hot Springs in New Mexico are favorable for enhanced geothermal exploration (Ahmed and Vesselinov, 2022a; Ahmed et al., 2021; Levitte and Gambill, 1980). Characterizing the geothermal source and governing mechanisms (e.g., identifying which data attributes contribute to the heat source) that make the water hot in these springs can help us better explore this region. NMF k is a suitable method to better understand the resource type. We investigated 19 data attributes observed at 41 wells and Tohatchi's hydrogeology to assess if groundwater circulates from a shallow depth (400 m due to shallow geothermal system, secondary fracture permeability effects) or from a greater depth (e.g., deep circulation of water, forced-convection geothermal systems). Figure 5 shows the result of our NMF k analysis on the well data, which identified four hidden signatures. Signature C defines the hidden potential geothermal resources, whose key attributes are pH, Li^+ , HCO_3^- , F^+ , quartz-water-vapor geothermometer, and Na-K-Ca geothermometer. Our ML analysis of the well data attributes and its signature may suggest that

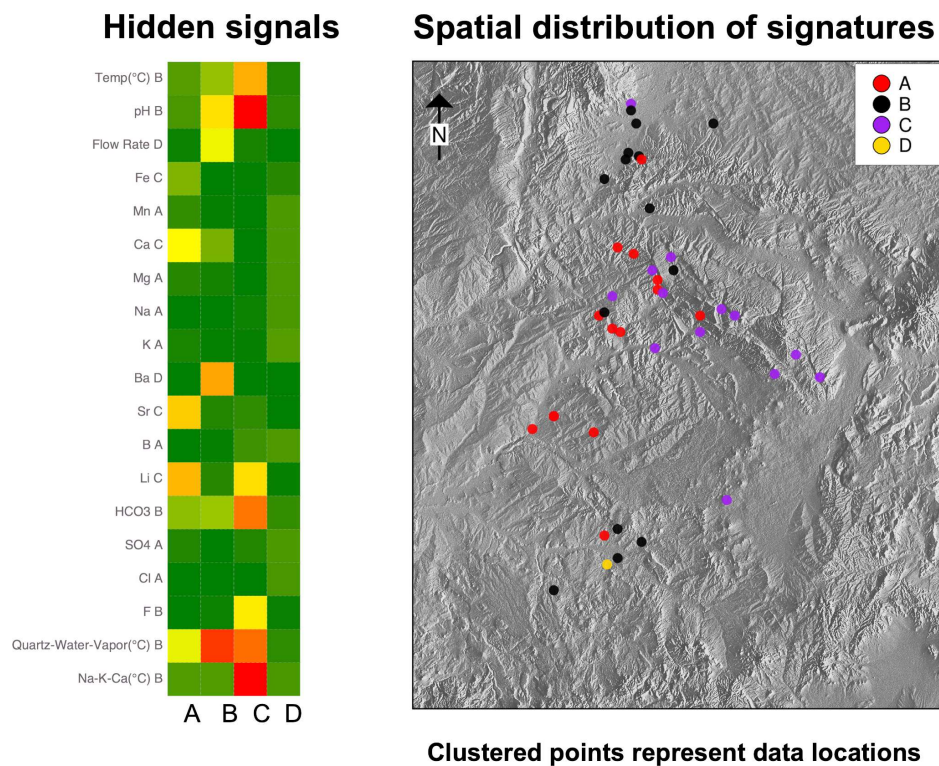


FIG. 5: ML analysis of Tohatchi Hot Springs data: ML results based on geothermal data for the Tohatchi Hot Springs region in New Mexico. The left figure shows the four hidden geothermal signatures discovered by NMF k . The right figure provides the spatial distribution of these signatures, where signature C defines the hidden potential of high-temperature geothermal resources within this region.

Tohatchi Hot Springs' geothermal reservoir is large for long-term production of hot well water ($< 100^{\circ}\text{C}$). ML analysis in connection with the SME technical reports allows us to hypothesize that groundwater circulates from a shallow depth due to low geothermal gradients.

4.5 Hawaiian Islands

Hawaii is an ocean island hotspot environment with a magmatic geothermal heat source. Currently, geothermal resources within this region provide up to 3% of the state's energy needs. The PFA analysis conducted across the state of Hawaii by Ito et al. (2017) and Lautze et al. (2017) showed that these islands have huge geothermal prospects. We have applied NMF k to discover hidden features in hydrological data linked to Hawaii's geothermal resources. The discovered signatures and their spatial distribution are shown in Fig. 6. NMF k identified four signatures in the PFA data that characterized the islands and their resource types within the state of Hawaii. Each signature has unique physical significance (e.g., water type, alkalinity). Hidden signatures with high contributions from temperature are essential for ML-enhanced geothermal exploration. Specifically, signatures B and D define the hidden potential of geothermal resources within this region as they are linked to groundwater temperature and chemistry. Signature A is related to geochemistry (e.g., cations and anions), another proxy for blind systems. This extracted signature shows that further data analysis of each island in the state of Hawaii can help us better understand the geothermal resource type and its source (e.g., meteoric, magmatic, or connate).

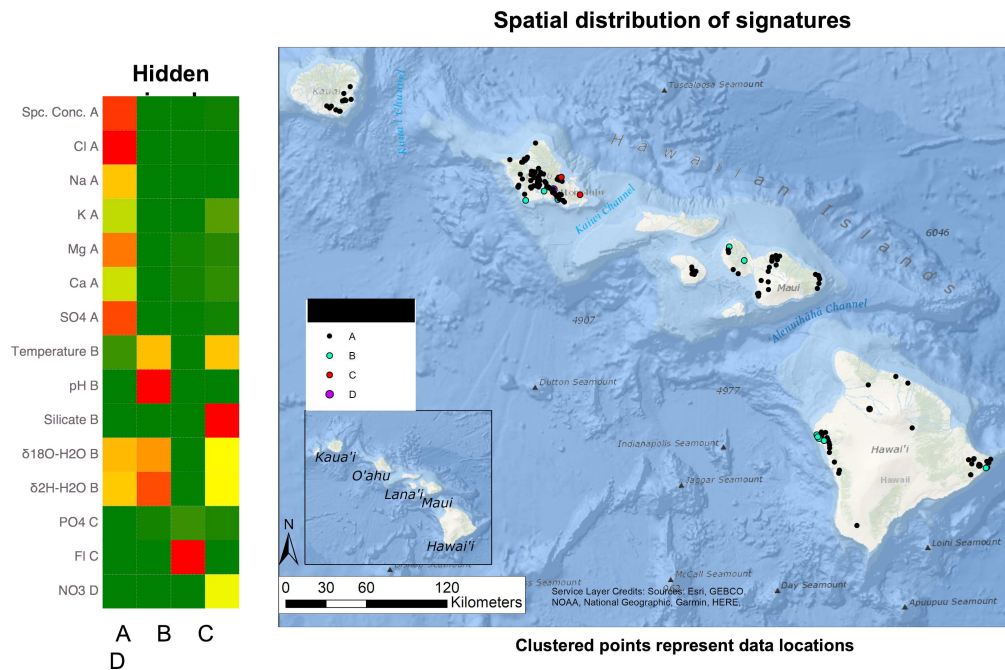


FIG. 6: ML analysis of Hawaiian islands data: ML results based on geothermal data for the Hawaiian islands. The left figure shows the four hidden geothermal signatures discovered by NMF k . Their dominant attributes include pH, $\delta^{18}\text{O}$, $\delta^2\text{H}$, and silicate. The right figure provides the spatial distribution of these signatures. Signatures B and D define the hidden potential of geothermal resources within this region as they are linked to groundwater temperature and chemistry.

Note that the groundwater chemistry data collected on Ca^{2+} and Mg^{2+} ions can be indicative of a hydrothermal system. For instance, Lautze et al. (2017) provide insights into why such ionic data are less reliable but still valuable for geothermal resource exploration. To summarize, the chemical signatures from Mg^{2+} are very sensitive to removal by geothermal processes. Hence, chemical signatures from such cations should be interpreted with caution as they can also be removed from secondary clays due to groundwater infiltration (Thomas, 1987).

5. CONCLUSIONS

This paper presented an machine learning (ML) methodology developed in our open-source framework, GeoThermalCloud <https://github.com/SmartTensors/GeoThermalCloud.jl> for efficient geothermal exploration of hidden geothermal resources. The GeoThermalCloud uses a series of unsupervised, supervised, and physics-informed ML methods in our SmartTensors AI platform. Here, the presented GeoThermalCloud analyses are performed using our unsupervised ML algorithm called NMF_k from the SmartTensors AI platform. NMF_k is applied to analyze and enhance existing public geothermal datasets. In this way, GeoThermalCloud provides enhanced ML-based exploration results consistent with those obtained from traditional SME-based PFA methods. The NMF_k discovers hidden signatures embedded in the geothermal data, and then SMEs analyze the extracted hidden signatures to understand the discovered knowledge. We applied the proposed GeoThermalCloud framework to various regions in the U.S. to better characterize the resource types based on the dominant data attributes discovered by NMF_k . These regions include the Great Basin in Nevada, UtahFORGE, Tularosa Basin, Tohatchi Hot Springs in New Mexico, and Hawaiian Islands. In Great Basin, NMF_k identified modestly, highly, and moderately prospective hydrothermal systems; their dominant characterization attributes; and their spatial distribution using geochemistry data (Ahmed and Vesselinov, 2022b). They also compared their results with a comprehensive PFA study conducted in the same study area by Faults et al. (2016). Faults et al. used geophysical, geological, geochemical, etc., data yet the prospectivities of the two studies are similar with much fewer data using our ML. We analyzed site prospectivity within UtahFORGE and potential exploration options (e.g., drilling location for future geothermal field exploration). In the Tularosa Basin, Tohatchi Hot Springs, and Hawaiian Islands, our NMF_k was able to identify low-, medium-, and high-temperature hydrothermal systems and found dominant attributes and spatial distribution for each hydrothermal system. To conclude, these applications of our GeoThermalCloud for different types of available PFA data instill confidence in our approach to analyze new datasets (e.g., airborne geophysical, 3DEP LiDAR surveys) such as Geoscience Data Acquisition for Western Nevada (GeoDAWN) (Rhodes et al., 2021), currently being collected over the parts of Nevada and California. Figure 7 shows a pictorial description of ML-enhanced PFA (ePFA) workflows using our GeoThermalCloud, which we plan to enhance by combining GeoDAWN data with multi-physics simulations. ML-enhanced GeoDAWN data will provide a deeper understanding of the geothermal conditions (e.g., fault patterns, stress regime) in the Great Basin that represent blind hydrothermal systems, thereby accelerating the discovery of new and undiscovered geothermal resources for domestic use.

ACKNOWLEDGMENTS

This research is based upon work supported by the U.S. Department of Energy's Office of energy efficiency and renewable energy under the Geothermal Technology Office Machine Learning

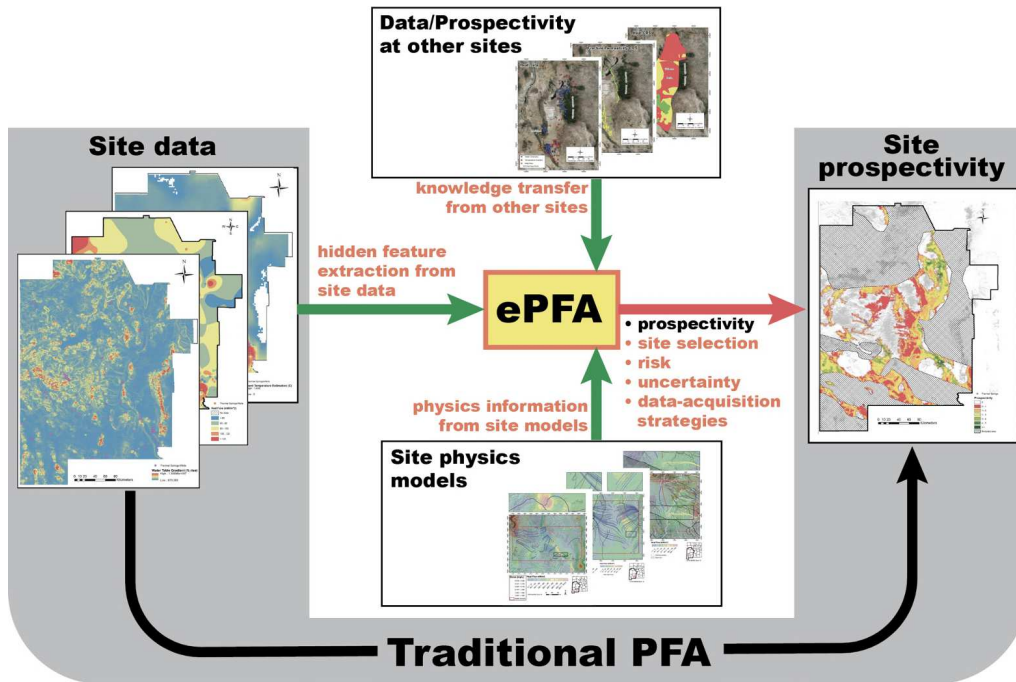


FIG. 7: Traditional vs. ePFA: This conceptual figure compares the traditional and proposed workflow for geothermal PFA using the GeoThermalCloud framework. In traditional PFA, geothermal play fairway models are generated using weighted products or sums of geothermal attributes curated by SMEs from raw field datasets. SMEs also perform data weighting in traditional PFA models. In ePFA, the final prospectivity maps are generated by combining ML analysis with SME. ML is used at various stages in the PFA analysis, such as extracting hidden signatures from curated field data, transferring knowledge across sites, and discovering physics information from site-specific process models.

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). Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. The authors thank the reviewers whose feedback helped in substantially improving the manuscript.

REFERENCES

- Ahmed, B., Non-Negative Matrix Factorization to Discover Dominant Attributes in Utah FORGE Data, *GRC Trans.*, vol. **44**, p. 1281, 2020a.
- Ahmed, B., Unsupervised Machine Learning to Extract Dominant Geothermal Attributes in Hawaii Island Play Fairway Data, in *Proc. of the Geothermal Resources Council's Annual Meeting & Expo*, Reno, NV, pp. 18–21, 2020b.
- Ahmed, B. and Vesselinov, V.V., Exploration of Groundwater and Geothermal Characteristics of Tohatchi

- Hot Springs Aquifer at Local and Regional Scales, Tech. Rep., Los Alamos National Lab., Los Alamos, NM, 2022a.
- Ahmed, B. and Vesselinov, V.V., Machine Learning and Shallow Groundwater Chemistry to Identify Geothermal Prospects in the Great Basin, USA, *Renew. Energy*, vol. **197**, pp. 1034–1048, 2022b.
- Ahmed, B., Vesselinov, V.V., Woods, S., Singer, S., and Lake, S., Geothermal Resource Analysis at Tohatchi Hot Springs, New Mexico, Tech. Rep., Los Alamos National Lab., Los Alamos, NM, 2021.
- Alexandrov, B.S. and Vesselinov, V.V., Blind Source Separation for Groundwater Pressure Analysis Based on Nonnegative Matrix Factorization, *Water Resour. Res.*, vol. **50**, no. 9, pp. 7332–7347, 2014.
- Allis, R., Moore, J., Davatzes, N., Gwynn, M., Hardwick, C., Kirby, S., McLennan, J., Pankow, K., Potter, S., and Simmons, S., EGS Concept Testing and Development at the Milford, Utah FORGE Site, in *Proc. of 41st Workshop on Geothermal Reservoir Engineering*, Stanford, CA, 2016.
- Brandt, A., Tularosa Basin Play Fairway Analysis: Methodology Flow Charts, Tech. Rep., USDOE Geothermal Data Repository, University of Utah, Salt Lake, UT, 2015.
- Brown, D.W., Duchane, D.V., Heiken, G., and Hrisco, V.T., *Mining the Earth's Heat: Hot Dry Rock Geothermal Energy*, New York: Springer Science & Business Media, 2012.
- Cheng, A., President's Page: Geothermal Energy: Current and Future, *Leading Edge*, vol. **41**, no. 9, pp. 588–589, 2022.
- Cichocki, A., Zdunek, R., Phan, A.H., and Amari, S.I., *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*, Hoboken, NJ: John Wiley & Sons, 2009.
- Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., and Piccialli, F., Scientific Machine Learning through Physics-Informed Neural Networks: Where We Are and What's Next, arXiv: 2201.05624, 2022.
- Dobson, P.F., A Review of Exploration Methods for Discovering Hidden Geothermal Systems, *GRC Trans.*, vol. **40**, pp. 695–706, 2016.
- Faulds, J., Hinz, N., Coolbaugh, M., Shevenell, L., and Siler, D., The Nevada Play Fairway Project – Phase II: Initial Search for New Viable Geothermal Systems in the Great Basin Region, Western USA, *GRC Trans.*, vol. **40**, pp. 535–540, 2016.
- Fournier, R., Chemical Geothermometers and Mixing Models for Geothermal Systems, *Geothermics*, vol. **5**, nos. 1-4, pp. 41–50, 1977.
- Frash, L.P., Geothermal Design Tool (GeoDT), in *Proc. of the 46th Workshop on Geothermal Reservoir Engineering*, Stanford, CA, vol. **17**, February 15, 2021.
- Fridriksson, T. and Ármannsson, H., Application of Geochemistry in Geothermal Resource Assessments, Short Course on Geothermal Development in Central America Resource Assessment and Environmental Management, El Salvador, 2007.
- Gehring, M. and Loksha, V., Geothermal Handbook: Planning and Financing Power Generation, Tech. Rep., The World Bank, 2012.
- Geothermal Technologies Office, GeoVision: Harnessing the Heat Beneath Our Feet, accessed from <https://www.energy.gov/eere/geothermal/articles/geovision-harnessing-heat-beneath-our-feet>, 2019.
- Hamm, S.G., Anderson, A., Blankenship, D., Boyd, L.W., Brown, E.A., Frone, Z., Hamos, I., Hughes, H.J., Kalmuk, M., and Marble, A., Geothermal Energy R&D: An Overview of the US Department of Energy's Geothermal Technologies Office, *J. Energy Resour. Technol.*, vol. **143**, no. 10, p. 100801, 2021.
- Hammond, G.E., Lichtner, P.C., and Mills, R., Evaluating the Performance of Parallel Subsurface Simulators: An Illustrative Example with PFLOTRAN, *Water Resour. Res.*, vol. **50**, no. 1, pp. 208–228, 2014.
- Holmes, R.C. and Fournier, A., Machine Learning-Enhanced Play Fairway Analysis for Uncertainty Characterization and Decision Support in Geothermal Exploration, *Energies*, vol. **15**, no. 5, p. 1929, 2022.

- Iliev, F.L., Stanev, V.G., Vesselinov, V.V., and Alexandrov, B.S., Nonnegative Matrix Factorization for Identification of Unknown Number of Sources Emitting Delayed Signals, *PLoS One*, vol. **13**, p. e0193974, 2018.
- Imolauer, K., Richter, B., and Berger, A., Non-Technical Barriers of Geothermal Projects, in *Proc. World Geothermal Congress*, Bali, Indonesia, 2010.
- Innes, M., Edelman, A., Fischer, K., Rackauckas, C., Saba, E., Shah, V.B., and Tebbutt, W., A Differentiable Programming System to Bridge Machine Learning and Scientific Computing, arXiv: 1907.07587, 2019.
- Ito, G., Frazer, N., Lautze, N., Thomas, D., Hinz, N., Waller, D., Whittier, R., and Wallin, E., Play Fairway Analysis of Geothermal Resources Across the State of Hawaii: 2. Resource Probability Mapping, *Geothermics*, vol. **70**, pp. 393–405, 2017.
- Klein, C., Advances in the Past 20 Years: Geochemistry in Geothermal Exploration, Resource Evaluation and Reservoir Management, *GRC Trans.*, vol. **31**, pp. 17–22, 2007.
- Lautze, N., Thomas, D., Hinz, N., Apuzen-Ito, G., Frazer, N., and Waller, D., Play Fairway Analysis of Geothermal Resources Across the State of Hawaii: 1. Geological, Geophysical, and Geochemical Datasets, *Geothermics*, vol. **70**, pp. 376–392, 2017.
- Lee, D.D. and Seung, H.S., Learning the Parts of Objects by Non-Negative Matrix Factorization, *Nature*, vol. **401**, p. 788–791, 1999.
- Levine, A. and Young, K.R., Crossing the Barriers: An Analysis of Land Access Barriers to Geothermal Development and Potential Improvement Scenarios, *GRC Trans.*, vol. **41**, pp. 2164–2192, 2017.
- Levitte, D. and Gambill, D., Geothermal Potential of West-Central New Mexico from Geochemical and Thermal Gradient Data, Tech. Rep., Los Alamos Scientific Lab., NM, 1980.
- Lichtner, P.C., Hammond, G.E., Lu, C., Karra, S., Bisht, G., Andre, B., Mills, R.T., Kumar, J., and Frederick, J.M., PFLOTRAN User Manual, Tech. Rep., 2020.
- Lindsey, C.R., Ayling, B.F., Asato, G., Seggiaro, R., Carrizo, N., Larcher, N., Marquetti, C., Naon, V., Serra, A.C., and Faulds, J.E., Play Fairway Analysis for Geothermal Exploration in North-Western Argentina, *Geothermics*, vol. **95**, p. 102128, 2021.
- Moore, J., McLennan, J., Allis, R., Pankow, K., Simmons, S., Podgorney, R., Wannamaker, P., Bartley, J., Jones, C., and Rickard, W., The Utah Frontier Observatory for Research in Geothermal Energy (FORGE): An International Laboratory for Enhanced Geothermal System Technology Development, in *Proc. of 44th Workshop on Geothermal Reservoir Engineering*, Stanford, CA, 2019.
- Moraga, J., Duzgun, H., Cavur, M., and Soydan, H., The Geothermal Artificial Intelligence for Geothermal Exploration, *Renew. Energy*, vol. **192**, pp. 134–149, 2022.
- Mudunuru, M.K., Ahmmed, B., Karra, S., Vesselinov, V.V., Livingston, D.R., and Middleton, R.S., Site-Scale and Regional-Scale Modeling for Geothermal Resource Analysis and Exploration, in *Proc. of 45th Workshop on Geothermal Reservoir Engineering*, Stanford, CA, 2020.
- Nash, G., Tularosa Basin Play Fairway Analysis: White Sands Missile Range Main Cantonment and NASA Area Faults, New Mexico, Tech. Rep., USDOE Geothermal Data Repository, Energy and Geoscience, 2017.
- Nash, G.D., Brandt, A., Pfaff, B., Hardwick, C., Gwynn, M., Blake, K., Simmons, S., and Bennett, C.R., Phase 2: Updated Geothermal Play Fairway Analysis of the Tularosa Basin, New Mexico, *GRC Trans.*, vol. **41**, pp. 2312–2327, 2017.
- Rackauckas, C., Edelman, A., Fischer, K., Innes, M., Saba, E., Shah, V.B., and Tebbutt, W., Generalized Physics-Informed Learning through Language-Wide Differentiable Programming, accessed from https://ceur-ws.org/Vol-2587/article_8.pdf, 2021.
- Raissi, M., Perdikaris, P., and Karniadakis, G.E., Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations, *J. Comput. Phys.*, vol. **378**, pp. 686–707, 2019.

- Rhodes, G., Roberts, B., Pauling, H., Taverna, N., and Warren, P., Analysis of Selected Publicly Available Geothermal Exploration Data Gaps, Tech. Rep., National Renewable Energy Lab., Golden, CO, 2021.
- Rousseeuw, P.J., Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis, *J. Comput. Appl. Math.*, vol. **20**, pp. 53–65, 1987.
- Siler, D.L., Pepin, J.D., Vesselinov, V.V., Mudunuru, M.K., and Ahmed, 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*, vol. **9**, no. 1, pp. 1–17, 2021.
- Siler, D.L., Zhang, Y., Spycher, N.F., Dobson, P.F., McClain, J.S., Gasperikova, E., Zierenberg, R.A., Schiffman, P., Ferguson, C., and Fowler, A., Play-Fairway Analysis for Geothermal Resources and Exploration Risk in the Modoc Plateau Region, *Geothermics*, vol. **69**, pp. 15–33, 2017.
- Smith, C.M., Machine Learning Techniques Applied to the Nevada Geothermal Play Fairway Analysis, PhD, University of Nevada, Reno, 2021.
- Thomas, D.M., A Geochemical Model of the Kilauea East Rift Zone, U.S. Geological Survey Professional Paper 1350, 1987.
- Vesselinov, V., Ahmed, B., Mudunuru, M., Pepin, J., Burns, E., Siler, D., Karra, S., and Middleton, R., Discovering Hidden Geothermal Signatures Using Non-Negative Matrix Factorization with Customized k -Means Clustering, *Geothermics*, vol. **106**, p. 102576, 2022a.
- Vesselinov, V.V., Ahmed, B., Frash, L., and Mudunuru, M.K., GeoThermalCloud: Machine Learning for Discovery, Exploration, and Development of Hidden Geothermal Resources, Tech. Rep., 2022b.
- Vesselinov, V.V., Alexandrov, B.S., and O'Malley, D., Contaminant Source Identification Using Semi-Supervised Machine Learning, *J. Contaminant Hydrol.*, vol. **212**, pp. 134–142, 2018.
- Vesselinov, V.V., Mudunuru, M.K., Ahmed, B., Karra, S., and Middleton, R.S., Discovering Signatures of Hidden Geothermal Resources Based on Unsupervised Learning, in *Proc. of 45th Workshop on Geothermal Reservoir Engineering*, Stanford, CA, 2020.
- Vesselinov, V.V., Mudunuru, M.K., Karra, S., O'Malley, D., and Alexandrov, B.S., Unsupervised Machine Learning Based on Non-Negative Tensor Factorization for Analyzing Reactive-Mixing, *J. Comput. Phys.*, vol. **395**, pp. 85–104, 2019.
- Vesselinov, V.V., O'Malley, D., Frash, L.P., Ahmed, B., Rupe, A.T., Karra, S., Middleton, R.S., Alexandrov, B., Mudunuru, M.K., and Mims, M., GeoThermalCloud: Cloud Fusion of Big Data and Multi-Physics Models Using Machine Learning for Discovery, Exploration, and Development of Hidden Geothermal Resources, Tech. Rep., Los Alamos National Lab., Los Alamos, NM, 2021.
- Wagstaff, K., Cardie, C., Rogers, S., and Schrödl, S., Constrained k -Means Clustering with Background Knowledge, in *Proc. of the Eighteenth International Conference on Machine Learning*, Williams College, Williamstown, MA, pp. 577–584, June 28–July 1, 2001.
- Weers, J. and Anderson, A., The DOE Geothermal Data Repository and the Future of Geothermal Data, Tech. Rep., National Renewable Energy Lab., Golden, CO, 2016.
- Weers, J., Anderson, A., and Taverna, N., The Geothermal Data Repository: Ten Years of Supporting the Geothermal Industry with Open Access to Geothermal Data, Tech. Rep., National Renewable Energy Lab., Golden, CO, 2022.
- Wicks, F., New Horizons for Geothermal Energy, *Mech. Eng.*, vol. **144**, no. 5, pp. 34–39, 2022.
- Williams, C.F., Reed, M.J., Mariner, R.H., DeAngelo, J., and Galanis, S.P., Assessment of Moderate and High-Temperature Geothermal Resources of the United States, Tech. Rep., US Geological Survey, 2008.
- Young, K.R., Levine, A., Cook, J., Hernandez, K., Ho, J., and Heimiller, D., Crossing the Barriers: An Analysis of Market Barriers to Geothermal Development and Potential Improvement Scenarios, Golden, CO, National Renewable Energy Laboratory Tech. Rep. NREL/CP-6A20-68996, 2017.