Proceedings World Geothermal Congress 2020+1
Reykjavik, Iceland, April - October 2021

Discovering the Hidden Geothermal Signatures of Southwest New Mexico

Velimir V. Vesselinov[1], Bulbul Ahmmed[1,*], Maruti K. Mudunuru[2], Satish Karra[1], and Richard Middleton[1]

[1]Computational Earth Science Group, Los Alamos National Laboratory, Los Alamos, NM 87545

[2]Watershed & Ecosystem Science, Pacific Northwest National Laboratory, Richland, WA 99352

vvv@lanl.gov, ahmmedb@lanl.gov, maruti@pnnl.gov, rsm@lanl.gov, satkarra@lanl.gov

*Corresponding author: ahmmedb@lanl.gov

Keywords: geothermal, hidden geothermal resource, unsupervised machine learning, SmartTenosrs, NMF*k*, geothermal play, New Mexico

ABSTRACT

Southwest New Mexico (SWNM) has high potential for further geothermal exploration as demonstrated by existing Play Fairway Analysis (PFA). The region includes four physiographic provinces: the Colorado Plateau, the Mogollon-Datil Volcanic Field, the Rio Grande Rift, and the Basin and Range provinces. These provinces are known to have low-, medium/moderate-, and high-temperature geothermal resources, and some of them are already utilized and in operation. This study aims (1) to characterize the existing geothermal resources and (2) to propose future data collection plans to efficiently explore and develop the hidden geothermal resources in the SWNM region. To achieve this, we jointly analyze datasets from three geothermal studies conducted in SWNM. Next, an unsupervised machine learning technique, non-negative matrix factorization with customized *k*-means clustering (NMF*k*), is applied. NMF*k* is a robust method for finding hidden patterns or signatures in data, which can not be easily detectable or interpretable with other exploratory data methods. The NMF*k* analysis reveals (1) the optimal numbers of signal capturing hidden structures in the data, (2) key parameters that characterize the geothermal systems in SWNM and (3) prospectivity of hidden geothermal resources. Furthermore, a detailed guideline to collect more data, including their corresponding locations, is provided. NMF*k* finds 8 geothermal signatures (A, B, C, D, E, F, G, and H) in the data. The locations associated with three signatures (A, D, and H) potentially have high geothermal prospectivity. *BHT, Quartz-water-vapor, Temperature @ 2m depth, Temperature @ 250m depth, Silica, K-Mg, Na-Ca-K,* and *NaK-Giggenbach geothermometers, Ca, SO_4, Na, and Spring density* are key parameters defining geothermal prospectivity in SWNM. Therefore, data of these attributes at all locations might facilitate better characterization and predictions of geothermal resources.

1. INTRODUCTION

The study area is located in the southwest of New Mexico (SWNM), which has Known Geothermal Resources (KGRs) (Bielicki et al., 2015). The KGRs range from low- to high-temperature, and their characteristics are controlled by the local and regional geology and hydrology (Fleischmann, 2006; Bielicki et al., 2015). The geology of SWNM is broadly divided by the four physiographic provinces: the Colorado Plateau, the Mogollon-Datil Volcanic Field, Basin and Range, and the Rio Grande rift (Bielicki et al., 2015). The Basin and Range province has a geothermal based power plant (Lightning dock), a few greenhouses based farms, and several medium temperature wells or springs (Kelley, 2010). The Volcanic Field province has a few hot springs. Play Fairway Analysis (PFA) of SWNM revealed more potential geothermal resources (Bielicki et al., 2015). However, what subsurface features control the geothermal occurrence in the study region is poorly understood. This study aims to discover hidden data signatures in the existing SWNM data that will elucidate the relation between subsurface features and the geothermal occurrence. Furthermore, discovered signatures can help us to characterize and identify hidden geothermal resources.

Hidden/blind geothermal resources lack surface thermal manifestation such as geysers, springs, fumaroles, etc. There are several assumptions regarding hidden geothermal resources. For instance, these blind resources might be smaller or equivalent in size when compared to the discovered resources; they might also be located predominantly in igneous and sedimentary rocks (Dobson et al., 2016; Porro 2012; Anderson et al., 2013; Brott 1981; Williams et al., 2009; Smith et al., 2004). The common characteristic is that the hidden geothermal resources lie at depth and are sealed by overlying thick strata that precludes the flow of hot water and heat towards the ground surface. Due to the lack of thermal signatures near the surface, identifying hidden geothermal resources is a challenge. There are several exploration methods (e.g., PFA, direct evidence) (Siler et al., 2019; Faulds et al., 2018; Faulds and Hinz, 2015; Faulds et al., 2016, 2015), which analyze geothermal, hydrogeochemical, geological, and geophysical data to characterize hidden geothermal resources. Geothermal data include thermal gradient, bottom-hole temperature, shallow and deep heat flow, and temperature measurements at certain depths. Water geochemistry data include major cations, anions, and tracer elements. Geological data include structural folds, faults, deep exploration well logs, and geologic framework models. Commonly used geophysical data are provided by gravity, magnetics, and electrical resistivity surveys. Seismic reflection, self-potential (SP), soil gas surveys, LiDAR, hyperspectral surveys, and microseismic monitoring are also used to explore geothermal resources, but they are often less informative, and their field execution can be expensive. However, none of these data types are unique for identifying hidden geothermal resources. Researchers often apply multiple methods on available data to assess geothermal resources; yet, currently there are no widely accepted methods that are generally applicable for discovery of unknown hidden geothermal resources. This study demonstrates the applicability of a novel unsupervised machine learning method for characterizing the geothermal resources and finding corresponding dominant data attributes in SWNM. The method is applied to analyze diverse geothermal, geological, geophysical, and geochemical datasets. This study is a key step towards developing a general ML framework for identifying hidden geothermal resources.

Applied machine learning framework is called SmartTensors and it includes a series of unsupervised and physics-informed methods developed by our team (http://tensors.lanl.gov). Here, we apply one of the SmartTensors tools called NMF*k*, which is based on non-

negative matrix factorization with customized *k*-means clustering (Alexandrov et al., 2018a,b; Vesselinov et al., 2018, 2019). NMF separates a data matrix into two smaller matrices called signature and mixing matrices, while k-means clustering finds the optimal number of signatures in the dataset. Both matrices represent hidden/latent data structure in the data. The mixing matrix captures how each of the extracted signatures are represented in the measurement locations, while the signature matrix relates attributes with signatures. Note, these explicit relations among signatures, locations, and attributes are not directly surmisable in the data. Alternative unsupervised machine learning methods such as singular value decomposition (Klema and Laub, 1980), principal component analysis (Wold, 1987), and independent component analysis (Comon, 1994) can also be applicable. However, NMF*k* provides important benefits over these methods. It can handle both real and categorical variables, sparse datasets (with missing data entries), and provides interpretable results (Alexandrov et al. 2018a,b; Vesselinov et al. 2018, 2019). Furthermore, NMF*k* has already been applied to discover hidden geothermal signatures at eight unique geothermal sites in the U.S. (Ahmmed et al. 2020a,b,c,d; Vesselinov et al. 2020a,b). All these analyses successfully discovered hidden geothermal signatures.

During the SWNM PFA study, a total of 18 attributes were collected at 44 locations to study geothermal resources in SWNM (GDR, 2019; Bielicki et al., 2015). Another PFA work collected and analyzed a total of 23 attributes at 120 locations to study geothermal resources in the Tularosa Basin, which is also part of SWNM (Bennett & Nash, 2017). In the third study, 24 attributes at 43 locations were collected to assess geothermal prospects in the Tohatchi springs area (Levitte and Gambill, 1980). Here, we combined the datasets from all these three studies to generate a dataset that includes 42 unique attributes at 207 locations (Figure 1). It is important to note that the three studies have collected data for some common but also many different data attributes. As a result, there are many data gaps. Next, we applied NMF*k* on this joint non-uniform dataset to identify hidden signatures characterizing geothermal resources, the optimal number of these signatures, the dominant attributes associated with each signature, and locations associated with each signature. Furthermore, NMF*k* provides estimates for the missing data entries based on the extracted signatures. Finally, we offer data acquisition guidelines and suggestions for future work.

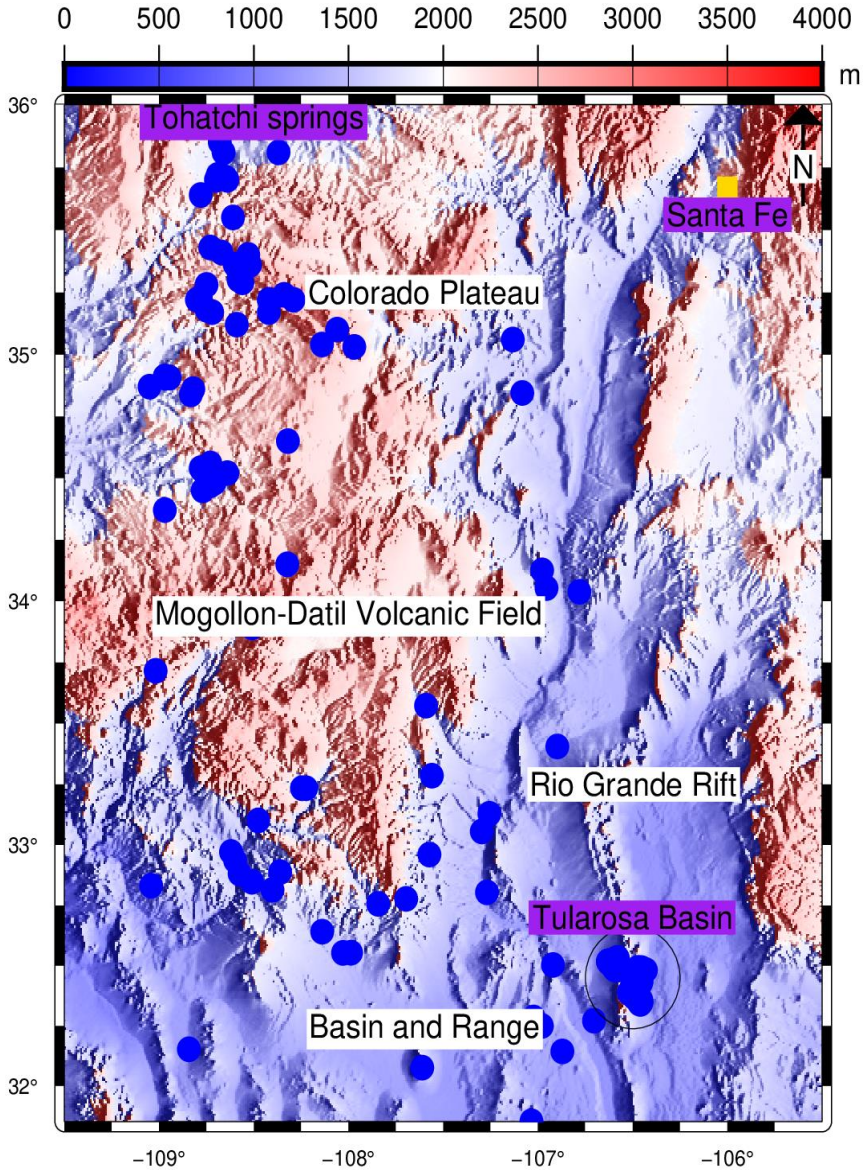

Figure 1: Map of the study area with colors representing elevation and blue circles represent locations of the data for this study, text with white background represents physiographic provinces, and text with purple background represents a specific location.

2. METHOD

2.1 NON-NEGATIVE MATRIX FACTORIZATION

Given observational data X of size (n, m) with non-negative values, where m is the number of observational locations from which data are sampled, and n is the number of data attributes that are observed at each location. The first step in the NMFk analyses decompose this matrix X into a non-negative "attribute" matrix W of size (n, k) and non-negative "location" matrix V of size (k, m):

$$X = W \times V + \varepsilon(k)$$

where k is the unknown number of signals (features) present in the data and $\epsilon(k)$ is a matrix representing the reconstruction error for k^{th} signals. The "attribute" matrix W captures how the extracted features are related to the attributes. The "location" matrix V represents how the hidden features are related to the locations.

The optimal number of hidden signals k_{opt} is unknown and is estimated by performing a series of non-negative matrix factorization for different values of k; $k = 1, 2, 3, \ldots, d$. The maximum value d cannot exceed n or m. This is achieved by minimizing the following objective function O based on Frobenius norm for all possible values of k(Lee and Sung, 1999)

$$O = ||X - W \times V||_F;\ W, V > 0\ \forall\ m, n, k$$

For each k value in the range $1, 2, 3, \ldots, d$, non-negative matrix factorization is performed multiple times (typically on the order of 1000 times) based on random initial guesses for W and V matrices. The best estimate of O for a given k from all these runs is applied to define the reconstruction error for each k value: $O(k)$. The resulting multiple solutions of W (or alternatively V; typically, it is preferred to cluster the smaller matrix) are clustered into k clusters using a customized k-means clustering. During clustering, we enforce the condition that each of the k rows of V (or, respectively, the k columns of W) are assigned to one of the k clusters in m-dimensional space (or n-dimensional space, if W is clustered). As a result, it is enforced that each of the k clusters contains an equal number of members, which is equal to the number of performed multiple random runs (e.g., 1000 solutions). After clustering, the average silhouette width $S(k)$ of the k clusters is computed (Rousseeuw, 1987). This metric (see Vesselinov et al. 2018) measures how well the random NMF solutions are clustered for a given value of k. When a cosine norm is applied, which is typically preferred, the values of $S(k)$ theoretically can vary between -1 and 1. Commonly, $S(k)$ is close to 1 for a small number of signals k and declines sharply, typically below zero, after an optimal number of signal, k_{opt}, is reached. As a result, k_{opt} value defines the maximum number of signals that relatively accurately reconstruct the observational data matrix X as estimated by $O(k_{\text{opt}})$ without overfitting. The overfitting occurs for k values higher than k_{opt}; the underfitting region is for k values less than k_{opt}.Typically, the average silhouette width $S(k_{\text{opt}})$ is close to 1 or at least greater than 0.25. NMFk algorithm estimates k_{opt}automatically. More details on the NMFk algorithm and its implementation are discussed in Alexandrov and Vesselinov 2014, Vesselinov et al. 2018, Vesselinov et al. 2019 (cf. http://tensors.lanl.gov, https://github.com/TensorDecompositions).

2.2 DATA DESCRIPTION

The analyzed data matrix contains 42 geological, geophysical, water geochemistry, and geothermal attributes at 207 locations (Figure 1). The 42 attributes are specified in Table 1, including their minimum, maximum, standard deviation. The first 16 attributes in Table 1 contain geological and geophysical attributes, the next 16 attributes include water geochemistry attributes, and the last 10 attributes represent geothermal attributes. Data are heavily sparse (contains many missing values). Also, they significantly vary from each other, and the range of minimum and maximum is large. To make the ranges smaller, several attributes are log-transformed. Table 1 shows the number of data entries (non-missing) values and a flag if any variables are log-transformed.

The data attributes *Gravity anomaly* and *Magnetic intensity* express information about geologic setup such as faults, uplifts, volcanism, magmatic emplacement, and secondary mineralization (Biehler, 1971; Kohrn et al. 2011). *Volcanic dyke density* and *Volcanic vent density* represent volcanic occurrence. *Fault intersection density*, *Quaternary fault density*, and *State map fault density*, and *Fault distance* capture reservoir characteristics in the subsurface. *Seismicity* is a measure encompassing earthquake magnitude at a given location and, in general, captures the seismotectonic properties of the subsurface. *Drainage density*, *Spring density*, and *Hydraulic gradient* control lateral and vertical groundwater flow in the subsurface. *Precipitation* represents water available for surface runoff and subsurface infiltration. Groundwater *Flow rate* provides information about groundwater flow and sustainability of the reservoirs. *Crustal thickness* and *Depth to basement* impact heat flow from Earth's mantle to the ground surface. Hydrogeochemical constituents such as B^{+}(*B*), Li^{+}(*Li*), *pH*, Fe^{2+}(*Fe*), Mn^{2+}(*Mn*), Ca^{2+}(*Ca*), Mg^{2+}(*Mg*), Na^{+}(*Na*), K^{+}(*K*), Ba^{2+}(*Ba*), Sr^{2+}(*Sr*), HCO_3^{-} (*HCO3*), SO_4^{2-} (*SO4*), Cl^{-} (*Cl*), Fl^{-} (*Fl*), and SiO_2(*SiO2*) may indicate groundwater mixing and reservoir temperatures (Fourneir et al. 1970; Fridriksson and Ármannsson, 2007). *Heat flow*, bottom-hole temperature (*BHT*), *Temperature at 2m* and *250m* depths are geothermal attributes and indicators of geothermal reservoir conditions (Dobson, 2016; Kratt et al. 2008, 2009, 2010). *Silica*, *Quartz-water-vapor*, *K-Mg*, *Na-K-Ca*, *NaK-Giggenbach*, *NaK-Fourneir* geothermometers estimate potential reservoir temperature based on a few assumptions (Fourneir et al. 1970). Every attribute in the dataset is not equally important; nonetheless, we used all available attributes to examine their significance in characterizing geothermal conditions in SWNM.

3. RESULTS AND DISCUSSION

NMFk was applied to SWNM data exploring the presence of 15 hidden signals. Figure 2 shows the estimated $O(k)$and $S(k)$for k ranging from 2 to 15 signals after 1000 NMF iterations with random initial guesses. $O(k)$ reduces as the number of signals increases. This is typical because by increasing the number of signals, we increase the degrees of freedom (the number of adjustable unknown variables which are the elements of the W and V matrices) of the solved NMF optimization problem; however, $S(k)$values fluctuate. $S(k)$depends on the k-means clustering of the Vor the attribute matrix. The optimal solution is for a k value, which satisfies reasonable low and high values of $O(k)$and $S(k)$, respectively. Here, the solution for 8 signals shows reasonable low $O(k)$and high $S(k)$values; therefore, the optimal number of hidden geothermal signals in the analyzed dataset is 8.

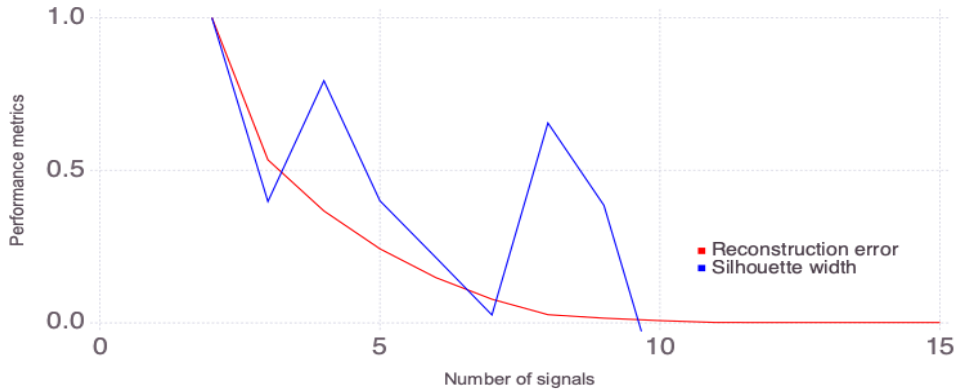

Figure 2: Estimated reconstruction error $O(k)$ and silhouette width $S(k)$ for different numbers of signals k.

Table 1: Attributes used in the study and their properties (minimum, maximum, standard deviation, count of actual (non-missing) data entries, and log-transformation flags).

Attributes	Unit	Minimum	Maximum	Standard deviation	Count of (non-missing) values	Log transformed
Gravity anomaly	mGal	0	121.8	31.34	164	N
Magnetic intensity	A/m	31.77	683.95	135.43	44	Y
Volcanic dike density	m/m^2	0	4.91	1.18	44	Y
Volcanic vent density	m/m^2	0	0.02	0	44	N
Fault intersection density	m/m^2	0	0.07	0.02	44	Y
Quaternary fault density	m/m^2	0	28.89	5.12	164	N
State map fault density	m/m^2	0	41.09	10.1	44	Y
Fault distance	m	419.4	2239.76	481.89	68	N
Crustal thickness	m	57.27	73.88	4.77	44	N
Depth to basement	m	26.56	5751.75	1355.6	44	Y
Seismicity	Richter	0	0.04	0.01	44	N
Flow Rate	gpm	0.84	1135.44	267.57	22	Y
Precipitation	inch	511.9	806.7	81.22	44	N
Hydraulic gradient	[]	0.59	14.2	3.17	44	N
Drainage density	m/m^2	4.33	26.97	6.57	44	N
Springs density	m/m^2	0	0.02	0	44	N
B	mg/L	0	8.32	1.64	87	Y
Li	mg/L	0	150.45	64.16	197	Y
pH	mg/L	6.6	9.7	0.77	38	N
Fe	mg/L	0	12.1	2.22	43	Y
Mn	mg/L	0	3.8	0.58	43	Y
Ca	mg/L	2.52	339.62	70.23	43	Y
Mg	mg/L	0	1115	168.15	43	Y
Na	mg/L	5.46	44002.86	6694.42	43	Y
K	mg/L	1.09	287.57	45.21	43	N
Ba	mg/L	0	0.54	0.12	43	Y
Sr	mg/L	0	4.22	0.89	39	Y
HCO3	mg/L	240.83	775.19	116.76	43	N
SO4	mg/L	28.35	11435.85	1722.27	43	Y
Cl	mg/L	5.67	62702.97	9558.9	43	Y
Fl	mg/L	0.23	4.97	1.04	43	N
SiO2	mg/L	35.05	63.47	6.92	120	N
Heat flow	mW/m^2	7.98	47.42	12.85	129	N
BHT	°C	13.86	52.76	7.6	39	N
Temperature @ 2m depth	°C	23.45	32.62	1.68	120	N
Temperature @ 250m depth	°C	59.69	64.61	1.36	120	N
K-Mg geothermometer	°C	53.54	69.05	4.37	120	N
NaK-Giggenbach geothermometer	°C	356.14	498.72	27.29	120	N
NaK-Fourneir geothermometer	°C	314.95	466.37	29.01	120	N
Silica geothermometer	°C	28.13	185.26	54.97	118	N
Quartz-Water-Vapor geothermometer	°C	10.5	88.5	13.2	43	N
Na-K-Ca geothermometer	°C	0	133	28.72	38	N

Table 2: The dominant and secondary attributes, physical significance, and potential geothermal prospectivity of the geothermal signatures.

Signatures	Dominant (primary) attributes	Secondary attributes	Physical significance	Potential geothermal prospectivity
A	*K-Mg* and *NaK-Giggenbach geothermometers*	*Na-Ca-K geothermometer, Ca*	High reservoir temperature	High
B	*Mg*	*BHT, Na-Ca-K, SO_4, Na, Silica geothermometer, pH, Ba*	Groundwater mixing	Low
C	*Mn, Fe, pH*	*BHT, Ba, Heat flow*	Groundwater mixing	Low
D	*Temperature @ 2m depth*	*Silica geothermometer, Ca*	Good connection between deep and surface heat flow path	High
E	*Fault intersection density, Spring density*	*SO_4, Precipitation*	Vertical water flow	Very low
F	*Fault intersection density, B*	*Depth to basement, Hydraulic gradient, Quartz-water-vapor*	Deep heat circulation	Low
G	*Flow rate, Li*	*Crustal thickness*	Sustainable reservoir and deep heat source	Low
H	*Temperature @ 250m depth*	*SO_4, Na, Spring density*	Potential heat source at depth	High

This study deals with a heavily sparse dataset (see Table 1). Nevertheless, the NMF*k* results are well representative of the local/regional physiography and geology. Figure 3 shows the extracted hidden geothermal signatures. Table 2 demonstrates both primary and secondary dominant attributes, the physical significance of each signature, and potential geothermal prospectivity. Moreover, Figures 4a and 4b show the spatial distribution of locations associated predominantly with the extracted 8 geothermal signatures.

Signature A appears to be represented predominantly by locations in the Tularosa Basin, which is also in the Rio Grande rift physiographic province. Signature A locations are also detected within the Colorado Plateau and Volcanic Field provinces. The area (the Tularosa Basin) associated with this geothermal signature went through a few volcanic events suggesting the heat source is relatively close to the ground surface. Also, the primary and secondary dominant attributes suggest that this area might have geothermal prospectivity associated with high-temperature geothermal resources (Table 2). However, further investigations are needed in this area to better understand the geothermal conditions, refine prospects, and propose extraction approaches.

Signature B also represents locations in the Tularosa Basin with additional three locations in the vicinity of the Colorado Plateau and Volcanic Field provinces. Five of eight primary and secondary attributes (*Mg*, *pH*, *Na*, *Ba*, and *SO_4*) of this signature are hydrogeochemical attributes; hence, this signature represents groundwater mixing (Table 2). The locations associated with this signature potentially indicate mixing of hot deep and cold shallow groundwater. However, the absence of geothermal attributes in the dominant attributes in this signature suggests very low geothermal prospectivity.

Locations of Signature C are similar to both Signatures A and B. Four of the six primary and secondary attributes (*Mn*, *Fe*, *pH*, and *Ba*) primarily represent groundwater mixing. *BHT* and *Heat flow* both are geothermal attributes with low significance. However, these are critical attributes to identify geothermal resource types. It is important to mention that all locations in the study area do not have data regarding *BHT* and *Heat flow*. More *BHT* and *Heat flow* data would assist to make a robust assessment about the type of geothermal resources. In this signature, *Mn*, *Fe*, *pH*, and *Ba* are highly correlated with *BHT* and *Heat flow*; therefore, more data of these features at all locations in the study area may also assist us to better characterize the geothermal resources. Nonetheless, the current data and analysis suggest low geothermal prospectivity for the locations associated with Signature C and have better promise than the locations associated with Signature B.

Signature D represents locations in the Basin and Range province where the *Temperature @ 2m depth* is the primary dominant attribute along with *Silica geothermometer* as a secondary dominant attribute (Table 2). Both attributes are good indicators of geothermal prospects. Therefore, our analyzes demonstrate that the locations associated with this signature have high geothermal prospectivity. For example, *Temperature @ 2m depth* assisted in discovering the Tungsten Mountain geothermal site (Kratt et al. 2008). Note, *Temperature @ 2m depth* is not available in all locations in the study area. Data for this attribute in all locations would assist in making a robust assessment of the geothermal prospectivity. Also, the highly correlated and easily accessible attributes *Silica geothermometer* and *Ca* at all locations may assist to better characterize the geothermal resources in this area.

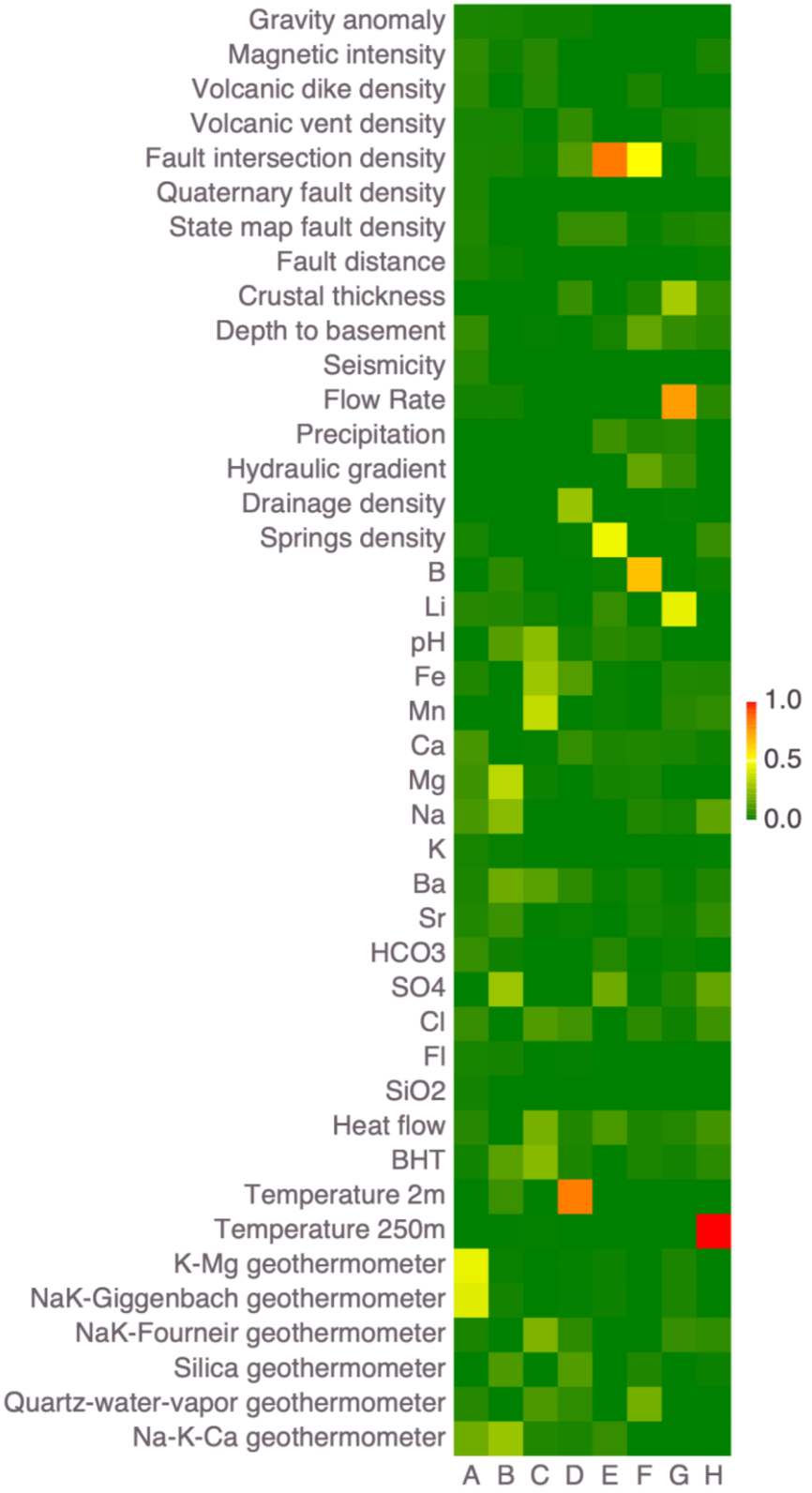

Figure 3: Attribute matrix showing the significance of attributes on the extracted geothermal signatures.

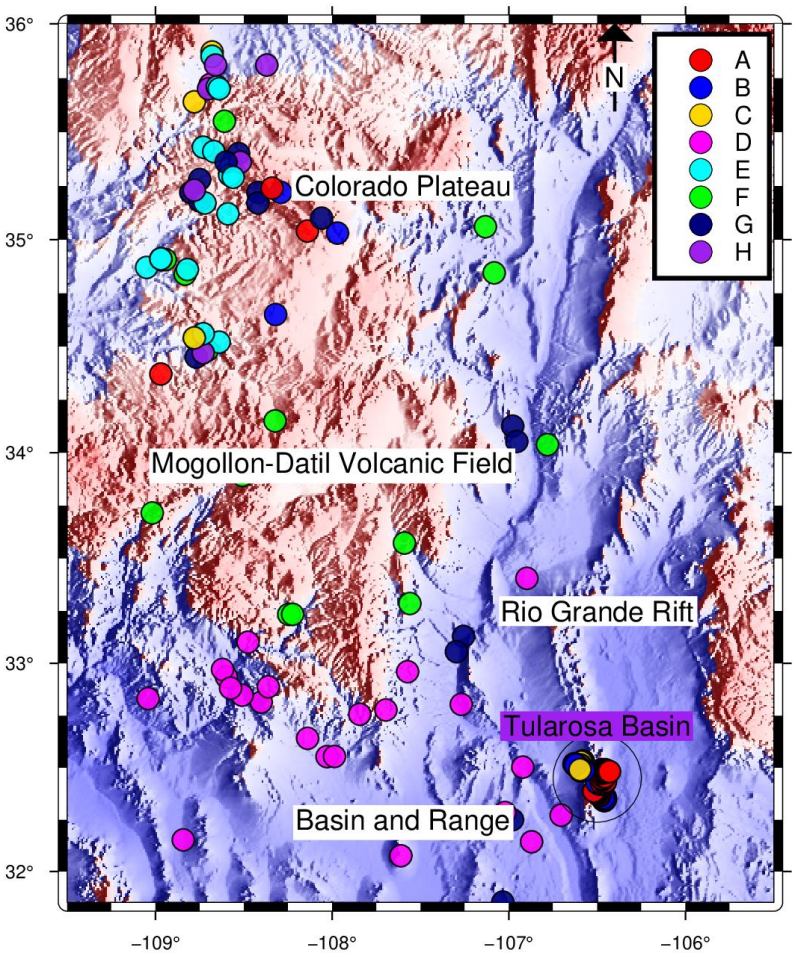

Figure 4a: Eight spatial clusters corresponding to eight signatures of the optimal signal.

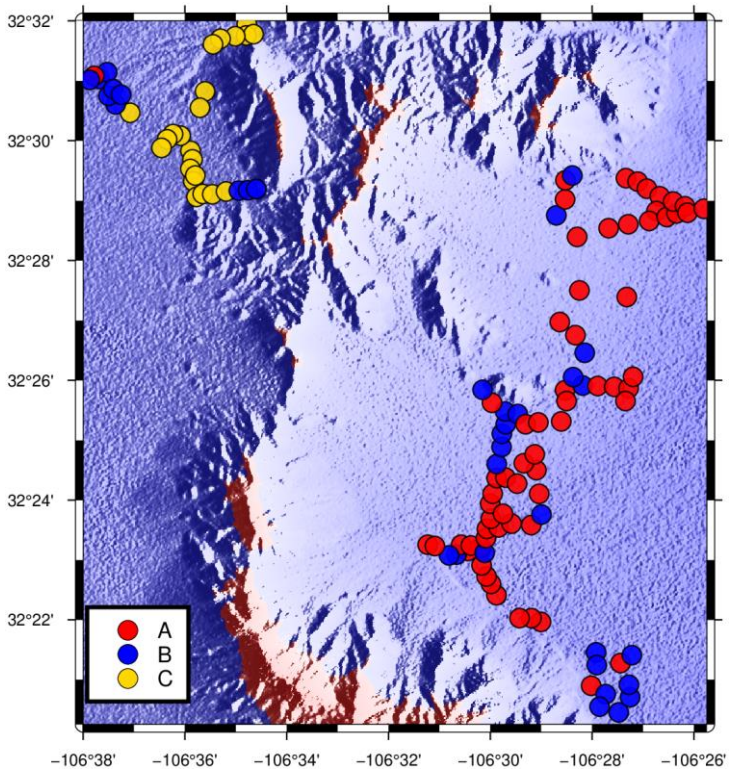

Figure 4b: Spatial clusters in the Tularosa Basin for the signatures A, B, and C.

Signature E represents locations in the western Colorado Plateau. The primary dominant attributes *Fault intersection density* and *Spring density* as well as one of the secondary dominant attributes (*Precipitation*) suggest that this region has good subsurface to

surface hydraulic connection. However, the signature fails to capture any geothermal attributes. Therefore, these locations may have very low geothermal prospectivity but potentially good groundwater resources.

Signature F represents locations in the Volcanic field. The signature attributes *Fault intersection density* and *Hydraulic gradient* indicate deep vertical circulation; *B* might be released from the deep subsurface because of high heat source; *Depth to basement* suggests thick basement. *Quartz-water-vapor* in the secondary dominant attributes suggests that the reservoir temperature may be high. High heat source in the deep, potentially high reservoir temperature but thick basement indicates that the locations associated with Signature F may have low geothermal prospectivity. It is critical to mention that *Quartz-water-vapor* data are not available at all locations. However, this attribute is still represented in this geothermal signature. Therefore, the collection of *Quartz-water-vapor* or attributes with high significance in the signal (i.e., *Fault intersection density, B, Depth to basement, Hydraulic gradient*) at all locations will be essential to make robust predictions about the geothermal prospectivity in this region.

Signature G represents locations of the Rio Grande rift zone and part of the western Colorado Plateau. The primary attributes include *Flow rate* and *Li*. High significance in *Flow rate* suggests that the locations associated with this signature have sustainable reservoirs. *Li* tracer is believed to indicate vertical mass transport from a crustal heat source, while the secondary dominant attribute, *Crustal thickness*, indicates a substantial vertical distance of propagation for the heat/mass flow from depth. All these attributes suggest that there is a hot deep resource which might not be amenable to efficient geothermal extraction because of the thick crust.

Signature H represents locations of the western Colorado Plateau. The primary dominant attribute of this signature is the *Temperature @ 250m depth* that indicates potential heat source in depth. The secondary dominant attributes (*SO_4* and *Na*) indicate groundwater mixing, while *Spring density* suggests vertical water flow. Because of the high significance of *Temperature @ 250m depth*, the locations associated with this signature may have a high-geothermal prospectivity. However, *Temperature @ 250m depth* was not available at all locations in the analyzed dataset. Additional data of *Temperature @ 250m depth* at all locations in the study area would assist in making robust predictions of geothermal prospectivity. Also, *SO_4, Na, Spring density* may facilitate better prediction because of their high significance in Signature H.

4. CONCLUSIONS

Primary objective of this study was to find optimal hidden geothermal signatures in the analyzed data, characterize the potential geothermal prospects of SWNM, and provide a future data collection plan that will enhance robustness of geothermal prospectivity predictions. We combined geothermal datasets of three unique studies in SWNM. The size of the data matrix is 207×42 (locations×attributes), and the data matrix is heavily sparse. To characterize the dataset, we applied NMF*k* to find the optimal signal that characterizes the dataset. NMF*k* is specifically designed to be capable of processing sparse datasets. NMF*k* extracted 8 hidden (latent) geothermal signatures in the dataset labeled A, B, C, D, E, F, G, and H. Our ML analysis also extracted dominant attributes and locations associated with each geothermal signature. The dominant attributes of Signature A are *K-Mg, NaK-Giggenbach* and *Na-Ca-K* geothermometers, and *Ca*. The dominant attributes of Signature B are *Mg, BHT, Na-Ca-K geothermometer, SO_4, Na, Silica geothermometer, pH,* and *Ba*. The dominant attributes of Signature C are *Mn, Fe, BHT, Ba,* and *Heat flow*. The dominant attributes of Signature D are *Temperature @ 2m depth, Silica geothermometer,* and *Ca*. The dominant attributes of Signature E are *Fault intersection density, Spring density, SO_4,* and *Precipitation.* The dominant attributes of Signature F are *Fault intersection density, B, Depth to basement, Hydraulic gradient,* and *Quartz-water-vapor*. The dominant attributes of Signature G are *Flow rate, Li,* and *Crustal thickness*. Finally, the dominant attributes of Signature H are *Temperature @ 250m depth, SO_4, Na,* and *Spring density*.

Furthermore, these attributes associated with the extracted signatures indicate the physical significance captured by each signature and their geothermal prospectivity.

Signature A shows that the locations associated with it may have potential high-reservoir temperature; therefore, they may have high-geothermal prospects.

Signature B captures groundwater mixing with little effect from geothermal attributes; hence, locations associated with this signature potentially have very low geothermal prospects.

Signature C also captures groundwater mixing but gets higher significance from critical geothermal attributes; therefore, locations associated with this signature may have better geothermal prospects than locations associated with Signature B.

Signature D captures heat flow path between depth and the surface; consequently, the locations associated with this signature potentially have potentially high geothermal prospects.

Signature E captures a well-connected flow path but not a single geothermal attribute; therefore, the locations associated with it may have low geothermal prospects.

Signature F captures deep circulation features but fails to capture critical geothermal attributes; resultantly, the locations associated with it potentially have low geothermal prospects.

Signature G captures sustainable reservoirs feature and deep heat sources but not a single geothermal attribute. Thus, the locations associated with it may have low geothermal prospects.

Signature H captures potential heat sources at 250m depth; therefore, the locations associated with Signature H potentially have high geothermal prospects.

Moreover, this study also demonstrated that several attributes at all locations will help to make robust predictions about geothermal prospectivity in SWNM. Critical geothermal attributes include *BHT, Quartz-water-vapor, Temperature @ 2m depth,* and *Temperature @ 250m depth*. Hydrogeochemical and geological attributes for instance *Silica, K-Mg, Na-Ca-K,* and *NaK-Giggenbach*

geothermometers, Ca, SO_4, Na, Spring density might also facilitate better assessment of geothermal resources in this area. All these attributes are not available at all locations in the SWNM, but they are key parameters to assess geothermal resources in any area. Therefore, collecting additional measurements of these attributes will make the current dataset more comprehensive and presented ML analyses more robust. Based on the presented SWNM results, it is clear that ML methods can be applied to make robust predictions regarding geothermal prospectivity.

ACKNOWLEDGMENTS 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). Additional information regarding the datasets and codes can be obtained from Velimir V. Vesselinov (Monty) (vvv@lanl.gov) and Bulbul Ahmmed (ahmmedb@lanl.gov).

REFERENCES

Ahmmed, B., Lautze, N., Vesselinov, V., Dores, D., and Mudunuru, M. (2020a). Unsupervised machine learning to extract dominant geothermal attributes in Hawaii Island Play Fairway data. Geothermal Resources Council, Reno, NV, October 18–23.

Ahmmed, B., Vesselinov, V., and M.K., M. (2020b). Machine learning to characterize regional geothermal reservoirs in the western USA. Fall Conference, Geological Society of America, Abstract T185-358249, October 26–29.

Ahmmed, B., Vesselinov, V., and M.K., M. (2020c). Non-negative matrix factorization to discover dominant attributes in Utah FORGE Data. Geothermal Resources Council, Reno, NV, October 18–23.

Ahmmed, B., Vesselinov, V., and Mudunuru, M. (2020d). Integration of data, numerical inversion, and unsupervised machine learning to identify hidden geothermal resources in southwest New Mexico. Fall Conference, American Geophysical Union, San Francisco, CA, December 1–17.

Alexandrov, B. and Vesselinov, V. V. (2014). Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization. *Water Resources Research*, 50(9):7332– 7347.

Alexandrov, B., Vesselinov, V. V., and Djidjev, H. N. (2018a). Non-negative tensor factorization for robust exploratory big data. Technical report, LA-UR-18-20307, Los Alamos National Lab, Los Alamos, NM (United States).

Alexandrov, B., Vesselinov, V. V., Stanev, V. G., and Rasmussen, K. O. (2018b). Tensor decomposition analysis of block copolymer microphase separation. Technical report, LA-UR-18-21907, Los Alamos National Lab, Los Alamos, NM (United States).

Anderson, T. (2013). Geothermal potential of deep sedimentary basins in the United States. In *Unconventional Resources Technology Conference*, pages 1969–1978. Geothermal Resources Council Transaction.

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

Biehler, S. (1971). Gravity studies in the Imperial Valley. *Cooperative Geological-Geophysical- Geochemical Investigations of Geothermal Resources in the Imperial Valley of California: Riverside, California, University of California–Riverside Education Research Service*, pages 29–41.

Brott, C., Blackwell, D., and Ziagos, J. (1981). Thermal and tectonic implications of heat flow in the eastern Snake River Plain, Idaho. *Journal of Geophysical Research: Solid Earth*, 86(B12):11709–11734.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, 36(3):287– 314.

Dobson, P. (2016). A review of exploration methods for discovering hidden geothermal systems. *Geothermal Resources Council Transactions*, pages 695–706.

Faulds, J., Craig, J., Hinz, N., Coolbaugh, M., Glen, J., Earney, T., Schermerhorn, W., Peacock, J., Deoreo, S., and Siler, D. (2018). Discovery of a Blind Geothermal System in southern Gabbs Valley, Western Nevada, through Application of the Play Fairway Analysis at Multiple Scales. *Geothermal Resources Council Transactions*, 42.

Faulds, J. and Hinz, N. (2015). Favorable tectonic and structural settings of geothermal systems in the Great Basin region, western USA: Proxies for discovering blind geothermal systems. In *Proceedings of the World Geothermal Congress*, volume 6.

Faulds, J., Hinz, N., Coolbaugh, M., DePolo, C., Siler, D., Shevenell, L., Hammond, W., Kreemer, C., and Queen, J. (2016). Discovering geothermal systems in the Great Basin region: An integrated geologic, geochemical, and geophysical approach for establishing geothermal play fairways: Proceedings. In *41st Workshop on Geothermal Reservoir Engineering, SGP-TR-209*, volume 15.

Faulds, J., Hinz, N., Coolbaugh, M., Shevenell, L., Siler, D., dePolo, C., Hammond, W., Kreemer, C., Oppliger, G., and Wannamaker, P. (2015). Integrated geologic and geophysical approach for establishing geothermal play fairways and discovering blind geothermal systems in the Great Basin region, western USA: A progress report. *Geothermal Resources Council Transactions*, 39:691–700.

Fleischmann, D. (2006). Geothermal Resource Development Needs in New Mexico. A Publication by the Geothermal Energy Association (GEA) for the U.S. Department of Energy. Technical report, the Geothermal Energy Association.

Fournier, R. (1977). Chemical geothermometers and mixing models for geothermal systems. *Geothermics*, 5(1-4):41–50.

GDR (2015). https://gdr.openei.org/submissions/597, Geothermal Data Repository, Accessed on 2019-12-01.

Kelley, S. (2010). Geothermal energy, lite geology. *New Mexico Bureau of Geology & Mineral Resources, a Division of New Mexico Tech*.

Klema, V. and Laub, A. (1980). The singular value decomposition: Its computation and some applications. *IEEE Transactions on automatic control*, 25(2):164–176.

Kohrn, B., Bonet, C., DiFrancesco, D., and Gibson, H. (2011). Geothermal exploration using gravity gradiometry–a Salton Sea example. *Geothermal Resources Council Transactions*, 35:1699– 1702.

Kratt, C., Coolbaugh, M., Peppin, B., and Sladek, C. (2009). Identification of a new blind geothermal system with hyperspectral remote sensing and shallow temperature measurements at Columbus Salt Marsh, Esmeralda County, Nevada. *Geothermal Resources Council Transactions*, 33:481–485.

Kratt, C., Coolbaugh, M., Sladek, C., Zehner, R., Penfield, R., and Delwiche, B. (2008). A new gold pan for the west: discovering blind geothermal systems with shallow temperature surveys. *Geothermal Resources Council Transactions*, 32:153–158.

Kratt, C., Sladek, C., and Coolbaugh, M. (2010). Boom and bust with the latest 2 m temperature surveys: Dead Horse Wells, Hawthorne Army Depot, Terraced Hills and other areas in Nevada. *Geothermal Resources Council Transactions*, 34:567–574.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791.

Levitte, D. and Gambill, D. (1980). Geothermal potential of west-central New Mexico from geochemical and thermal gradient data. Technical report, Los Alamos Scientific Lab., NM (USA).

Porro, C., Esposito, A., Augustine, C., and Roberts, B. (2012). An estimate of the geothermal energy resource in the major sedimentary basins in the United States. *Geothermal Resources Council Transactions*, 36:1359–1369.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

Siler, D., Faulds, J., Hinz, N., Dering, G., Edwards, J., and Mayhew, B. (2019). Three-dimensional geologic mapping to assess geothermal potential: examples from Nevada and Oregon. *Geothermal Energy*, 7(1):1–32.

Smith, R. (2004). Geologic setting of the Snake River Plain aquifer and vadose zone. *Vadose Zone Journal*, 3(1):47–58.

Vesselinov, V., Ahmmed, B., and M.K., M. (2020a). Unsupervised machine learning to discover attributes that characterize low, moderate, and high-temperature geothermal resources. Geothermal Resources Council, Reno, NV, October 18–23.

Vesselinov, V., Mudunuru, M., Ahmmed, B., S., K., and R.S, M. (2020b). Discovering signatures of hidden geothermal resources based on unsupervised learning. *45th Annual Stanford Geothermal Workshop*.

Vesselinov, V., Mudunuru, M., Karra, S., O'Malley, D., and Alexandrov, B. (2019). Unsupervised machine learning based on non-negative tensor factorization for analyzing reactive mixing. *Journal of Computational Physics*, 395:85 – 104.

Vesselinov, V. V., Alexandrov, B. S., and O'Malley, D. (2018). Contaminant source identification using semi-supervised machine learning. *Journal of contaminant hydrology*, 212:134–142.

Williams, C., Reed, M., DeAngelo, J., and Galanis, S. (2009). Quantifying the undiscovered geothermal resources of the United States. In *Geothermal Resources Council 2009 Annual Meeting*, volume 33.

Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.