LA-UR-21-24126
Approved for public release; distribution is unlimited.

Title: Discovering Hidden Geothermal Signatures using Unsupervised Machine Learning

Author(s): Vesselinov, Velimir Valentinov
Ahmmed, Bulbul
Mudunuru, Maruti K.
Pepin, Jeff D.
Burns, Erick R
Siler, D. L.
Karra, Satish
Middleton, Richard S

Intended for: Report

Issued: 2021-04-28

Disclaimer:
Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Discovering Hidden Geothermal Signatures using Unsupervised Machine Learning

V. V. Vesselinov[1,*], B. Ahmmed[1], M. K. Mudunuru[2], J. D. Pepin[3], E. R. Burns[4], D. L. Siler[5], S. Karra[1], and R. S. Middleton[1]

[1]Computational Earth Science Group, Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545

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

[3]USGS New Mexico Water Science Center, Albuquerque, NM 87113

[4]USGS Oregon Water Science Center, Portland, OR 97201

[5]USGS Geology, Minerals, Energy, and Geophysics Science Center, Moffett Field, CA 94035

[*]**Corresponding author:** Velimir V. Vesselinov, Senior Scientist, Computational Earth Science Group (EES-16), Earth and Environmental Sciences Division (EES), Los Alamos National Laboratory (LANL), Los Alamos, NM - 87545. Email: vvv@lanl.gov

Abstract: Discovering hidden geothermal resources is a very challenging task. It requires the mining of large datasets, including various diverse data attributes representing subsurface hydrogeological and geothermal conditions. The commonly used Play Fairway Analysis (PFA) typically relies on subject-matter expertise to analyze site or regional data to estimate geothermal conditions and prospectivity. Here, we demonstrate an alternative approach based on machine learning (ML) to process a geothermal dataset of Southwest New Mexico (SWNM). The study region includes low- and medium-temperature hydrothermal systems. However, most of these systems are poorly characterized because of insufficient existing data and limited past explorative studies. This study aims to discover hidden patterns and relationships in the SWNM geothermal dataset to better understand regional hydrothermal conditions. This is achieved by applying an unsupervised machine learning algorithm based on non-negative matrix factorization coupled with customized *k*-means clustering (NMF*k*). NMF*k* can automatically identify (1) hidden (latent) signatures characterizing datasets, (2) the optimal number of these signatures, (3) dominant data attributes associated with each signature, and (4) spatial distribution of the extracted signatures. Here, NMF*k* is applied to analyze 18 geological, geophysical, hydrogeological, geothermal attributes at 44 locations in SWNM. NMF*k* successfully finds data patterns and identifies the spatial associations of hydrothermal signatures with the four physiographic provinces in SWNM (Colorado Plateau, Volcanic Field, Basin and Range, and the Rio Grande rift). The algorithm identified up to 5 hydrothermal signatures in the SWNM datasets that differentiate between low- and medium-temperature hydrothermal systems in different provinces. Also, the algorithm identifies two medium-temperature hydrothermal systems in SWNM that require further exploration for geothermal resource development. Based on our analyses, 12 of the attributes are important to identify medium-temperature hydrothermal systems, and the remaining six attributes are critical to characterize low-temperature hydrothermal systems. Based on the obtained results, we identify potential physiographic provinces for further exploration to characterize them as geothermal resources. The resulting NMFk model can be applied to predict geothermal conditions and their uncertainties at new SWNM locations based on limited data from unexplored areas.

Keywords: Geothermal energy, unsupervised machine learning, non-negative matrix factorization, custom *k*-means clustering, feature extraction, hidden (latent) signatures, hidden geothermal resources.

1. Introduction

Typically, hidden hydrothermal systems lie at a few kilometer depths below the ground surface, and they may have deep water tables or being sealed by overlying thick rock that precludes the flow of hot water and heat towards the ground surface (Dobson et al., 2016; Porro 2012; Anderson et al., 2013, Brott, 1981; Williams et al., 2009; Smith et al., 2004). Due to the lack of hydrothermal features at the ground surface, identifying hidden geothermal resources is challenging. A commonly used approach for geothermal exploration of hidden resources is based on Play Fairway Analysis (Siler et al., 2017, 2019; Faulds et al., 2015, 2018, 2019; Lautze et al., 2017, 2020; McClain et al., 2015; Shervais et al., 2015, 2017).

PFA evaluates geothermal prospectivity by assimilating various geological, geophysical, geochemical, and geothermal attributes at site and regional scales. The latter attributes provide direct geothermal evidence and are based on actual measurements such as temperatures at different depths, heat flow, thermal gradients, etc. However, the direct measurements of geothermal attributes are often challenging and expensive to acquire. PFA typically relies on subject-matter expertise to process and analyze the available data and make conclusions about geothermal prospectivity. Ultimately, this can produce bias in the interpretations and limits the amount of data that can be efficiently mined. The general challenges of PFA applications relate to how to (1) unravel unknown relationships between analyzed geothermal data attributes, and (2) identify important easy-to-measure attributes that can be applied to estimate geothermal reservoir properties and prospectivity at new unexplored locations.

To address these challenges, here, we propose an alternative PFA approach that incorporates artificial intelligence and machine learning (ML) methods to process the existing data and to find the hidden relationships in the data without interpretive biases. To achieve this, we need an ML methodology that can efficiently analyze all available site or regional data to learn unknown hidden (latent) relationships between geothermal reservoir properties and other more accessible and cheaper to acquire attributes (e.g., drainage density, shallow groundwater geochemistry, etc.). Also, the methodology should (1) provide a better understanding and robust prediction of geothermal conditions, and (2) allow for discovering hidden geothermal resources without human intervention. This alternative PFA approach is demonstrated here to a geothermal dataset of southwest New Mexico (SWNM). Through this approach, we discover hidden geothermal signatures, their dominant attributes, and the spatial association of each hidden signature. Moreover, the spatial distribution of low- and medium-temperature hydrothermal systems has also been identified.

Section 2 below provides background for this study. In Section 2.1, we discuss various ML methods that can be applied to discover hidden geothermal resources and their strengths and weaknesses. Section 2.2 presents the applied ML methodology called NMFk. Section 2.3 discusses the regional dataset involved in this study. Section 3 summarizes the obtained ML results. Subsection 3.1 describes how the optimal number of signatures is chosen. Subsection 3.2 explains the geothermal significance of the extracted hidden signatures. Subsection 3.3 characterizes two medium-temperature hydrothermal systems that require further exploration. Finally, we present our conclusions in Section 4.

2. Background

2.1 Machine Learning

ML methods, in general, can be subdivided into supervised and unsupervised. The supervised methods require attributes and corresponding labels of the analyzed data (Muller et al., 2016; Rouet-Leduc et al., 2020; Johnson et al., 2021). The labeling should be done by subject-matter experts who can identify, for example, locations with high-, intermediate-, and low-temperature geothermal prospectivity or specific geologic features such as fault offsets. The labeling process can also be automated by unsupervised ML (Muller et al., 2016). The supervised methods are then applied to learn geothermal prospectivity based on the available data. However, the successful training of supervised methods requires large, continuous (without data gaps), non-noisy (with small measurement errors) training datasets that are typically not available for geothermal exploration. Commonly-used supervised methods include deep neural networks (Yoshinki et al., 2014), convolutional neural networks (Gu et al., 2018), recurrent neural networks (Medsker & Jain, 1999), and random forest (Breiman, 2001).

In contrast, the unsupervised ML techniques extract information from existing datasets without any prior labeling or subject-matter preprocessing. The extracted information is post-processed by subject-matter experts to identify the physical meaning of the results. Commonly used unsupervised methods include singular value decomposition (SVD) (Klema & Laub, 1980), principal component analysis (PCA) (Wold et al., 1987), independent component analysis (ICA) (Comon, 1994), *k*-means clustering (Hartigan & Wong, 1979), Gaussian mixture models (Friedman et al., 2001), non-negative matrix/tensor factorization (NMF/NTF) (Lee & Sung, 1999), and non-negative matrix/tensor factorization with customized *k*-means clustering (NMF*k*/NTF*k*; https://smarttensors.github.io; Alexandrov & Vesselinov, 2014; Vesselinov et al., 2018).

One or more unsupervised machine learning methods can be applied for finding hidden patterns in a geologic/geothermal dataset. For example, Watson (2020) utilized *k*-means clustering on infrasound signals to characterize volcanic eruption activity; Anzieta et al. (2019) used *k*-means clustering, correntropy, and dynamic time warping to understand the precursor of the 2015 Cotopaxi volcano eruption. Alexandrov & Vesselinov (2014) and Vesselinov et al. (2018, 2019) applied NMF*k* for blind source separation and extraction of physics insights about complex geologic systems. Unsupervised ML also has been used to characterize hydrothermal systems. For example, PCA and NMF*k* have been used to identify geologic factors that control flow in the Brady, Nevada, geothermal site (Siler & Pepin, 2021; Siler et al., 2021). Ahmmed et al. (2020 a-d) identified hidden geothermal signatures at the Utah FORGE site, the Great Basin, and Hawaii Islands, while Vesselinov et al. (2020 a, b) successfully identified hidden geothermal signatures in eight geothermal datasets of the U.S. geothermal reservoirs. As our recent work suggest, the application of NMF*k* to diverse multi-source, multi-scale, and multi-physics geothermal datasets may lead to the discovery of funknown geothermal signatures. These discovered signatures can be applied to improved detection of hidden geothermal resources.

Here, we applied NMF*k* to analyze an existing SWNM geothermal dataset. NMF*k* is capable of identifying (i) the optimal number of hidden signatures in data, (ii) the dominant set of attributes in data that correspond to identified hidden signatures, and (iii) locations associated with each hidden signature. To discover hidden signatures along with their optimal number in large geothermal datasets, NMF*k* is at the forefront among various unsupervised ML methods such as NMF, PCA, ICA, SVD and its variants, *k*-means clustering, and

Gaussian mixture models. In contrast, with traditional NMF (Lee & Seung 1999), NMF*k* allows for automatic identification of the optimal number of signatures (features) present in the data (Vesselinov et al., 2018). Since the data attributes analyzed here are transformed to be non-negative, NMF*k* preserves non-negativity when extracting hidden signatures. The non-negativity constraint makes the decomposed matrices easier to interpret than PCA, SVD, and ICA because the extracted signatures are additive. Moreover, NMF*k* can handle real, categorical, and missing data (challenging or impossible with other supervised and unsupervised ML methods). Even more importantly, the missing data (some or all of it) can be reconstructed from available data based on the estimated matrix factorization.

2.2 NMF*k*

NMF*k* performs matrix factorization of a data matrix, $X_{m\times n}$, where the *m* rows represent here measurement locations, and the *n* columns are the values of the geothermal attributes. The goal of NMFk is to find the optimal number of signatures k that describe the analyzed dataset. This is accomplished by matrix factorization, which can be represented as using:

$$X \cong W \times H \tag{1}$$

where $W_{m\times k}$ is an "attribute" matrix characterizing the significance of attributes and $H_{k\times n}$ a "location" matrix captures the significance of locations and their spatial association. It is important to note that all the elements of matrices W and H are unknown. The number of signatures k is also unknown. The matrix factorization in (1) provides an approximate representation of the data X. To solve for all the unknowns, NMF*k* performs a series of matrix factorization with random initial guesses for W and H elements and for a range of values of k; theoretically, k can range between 2 and min(m,n). For a given number of signatures k, Equation 1 is solved iteratively by minimizing the reconstruction error $O(k)$:

$$O(k) = || X - W \times H||_F \tag{2}$$

by constraining the W and H elements to be greater or equal to zero and F defines the Frobenius matrix norm (Böttcher & Wenzel, 2008). Under the NMF*k* algorithm, NMF is executed numerous times (typically 1,000), which generates a series of solutions for W and H matrices for a given k value. The resulting multiple solutions of H are clustered into k clusters using a customized k-means clustering. The average silhouette width $S(k)$ based on cosine norm is computed for all k clusters. This metric (see Vesselinov et al. 2018) measures how well the random NMF solutions are clustered for a given value of k. The values of $S(k)$ theoretically can vary from -1 to 1.

These operations are repeated for a series of k values. The optimal number of signatures, k, is estimated on how the reconstruction error$O(k)$ and the average silhouette width, $S(k)$, vary with the increase of k. The reconstruction error decreases as the number of signatures increases. The average silhouette width behavior is more complicated; $S(k)$generally declines as the number of signatures increases from 1 to -1. However, $S(k)$ values frequently spike up for specific k values, indicating that these k values are potentially optimal. In an ideal case, a given k value is considered optimal when adding another signature does not significantly improve the estimate of X (i.e., lower $O(k)$) and does not lower $S(k)$. In practice, a solution with $S(k)$ greater than 0.5 and the lowest $O(k)$ value can be chosen as an optimal solution. The solutions with k values less than the optimal value and S(k) values > 0.5 are acceptable; they provide underfitting representations of the data matrix X. All the solutions with k values greater than the optimal value are not acceptable; they provide overfitting representations of the data matrix X. Implementation of the NMF*k* algorithm and details

related to the selection of the optimal solution are further discussed in Alexandrov and Vesselinov (2014), and Vesselinov et al. (2018, 2019). The NMF*k* results are summarized using different analytical and visual methods discussed in the discussion section below.

2.3 Test dataset

This study analyzes a geothermal dataset of SWNM. SWNM is broadly divided into four physiographic provinces: the Colorado Plateau, the Mogollon-Datil Volcanic Field (MDVF), the Basin and Range, and the Rio Grande rift (Bielicki et al. 2015, 2016; Pepin, 2019). Each physiographic province is associated with different types of unique hydrothermal systems with temperatures ranging from low (<90°C) to medium (90-150°C) (Vesselinov et al., 2020, 2021; Pepin, 2019; Bielicki et al., 2015). Some of the SWNM systems are already utilized for commercial and recreational purposes. At 23 locations, energy-extraction facilities are providing both electricity and direct-use heating. For example, the Basin and Range province has one geothermal power plant (Lightning dock) of gross ~14 MWe power, five greenhouse farms (Kelley, 2010), and numerous medium temperature wells and springs. There are 14 spas and recreational facilities utilizing the SWNM geothermal resources (Kelley, 2010). Recent Play Fairway Analysis (PFA) Phase I study of SWNM revealed more potential geothermal resources (Bielicki et al. 2015, Bennett & Nash, 2017, Levitte & Gambill, 1980).

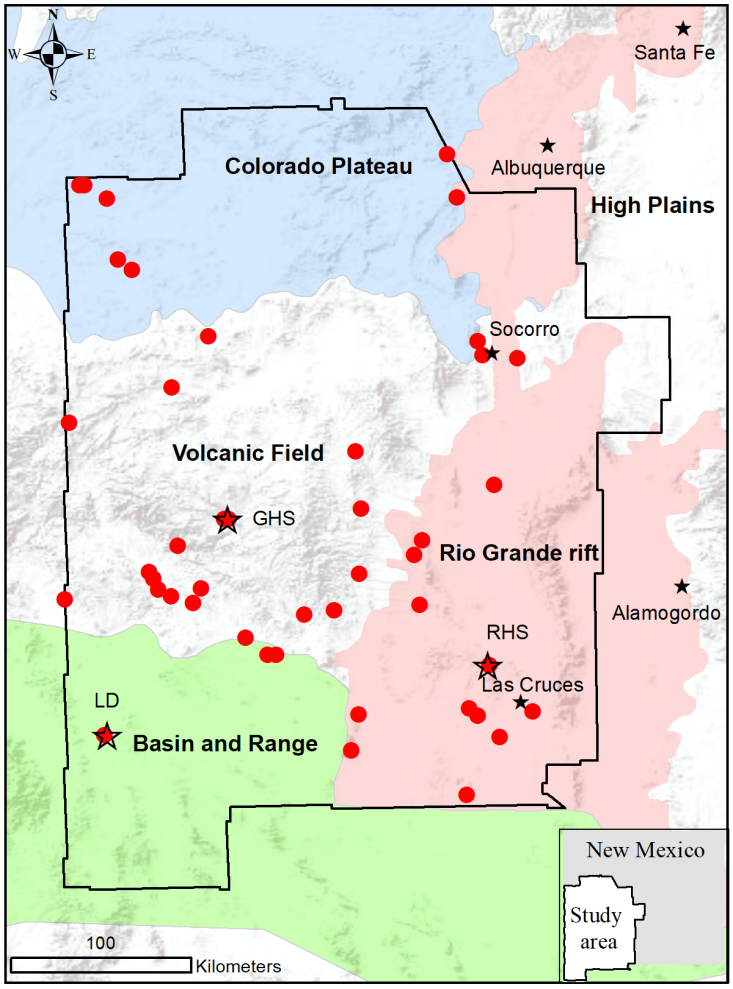

Figure 1: The study area is southwest New Mexico. The red dots show known 44 locations where data were collected for this study (after Pepin 2019). Filled black color stars represent cities, while open stars represent moderate-temperature hydrothermal systems. GHS, LD, and RHS stand for Gila hot springs, lightning dock geothermal plant, and Radium hot springs, respectively.

The analyzed dataset includes two geochemical, two geophysical, five geological, four hydrogeological, and four geothermal attributes (total 18) at 44 locations in SWNM (Figure 1). Each attribute has significance for discovering geothermal signatures, as discussed in detail in Table 1. One of the attributes is the reservoir temperature estimates based on *silica geothermometry* (Bielicki et al., 2015). *Boron* (Bielicki et al., 2015) and *lithium* (Bielicki et al., 2015) are tracer elements in thermal water. *Drainage density* (Bielicki et al., 2015), *spring density* (USGS, 2018a), *hydraulic gradient* (Bielicki et al., 2015), and *precipitation* (Bielicki et al., 2015) are hydrogeological attributes. *Gravity anomaly* (Bielicki et al., 2015), *magnetic intensity* (Bielicki et al., 2015), and *seismicity* (Bielicki et al., 2015) are geophysical attributes. *Silica geothermometer* (Bielicki et al., 2015) temperature and *heat flow* (Bielicki et al., 2015) are geothermal attributes. Geological attributes include *crustal thickness* (Keller et al., 1991), *depth to the basement* (Person et al., 2013), *fault intersection density* (Pepin, 2019), *quaternary fault density* (Pepin, 2019), *state map fault density* (USGS, 2018b), *volcanic dike density* (USGS, 2018b), and *volcanic vent density* (UNM, 2018). The data are preprocessed prior to the ML analyses. The *boron* and *lithium* concentration values are log-transformed to narrow down the distribution of values. The values of each attribute at each location are shown in Table 2. Next, all attributes are rescaled within the range of 0.0 to 1.0 using unit range transformation. To apply NMF*k*, we create a 44x18 matrix ($X_{m\times n}$) where the m = 44 rows represent each location, and the n = 18 columns are the values of the measured attributes.

Table 1. List of geothermal data attributes and their significance for geothermal resource exploration, units, and log-transformation flags.

Attribute number	Attribute	Measurement type	Significance for geothermal resource exploration	Unit	Log transformation
1	*B^+ concentration*	Geochemical	Potentially represents deep heat source	mg/L	Y
2	*Li^+ concentration*	Geochemical	Potentially represents deep heat source	mg/L	Y
3	*Drainage density*	Hydrogeological	Represents the structure of surface water flow; it may also represent groundwater recharge areas and the existence of geologic structures influencing the shape of the drainage network	m/m^2	N
4	*Springs density*	Hydrogeological	Represents the occurrence of conduits of groundwater from depth to the ground surface	m/m^2	N
5	*Hydraulic*	Hydrogeological	Slope of the water table along the	[-]	N

	gradient		direction of maximum head decrease; it characterizes the magnitude of groundwater flow		
6	*Precipitation*	Hydrogeological	The primary source of groundwater recharge	inch	N
7	*Gravity anomaly*	Geophysical	It may represent secondary mineralization. It also characterizes the geologic structure (Beihmer 1971; Kratt et al. 2011)	mGal	N
8	*Magnetic intensity*	Geophysical	May represent secondary mineralization, and in some events, they characterize the geologic structure	A/m	N
9	*Seismicity*	Geophysical	Represents seismotectonic conditions	Richter	N
10	*Silica geothermometer*	Geothermal	The potential temperature of subsurface reservoirs	°C	N
11	*Heat flow*	Geothermal	Represents deep heat source properties	mW/m^2	N
12	*Crustal thickness*	Geological	Represents proximity of the deep heat source (Earth's mantle)	m	N
13	*Depth to the basement*	Geological	Represents the thickness of the potential geothermal reservoir and the depth of a potential deep heat source	m	N
14	*Fault intersection density*	Geological	Represents connection of fault networks; the higher the density, the better for sustainable geothermal play development (Faulds et al. 2018)	m/m^2	N
15	*Quaternary fault density*	Geological	Faults may act as conduits of: (1) groundwater flow water from depth to the ground surface as well as (2) groundwater recharge	m/m^2	N
16	*State map fault density*	Geological	Significance is the same as above	m/m^2	N
17	*Volcanic dike density*	Geological	Represent subsurface manifestation of volcanic events	m/m^2	N
18	*Volcanic vent density*	Geological	Occurrence of volcanic eruptions	m/m^2	N

Table 2. Geothermal data applied for NMF*k* analysis. The table shows a transposed version of the data matrix (X) representing observations of 18 attributes (columns) over 44 locations in the study

area (rows). The values along each column are color coded (green, yellow, and red colors represent minimum, intermediate and maximum values, respectively).

Location	Boron	Gravity	Magnet	Dikes	Drain	Fault	Qfault	Seism	NMFlt	Springs	Vents	Lithium	Precip	Silica	Δh	Qheat	Crust	Bsmt
Alamos Spring	-0.2	-203.3	136.2	0.431	7.4	0.000	0.00	0.004	16.2	0.010	0.003	-3.1	264.8	16.5	5.6	4.6	38.7	1439
Allen Springs	-3.2	-189.3	184.6	3.625	17.3	0.000	0.01	0.002	15.6	0.003	0.001	-4.0	514.5	24.0	13.9	4.4	32.5	51
Apache Tejo Warm Springs well	-1.8	-181.2	15.0	3.807	17.3	0.001	0.03	0.001	0.7	0.003	0.000	-8.6	326.3	52.0	4.7	4.6	30.7	24
Aragon Springs	1.5	-229.1	-317.7	0.010	19.0	0.000	0.00	0.000	41.1	0.005	0.003	-7.5	387.0	56.5	4.0	4.5	38.8	1486
Ash Spring	-2.7	-193.2	66.6	4.914	17.0	0.000	0.00	0.002	9.3	0.003	0.000	-5.0	492.0	29.3	4.1	4.4	32.2	-92
B. Iorio 1 well	-2.1	-196.5	-48.2	1.936	18.8	0.057	21.02	0.000	9.1	0.003	0.003	-2.6	260.4	59.4	0.9	4.0	30.9	-188
Cliff Warm Spring	-2.5	-199.1	-47.1	1.290	22.8	0.001	2.58	0.002	11.0	0.002	0.001	-6.9	364.2	64.2	1.8	4.2	33.1	-191
Dent windmill well	-2.1	-230.8	89.3	0.000	13.4	0.000	0.00	0.000	0.0	0.005	0.000	-7.3	341.7	19.7	2.4	4.7	43.5	865
Derry Warm Springs	-1.5	-161.6	197.0	0.659	18.3	0.007	9.16	0.000	15.9	0.002	0.000	-7.5	276.1	37.4	3.0	4.6	30.0	-120
Faywood Hot Springs	-2.6	-172.1	-49.8	0.939	16.6	0.002	2.81	0.000	1.9	0.003	0.000	-4.8	346.4	67.2	4.2	5.5	30.0	619
Federal H 1 well	-0.4	-132.0	35.0	0.000	5.8	0.004	20.31	0.001	7.2	0.000	0.015	-5.0	253.8	78.7	2.7	4.9	27.3	2906
Freiborn Canyon Spring	-2.5	-225.0	-242.0	0.401	13.1	0.000	0.00	0.001	19.8	0.001	0.004	-12.6	538.6	49.8	13.0	4.6	38.4	1138
Garton well	-3.2	-196.8	35.6	0.150	18.0	0.000	0.00	0.000	28.9	0.002	0.001	-5.0	489.9	70.0	4.3	3.9	30.9	-266
Gila Hot Springs 1	-1.9	-221.6	-149.3	0.127	24.2	0.000	0.00	0.001	25.5	0.003	0.003	-7.8	422.6	69.9	6.6	4.4	34.0	413
Gila Hot Springs 2	-1.8	-222.9	-138.8	0.112	24.7	0.000	0.00	0.001	23.7	0.003	0.003	-6.7	425.9	70.8	3.2	4.6	33.9	519
Goat Camp Spring	-2.1	-159.2	-29.7	0.751	10.0	0.001	2.22	0.007	10.6	0.002	0.001	-8.0	344.0	68.9	5.8	4.4	32.4	19
Jerry well	-0.8	-219.6	172.4	0.111	15.5	0.000	0.00	0.000	6.3	0.004	0.005	-7.9	243.9	13.4	1.0	4.4	42.3	1190
Kennecott Warm Springs well	-2.4	-178.3	-69.9	1.422	17.8	0.002	1.76	0.000	1.1	0.003	0.000	-6.9	355.0	66.1	4.3	5.0	30.0	409
Laguna Pueblo	0.4	-204.2	62.5	0.406	8.6	0.004	4.58	0.006	14.6	0.018	0.005	-3.3	259.7	42.9	2.6	4.4	37.2	1506
Lightning Dock	-1.0	-168.0	-168.1	0.086	4.6	0.008	8.40	0.002	4.3	0.000	0.000	-3.9	291.5	107.3	0.8	5.0	29.8	1800
Los Alturas Estates	-1.5	-141.4	-127.5	0.004	7.6	0.003	0.05	0.002	6.6	0.001	0.000	-12.7	265.3	71.9	2.2	6.3	27.4	4321
Mangas Springs	-2.6	-201.0	-227.1	3.503	20.2	0.000	0.91	0.002	11.5	0.002	0.000	-4.5	393.5	53.6	0.3	4.2	32.4	-178
Mimbres Hot Springs	-2.3	-200.6	43.4	0.670	15.4	0.002	1.13	0.000	19.0	0.004	0.000	-3.8	445.9	68.3	9.1	4.9	31.0	50
Ojitos Springs	-1.6	-202.1	-7.5	1.342	19.6	0.044	19.74	0.037	31.0	0.020	0.005	-4.5	257.5	57.6	7.2	4.5	33.0	-255
Ojo Caliente	-2.6	-226.5	-168.4	0.000	20.5	0.000	0.00	0.000	8.3	0.004	0.000	-2.9	333.6	48.4	3.5	5.5	33.8	2415
Ojo De las Canas	-1.7	-188.5	-85.8	0.839	22.3	0.036	12.55	0.036	28.0	0.013	0.003	-6.0	270.5	14.2	4.0	4.5	31.8	101
Pueblo windmill well	-1.2	-228.8	315.9	0.029	15.2	0.000	0.00	0.000	6.1	0.004	0.003	-12.0	265.8	18.3	2.9	4.3	42.5	1027
Radium Hot Springs	-0.8	-151.4	-7.8	0.010	8.8	0.013	11.40	0.003	10.6	0.001	0.000	-5.3	264.2	63.6	0.3	5.4	28.2	1191
Rainbow Spring	-1.7	-227.1	-48.5	0.000	11.0	0.000	0.00	0.001	0.0	0.006	0.000	-7.0	307.8	21.7	3.3	4.7	43.9	755
Riverside Store well	-1.3	-196.1	-102.9	1.562	22.6	0.000	2.50	0.002	11.7	0.002	0.001	-2.4	356.1	60.8	0.9	4.3	32.9	-165
Sacred Spring	-1.8	-228.4	-80.4	0.000	10.9	0.000	0.00	0.001	0.0	0.006	0.000	-7.0	298.4	21.2	1.3	4.6	43.9	742
Socorro Canyon	-1.8	-204.7	-136.5	1.203	21.1	0.051	28.88	0.034	33.8	0.020	0.005	-6.7	284.1	44.6	11.1	5.0	32.6	-229
Spring	-4.1	-183.5	334.5	0.218	20.1	0.011	1.81	0.000	20.1	0.001	0.006	-6.8	361.9	117.2	5.1	3.8	31.5	-104
Spring Canyon Warm Spring	-2.1	-194.2	117.3	2.293	21.9	0.000	1.50	0.002	12.7	0.002	0.000	-8.3	361.7	51.6	5.8	4.2	32.6	-57
Truth or Consequences spring	-1.1	-168.2	-54.3	2.175	18.4	0.064	20.51	0.000	10.3	0.003	0.002	-3.3	265.9	55.3	0.6	4.3	31.0	304
Turkey Creek Spring	-3.2	-196.4	54.8	0.984	19.2	0.001	3.69	0.002	28.1	0.002	0.002	-3.7	493.4	81.3	5.8	4.4	33.6	56
Victoria Land and Cattle Co. well	-1.8	-165.9	-65.4	0.478	6.4	0.003	0.06	0.001	0.9	0.001	0.000	-2.9	253.0	43.0	1.9	4.1	30.7	2014
Warm Springs	-2.1	-193.3	113.5	0.220	19.0	0.029	2.63	0.000	16.5	0.004	0.003	-2.5	314.6	56.0	5.4	4.3	32.7	1252
Well 1	-1.4	-230.7	-31.3	1.190	15.7	0.000	0.75	0.001	22.1	0.004	0.002	-6.6	345.4	49.0	1.7	4.4	40.0	1961
Well 2	-1.2	-162.5	0.8	0.000	4.5	0.008	24.24	0.003	11.8	0.000	0.006	-10.1	279.5	70.5	1.7	4.8	27.8	2993
Well 3	-2.5	-140.0	31.7	0.839	2.1	0.001	2.11	0.001	5.0	0.001	0.000	-7.3	369.0	51.0	4.1	4.3	28.0	3073
Well 4	-1.3	-161.7	-56.1	0.000	3.4	0.008	28.49	0.003	10.6	0.000	0.006	-10.0	274.3	94.0	1.9	4.7	27.7	3373
Well 5	-1.9	-167.2	-29.9	0.000	2.5	0.008	15.48	0.002	3.1	0.000	0.005	-6.8	243.8	47.0	0.3	4.0	27.4	5460
Well south of Carne	-2.4	-156.7	-129.6	0.457	4.3	0.000	2.11	0.002	6.0	0.001	0.000	-6.8	269.7	87.1	1.4	4.5	28.4	2761

3. Results and Discussion

3.1 Identification of the Optimal Number of Signatures

For the problem analyzed here, the reconstruction error, $O(k)$, decreases as the number of signatures increases (Figure 2). However, the average silhouette width, $S(k)$, does not follow this pattern. $S(k)$ fluctuates over the number of signatures, as shown in Figure 2. In general, solutions with $S(k)$greater than 0.5 can be considered to be acceptable. Based on this criteria, the solutions for k=2, 3, 4, and 5 were accepted, while the $k > 5$ solutions were rejected by the algorithm. The solutions for k > 5 are overfitting the analyzed dataset. The solution for k=5 is the optimal one because of its reasonably low $O(k)$ and high$S(k)$values. Still, there are general consistencies in the solutions for k=2, 3, 4, 5, and 8, where all these solutions have relatively high $S(k)$value (>0.5). The relations between signatures extracted by these solutions are further discussed in Appendix A. The analyses in the appendix confirm our conclusion that the k=5 solution is optimal for the studied problem. However, all acceptable solutions (for k=2, 3, 4, and 5) can be applied to describe the dataset. Furthermore, the solution with the optimal number of signatures is

expected to provide the best physical interpretability of the analyzed data matrix. Below, we focus on the spatial association of the extracted signatures within the study area.

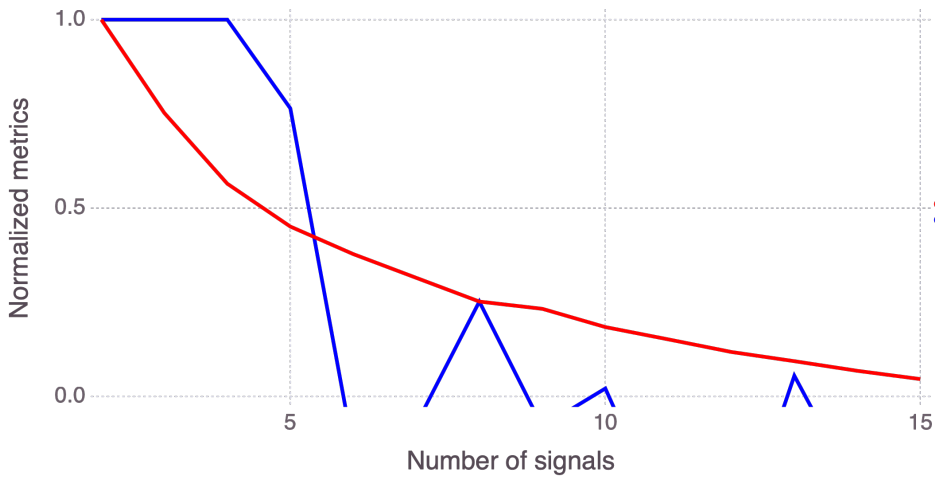

Figure 2: NMF*k* results for normalized reconstruction error (fit) $O(k)$in red color and solution robustness (based on the average silhouette $S(k)$ width of the clusters) in blue color for different numbers of signatures k.

Figure 3 shows the predominant association of the 44 measurement locations with the extracted geothermal signatures for solutions with *k*=2, 3, 4, 5, and 8.

The NMF*k* solution for *k*=2 separates the Colorado Plateau and the Volcanic Field (Signature A) from the Basin and Range and the Rio Grande Rift (Signature B) provinces (Figure 3a).

The *k*=3 solution combines the Colorado Plateau and the Volcanic Field in Signature A; however, Signatures B and C separate the Basin and Range and the Rio Grande Rift provinces, respectively (Figure 3b).

Signature A of the *k*=4 solution (Figure 3c) represents the Volcanic Field. Signature B captures the Basin and Range province. Signature C coincides with the Colorado Plateau. Signature D encompasses the Rio Grande Rift zone (Figure 3c).

The *k*=5 solution (Figure 3d), regrouped the four signatures of the *k*=4 solution into five signatures. Signatures A and E cover MDVF; Signatures B, C, and D capture the remaining three provinces: the Basin and Range, the Colorado Plateau, and the Rio Grande rift provinces, respectively (Figure 3d). Signature A encompasses the area below the Jemez lineament (we call it the southern MDVF). In contrast, Signature E covers the Jemez lineament and its contiguous north area (we call it the northern MDVF).

In the *k*=8 solution (Figure 3e), Signature B captures the Colorado Plateau province. Signatures G and H encompass two separate areas in the Rio Grande rift zone (Figure 3e). Signatures A, C, and D capture the spatial variability of geothermal conditions within MDVF, while Signatures E and F do the same for the Basin and Range province.

The solution progression with increasing k demonstrates the power of our method to further refine the spatial characterization of the analyzed geothermal data. There is a general consistency between all the acceptable solutions (k=2, 3, 4, 5, and 8). The solutions for k = 2, 3, and 4 provide a higher-level generalization of the geothermal signatures, while the k=8 solution allows us to further refine the characterization of the extracted geothermal signatures. It should be noted that none of the signatures of the k=2, 3, 4, 5, and 8 solutions perfectly represent physiographic associations. It appears that some of the locations outside but in close vicinity of a given province have associated signatures. This observation signifies that the extracted signatures are less distinct in terms of their physiographic province association as the number of signatures increases. However, the addition of more signatures in the NMFk solutions refines the characterization of geothermal characteristics in the study area; therefore, the solutions for k=2, 3, 4, 5, and 8 can be applied to explain the dataset. Because the k=5 solution captures sufficiently well characteristics of both hydrothermal systems and physiographic provinces, this confirms our analyses above that this solution can be selected to be the optimal one.

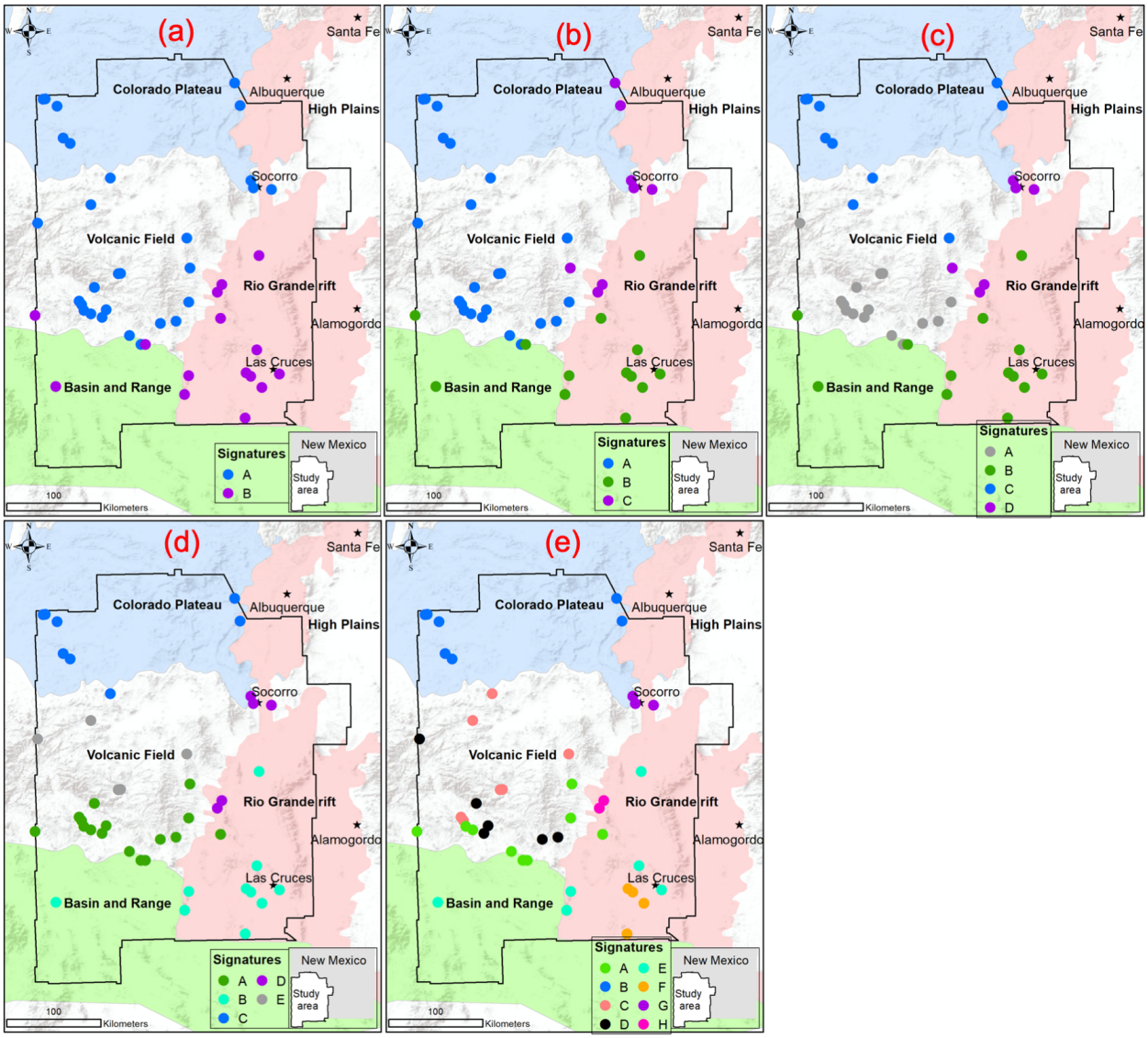

Figure 3: Spatial distribution of signatures associated with the NMF*k* solutions for signature (a) 2, (b) 3, (c) 4, (d) 5, and (e) 8.

3.2 Interpretation of the NMFk solution with the optimal number of signatures

This subsection provides a high-level interpretation of each signature in the optimal *k*=5 solution (Figure 4-5). We categorize each signature as low- or medium-temperature hydrothermal systems based on the contribution of *silica geothermometer* in each signature: low and high silica values define low- and medium-temperature systems, respectively). We also describe how geology, hydrogeology, and geothermal attributes relate to each other in each signature and how they define the hydrothermal systems within the study area.

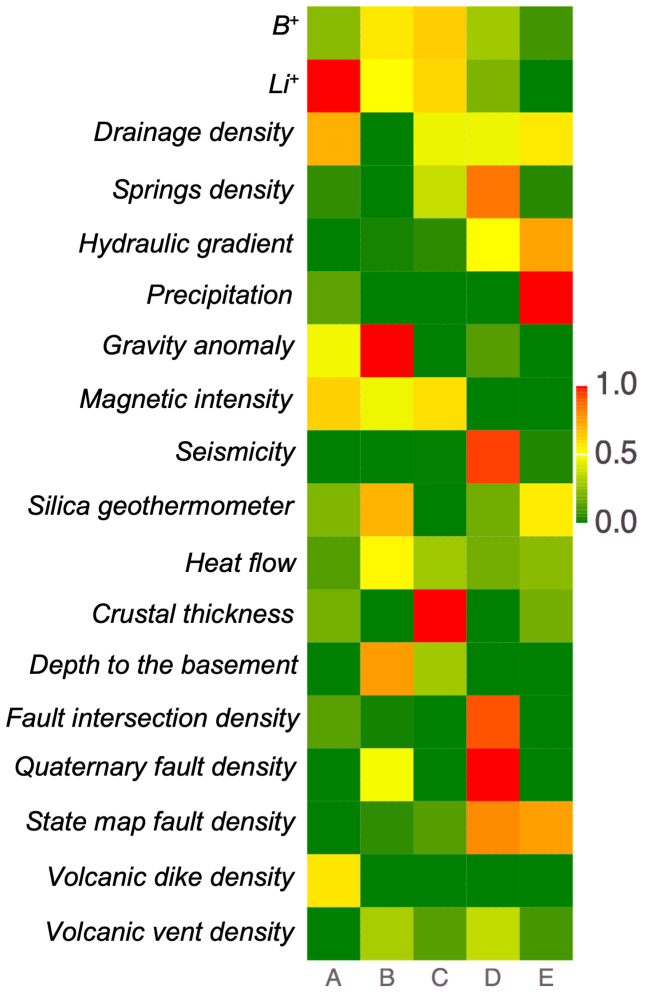

Figure 4: NMF*k* attribute matrix for the *k*=5 solution. High-value matrix entries (red) define high significance, while low-value matrix entries (green) represent low significance. High and low values of *silica geothermometer* indicate medium- and low-temperature geothermal resources, respectively.

Table 3: Geothermal hidden signatures and their association with a geothermal resource type, physical significance, dominant data attributes (Figure 4), and physiographic provinces (Figures 1 and A-1).

Signature	Hydrothermal system type	Dominant attributes	Physical significance	Physiographic province
A	Low temperature	*Gravity anomaly* *Magnetic intensity* *Volcanic dike density* *Drainage density* *Li^+ concentration*	Shallow heat flow	Southern MDVF
B	Medium temperature	*B^+ and Li^+ concentrations* *Gravity anomaly* *Magnetic intensity* *Quaternary fault density* *Silica geothermometer* *Heat flow* *Depth to the basement*	Deep heat flow	Rio Grande rift
C	Low temperature	*B^+ and Li^+ concentrations* *Magnetic intensity* *Drainage density* *Crustal thickness*	Deep heat source	Colorado Plateau
D	Low temperature	*Drainage density* *Fault intersection density* *Seismicity* *State map fault density* *Spring density* *Hydraulic gradient*	Tectonics	Rio Grande rift
E	Medium temperature	*Drainage density* *State map fault density* *Precipitation* *Silica geothermometer* *Hydraulic gradient*	Vertical hydraulics	Northern MDVF

Signature A is potentially representative of low-temperature hydrothermal systems because of the low contribution of *silica geothermometer*. This signature's dominant attributes are *gravity anomaly*, *magnetic intensity*, *volcanic dike density*, *drainage density*, and *Li^+ concentration* (Table 3). *Volcanic dike density*, *gravity anomalies*, *magnetic intensity* indicate the manifestation of plutonic mafic rocks due to Tertiary volcanic events (Nakai et al., 2017, Figure 5). The locations associated with Signature A are in the southern MDVF. This portion of the MDVF has a history of active volcanism in the past (Cather, 1990; McIntosh et al. 1992; Chapin et al. 2004; Ratté & Grotbo, 1979) that might have further enhanced *Volcanic dike density* and secondary mineralization. The resultant secondary mineralization is expected to elevate *gravity anomaly* and *magnetic intensity* in this region. The dominant attributes except *drainage density* are

indicators that the hydrothermal systems in this region are prospective geothermal resources. Yet, NMF*k* did not diagnose the locations associated with Signature A as medium-temperature hydrothermal systems because of low *silica geothermometer* contribution to this geothermal signature. A possible explanation might be the lack of high-temperature gradient in this area due to increased infiltration caused by high *drainage density*.

Signature B potentially represents medium-temperature hydrothermal systems because of the high contribution of *silica geothermometer*. The other dominant attributes of this signature are B^+ and Li^+ *concentrations*, *gravity anomaly*, *magnetic density*, *quaternary fault density*, *heat flow*, and *depth to the basement* (Table 3). *Heat flow* and *depth to the basement* are unique dominant attributes of this signature. *Heat flow* is also an indicator of temperature gradient, while *depth to the basement* is an indicator of a high reservoir depth and a large distance of the heat source distance from the ground surface. The locations associated with Signature B fall in the southern Rio Grande Rift; there is also one location in the Basin and Range province, suggesting an extension of this rift signature within the Basin and Range. The area covered by Signature B went through frequent Tertiary and some Quaternary volcanic events (Nakai et al., 2017). Therefore, it is not surprising that *magnetic intensity*, *gravity anomaly*, and *volcanic dike density* are dominant attributes. This area also has a low crustal thickness (Elston et al. 1976; Nakai et al. 2017; Olson, 1979; Sanford, 2002), which indicates that this area is also closer to the mantle heat source. *Depth to the basement* is the deepest in the study area that may assist in entrapping the heat originating from the mantle. The high-temperature gradient, deep basement, and lower-crustal thickness may be the potential cause of the medium-temperature hydrothermal systems in this region. Further field explorations and data collection activities are required to better define the locations associated with this geothermal signature and associated hydrothermal resources.

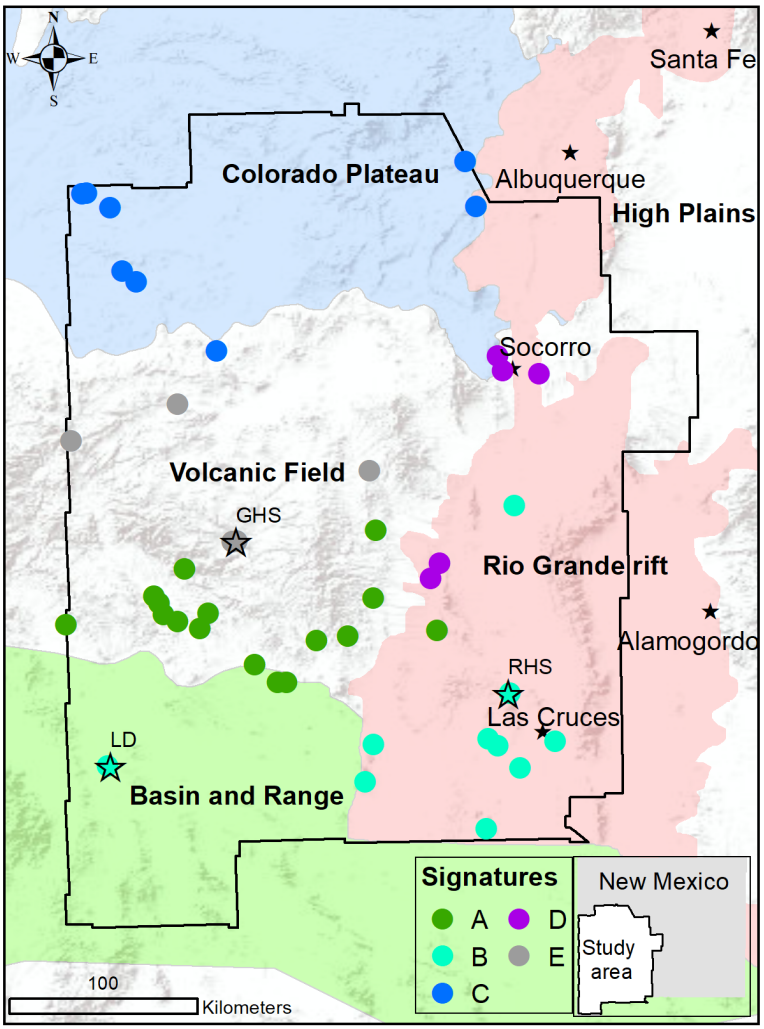

Figure 5: Locations associated predominantly with the five signatures A, B, C, D, and E. Filled black color asterisks represent cities while open asterisks represent moderate-temperature hydrothermal locations. GHS, LD, and RHS stand for Gila hot springs, Lightning Dock geothermal plant, and Radium hot springs, respectively.

Signature C represents low-temperature hydrothermal systems because of the low contribution of *silica geothermometer*. The dominant attributes of this signature are B^+ and Li^+ *concentrations*, *magnetic intensity*, *drainage density*, and *crustal thickness* (Table 3). B^+ and Li^+ may be released from the subsurface due to the nearby heat source, while *magnetic intensity* may indicate secondary mineralization due to Tertiary volcanic events, which may produce plutonic mafic rocks (Hunt, 1956; Thompson & Zoback, 1979; Lucchitta, 1979). These three attributes suggest potential heat sources at depth. *Drainage density* and *crustal thickness* are a unique combination of attributes for this signature, indicating that it might represent lateral hydraulics. The locations associated with Signature C are within the Colorado Plateau. However, the high significance of B^+ and Li^+ *concentrations*, *magnetic intensity*, *drainage density* in this signature are good indicators of geothermal resources. NMF*k* did not designate this signature as medium-temperature hydrothermal systems due to low *silica geothermometer* impact on this geothermal signature . The large

crustal thickness may preclude heat flow in this region that might be a potential explanation for the designation as low-temperature hydrothermal systems.

Signature D represents low-temperature hydrothermal systems because of the low contribution of *silica geothermometer*. The dominant attributes of this signature are *drainage density*, *fault intersection density*, *seismicity*, *state map fault density*, *spring density*, and *hydraulic gradient* (Table 3). *Fault intersection density*, *seismicity*, and *state map fault density* suggest that this signature represents tectonic features. The locations associated with this Signature D are in the Rio Grande Rift and the Jemez lineament, which went through both extension and subduction tectonic events (Nakai et al., 2017; Olson, 1979; Sanford, 2002). Active tectonic events increase *fault intersection density*, which increases *drainage density*, *fault intersection density*, *spring density*, and *hydraulic gradient*. Also, *seismicity* indicates the presence of active faults. This signature did not get high contribution from attributes, which are well indicators of medium-temperature hydrothermal systems such as B^+, Li^+, *gravity anomaly*, *magnetic intensity*, *heat flow*, and *silica geothermometer*. The dominant attributes indicate that the locations associated with this signature have good groundwater flow characteristics. They do not necessarily indicate amenable to being good hydrothermal systems for further exploration.

Signature E represents medium-temperature hydrothermal systems because of the high contribution of *silica geothermometer*. The remaining dominant attributes of this signature are *drainage density*, *state map fault density*, *precipitation*, and *hydraulic gradient* (Table 3). *State map fault density* and *hydraulic gradient* represent deep flow circulation, meaning that this signature may capture vertical groundwater flow characteristics. The locations associated with this signature are in or around the Jemez lineament within the northern part of the MDVF. The Jemez lineament went through rigorous Tertiary and Quaternary volcanism events. Also, this region has an intensive tectonic history when compared to the surrounding areas. Both high volcanism and tectonic events make this area more amenable to have geothermal resources. In addition, this area has high precipitation, which increases deep flow circulation. A fluid circulation from the depth where hot water exists is a good indicator of good hydrothermal systems. Another possible explanation is that crustal thickness is lower than the Colorado Plateau (Cather, 1990; McIntosh et al., 1992; Chapin et al., 2004; Ratte, 1989; Pepin, 2019). Another critical factor is that the *depth to the basement* is high, although it is not visible in this signature. Deep *depth to the basement* along with deep fluid circulation characteristics may be the potential cause of medium-temperature hydrothermal systems in the northern MDVF. Further field explorations and data collection activities are required to better define the locations associated with this geothermal signature and associated hydrothermal as geothermal resources.

Based on the discussion presented above, among 18 analyzed attributes, the 12 dominant attributes related to medium-temperature hydrothermal systems are B^+ and Li^+ *concentrations*, *silica geothermometer*, *heat flow*, *gravity anomaly*, *magnetic intensity*, *quaternary fault density*, *state map fault density*, *depth to the basement*, *drainage density*, *precipitation*, and *hydraulic gradient*. All of these attributes are related to geothermal signatures B and E (Table 3). The remaining six attributes are *volcanic dike density*, *volcanic vent density*, *fault intersection density*, *springs density*, *crustal thickness*, and *seismicity*. These attributes are dominant in the signatures A, C and D (Table 3), which represent low-temperature hydrothermal systems.

3.3 Characterization of Medium-temperature Hydrothermal Systems

The subsection provides a high-level explanation of the uniqueness of two medium-temperature

hydrothermal systems designated by our NMF*k* analyses. NMF*k* extracts geothermal signatures and also estimates the significance of attributes and locations to define these signatures. To show the correlations between the signatures, a series of biplots are generated by pairing each of the extracted signatures. In this analyses, the signatures are viewed as basis vectors similarly to how eigenvectors under PCA analyses are defined (Wold et al., 1987). A biplot is an exploratory scatterplot showing the mutual relation between two signatures based on how attributes and locations associated with these signatures are weighted. In a biplot, an attribute is well correlated if its significance is high for both signatures. In contrast, some attributes might be important for one signature but not for another signature. A biplot is also a good indicator if an attribute is not critical for both signatures; these attributes will be located close to the plot origin. In a biplot, well-correlated attributes lie on the diagonal between the two axes and away from the origin, while uncorrelated attributes lie close to the axes.

Figure 6 demonstrates the correlation of attributes between two medium-temperature geothermal Signatures B and E. Here, the only correlated attribute is the *silica geothermometer*, which classifies these signatures as medium-temperature hydrothermal systems. This lack of correlation among the other geothermal attributes reveals that they are uniquely associated with Signatures B and E, respectively. Because the geology of each province is unique, their controls on hydrothermal systems also vary. Signature B falls mainly in the southern Rio Grande Rift zone. The hydrothermal system of this area is primarily defined by *gravity anomaly*, *depth to the basement,* B^+ and Li^+ *concentrations, and heat flow*. On the other hand, Signature E falls in the northern MDVF, which is in the northern portion of the study area. Hydrothermal systems in this area are defined by *precipitation*, *hydraulic gradient, state map fault density*, *state map fault density*, and *drainage density*.

As mentioned above, the geological, geophysical, hydrological, and geothermal characteristics of Signatures B (the northern MDVF) and E (Rio Grande rift) are distinct. Both areas went through Tertiary and Quaternary volcanisms, but the northern MDVF went through more frequent volcanic events than the Rio Grande rift zone. Also, the northern MDVF is tectonically more active than the Rio Grande Rift zone. However, a tectonic extension feature is present between the western and easter portions of the Rio Grande rift zone that is absent in the northern MDVF. Moreover, the Rio Grande Rift zone has a lower crustal thickness than the northern MDVF. All these observations demonstrate the unique geological and hydrological characteristics of the two regions. Therefore, they represent unique hydrothermal systems, and these differences were successfully captured by NMF*k* in the extracted geothermal signatures.

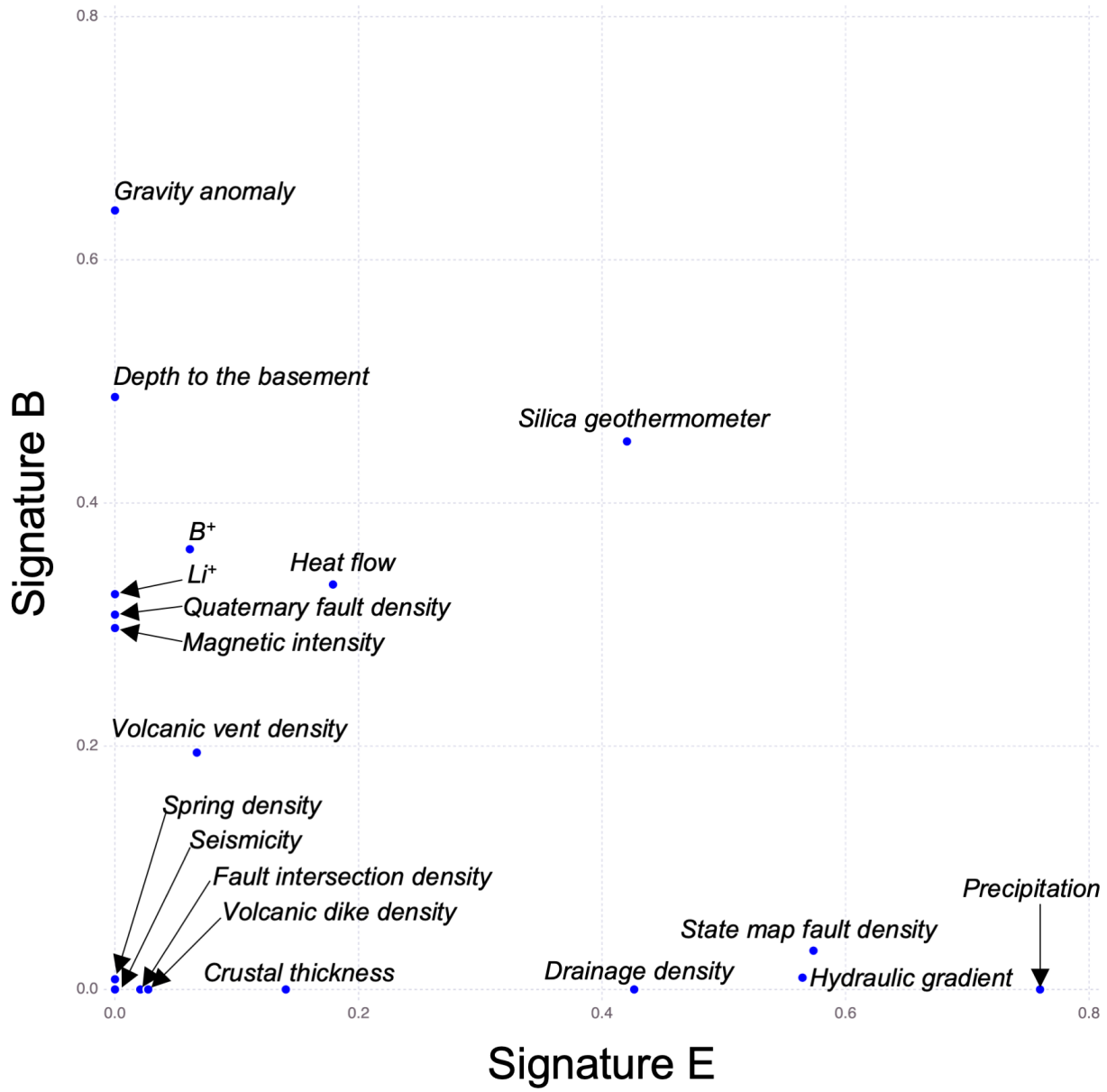

Figure 6: Biplot showing the correlation between attributes for medium-temperature resources as defined by Signatures B and E (Table 3). Attributes that fall along the axes are not correlated. If attributes lie near the origin, they are low correlated; if diagonally away from the origin, they are highly correlated. Arrows indicate the locations of attributes.

4. Conclusions

Using an unsupervised ML, this study characterized a geothermal dataset of SWNM to (1) identify hidden geothermal signatures, (2) estimate the optimal number of signatures, (3) find dominant attributes associated with each signature, (4) map spatial association of the signatures. Based on the obtained results, we identify potential physiographic provinces for further exploration to characterize them as geothermal resources. The dataset consists of 18 geothermal attributes measured at 44 locations. The locations represent

hydrothermal systems within the four physiographic provinces present within the SWNM study domain. On purpose, the analyses did not include the labeling of the hydrothermal systems based on their type and their association with physiographic provinces. In this way, we tested if the algorithm can blindly group the locations based on their type and province association just utilizing the provided data about observed geothermal attributes. The applied unsupervised ML tool is called NMF*k,* and our analyses extracted a series of geothermal signatures. The solutions for the number of signatures, *k,* equal to 2, 3, 4, 5, and 8, can explain the dataset with respect to low/medium-temperature resources. Among these solutions, the *k*=5 result provides the best characterization of the SWNM geothermal conditions. The optimal number of signatures is automatically identified by the algorithm. The five signatures under this solution are labeled as A, B, C, D, and E, and based on our analyses, they directly associate with the geothermal prospectivity of the SWNM study area.

Each signature of *k*=5 solutions has been categorized as low- or medium-temperature hydrothermal systems based on the contribution of *silica geothermometer* on the corresponding signature. In summary, Signature A represents low-temperature hydrothermal systems, and it includes locations in the southern MDVF. The dominant attributes are *gravity anomaly*, *magnetic intensity*, *volcanic dike density*, *drainage density*, and Li^+ *concentration* that are expected to characterize shallow heat flow. Signature B depicts medium-temperature hydrothermal systems, and it covers locations in the southern Rio Grande Rift zone and the Basin and Range province. The dominant attributes are B^+ and Li^+ *concentrations*, *gravity anomaly*, *magnetic density*, *quaternary fault density*, *silica geothermometer*, *heat flow*, and *depth to the basement* that potentially indicate deep heat flow. Signature C defines low-temperature systems, and it captures locations in the Colorado Plateau. The dominant attributes are B^+ and Li^+ *concentrations*, *magnetic intensity*, *drainage density*, and *crustal thickness* that likely demonstrate the existence of deep heat sources. Signature D represents low-temperature hydrothermal systems, and it covers locations in the Rio Grande Rift and the Jemez lineament in MDVF. The dominant attributes are *drainage density*, *fault intersection density*, *seismicity*, *state map fault density*, *spring density*, and *hydraulic gradient* that predominantly capture the occurrence of tectonic activities and the potential of upward groundwater flow. Signature E is associated with medium-temperature hydrothermal systems, and it covers the northern MDVF. The dominant attributes are *drainage density*, *state map fault density*, *precipitation*, *silica geothermometer*, and *hydraulic gradient* that likely portray vertical downward (recharge) fluid flow.

Out of five signatures, only two signatures (B and E) are associated with medium-temperature features. Those two signatures are connected to a heat source such as Signature B is located mainly in the Rio Grande Rift zone, including one location in the Basin and Range province where there is a high heat gradient. It is critical to mention here that only powerplant in New Mexico is located in the Basin and Range province. This successful identification of correct hydrothermal system types without prior knowledge demonstrates the usefulness of the proposed ML methodology based on NMF*k*. Signature E, another medium-temperature geothermal signature, is located in the northern MDVF, where heat and hot groundwater may originate from the depth and reach the ground surface through faults.

The northern MDVF and the Rio Grande rift zones required further exploration to designate them as geothermal resources. The PFA work by Bielicki et al. (2015) generated a preliminary geothermal prospectivity map. In the future, we will combine PFA results and the knowledge accumulated in this study to make an ML-enhanced geothermal prospectivity map of the SWNM region. This map will assist in discovering hidden resources and their accurate locations for geothermal heat extraction using well drilling.

To conclude, the extracted dominant signatures by NMF*k* indicate dominant attributes to identify hydrothermal systems in each province. Moreover, the proposed NMF*k* analysis is widely applicable to extract signatures (signals) from large-scale geothermal data (including observational and simulation outputs). This broad applicability of our ML tools makes it attractive for researchers in the geothermal industry and institutions to use these tools to discover, quantify, and assess hidden geothermal energy resources. Our algorithms are open source, and examples, test problems, notebooks, and documentation are available at https://smarttensors.github.io

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). Support for Jeff Pepin, Erick Burns, Drew Siler, and Jake DeAngelo was provided by the U.S. Geological Survey (USGS) Energy Resources Program.

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.

Appendix A: Discussion of the NMF*k* solutions for different number of signatures

NMF*k* analyses provided solutions for a different number of signatures. The optimal number of signatures is equal to 5, as discussed in Section 3.1. However, there is a general consistency between the extracted geothermal signatures. Here, we demonstrate these consistencies in the solutions for k=2, 3, 4, 5, and 8; all these solutions have relatively high $S(k)$values (>0.25). The solutions for k=2, 3, and 4 provide a higher-level generalization of the geothermal signatures (Figure 3), while the k=8 solution allows us to further refine the characterization of the extracted geothermal signatures (Figure 3).

For these 5 solutions, the locations associated with each geothermal signature share a portion of a physiographic province or neighbouring physiographic provinces. If more than one geothermal signatures are within a given physiographic province, they either characterize a spatial complexity or hydrothermal impacts from adjacent provinces (Figure 3).

The k=2 solution subdivides the region into two groups (Figure A-1a). Signatures A and B of the k=3 solution (Figure A-1b) are split up into Signatures A, B, and C of the k=4 solution (Figure A-1b). Signature

C for *k*=3 (Figure A-1c) and Signature D for *k*=4 (Figure A-1b) share similar very properties. Signatures A, B, C, and D of both the *k*=4 and 5 solutions (Figure A-1b,d) also possess similar properties. However, the *k*=5 solution got an entirely new signature (Signature E) (Figure A-1d). The *k*=8 solution (Figure A-1e) regrouped the *k*=5 solution (Figure A-1d). Signature A of the *k*=5 solution possesses similar properties to Signatures A and D of the *k*=8 solution. Signature B of the *k*=5 solution shares similar values to Signatures E and F of the *k*=8 solution. Signature C of the *k*=5 solution has similarities to both Signatures B and C of the *k*=8 solution. Signature D for *k*=5 and both Signatures G and H for *k*=8 also have similar values. These associations among signatures for *k*=2, 3, 4, 5, and 8 solutions are best visualized in Figures 3(a)-(e).

It is critical to mention that although the 44 locations in the W matrices are labeled (Figures 3 and A-1) to be associated predominantly with a given geothermal signature (i.e., a specific region; A, B, etc.), it does not mean the locations are related with only one signature. Instead, it means that those locations predominantly dominate commensurate signatures with contributions from other signatures too.

Figure A-2 shows the *H* matrices for signatures of the *k*=2, 3, 4, 5, and 8, which show the progression of the extracted signatures related to the observed 18 geothermal attributes. This progression also represents the transformation of signatures as the number of signatures increases. For example, Signatures A, B, and C of the *k*=3 solution (Figure A-2b) have similar properties to Signatures A, B, and both C and D of the *k*=4 solution (Figure A-2c), respectively. Signatures of A, B, C, and D for *k*=4 (Figure A-2c) possess similarities to signatures both A and E, B, C, and D for *k*=5 (Figure A-2d), respectively. Signatures A, B, C, D, and E of the *k*=5 solution share similar values with (1) A and E, F, B, (2) G and H, and (3) C and D of the *k*=8 solution (Figure A-2e), respectively.

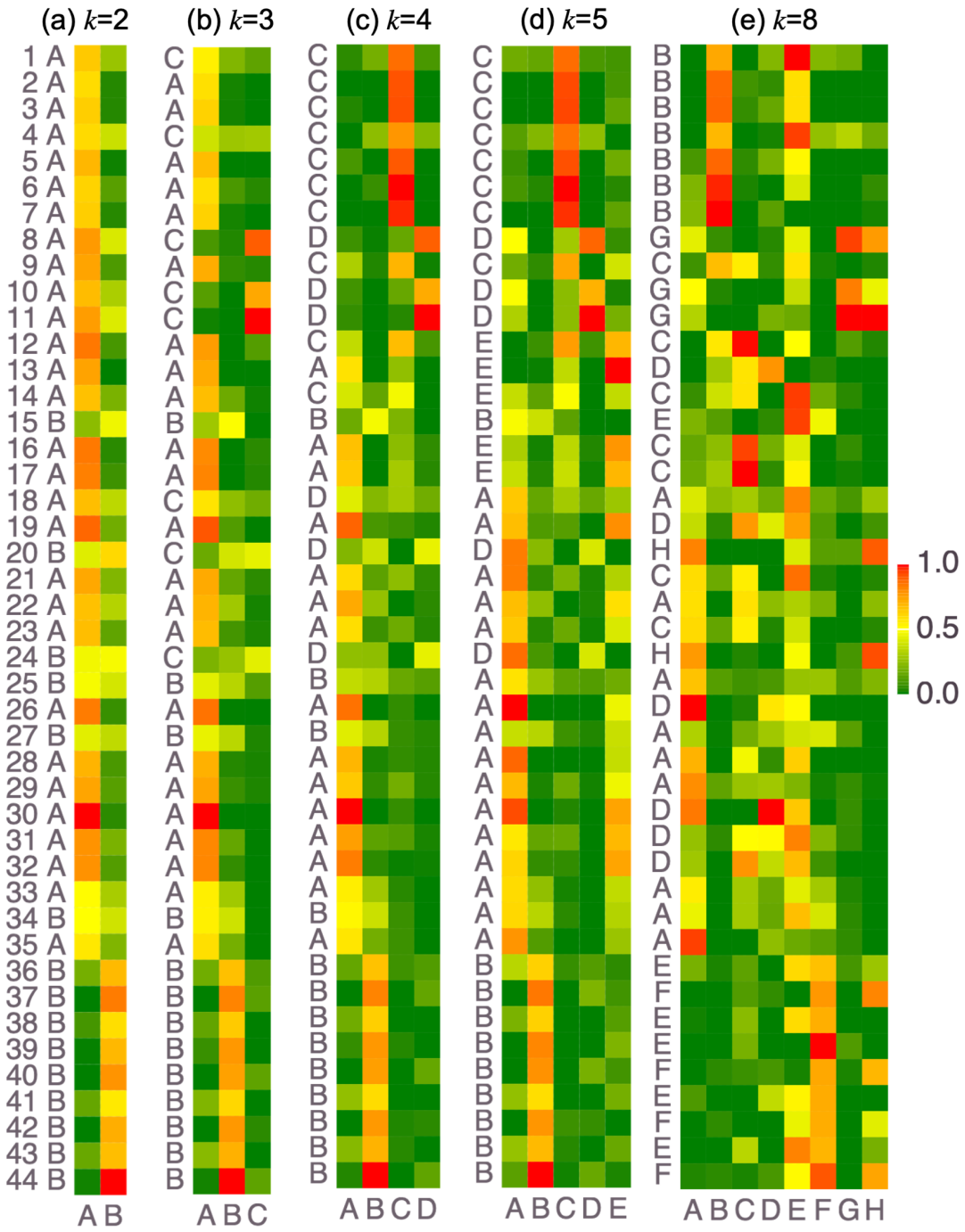

Figure A-1: NMF*k* location (*W*) matrices for (a) *k*=2, (b) *k*=3, (c) *k*=4, (d) *k*=5, and (e) *k*=8. These matrix plots show the association of each location to the extracted geothermal signatures. High-value matrix entries (red) define high significance, while low-value matrix entries (green) represent low significance.

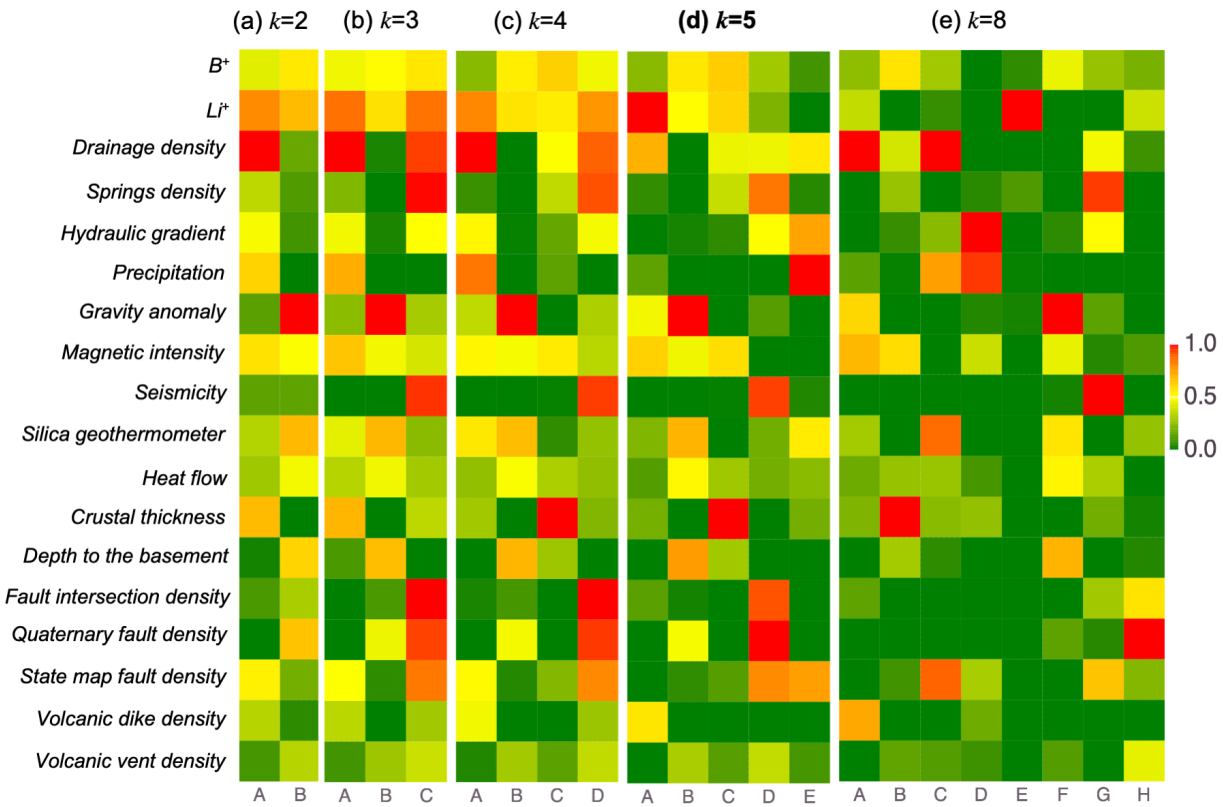

Figure A-2: NMF*k* attribute matrices for (a) *k*=2, (b) *k*=3, (c) *k*=4, (d) *k*=5 , and (e) *k*=8. These matrix plots show the contribution of each attribute on signatures. High-value matrix entries (red) define high significance, while low-value matrix entries (green) represent low significance.

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). 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. (2020c). 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., Mudunuru, M., Middleton, R., and Karra, S. (2021). Geochemical characteristics of low-, medium-, and hot-temperature geothermal resources of the Great Basin, USA. In *World Geothermal Congress, Reykjavik, Iceland*.

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.

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.

Anzieta, J., Ortiz, H., Arias, G., and Ruiz, M. (2019). Finding possible precursors for the 2015 Cotopaxi Volcano eruption using unsupervised machine learning techniques. *International Journal of Geophysics*, 2019.

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.

Bielicki, J., Blackwell, D., Harp, D., Karra, S., Kelley, R., Kelley, S., Middleton, R., Person, M., Sutula, G., and Witcher, J. (2016). Hydrogeologic windows and estimating the prospectivity of geothermal resources. In *Proceedings, 41st Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, California, February*, pages 22– 24.

Bielicki, J., Blackwell, D., Harp, D., Karra, S., Kelley, R., Kelly, S., Middleton, R., Pepin, J., Person, M., and Sutula, G. (2015). Hydrogeolgic Windows: Regional Signature Detection for Blind and Traditional Geothermal Play Fairways, Final Report, Los Alamos National Laboratory. Technical report, LA-UR-15-28360.

Böttcher, A. and Wenzel, D. (2008). The frobenius norm and the commutator. *Linear algebra and its applications*, 429(8-9):1864–1885.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

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.

Cather, S. (1990). Stress and volcanism in the northern Mogollon-Datil volcanic field, New Mexico: Effects of the post-Laramide tectonic transition. *GSA Bulletin*, 102(11):1447– 1458. Publisher: GeoScienceWorld.

Chapin, C., Wilks, M., McIntosh, W., and Cather, S. (2004). Space-time patterns of Late Cretaceous to present magmatism in New Mexico—Comparison with Andean volcanism and potential for future volcanism. *New Mexico Bureau of Geology and Mineral Resources Bulletin*, 160:13–40.

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

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

Elston, W., Rhodes, R., Coney, P., and Deal, E. (1976). Progress Report on the Mogollon Plateau Volcanic Field Southwestern New Mexico, no. 3—Surface Expression of a Pluton. Technical report.

Faulds, J., Craig, J., Hinz, N., Coolbaugh, M., Glen, J., Earnery, 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. *GRC Transactions*, 42. Institution: Nevada Bureau of Mines and Geology, University of Nevada, Reno, Number: DOE-UNR-06731-08.

Faulds, J., Hinz, N., Coolbaugh, M., Ramelli, A., Glen, J., Ayling, B., Wannamaker, P., Deoreo, S., Siler, D., and Craig, J. (2019). Vectoring into Potential Blind Geothermal Systems in the Granite Springs Valley Area, Western Nevada: Application of the Play Fairway Analysis at Multiple Scales. *PROCEEDINGS, 44th Workshop on Geothermal Reservoir Engineering, Stanford University*. Institution: Nevada Bureau of Mines and Geology, University of Nevada, Reno, Number: DOE-UNR-06731-03.

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. *GRC Transactions*, 39(DOE-UNR-06731-05). Nevada Bureau of Mines and Geology, University of Nevada, Reno.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics, New York, New York.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G., and Cai, J. (2018). Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377.

Hartigan, J. and Wong, M. (1979). A k-means clustering algorithm. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(1):100–108.

Hunt, C. (1956). Cenozoic Geology of the Colorado Plateau. *US Geological Survey Professional Paper*.

Johnson, P., Rouet-Leduc, B., Pyrak-Nolte, L., Beroza, G., Marone, C., Hulbert, C., Howard, A., Singer, P., Gordeev, D., Karaflos, D., Levinson, C., Pfeiffer, P., Ming Puk, K., and Reade, W. (2021). Laboratory earthquake forecasting: A machine learning competition. *Proceedings of the National Academy of Sciences*, 118(5).

Keller, G. R., Khan, M. A., Morganc, P., Wendlandt, R. F., Baldridge, W. S., Olsen, K. H., Prodehl, C., and Braile, L. W. (1991). A comparative study of the Rio Grande and Kenya rifts. *Tectonophysics*, 197(2-4):355–371.

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.

Lautze, N., Ito, G., Thomas, D., Frazer, N., Martel, S. J., Hinz, N., Tachera, D., Hill, G., Pierce, H. A., Wannamaker, P. E., and Martin, T. (2020). Play Fairway analysis of geothermal resources across the State of Hawai‘i: 4. Updates with new groundwater chemistry, subsurface stress analysis, and focused geophysical surveys. *Geothermics*, 86:101798.

Lautze, N., Thomas, D., Waller, D., Frazer, N., Hinz, N., and Apuzen-Ito, G. (2017). Play fairway analysis of geothermal resources across the state of Hawaii: 3. Use of development viability criterion to prioritize future exploration targets. *Geothermics*, 70:406–413.

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).

Lucchitta, I. (1979). Late Cenozoic uplift of the southwestern Colorado Plateau and adjacent lower Colorado River region. *Tectonophysics*, 61(1-3):63–95.

McClain, J. S., Dobson, P., Cantwell, C., Conrad, M., Ferguson, C., Fowler, A., Gasperikova, E., Glassley, W., Hawkes, S., Schiffman, P., Siler, D., Spycher, N., Ulrich, C., Zhang, Y., and Zierenberg, R. (2015). Geothermal play fairway analysis of potential geothermal resources in NE California, NW Nevada, and Southern Oregon: A transition between extension-hosted and volcanically-hosted geothermal fields. In *Geothermal Resources Council Annual Meeting, GRC 2015 - Geothermal: Always On*, pages 739–742. Geothermal Resources Council.

McIntosh, W., Geissman, J., Chapin, C., Kunk, M., and Henry, C. (1992). Calibration of the latest Eocene-Oligocene geomagnetic polarity time scale using 40Ar/39Ar dated ignimbrites. *Geology*, 20(5):459–463.

Medsker, L. and Jain, L. (1999). *Recurrent neural networks: design and applications*. CRC press.

Müller, A. C. and Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Inc.

Nakai, J. S., Sheehan, A. F., and Bilek, S. L. (2017). Seismicity of the rocky mountains and Rio Grande Rift from the EarthScope Transportable Array and CREST temporary seismic networks, 2008–2010. *Journal of Geophysical Research: Solid Earth*, 122(3):2173– 2192.

Olsen, K., Keller, G., and Stewart, J. (1979). Crustal structure along the Rio Grande rift from seismic refraction profiles. *Rio Grande Rift: Tectonics and Magmatism*, 14:127–144.

Pepin, J. (2019). *New approaches and insights to geothermal resource exploration and characterization*. PhD thesis, New Mexico Institute of Mining and Technology.

Person, M., Phillips, F., Kelley, S., Timmons, S., Pepin, J., Blom, L., Haar, K., and Murphy,
M. (2013). Assessment of the sustainability of geothermal development within the Truth or Consequences hot-springs district. 551(65).

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.

Ratté, J. and Grotbo, T. (1979). *Chemical analyses and norms of 81 volcanic rocks from part of the Mogollon-Datil volcanic field, southwestern New Mexico*. US Geological Survey.

Rouet-Leduc, B., Hulbert, C., McBrearty, I., and Johnson, P. (2020). Probing slow earthquakes with deep learning. *Geophysical research letters*, 47(4):e2019GL085870.

Sanford, A., Lin, K., Tsai, I., and Jaksha, L. (2002). Earthquake catalogs for New Mexico and bordering areas: 1869–1998. *New Mexico Bureau of Geology and Mineral Resources Circular*, 210:1–9.

Shervais, J. W., Glen, J. M., Liberty, L. M., Dobson, P., and Gasperikova, E. (2015). Snake River Plain Play Fairway Analysis – Phase 1 Report. *Transactions - Geothermal Resources Council*. Institution: Utah State Univ., Logan, UT (United States), Geothermal Resources Council.

Shervais, J. W., Glen, J. M. G., Nielson, D. L., Garg, S., Liberty, L. M., Siler, D., Dobson, P., Gasperikova, E., Sonnenthal, E., Neupane, G., DeAngelo, J., Newell, D. L., Evans, Snyder, N. (2017). Snake River Plain Play Fairway Analysis – Phase 1 Report. Idaho National Lab. (INL), Idaho Falls, ID (United States).

J. P., and Snyder, N. (2017). Geothermal Play Fairway Analysis of the Snake River Plain: Phase 2. Technical report, Idaho National Lab. (INL), Idaho Falls, ID (United States).

Siler, D. and Pepin, J. (2021). 3-D Geologic Controls of Hydrothermal Fluid Flow at Brady geothermal field, Nevada, USA. *Geothermics*. in review.

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

Siler, D., Zhang, Y., Spycher, N., Dobson, P., McClain, J., Gasperikova, E., Zierenberg, R., Schiffman, P., Ferguson, C., Fowler, A., and Cantwell, C. (2017). Play-fairway analysis for geothermal resources and exploration risk in the Modoc Plateau region. *Geothermics*, 69:15–33.

Siler, D. L., Faulds, J. E., Hinz, N. H., Dering, G. M., Edwards, J. H., and Mayhew, B. (2019). Three-dimensional geologic mapping to assess geothermal potential: examples from Nevada and Oregon. *Geothermal Energy*, 7(1):2.

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

Thompson, G. and Zoback, M. (1979). Regional geophysics of the Colorado Plateau. *Tectonophysics*, 61(1-3):149–181.

UNM (2018). Digital Geologic Map of New Mexico – Volcanic Vents. https://catalog.data.gov/dataset/digital-geologic-map-of-new-mexicovolcanic-vents

USGS (2018a). Energy and Environment in the Rocky Mountain Area. https://waterdata.usgs.gov/nwis.

USGS (2018b). National Water Information System. https://my.usgs.gov/eerma/.

Vesselinov, V., Ahmmed, B., Mudunuru, M., Karra, S., and Middleton, R. (2021). Hidden geothermal signatures of southwest New Mexico. In *Proceedings of the World Geothermal Congress, Reykjavik, Iceland*.

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

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.

Watson, L., Johnson, J., Sciotto, M., and Cannata, A. (2020). Changes in Crater Geometry Revealed by Inversion of Harmonic Infrasound Observations: 24 December 2018 Eruption of Mount Etna, Italy. *Geophysical Research Letters*, 47(19):e2020GL088077.

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. Publisher: Elsevier.

Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? *arXiv preprint arXiv:1411.1792*.