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

Machine Learning on the Geochemical Characteristics of Low-, Medium-, and Hot-temperature Geothermal Resources in the Great Basin, USA

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

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

*Corresponding author: ahmmedb@lanl.gov

Keywords: geothermal, hidden geothermal resource, unsupervised machine learning, NMF*k*, geochemistry

ABSTRACT

The Great Basin is the largest area of contiguous endorheic watersheds located in the western USA. It covers Nevada, much of Oregon and Utah, and portions of California, Idaho, and Wyoming. It includes multiple geothermal reservoirs ranging from low- to high-temperature resources and a vast area that is yet to be explored to discover new geothermal resources. In this study, we aim to characterize the geochemical data features of low-, medium-, and hot-temperature geothermal resources of the Great Basin. Geochemistry of the groundwater provides critical information about groundwater types, sources, recharge areas, mixing, flow paths, hydrogeochemical interaction between water and reservoir rock. Geochemistry is expected to provide critical information about the geothermal properties of the reservoirs, including temperature, heat flow, flow boundary conditions, and spatial extent (lateral and vertical). The geochemical data are also expected to include hidden information that is a proxy for geothermal anomalies. The relationships between geothermal conditions and chemistry have been extensively explored in the past Play Fairway Analysis. However, these relationships are not always efficient to discover hidden/blind geothermal systems. Recent advances in machine-learning (ML) methods provide novel approaches for fast, efficient, and robust exploration and mining of existing large datasets. As a result, analyses of this proxy data with ML methods can provide insights on the characteristics and locations of geothermal resources. We processed geochemical data in the Great Basin at 14341 locations for 15 attributes: pH, TDS (total dissolved solids), Al^{3+}, B^{+}, Ba^{2+}, Be^{2+}, Br^{-}, Ca^{2+}, Cl^{-}, HCO_3^{-}, K^{+}, Li^{+}, Mg^{2+}, Na^{+}, and $\partial^{18}O$. Three datasets are generated based on the labeling of low-temperature, medium-temperature, and high-temperature resources by the Nevada Bureau of Mines and Geology. An unsupervised ML method called non-negative matrix factorization with customized k-means clustering (NMF*k*) is applied to each dataset. NMF*k* identifies hidden signals in the dataset that are representative of hidden geothermal resources. In this study, ML analyses define key attributes that best characterize each type of geothermal resource. Results show that major cations/anions define low-temperature geothermal resources while relatively fewer major cations/anions define medium-temperature geothermal resources. Tracer elements are key attributes in defining high-temperature geothermal resources.

1. INTRODUCTION

Tester et al. (2007) estimated that the U.S. has ~13 million exajoule (EJ) geothermal energy resources, and ~200,000 EJ is extractable. William et al. (2008) estimated that hidden geothermal resources (~30GW) are three times more than discovered resources (~9GW); furthermore, additional 500 GW are potentially tappable through enhanced geothermal systems. Regardless of the discrepancy between the two estimations, it can be concluded that the U.S. has huge geothermal reserves. Currently, the U.S. has an installed capacity of ~3.7 GW and no new geothermal wells were drilled between 2015-2019 (Huttrer, 2020). This huge energy reserve is underutilized because (1) most of the resources are hidden/blind, (2) exploration is expensive and risky. Moreover, geothermal sites and their data are unique, which makes exploring hidden geothermal resources even more challenging.

Key geothermal exploration methods include water chemistry, temperature gradient wells, shallow heat flow, gravity anomaly, magnetic intensity, magnetotelluric or electrical resistivity survey, geology (e.g., fault), deep exploration drilling, etc. Other less useful methods include reflection seismic, self-potential (SP), soil gas surveys, LiDAR and hyperspectral surveys, and microseismic monitoring (Dobson et al. 2016). Also, well drilling during oil/gas exploration may provide information about geothermal conditions and help identify geothermal resources. Recently, Faulds and co-workers carried out a detailed Play Fairway Analysis (PFA) in the Great Basin to discover hidden geothermal resources (Faulds et al., 2015,2016,2018, Siler et al., 2019). They explored nine critical parameters for mapping hidden geothermal resources that are structural settings (e.g., normal/strike slip faults), age of recent faulting, slip rates on recent faults, regional-scale strain rates, slip and dilation tendency on faults, earthquake density, increased gravity, temperature distribution map at 3 km depth, and geochemistry from springs and wells. Among the preceding nine parameters, water geochemistry is frequently used in the earlier stage of exploration. Shallow water geochemistry data is cheaper to collect and easily accessible than the other geothermal data. Besides, they are useful in many ways such as: inferring reservoir temperatures from the composition of geothermal fluids, delineating boundaries of geothermal reserves, defining heat sources such as meteoric, magmatic, and mixed, estimating the degree of water/rock interaction, etc (Fridriksson, and Ármannsson, 2007; Klein, 2007). Generally, water geochemistry is used to estimate reservoir temperature before drilling deep wells using classical geothermometry (Fournier et al. 1970) or composite geothermometry (Spycher et al. 2014). Water geochemistry is also used to understand site water characteristics (Fowler et al. 2018 and Mao et al. 2015). However, to the authors' knowledge, there are no studies attempting to investigate geochemical data associated with low-, medium, and high-temperature geothermal resources in the Great Basin using ML. In this study, we apply unsupervised machine learning to differentiate water geochemical characteristics that can be applied to characterize and identify low-, medium, and high-temperature geothermal resources.

Hydrogeochemical processes of water chemistry are a complex phenomenon to generalize. However, existing novel machine-learning (ML) methods provide novel approaches for fast, efficient, and robust exploration and mining of existing large geochemical datasets. They are also good tools to gain an understanding of such complex systems. Here, we apply a novel unsupervised machine learning method called non-negative matrix factorization coupled with *k*-means clustering (NMF*k*) (Alexandrov and Vesselinov 2014; Vesselinov et al. 2018; Vesselinov et al. 2019). NMF factorizes a data matrix into two smaller matrices differentiating its latent (hidden) signatures (features) while *k*-means clustering finds the optimal number of signatures in the dataset. Here, NMF*k* identifies hidden signatures characterizing geothermal resources, the optimal number of these signatures, and the dominant attributes associated with each signature. Note that the extracted features are encapsulated in the data but cannot be observed explicitly; they represent some unknown relationships between observed attributes. Alternative unsupervised machine learning methods such as principal component analysis (Wold, 1987), independent component analysis (Comon, 1994), and singular value decomposition (Klema, 1980) can be also applicable. However, NMF*k* provides important benefits over preceding methods. It can handle sparse datasets, learns from parts of an object, and provides interpretable results (Lee & Sung, 1999; Vesselinov et al. 2018). Furthermore, NMF*k* has already been used 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.

In this study, we analyzed geochemistry data of the Great Basin (Figure 1). The Great Basin covers Nevada, much of its neighboring states Oregon, Utah, California, Idaho, and Wyoming. It has multiple geothermal reservoirs ranging from low- to high-temperature resources and a vast area is yet to be explored to discover hidden geothermal resources. To explore its geothermal resources, plenty of data have been collected over several decades. The data size for this study is 14341 x 15 where 14341 represents locations while 15 are geochemical attributes (water cations/anions) (Goff, et al., 2002; Zehner, et al., 2006; Nevada Bureau of Mines and Geology, 2012). Among 14341 locations, 8652, 3413, and 2276 locations represent low-, medium-, and high-temperature geothermal resources. Questions key to this research are (*1) do geochemical signatures vary from one resource to another? and (2) what chemical elements differentiate each geothermal resource?*

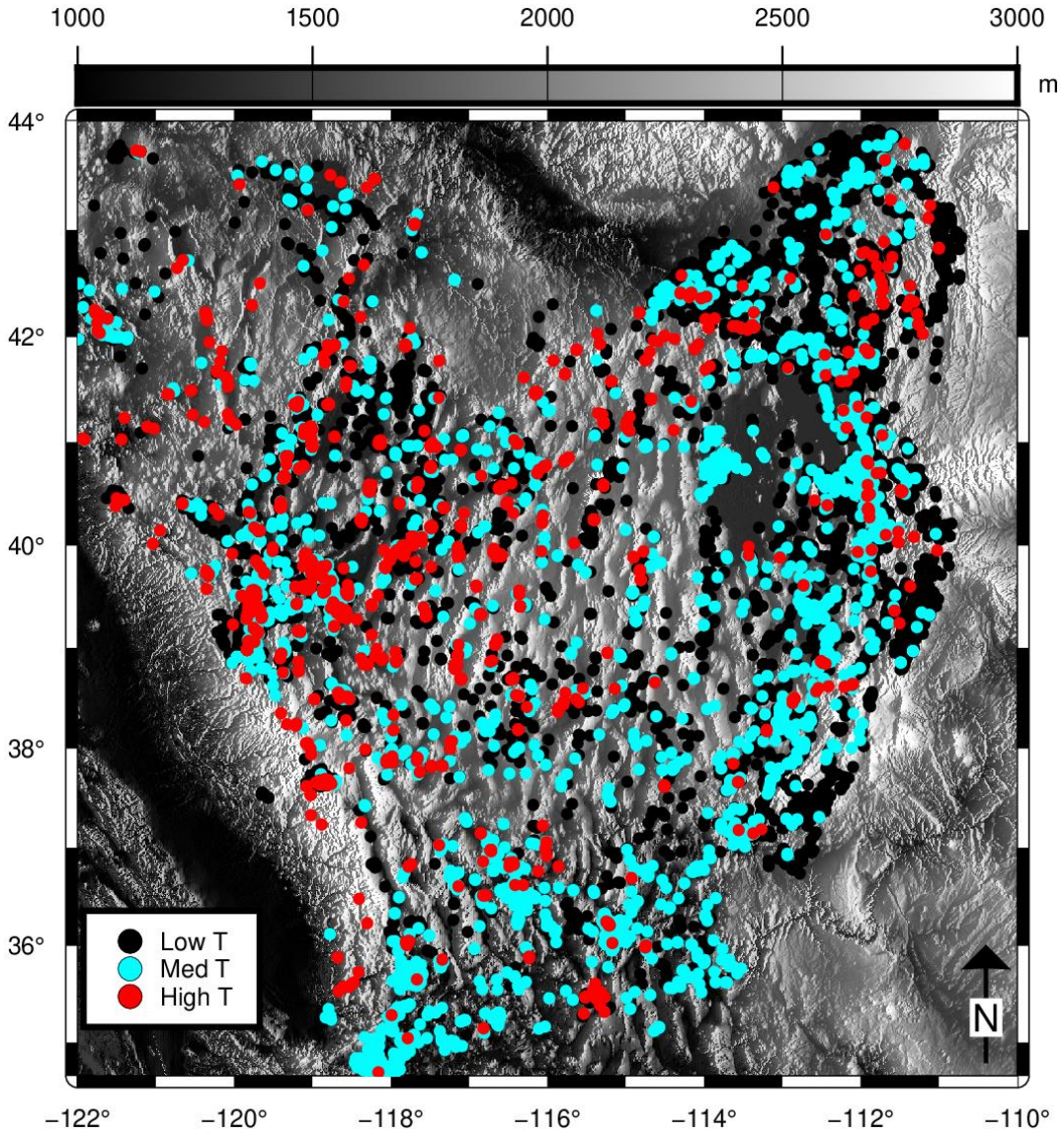

Figure 1: Locations of low-, medium-, and high-temperature geothermal resources.

2. METHODS

2.1 Non-negative matrix factorization

Given observational data D 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 is to decompose this matrix D into a non-negative "attribute" matrix A of size (n, k) and non-negative "location" matrix B of size (k, m):

$$D = A \times B + \varepsilon(k)$$

where k is the unknown number of signals (features) present in the data and $\epsilon(k)$ is error for k^{th} signals. The "attribute" matrix A represents how the extracted features are related to the attributes. The "location" matrix B 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 be expected to 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 & Sung, 1999)

$$O = || D - A \times B||_F;\ A, B > 0\ \forall\ m, n, k$$

For each k value in the range $1, 2, 3, \ldots, d$, non-negative matrix factorization is performed for multiple times (typically on the order of 1000 times) based on random initial guesses for A and B 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 B (or alternatively A; 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 clusters contain 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 $S(k)$ 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. The values of $S(k)$ theoretically can vary from -1 to 1. Typically, $S(k)$ declines sharply after an optimal number, k_{opt} , is reached. k_{opt} value is selected to be equal to the maximum number of signals that accurately reconstructs the observational data matrix D as estimated by $O(k_{\text{opt}})$ and the average silhouette width $S(k_{\text{opt}})$ is close to 1.

2.2 Data Description

The size of the total dataset is14341×15(locations x attributes). The size of the datasets for low-, medium-, and high-temperature resources are 8652×15, 3413×15, and 2276×15, respectively. The 15 attributes are pH, TDS (total dissolved solids), Al^{3+}, B^{+}, Ba^{2+}, Be^{2+}, Br^{-}, Ca^{2+}, Cl^{-}, HCO_3^{-}, K^{+}, Li^{+}, Mg^{2+}, Na^{+}, and $\partial^{18}O$. pH represents alkalinity of water, TDS represents total amount of major and tracer cations/anions, Ca^{2+}, K^{2+}, Mg^{2+}, Na^{+} are major cations, HCO3 and Cl are major anions, Al^{3+}, B^{+}, Ba^{2+}, Be^{2+}, Br^{-}, are Li^{+} trace elements, and $\partial^{18}O$ is an oxygen isotope. Major anions/cations define the ionic type of water. The $\partial^{18}O$ defines the origin (e.g., meteoric, magmatic, connate) of the water.

3. RESULTS AND DISCUSSION

Low-temperature geothermal resources: NMFk was applied to low-temperature geothermal data for eight signals (Figure 2(a)). Solutions for 2, 3, and 4 signals have low $O(k)$and high$S(k)$values. The algorithm rejects the solutions for 5, 7, and 8 signals. Signal 6 shows low $S(k)$. Therefore, signals 2, 3, and 4 are potential optimal signals. Because Signal 3 yields low $O(k)$and highest$S(k)$values, it is the optimal signal for this dataset suggesting that the dataset can be explained using three signatures A, B, and C (Figure 2b) instead of the whole dataset. Dominant attributes of Signature A are pH, TDS, B^{+}, HCO_3^{-}, and Na^{+}. Dominant attributes of Signature B are TDS, B^{+}, Br^{-}, Ca^{2+}, Cl^{-}, HCO_3^{-}, Li^{+}, and Mg^{2+}. Dominant attributes of Signature C are TDS, Ba^{2+}, Ca^{2+}, HCO_3^{-}, Mg, and δO^{18}. Figure 2(c) shows locations relating to each signature. By number, Signature A represents most of the wells (5136) followed by Signature B (2344) and C (1172).

Medium-temperature geothermal resources: Figure 3a shows the progression of$O(k)$and$S(k)$over signal number. Like in low-temperature resources, only eight signals were used in the NMFk algorithm. Here, $O(k)$decreased over signal number but $S(k)$varies. Because Signal 2 yields relatively low $O(k)$and highest$S(k)$values, it is the optimal signal for this dataset suggesting that the dataset can be explained using two signatures A and B (Figure 3b). Dominant attributes of Signature A are pH, TDS, Ca, HCO3, and Mg. Dominant attributes of Signature B are Br, Li, and Na. Figure 3(c) shows locations relating to each signature of Signal 2. Signature A captures the majority (1869 or ~55%) of wells while Signature B captures 1544 locations

High-temperature geothermal resources: Figure 4a shows the progression of$O(k)$and$S(k)$over signal number. NMFk algorithm is run for eight signals. Here, Signal 3 is the optimal signal because of reasonably low $O(k)$and highest $S(k)$values. Figure 4(b) shows the attribute matrix representing the contribution of each attribute on each signal. Dominant attributes of Signature A are pH, TDS, B^{+}, and HCO_3^{-}. Dominant attributes of Signature B^{+} are Ca^{2+} and Mg^{2+}. Dominant attributes of Signature C are Al^{3+} and Be^{2+}. Figure 4(c) shows locations relating to each signature of Signal 3. By number, Signature A captures the majority (1702) of wells followed by Signature B (541) and Signature C (33).

Table 1 shows dominant attributes for all signatures of the optimal signal for low-, medium-, and high-temperature resources. All major cations/anions, ∂O^{18} isotope, and a few trace elements have high dominance in signatures of low-temperature resources captures suggesting that it represents groundwater chemistry. The presence of B^{+}, Ba^{2+}, Br^{-}, Li^{+} indicates that tracer elements also control the characteristics of the groundwater that might arrive from the deep. Signatures of medium-temperature resources do not capture all major anions/cations of the groundwater. Dominant attributes are a subset of dominant attributes of low-temperature resources. Very few major ions dominate the signatures of high-temperature geothermal resources that are unique to this type of resource. Also, Al and Be are present in signatures that are not visible in either low- and medium-temperature resources.

Table 1: Uniqueness of each type of geothermal resources based on all signatures.

Geothermal resource type	Common dominant attributes across signatures
Low temperature	B^{+}, **Ba^{2+}**, Br^{-}, Ca^{2+}, Cl^{-}, HCO_3^{-}, Li^{+}, Mg^{2+}, Na^{+}, **δO^{18}**, pH, and TDS
Medium temperature	Br^{-}, Ca^{2+}, HCO_3^{-}, Li^{+}, Mg^{2+}, Na^{+}, pH, and TDS
High temperature	**Al^{3+}**, B^{+}, **Be^{2+}**, Ca^{2+}, HCO_3^{-}, Mg, pH, and TDS

Table 2: Uniqueness of each type of geothermal resources based on Signature A.

Type of geothermal resources	Dominant attributes of Signature A
Low temperature	B^{+}, HCO_3^{-}, Na^{+}, pH, and TDS
Medium temperature	Ca^{2+}, HCO_3^{-}, Mg^{2+}, pH, and TDS
High temperature	B^{2+}, and HCO_3^{-}, pH, and TDS

Figures 2-4(c) show locations associated with each signature of the optimal signals of low-, medium-, and high-temperature resources. Signature A for each resource type captures most of the well followed by B and/or C. Percentages of Signature A in each resource type are ~60, ~55, and ~75%, respectively. The dominant attributes of all signatures are different than the dominant attributes of Signature A. Additionally, dominant attributes of Signature A vary over types of geothermal resources. For example, Na of low-temperature resources is not dominant in medium- and high-temperature resources; B^{+} is not dominant in medium-temperature resources. Ca^{2+} and Mg^{2+} are unique in Signature A of medium-temperature resources. Na^{+}, Mg^{2+}, and Ca^{2+} (major cations) are absent in high-temperature resources as dominant elements. All these analyses support our hypothesis that geochemical signatures vary for types of geothermal resources. Specifically, major cations/anions define low-temperature geothermal resources while relatively fewer major cations/anions define medium-temperature geothermal resources. Finally, tracer elements provide the main representatives of high-temperature geothermal resources.

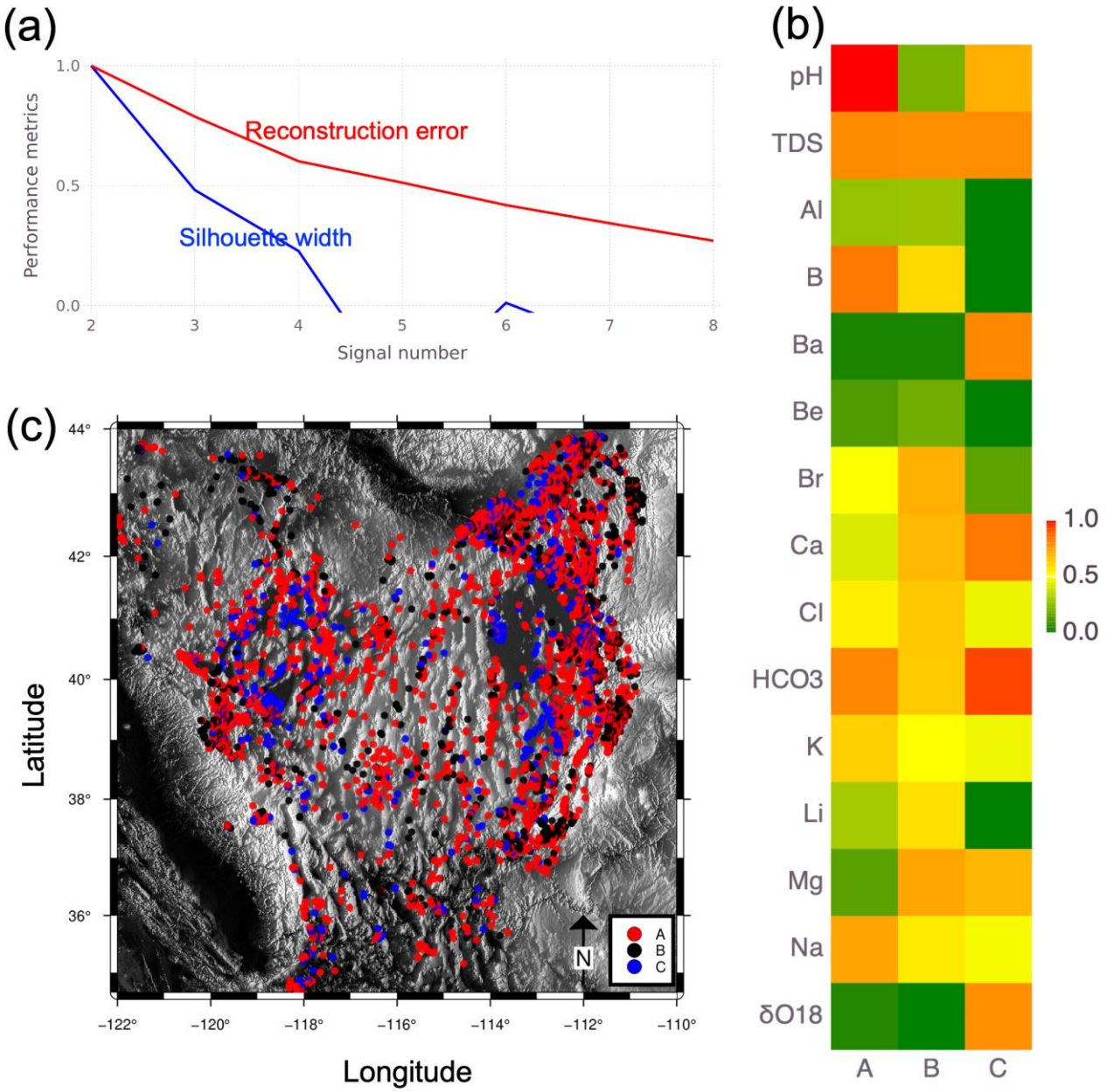

Figure 2: Low-temperature geothermal resources: (a) Estimated $O(k)$ and $S(k)$ over number of signals (k), (b) attribute matrix showing the significance of each attribute on each signature, and (c) locations relating to each signature.

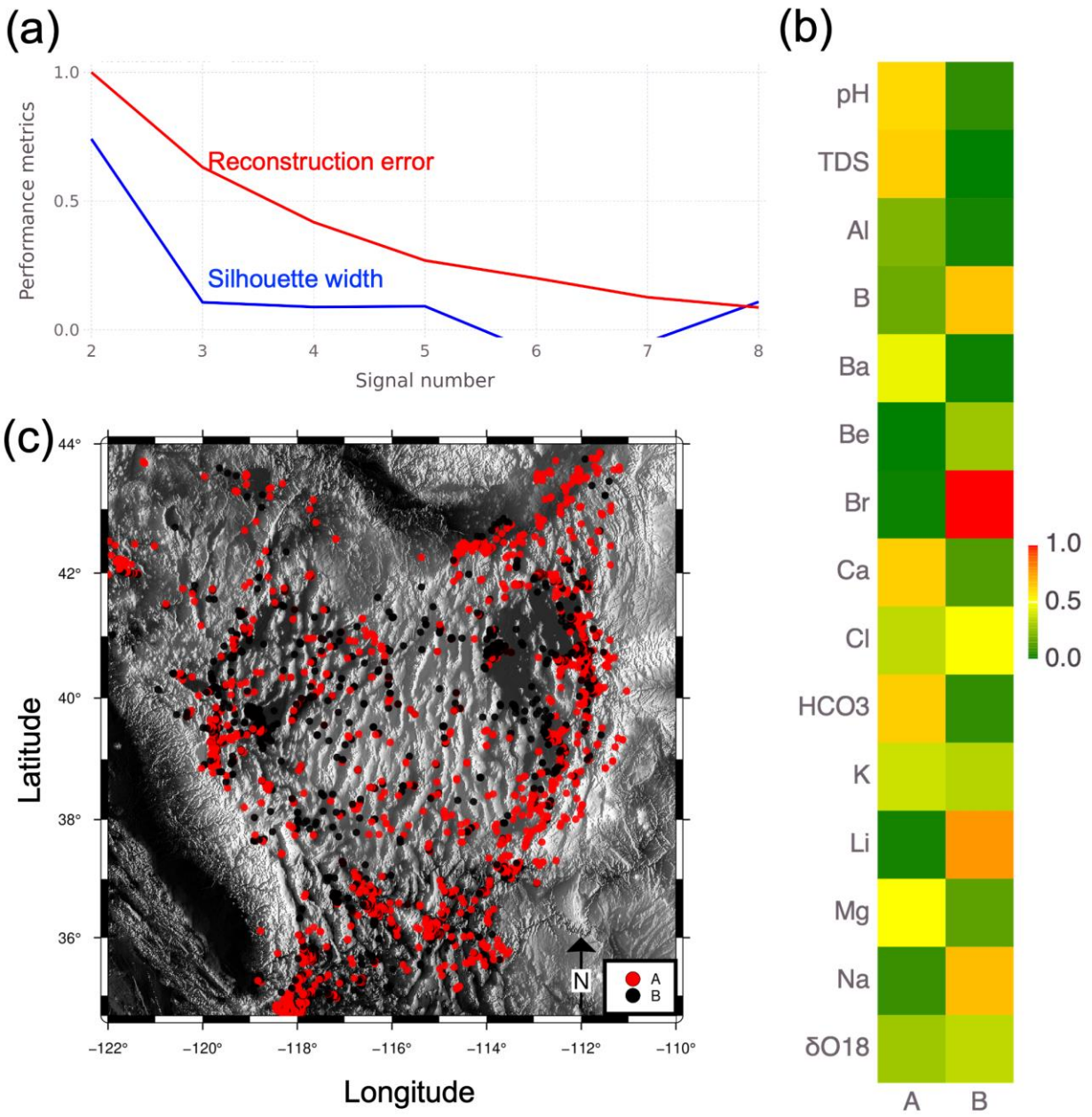

Figure 3: Medium-temperature geothermal resources: (a) Estimated $O(k)$and$S(k)$over number of signals (k), (b) attribute matrix showing the significance of each attribute on each signature, and (c) locations relating to each signature.

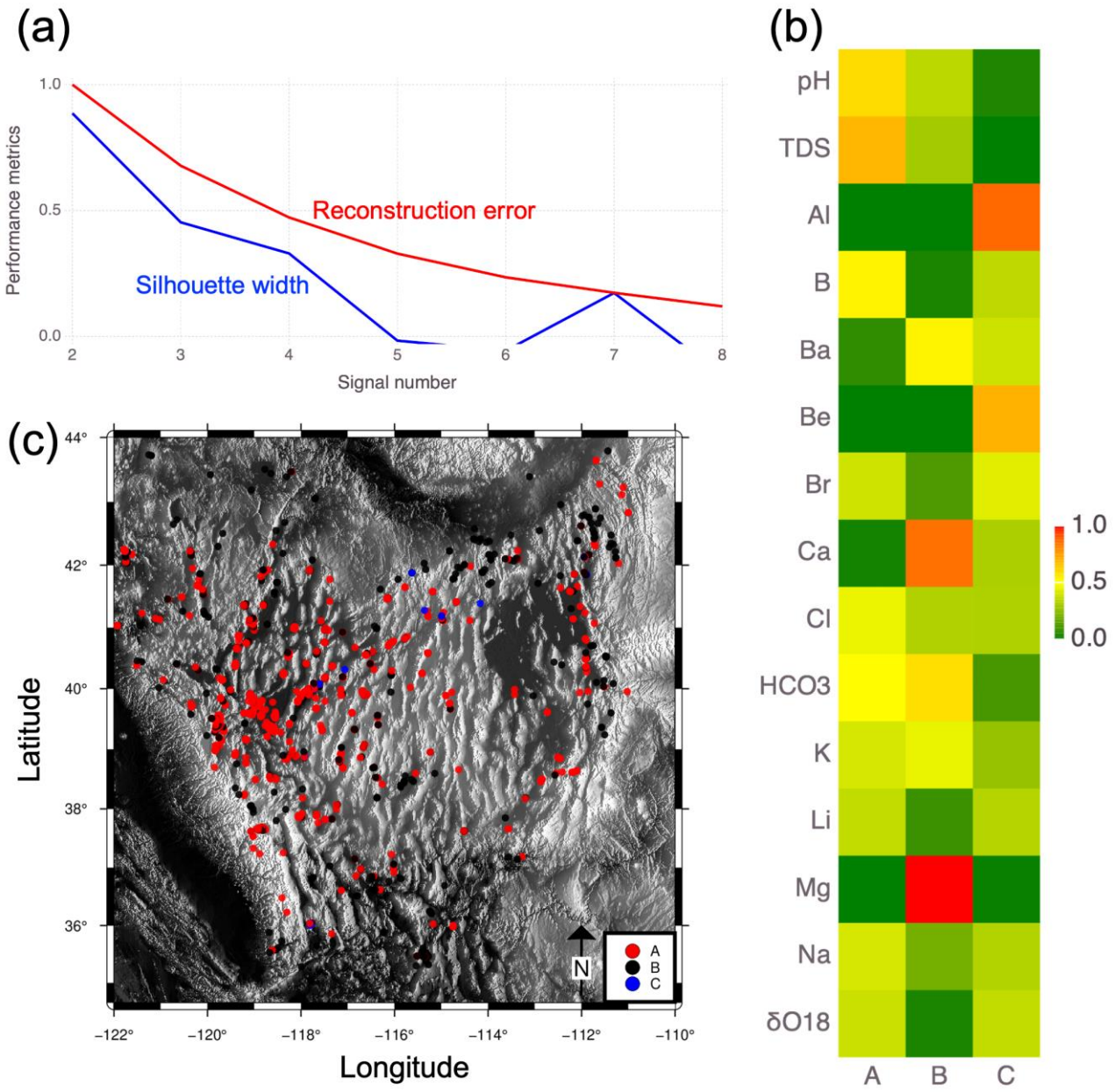

Figure 4: High-temperature geothermal resources: (a) Estimated $O(k)$ and $S(k)$ over number of signals (k), (b) attribute matrix showing the significance of each attribute on each signature, and (c) locations relating to each signature.

4. CONCLUSIONS

This study applied unsupervised machine learning tool NMFk to understand geochemical characteristics for low-, medium-, and high-temperature geothermal resources. NMFk was applied to a rich geochemistry dataset for the Great Basin. The size of the total dataset is 14341×15 while the size of the datasets specifically related to low-, medium-, and high-temperature resources are 8652×15, 3413×15, and 2276×15, respectively. NMFk found the optimal number of hidden signals associated with low-, medium-, and high-temperature geothermal resources are 3, 2, and 3, respectively. The dominant attributes of low-temperature resources are B^{+}, Ba^{2+}, Br^{-}, Ca^{2+}, Cl^{-}, HCO_3^{-}, Li^{+}, Mg^{2+}, Na^{+}, δO^{18}, pH, and TDS; the dominant attributes of medium-temperature resources are Br^{-}, Ca^{2+}, HCO_3^{-}, Li^{+}, Mg^{2+}, Na^{+}, pH, and TDS; and the dominant attributes of high-temperature resources are Al^{3+}, B^{+}, Be^{2+}, Ca^{2+}, HCO_3^{-}, Mg^{2+}, pH, and TDS.

Results indicate that the dominance of geochemical attributes in extracted hidden geothermal signatures vary. Major cations/anions define low-temperature geothermal resources while relatively fewer major cations/anions define medium-temperature geothermal resources. Contrary to the current belief that tracer elements have nothing to do with groundwater temperature, we found that in high-temperature resources, tracer elements are dominant. pH and TDS are present in all types of resources as dominant attributes indicating that they are not critical variables to classify resource types. These findings suggest that water geochemistry may define geothermal resource types. This understanding is a massive gain towards estimating geothermal prospects using supervised machine learning methods. In the future, a machine learning model will be built that can predict geothermal prospects using the geochemical constituents of this dataset excluding pH and TDS.

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).

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.

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., 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.

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

Fowler, A. P., Ferguson, C., Cantwell, C. A., Zierenberg, R. A., McClain, J., Spycher, N., and Dobson, P. (2018). A conceptual geochemical model of the geothermal system at Surprise Valley, CA. *Journal of Volcanology and Geothermal Research*, 353:132 – 148.

Fridriksson, T. and Ármannsson, H. (2007). Application of geochemistry in geothermal resource assessments. *Short course on geothermal development in Central America resource assessment and environmental management, El-Salvador*.

Goff, F., Bergfeld, D., and Janik, C. (2002). Geochemical data on waters, gases, scales, and rocks from the Dixie Valley region, Nevada (1996-1999). Technical report, DOE–EEGTP (USDOE Office of Energy Efficiency and Renewable Energy Geothermal.

Klein, C. (2007). Advances in the past 20 years: Geochemistry in geothermal exploration, resource evaluation and reservoir management. *Geothermal Resources Council Transactions*, 31:17–22.

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

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).

Mao, X., Wang, Y., Zhan, H., and Feng, L. (2015). Geochemical and isotopic characteristics of geothermal springs hosted by deep-seated faults in Dongguan Basin, southern China. *Journal of Geochemical Exploration*, 158:112–121.

Nevada Bureau of Mines and Geology (2012). Great Basin Groundwater Geochemical Database. http://www.nbmg.unr.edu/Geothermal/GeochemDatabase.html, Accessed on 2020-12-02.

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. *Geother- mal Energy*, 7(1):1–32.

Spycher, N., Peiffer, L., Sonnenthal, E., Saldi, G., Reed, M., and Kennedy, B. (2014). Integrated multicomponent solute geothermometry. *Geothermics*, 51:113–123.

Tester, J. (2007). The future of geothermal energy, impact of enhanced geothermal system (EGS) on the United State in the 21st century: An assessment by an MIT-led interdisciplinary panel. Technical report, Massachusetts Institute of Technology, Report ISBN 0-615-13438-6.

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.

Williams, C., Reed, M., Mariner, R., DeAngelo, J., and Galanis, S. (2008). Assessment of moderate- and high-temperature geothermal resources of the United States. Technical report, Geological Survey (US).

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

Zehner, R., Coolbaugh, M., and Shevenell, L. (2006). Regional groundwater geochemical trends in the Great Basin: implications for geothermal exploration. *Geothermal Resources Council Transactions*, 30:117–124.