

Machine learning to identify geologic factors associated with production in geothermal fields: A case-study using 3D geologic data, Brady geothermal field, Nevada

Drew L. Siler¹, Jeff D. Pepin², Velimir V. Vesselinov³, Maruti K. Mudunuru⁴ Bulbul Ahmmmed³

¹U.S. Geological Survey, Moffett Field, California, USA

²U.S. Geological Survey, Albuquerque, New Mexico, USA

³Computational Earth Science Group, Los Alamos National Laboratory, Los Alamos, NM 87545

⁴Watershed & Ecosystem Science, Pacific Northwest National Laboratory, Richland, WA 99352

Abstract

In this paper, we present an analysis using unsupervised machine learning (ML) to identify the key geologic factors that contribute to the geothermal production in the Brady geothermal field. Brady is a hydrothermal system in northwestern Nevada that supports both electricity production and direct use of hydrothermal fluids. Transmissive fluid flow pathways are relatively rare in the subsurface but are critical components of hydrothermal systems like Brady and many other types of fluid flow systems in fractured rock. The ML method, non-negative matrix factorization with k-means clustering (NMFk), is applied to a library of fourteen 3D geologic characteristics hypothesized to control hydrothermal circulation in the Brady geothermal field. Our results indicate the macro-scale faults and a local step-over in the fault system preferentially occur along with production wells when compared to injection wells and non-productive wells. We infer that these are the key geologic characteristics that control the through-going hydrothermal transmission pathways at Brady. Our results demonstrate 1) the specific geologic controls on the Brady hydrothermal system and 2) the efficacy of pairing ML techniques with 3D geologic characterization to enhance the understanding of subsurface processes.

1. Introduction

Crustal permeability is a key parameter in geothermal process models used in exploration and development of geothermal systems. Permeability is, however, highly variable in space (Caine et al., 1996; Caine and Forster, 1999; Fairley et al., 2003; Fairley and Hinds, 2004; Sanderson and Zhang, 2004) and this complicates characterization of subsurface hydrothermal processes. Accordingly, it is common in developed geothermal systems to produce fluid flow from a few relatively small (sub-meter- to meter-long) intervals of a borehole that may be 100s or 1000s of meters in total length (based on Nevada Division of Minerals, publicly available data). This compartmentalization of hydrothermal fluid flow means that the volume of rock that transmits fluids at rates suitable for power production or direct use is much smaller than the volume of rock that does not transmit fluid (or transmits at sub-commercial rates). This presents a significant challenge to efficient exploration, development, and management of these renewable energy resources.

Compartmentalization of the fluid flow system may be associated with a variety of geologic characteristics. For instance, spatial changes in fracture permeability throughout a fault network, and/or permeability variation in the stratigraphic succession may control compartmentalization. The purpose and innovation of this study is to reveal the geologic factors that influence this compartmentalization of fluid

flow in hydrothermal systems. We evaluate three-dimensional (3D) geologic characteristics through an unsupervised machine learning (ML) method called non-negative matrix factorization with k-means clustering (NMF k). Specifically, NMF k is applied to a suite of geologic factors that have been calculated along production, injection, and non-productive wells at Brady geothermal field in northwestern Nevada. The ML results indicate that the mapped macro-scale faults and the ~km-scale step-over in the fault system are closely associated with production wells relative to injection wells and non-productive wells. Tracking the 3D distribution of these factors in geothermal prospects, developed geothermal fields, and other types of fluid flow systems may help promote more efficient resource development and management.

2. Background

2.1 The Brady geothermal system

Brady geothermal field, in northwestern Nevada, USA (Figure 1), has seen geothermal electricity production since 1992 and research or exploration since at least 1959 (Benoit and Butler, 1983). The hydrothermal system supplies hot fluid to two power stations and a direct-use vegetable drying facility. Electricity production capacity at Brady is 26.1 MWe, and ~7 MWth is delivered to the drying facility (Ayling, 2020). Temperatures of produced fluid have been ~130-185°C during this time (based on Nevada Division of Minerals, publicly available data), though temperatures as high as 219°C have been measured (Shevenell et al., 2012). These relatively high temperatures occur at relatively shallow levels (300-600 depth for some production wells) as a result of convective upwelling driven by temperature-related differences in fluid density, and/or hydraulic head driven circulation through hot rock. In either case, relatively high heat flow in the Basin and Range physiographic province, which is associated with Miocene to recent crustal thinning (Lachenbruch and Sass, 1977; Blackwell, 1983), provides the heat. The fluids circulate through transmissive pathways that have been primarily attributed to a network of fractures within a step-over in the Basin and Range-type (e.g., Wernicke, 1992; Colgan et al., 2006) normal fault system, the Brady fault zone (Faulds et al., 2003, 2010a, b; 2017; Siler et al., 2018; Figure 1).

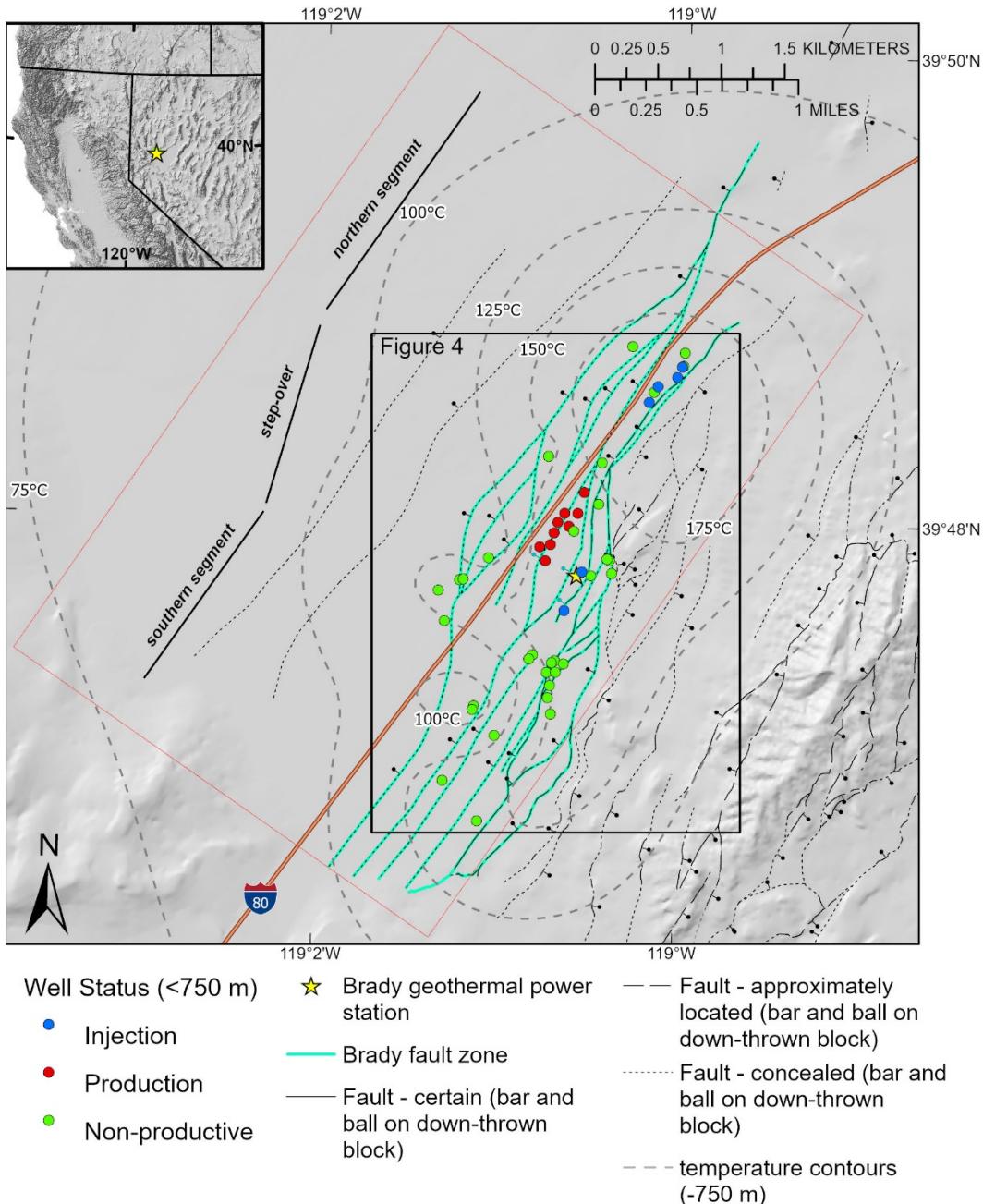


Figure 1. Map of Brady geothermal area. The fault strands that constitute the Brady fault are shown in green. Contours represent modeled temperature at 750 m depth. Wells are colored by their usage (production, injection, and non-productive) for depths shallower than 750 m. The general geometry of the step-over is shown to the left of the fault system. Interstate 80 (orange line) is shown for reference.

The stratigraphic section at Brady consists of metamorphic basement rocks overlain by Oligocene to late Miocene volcanic rocks, and late Miocene to Holocene sedimentary rocks. The Brady fault zone is a west-dipping, north-northeast-striking system of normal faults that cuts this stratigraphic section (Figure 1).

The step-over (Faulds et al., 2010a, b; 2017; Siler et al., 2013, 2016) is an area where parallel but non-collinear strands of the Brady fault zone come together (e.g., Peacock and Sanderson, 1991, 1994; Fossen and Rotevatn, 2016). The southern segment of the Brady fault zone steps to the left to meet the northern segment ([Figure 1](#); Faulds et al., 2017). Siler et al. (2018) suggested the occurrence of the hydrothermal system within the step-over is related to focused stress and strain that periodically occur at the step-over during fault slip, resulting in progressive generation and maintenance of a dense fracture network over geologic time. Advection of heat to shallow levels by hydrothermal circulation within this fracture network is evident from a ~3 km-wide × 6 km-long (across strike × along strike) temperature anomaly centered on the step-over ([Figure 1](#)). Geothermal production wells at Brady are situated within the step-over. Fluids are produced from two levels; ~300-600 m and ~1750 m depth (based on Nevada Division of Minerals, publicly available data).

3. Methods

An existing 3D geologic map of Brady, synthesizing a variety of geologic and geophysical data (Siler and Faulds, 2013a; Jolie et al., 2015; Siler et al., 2016, [2021](#); Witter et al., 2016) was used to develop a suite of geologic variables that may control the distribution/localization of transmissive pathways and production-grade hydrothermal fluid flow. The fourteen different variables described below are calculated from the 3D geologic map and projected to forty-seven production, injection, and non-productive wells within the field ([Figure 1](#)).

3.1 Geothermal Well Data

The Brady well dataset is much denser at shallow levels than at deep levels (just eight of forty-seven wells extend to ~1750 m, the deeper of the two geothermal reservoirs). The data at deeper levels are too sparse to be used in the NMFK analysis. As a result, this analysis focuses on the shallow (~300-600-m-deep) reservoir, where the data are sufficiently dense for NMFK. 750 m is used as the cutoff depth to ensure that the full length of all wells that produce from the shallow reservoir is included in the dataset. There are nine production wells and six injection wells that have been used for production or injection at depths of less than 750 meters since the geothermal power station was brought online in 1992. Wells producing from the shallow reservoir account for ~57% of the total produced volume (June 1992-August 2019; based on publicly available data Nevada Division of Minerals). We consider the remaining 32 wells to be non-productive wells. Four of these remaining thirty-two wells are used for production or injection; however, these wells produce or inject in the deeper reservoir. At depths less than 750 m, these four wells are cased and cemented. We assume that these wells would not transmit fluids to or from the formation at <750 m regardless of well completion. These four wells are therefore considered ‘non-productive’ for our purposes. Each of the below 14 geologic factors is calculated at 1m-intervals along all 47 wells. The resultant database of 336,784 entries (24,056 locations with 14 variables) is used as input data for NMFK analysis. [Figure 2](#) shows the values for each of the 14 variables for one of the production wells.

3.2 Geothermal variables

Fault factors (*faults*, *curve*, *td*, *ts*, *faultnear*):

For each of the thirty-two faults defined by the 3D geologic map (Siler and Faulds, 2013a; Siler et al., 2016, [2021](#); Witter et al., 2016), a 30-meter-wide fault zone is generated. This zone approximates the effective width of secondary faulting and fracturing around each fault. This width is consistent with empirically derived fault zone widths for km-long faults, like the Brady fault zone (Scholz et al., 1993; Anders and

Wiltschko, 1994). The *fault* variable has a value of '1' where a well is located within a fault zone and '0' for well intervals not located within a fault zone. The *curve* variable is the along-strike and down-dip curvature calculated along each fault. The *td* and *ts* variables are the dilation tendency and slip tendency, respectively for each fault. These values are calculated using methods of Morris et al. (1996) and Ferrill et al. (1999) and a local stress model calculated at Brady (Jolie et al., 2015). The 30-meter-wide fault zone for each fault is populated with the estimated *curve*, *ts*, and *td* values. Segments of faults with a high value for *curve* are postulated to be associated with accentuated faulting and fracturing as a result of stress loading at the highly curved fault segments (e.g., Sibson, 1994), and may therefore preferentially host fluid flow. Dilation tendency (*td*) and slip tendency (*ts*) are the ratios of the resolved normal stresses and the normal to shear stress ratio on faults, respectively. Fault segments that are either highly dilatant (high *td*) or stress loaded for slip (high *ts*) are likely to host fluid flow (Siler et al., 2021). For all wells, the *faultnear* variable is calculated as the difference between the distance to the nearest 3D mapped fault plane and the maximum distance to a fault in the dataset. This is done so that high *faultnear* values occur at intervals of wells that are near to faults (e.g., see Figure 2), in the same way that high values for the other variables occur where hydrothermal processes are expected.

Fault network factors (*faultdense*, *faultintdense*):

Areas in the subsurface with especially dense faulting and fracturing are expected to have relatively high permeability, and thus host hydrothermal circulation. The spatial density of fault planes (*faultdense*) and the spatial density of the lines of intersection between faults and the lines of termination of faults (*faultintdense*) are calculated in 3D space (Siler et al., 2021). Fault intersections and terminations represent structural discontinuities, where stresses become concentrated and accentuated fracturing is expected (Peacock and Sanderson, 1991; Fossen and Rotevatn, 2016). Similarly, areas with many closely spaced faults are also expected to have a relatively high density of fractures, i.e. high permeability.

Stress and strain factors (*dilation*, *normal*, *coulomb*):

The step-over in the Brady fault is an important factor controlling the presence of hydrothermal circulation at Brady (Faulds et al., 2003, 2010a, b; 2017; Siler et al., 2013; 2016; 2018; 2021). Stress and strain become concentrated at the step-over when slip occurs on the Brady fault, and the location of the stress and strain perturbation is largely concomitant with the production well field and the local temperature anomaly (Siler et al., 2018). The 2D modeled dilation (*dilation*), normal stress reduction (i.e., 'unclamping' of a fault) (*normal*) and coulomb shear stress increase (*coulomb*) as a result of 1-meter normal slip on the Brady fault are calculated at 250-meter-depth intervals from the surface to 750 m depth (Siler et al., 2018). *dilation*, *normal* and *coulomb* values between the calculated intervals are linearly interpolated, approximating the volumetric dilation, normal stress reduction, and coulomb shear stress increase in the study area. Siler et al. (2018) suggest that the stress and strain perturbations that occur with fault slip results in a zone of accentuated secondary faulting and fracturing that is an important factor in localizing hydrothermal circulation in the step-over.

Stratigraphic factors (*contactnear*, *unitthick*, *goodlith*):

In addition to the above structural variables, permeability associated with stratigraphic factors may also play an important role in localizing hydrothermal circulation. Stratigraphic contacts can be manifested as zones of breccia. These brecciated contact zones in successions of volcanic rocks, like occurring at Brady, may have matrix porosity and permeability that are important aspects of the flow system. The distance from the nearest stratigraphic contact (*contactnear*) is calculated as the difference between the distance

to the nearest stratigraphic contact along each well and the maximum distance to a contact in the dataset. In this case, high values of *contactnear* would be expected to correlate with hydrothermal fluid flow. Alternatively, relatively thick geologic units, i.e. relatively large, intact volumes of rock distal to stratigraphic contacts may focus strain on a relatively small number of dominant, high-aperture fractures. Areas with high values for the thickness of each stratigraphic unit (*unitthick*) from the 3D geologic map could be favorable for localizing hydrothermal circulation in this case. The ~300-600-m-depth production reservoir at Brady occurs in Miocene mafic to intermediate volcanic rocks. It is possible that these specific stratigraphic units have high matrix porosity and permeability and/or are particularly favorable for developing highly transmissive fracture systems when faulted. The *goodlith* variable is '1' for well intervals with these stratigraphic units and '0' for intervals in other units.

Temperature (*modeltemp*):

Advection is a much more efficient means of heat transport than conduction. Higher temperatures, therefore, are expected within or near transmissive fluid flow conduits. Equilibrated temperature logs from thirty-nine deep (as deep as ~2 km) geothermal wells and seventy-nine shallow (~150 m) temperature gradient wells (Shevenell et al., 2012) were utilized to build a 3D temperature model (Siler et al., 2021). The modeled temperature (*modeltemp*) is projected to each of the forty-seven wells.

3.3 NMFk methods

NMFk combines two unsupervised machine learning (ML) methods, non-negative matrix factorization (NMF) and customized *k*-means clustering. NMF factorizes a non-negative data matrix, *X*, into two components *W* and *H*, where *W* is a location matrix and *H* is an attribute matrix.

Given a non-negative data matrix $X = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathbb{R}^{m \times n}$, each column of *X* is a variable/sample vector, where *m* and *n* are the number of locations and attributes, respectively. NMF factorizes or decomposes *X* based on user-specified number of dimensions *k* into *W* and *H* matrices by minimizing the following loss function (Lee and Seung, 1999):

$$\mathcal{L} = \|X - WH^T\|_F^2 \quad (1)$$

where $\|\cdot\|_F$ denotes Frobenius norms. *H* can be considered as a basis matrix of *X* that is optimized for the linear approximation of *X*. Because only a few basis vectors represent all data vectors, good approximation vectors are those that capture the latent structure of *X*.

After completion of NMF process, 1,000 estimated *H* are clustered into *k* clusters using customized *k*-means clustering. Because *k* is also unknown in *k*-means clustering, the algorithm consecutively examines specified *k* by obtaining 1,000 *H* for each feature/variable. During clustering, the similarity between two variables is assessed according to the cosine norm. After clustering, the Silhouette value (Silhouettes, 1987) are calculated and used to estimate a particular choice of *k*. The Silhouette value quantifies how similar an object is to its own cluster compared to other clusters and varies from -1 to +1; high values indicate that the object is well matched to its own cluster and poorly matched to neighboring clusters. The combination of the least \mathcal{L} and the highest Silhouette value is used to determine the number of optimal clusters, or hidden signals. If *k* is low, the Silhouette value will be high, but so may be \mathcal{L} because of under-fitting. For high *k*, the Silhouette value will be low and the solution may be over-fit. So, the best estimate for *k* is a number that optimizes both the \mathcal{L} and Silhouette values.

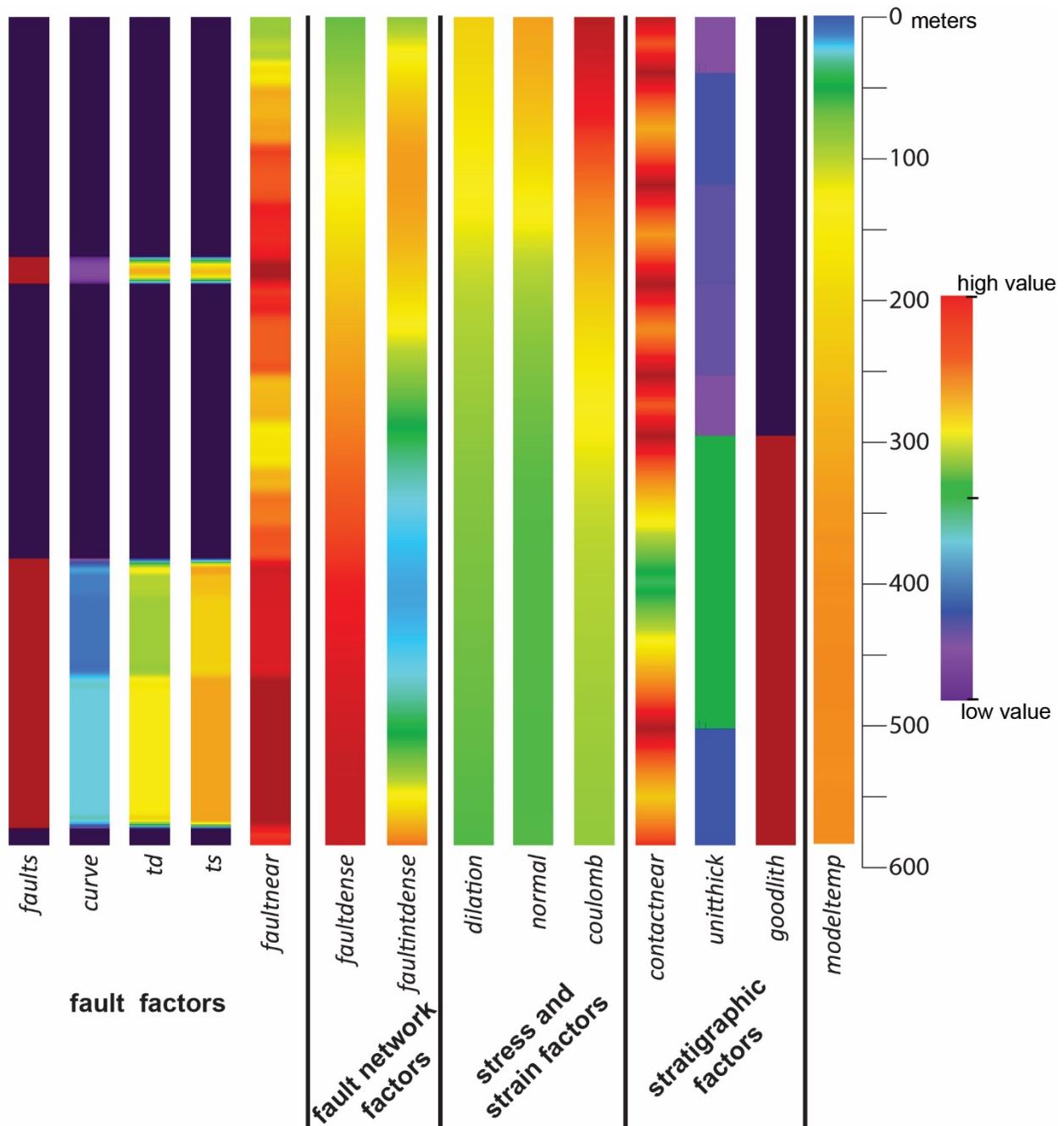


Figure 2. The 14 variables used in this study along one of the Brady production wells. Cool colors correspond to low values and warm colors correspond to high values for each variable. For the binary variables (fault and goodlith), red corresponds to a value of one (a fault or the producing lithology), purple corresponds to a value of zero. For faultnear and contactnear warm colors indicate nearness to faults or contacts

There are other matrix factorization tools such as singular value decomposition (SVD) (Klema & Laub, 1980) and principal and independent component analyses (PCA and ICA) (Wold & Geladi, 1987; Comon, 1994). There are few advantages of NMF k over the preceding matrix factorization tools, for instance, it handles both real and categorical variables, sparse datasets (with missing data entries), and provides interpretable results (Alexandrov et al. 2014; Vesselinov et al. 2018). Furthermore, NMF k has been used to discover hidden geothermal signatures at several geothermal sites in the U.S. (Ahmmmed et al. 2020a,b,c,d; Vesselinov et al. 2020a,b).

4. Results

As outlined above, NMF k analysis reveals associations within a complex dataset. In our case, these associations characterize interdependencies among geologic attributes and production, injection, and non-productive well locations within the analyzed 3D domain. In this analysis, the results with values of 2, 4, 5, and 6 for k are considered. The 2-cluster result was underfit. The 3-cluster result is rejected by NMF k algorithm because of its low Silhouette value. The 4, 5, and 6 cluster results are all robust solutions based on their \mathcal{L} and Silhouette values. Numbers of clusters larger than 6 were overfit. Each of the 4, 5, and 6 cluster results is a permissible solution and the H matrices for 4, 5, and 6 signals show very similar weighting patterns between variables (Figure 3). We chose to interpret the 4-cluster herein, since the smaller number of clusters is more easily interpretable in the framework of the geologic controls of hydrothermal fluid flow at Brady. Figure 3A shows the 4-cluster H (attribute) -matrix. Below, we use ‘signal’ to refer to the H -matrix rows and W -matrix columns (S1, S2, S3, and S4).

In addition to defining the four signals, a ‘cluster’ is defined for each well. Figure 4 shows the W (location) -matrix, the four signals (S1, S2, S3, S4) relative to each of the 47 geothermal wells, and the cluster (A, B, C, or D) that each well belongs to. Figure 4A and Figure 4B are the same matrix, Figure 4A is sorted by the cluster label, Figure 4B is sorted by S2 value. For both Figures 3 and 4, warm colors indicate a relatively high (strong) weight between the signal and the variable (Figure 3) or the signal and the well (Figure 4) and cool colors indicate a relatively low weight. Figure 5 shows the Brady well field and the cluster that each well falls into. Figure 6 shows a biplot of S1 vs S2. These two signals most effectively separate the production wells from the injection and non-productive wells. On Figure 6, the wells and variables are plotted by their S1 and S2 values from Figure 3A and Figure 4, respectively. The production wells (red) have relatively high S2 values and relatively low S1 values. The variables (plotted as asterisks) that plot in the same quadrant, i.e. also have relatively high S2 values and relatively low S1 values, are those that most effectively separate the production wells from the other wells.

5. Discussion

These results show that six of the nine wells that have been used for geothermal production at Brady from the shallow (~300-600 m depth) reservoir (June 1992-August 2019) fall in cluster C (Figures 4A and 5). Cluster C is associated with relatively high S2 values, and relatively low values for S1, S3, and S4 (Figure 3A). Additionally, all nine of the production wells fall within the seventeen highest S2 values (Figure 4B). This further indicates that relatively high S2 values are strongly associated with production wells relative to the other wells. NMF k results indicate that the fault factors, fault network factors, and stress and strain factors that are more dominant in S2 (and therefore predominate along the production wells) relative to injection wells and non-productive wells.

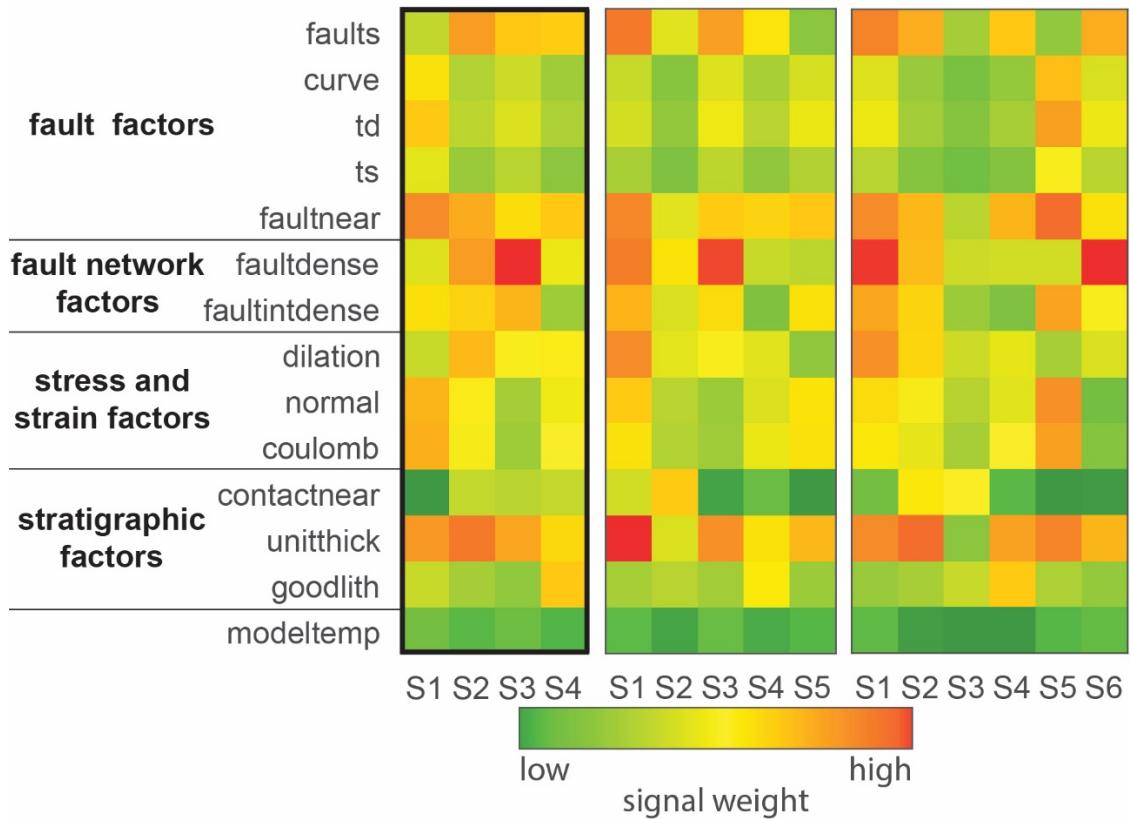


Figure 3. H (attribute) matrix for 4-signals, 5-signals and 6-signals. The 4-signal solution, which is interpreted here, is highlighted. Warm colors indicate that the variable has a high weight with that particular signal, cool colors indicate that the variable has a low weight with that specific signal.

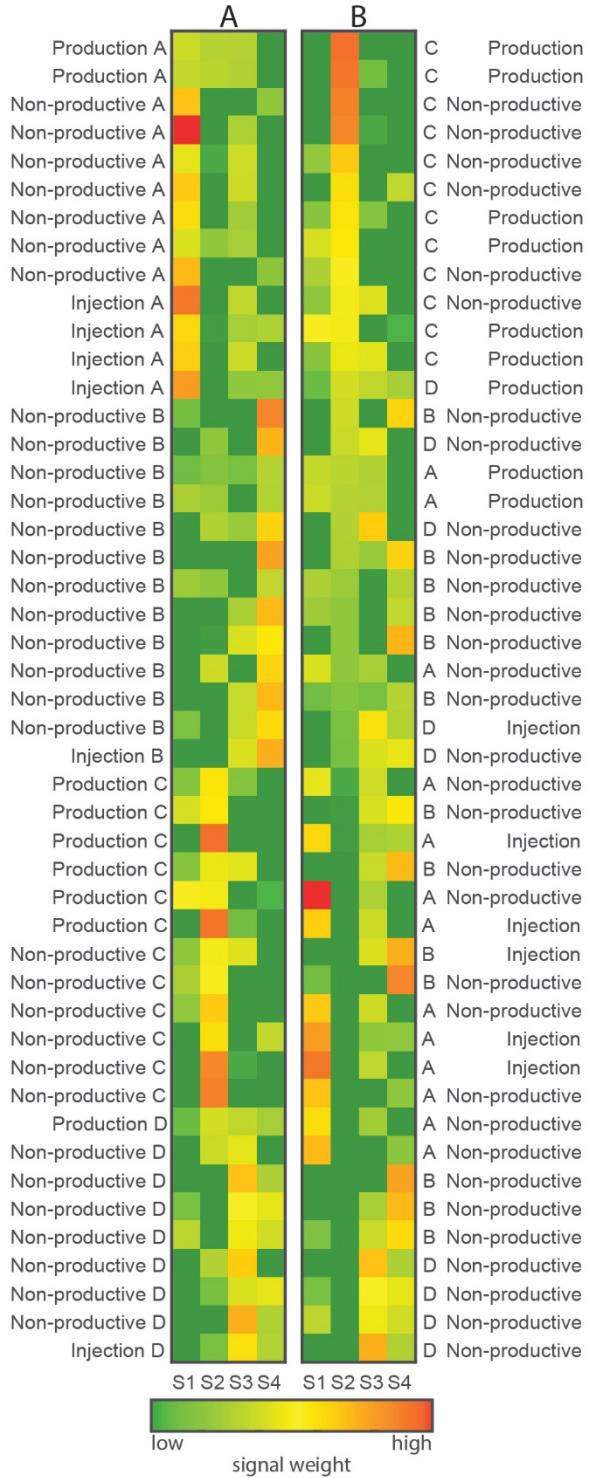


Figure 4 W (location) matrix. The four signals from the NMfk results relative to the forty-seven geothermal wells A) sorted by well clusters, B) sorted by S_2 values. Warm colors indicate that the variable has a high weight with that particular signal, cool colors indicate that the variable has a low weight with that particular signal. The well usage (production, injection or non-productive) and the well cluster labels (A, B, C, or D) are listed.

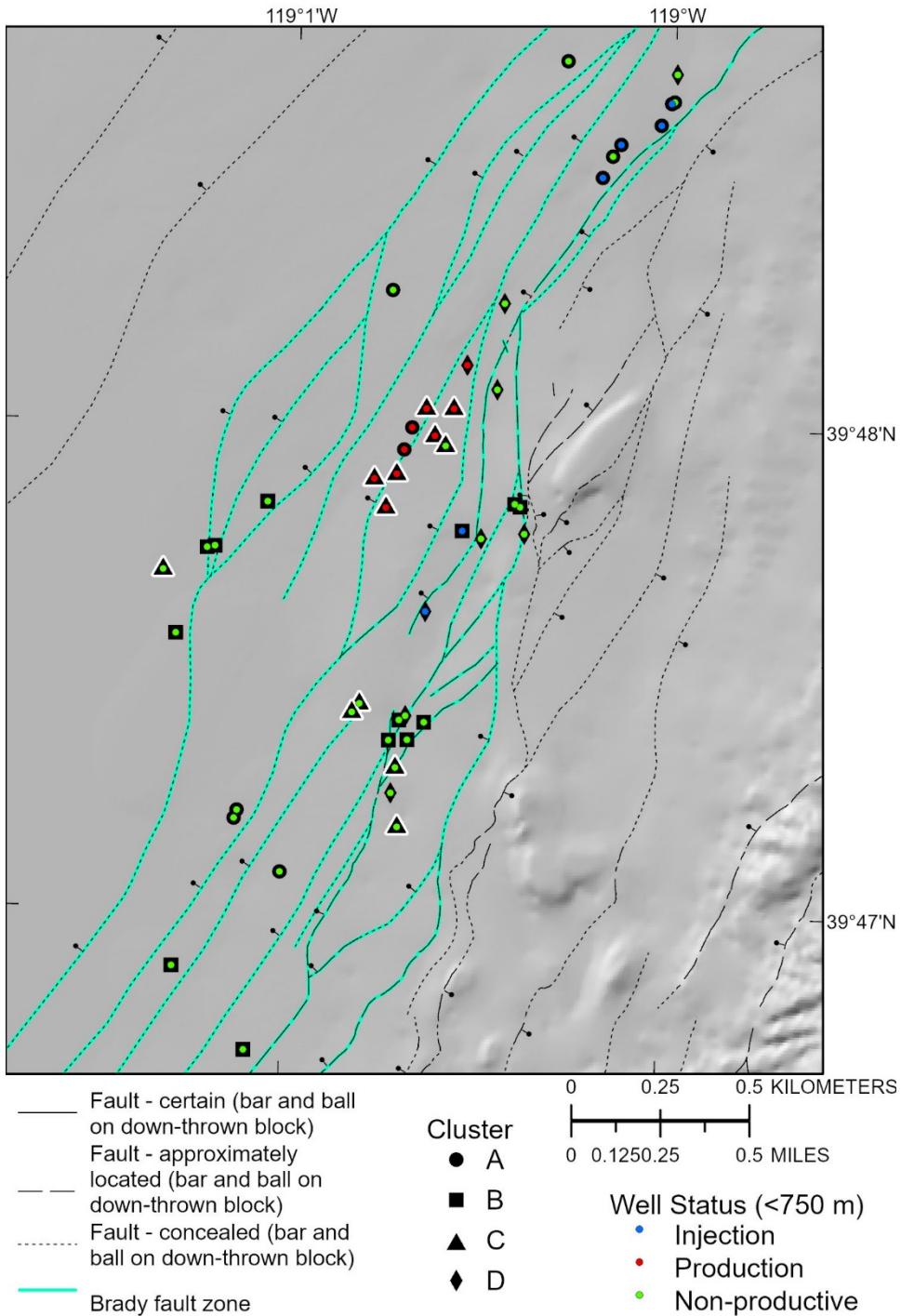


Figure 5. Map of the Brady well field and fault system. Wells are shown by their use (production, injection, or non-productive) and their cluster (A, B, C, or D). Cluster C (triangles), which contains six of the nine production wells is highlighted with a white halo.

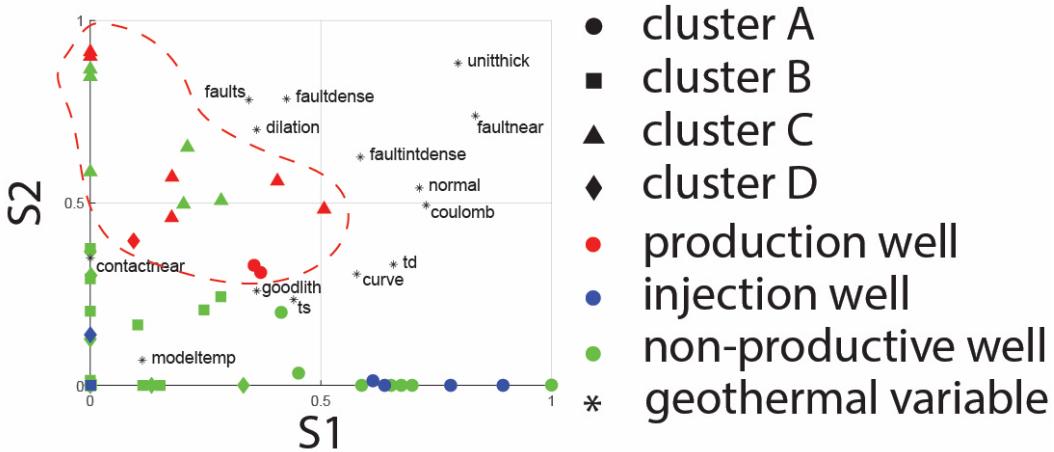


Figure 6. Biplot of signal-1 (S1) vs signal-2 (S2). Production wells (red) and the faults, dilation, and faultdense variables have high S2 values and relatively low S1 values indicating that these variables control the separation of the production wells from the rest of the data set.

Fault factors

The occurrence of faults intersecting a well, i.e. the *faults* variable, is the predominant faulting related factor associated with S2 (Figure 3A). This is evident on Figure 6 in which *faults* plots in the upper left quadrant, with relatively high S2 and relatively low S1; a similar pattern is observed to with the majority of the production wells. Though *faultnear* also has a high S2 values, it has relatively high values in S1, S3, and S4 (Figure 3A), so it less distinctly related to S2 relative to *faults*. On Figure 6 this is evident from *faultnear* plotting in the upper-right quadrant, farther to the right relative to the production wells. These results suggest that the presence or absence of distinct, macro-scale fault zones is strongly related to production wells, more so than to the other fault factors such as nearness to faults (*faultnear*), the curvature of faults (*curve*), slip tendency (*td*), or dilation tendency (*ts*).

Fault network factors

The spatial density of fault planes (*faultdense*) is the predominant fault network factor related to S2 and the production wells (Figure 3A). This is evident on Figure 6 in which *faultdense* plots in the upper left quadrant, with relatively high S2 and relatively low S1, similar to the majority of the production wells. Though *faultintdense* has a high S2 value, it also has relatively high S1 and S3 values (Figure 3A), and plots to the right of the production wells relative to *faultdense*. This indicates that fault density is more strongly related to production wells than fault intersection density.

Stress and strain factors

Dilation occurring as a result of modeled fault slip (*dilation*) is the predominant stress/strain factor related to S2 and the production wells. This is evident on Figure 6 in which *dilation* plots in the upper left quadrant, with relatively high S2 and relatively low S1, similar to the majority of the production wells. These results indicate that *dilation* is strongly related to production wells relative to *normal* or *coulomb*, the other stress/strain network factors examined herein.

Stratigraphic factors

The thickness of geologic units (*unitthick*) has high S2 values relative to S1 values (Figure 3A). Relative to the nearness to geologic contacts (*contactnear*), and the specific geologic units that are associated with geothermal production (*goodlith*), *unitthick* is more strongly related to S2 and the production wells (Figure 2A, 6). However, S1 values for *unitthick* are high relative to the production wells (*unitthick* plots in the upper right on Figure 6), so *unitthick* appears to be less strongly related to the production wells relative to *dilation*, *faultdense*, and *faults*.

Temperature

Temperature (*modeltemp*) has relatively low values for all signals. This suggests that the modeled temperature is not significantly higher or lower along any subset of wells relative to the others.

Geologic controls on hydrothermal processes at Brady

The NMFK results suggest that there are two dominant characteristics of the geologic structure that control hydrothermal processes at <750 m depth at Brady: the distinct, macro-scale faults and the step-over in the Brady fault system. The macro-scale faults are those that are mappable in 3D at the local scale using geologic and geophysical evidence (Siler et al. 2021). Interestingly, this relatively simple *fault* variable, the binary occurrence or non-occurrence of a 30-m-wide fault zone (Figure 2), is more closely related to the production wells than the static stress state of the faults (*td* or *ts*), the geometry of the faults (*curve*), or the nearness to the mapped fault planes (*faultnear*).

The step-over in the Brady fault zone (Figure 1) is the other dominant geologic factor controlling hydrothermal processes at Brady. The geometry and location of the step-over controls the spatial density of fault planes (*faultdense*), since faults are most dense in the step-over (Figure 1). The step-over also controls the dilatational strain that occurs as a result of modeled fault slip on the Brady fault zone (*dilation*) (Siler et al., 2018). The NMFK results suggest that the spatial density of fault planes and the modeled dilation are the most effective indicators of step-over's control on hydrothermal processes, relative to the other stress and strain variables (*normal* and *coulomb*) and the spatial density of fault intersections (*faultintdense*). Thicker geologic units (*unitthick*) may also influence hydrothermal processes. This may indicate that faults cutting through thicker geologic units preferentially transmit the high flow rates necessary for geothermal production relative to faults cutting thinner units. However, based on the ML analyses, this control appears to be secondary to the macro-scale faults and the step-over. The *modeltemp* variable is not strongly related to production wells relative to the other wells. It is likely that our extrapolation of the existing temperature data does not sufficiently resolve advective relative to conductive heat transport, and thus modeled temperature is of relatively ineffective for resolving discrete fluid flow pathways.

6. Conclusions

Non-negative matrix factorization with k-means clustering (NMFK) analyses was conducted on a 3D geological dataset from Brady geothermal field in order to elucidate the geologic characteristics that control hydrothermal circulation in the shallow (~300-600 m depth) geothermal reservoir. These analyses show that known, macro-scale faults, i.e. those that have been mapped in 3D based on geological and geophysical evidence, are strongly associated with the production wells at Brady. Geologic factors that occur most prominently within the Brady step-over, such as high spatial densities of faults, and dilatation brought on by modeled fault slip, are also associated predominantly with production wells relative to

other wells. These results suggest that the shallow hydrothermal reservoir at Brady is hosted by relatively prominent faults. Locations where such faults lie within the subsurface projection of the step-over; i.e. the volume of rock with relatively high fault and fracture density and where fractures tend dilate as a result of periodic fault slip, are especially well suited for geothermal production. These two factors, in concert and not either independently, control the presence of the Brady hydrothermal system that has been developed for electricity production and direct uses. The NMFK methodology successfully differentiates production wells from amongst a larger number of non-productive wells using these geologic data. This suggests that these geologic parameters may be effective as training data for using NMFK or other machine learning techniques to identify areas within an unexplored volume of the subsurface that have the geologic characteristics that constitute productive geothermal wells.

Acknowledgements

We thank Lowell Price at the Nevada Division of Minerals his help obtaining some of the data used in this study. This research was supported by the U.S. Geological Survey Energy and Minerals Program Geothermal Resources Investigations Projection (GRIP) and U.S. Geological Survey Mendenhall Research Fellowship Program. This work was also supported DOE EERE Geothermal Technologies Program (award number DE-EE3.1.8.1) to Los Alamos National Lab. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- Ahmmmed, 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.
- Ahmmmed, 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.
- Ahmmmed, 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.
- Ahmmmed, 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.
- Alberti M., 2011, 3D point cloud density calculation: a C++ program,
<https://gisoftw.blogspot.com/2011/05/3d-point-density-calculation-c-program.html>
- Alexandrov, B. S., & Vesselinov, V. V., 2014, Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization. Water Resources Research, 50(9), 7332-7347.
- Anders, M.H., and Wiltschko, D.V., 1994, Microfracturing, paleostress and the growth of faults: Journal of Structural Geology, v. 16, p. 795–815, [https://doi.org/10.1016/0191-8141\(94\)90146-5](https://doi.org/10.1016/0191-8141(94)90146-5)
- Ayling, B.F., 2020. 35 Years of Geothermal Power Generation in Nevada, USA: A Review of Field Development. Proceedings, Forty-Fifth Workshop on Geothermal Reservoir Engineering, Stanford University. V. 45, 12 p.

- Benoit, W.R. and Butler, R.W., 1983, A Review of High-Temperature Geothermal Developments in the Northern Basin and Range Province; Geothermal Resources Council, Special Report, No. 13, pp. 57-80.
- Blackwell, D.D., 1983. Heat flow in the northern Basin and Range province. Geothermal Resources Council, Special Report 13 81–93.
- Caine, J.S., Evans, J.P., Forster, C.B., 1996. Fault zone architecture and permeability structure. *Geology* 24, 1025–1028, [https://doi.org/10.1130/0091-7613\(1996\)024%3C1025:FZAAPS%3E2.3.CO;2](https://doi.org/10.1130/0091-7613(1996)024%3C1025:FZAAPS%3E2.3.CO;2)
- Caine, J.S., Forster, C.B., 1999. Fault Zone Architecture and Fluid Flow: Insights From Field Data and Numerical Modeling. *Geophysical Monograph: American Geophysical Union* 113, 101–128.
- Colgan, J.P., Dumitru, T. a., Reiners, P.W., Wooden, J.L., Miller, E.L., 2006. Cenozoic Tectonic Evolution of the Basin and Range Province in Northwestern Nevada. *American Journal of Science* 306, 616–654. <https://doi.org/10.2475/08.2006.02>
- Comon, P., 1994, Independent component analysis, A new concept? *Signal Processing*, 36(3).
- Fairley, J.P., Heffner, J., Hinds, J., 2003. Geostatistical evaluation of permeability in an active fault zone. *Geophysical Research Letters* 30. <https://doi.org/10.1029/2003GL018064>
- Fairley, J.P., Hinds, J.J., 2004. Rapid transport pathways for geothermal fluids in an active Great Basin fault zone. *Geology* 32, 825–828. <https://doi.org/10.1130/G20617.1>
- Faulds, J.E., Garside, L.J., Oppliger, G., 2003. Structural analysis of the Desert Peak–Brady geothermal fields, northwest Nevada: implications for understanding links between northeast-trending structures and geothermal reservoirs in the Humboldt structural zone: Geotherm Resources Council Transactions, v. 27, pp. 859–64.
- Faulds, J.E., Coolbaugh, M.F., Benoit, W.R., Oppliger, G.L., Perkins, M., Moeck, I., Drakos, P.S., 2010a. Structural Controls of Geothermal Activity in the Northern Hot Springs Mountains, Western Nevada: The Tale of Three Geothermal Systems (Brady's, Desert Peak, and Desert Queen). *Geothermal Resources Council Transactions* 34, 675–684.
- Faulds, J.E., Moeck, I., Drakos, P.S., Zemach, E., 2010b. Structural Assessment and 3D geologic modeling of the Brady's geothermal area, Churchill County (Nevada, USA): A preliminary report. *Proceedings, Thirty-Fifth Workshop on Geothermal Reservoir Engineering*, Stanford University. 298–302.
- Faulds, J.E., Ramelli, A.R., Coolbaugh, M.F., Hinz, N.H., Garside, L.J., Queen, J.H., 2017. Preliminary Geologic Map of the Bradys Geothermal Area, Churchill County, Nevada. *Nevada Bureau of Mines and Geology, Open-File Report 17-4*, scale 1:12,000.
- Ferrill, D.A., Winterle, J., Wittmeyer, G., Sims, D., Colton, S., Armstrong, A., Horowitz, A.S., Meyers, W.B., Simons, F.F., 1999, Stressed rock strains groundwater at Yucca Mountain, Nevada: *GSA Today*, v. 9 p. 1–8.
- Fossen, H., Rotevatn, A., 2016. Fault linkage and relay structures in extensional settings—A review. *Earth-Science Reviews* 154, 14–28.

- Jolie, E., Moeck, I., and Faulds, J.E., 2015, Quantitative structural-geological exploration of fault-controlled geothermal systems—A case study from the Basin-and-Range Province, Nevada (USA): *Geothermics*, v. 54, p. 54–67, <https://doi.org/10.1016/j.geothermics.2014.10.003>
- Klema, V. C. & Laub, A. J., 1980, The singular value decomposition: Its application and some applications. *IEEE Transactions on Automatic Control*, 25:2, 164-176.
- Lachenbruch, A.H., Sass, J.H., 1977. Heat flow in the United States and the thermal regime of the crust. In: Heacock, J.G. (Ed.), *The Earth's Crust*, 20. American Geophysical Union Monograph, 626–675.
- Lee, D.D., Seung, H.S. 1999, Learning the parts of objects by non-negative matrix factorization: *Nature*, v. 401, no. 6755, p. 788–791. <https://doi.org/10.1038/44565>
- Morris, A., Ferrill, D.A., and Henderson, D.B., 1996, Slip-tendency analysis and fault reactivation: *Geology*, v. 24, p. 275–278.
- Peacock, D.C.P., Sanderson, D.J., 1991. Displacements, segment linkage and relay ramps in normal fault zones. *Journal of Structural Geology* 13, 721–733. [https://doi.org/10.1016/0191-8141\(91\)90033-F](https://doi.org/10.1016/0191-8141(91)90033-F)
- Sanderson, D.J., Zhang, X., 2004. Stress-controlled localization of deformation and fluid flow in fractured rocks. Geological Society, London, Special Publications 231, 299–314.
- Scholz, C.H., Dawers, N.H., Yu, J., Anders, M.H., and Cowie, P.A., 1993, Fault growth and fault scaling laws—Preliminary Results: *Journal of Geophysical Research*, v. 98, p. 951–961.
- Shevenell, L.A., Oppliger, G., Coolbaugh, M.F., Faulds, J.E., 2012. Bradys (Nevada) InSAR Anomaly Evaluated with Historical Well Temperature and Pressure Data. *Geothermal Resources Council Transactions* 36, 1383–1390.
- Sibson, R.H., 1994, Crustal stress, faulting and fluid flow, in: *Geofluids: origin, migration and evolution of fluids in sedimentary basins*, editor Parnell, J., Geological Society, London, Special Publications, pp. 69–84.
- Siler, D.L., Faulds, J.E., Hinz, N.H., Queen, J., M., 2021 *in review*, Three-dimensional geologic of the Brady Geothermal Area, Nevada: U.S. Geological Survey Scientific Investigations Map xxxx, pamphlet xx p.
- Siler, D.L., Faulds, J.E., 2013. Three-dimensional geothermal fairway mapping: Examples from the western Great Basin, USA. *Geothermal Resources Council Transactions*.
- Siler, D.L., Hinz, N.H., Faulds, J.E., 2018. Stress concentrations at structural discontinuities in active fault zones in the western United States: Implications for permeability and fluid flow in geothermal fields. *Geological Society of America Bulletin* 130. <https://doi.org/10.1130/B31729.1>
- Siler, D.L., Hinz, N.H., Faulds, J.E., Queen, J., 2016. 3D Analysis of Geothermal Fluid Flow Favorability: Brady's, Nevada, USA. The 41st Workshop on Geothermal Reservoir Engineering, Stanford University 41, 10.
- Silhouettes, P.J.R., 1987. A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

- Vesselinov, V., Ahmmmed, 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., Ahmmmed, 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. V., Mudunuru, M. K., Karra, S., O'Malley, D., & Alexandrov, B. S., 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., & O'Malley, D., 2018, Contaminant source identification using semi-supervised machine learning. *Journal of contaminant hydrology*, 212, 134-142.
- Wernicke, B., 1992. Cenozoic extensional tectonics of the U.S. Cordillera. In: Burchfiel, B.C., Lipman, P.W., Zoback, M.L. (Eds.), *The Cordilleran Orogen: Conterminous U.S.: The Geology of North America*. Geologic Society of America, Boulder, CO, 553–581.
- Witter, J.B., Siler, D.L., Faulds, J.E., and Hinz, N.H., 2016, 3D geophysical inversion modeling of gravity data to test the 3D geologic model of the Bradys geothermal area, Nevada, USA: *Geothermal Energy*, v. 4, no. 14, 21 p., <https://doi.org/10.1186/s40517-016-0056-6>
- Wold, E., and Geladi, S., 1987, Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3), 37-52.