**Which geologic factors are associated with production in geothermal fields? Non-negative matrix factorization with k-means clustering (NMF*k*) analysis of 3D geologic data, Brady geothermal field, Nevada**

Drew L. Siler1, Jeff D. Pepin2, Velimir V. Vesselinov3, Mudunuru4 Bulbul Ahmmed3

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

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

# Abstract

# Introduction

Permeability in the crust can be highly variable in space (e.g., Caine et al., 1996; Caine and Forster, 1999; Fairley et al., 2003; Fairley and Hinds, 2004; Sanderson and Zhang, 2004). Accordingly, it is common in developed geothermal systems for production fluid flow to occur 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 within geothermal fields, the volume of rock that transmits fluids at volumes suitable for power production is much smaller than the volume of rock that does not transmit fluid (or transmits at sub-commercial volumes). This presents a significant challenge to the efficient exploration, development, and maintenance of these renewable energy resources. The compartmentalization of the fluid flow system may be associated with spatial changes in fracture permeability along faults, and/or with permeability variation in the stratigraphic succession. The purpose of this study is to identify the geologic factors that control the compartmentalization of fluid flow in hydrothermal systems. We evaluate this through unsupervised machine learning based on non-negative matrix factorization coupled with k-means clustering (NMF*k*)analysis of three-dimensional (3D) geologic characteristics. 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, N.V. The results elucidate the geologic factors that are most closely associated with production wells. Tracking the 3D distribution of these factors in new geothermal prospects and in developed geothermal fields may help promote more efficient resource development and reservoir maintenance. The developed ML analyses facilitate the classification (labeling) of the subsurface geothermal conditions. Furthermore, they can also be applied to make predictions of geothermal conditions at new drilling locations based on the available

# 2. Background

## 2.1 The Brady geothermal system

Brady geothermal field, in northwestern Nevada, U.S.A., has seen geothermal electricity production since 1992 (Figure 1) 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 supplied 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 at relatively shallow levels (300-600 depth for some production wells) occur due to either convective circulation driving upwelling, or hydraulic-head driven circulation through the hot rock. In either case, relatively high heat flow in the Basin and Range, which is associated with Miocene to recent crustal thinning (Lachenbruch and Sass, 1977; Blackwell, 1983), provides that heat for the Brady geothermal field. The fluids circulate through a system of transmissive pathways that have been primarily attributed to a fracture system in a step-over in the Basin and Range-type (e.g., Wernicke, 1992; Colgan et al., 2006) normal fault, the Brady fault (Faulds et al., 2003, 2010a,b; 2017; Siler et al., 2018).

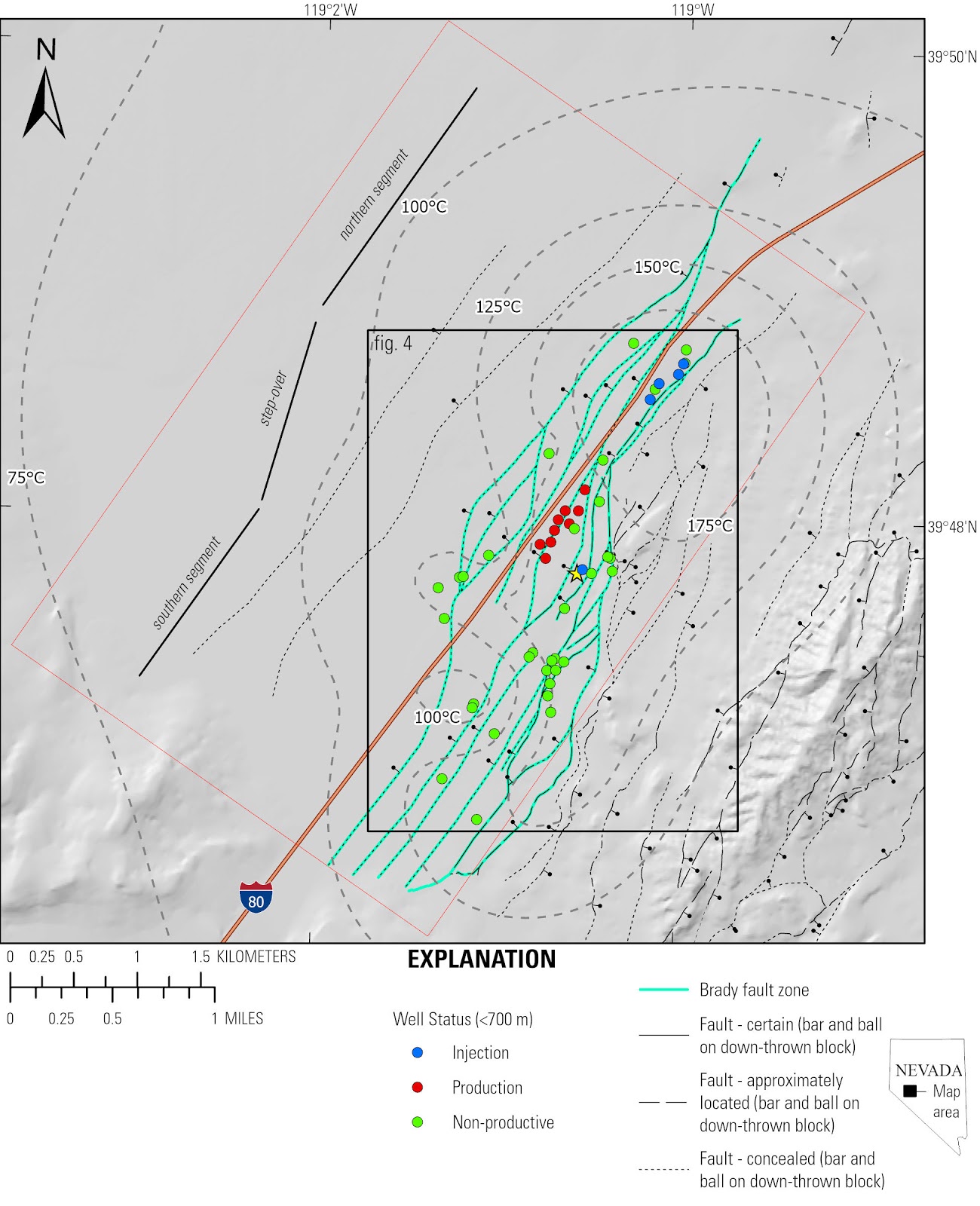


Figure 1. Map of Brady geothermal area. The fault strands that constitute the Brady fault are shown in green. Contours represent modeled temperature at 750 depth. Wells are colored by their usage (production, injection, and non-productive) for depths shallower than 750 m. The general geometry of the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

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 Brady fault zone steps to the left to meet the northern segment (Figure 1; Faulds et al., 2017). Siler et al. (2018) suggested the localization 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).

## 2.2 NMF*k* analysis

TEXT ON NMFk background here

# 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, 2020; Witter et al., 2016) was used to develop a suite of geothermal variables that may control the distribution/localization of production-grade fluid flow. The fourteen different variables described below are calculated and projected to forty-seven production, injection, and non-productive wells within the field (Figure 1).

## 3.1 Geothermal variables

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

For each of the thirty-two faults defined by the 3D geologic map (Siler and Faulds, 2013a; Siler et al., 2016, 2020; 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 and is consistent with empirically derived fault zone widths for kms-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 segment is located within a fault zone and ‘0’ for segments of well 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, calculated for each fault. These values are computed using methods of Morris et al. (1996) and Ferrill et al. (1999) and a local stress model developed at Brady (Jolie et al., 2015). The 30-meter-wide fault zones for each fault are populated with *curve*, *ts*, and *td* values. Segments of faults with a high value for *curve* may be associated with accentuated faulting and fracturing as a result of stress loading at the highly curved fault segments (Sibson, 1994), and may host fluid flow. Dilation tendency and slip tendency are the ratios of the resolved normal stress and total stress on faults, respectively. Fault segments that are either highly dilatant (high *td*) or stress loaded for slip (high *ts*) are likely to host preferential fluid flow (Siler et al., 2020). For all locations along the wells, the *faultnear* variable is reported as the difference between the distance to the nearest 3D mapped fault and the maximum distance to a fault in the dataset. This is done so that the high *faultnear* values occur at segments of wells near to faults (e.g., Figure 2), in the same way that high values for the other variables occur where hydrothermal processes are expected.

### Fault density factors (*faultdense*, *faultintdense*):

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 3-D space (Siler et al., 2020). 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 fracture density.

### 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; 2020). 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 concomitant with the geothermal field and the local temperature anomaly (Siler et al., 2018). Siler et al. (2018) suggest that the stress and strain perturbation results in a zone of accentuated secondary faulting and fracturing that is an important factor in localizing hydrothermal circulation in the step-over. The 2-D modeled dilation (*dilation*), normal stress reduction (*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 1000 m depth. These surfaces approximate the volumetric dilation, normal stress reduction, and coulomb shear stress increase in the study area.

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

In addition to the above structural variables, permeability associated with stratigraphic factors may play an important role in localizing hydrothermal circulation. Stratigraphic contacts in volcanic sequences can be manifest as zones of breccia. These brecciated contact zones 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 closest 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 production. Alternatively, intact volumes of rock, i.e., areas in thick geologic units, far from stratigraphic contacts, may focus strain on a relatively small number of dominant, high-aperture, fractures. Areas with high values for *unitthick,* the thickness of each stratigraphic unit from the 3D geologic map, would be favorable for localizing hydrothermal circulation in this case. As another alternative, the two discrete production reservoirs at Brady (~300-600 m and ~1750 m depth) occur in particular stratigraphic units. The shallower reservoir occurs in a relatively thick section Miocene mafic to intermediate volcanic rocks. It is possible that these units have high matrix porosity and/or permeability and/or are particularly favorable for developing highly transmissive fracture systems when faulted. The *goodlith* variable returns ‘1’ for well intervals in one of 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 very 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., 2020). The modeled temperature (*modeltemp*) is projected to each of the forty-seven wells.

## 3.2 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 at the site). As a result, this analysis focuses on the shallow (~300-600 m deep) reservoir, where the well data are denser. 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 analyzed 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 managed by Nevada Division of Minerals). We consider the remaining thirty-two wells to be non-productive wells. Though four of these remaining thirty-two wells have significant volumes of production or injection, these wells produce or inject in the deeper reservoir, where the data are not sufficiently dense to perform rigorous ML analyses. Since the present study stops at 750 m depth, where these four wells are cemented and do not transmit fluids to or from the formation, these wells are interpreted here as ‘non-productive.’ Each of the above fourteen geologic factors is calculated at 1-m-intervals along all the fifty-one wells. The resultant dataset of 336,784 entries (24,056 locations with fourteen variables) is used as input for NMF*k* analysis. Figure 2 shows the values for each of the 14 variables for one of the production wells.

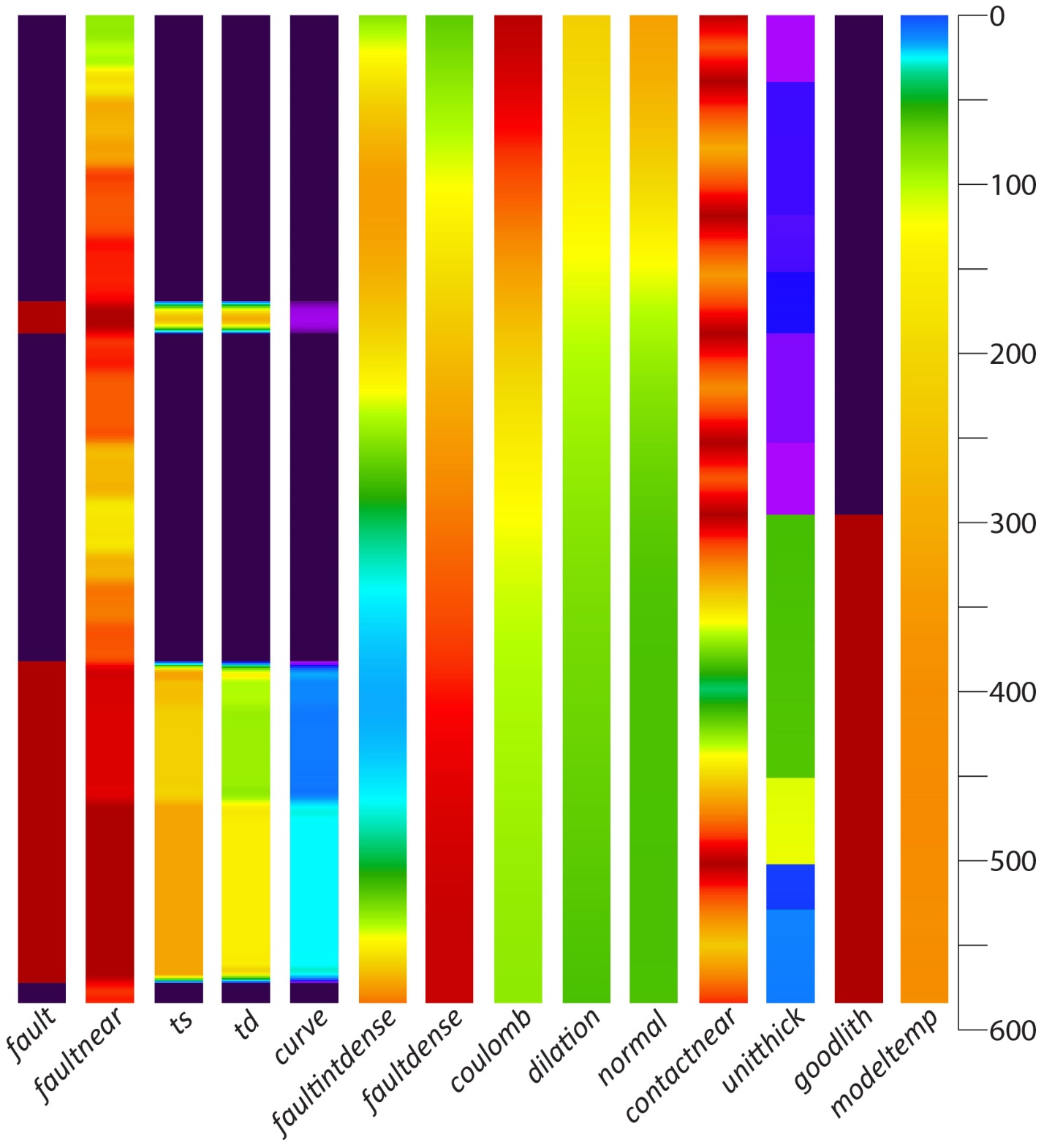


Figure 2. The 14 variables used in this study along one of the Brady production wells. Cold colors correspond to low values and warm colors correspond to high values for each variable. For faultnear and contactnear warm colors indicate nearness to faults or contacts

## 3.3 NMF*k* methods

NMF*k* combines two unsupervised machine learning methods, such as non-negative matrix factorization (N.M.F.) and customized *k*-means clustering methods. N.M.F. factorizes non-negative data matrix, X, into two reduced-order matrices W and H where W is a mixing matrix while H is an attribute matrix.

Given a non-negative data matrix , each column of *X* is a variable / sample vector, where and are the number of rows and columns, respectively. N.M.F. factorizes or decomposes *X* based on user-specified number of dimensions into and matrices by minimizing the following loss function (Lee and Seung, 1999):

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

where denotes Frobenius norms. Equation 1 is iteratively (for this work 1,000) solved as follows (cite Lee and Sung):

|  |  |  |
| --- | --- | --- |
|  |  | (2) |
|  |  | (3) |

In reality, and , so N.M.F. finds compressed approximation of X and column by column approximation is as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

where is the column vector of matrix *W* suggesting that is approximated by the linear combination of the columns of W, weighted by the components of H. Therefore, W 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; therefore, good approximation vectors are those that capture the latent structure of *X*. There are few advantages of NMF over analogous matrix factorization tools such as singular value decomposition (SVD), principal and independent component analyses (PCA and ICA). First, NMF is an additive combination of many different basis and second, it is a part-based representation. These two factors assist in interpreting the NMF outcomes better than SVD, PCA, and ICA

After completion of the NMF process, the 1,000 initial guesses of *H* are clustered into *k* clusters using customized *k*-means clustering. Because *k* is also unknown in *k* -means clustering, the algorithm consecutively examined 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 (Silhouttes, 1987) is calculated and uses 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 the neighboring clusters. The combination of the least and the highest Silhouette value are used to determine the number of optimal hidden signals. If *k* is low, the Silhouette value will be high, but so may be 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 and Silhouette values.

# 4. Results

As outlined above, NMFk analysis reveals associations within a complex dataset. These associations characterize interdependencies among geothermal attributes and well locations within the analyzed 3-D domain. The analysis describes these associations in terms of *signals* (or hidden features) associated with the analyzed geothermal variables and *clusters associated with the well locations*. Typically, the numbers of signals and clusters are equal. NMFk estimates the optimate number of signals/clusters. Based on the NMFk output, for the Brady dataset, four signals represent a robust solution with a relatively high fit to the data. For each geothermal variable and each well, a cluster is defined based on the extracted signals. Figure 3 shows the H-matrix, a plot of the four signals (S1, S2, S3, or S4) and the relative influence of each geothermal variable on that signal. Figure 3 also shows the cluster (A, B, C, or D) to which each geothermal variable belongs.

Figure 4 shows the W matrix, a plot of the four signals relative to each of the forty-seven geothermal wells, and the cluster that each well belongs to. For both Figures 3 and 4, warm colors indicate a relatively strong correlation between the signal and the variable or well, and cool colors initiate a relatively weak correlation between the signal and the variable or well. For instance, Figure 3 shows that *normal* has a relatively strong correlation with S3 and relatively weak correlation with S1, whereas Figure 4 shows that the cluster A wells are defined by high values for S1 and low values for S2.

Figure 5 shows the NMF*k* results as biplots; one signal relative to another. In each of the Figure 5 plots, the forty-seven wells are plotted by normalized W-matrix value (Figure 4 value), their cluster association (A, B, C, or D), and their usage (production, injection, non-productive). For each variable, the normalized H-matrix values (i.e., Figure 3) are plotted for each signal. Variables that plot in the vicinity of a particular cluster of wells control the differentiation of that cluster relative to the other wells. Figure 6 shows the clustering for each of the fifty-one wells by well usage. Figure 7 shows a map of the Brady area with the locations of the wells, their use (production, injection, or non-productive), and their cluster (A, B, C, and D).

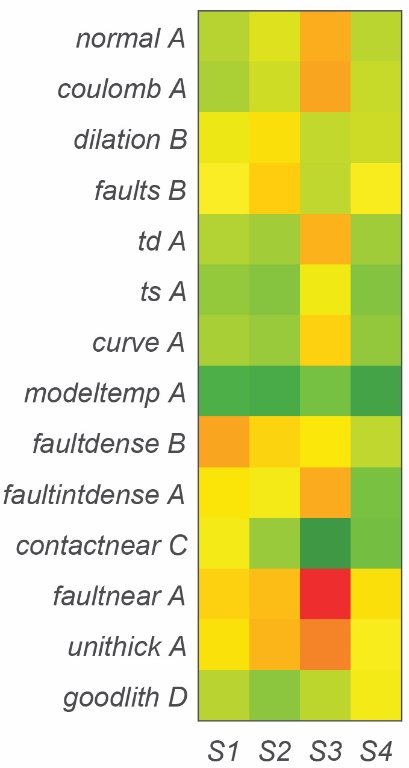


Figure 3. H-matrix. The four signals from the NMFk results relative to the fourteen geothermal variables. Warm colors (red) indicate that the variable is correlated with that particular signal, cold colors (green) indicate that the variable is not correlated with that specific signal. The cluster association (A, B, C, or D) for each variable is listed.

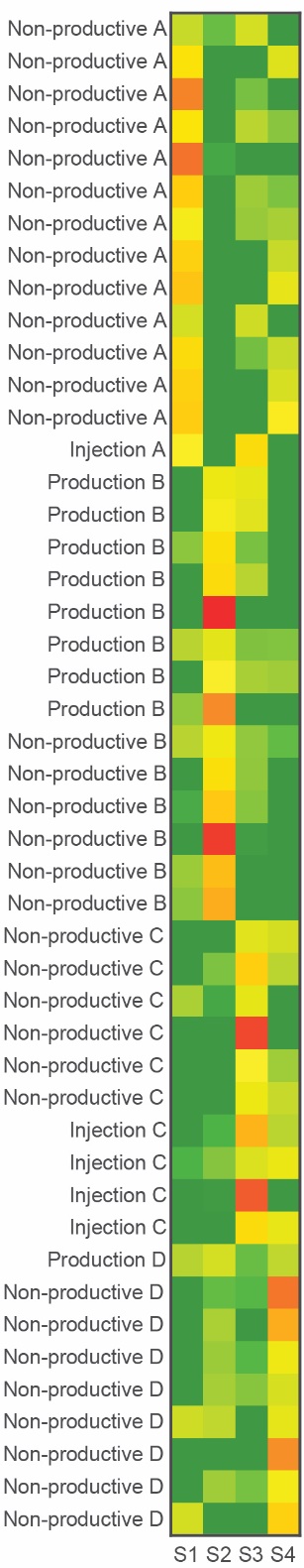


Figure 4 W-matrix. The four signals from the NMFk results relative to the forty-seven geothermal wells. Warm colors indicate that the well is correlated with a particular signal; cold colors suggest that the well is not correlated with a particular signal. The usage of each well (production, injection or non-productive) and the cluster (A, B, C, or D) for wells is listed.

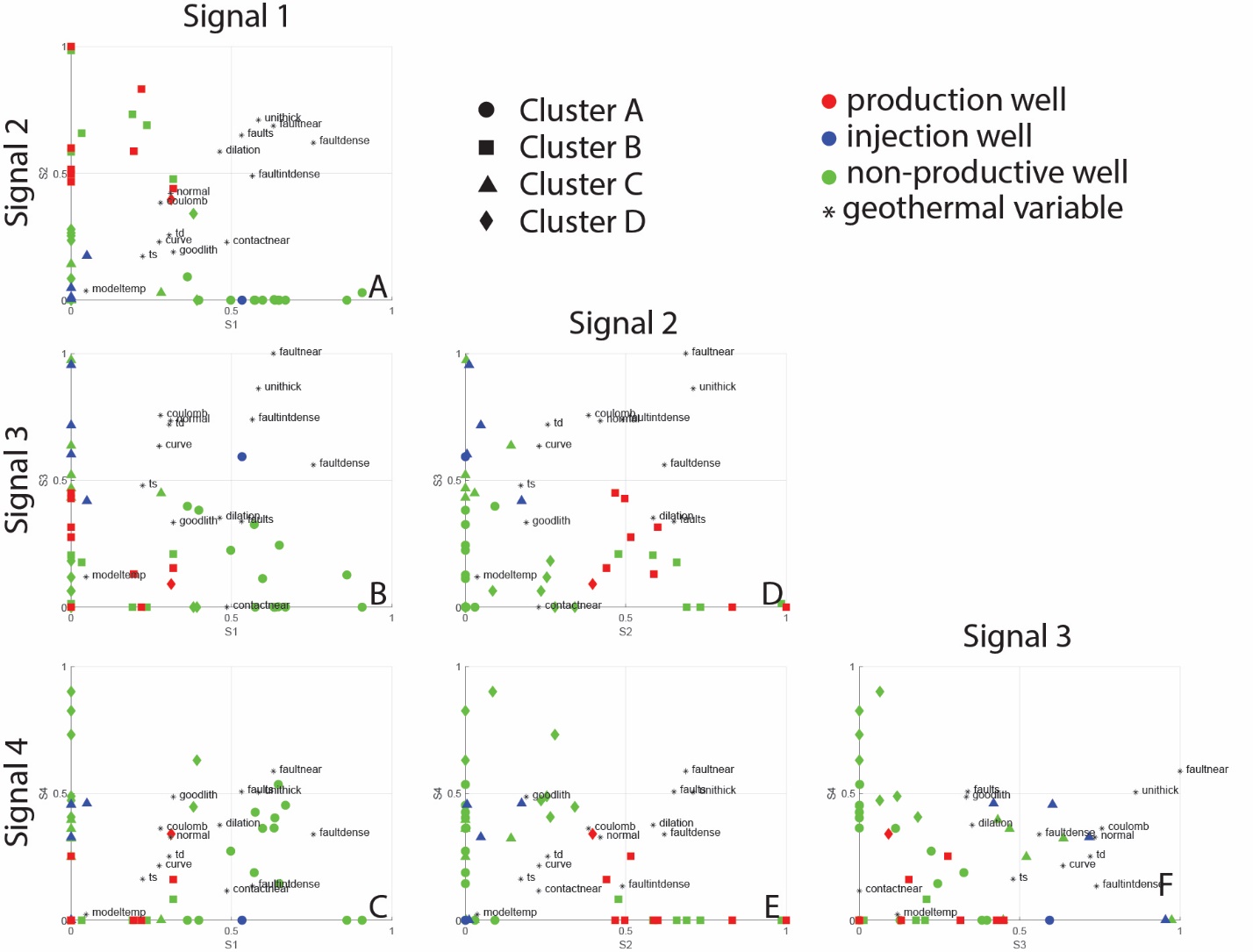


Figure 5. Biplots showing the normalized signal value for each well and each variable. The symbol for each well corresponds to their cluster (A, B, C, or, D) and each well is colored by usage (production, non-productive, injection). The variables are plotted as (\*) symbols. Where a cluster of wells plots in the same area as variable(s), those variable(s) control the relative differentiation of that cluster of wells from other wells.

# 5. Discussion

NMF*k* results show that eight of the nine wells that have been used for geothermal production at Brady from the shallow (~300-600 m depth) reservoir are in cluster B (Figures 4, 5, and 7). The production well that is not in cluster B was used for production only during the first five years after the power station was brought online (1992-1997). Thus, cluster B contains all the wells that have produced from the shallow reservoir at Brady for the most recent twenty-one years (1997-2018). Cluster A consists of one injection well and thirteen non-productive wells. Cluster C consists of four of the five injection wells, and six non-productive wells. Cluster D contains one production well and eight non-productive wells (Figure 6). The clustering of production wells (cluster B), injection wells (cluster C), and non-productive wells (clusters A and D) suggests that the three types of wells are largely distinct from one another with respect to the geologic characteristic examined herein.

## Cluster A

Cluster A contains thirteen non-productive wells and one injection well (Figure 6). The cluster A wells are mostly located on the eastern side of the Brady fault zone, though one is located on the western side (Figure 7). Cluster A is most strongly associated with high values for S1 (Figures 4, 5A, 5B, 5C) and relatively low values for S2, S3, and S4). The H-matrix (Figure 3) shows that *normal*, *coulomb, td, ts,* *curve,* *modeltemp, faultintdense, faultnear,* and *unitthick* are dominant variables of cluster A. Figure 5A, 5B, and 5C show that *contactnear, faultdense, faults, unitthick,* and *faultnear* are the variables that plot near, or in the same quadrant as cluster A. The location of many of the cluster A wells on the east side of the west-dipping Brady fault zone may explain the low values for *ts, td, curve, faultnear,* and *faultintdense*. These wells penetrate through the west-dipping Brady fault zone at relatively shallow levels, and thus much of their lengths intersect relatively few faults and fault intersections. For the cluster A well that is on the west side of the Brady fault zone (Figure 7), this well appears to be too far west to reach the main strands of the fault zone at less than 750 m depth, the depth cut off for this study. Thus, it may also penetrate relatively few faults like the rest of cluster A.

## Cluster B

Cluster B contains eight production wells and six non-productive wells (Figure 6). Four of the cluster B production wells have been the dominant producers from the shallow reservoir since 1997. The other four are wells that produced at various times 1992-2018, though predominantly in the earlier years of field operation (based on Nevada Division of Minerals, publicly available data). Nine of the fourteen cluster B wells lie within the step-over in the Brady fault zone, four lie to the south, and one is located to the west-southwest of the step-over. Cluster B is most strongly associated with high values for S2 (Figure 4, 5A, 5D, 5E) and low values for S1, S3, and S4. Figures 5A, 5D, 5E show that *dilation*, *faults*, *normal*, *coulomb,* *faultdense*, and *faultintdense* variables plot in the vicinity of, or in the same quadrant as the cluster B wells. The H-matrix (Figure 3) indicates that *dilation, faults,* and *faultdense* are dominant variables for cluster B. The step-over in the Brady fault zone probably accounts for the high *dilation*, *faults*, *normal*, *coulomb* *faultdense*, and *faultintdense* values for cluster B wells, since these variables are expected to be high in the step-over, where there are the most fault strands and where stress and strain effects as a result of modeled slip are focused (Siler et al., 2018).

## Cluster C

Cluster C contains six non-productive wells, four injection wells. The four injection wells include the two injection wells at the northern end of the Brady fault zone that have accounted for the majority of injection in the field since ~1997 and two others that have been used for re-injection at various times 1992-2018 (based on Nevada Division of Minerals, publicly available data). The cluster C wells are located at the southern and northern ends of the Brady fault zone, all outside of the Brady step-over. Cluster C is associated with relatively high S3 values and low values for S1, S2, and S4 (Figure 4, 5C, 5E, 5F). *contactnear* is the only S3 variable (Figure 3). Figure 5 shows *curve, td, normal, coulomb, faultdense* are the variables that plot near or in the same quadrant at Cluster C.

## Cluster D

Cluster D contains eight non-productive wells and one production well. The cluster D production well was used between 1992-1994 but has not been used for production since (based on Nevada Division of Minerals, publicly available data). Four of the cluster D wells are on the western side, two at the southern end, and two are located on the eastern side of the Brady fault zone. One is located in the center of the Brady fault zone, at the northern end of the step-over. Cluster D is associated with high S4 and low S1, S2, and S3 (Figure 4). Figure 4 shows that *goodlith* is a cluster D variable. The biplots show *goodlith* and *faults* are the variables that plot near or in the same quadrant at cluster D (Figures 5C, 5E, 5F).

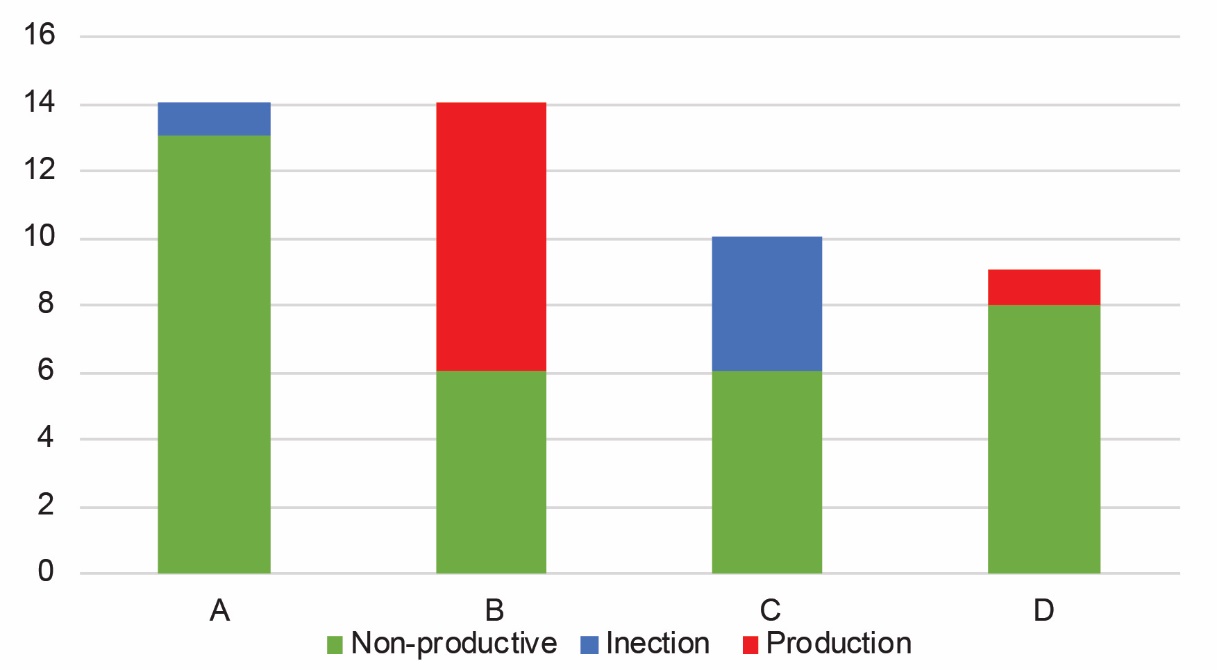


Figure 6. Clustering of wells by well use.

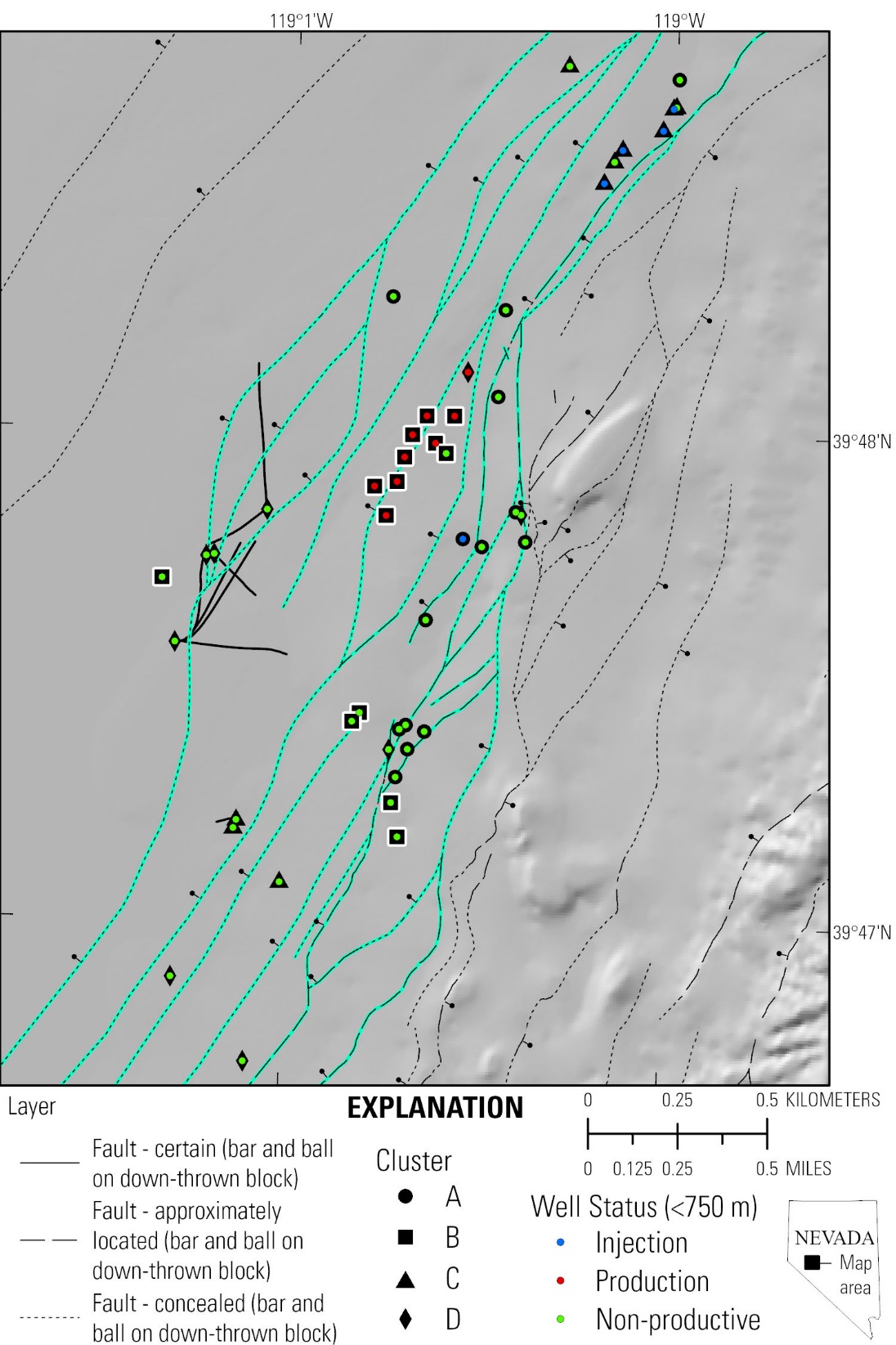


Figure 7. Map of the Brady well field and fault system. Wells are shown by their use (production, injection, or non-productive) and their cluster (A, B, C, or D). Cluster B, which contains eight of the nine production wells, is highlighted with a white halo.

## Structural factors

The clustering of eight of the nine production wells, and all of the wells that have been used for production 1997-2018 into cluster B, strongly indicates that there are geologic characteristics that are shared between production wells that are not as prevalent in other wells. The H-matrix (Figure 3) indicates that several faulting related (*dilation, faults, faultdense, faultnear*, and *unitthick)* variables are the most important in grouping cluster B wells. Figure 5A, 5D, and 5E show *normal*, *coulomb*, and *faultintdense*, along with *dilation, faults, and faultdense,* plot in the same general areas (relatively high S2 and low S1, S3, and S4) as the cluster B wells, and thus differentiate these wells from the rest of the dataset.

The definition of *faults* as a cluster B variable (Figure 3) and its clear association with the cluster B production wells (Figure 5A, 5E, and 5D) indicates the macro-scale faults, those that were identified in the 3D geologic map based on geologic mapping, analysis of geophysical maps, seismic reflection interpretation, and downhole lithologic data, are an important control on hydrothermal circulation in the shallow (~300-600 m depth) reservoir. Interestingly, *faultnear*, *td*, *ts*, and *curve* do not strongly correlate with cluster B or the cluster B production wells. These classifications of faults (i.e., by proximity, by resolved stress, or by geometry) are actually more weakly correlated with cluster B than the simple ‘1’ for fault ‘0’ for no fault classification used for the *faults* variable.

Relatively high values of *faultdense*, *faultintdense*, *normal*, *dilation*, and *coulomb* all occur in the Brady step-over. *dilation*, *coulomb*, and *normal* are stress and strain effects calculated based on slip modeling on the Brady fault zone (Siler et al., 2018). Though they have varying spatial distribution, dilation (*dilation)*, coulomb shear stress increase (*coulomb)*, and normal stress decrease (unclamping) (*normal*) all are maximized in different parts of the Brady step-over. Furthermore, a high density of fault strands (*faultdense*) and a high fault intersection density (*faultintdense*) also occur in the step over. The importance of these variable in define cluster B suggests that the productivity of the cluster B production wells is controlled by their intersection with the down-plunge projection of the step-over at depth. Wells in other clusters are 1) too far east of the step-over and therefore only intersect it at very shallow levels (e.g., several Cluster A wells), 2) north (the majority of Cluster C wells) or south (several Cluster A, C, and D wells) of the step-over and do not intersect it, or 3) west of the step-over (some Cluster A and D wells), such that they do not reach the west-plunging step-over at less than 750 m depth.

Since *faults* and the variables that are controlled by the Brady step-over (*faultdense*, *faultintdense*, *normal*, *dilation*, and *coulomb)* effectively cluster the production wells from the rest of the dataset, known faults within the relatively dense network of faulting and fracturing that occur in down dip projection of the Brady step-over are inferred as the dominant controls on the shallow (~300-600 m depth) hydrothermal reservoir at Brady geothermal field.

## Stratigraphic and temperature variables

The stratigraphic variables (unitthick, contactnear, goodlith) and *modeltemp* appear to be relatively minor controls on the clustering of wells (Figure 5) as compared to the faulting related variables described above. Still, *contactnear* is the cluster C variable and cluster C contains four of the five injection wells. Cluster C is defined by high S3 values (Figure 4) so the cluster C wells, including the injection wells, are differentiated by being near to geologic contacts relative to the other wells. Matrix permeability along geologic contacts may be an important component in the injectivity of the injection wells. *unitthick* and *goodlith* are cluster A and D variables, respectively. These clusters are dominated by non-productive wells, indicating that neither thick geologic units (*unitthick*) nor a thick section of the Miocene mafic to intermediate volcanic rocks which the shallow (~300-600 m depth) production wells produce from (i.e., the *goodlith* variable) differentiate production or injection wells from non-productive wells. These two geologic factors are, therefore, not likely important controls on the permeability of the hydrothermal reservoir. The occurrence of the reservoir in the Miocene mafic to intermediate volcanic rocks is probably more a factor of the fault system (and fault step-over) that cuts through this stratigraphic sequence rather than characteristics of the lithologic unit itself. *modeltemp* also has relatively little control on the clustering of the wells. Figure 5 shows that *modeltemp* has very low values for all signals and is not associated with any particular cluster, or production or injection wells more so than non-productive wells. This is probably because the entire Brady area is relatively hot at shallow (<750 m) depths. The existing temperature data, and the 3D temperature model calculated from it, are apparently insufficient to distinguish areas of advective heat transport (by hydrothermal circulation) from areas of conductive transport of heat.

## Geologic controls on hydrothermal processes at Brady

In addition to the eight production wells in cluster B, there are six of the thirty-two non-productive wells (~19%) in cluster B. Though this is less than if the non-productive wells were evenly distributed between the four clusters, the fact that there are a significant number of non-productive wells in cluster B underscores the challenge of differentiating volumes of the subsurface that host hydrothermal circulation from volumes that do not. Still, the clustering of the production wells suggests geologic-based variables used in the analysis, particularly faults, *faultdense*, *faultintdense*, *normal*, *dilation*, and *coulomb,* can effectively distinguish productive areas from amongst the much greater volume of the subsurface that is non-productive.

# 6. Conclusions

Unsupervised ML analyses using a novel method based on non-negative matrix factorization coupled k-means clustering (NMFk) were conducted on a 3-D geological dataset from Brady geothermal field in order to elucidate the geologic characteristics that control hydrothermal circulation in the shallow (~300-600 m depth) reservoir. These analyses show that known faults, i.e. those that have been mapped in 3D based on geological and geophysical evidence are well correlated with production wells. Geologic factors that occur exclusively within the Brady step-over, such as high densities of faults and fault intersections, and strain and stress effects brought on by slip modeling, also correlate with production wells. These results suggest that the shallow hydrothermal reservoir at Brady is probably hosted within known, relatively prominent faults, that lie within a volume of the subsurface that is expected to have relatively high fault and fracture density, i.e. the step-over. These two factors, in concert and not either independently, that control the presence of the hydrothermal system. The NMF*k* methodology also successfully differentiates production wells and injection wells from amongst a larger number of non-productive wells using just these geologic data. This suggests that the NMF*k* method and these geologic parameters may be effective at training machine learning methods to diagnose areas that have similar characteristics to the production and injection wells within an unexplored volume of rock at the Brady site.

# Acknowledgements

# References

Alberti M., 2011, 3D point cloud density calculation: a C++ program, <https://gisoftw.blogspot.com/2011/05/3d-point-density-calculation-c-program.html>

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>

Ayling, B.F., 2020. 35 Years of Geothermal Power Generation in Nevada, U.S.A.: 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>

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>

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, U.S.A.): 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: G.S.A. 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 (U.S.A.): Geothermics, v. 54, p. 54–67, <https://doi.org/10.1016/j.geothermics.2014.10.003>

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,401:788–791, 10.

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>

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., 2013. Three-dimensional geothermal fairway mapping: Examples from the western Great Basin, U.S.A. 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, U.S.A. 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.

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, U.S.A.: Geothermal Energy, v. 4, no. 14, 21 p., <https://doi.org/10.1186/s40517-016-0056-6>