Siler *et al. Geotherm Energy* (2021) 9:17
https://doi.org/10.1186/s40517-021-00199-8

Geothermal Energy

RESEARCH Open Access

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. Siler[1*], Jeff D. Pepin[2], Velimir V. Vesselinov[3], Maruti K. Mudunuru[4] and Bulbul Ahmmed[3]

*Correspondence: dsiler@usgs.gov
[1] U.S. Geological Survey, Moffett Field, Mountain View, CA, USA
Full list of author information is available at the end of the article

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 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. Here, we analyze geologic data with ML methods to unravel the local geologic controls on these pathways. The ML method, non-negative matrix factorization with *k*-means clustering (NMF*k*), is applied to a library of 14 3D geologic characteristics hypothesized to control hydrothermal circulation in the Brady geothermal field. Our results indicate that macro-scale faults and a local step-over in the fault system preferentially occur along 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.

Introduction

Crustal permeability is a key parameter in hydrothermal 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 from a few, relatively small (sub-meter- to meter-long) intervals of a borehole that may be 100 s or 1000 s of meters in total length (based on Nevada Division of Minerals, publicly available data http://www.nbmg.unr.edu/Geothermal/ProductionInjection/index.html). This compartmentalization of hydrothermal processes means that

Springer Open

© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

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 or temperature). 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 the compartmentalization of fluid flow this hydrothermal system. 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 macro-scale faults and the ~km-scale step-over in the fault system are closely spatially associated with production wells relative to injection wells and non-productive wells. Mapping 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.

Background

The Brady geothermal system

Brady geothermal field is located ~80 km northwest of Reno, Nevada, USA (Fig. 1). Brady 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). Fluids are produced from two levels in the subsurface ~300–600 m depth and ~1750 m depth. Temperatures of produced fluid have been ~130–185 °C during this time (Benoit 2014, and based on Nevada Division of Minerals, publicly available data http://www.nbmg.unr.edu/Geothermal/ProductionInjection/index.html), though temperatures as high as 219 °C have been measured (Shevenell et al. 2012). These relatively high temperatures occur at relatively shallow levels (as shallow as 300–600 m depth for some production wells) as a result of convective upwelling driven by temperature-related differences in fluid density, and/or hydraulic head driven fluid-flow through hot rock (advection). 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 in the rock. At Brady, these have been primarily attributed to a network of fractures within a step-over in the Basin and Range-type normal fault (e.g., Wernicke 1992; Colgan et al. 2006) called the Brady fault (Faulds et al. 2003, 2010a, b, 2017; Siler et al. 2018; Siler and Pepin 2021; Fig. 1). The production well field; an elliptical, ~3 km wide × 6 km long (across strike × along strike, relative to the north-northeast fault strike) temperature anomaly; surface geothermal features include active fumaroles, silica sinter, silicified sediments, and warm ground (Kratt et al. 2006; Lechler and

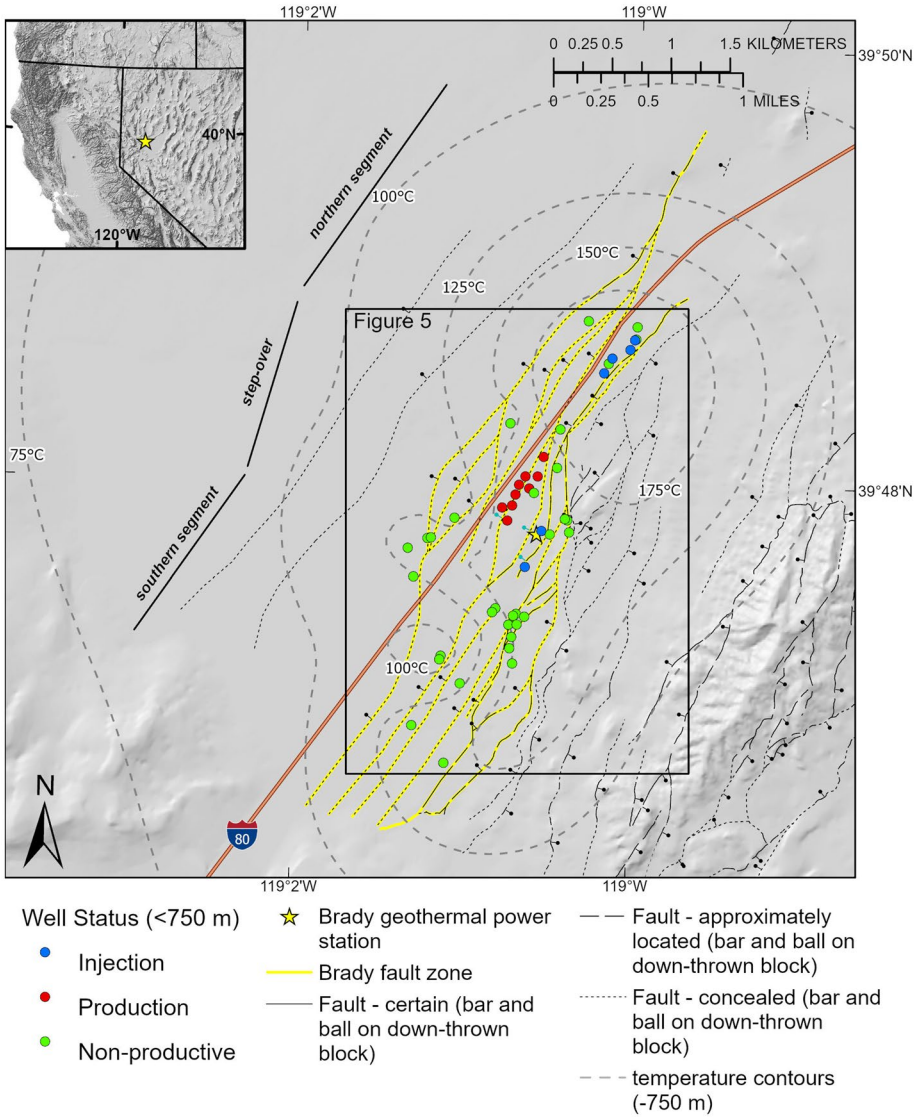

Fig. 1 Map of Brady geothermal area. The fault strands that constitute the Brady fault (data from Faulds et al. 2017) are shown in yellow. Contours represent modeled temperature, a linear interpolation of the temperature data at 750 m depth. These are the same data as the modeltemp variable. Wells are 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

Coolbaugh 2007; Faulds et al. 2010a, b, 2017); and diffuse degassing of anomalous CO_2, H_2S, and Rn are also centered on the Brady step-over (Jolie et al. 2015a, b, 2016).

The stratigraphic section at Brady consists of metamorphic basement rocks at ~1400 m and deeper in the well field, overlain by Oligocene-to-Late Miocene volcanic rocks from ~1400 to ~300 m depth, and Late Miocene-to-Holocene sedimentary rocks from ~300 m depth to the surface. The Brady fault is a west-dipping, north-northeast-striking system of normal faults that cuts this stratigraphic section (Fig. 1). The step-over of the Brady fault zone (Faulds et al. 2010a, b, 2017; Siler and Faulds 2013; Siler

et al. 2016) is an area where parallel but non-collinear strands 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 (Fig. 1; Faulds et al. 2017). Production wells are located in the step-over, injection wells are both in the step-over and ~ 1.5 km away, at the north-northeast extent of the thermal anomaly (Fig. 1).

Methods

An existing 3D geologic map of Brady, synthesizing geologic map, downhole lithologic and structural data from well cuttings and core, and geophysical data (Siler and Faulds 2013; Jolie et al. 2015b; Siler et al. 2016, 2021; Witter et al. 2016) was used to develop a suite of geologic variables that may control the distribution and localization of transmissive pathways and production-grade hydrothermal fluid flow. The 14 different variables described below are calculated from the 3D geologic map and projected to 47 production, injection, and non-productive wells within the field (Fig. 1).

Geothermal well data

The Brady well dataset is much denser at shallow levels than at deep levels (just eight of 47 wells extend to ~ 1750 m, the deeper of the two geothermal reservoirs). The well data at deeper levels are too sparse to be used in the NMF*k* analysis. As a result, this analysis focuses on the shallow (~ 300–600 m deep) reservoir, where the data are sufficiently dense for NMF*k*. 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 m 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 http://www.nbmg.unr.edu/Geothermal/ProductionInjection/index.html). We consider the remaining 32 wells to be non-productive wells. Four of these remaining 32 wells are used for production or injection; however, these wells produce or inject in the deeper reservoir. At depths less than 750 m they are not open to flow to or from the formation. These four deeper production wells are therefore considered "non-productive" for our purposes. The other 28 wells have never been used for production or injection so we assume they have sub-commercial temperatures and/or flow rates. Each of the 14 geologic factors described below is calculated at 1-m intervals along all 47 wells. The resultant database of 336,784 entries (24,056 locations with 14 variables) is used as input data for NMF*k* analysis. These are the input data for the NMF*k* analysis. Figure 2 shows the values for each of the 14 variables for one of the production wells.

Geothermal variables

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

For each of the 32 faults defined by the 3D geologic map (Siler and Faulds 2013; Siler et al. 2016, 2021; Witter et al. 2016), a 30-m-wide fault zone is generated. This zone approximates the effective width of secondary faulting and fracturing around each fault. Though not constrained by field data, this width is consistent with empirically derived fault zone widths for km-long faults, like the Brady fault zone (Scholz et al. 1993; Anders

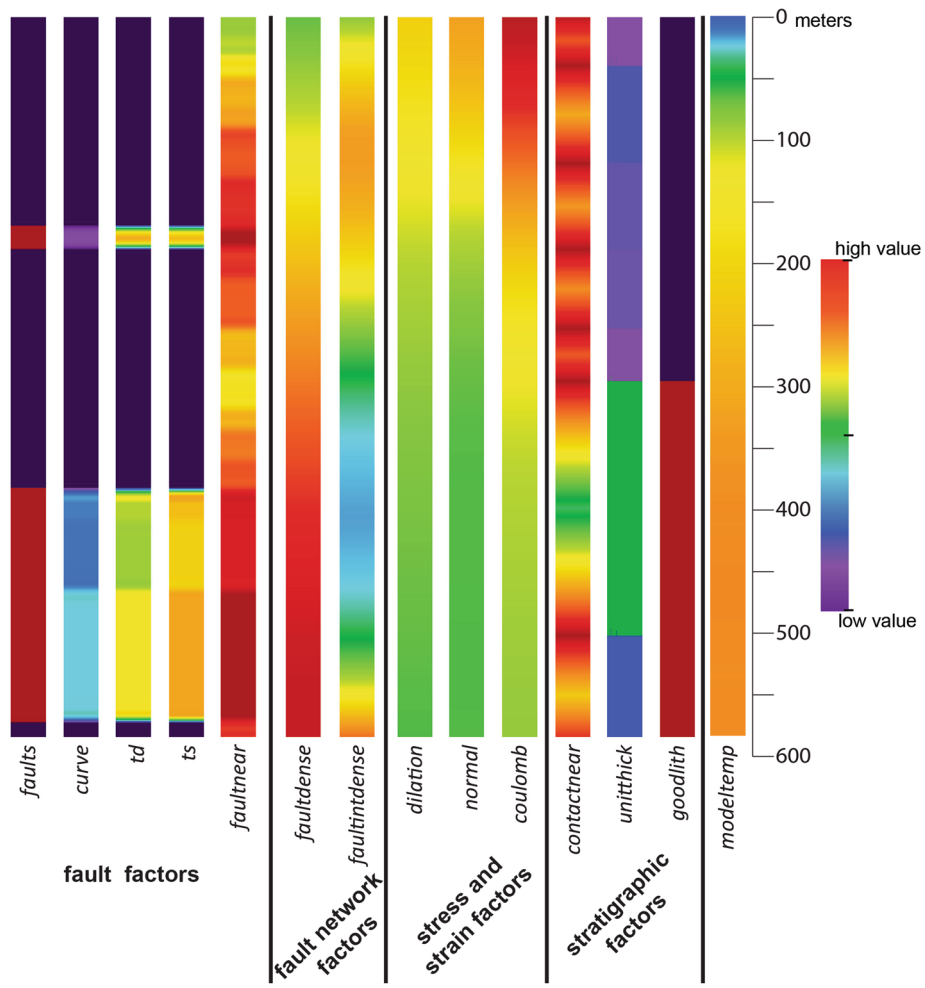

Fig. 2 The 14 variables used in this study along one of the Brady production wells. These are the input data for the NMFk algorithm. Cool colors correspond to low values and warm colors correspond to high values for each variable. For the binary variables (fault and goodlith), red corresponds to a value 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

and Wiltschko 1994). The *fault* variable has a value of '1' where a well is located within a 30-m-wide 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. 2015b). The 30-m-wide fault zone for each fault is populated with the calculated *curve*, *ts*, and *td* values. The *td*, *ts*, and *curve* values are proxies for the occurrence of conductive fractures in each fault zone. 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 ratio of normal stress to shear stress on faults, respectively. Fault segments that are either highly dilatant (high *td*) and/or stress loaded for slip (high *ts*) are likely to host fluid flow (Ferrill et al. 2019, 2020). 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 Fig. 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 as faults per unit volume and fault intersections or terminations per unit volume, respectively (Alberti 2011; 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, and either may host 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, 2016, 2017, 2018, 2021; Jolie et al. 2015a, b; Siler and Faulds 2013; Siler et al. 2016). 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 m normal slip on the Brady fault are calculated at 250-m-depth intervals from the surface to 750 m depth (Stein et al. 1992; King et al. 1994; Lin and Stein 2004). *dilation*, *normal* and *coulomb* values between the calculated depth slices are linearly interpolated, approximating the volumetric dilation, normal stress reduction, and Coulomb shear stress increase in the study area. The stress and strain perturbations that occur with fault slip result in zones of accentuated secondary faulting and fracturing and may be important factors in localizing hydrothermal circulation in the step-over (Siler et al. 2018).

Stratigraphic factors (contactnear, unitthick, goodlith)

In addition to the above geologic and 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 those that occur 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. Though the 3D geologic map (Siler et al. 2021) contains simplified stratigraphy relative to the more detailed geologic map and cross-section published for the Brady area (Queen et al. 2016; Faulds et al. 2012, 2017), the stratigraphic contacts from the 3D geologic map used to calculate *contactnear* and *unitthick*, represent major lithologic boundaries. It is probable that if there were a stratigraphic effect on the production zones, it would come from these boundaries between the major lithologic packages. 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 favorable for developing highly transmissive fracture systems when faulted relative to other lithologic units. The *goodlith* variable is '1' for well intervals with these stratigraphic units and '0' for intervals in other units.

Temperature (modeltemp)

Advection or convection is 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 39 deep (as deep as ~2 km) geothermal wells and 79 shallow (~150 m) temperature gradient wells (Shevenell et al. 2012) were used to build a 3D temperature model (Siler et al. 2021). The modeled temperature (*modeltemp*) is projected to each of the 47 wells.

Machine learning methods: NMF*k*

Machine learning (ML) approaches are mainly classified as unsupervised or supervised; the difference being whether the algorithm is formally trained by using labeled data (supervised) or not trained (unsupervised). Both supervised and unsupervised methods have been employed to address geothermal-related concepts. Faulds et al. (2020), applied supervised methods to regional-scale data to reveal hidden hydrothermal systems. Supervised methods have also been used to inform geothermal reservoir modeling (Gudmundsdottir and Horne 2020; Beckers et al. 2021), predict loss circulation zones in wells (Kiran and Salehi 2020), and identify small-scale fractures in seismic reflection data (Zheng et al. 2021). Unsupervised ML has been applied to identify hidden geothermal signals in both regional-scale data (Vesselinov et al. 2020a, b; Ahmmed et al. 2020a, b, d) and local-scale data (Vesselinov et al. 2019; Ahmmed et al. 2020c; Siler and Pepin 2021). Unsupervised learning was chosen for this study to infer the natural structure present within the dataset so that new information regarding the geologic controls on hydrothermal processes may readily be discovered.

The NMF*k* algorithm we employ combines two unsupervised machine learning (ML) methods, non-negative matrix factorization (NMF) and customized *k*-means clustering. NMF factorizes a non-negative data matrix, $\mathbf{X}$, into two components $\mathbf{W}$ and $\mathbf{H}$, where $\mathbf{W}$ is a location matrix and $\mathbf{H}$ is an attribute matrix.

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

$$\mathcal{L} = \left\| \mathbf{X} - \mathbf{W}\mathbf{H}^{\mathrm{T}} \right\|_F^2, \tag{1}$$

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

After completion of NMF process, 1000 estimated $\mathbf{H}$ are clustered into k clusters using k-means clustering that has been customized using blind source detection (Alexandrov and Vesselinov 2014). Because k is unknown in k-means clustering, the algorithm consecutively examines specified k by obtaining 1000 $\mathbf{H}$ for each feature/variable. During clustering, the similarity between two variables is assessed according to the cosine norm. After clustering, the Silhouette values (Rousseeuw 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 $+1$; positive 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 optimal number of clusters, or hidden signals. If k is low, the Silhouette value will be high, but so may be $\mathcal{L}$ because of under-fitting (e.g., Lever et al. 2016). For high k, the Silhouette value will be low and the solution may be over-fit (e.g., Lever et al. 2016). So, the best estimate for k is a number that optimizes both the $\mathcal{L}$ and Silhouette values.

Other existing matrix factorizations include singular value decomposition (SVD) (Klema and Laub 1980) and principal and independent component analyses (PCA and ICA) (Wold and Geladi 1987; Comon 1994; Ritchie and Pepin 2020; Siler and Pepin 2021). There are a few advantages of NMF*k* over these other tools. For instance, NMF*k* handles both real and categorical variables and datasets with missing entries, yet still provides interpretable results (Alexandrov and Vesselinov 2014; Vesselinov et al. 2018). For geological applications of ML like this one, the ability to handle categorical data (e.g., the *faults* and *goodlith* variables) and datasets with missing entries is critical. These parameters may not be constrainable with numerical values in all locations of interest, yet still constitute an important part of the solution. In our case, the categorical *faults* variable is a key part of the solution. Our dataset does not have any missing entries, so this advantage of NMF*k* was not utilized in this study.

Results

As outlined in "Machine learning methods: NMFk" section, NMF*k* analysis reveals hidden associations within a dataset. In our case, these associations characterize interdependencies among geologic attributes and production, injection, and non-productive well locations within the analyzed 3D domain. Results were calculated for k of values

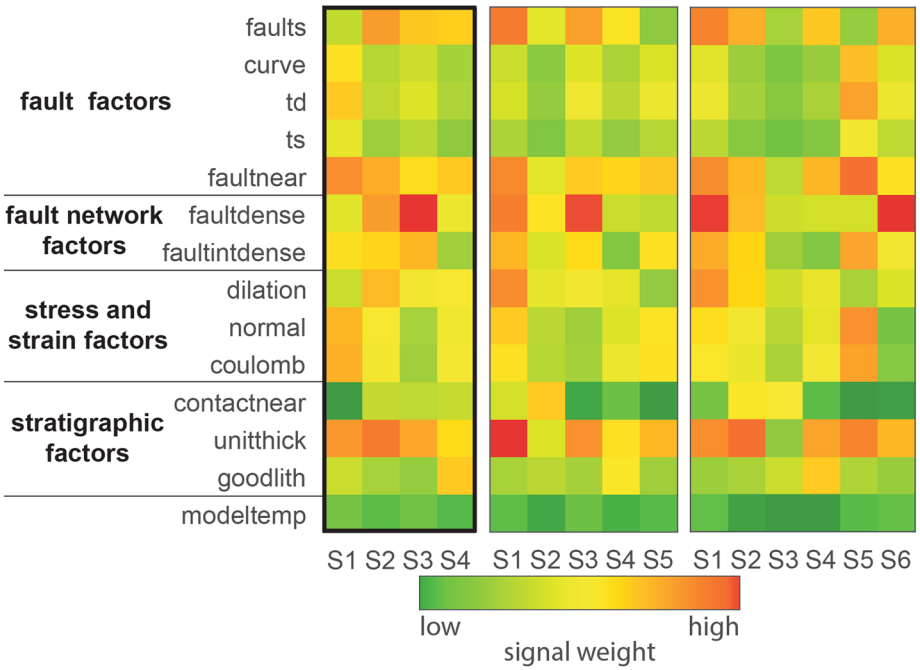

Fig. 3 *H* (attribute) matrix for 4-signals, 5-signals and 6-signals. The 4-signal solution, which is interpreted here, is highlighted. Warm colors indicate that the variable has a high weight with that particular signal; cool colors indicate that the variable has a low weight with that specific signal

of 2, 4, 5, and 6. The 2-cluster result was under-fit. The 3-cluster result is rejected by the 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 over-fit. For each of the 4, 5, and 6 cluster solutions the *H* matrices show similar weighting patterns between variables (Fig. 3). We chose to interpret the 4-cluster solution 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 and *W*-matrix columns (S1, S2, S3, and S4). The signal weight for each factor is calculated by the NMF*k* algorithm. In addition to defining the four signals, the NMF*k* results define "cluster" 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 is defined by the NMF*k* algorithm. Figure 4A, B is the same matrix, Fig. 4A is sorted by the cluster label, Fig. 4B is sorted by S2 value. For both Figs. 3 and 4, warm colors indicate a relatively high (strong) weight between the signal and the variable (Fig. 3) or the signal and the well (Fig. 4) and cool colors indicate a relatively low weight. Figure 5 shows the Brady well field and the cluster

(See figure on next page.)
Fig. 4 *W* (location) matrix. The four signals from the NMFk results relative to the 47 geothermal wells: **A** sorted by well clusters, **B** sorted by S2 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

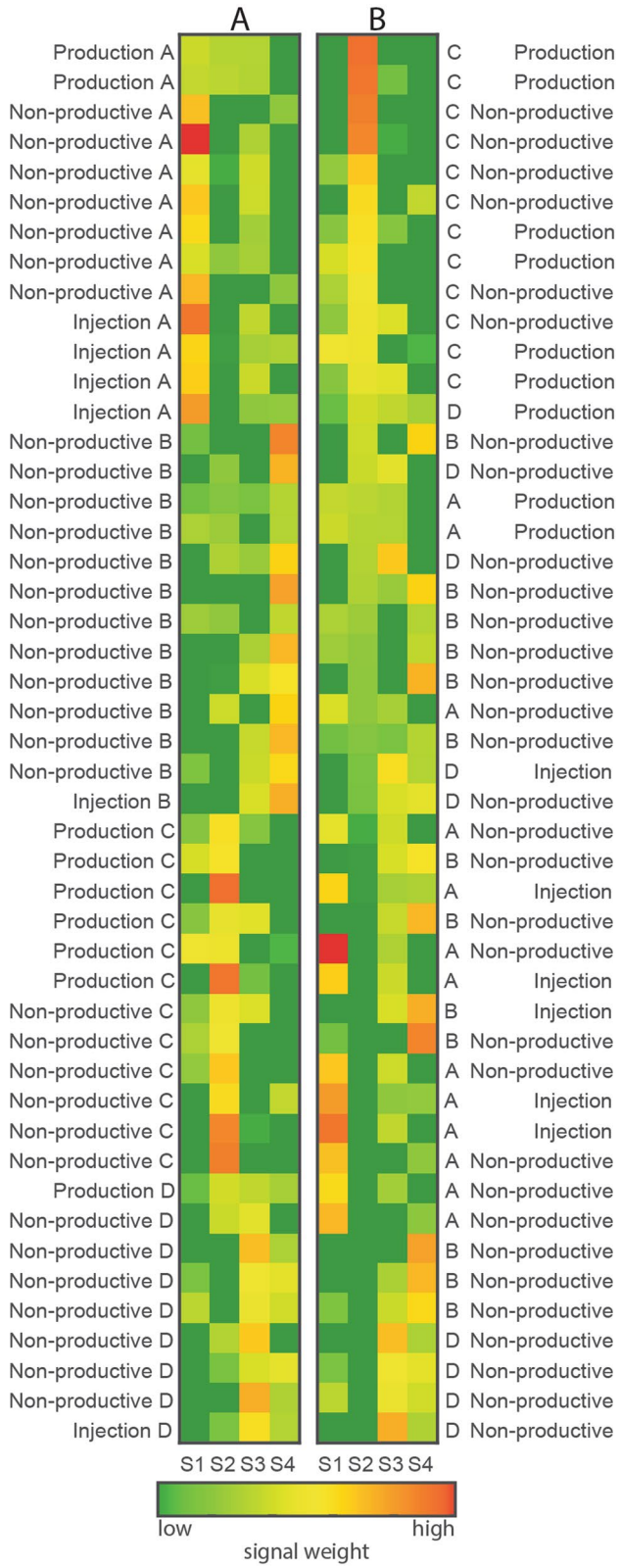

Fig. 4 (See legend on previous page.)

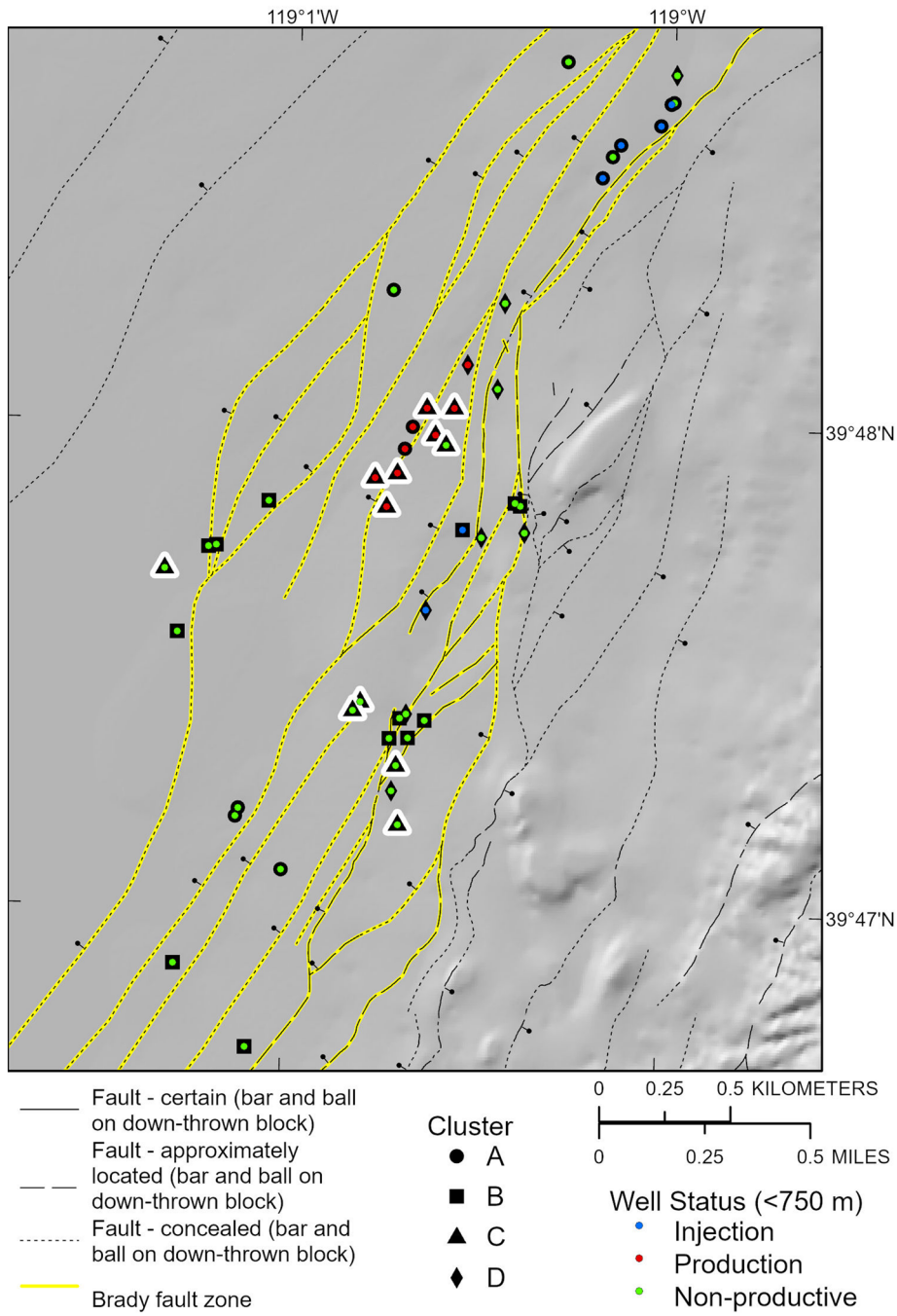

Fig. 5 Map of the Brady well field and fault system. Wells are shown by their use (production, injection, or non-productive) and their cluster (**A**, **B**, **C**, or **D**) as assigned by the NMFk results. Six of the nine production wells are in Cluster C (triangles). Cluster C is highlighted with a white halo

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 Fig. 6, the wells and variables are plotted by their S1 and S2 values from Figs. 3A, 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

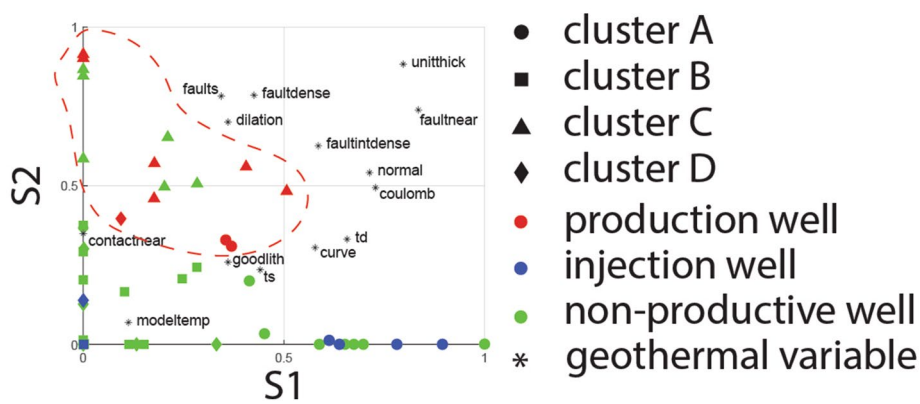

Fig. 6 Biplot of signal-1 (S1) vs signal-2 (S2). Production wells (red) are enclosed by the red dashed polygon. The production wells and the faults, dilation, and faultdense variables have relatively high S2 values and relatively low S1 values indicating that these variables control the separation of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

effectively separate the production wells from the injection wells and non-productive wells (Fig. 6).

Discussion

The NMF*k* 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 (Figs. 4A and 5). Cluster C is associated with relatively high (red) S2 values, and relatively low values (green) for S1, S3, and S4 (Fig. 4B). Additionally, all 9 of the production wells fall within the 17 highest S2 values (Fig. 4B). This further indicates that relatively high S2 values are strongly associated with production wells relative to the other wells. Additionally, the NMF*k* results indicate which fault factors, fault network factors, and stress and strain factors are more dominant in S2 (and therefore predominate along the production wells) relative to injection wells and non-productive wells.

Fault factors

The occurrence of faults intersecting a well, i.e., the *faults* variable, is the predominant faulting related factor associated with S2 (Figs. 3A and 6). This is evident in Fig. 6 on which *faults* plots in the upper left quadrant, with relatively high S2 and relatively low S1; a similar pattern is observed for the majority of the production wells. Though *faultnear* also has high S2 values, it also has relatively high values for S1, S3, and S4 (Fig. 3A), so it less distinctly related to S2 relative to *faults*. On Fig. 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 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 3D spatial density of fault planes (*faultdense*) is the predominant fault network factor related to S2 and the production wells (Figs. 3A and 6). This is evident in Fig. 6 on

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 *faultindense* has a high S2 value, it also has relatively high S1 and S3 values (Fig. 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 (Figs. 3A and 6). This is evident in Fig. 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 more strongly related to production wells relative to *normal* or *coulomb,* the other stress/strain network factors examined herein.

Stratigraphic factors

Overall, no stratigraphic factor is particularly strongly related to the production wells. Of the three stratigraphic factors, the thickness of geologic units (*unitthick*) has high S2 values (Figs. 3A and 6) relative to the nearness to geologic contacts (*contactnear*), and the specific geologic units that are associated with geothermal production (*goodlith*). However, S1 values for *unitthick* are high relative to the production wells (*unitthick* plots in the upper right in Fig. 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 other wells.

Geologic controls on hydrothermal processes at Brady

The NMF*k* 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 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 crossing a well (Fig. 2), is more closely related to the production wells than the static stress state of the faults (*td* or *ts*), the curvature of the faults (*curve*), or the nearness to the mapped fault planes (*faultnear*).

The step-over in the Brady fault (Fig. 1) is the other dominant geologic factor controlling hydrothermal processes at Brady. This concept is well established (Faulds et al. 2003, 2010a, b, 2016, 2017; Kratt et al. 2006; Lechler and Coolbaugh 2007; Jolie et al. 2015a, b; Siler et al. 2018), but this study and a similar unsupervised machine learning study (Siler and Pepin 2021) directly link the 3D geologic attributes of the step-over to the production wells and production intervals. The geometry and location of the step-over controls the spatial density of fault planes (*faultdense*) because faults are most dense in the step-over (Fig. 1). The step-over also controls the dilatational strain that occurs as a result of

modeled fault slip on the Brady fault (*dilation*) (Siler et al. 2018). The NMF*k* results suggest these two factors are more 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*). A principal component analysis (PCA) study conducted on a similar dataset but examining the entirety of the Brady production wells (rather than < 750 m depth as we do here) also indicate the macro-scale faults and 3D density of faults are the dominant controls on geothermal production zones (Siler and Pepin 2021). Interestingly, this study suggests that dilation is more strongly associated with geothermal production than coulomb stress change, whereas Siler and Pepin (2021) show the opposite. The magnitude of dilation occurring as a result normal fault slip is highest at shallow levels and decreases with depth. This may explain why this study, which focuses on the shallow reservoir (< 750 m), reveals dilation as a primary control, whereas Siler and Pepin (2021), which includes analysis to ~ 1750 m depth suggests that changes in coulomb shear stress are a primary control.

Of the stratigraphic factors, thicker geologic units (*unitthick*) the strongest 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 tertiary 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 or convective relative to conductive heat transport, and thus modeled temperature is relatively ineffective for resolving discrete fluid-flow pathways. In future work, a different strategy for considering temperature data is needed.

Conclusions

Non-negative matrix factorization with *k*-means clustering (NMF*k*) analyses was conducted on a 3D geological dataset from Brady geothermal field 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 strongly associated with production wells relative to other wells. These results indicate that the shallow hydrothermal reservoir at Brady is hosted by relatively prominent faults and that 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 to dilate as a result of periodic fault slip, are especially well suited for geothermal production. These two factors, macro-scale faults and the Brady step-over, together, and not either independently, control the presence of the Brady hydrothermal system that has been developed for electricity production and direct thermal uses.

The NMF*k* methodology successfully differentiates production wells from among a larger number of non-productive wells using these geologic data. This suggests that these geologic parameters may be assessed at other geothermal sites in order to help evaluate site prospectively and/or reservoir processes. In sites with limited subsurface

data, the 3D geologic factors may be less well constrained, but analysis similar to what has been presented herein may still indicate areas of the subsurface that are likely to have the appropriate geologic characteristics for fluid-flow. These types of analysis may also be effective as training data for using NMF*k* or supervised machine learning techniques to identify areas within an unexplored volume of the subsurface that have the geologic characteristics that are expected to host productive geothermal wells.

Acknowledgements
We thank Lowell Price at the Nevada Division of Minerals for his help obtaining some of the data used in this study. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We thank Jacob DeAngelo for his very helpful review of the paper and we thank three anonymous reviewers for their reviews.

Authors' contributions
DLS conceived the research, prepared the data, prepared the figures, and wrote the paper. VVV conducted the NMF*k* analysis. DLS, JDP, VVV, MKM, and BA analyzed the results and edited the paper. All authors read and approved the final manuscript.

Funding
Funding for this research was provided 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.

Availability of data and materials
The NMF*k* code and documentation is freely available at https://github.com/SmartTensors/GTcloud.jl and https://github.com/SmartTensors. The input data are available in the Geothermal Energy data repository.

Declarations

Competing interests
The authors have no competing interests.

Author details
[1]U.S. Geological Survey, Moffett Field, Mountain View, CA, USA. [2]U.S. Geological Survey, Albuquerque, NM, USA. [3]Computational Earth Science Group, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [4]Watershed & Ecosystem Science, Pacific Northwest National Laboratory, Richland, WA 99352, USA.

Received: 1 February 2021 Accepted: 4 June 2021
Published online: 30 June 2021

References

Ahmmed B, Lautze N, Vesselinov V, Dores D, Mudunuru M. Unsupervised machine learning to extract dominant geothermal attributes in Hawaii Island Play Fairway data. Geotherm Resour Council Trans. 2020a;44:1282.
Ahmmed B, Vesselinov V, Mudunuru M. Machine learning to characterize regional geothermal reservoirs in the western USA. In: Fall conference, Geological Society of America, Abstract T185-358249, October 26–29. 2020b.
Ahmmed B, Vesselinov V, Mudunuru M. Non-negative matrix factorization to discover dominant attributes in Utah FORGE Data. Geotherm Resour Council Trans. 2020c;44:1281.
Ahmmed B, Vesselinov V, Mudunuru M. Integration of data, numerical inversion, and unsupervised machine learning to identify hidden geothermal resources in southwest New Mexico. In: Fall conference, American Geophysical Union, Abstract H166-0022, December 1–17. 2020d.
Alberti M. 3D point cloud density calculation: a C++ program. 2011. https://gisoftw.blogspot.com/2011/05/3d-point-density-calculation-c-program.html.
Alexandrov BS, Vesselinov VV. Blind source separation for groundwater pressure analysis based on nonnegative matrix factorization. Water Resour Res. 2014;50(9):7332–47. https://doi.org/10.1002/2013WR015037.
Anders MH, Wiltschko DV. Microfracturing, paleostress and the growth of faults. J Struct Geol. 1994;16:795–815. https://doi.org/10.1016/0191-8141(94)90146-5.
Ayling BF. 35 years of geothermal power generation in Nevada, USA: a review of field development. In: Proceedings, forty-fifth workshop on geothermal reservoir engineering. Stanford University. 2020;45:12.
Beckers KF, Duplyakin D, Martin MJ, Johnston HE, Siler DL. Subsurface characterization and machine learning predictions at brady hot springs. In: Proceedings, forty-sixth workshop on geothermal reservoir engineering, Stanford University; 2021.
Benoit WR. The long-term performance of Nevada geothermal fields utilizing flash plant technology. Geotherm Resour Council Trans. 2014;38:977–84.
Benoit WR, Butler RW. A review of high-temperature geothermal developments in the Northern Basin and Range Province. Geotherm Resour Council Spec Rep. 1983;13:57–80.

Blackwell DD. Heat flow in the northern Basin and Range province. Geotherm Resour Council Spec Rep. 1983;13:81–93.
Caine JS, Forster CB. Fault zone architecture and fluid flow: insights from field data and numerical modeling. Geophys Monogr Am Geophys Union. 1999;113:101–28. https://doi.org/10.1029/GM113p0101.
Caine JS, Evans JP, Forster CB. Fault zone architecture and permeability structure. Geology. 1996;24:1025–8. https://doi.org/10.1130/0091-7613(1996)024%3C1025:FZAAPS%3E2.3.CO;2.
Colgan JP, Dumitru TA, Reiners PW, Wooden JL, Miller EL. Cenozoic tectonic evolution of the Basin and Range Province in Northwestern Nevada. Am J Sci. 2006;306:616–54. https://doi.org/10.2475/08.2006.02.
Comon P. Independent component analysis, A new concept? Signal Process. 1994;36(3):287–314. https://doi.org/10.1016/0165-1684(94)90029-9.
Fairley JP, Hinds JJ. Rapid transport pathways for geothermal fluids in an active Great Basin fault zone. Geology. 2004;32:825–8. https://doi.org/10.1130/G20617.1.
Fairley JP, Heffner J, Hinds J. Geostatistical evaluation of permeability in an active fault zone. Geophys Res Lett. 2003. https://doi.org/10.1029/2003GL018064.
Faulds JE, Garside LJ, Oppliger G. 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 Resour Council Trans. 2003;27:6.
Faulds JE, Coolbaugh MF, Benoit WR, Oppliger GL, Perkins M, Moeck I, Drakos PS. 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). Geotherm Resour Council Trans. 2010a;34:675–84.
Faulds JE, Moeck I, Drakos PS, Zemach E. Structural assessment and 3D geologic modeling of the Brady's geothermal area, Churchill County (Nevada, USA): a preliminary report. In: Proceedings, thirty-fifth workshop on geothermal reservoir engineering. Stanford University. 2010b. p. 298–302.
Faulds JE, Ramelli AR, Garside LJ, Coolbaugh MF, Green HL. Preliminary geologic map of the desert Peak Quadrangle, Churchill County, Nevada: Nevada Bureau of Mines and Geology, Open-File Report 12-5, scale 1:24,000; 2012.
Faulds JE, Ramelli AR, Coolbaugh MF, Hinz NH, Garside LJ, Queen JH. Preliminary geologic map of the Brady's geothermal area, Churchill County, Nevada: Nevada Bureau of Mines and Geology, Open-File Report 17-4, scale 1:12,000. 2017.
Faulds JE, Brown S, Coolbaugh M, DeAngelo J, Queen JH, Treitel S, Fehler M, Mlawsky E, Glen JM, Lindsey C, Burns E. Preliminary report on applications of machine learning techniques to the Nevada geothermal play fairway analysis. In: Proceedings, forty-fifth workshop on geothermal reservoir engineering. Stanford University. 2020. p. 229–34.
Ferrill DA, Winterle J, Wittmeyer G, Sims D, Colton S, Armstrong A, Horowitz AS, Meyers WB, Simons FF. Stressed rock strains groundwater at Yucca Mountain, Nevada. GSA Today. 1999;9:1–8.
Ferrill DA, Smart KJ, Morris AP. Fault failure modes, deformation mechanisms, dilation tendency, slip tendency, and conduits v. seals. Geol Soc Lond Spec Publ. 2019;496(1):75–98. https://doi.org/10.1144/SP496-2019-7.
Ferrill DA, Smart KJ, Morris AP. Resolved stress analysis, failure mode, and fault-controlled fluid conduits. Solid Earth. 2020;11(3):899–908. https://doi.org/10.5194/se-11-899-2020.
Fossen H, Rotevatn A. Fault linkage and relay structures in extensional settings—a review. Earth-Sci Rev. 2016;154:14–28. https://doi.org/10.1016/j.earscirev.2015.11.014.
Gudmundsdottir H, Horne RN. Prediction modeling for geothermal reservoirs using deep learning. In: Proceedings, forty-fifth workshop on geothermal reservoir engineering. Stanford University. 2020. p. 229–34.
Jolie E, Klinkmueller M, Moeck I. Diffuse surface emanations as indicator of structural permeability in fault-controlled geothermal systems. J Volcanol Geotherm Res. 2015a;290:97–113. https://doi.org/10.1016/j.jvolgeores.2014.11.003.
Jolie E, Moeck I, Faulds JE. Quantitative structural-geological exploration of fault-controlled geothermal systems—a case study from the Basin-and-Range Province, Nevada (USA). Geothermics. 2015b;54:54–67. https://doi.org/10.1016/j.geothermics.2014.10.003.
Jolie E, Klinkmueller M, Moeck I, Bruhn D. Linking gas fluxes at Earth's surface with fracture zones in an active geothermal field. Geology. 2016;44:187–90. https://doi.org/10.1130/G37412.1.
King GCP, Stein RS, Lin J. Static stress changes and the triggering of earthquakes. Bull Seismol Soc Am. 1994;84:935–53.
Kiran R, Salehi S. Assessing the relation between petrophysical and operational parameters in geothermal wells: a machine learning approach. In: Proceedings, forty-fifth workshop on geothermal reservoir engineering, Stanford University. 2020. p. 10.
Klema VC, Laub AJ. The singular value decomposition: Its application and some applications. IEEE Trans Autom Control. 1980;25(2):164–76. https://doi.org/10.1109/TAC.1980.1102314.
Kratt C, Calvin W, Coolbaugh MF. Geothermal exploration with Hymap hyperspectral data at Brady-Desert Peak, Nevada. Remote Sens Environ. 2006;104:313–24. https://doi.org/10.1016/j.rse.2006.05.005.
Lachenbruch AH, Sass JH. Heat flow in the United States and the thermal regime of the crust. In: Heacock JG, editor. The Earth's Crust, 20. American Geophysical Union Monograph. Hoboken: Wiley; 1977. p. 626–75. https://doi.org/10.1029/GM020p0626.
Lechler PJ, Coolbaugh MF. Gaseous emissions from steamboat Springs, Brady's Hot Springs, and Desert Peak Geothermal Systems, Nevada. Geotherm Resour Council Trans. 2007;31:359–61.
Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. Nature. 1999;401(6755):788–91. https://doi.org/10.1038/44565.
Lever J, Krzywinski M, Altman N. Model selection and overfitting. Nat Methods. 2016;13:703–4. https://doi.org/10.1038/nmeth.3968.
Lin J, Stein RS. Stress triggering in thrust and subduction earthquakes and stress interaction between the southern San Andreas and nearby thrust and strikeslip faults. J Geophys Res. 2004;109:B02303. https://doi.org/10.1029/2003JB002607.
Morris A, Ferrill DA, Henderson DB. Slip-tendency analysis and fault reactivation. Geology. 1996;24:275–8. https://doi.org/10.1130/0091-7613(1996)024%3c0275:STAAFR%3e2.3.CO;2.
Peacock DCP, Sanderson DJ. Displacements, segment linkage and relay ramps in normal fault zones. J Struct Geol. 1991;13:721–33. https://doi.org/10.1016/0191-8141(91)90033-F.
Peacock DCP, Sanderson DJ. Geometry and development of relay ramps in normal fault systems. Am Assoc Pet Geol Bull. 1994;78:147–65.

Queen JH, Daley TM, Majer EL, Nihei KT, Siler DL, Faulds JE. Surface reflection seismic and vertical seismic profile at Brady's Hot Springs, NV, USA. In: Proceedings, forty-first workshop on geothermal reservoir engineering, Stanford University; 2016.
Ritchie AB, Pepin JD. Optimization assessment of a groundwater-level observation network in the Middle Rio Grande Basin, New Mexico (ver. 2, December 2020). U.S. Geological Survey Scientific Investigations Report 2020–5007. 2020. 113 p. https://doi.org/10.3133/sir20205007.
Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
Sanderson DJ, Zhang X. Stress-controlled localization of deformation and fluid flow in fractured rocks. Geol Soc Lond Spec Publ. 2004;231:299–314. https://doi.org/10.1144/GSL.SP.2004.231.01.18.
Scholz CH, Dawers NH, Yu J, Anders MH, Cowie PA. Fault growth and fault scaling laws—preliminary results. J Geophys Res. 1993;98:951–61. https://doi.org/10.1029/93JB01008.
Shevenell LA, Oppliger G, Coolbaugh MF, Faulds JE. Brady's (Nevada) InSAR anomaly evaluated with historical well temperature and pressure data. Geotherm Resour Council Trans. 2012;36:1383–90.
Sibson RH. Crustal stress, faulting and fluid flow. In: Parnell J, editor. Geofluids: origin, migration and evolution of fluids in sedimentary basins. Geological Society, London, Special Publications. 1994. p. 69–84. https://doi.org/10.1144/GSL.SP.1994.078.01.07.
Siler DL, Faulds JE. Three-dimensional geothermal fairway mapping: examples from the western Great Basin, USA. Geotherm Resour Council Trans. 2013;37:327–32.
Siler DL, Pepin JD. 3-D geologic controls of hydrothermal fluid flow at Brady geothermal field, Nevada, USA. Geothermics. 2021;94:102112. https://doi.org/10.1016/j.geothermics.2021.102112.
Siler DL, Hinz NH, Faulds JE, Queen J. 3D analysis of geothermal fluid flow favorability: Brady's, Nevada, USA. In: The forty-first workshop on geothermal reservoir engineering, Stanford University. 2016. p. 10.
Siler DL, Hinz NH, Faulds JE. Stress concentrations at structural discontinuities in active fault zones in the western United States: implications for permeability and fluid flow in geothermal fields. Geol Soc Am Bull. 2018. https://doi.org/10.1130/B31729.1.
Siler DL, Faulds JE, Hinz NH, Queen JH. Three-dimensional geologic map of the Brady geothermal area, Nevada. U.S. Geological Survey Scientific investigations Map 3469, 2 sheets, pamphlet 32 p. 2021. https://doi.org/10.3133/sim3469.
Stein RS, King GC, Lin J. Change in failure stress on the southern San Andreas fault system caused by the 1992 magnitude = 7.4 Landers earthquake. Science. 1992;258:1328–32. https://doi.org/10.1126/science.258.5086.1328.
Vesselinov VV, Alexandrov BS, O'Malley D. Contaminant source identification using semi-supervised machine learning. J Contam Hydrol. 2018;212:134–42. https://doi.org/10.1016/j.jconhyd.2017.11.002.
Vesselinov VV, Mudunuru MK, Karra S, O'Malley D, Alexandrov BS. Unsupervised machine learning based on non-negative tensor factorization for analyzing reactive-mixing. J Comput Phys. 2019;395:85–104. https://doi.org/10.1016/j.jcp.2019.05.039.
Vesselinov V, Ahmmed B, Mudunuru MK. Unsupervised machine learning to discover attributes that characterize low, moderate, and high-temperature geothermal resources. Geotherm Resour Council Trans. 2020a;44:1363.
Vesselinov VV, Mudunuru MK, Ahmmed B, Karra S, Middleton RS. Discovering signatures of hidden geothermal resources based on unsupervised learning. In: Proceedings, forty-fifth workshop on geothermal reservoir engineering, Stanford University; 2020b.
Wernicke B. Cenozoic extensional tectonics of the U.S. Cordillera. In: Burchfiel BC, Lipman PW, Zoback ML, editors. The Cordilleran orogen: conterminous U.S.: the geology of North America. Boulder: Geologic Society of America; 1992. p. 553–81.
Witter JB, Siler DL, Faulds JE, Hinz NH. 3D geophysical inversion modeling of gravity data to test the 3D geologic model of the Brady's geothermal area, Nevada, USA. Geotherm Energy. 2016;4(14):21. https://doi.org/10.1186/s40517-016-0056-6.
Wold E, Geladi S. Principal component analysis. Chemometr Intell Lab Syst. 1987;2(1–3):37–52. https://doi.org/10.1016/0169-7439(87)80084-9.
Zheng Y, Li J, Lin R, Hu H, Gao K, Huang L. Physics-guided machine learning approach to characterizing small-scale fractures in geothermal fields. In: Proceedings, forty-sixth workshop on geothermal reservoir engineering, Stanford University; 2021.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen journal and benefit from:

- ► Convenient online submission
- ► Rigorous peer review
- ► Open access: articles freely available online
- ► High visibility within the field
- ► Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com