Geothermics 72 (2018) 358–370

Contents lists available at ScienceDirect

Geothermics

journal homepage: www.elsevier.com/locate/geothermics

Cluster analysis as a tool for evaluating the exploration potential of Known Geothermal Resource Areas

Cary R. Lindsey[a], Ghanashym Neupane[b], Nicolas Spycher[c], Jerry P. Fairley[a,*], Patrick Dobson[c], Thomas Wood[a], Travis McLing[b], Mark Conrad[c]

[a] Department of Geological Sciences, University of Idaho, 875 Perimeter Drive, MS 3022, Moscow, ID 83844-3022, United States
[b] Idaho National Laboratory, Science Center Drive, Idaho Falls, ID 83402, United States
[c] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, M/S 74R316C, Berkeley, CA 94720, United States

ARTICLE INFO

Keywords:
KGRAs
Geothermal exploration
Idaho/Oregon
PCA
Cluster analysis

ABSTRACT

Although many Known Geothermal Resource Areas in Oregon and Idaho were identified during the 1970s and 1980s, few were subsequently developed commercially. Because of advances in power plant design and energy conversion efficiency since the 1980s, some previously identified KGRAs may now be economically viable prospects. Unfortunately, available characterization data vary widely in accuracy, precision, and granularity, making assessments problematic. Here we suggest a procedure for comparing test areas against proven resources using Principal Component Analysis and cluster identification. The result is a low-cost tool for evaluating potential exploration targets using uncertain or incomplete data.

1. Introduction

In the early 1970s, amidst a national energy crisis, the US Energy Research and Development Administration (which later became the United States Department of Energy, or USDOE) partnered with the United States Geological Survey (USGS) to identify and inventory the geothermal resources of the United States. As defined by the Geothermal Steam Act of 1970, a Known Geothermal Resource Area (KGRA) is an area where "…the prospects for extraction of geothermal steam or associated geothermal resources from an area are good enough to warrant expenditures of money for that purpose" (Godwin et al., 1971). The USDOE/USGS program of geothermal exploration identified a number of KGRAs, many of which are located in southern Idaho and eastern Oregon. Unfortunately, as the energy crisis eased during the 1980s, so did federal funding for geothermal exploration, and many of the identified KGRAs did not receive the follow-on studies that would have been required to evaluate their economic potential.

In the 40+ years since the Geothermal Steam Act of 1970, innovations in power plant design have increased the overall conversion efficiency of geothermal power developments. Although the average conversion efficiency of geothermal power plants is still the lowest of all thermal plants (Zarrouk and Moon (2014) cite an average conversion efficiency of 12%, on the basis of a worldwide review of published data), technological improvements such as double flash, triple flash, hybrid geopressure/geothermal, and binary plant designs have allowed an expansion of installed geothermal capacity to a worldwide total in 2015 of about 12,635 MWe (Bertani, 2015). In particular, binary plants, first introduced in the early 1980s, and the optimization of working fluids (e.g., ammonia, HCFC123, n-Pentane, PF5050) for a wide range of evaporation and condensation temperatures, enthalpy fluxes, and coolant velocities, have improved the performance of power plants and decreased the required resource temperatures, allowing economic development of resources that had previously not been considered viable (Hettiarachchi et al., 2007). Changes in legislation have also led to increased opportunities for geothermal development. The Energy Policy Act of 2005 amended the Geothermal Steam Act of 1970, modifying how royalties are calculated, how land is leased, and providing tax incentives and loan guarantees for certain types of energy resources in an effort to make geothermal (and other renewable resources) more competitive with fossil fuel electrical power generation.

As a result of the changing technological and economic landscape, KGRAs that were previously identified as not economically exploitable may now be commercially viable. Unfortunately, efforts to reevaluate data collected during earlier phases of exploration have been hampered by heterogeneous quality and granularity, as well as by site-to-site variations in observed parameters. To address these challenges, researchers at Lawrence Berkeley National Laboratory and Idaho National Laboratory, in collaboration with scientists at the University of Idaho, have been working to develop an approach to making between-area comparisons that can be used with incomplete and/or uncertain data.

* Corresponding author.
E-mail address: jfairley@uidaho.edu (J.P. Fairley).

https://doi.org/10.1016/j.geothermics.2017.12.009
Received 25 July 2017; Received in revised form 31 October 2017; Accepted 18 December 2017
Available online 03 January 2018

Here, we present one possible approach to such between-site comparisons. Our method applies Principal Component Analysis (PCA), Hierarchal Cluster Analysis (HCA), and *K*-Means Cluster Analysis (KMCA) to compare existing data from a group of candidate areas to data from multiple (high- and low-geothermal potential) control groups. The final result of the analysis is a dendrogram of related sites (see supplemental material S1 and S2) that can be used to help prioritize future exploration and characterization efforts.

2. Study area description

The KGRAs evaluated in this study are all located in either eastern Oregon or southern Idaho in the northwest region of the United States. With one exception,[1] the areas fall into one of two geological provinces: the Basin and Range province, or the Snake River Plain. We first present general background information on the geology and geothermal setting of these two provinces, followed by a brief description of the KGRAs included in our investigation.

2.1. The Basin and Range

The Basin and Range province is an extensional terrain comprising a large number of horst and graben structures distributed across the western United States (Fig. 1). Thinning of the crust due to east–west extension allows for variable, but generally high, heat flow (from about 60 mW/m^2 to > 100 mW/m^2; Blackwell, 1983). Northwest–southeast oriented bulk regional extension in this area is generally manifested along northeast striking faults (Pezzopane and Weldon, 1993; Blewitt et al., 2003), and high-temperature geothermal systems are preferentially located along northeast-striking lineaments (Koenig and McNitt, 1983; Coolbaugh et al., 2003), or are associated with accommodation zones and other structurally favorable settings (Faulds et al., 2011, 2013). Geothermal activity in the region is generally assumed to derive from topographically-driven deep circulation of groundwater, although magmatic heat sources are likely responsible for a subset of areas (Koenig and McNitt, 1983). The high heat flow of the region, coupled with active extensional to transtensional faulting, creates a favorable environment for geothermal development. Apart from the large number of known and potential conventional geothermal resource areas in the US Basin and Range province, it is also believed that the region presents opportunities for “unconventional” (Engineered/Enhanced Geothermal Systems, or EGS) resources. A review of US geothermal potential cited the Great Basin (a subset of the US Basin and Range) as first in a list of high-grade EGS resources (Tester et al., 2006).

2.2. The Snake River Plain

The Snake River Plain (SRP) is a large igneous province that stretches some 640 km across southern Idaho, from the Idaho–Oregon border to the northwest corner of Wyoming (Fig. 1). The region is a shallow physiographic depression that cross-cuts pre-existing Basin and Range topography (Pierce and Morgan, 1992; Rodgers et al., 2002; Smith et al., 2009). The western part of the plain is a large tectonic graben filled with thick (1000s of meters) lacustrine deposits that are underlain by rhyolitic ignimbrites and basalt flows. In contrast, the Eastern Snake River Plain (ESRP) was formed by a string of large calderas associated with the migration of the North American Plate over the Yellowstone hot spot during the past 17 My (Pierce and Morgan, 1992). Up to 2 km of Holocene to early Pliocene basalts underlie the plain, which were erupted from shield volcanoes and NW-striking volcanic rifts. Beneath, and largely obscured by the basalts, are extensive rhyolitic ignimbrites and lava flows that are known from

Fig. 1. Map of the western United States, showing the general area of the northern Basin and Range Province (darker shaded region) and the Snake River Plain (lighter shaded region) in southern Idaho. Note that some authors define the Basin and Range to include much of Idaho and, in some cases, as far north as eastern Washington.

boreholes and exposures along the margins of the plain (Morgan et al., 2008; Podgorney et al., 2013). The Snake River Plain represents one of the highest heat flow provinces in North America (Blackwell, 1989; Blackwell and Richards, 2004), and was listed second (behind the Great Basin) in a recent survey of high-grade EGS prospects in the US (Tester et al., 2006). Near-surface heat flow is suppressed by groundwater in the high permeability eastern Snake River Plain aquifer (McLing et al., 2016), but thermal gradients are high along the margins of the plain, and heat flow below the SRP aquifer is believed to be high as a result of the intrusion of mafic magmas in a mid-crustal sill complex (Blackwell, 1989; Shervais et al., 2006; Nielson et al., 2017).

2.3. KGRAs

We examined 14 KGRAs or IHRAs (Identified Hydrothermal Resource Areas; Burkhardt et al., 1980) in southeast Oregon and southern Idaho. Three of these areas currently host geothermal power plants (Raft River, Neal Hot Springs, and Summer Lake/Paisley Hot Springs) for a combined electrical output of about 37 MWe, and these areas were used as a high-potential control group (i.e., high geothermal potential). In addition to the KGRAs/IHRAs, we also included groundwater samples from one area with no known geothermal potential to serve as a low-potential control group, for a total of 15 areas included in the analysis (Table 1). The low-potential control group samples were taken from shallow wells producing Ca-HCO_3-type waters from the Eastern Snake River Plain that differ markedly from the deeper, Na-HCO_3-type waters of thermal origin (McLing et al., 2002). Brief descriptions of the 14 thermal areas included in this study are given in the following paragraphs, and their approximate locations are shown in Fig. 2.

Alvord Basin Geothermal Area. The Alvord Basin is a north-northeast trending structural graben located in Harney County, southeast Oregon. The area is a KGRA comprising three groups of hot springs, with

[1] Vulcan Hot Springs KGRA is located in the Atlanta lobe of the Idaho batholith, somewhat off the margin of the western Snake River Plain near Cascade, Idaho.

Table 1
Names and sizes of KGRAs included in this study. Areas listed are taken from Burkhardt et al. (1980). The Camas, Neal Hot Springs, and Preston areas have not been officially designated as KGRAs, but contain several Identified Hydrothermal Resource Areas (IHRAs). The area listed for the Mountain Home site is for the officially designated KGRA; however, for this study we extended the boundaries of the area to the west as far as the Mountain Home Air Force Base. *N* is the number of samples used for the indicated location; of the 22 samples from the Raft River KGRA, 11 are associated with geothermal power production at the Raft River plant, and are included in the high-potential control group. High-potential controls (i.e., areas with producing geothermal fields) are listed in **bold** type. Latitude and Longitude (WGS84) are approximate picks from Google Earth.

Name	Area (km^2)	*N*	Latitude	Longitude
Alvord Basin	715.6	32	42.5997	−118.4872
Bruneau	10.5	60	42.7931	−115.7249
Camas	–	37	43.3447	−114.8817
Castle Creek	322.6	26	43.0114	−116.2611
Crane Creek	17.6	3	44.3067	−116.7447
Crump Geyser	346.7	2	42.2265	−119.8812
Lakeview	49.2	7	42.2219	−120.3682
Mountain Home	38.5	10	43.1014	−115.4469
Neal Hot Springs	–	4	44.0232	−117.4703
Preston	–	6	42.1332	−111.9277
Raft River	122.3	22(11)	42.1021	−113.3843
Summer Lake/Paisley	39.5	4	42.7111	−120.6457
Vale Hot Springs	93.1	8	43.9827	−117.2331
Vulcan Hot Springs	15.5	1	44.5675	−115.6952
Control Group (low)	–	6	43.5780	−112.9667

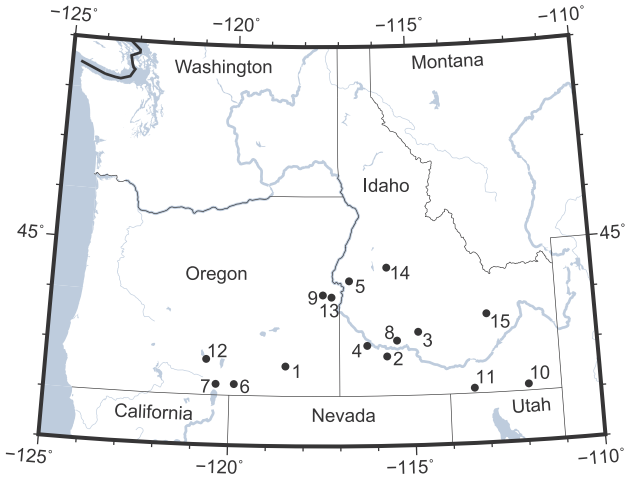

Fig. 2. Map of the northwest United States, showing the approximate locations of areas included in the analysis presented in this study. The numbered locations are: (1) Alvord Basin; (2) Bruneau, (3) Camas Prairie, (4) Castle Creek, (5) Crane Creek, (6) Crump Geyser, (7) Lakeview, (8) Mountain Home, (9) Neal Hot Springs, (10) Preston, (11) Raft River, (12) Summer Lake/Paisley, (13) Vale Hot Springs, (14) Vulcan Hot Springs, and (15) low-potential control group.

discharge temperatures ranging up to 98.5 °C (Anderson and Fairley, 2008). Phillips Geothermal conducted an extensive exploration program in the area during the 1970s, including geochemical surveys, geothermal gradient data collection, AMT and MT surveys, and geological mapping. An exploration well drilled by Anadarko Petroleum to a depth of 451 m in 1989 flowed at a rate of 400 gpm and a temperature of 152 °C (Cummings et al., 1993). Cation geothermometry indicates equilibration temperatures around 200 °C (Cummings et al., 1993; Koski and Wood, 2004), and studies indicate hydrothermal discharge in the basin is fault controlled (Fairley and Hinds, 2004; Anderson and Fairley, 2008; Hess et al., 2009), with strong lateral variations in fault permeability tending to channelize subsurface flow into localized "fast-flow" pathways (Fairley and Hinds, 2004; Fairley, 2009). Environmental restrictions on development in the Alvord Basin have discouraged further commercial exploration in the basin, but the area provides an important source of data for testing models of geothermal systems.

Bruneau Geothermal Area. The geothermal potential of the Bruneau region in southwestern Idaho was first recognized by Piper (1924), and the area was classified as a KGRA in the early 1970s (Young and Mitchell, 1973). Subsequent field studies focused on sampling hot springs and wells, with identified temperatures around the Bruneau and Grand View areas in the range of 22–54 °C (Young et al., 1975; Rightmire et al., 1976). Although early assessments indicated significant geothermal potential, with predicted reservoir temperatures up to 190 °C, the prospect failed to attract investors, and there have been no additional field studies in this area since the 1970s.

Camas Prairie Geothermal Area. Camas Prairie is an east–west elongated (about 50 km by 15 km) valley in Blaine, Camas, and Elmore Counties in south-central Idaho. The prairie is bounded by an east–west trending horst known as the Mount Bennett Hills to the south, and the Soldier Mountains to the north, which are a faulted mountain block associated with the Cretaceous-Miocene magmatic/volcanic activity of central Idaho. Although this area has not been designated a KGRA, the occurrence of several clusters of hot springs (IHRAs) within and along the margins of the Camas Prairie suggest the presence of geothermal resources (Mitchell, 1976). Specifically, locations within the larger region that have been identified as demonstrating geothermal potential include: Magic Hot Springs and Elk Creek Hot Springs, located on the eastern and northeastern margins of the prairie, Sheep and Wolf Hot Springs, located in the western part of the Camas Prairie, and having discharge temperatures of ~50 °C, Wardrop Hot Springs (60 °C) in the north-central part of the Prairie near the base of the Soldier Mountains, and Barron Hot Springs (73 °C) on the southern side of the prairie near the base of the Mount Bennett Hills. Recent work in this area has identified potential exploration targets in the southern portion of the prairie (Glen et al., 2017; Neupane et al., 2017; Shervais et al., 2017).

Castle Creek Geothermal Area. The Castle Creek geothermal area is a designated KGRA in Owyhee County, southwestern Idaho (Burkhardt et al., 1980). The area is located on the southern margins of the western Snake River Plain between the towns of Grand View and Oreana, Idaho, west of the Mountain Home geothermal area and northwest of the Bruneau geothermal area. The area is likely situated within the rhyolitic basement rocks of the Owyhee Mountains, on the downthrown blocks of a mapped series of NE-dipping, SE/NW-striking faults (Jenks et al., 1998). Several hot springs and artesian wells are known in the area, with water temperatures reported in the range of 23–83 °C (Young et al., 1975; Moore et al., 1979).

Crane Creek Geothermal Area. The Crane Creek geothermal area is located about 18 km northeast of Weiser, in Washington County, Idaho. The area lies along a fault zone that defines the northern margins of the western Snake River Plain, and includes two groups of hot springs: Crane Creek Hot Springs, with a discharge temperature of 92 °C, and Cove Creek Hot Springs, which discharge at 74 °C. The Crane Creek area displays a negative gravity anomaly, which contrasts with the positive gravity anomaly of the region west of Weiser, and the general western Snake River Plain. The high-density basaltic rocks present to the southwest are thought to thin or truncate near Crane Creek as a result of southwest-dipping fault structures, and the juxtaposition of higher- and lower-density rocks across the faults may give rise to the contrasting gravity measurements (Young and Whitehead, 1975). Reconnaissance audio-magnetotelluric and telluric current soundings of Crane Creek and nearby Vale KGRAs indicate low resistivity zones are associated with both areas (Long and Kaufmann, 1980).

Crump Geyser Geothermal Area. The Crump Geyser geothermal area is located in the Warner Valley of Lake County, southeast Oregon. In spite of its reputation as the largest continuously erupting geyser in the US, Crump Geyser was actually the site of an artesian well that discharged water of unknown (but apparently elevated) temperature, drilled in 1959 by the Nevada Thermal Power Company on land owned by Charles Crump. The area was the site of a series of geothermal investigations, including gravity, seismic, and aeromagnetic surveys, in

addition to hosting a number of exploration and thermal gradient wells that found maximum temperatures in the subsurface on the order of 120 °C (Fairbank and Smith, 2015). Studies have outlined a fault-controlled reservoir at the site (e.g., Casteel, 2010; Glen et al., 2015).

Lakeview Geothermal Area. The town of Lakeview, Oregon, is the site of an apparently low-temperature resource that has been developed for direct use, such as building heating and similar applications. A number of local establishments make direct use of the hydrothermal resources, including a commercial greenhouse, a tilapia farm, and the Oregon Department of Corrections Warner Creek facility, as well as local schools and a hospital (Sifford, 2014).

Mountain Home Geothermal Area. The Mountain Home KGRA is located on the western Snake River Plain in southwest Idaho, and includes the area surrounding the town of Mountain Home and the Mountain Home Air Force Base, as well as an expanse of area to the east. Findings from the Bostic 1-A well, drilled during the late 1970s, indicated thepresence of a high temperature (~200 °C) but low permeability resource at a depth of about 3 km (Arney and Goff, 1982; Arney, 1982; Arney et al., 1984). Later drilling of the HOTSPOT MH-2B well identified a permeable zone at a depth of approximately 1.8 km, with a temperature of about 150 °C (Shervais et al., 2013). Multiple hot springs are located along the foot of the Mount Bennett Hills, several kilometers to the north of the main area.

Neal Hot Springs Geothermal Area. Neal Hot Springs is located near Vale, Oregon, along faults that are likely to represent the northern extremity of Basin and Range extension in this area. The geothermal reservoir is thought to be located within a major step-over in a northerly-striking normal fault zone and bounded to the east and west by the west-dipping Neal and Sugarloaf Butte faults, respectively (Edwards, 2013). The presence of 90 °C springs drew attention to the region as a potential site for geothermal development, and initial exploratory drilling in the area by Chevron Minerals was conducted in 1979. Bottom hole temperatures in several production wells range from 135–145 °C, and US Geothermal currently operates a net 22 MWe power plant at the site.

Preston Geothermal Area. The Preston area is located in the northward trending Cache Valley of southeastern Idaho. It is situated at the confluence of several geologic terrains at the northeastern extent of the Basin and Range province, where the Basin and Range meets the Sevier orogenic belt and the Rocky Mountains. The juncture of these provinces is characterized by seismic activity and clusters of hot springs (Sbar et al., 1972). Although this area has not been officially designated a KGRA, it comprises several IHRAs (i.e., Battle Creek Hot Springs and Squaw Hot Springs). The area attracted attention in the 1970s as a result of work by Mitchell (1976), but interest in the area waned until 2014, when a 79 m deep well encountered a bottom hole temperature of 104 °C (Wood et al., 2015). Multicomponent geothermometry suggests a reservoir temperature in the range of 170–180 °C (Neupane et al., 2016).

Raft River Geothermal Area. The Raft River KGRA is located in southern Idaho, near the town of Malta. The highest measured bottom-hole temperature at the site is 149 °C (Dolene et al., 1981). Extensive exploration and development conducted by the USGS and the US Department of Energy in the mid- to late-1970s and early 1980s proved the viability of the site for commercial power generation (Dolene et al., 1981; Ayling and Moore, 2013); currently, US Geothermal operates a 13 MWe geothermal power plant at the site. The hydrothermal reservoir at Raft River is fracture-controlled; hydrothermal fluids circulate through fractures in the Precambrian basement rocks and rise along northwest-striking normal faults (Dolene et al., 1981), discharging into the Tertiary lacustrine sediments of the Salt Lake formation. The Bridge and Horse Wells fault zones are the two main fault zones in the area, and these are presumed to intersect a basement shear zone called the Narrows Structure to control the upflow of hydrothermal fluids at the site (Ayling and Moore, 2013; Bradford et al., 2015; Li et al., 2017).

Summer Lake/Paisley Geothermal Area. The Summer Lake KGRA is a relatively low-temperature resource located in the northwest portion of the Basin and Range Province near Paisley, Oregon. The resource is believed to be associated with the intersection of two *en echelon* normal faults to the west of Summer Lake; pathways created by this fault intersection likely control reservoir permeability and play an important role in localizing the resource (Mink et al., 2014). Reservoir temperature estimates based on geothermometry range between 95 and 166 °C, although a well drilled by Surprise Valley Electric produced hydrothermal fluids at a temperature of 111 °C (Makowsky, 2013). In spite of the apparently modest temperature of the resource, a 3.1 MWe plant currently operates at the site.

Vale Geothermal Area. The Vale geothermal area in southeast Oregon has a long history of direct use geothermal energy, and is characterized by the presence of hot springs and wells, with discharge temperatures around 73 °C. The area lies near the western margin of the western Snake River Plain, and borders the Basin and Range and Columbia Plateau geographical regions. Wisian et al. (1996) reported a maximum subsurface temperature in the Vale area of 143 °C. Geothermal resources in the area are likely controlled by a swarm of southeast-northwest trending faults known collectively as the Vale fault zone (Fern et al., 1993), and locally thermal anomalies coincide with gravity highs (Wisian et al., 1996) and low resistivity (Long and Kaufmann, 1980).

Vulcan Hot Springs Geothermal Area. The Vulcan Hot Springs geothermal area is located in Valley County, Idaho, about 30 km northeast of the town of Cascade. The area lies entirely within the Cretaceous Idaho Batholith granite/granodiorite, and is known by the presence of a group of hot springs that discharge water at temperatures up to 87 °C. Hot spring activity in the area appears to be structurally controlled, potentially by an east-dipping, north-south oriented fault, identified by Christopherson et al. (1980) as passing through the site. Although the surface expressions of any controlling faults are obscured at the site by surficial deposits, west- and northwest-dipping faults likely intersect the proposed east-dipping fault to the south and north of the site, possibly placing additional structural controls on the reservoir (Christopherson et al., 1980; Lewis et al., 2012).

3. Methods

3.1. Data compilation

We compiled data from a variety of published sources, including USGS reports, state water resource management agencies, conference papers, and journal articles. We chose data primarily gathered from wells and hydrothermal springs, and the number of observations we were able to obtain varied between areas. In total across all areas, 228 data points (observations) were used in the final analysis, although some of these were non-thermal (from the low-potential control group). The initial dataset was compiled with 26 “continuous” variables (i.e., “ratio” or “interval” variables, which are real number-valued variables either possessing or lacking a true zero, respectively), including coordinates, depth, temperature, pH, conductance, TDS, species concentrations, and selected concentration ratios. In addition, we attempted to include non-numerical information that may be helpful in distinguishing commercially-viable reservoirs by encoding such data as “categorical” variables; that is, variables that usually cannot be described by real numbers, but can be “binned,” or placed into categories. Categorical variables are either “ordinal” or “nominal” in type, depending on whether the bins can be ranked, or lack an intrinsic order, respectively. As discussed below, it was ultimately determined that the reporting of data that could be described by categorical variables was too inconsistent and uneven between sites for inclusion in the present study; however, we feel that categorical data are likely to be extremely valuable for future studies of the commercial viability of geothermal areas. In the following sections, we therefore discuss both the continuous and categorical variables that we considered for this study, and

Table 2
Complete list of 26 continuous variables compiled for this study.

pH	Temperature	TDS	Conductivity
Depth	T_{dev}	Latitude	Longitude
Fracturing	Na	K	Ca
Mg	SiO_2	Cl	HCO_3
SO_4	F	Li	B
Na/K	Na/Ca	K/Mg	Na/Li
$Cl/(HCO_3 + CO_3)$	$SO_4/(HCO_3 + CO_3)$		

review the evaluation process by which we arrived at the final set of variables used in the analysis. A compilation of the data used in this study, along with citations to the original source material and estimated reservoir temperatures from a number of different geothermometers on a per-sample basis, are provided in the supplemental materials S3.

3.1.1. Continuous variables

The initial set of continuous (ratio or interval) variables consisted of 26 parameters such as species concentrations, ratios of the concentration of two species (e.g., $Cl/(HCO_3 + CO_3)$, Na/K), depth of the resource, total dissolved solids (TDS), pH, temperature, etc. A complete list of all continuous variables compiled for the present study is provided in Table 2. Of the variables listed in Table 2, T_{dev} is the difference, in °C, between the measured temperature for a sample and the expected temperature at the depth from which the sample was taken, according to:

$$T_{\text{dev}} = T(z) - (T_{\text{sur}} + \Delta \text{Tz}), \quad (1)$$

where $T(z)$ is the measured temperature at the depth z where the sample was taken (equal to zero for springs or other surface samples), T_{sur} is the assumed average surface temperature (taken as 10 °C), and ΔT is an assumed average linear geothermal gradient (taken as 25 °C/1000 m). The parameter "fracturing" is a qualitative indication of the ability of a reservoir to hold and transmit hydrothermal fluids, and ranged from 0 (no fracturing) to 100 (ubiquitous fracturing). Because we primarily used "legacy" data (i.e., existing data from published literature), our analysis depended upon information collected and analyzed by multiple investigators at disparate times, often using different analytical methods or collection protocols. As a result, the parameters that were reported varied from site to site, as did the detection limits for chemical analyses. We removed from consideration parameters that were missing or below the reported detection limits for more than 25% of the samples. This eliminated a number of variables from the analysis, including several species and/or ratios that are commonly used in the investigation of geothermal systems, such as Li, B, and Na/Li.

3.1.2. Categorical variables

While reviewing published characterizations of hydrothermal sites, we considered a large amount of information for possible inclusion in the analysis in the form of categorical variables. Some of the categorical data commonly used to describe sites have no bearing on development potential; for example, the KGRA identifier (e.g., "the Alvord Basin" or "Raft River") is useful as a reference to an area, but is generally devoid of scientific content (i.e., it does not add discriminatory power to the analysis). Other items clearly contain information bearing on the viability of a site, but lack a consistent definition, are overly inclusive, or are subjective in nature, such that different observers can reasonably give contrasting descriptions of the same phenomenon. An example of this type of information is the "permeability" of a KGRA. Although the term "permeability" has an agreed-upon definition (i.e., the negative of the ratio of flux to gradient), the statement that a given site is "permeable" or "has high permeability" is of dubious value. In the study presented here, most of the sites included in the analysis can be said to demonstrate "high permeability," in the sense that there must perforce be channels of sufficient permeability to conduct discharge of elevated temperature to the land surface. More specifically, however, the existence of a permeability "window," above or below which hydrothermal discharge is unlikely to manifest at the surface, is well known (see, for example, Forster and Smith, 1989). Simulations of fault-controlled hydrothermal systems indicate that elevated surface discharge temperatures are associated with country rock permeability less than $\sim 10^{-15}\,m^2$ and effective fault permeabilities in the range of 10^{-14}–$10^{-11}\,m^2$; furthermore, peak discharge temperatures are particularly sensitive to fault permeability (López and Smith, 1995, 1996). Thus, the description of an area as "high permeability" is not in itself diagnostic, because a geothermal reservoir requires particular ranges and spatial distributions of permeability in order to be commercially viable.

Of all the categorical information examined in this study, probably the most relevant were the data describing structural setting with respect to faulting. On the basis of a worldwide survey of hydrothermal discharge areas, Curewitz and Karson (1997) determined that approximately 78% of all hot springs are located near mapped faults. It is generally accepted that fault-induced permeability enhancement and zones of fault dilation (tension) are closely related (Zhang et al., 2008), and that dilation zones are often associated with focused, high permeability flow paths (Curewitz and Karson, 1997; Connolly and Cosgrove, 1999; Zhang et al., 2008). Dilation and associated high permeability pathways can occur in a variety of fault settings, including jogs, step-overs, wing/tail cracks, and fault intersections, and conceptual models for the relationship between fault structural setting and hydrothermal discharge have been described by a number of investigators (e.g., Curewitz and Karson, 1997; Connolly and Cosgrove, 1999; Hickman et al., 2004; Micklethwaite and Cox, 2004; Zhang et al., 2008; Faulds et al., 2011). In the present study, we attempted to encode structural setting information for the sampled locations as a nominal (unranked) categorical variable using the classification system of Faulds et al. (2011). The Faulds et al. (2011) system is based on earlier classifications of the structural settings of hot springs (Curewitz and Karson, 1997; Micklethwaite and Cox, 2004), but offers a refined classification for the Basin and Range-type regions of western Turkey and the western Great Basin of the USA for which it was developed. The five settings identified by the authors as favorable for geothermal exploration, and used in the present study as structural setting categories, are: (A) discrete steps in normal fault zones; (B) intersections between normal faults and transversely oriented oblique-slip faults; (C) overlapping, oppositely-dipping normal fault zones; (D) terminations of major normal faults; and (E) transtensional pull-apart zones (Faulds et al., 2011).

Although we did attempt to include information on structural setting in the present analysis, we were ultimately unsuccessful, for reasons that are discussed in Section 4. As a result of our inability to assemble a sufficiently comprehensive dataset of structural setting information on a per-sample basis, we were forced to exclude this important information from the analysis described below.

3.2. Analysis

In this study, we used a three-step approach to arrive at the final clustering of KGRAs. After removing parameters for which there were too many (i.e., > 25%) missing values, we used the results of Principal Component Analysis (PCA) to identify which of the remaining variables were likely to offer the most independent information. In the second step, the variables identified by PCA were input to a Hierarchal Cluster Analysis, with the objective of identifying an optimal number of clusters. Finally, we applied *K*-means clustering to the reduced dataset, using the results from hierarchal clustering to constrain the choice of *K*. We describe each of these steps in the following sections.

3.2.1. PCA

We began the analysis by using PCA as a screening test to determine the type of variables and the number of principal components (i.e.,

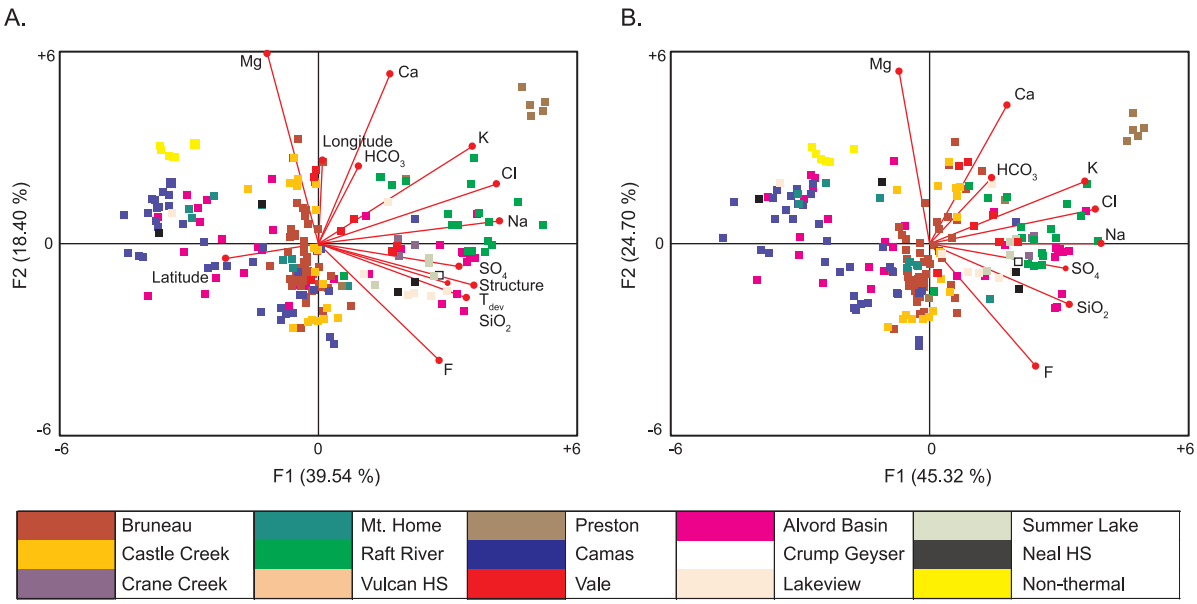

Fig. 3. Non-compositional distance biplots of first two principal components. (A) Full dataset, including both chemical and non-chemical variables. (B) Reduced dataset, excluding non-chemical data. Plotted points represent observation scores; the closer together two points are, the closer their common denominators. Rays (plotted in red) represent the loadings for each variable, with a smaller angle between rays indicating greater correlation, and opposing rays corresponding to inverse correlation.

common denominators) necessary to capture the essence of variability in the data. PCA is a method for defining a set of linearly uncorrelated variables (the *principal components*) from a set of variables, some or all of which are potentially correlated. The number of principal components, which must be less than or equal to the number of the original variables, form an orthogonal basis set. PCA is often useful in exploratory data analysis for screening large numbers of variables, some of which may be correlated (i.e., some variables may not be independent) to identify a subset of factors that best represents the behavior of the group. PCA has been widely applied in the social sciences, as well as in the geological sciences. Aitchison and Greenacre (2002) specifically cite its use in analyzing whole-rock oxide compositions. In hydrology and hydrogeology it is commonly applied to aqueous-phase geochemical variables (e.g., Otero et al., 2005; Fowler et al., 2017), as in this evaluation.

In the present study, we treated the data used in the PCA analysis as either compositional and non-compositional variables (Aitchison, 1986; Aitchison and Greenacre, 2002). Compositional data are components of a vector of non-negative values that sum to a constant and fixed value, and form a closed set (Aitchison, 1986). For example, the mass fractions of all the components of a system are compositional data, because the sum of the components is 1.0 (the vectors sum to a fixed and constant value), and the removal of some or all of one or more of the components does not change the sum (the set is closed). A set of compositional data form a finite-dimensional vector space, that provides a consistent definition of distance, norm, and scalar product, and satisfies the geometric requirements of a real space (Otero et al., 2005). However, compositional data are not truly independent, because a reduction in the fraction of one component results in a rescaling of all the other components (Fowler et al., 2017). On the other hand, non-compositional data form an open set, where the values of the individual components can vary independently. An aqueous chemistry dataset is always closed, and therefore compositional; however, a dataset consisting of measurements of depth to resource, or reservoir temperatures inferred from geothermometry, is open, and thus is non-compositional. Compositional and non-compositional data require different statistical treatments, and it has been known since the late 1800s that conventional multivariate statistical analyses cannot be naively applied to compositional datasets (Pearson, 1897).

Initial non-compositional PCA analyses were performed using the open set of continuous variables for our cataloged water analyses, including temperature, pH, TDS, Na, K, Ca, Mg, SiO_2, Cl, HCO_3, SO_4, F, Na/K, Na/Ca, and K/Mg. We also tested the impact of using log-transformed variables to represent species concentrations. When performing the initial analysis, we sought to maximize the percentage of variance captured by the first two common denominators (factors), while simultaneously maximizing the separation on biplots between samples from thermal waters and those from the non-thermal control group. A biplot is a 2D graphical representation of a singular value decomposition (Eckart and Young, 1936) that is commonly used to display the results of PCA (Gabriel, 1971). In biplots, data are represented as points, while variables are shown as rays. A ray is defined from the center of the plot to the ray's vertex (end-point), and the length of the ray is proportional to the fraction of the total variance explained by that variable (Otero et al., 2005). The cosine of the angle between any two rays is approximately equal to the correlation coefficient between the two variables; therefore, two rays that are separated by a small angle represent variables that are likely to be highly correlated, while two rays that are orthogonal to each other represent variables that are likely to be independent (or nearly so) (Otero et al., 2005).

On the basis of the initial PCA results, it was determined that the log-concentrations of nine variables (Na, K, Ca, Mg, SiO_2, Cl, HCO_3, SO_4, and F) could capture the variability between samples at least as well as, and possibly more clearly than, the full slate of variables. The analysis also revealed that the first two principal components accounted for about 70% of the variability in the data, while 97% of the variability was captured by the first six principal components.

After identifying the set of nine geochemical parameters that best captured the variability between samples, we tested several additional factors to see if their inclusion would improve the discriminatory power of the test. In particular, we examined the results of non-compositional PCA testing on the nine-variable geochemical dataset, augmented by three non-geochemical variables: (1) T_{dev}, or the difference between measured water temperature and the expected temperature at the measurement depth based on a conductive geothermal gradient for each location of interest; (2) the fracture "abundance," a qualitative representation of fracture density at a given site (varying between 0 and 100); and (3) the site longitude and latitude. We performed the PCA with all twelve variables (the nine geochemical and three non-geochemical variables), and on subsets in which each of the three non-

geochemical variables was successively dropped from the dataset. One interesting result from these tests was the identification of a strong correlation between SiO_2 concentration and difference in temperature from the conductive gradient (Fig. 3A); however, no combination of the non-geochemical variables and geochemical variables showed any essential change from the results obtained with the geochemical data alone, and dropping the three non-geochemical variables significantly improved the percentage of variance captured by the first two PCA factors (Fig. 3B). On the basis of these findings, we did not pursue the use of the non-geochemical variables further.

The non-compositional PCA analysis yielded a first component dominated by Cl, Na, K, SiO_2, and SO_4, and a second component dominated by Mg, Ca, and F. The corresponding biplots (Fig. 3) show a clear distinction between the control group of non-thermal waters and the rest of the areas included in the analysis. This distinction is particularly strong for the geothermal waters of the Raft River and Preston sites, one of which (Raft River) arguably represents the "most thermal" exemplar. These thermal waters align more strongly with a Na–K–Cl signature, in contrast to the non-thermal waters that tend to align with a Mg component. This is consistent with the observation that Mg concentrations are often elevated in groundwaters, relative to concentrations found in geothermal fluids (Giggenbach, 1988). From inspection of the biplots in Fig. 3, it may be concluded that waters from Alvord, Vale, and Summer Lake show the strongest relationship to the Preston and Raft River areas, even though these areas are located in widely separate geographic regions.

The non-compositional analyses, such as those presented in Fig. 3, were extremely useful for identifying potential signatures within the data, but may give spurious correlations when applied to compositional data (Pearson, 1897). For this reason, we also applied compositional PCA, using the set of nine concentration variables, as a check on the non-compositional results. We used a "centered-log-ratio" (clr) approach, following the method of Aitchison and Greenacre (2002), in which the concentration of each variable (Na, K, etc.) is normalized by dividing by the geometric mean of concentrations across variables x_1, x_2, …, x_D (Aitchison, 1986; Aitchison and Greenacre, 2002; Otero et al., 2005):

$$\mathrm{clr}_i(\mathbf{x}) = \log \frac{x_i}{\sqrt[D]{x_1 x_2 \ldots x_D}}, \quad i = 1, 2, \ldots, D. \tag{2}$$

This approach avoids the problem of spurious correlation that can arise when applying standard multivariate statistical methods of analysis to compositional data (Pearson, 1897), and is thought to reveal more clearly common denominators of compositional data. The biplot obtained from the compositional PCA (Fig. 4) shows three distinct groups: (1) a thermal "Raft River-like" group with a Na–Cl signature; (2) a thermal "Castle Creek-like" group with a strong F signature; and (3) a cold "groundwater-like" group, characterized primarily by the presence of Mg. As with the non-compositional PCA results, waters from Vale, Alvord, and Summer Lake appear to show the strongest common denominators with the Raft River thermal group. Many samples from Bruneau and Camas align with the Castle Creek group, showing a strong F signature that is possibly indicative of water-rock interaction with rhyolitic rocks at depth (Whelan, 2016). With the exception of samples from the Preston group, some samples from all of the KGRAs are classified with the groundwater (non-thermal) group. Samples from Preston appear to form their own group, with a character that deviates somewhat from the Raft River group.

Although the compositional PCA offered insights into relationships between areas examined in this study, the results did not differ in any substantial way from those of the non-compositional analysis. The two methods gave consistent results in terms of the variables that included the least amount of redundant information and provided the greatest discrimination power for analyzing between-area variations. From these outcomes, we chose to pursue the cluster analysis (Section 3.2.2) using the nine geochemical variables identified by the PCA analyses as offering the most independent information about the areas included in this investigation.

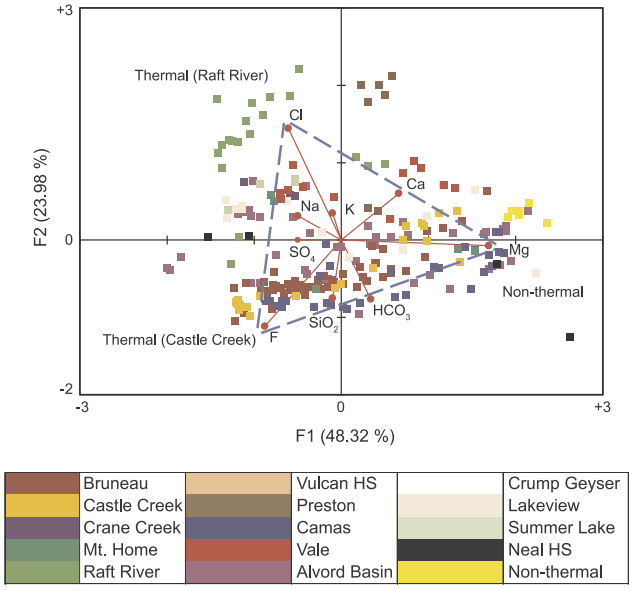

Fig. 4. Compositional distance biplot of first two principal components. Plotted points represent observation scores; the closer together two points are, the closer their common denominators. Rays (plotted in red) represent the loadings for each variable, with a smaller angle between rays indicating greater correlation, and opposing rays corresponding to inverse correlation. On this type of plot rays in which the end-points fall on the same link (blue dashed lines) may indicate the variables are controlled by a single process.

3.2.2. Cluster analysis

The techniques of statistical analysis collectively known as "cluster analysis" are a set of related algorithms for sorting N items or observations into K groups, usually as a means of discovering structure within complex data landscapes. In general, the aim of cluster analysis is to sort items into groups, or "clusters," that are internally homogeneous (i.e., the items in a cluster are "closely related," or "naturally associated"). At the same time, each cluster or associated group is distinct from other groups (Anderberg, 1973). The techniques of cluster analysis have been recognized at least since the 1960s; however, since the 1990s the method has found widespread applications in the field of molecular biology as a way to recognize patterns of gene expression in DNA data (Eisen et al., 1998).

In this study, we used two types of cluster analysis: hierarchal clustering and K-means clustering. For the analysis, we transformed all the continuous variables to their corresponding z-scores to allow comparison between variables of strongly contrasting magnitude. z-Score normalization is defined by:

$$z = \frac{x - \bar{x}}{s_x}, \tag{3}$$

where x is the variable to be transformed, $\bar{x}$ is the sample mean, s_x is the sample standard deviation, and z is the transformed variable. z-Score normalization transforms any sample distribution to a corresponding variable with a mean of 0 and a standard deviation of 1, while retaining rank information, thus allowing comparison between two (or more) variables of disparate magnitudes. As discussed in Section 4, the practical application of clustering techniques is computationally demanding, particularly with larger datasets. We used the statistical computing environment R (R Development Core Team, 2008) for the computation of clusters and visualization of results (i.e., dendrograms, scree plots), and performed our cluster analyses on the reduced set of nine variables identified by PCA as having the least redundant information about the sites.

Hierarchal Clustering. We chose to use hierarchal clustering as a preliminary tool to investigate the structure of the existing data. Because hierarchal clustering does not require the investigator to assign a target number of clusters, it is useful as an initial screening step (in addition to the PCA screening described in Section 3.2.1) for exploratory data analysis. Clustering algorithms use a defined "distance" (a measure of the difference) between objects in a dataset to quantify the strength of relationships. There are a number of potential choices as a measure of distance; for example, Euclidean distance, clr transform distance, etc. (Aitchison, 1986; Otero et al., 2005). For our investigation we used Euclidean distance, which is the "normal," or straightline distance in parameter space between two points. For any two objects x and y, characterized by p attributes, the Euclidean distance between them is given by (Rencher and Christensen, 2012):

$$d(x, y) = \left[\sum_{j=1}^{p} (x_j - y_j)^2\right]^{1/2}. \tag{4}$$

Hierarchal clustering uses the distance between objects to create a dendrogram. A dendrogram is a graphical representation of the relationships between objects. In the present instance, the objects are sampled sites or locations of hot springs or wells, and these objects are described by a vector of attributes such as temperature, depth, TDS, species concentrations, etc. Each object occupies a mutually exclusive position on the dendrogram "tree" (Everitt and Holthorn, 2010), and the separation between "branches" symbolizes the closeness or remoteness of relationship. The two most common methods used to define groups in hierarchal clustering are the complete linkage (often referred to as the "furthest neighbor" method) and the average linkage method. The complete linkage method uses the distance between the most remote objects to establish clusters. The maximum distance between two clusters A and B is (Rencher and Christensen, 2012):

$$D_{\max}(A, B) = \max_{x_i \in A,\, x_j \in B} d(x_i, x_j). \tag{5}$$

At each step of clustering, the maximum distance is found for all clusters, and the two clusters with the smallest maximum distance are merged.

Although complete linkage maintains the most remote clusters and merges closer relationships, average linkage groups objects by the closest relationships, or "highest cohesion" between groups, and builds higher-orders of the dendrogram by combining lower order groups. The cohesion between groups is defined as (Rencher and Christensen, 2012):

$$D_{\text{mean}}(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(x_i, x_j), \tag{6}$$

where n_A is the number of objects in cluster A and n_B is the number of objects in cluster B. Although the complete linkage method is more commonly used in practice, we applied both methods to the analysis of our datasets, and reconciled the results of the two methods to determine the optimum clustering of the data.

K-means Clustering. K-means clustering is one of the oldest methods for partitioning objects (collections of attributes or characteristics) into K related groups, and works by minimizing some objective function. Most commonly, the objective function to be minimized is a sum of squares, defined by:

$$\sum_{j} \sum_{i} (\bar{x}_{ij} - \bar{x}_j)^2, \tag{7}$$

where $\bar{x}_{ij}$ is the overall mean (the mean of all attributes for all groups) and $\bar{x}_j$ is the cluster mean (the mean distance for all attributes in the jth group). The objective of the analysis is to maximize the distance between, and minimize the distance within, a predetermined number of clusters, K. The optimum distribution of objects between clusters is accomplished by an iterative process (Everitt and Holthorn, 2010):

1. A number of clusters, K, is chosen by the investigator. In the present case, K was chosen on the basis of results obtained from hierarchal clustering.
2. k objects are chosen and placed into the K clusters, one object per cluster. The distances between all other, yet-to-be classified objects and the initial set of k classified objects are calculated, and each object is placed in the cluster to which it is the closest. Once all the objects have been binned, the overall sum of squares (Eq. (7)) is calculated for the entire dataset.
3. Objects are moved from one cluster to another cluster, and the sum of squares is evaluated for the new groupings.
4. If the sum of squares decreases for the new groupings, the groupings are retained. Otherwise, the objects in the original groups are replaced.
5. Steps 2–4 are repeated until convergence is achieved (i.e., until further reductions in the sum of squares are small enough to be considered negligible).

The results of our analysis, and the implications for targeting geothermal exploration in the study area, are discussed in the following sections.

4. Results and discussion

In Section 3.2.2, we alluded to the computationally demanding nature of cluster analysis, especially in situations involving large datasets. The difficulty of these types of sorting problems is discussed by Anderberg (1973), who noted the number of ways in which n observations can be placed into k groups is given by a Stirling number of the second kind (Abramowitz and Stegun, 1972):

$$\mathbb{S}_n^{(k)} = \frac{1}{k!} \sum_{j=0}^{j=k} (-1)^{k-j} \binom{k}{j} j^n, \tag{8}$$

where the 2×1 array indicates the binomial coefficients. Anderberg (1973) cites as an example the problem of putting 25 observations into 5 groups, for which there are more than 2×10^{15} possibilities. Furthermore, if the number of groups is not known *a priori*, it is necessary to sum the Stirling numbers over the potential numbers of groups (i.e., the Bell numbers). For the example of 25 observations, there are a minimum of one and a maximum of 25 groups, and Anderberg (1973) gives the number of possibilities as:

$$\sum_{k=1}^{25} \mathbb{S}_{25}^{(k)} > 4 \times 10^{18}. \tag{9}$$

Clearly, such a large number of possibilities cannot be searched exhaustively, and a great deal of effort has been put into finding computational methods that give near-optimum solutions with an achievable number of operations.

Apart from algorithmic advances (i.e., from the invention of new, more efficient ways to obtain near-optimum solutions to the clustering problem), there are several steps that may be taken to reduce the computational burden of cluster analysis. Because of the dramatic increase in the number of possibilities to be searched with an increasing number of clusters (i.e., Eq. (9)), the simplest way to reduce the computational burden is to limit the number of terms in the summation of Eq. (9) (i.e., only perform clustering with one – presumably optimum – value of k, rather than seeking optimum results by testing multiple possible values of k), and choosing the number of clusters to be a minimum, while still being large enough to be consistent with an optimum solution. We achieved this goal by first applying hierarchal clustering to the data to find an optimum number of clusters, then using this optimum number as the parameter for Step 1 in the K-Means Cluster Analysis.

The second thing that can be done to reduce the computational burden of cluster analysis is to reduce the number of attributes for each

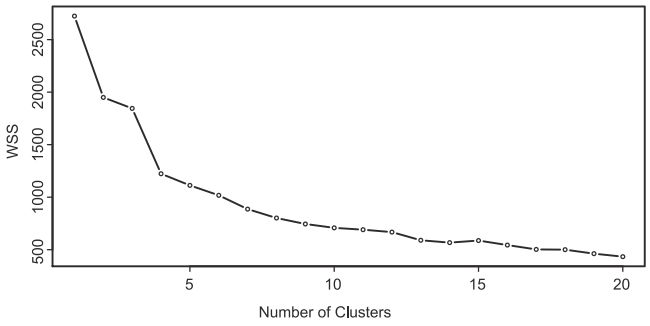

Fig. 5. Scree plot showing the weighted sum of squares (WSS) as a function of the number of clusters for *K*-means clustering. We chose $K = 10$ as the approximate value of the optimum *K*. The decrease in the WSS between *K* and $K + 1$ is greatest for values of $K < 10$; for values of $K \geq 10$, the incremental decreases were judged to be "small".

item. Although the Stirling number (Eq. (8)) gives the number of ways *n* items can be sorted into *k* groups, each item is characterized by *p* attributes. The calculation of Euclidean distance between items must be repeated for each trial grouping, and it is clearly inefficient to make distance calculations with, for example, ten attributes when seven of those attributes carry redundant information. To minimize the number of attributes used in the cluster analysis, we applied PCA to identify the attributes that provided the maximum ability to discriminate with the minimum number of factors. By eliminating from consideration linearly correlated factors, we were able to decrease the number of attributes in the distance calculations from 26 to 9 (see Section 3.2.1), with a commensurate savings in computational load.

From the Hierarchal Cluster Analysis results, we developed a "scree plot" (Fig. 5), which shows the sum-of-squares as a function of the number of clusters. From this plot, it is clear that the decrease in the sum-of-squares obtained for additional clusters becomes comparatively small in the vicinity of 10–12 clusters. From this result, we chose to pursue further analyses with *K*-means clustering, using $K = 10$. We also investigated the robustness of the clustering algorithm to changes in *K* by testing the outcome of using values in the range $10 \leq K \leq 12$.

In the final distribution of 10 clusters (see Fig. 6, and supplemental material S4), 5 Preston samples and 6 of the Raft River samples each binned together into their own clusters (Groups 2 and 6, respectively). However, 7 Alvord Basin samples, 40 from Bruneau, 11 from Camas, 7 from Castle Creek, and one sample each from Lakeview, Mountain Home, and Vale all binned together with 2 samples from production wells at the Raft River geothermal power plant to comprise the largest

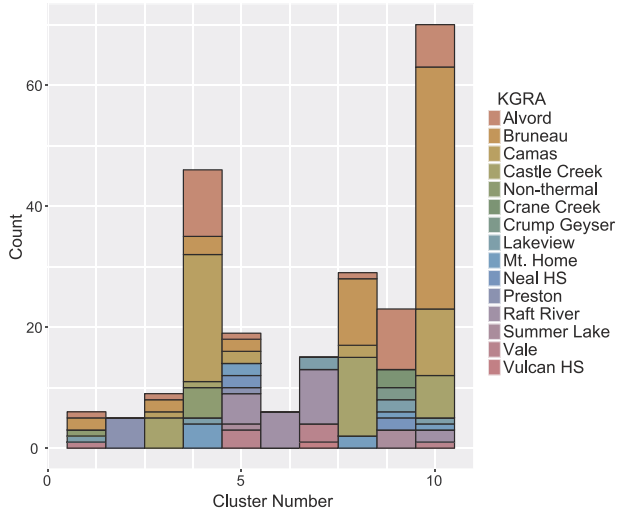

Fig. 6. Graphical representation of data points grouped by cluster for the case of 10 clusters (i.e., $K = 10$).

cluster of samples (Group 10; $N = 70$), which we propose as one of the high geothermal potential clusters. Another high potential cluster was Group 9 ($N = 23$), in which 10 Alvord Basin samples, 3 Crane Creek samples, 1 sample from Mountain Home, and 2 samples each from Crump Geyser and Lakeview bin together with the geothermal production wells from Neal Hot Springs (2 samples) and Summer Lake (3 samples). Other, smaller groups also showed development potential. For example, ~41% (9/22) of the Raft River samples binned together with 2 Lakeview samples, 3 Vale samples, and the sole sample from Vulcan Hot Springs in Group 7. Also, in Group 5 a scattering of samples from the Alvord Basin (1), Bruneau (2), Camas (2), Mountain Home (2), and Vale (3) grouped together with a well from Preston, one from Summer Lake, 2 production wells from Neal Hot Springs, and 5 wells from Raft River.

In contrast to the clusters containing the high geothermal potential sites, 5 of the 6 non-thermal control samples binned together with 11 Alvord Basin samples, 21 Camas samples, 4 Mountain Home samples, 3 Bruneau samples, and 1 sample each from Castle Creek and Lakeview in a low potential cluster (Group 4; $N = 46$). Of course, it is reasonable to expect that not all the samples from a particular area would tap the same resources, or show the same geothermal potential, even within an area of known high resource potential. This is illustrated by the 11 Alvord Basin samples that binned into the low potential Group 4, representing approximately 34% (11/32) of all the data points for this known high potential area. Inspection of the sample descriptions, however, indicates that all 11 Alvord Basin samples in this cluster were taken from low temperature springs and seeps that are clearly not associated with hydrothermal activity. Similarly, the 21 Camas samples in this group consist almost entirely of environmental (i.e., "cold") springs and water supply wells. The same is true for Group 1, another low potential group in which the remaining non-thermal control sample (USGS 134) binned with 1 Alvord Basin sample (an environmental-temperature seep on the SE side of Mickey Butte), 2 samples from the Bruneau area (both of which represent farm wells used for potable water), a sample from the Seventh-Day Adventist Church well in Vale, and a shallow thermal gradient hole in Lakeview (Hammersly Canyon No. 1).

It is worth pointing out that the "near optimal" nature of *K*-means (and other) clustering methods results in non-unique membership in the various clusters. The outcome of *K*-means clustering, in particular, is dependent on the initial choice of items for the first round of sorting, and different picks for the first *K* items will return somewhat different final groupings. Because different software implementations of the *K*-means clustering algorithm employ various strategies for choosing the first *K* items, the final groupings can shift between software packages and, in fact, even using a different ordering of the observations in the input deck can lead to variations in the final outcome. Fortunately, these differences are usually minor, commonly amounting to one or a few items trading groups, with no significant changes in the overall distribution.

More obviously, changing the number of clusters (i.e., changing the number of *K*) necessarily results in a shuffling and reorganization of the group memberships; however, in our experience the choice of *K* is not critical, provided the chosen value is greater than the point on the scree plot at which the decrease in the sum-of-squares for $K + 1$ groups becomes small. In the present study, we compared the results of *K*-means clustering for $K = 12$ with our choice for an "optimum" number of clusters, $K = 10$. In spite of the fact that group memberships (and, of course, group numbering) changed slightly for $K = 12$, the overall results of the cluster analysis changed very little (see supplemental material S4). The largest group (the new Group 11, with $N = 67$), which comprised the majority of high potential sites, included 64 of the members of the "original" (i.e., for $K = 10$) high potential Group 10. The remaining 6 members of the original high potential group moved to a new high potential group (the new Group 1), the membership of which roughly corresponded to the original ($K = 10$) Group 5, and

binned 3 Alvord Basin samples, 10 Camas samples, and 2 Bruneau samples with samples from Mountain Home (5), Neal Hot Springs (2), and Raft River (5). Interestingly, membership in the original Groups 2 and 6, which consisted solely of Preston (5) and Raft River (6) samples, respectively, was identical to the new ($K = 12$) Groups 12 and 4.

It must be pointed out that the results of the analysis presented here are essentially entirely dependent on geochemistry data; non-geochemical data that were included in the PCA analyses (e.g., resource depth, temperature, qualitative fracture density) were shown by the PCA to be highly correlated with geochemical parameters, and therefore largely redundant for the purposes of this investigation. Of the non-geochemical data that could potentially have made an important contribution to this analysis, the information on structural setting deserves additional discussion. As was mentioned in Section 3.1.2, the majority of hydrothermal discharge zones are fault-controlled (Curewitz and Karson, 1997), and the structural setting of an area with respect to faulting has been shown to be an important control on the distribution of permeable flow pathways, at least in Basin and Range-type systems (Faulds et al., 2011). The primary difficulty we encountered with assembling structural setting data is that such data are inconsistently reported in the literature, and it is often difficult to evaluate the uncertainty associated with this type of information. For many areas, detailed structural information is available for the KGRA as a whole; for example, structural setting data for Neal Hot Springs (Edwards, 2013), or Raft River (Dolene et al., 1981; Ayling and Moore, 2013), are excellent. For other areas, however, the structural setting may be uncertain or unknown (e.g., Vulcan Hot Springs). Most commonly, information on the structural setting of a KGRA is known in an overall sense, but the structural associations for individual samples are not reported, and this information may be difficult or impossible to determine for legacy data. In such cases, it is not clear whether the structural setting categories assigned to individual data points should reflect the overall structural setting of the KGRA, an educated guess by the analyst (both of which actions open the possibility of introducing bias into the dataset), or if the structural setting variable should be assigned a value of 'NA,' with the likely result being insufficient data to include the structural setting variable in the final analysis. This difficulty probably cannot be readily addressed for legacy datasets; however, the importance of structural setting in determining the commercial viability of a site, and the potential for such information to feed into studies such as the one presented here, suggest that new studies should always record information on structural setting on a per-sample level. Furthermore, future studies aimed at determining the structural settings of legacy data may also have value for understanding the development potential of KGRAs.

5. Conclusions

We have proposed a method for grouping a set of samples, each of which is characterized by a vector of attributes (temperature, TDS, species concentrations, etc.), so as to highlight areas that share characteristics with sites that are known to have a high potential for geothermal development. The method uses a set of statistical tools to sort sampled locations into related groups; those that cluster primarily with control sites for high geothermal potential (in this case, systems hosting producing geothermal power plants) are assumed to be targets of interest for future exploration and site characterization, while locations that cluster with non-thermal control locations are supposed to represent targets with lesser potential, or at least having greater risk. The methodology proposed here evaluates existing data using in sequence: (1) Principal Component Analysis (PCA) to identify those variables that demonstrate the least covariation (i.e., are closest to linear independence), and hence offer the greatest amount of new information with the smallest number of variables; (2) Hierarchal clustering, to determine the optimal number of clusters into which the locations may be sorted; and (3) *K*-means clustering, to establish a distribution of sites guided by their relationships to each other and the control locations. KGRAs that cluster with the high-potential locations are considered worthwhile targets for additional exploration and characterization, along with areas that fall in clusters that show the closest relationship to the high-potential groups. Conversely, areas grouped with non-thermal sites, and those clusters most closely related to the low potential controls, are supposed to be less attractive or more risky targets for exploration potential.

In the present study, we examined data from 14 KGRAs and IHRAs, and one non-thermal control group. All the sites examined in this study were located in eastern Oregon and southern Idaho (Table 1 and Fig. 2), and the data were collected primarily during the 1980s. In addition to the candidate areas for future exploration/characterization, we included three producing geothermal areas (Raft River, Neal Hot Springs, and Summer Lake/Paisley Hot Springs) as high geothermal potential control locations; the group of non-thermal areas was included as a low potential control site. After removing from consideration any variables that were not measured at more than 25% of the locations, our dataset consisted of 228 samples distributed throughout the 15 areas of Table 1, and characterized by a vector of 26 continuous variables. We refined this vector to a set of 9 variables (Na, K, Ca, Mg, SiO_2, Cl, HCO_3, SO_4, and F) that PCA analysis identified as best representing the variability and relationships between samples. For our dataset, hierarchal clustering indicated that the optimum number of groups into which the data could be sorted was approximately 10, and the final sorting of the 228 sample locations into 10 groups was accomplished with *K*-means clustering.

On the basis of our analysis, we consider the Bruneau KGRA, Castle Creek KGRA, and Alvord Basin KGRA to be targets of highest priority for future geothermal exploration and characterization. Crane Creek and Vale KGRAs are also likely high potential targets, although the relatively small number of samples available to characterize these areas (3 and 8 samples, respectively) add uncertainty to their assessment. Other areas, such as Camas Prairie and Lakeview KGRAs, may be good targets for exploration but are considered higher risk, due to the fact that locations within these areas are split between high-potential clusters and the low-potential group. Two of the areas included in this study, Crump Geyser and Vulcan Hot Springs, were represented by one (Vulcan Hot Springs) or two (Crump Geyser) samples. These areas clustered with high-potential groups (Groups 7 and 9, respectively) and may offer good opportunities for future exploration, but the small number of observations available makes it difficult to determine the significance of their placement.

It is difficult to offer a convincing "proof" of the efficacy of the proposed method, since any *ex ante* predictions would necessarily require verification by drilling and, ultimately, commercial development. However, in terms of *ex post facto* "predictions," it is instructive to review the results of the cluster analysis with respect to the Alvord Basin samples. Although socio-economic (distance to market, available transmission capacity) and socio-political (land-use designations) factors have conspired to suppress development in the Alvord Basin, geothermometry (Cummings et al., 1993; Koski and Wood, 2004) and exploratory drilling by Phillips Geothermal and Anadarko Petroleum all indicate the presence of viable resources in terms of enthalpy. The binning of thermal samples from the Alvord Basin with production wells from Raft River and Neal Hot Springs (and conversely, of non-thermal samples with the low-potential control group) provides some indication that the method yields reasonable results. The strength of the proposed analysis is that it makes efficient use of existing data that may vary widely in quality and granularity, and can therefore provide guidance on the prioritization of exploration targets in situations where data uncertainty would typically make site-to-site comparisons problematic. However, we emphasize that the method proposed here does not rank potential targets for exploration, or explicitly call-out high value resource areas. Many measures that are crucial for identifying the development potential of a geothermal area are not included in the

proposed analysis; for example, we have made no attempt to apply standard geothermometers (e.g., Giggenbach, 1988), or explicitly incorporate information on subsurface temperatures into the proposed methodology. The decision to proceed (or not to proceed) with the characterization or development of an area is complex, and cannot be reduced to an objective algorithm. Moreover, interpretation of the results of the PCA/cluster analyses requires expert judgment and a familiarity with the potential resource areas. High value targets may be spread over several groups, increasing uncertainty about a target's true potential. One reason for this is that geochemical data from any given area often include samples from water supply wells, environmental (i.e., non-thermal) springs and seeps, etc. These non-thermal samples are likely to be binned with non-thermal control sites; therefore, the grouping of some samples from an area into low potential clusters does not necessarily reflect the overall potential of the area. In the present analysis, this situation is perhaps best illustrated by the results for the Alvord Basin: 11 out of 32 samples from that area clustered with the non-thermal control samples, in spite of the fact that the Alvord Basin resource is known from geothermometry and drilling to be a high-value target. Thus, it is important to be familiar with the data being analyzed, and to differentiate between samples that are likely to indicate conditions in a geothermal reservoir and those that are non-thermal.

Other factors may also cause samples from a single site to bin into different groups. Many analyses utilizing legacy data are likely to depend on a mix of samples from surface water (i.e., hot springs) and wells, as is the case for the present study. There are, of course, differences in chemistry between samples obtained from deep wells and those from surface features, the discharge from which has often been modified by cooling, boiling, and mixing with dilute (non-thermal) surface fluids. These differences may or may not influence the conclusions drawn from this, or other, methods of analysis (e.g., Neupane et al., 2015), but the practitioner should be aware of the potential for such factors to impact the final result. In the present study, PCA showed that the available non-geochemical variables added little new information to the analysis; however, it is not clear that this would be true for all sets of potential target areas. If non-geochemical variables that added unique information to a PCA/cluster analysis could be identified, the inclusion of such data may help to minimize the potential for confounding factors associated with geochemical data (e.g., differences between deep well and surface discharge samples). In particular, we have noted that the inclusion of information on structural setting of individual samples (i.e., sample location with respect to fault structures) may offer critical insight into the development potential of an area. Although this type of information is not consistently reported on a per-sample basis in legacy data, the potential benefit of its inclusion in an analysis of viability suggests that it should always be reported in new data gathering efforts, and incorporated into development assessments whenever possible.

It is also worth noting that many non-geological factors also influence the development potential of a hydrothermal resource area. In the present study, we focused on factors that effectively correlate with the temperature and, to a lesser extent, the enthalpy of a resource. Notably absent is any consideration of economic factors; for example, distance to market for produced power, the presence of transmission lines or available transmission capacity, or access to cooling water. Similarly, restrictions due to land ownership or zoning are neglected. Because our analysis focuses on primarily geological factors (and in particular, on geochemical factors), and many of the other factors are social or economic in nature, they do not fit comfortably in the framework we have proposed here.

The strengths of the method proposed here for identifying potential geothermal exploration targets are that it is inexpensive, can use existing data of uneven quality and granularity, has the potential to incorporate “soft” data, and provides a rapid assessment of prospects. The semi-quantitative nature of the analysis, the need for expert judgment to interpret the results, and the lack of an absolute ranking of exploration targets suggest the value of the method is primarily heuristic. Although heuristic methods are not guaranteed to provide optimal solutions, they are valued for their ability to offer satisfactory approaches to problems that are impossible or impractical to solve by optimal methods. We believe the method described here offers an important tool for efficiently seeking relationships between data of variable or uncertain quality, and has significant potential for identifying geothermal resources that might otherwise be overlooked, particularly when used in combination with other metrics for geothermal resource analysis, such as play fairway analysis, (e.g., Ito et al., 2017; Shervais et al., 2017; Siler et al., 2017).

Acknowledgments

This work was supported by funding from the Assistant Secretary for Energy Efficiency and Renewable Energy, Geothermal Technologies Office of the US Department of Energy under the USDOE contract numbers DE-AC07-05ID14517 for Idaho National Laboratory and DE-AC02-05CH11231 for Lawrence Berkeley National Laboratory. The authors would like to thank J. Faulds and two anonymous reviewers, and Editor C. Bromley, for helpful and constructive comments that clarified and strengthened the manuscript; in addition, we thank Dick Benoit for sharing data from his early work at the Alvord Basin KGRA.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.geothermics.2017.12.009.

References

Abramowitz, M., Stegun, I., 1972. Handbook of Mathematical Functions with Formulas, Graphs, and Tables. Dover Publications.
Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London.
Aitchison, J., Greenacre, M., 2002. Biplots of compositional data. Appl. Stat. 51 (4), 375–392.
Anderberg, M.R., 1973. Cluster Analysis for Applications. Probability and Mathematical Statistics. Academic Press, Inc.
Anderson, T.R., Fairley, J.P., 2008. Relating permeability to the structural setting of a fault-controlled hydrothermal system in southeast Oregon, USA. J. Geophys. Res. 113, B05402.
Arney, B.H., 1982. Evidence of former higher temperatures from alteration minerals, Bostic 1-A well, Mountain Home, Idaho. GRC Transactions, vol. 6. Geothermal Resources Council, pp. 3–6.
Arney, B.H., Gardner, J.N., Bellnomi, S.G., 1984. Petrographic analysis and correlation of volcanic rocks in Bostic 1-A well near Mountain Home, Idaho. Report LA-9966-HDR. Los Alamos National Laboratory.
Arney, B.H., Goff, G., Harding Lawson Associates, 1982. Evaluation of the hot dry rock geothermal potential of an area near Mountain Home Idaho. Report LA-9365-HDR. Los Alamos National Laboratory.
Ayling, B., Moore, J., 2013. Fluid geochemistry at the Raft River geothermal field, Idaho, USA: new data and hydrogeological implications. Geothermics 47, 116–126.
Bertani, R., 2015. Geothermal power generation in the world, 2010–2014 update report. In: Proceedings World Geothermal Congress. Melbourne, Australia. pp. 19.
Blackwell, D.D., 1983. Heat flow in the northern Basin and Range Province. The Role of Heat in the Development of Energy and Mineral Resources in the Northern Basin and Range Province, vol. 13. Geothermal Resources Council, pp. 81–93.
Blackwell, D.D., 1989. Regional implications of heat flow on the Snake River Plain, Northwestern United States. Tectonophysics 164, 323–343.
Blackwell, D.D., Richards, M., 2004. Geothermal Map of North America. Map (1 Sheet), Scale 1:6,500,000. American Association of Petroleum Geologists, Tulsa, OK.
Blewitt, G., Coolbaugh, M.F., Sawatzky, D.L., Holt, W., Savis, J., Bennett, R., 2003. Targeting of potential geothermal resources in the Great Basin from regional to basin-scale relationships between geodetic strain and geological structures. GRC Transactions, vol. 27. Geothermal Resources Council, pp. 3–7.
Bradford, J., McLennan, J., Moore, J., Podgorney, R., Tiwari, S., 2015. Hydraulic and thermal stimulation program at Raft River Idaho, a DOE EGS. GRC Transactions, vol. 39. Geothermal Resources Council, pp. 261–268.
Burkhardt, H.E., Brook, C.A., Smith, F.W., 1980. Selected administrative, land, and resource data for Known Geothermal Resource Areas in Arizona, California, Nevada, Oregon, and Washington. Open File Report 80-1290. United States Geological Survey.
Casteel, J., 2010. An overview and conceptual model of the Crump Geyser geothermal area. GRC Transactions, vol. 34. Geothermal Resources Council, pp. 1131–1134.
Christopherson, K.R., Senterfit, R.M., Dalati, M., 1980. Telluric profiles and location map

for Vulcan Hot Springs Known Geothermal Resource Area, Idaho. Open File Report 80-518. United States Geological Survey.
Connolly, P.T., Cosgrove, J.W., 1999. Prediction of static and dynamic fluid pathways within and around dilational jogs. Fractures, Fluid Flow and Mineralization, vol. 155. Geological Society, London, pp. 105–121.
Coolbaugh, M.F., Sawatzky, D.L., Oppliger, G.L., Minor, T.B., Raines, G.L., Shevenell, L.A., Blewitt, G., Louie, J.N., 2003. Geothermal GIS coverage of the Great Basin, USA. GRC Transactions, vol. 27. Geothermal Resources Council, pp. 9–11.
Cummings, M.L., John, A.M.S., Sturchio, N.C., 1993. Hydrgeochemical characterization of the Alvord Basin geothermal area, Harney County, Oregon, USA. Proceedings of the 15th New Zealand Geothermal Workshop. University of Auckland, Auckland, New Zealand.
Curewitz, D., Karson, J.A., 1997. Structural settings of hydrothermal outflow: fracture permeability maintained by fault propagation and interaction. J. Volcanol. Geotherm. Res. 79, 149–168.
Dolene, M.R., Hull, L.C., Mizell, S.A., Russell, B.F., Skiba, P.A., Strawn, J.A., Tullis, J.A., 1981. Raft River geoscience case study Tech. Rep. EGG-2125. US Department of Energy.
Eckart, C., Young, G., 1936. The approximation of one matrix by another of lower rank. Psychometrika 1, 211–218.
Edwards, J.H., 2013. Structural Controls of the Neal Hot Springs Geothermal System, Eastern Oregon (Master's thesis). University of Nevada, Reno, Nevada.
Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expressions patterns. Proc. Natl. Acad. Sci. U. S. A. 95, 14863–14868.
Everitt, B., Holthorn, T., 2010. A Handbook of Statistical Analyses Using R. CRC Press, Boca Raton, FL.
Fairbank, B.D., Smith, N., 2015. Crump Geyser final report. Tech. Rep. DE-EE0002835. US Department of Energy.
Fairley, J.P., 2009. Modeling fluid flow in a heterogeneous, fault-controlled hydrothermal system. Geofluids 9, 153–166.
Fairley, J.P., Hinds, J.J., 2004. Rapid transport pathways for geothermal fluids in an active Great Basin fault zone. Geology 32, 825–828.
Faulds, J.E., Hinz, N.H., Coolbaugh, M.F., Cashman, P.H., Kratt, C., Dering, G., Edwards, J., Mayhew, B., McLachlan, H., 2011. Assessment of favorable structural settings of geothermal systems in the Great Basin, Western USA. GRC Transactions, vol. 35. Geothermal Resources Council, pp. 777–783.
Faulds, J.E., Hinz, N.H., Dering, G.M., Siler, D.L., 2013. The hybrid model – the most accommodating structural setting for geothermal power generation in the Great Basin, western USA. GRC Transactions, vol. 37. Geothermal Resources Council, pp. 3–10.
Fern, M.L., Brooks, H.C., Evans, J.G., Cummings, M.L., 1993. Geologic map of the Vale 30 × 60 minute quadrangle, Malheur County, Oregon, and Owyhee County, Idaho. Tech. Rep. GMS-77. Oregon Department of Geology and Mineral Industries.
Forster, C., Smith, L., 1989. The influence of groundwater flow on thermal regimes in mountainous terrain: a model study. J. Geophys. Res. 94, 9439–9451.
Fowler, A.P.G., Spycher, N., Zierenberg, R.A., Cantwell, C.A., 2017. Identification of blind geothermal resources in Surprise Valley, CA, using publicly available groundwater well water quality data. Appl. Geochem. 80, 24–48.
Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. Biometrika 58 (3), 453–467.
Giggenbach, W.F., 1988. Geothermal solute equilibria: derivation of Na–K–Mg–Ca geoindicators. Geochim. Cosmochim. Acta 52, 2749–2765.
Glen, J.M.G., Connard, G., Casteel, J., Walsh, P., 2015. Assessing structural controls on geothermal fluids from a three-dimensional geophysical model of Warner Valley, Oregon, USA. In: Proceedings World Geothermal Congress. Melbourne, Australia. pp. 12.
Glen, J.M.G., Liberty, L., Gasperikova, E., Siler, D., Shervais, J., Ritzinger, B., Athens, N., 2017. Geophysical investigations and structural framework of geothermal systems in west and southcentral Idaho; Camas Prairie to Mountain Home. Proceedings, 42nd Workshop on Geothermal Reservoir Engineering. Stanford University, Stanford, CA, pp. 14.
Godwin, L.H., Haigler, L.B., Rious, R.L., White, D.E., Muffler, L.J.P., Wayland, R.G., 1971. Classification of public lands valuable for geothermal steam and associated geothermal resources. USGS Circular 647. United States Geological Survey.
Hess, S., Fairley, J.P., Bradford, J., Lyle, M., Clement, W., 2009. Evidence for composite hydraulic architecture in an active fault system based on 3D seismic reflection, time-domain electromagnetics and temperature data. Near Surf. Geophys. 2009, 341–352.
Hettiarachchi, H.D.M., Golubovic, M., Worek, W.M., Ikegami, Y., 2007. Optimum design criteria for an organic Rankine cycle using low-temperature geothermal heat sources. Energy 32 (9), 1698–1706.
Hickman, R.G., Dobson, P.F., vanGerven, M., Sagala, B., Gunderson, R.P., 2004. Tectonic and stratigraphic evolution of the Sarulla Graben region, North Sumatra, Indonesia. J. Asian Earth Sci. 23, 435–448.
Ito, G., Frazer, N., Lautze, N., Thomas, D., Hinz, N., Waller, D., Whittier, R., Wallin, E., 2017. Play fairway analysis of geothermal resources across the state of Hawaii: 2. Resource probability mapping. Geothermics 70, 393–405.
Jenks, M.D., Bonnichsen, B., Godchaux, M.M., 1998. Geologic maps of the Grand View-Bruneau area, Owyhee County, Idaho. Technical Report 98-1. Idaho Geological Survey.
Koenig, J.B., McNitt, J.R., 1983. Controls on the location and intensity of magmatic and non-magmatic geothermal systems in the Basin and Range Province. The Role of Heat in the Development of Energy and Mineral Resources in the Northern Basin and Range Province, vol. 13. Geothermal Resources Council, pp. 93.
Koski, A.K., Wood, S.A., 2004. The geochemistry of geothermal waters in the Alvord Basin, southeastern Oregon. In: Wanty, Richard B., Seal IIRobert R. (Eds.), Proceedings of the 11th International Symposium on Water-Rock Interaction. Saratoga Springs, Taylor and Francis Group, New York, pp. 149–152.
Lewis, R.E., Link, P.K., Stanford, L.R., Long, S.P., 2012. Geologic map of Idaho, Tech. Rep. M-9. Idaho Geological Survey.
Li, D., Huang, L., Jones, C., Moore, J., Freeman, K., Majer, E., 2017. Characterizing Narrows Zone at the Raft River EGS site using 2010 VSP data and microseismic data. GRC Transactions, vol. 41. Geothermal Resources Council, pp. 1669–1679.
Long, C.L., Kaufmann, H.E., 1980. Reconnaissance geophysics of a Known Geothermal Resource Area, Weiser, Idaho, and Vale, Oregon. Geophysics 45 (2), 312–322.
López, D.L., Smith, L., 1995. Fluid flow in fault zones: analysis of the interplay of convective circulation and topographically driven groundwater flow. Water Resour. Res. 31 (6), 1489–1503.
López, D.L., Smith, L., 1996. Fluid flow in fault zones: influence of hydraulic anisotropy and heterogeneity on the fluid flow and heat transfer regime. Water Resour. Res. 32 (10), 3227–3235.
Makowsky, K., 2013. The Geothermal System Near Paisley, Oregon: A Tectonomagmatic Framework for Understanding the Geothermal Resource Potential of Southeastern Oregon (Master's thesis). Boise State University, Boise, Idaho.
McLing, T.L., Smith, R.P., Smith, R.W., Blackwell, D.D., Roback, R.C., Sondrup, A.J., 2016. Wellbore and groundwater temperature distribution eastern Snake River Plain, Idaho: implications for groundwater flow and geothermal potential. J. Volcanol. Geotherm. Res. 320, 144–155.
McLing, T.L., Smith, R.W., Johnson, T.M., 2002. Chemical characteristics of thermal water beneath the eastern Snake River Plain, Idaho. In: Link, P.K., Mink, L.L. (Eds.), Geology, Hydrogeology, and Environmental Remediation: Idaho National Engineering and Environmental Laboratory, Eastern Snake River Plain, Idaho, vol. Special Paper 353. Geological Society of America, pp. 205–211.
Micklethwaite, S., Cox, S.F., 2004. Fault-segment rupture, aftershock-zone fluid flow, and mineralization. Geology 32 (9), 813–816.
Mink, L.L., Pezzopane, S.K., Culp, E.L., 2014. Surprise Valley Electric Paisley, Oregon geothermal power project. GRC Transactions, vol. 38. Geothermal Resources Council, pp. 85–89.
Mitchell, J.C., 1976. Geothermal investigations in Idaho, Part 5. Geochemistry and geologic setting of the thermal waters of the northern Cache Valley area, Franklin County, Idaho. Water Information Bulletin 30. Idaho Department of Water Resources.
Moore, B., Savage, N., Gladwell, J.S., Warnick, C.C., 1979. A summary of the assessment of geothermal resource use limitations of Bruneau KGRA, Castle Creek KGRA, Crane Creek KGRA, Mountain Home KGRA, Vulcan KGRA. Tech. rep. Idaho Water Resources Research Institute, University of Idaho, Moscow, ID.
Morgan, L.A., Pierce, K.L., Shanks, W.C.P., 2008. Track of the Yellowstone hotspot: Young and ongoing geologic processes from the Snake River Plain to the Yellowstone Plateau and Teton. In: In: Raynolds, R.G. (Ed.), Roaming the Rocky Mountains and Environs, vol. 10. Geological Society of America, pp. 139–173.
Neupane, G., Baum, J.S., Mattson, E.D., Mines, G.L., Palmer, C.D., Smith, R.W., 2015. Validation of multicomponent equilibrium geothermometry at four geothermal power plants. Proceedings Fortieth Workshop on Geothermal Reservoir Engineering. Stanford University, Stanford, CA.
Neupane, G., Mattson, E.D., McLing, T.L., Palmer, C.D., Smith, R.W., Wood, T.R., Podgorney, R.K., 2016. Geothermometric evaluation of geothermal resources in southeastern Idaho. Geotherm. Energy Sci. 4 (1), 11.
Neupane, G., Mattson, E.D., Spycher, N., Dobson, P.F., Conrad, M.E., Newell, D.L., McLing, T.L., Wood, T.R., Cannon, C.J., Atkinson, T.A., Brazell, C.W., Worthing, W.C., 2017. Geochemical evaluation of the geothermal resources of Camas Prairie, Idaho. Proceedings 42nd Workshop on Geothermal Reservoir Engineering. Stanford University, Stanford, CA, pp. 12.
Nielson, D.L., Sonnenthal, E., Shervais, J.W., Garg, S.K., 2017. Mafic heat sources for Snake River Plain geothermal systems. Proceedings, 42nd Workshop on Geothermal Reservoir Engineering. Stanford University, Stanford, CA, pp. 8.
Otero, N., Tolosana-Delgado, R., Soler, A., Pawlowsky-Glahn, V., Canals, A., 2005. Relative vs. absolute statistical analysis of compositions: a comparative study of surface waters of a Mediterranean river. Water Res. 39, 1404–1414.
Pearson, K., 1897. Mathematical contributions to the theory of evolution. On the form of spurious correlation which may arise when indices are used in the measurement of organs. In: Proceedings of the Royal Society of London, vol. LX. The Royal Society of London. pp. 489–502.
Pezzopane, S.K., Weldon, R.J., 1993. Tectonic role of active faulting in central Oregon. Tectonics 12, 1140–1169.
Pierce, K.L., Morgan, L.A., 1992. The track of the Yellowstone Hotspot: volcanism, faulting, and uplift. Regional Geology of Western Idaho and Western Wyoming, vol. 179. Geological Society of America, pp. 1–53.
Piper, A.M., 1924. Geology and water resources of the Bruneau River Basin, Owyhee County. Tech. Rep. 11. Idaho Bureau of Mines and Geology.
Podgorney, R.K., McCurry, M., Wood, T., McLing, T., Ghassemi, A., Welhan, J., Mines, G., Plummer, M., Moore, J., Fairley, J.P., Wood, R., 2013. Enhanced geothermal system potential for sites on the Eastern Snake River Plain, Idaho. GRC Transactions, vol. 37. Geothermal Resources Council, pp. 191–197.
R Development Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.
Rencher, A.C., Christensen, W.F., 2012. Methods of Multivariate Analysis. John Wiley & Sons, Inc., Hoboken, NJ.
Rightmire, C.T., Young, H.W., Whitehead, R.L., 1976. Geothermal investigations in Idaho, Part 4. Isotopic and geochemical analyses of water from the Bruneau-Grand View and Weiser areas, southwest Idaho. Water Information Bulletin 30. Idaho Department of Water Resources.
Rodgers, D.W., Ore, H.T., Bobo, R.T., McQuarrie, N., Zentner, N., 2002. Extension and

subsidence of the eastern Snake River Plain, Idaho. Tectonic and Magmatic Evolution of the Snake River Plain Province, vol. 30. Idaho Geologic Survey, pp. 121–155.
Sbar, M.L., Barazangi, M., Dorman, J., Scholz, C.H., Smith, R.B., 1972. Tectonics of the Intermountain Seismic Belt, western United States: microearthquake seismicity and composite fault plane solutions. Geol. Soc. Am. Bull. 83 (1), 13–28.
Shervais, J.W., Glen, J.M.G., Nielson, D.L., Garg, S., Liberty, L.M., Siler, D., Dobson, P., Gasperikova, E., Sonnenthal, E., Neupane, G., DeAngelo, J., Newell, D.L., Evans, J.P., Snyder, N., 2017. Geothermal play fairway analysis of the Snake River Plain: Phase 2. GRC Transactions, vol. 41. Geothermal Resources Council, pp. 2328–2345.
Shervais, J.W., Schmitt, D.R., Nielson, D., Evans, J.P., Christiansen, E.H., Morgan, L., Shanks, W.C.P., Blackwell, D.D., Glen, J.M., Champion, D., Potter, K.E., Kessler, J.A., 2013. First results from HOTSPOT: the Snake River Plain scientific drilling project, Idaho, USA. Sci. Drill. 15, 36–45.
Shervais, J.W., Vetter, K.K., Hanan, B.B., 2006. Layered mafic sill complex beneath the eastern Snake River Plain: evidence from cyclic geochemical variations in basalt. Geology 34 (5), 365–368.
Sifford, A., 2014. Geothermal energy in Oregon. GRC Transactions, vol. 38. Geothermal Resources Council, pp. 99–106.
Siler, D.L., Zhang, Y., Spycher, N.F., Dobson, P.F., McClain, J.S., Gasperikova, E., Zierenberg, R.A., Schiffman, P., Ferguson, C., Fowler, A., Cantwell, C., 2017. Play-fairway analysis for geothermal resources and exploration risk in the Modoc Plateau region. Geothermics 69, 15–33.
Smith, R.B., Jordan, M., Steinberger, B., Puskas, C.M., Farrell, J., Waite, G.P., Husen, S., 2009. Geodynamics of the Yellowstone hotspot and mantle plume: seismic and GPS imaging, kinematics, and mantle flow. J. Volcanol. Geotherm. Res. 188 (1–3), 26–56.
Tester, J.W., Anderson, B.J., Batchelor, A.S., Blackwell, D.D., DiPippo, R., Drake, E.M., Garnish, J., Livesay, B., Moore, M.C., Nichols, K., Petty, S., Toksöz, M.N., Veatch, R.W., 2006. The future of geothermal energy: impact of enhanced geothermal systems (EGS) on the United States in the 21st Century. INL report INL/EXT-06-11746. Idaho National Laboratory.
Whelan, J.A., 2016. Thermal and geochemical anomalies in the Eastern Snake River Plain Aquifer: contributions to a conceptual model of the proposed FORGE test site. In: Proceedings, 41st Workshop on Geothermal Reservoir Engineering, vol. 41. February 22–24. Stanford University, Stanford, CA, pp. 20.
Wisian, K.W., Blackwell, D.D., Teplow, W.J., Meidav, T., 1996. Interpretation of geophysical data for the Vale, Oregon geothermal system. GRC Transactions, vol. 20. Geothermal Resources Council, pp. 435–438.
Wood, T., Worthing, W., Cannon, C., Palmer, C., Neupane, G., McLing, T.L., Mattson, E., Dobson, P., Conrad, M., 2015. The Preston geothermal resources: renewed interest in a Known Geothermal Resource Area. Proceedings of the 40th Workshop on Geothermal Reservoir Engineering. Stanford University, Stanford, CA.
Young, H.W., Mitchell, J.C., 1973. Geothermal investigations in Idaho, Part I. Geochemistry and geologic setting of selected thermal waters. Water Information Bulletin 30. Idaho Department of Water Resources.
Young, H.W., Whitehead, R.L., 1975. Geothermal investigations in Idaho, Part 3. An evaluation of thermal water in the Weiser area, Idaho. Water Information Bulletin 30. Idaho Department of Water Resources.
Young, H.W., Whitehead, R.L., Hoover, D.B., Tippens, C.L., 1975. Geothermal investigations in Idaho, Part 2. An evaluation of thermal water in the Bruneau-Grand View area, southwest Idaho. Water Information Bulletin 30. Idaho Department of Water Resources.
Zarrouk, S.J., Moon, H., 2014. Efficiency of geothermal power plants. a worldwide review. Geothermics 51, 142–153.
Zhang, Y., Schaubs, P.M., Zhao, C., Ord, A., Hobbs, B.E., Barnicoat, A.C., 2008. Fault-related dilation, permeability enhancement, fluid flow and mineral precipitation patters: numerical models. The Internal Structure of Fault Zones, vol. 229. Geological Society, London, pp. 239–255.