PROCEEDINGS, 51st Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 9-11, 2026
SGP-TR-230

Explainable EGS Targeting with Self-Organizing Maps: Multivariate Analysis of Proxies at Utah FORGE and Roosevelt Hot Springs

Zeynep Ankut Aydar* and Sebnem Düzgün*

Colorado School of Mines, Mining Engineering Department, Golden, CO 80401

zeynep_ankut@mines.edu

Keywords: Enhanced Geothermal Systems, Self-Organizing Maps, Machine Learning, 3D Subsurface Modeling, Explainable AI, Statistical Validation, Utah FORGE, Geothermal Prospectivity

ABSTRACT

Enhanced Geothermal System (EGS) target identification requires integration of multimodal geoscientific data to capture thermal state, rock properties, and structural controls across three-dimensional (3D) subsurface volumes. Understanding the non-linear relationships between EGS favorability proxies is essential for targeted data collection and informed prospectivity analysis, yet traditional approaches lack robust mechanisms to reveal these critical interdependencies. This study presents a framework for EGS characterization that combines Self-Organizing Maps (SOM) with depth-resolved statistical validation to distinguish EGS from naturally conventional hydrothermal systems. We developed integrated 3D subsurface models at Utah Frontier Observatory for Research in Geothermal Energy (Utah FORGE) and Roosevelt Hot Springs (RHS) comprising 4.5 million blocks at 100 m resolution, incorporating 11 attributes: P-wave velocity (Vp), S-wave velocity (Vs), Vp/Vs ratio, density inferred from 3D magnetic gravity inversion, temperature, fault density, and five lithology indicators.

A 3×3 SOM was trained using a representative subset of the 4.5M valid blocks and then applied to the full block volume to obtain 3D cluster assignments. Component plane analysis revealed systematic correlations: high-temperature domains consistently co-locate with elevated Vp/Vs ratios and intermediate densities, confirming coupling between thermal anomalies, fluid presence, and enhanced permeability. Within the RHS footprint, Cluster C7 represents the dominant high-temperature regime and shows negligible within-cluster sampling bias (Δmean $\approx$ 0). This supports that the footprint-labeled blocks are representative of the underlying SOM prototype.

Depth-resolved statistical validation using dual parametric (Welch's t-test, Hedges' g) and non-parametric (Mann–Whitney U, Cliff's δ) frameworks quantified system-level contrasts across 500 m depth bins. RHS exhibits very large thermal effect sizes at shallow depths ($g > 2$, $|\delta| > 0.7$) reflecting strong convective upflow, while FORGE shows comparable or elevated temperatures below 3000 m depth consistent with deep EGS targeting.

This work establishes a reproducible, objective framework for 3D EGS prospectivity assessment that eliminates subjective weighting, delivers quantitative explainability of proxy interdependencies, and provides spatially explicit volumetric delineation of prospect quality. Unlike black-box approaches, the SOM architecture preserves interpretability while handling high-dimensional data, enabling geoscientists to interrogate proxy relationships, validate assignments against domain knowledge, and extract transferable insights. Explainability is provided via SOM component planes (proxy co-variation), cluster prototypes, and depth-binned effect sizes that quantify where and why system contrasts occur. The integrated methodology advances EGS exploration by demonstrating that combined unsupervised learning and rigorous statistical validation can identify subsurface targets in structurally complex basement environments while maintaining full transparency of decision criteria.

1. INTRODUCTION

EGS represent a transformative pathway for expanding geothermal energy production beyond naturally permeable hydrothermal reservoirs by engineering permeability in hot, low-permeability crystalline basement rocks. Successful EGS development requires accurate identification of subsurface volumes that combine favorable thermal conditions, appropriate lithological and mechanical properties, existing fracture networks, and structural geometries amenable to stimulation. Unlike conventional hydrothermal systems characterized by natural convection, surface thermal manifestations, and self-sustaining permeability, EGS targets are defined by engineered fluid pathways within conduction-dominated thermal patterns at depths often exceeding 3000 m.

Traditional geothermal exploration methods rely heavily on proxy analyses, such as temperature gradient mapping, structural interpretation, and geophysical anomaly identification, followed by expert judgment to synthesize these proxies into prospect rankings. While these approaches have proven effective for shallow, high-temperature hydrothermal systems with clear surface manifestation, they face fundamental limitations when applied to deep EGS environments where target signatures are subtle, multivariate, and depth-dependent. The proxy thresholds cannot adequately capture the complex interplay between thermal state, elastic properties, structural architecture, and lithological heterogeneity that governs EGS reservoir potential Ankut and Düzgün (2025).

Sites like the Utah Frontier Observatory for Research in Geothermal Energy (FORGE) offer rich subsurface datasets (e.g 3D seismic velocity models from distributed acoustic sensing and borehole geophones, thermal fields interpolated from wellbore measurements, density distributions from magnetic gravity inversion, mapped fault networks, and detailed lithological frameworks from drilling) that can advance our understanding of the relationships between proxies and their thresholds for EGS site selection. However, integrating these

heterogeneous datasets into unified prospectivity assessments remains challenging due to scale inconsistencies, proxy interdependencies, and the lack of robust frameworks for multivariate pattern recognition in 3D geological contexts.

Machine learning (ML) offers powerful tools for discovering patterns in high-dimensional datasets, yet many contemporary approaches suffer from a lack of interpretability (the "black-box" problem), which can limit adoption in geosciences where physical understanding and stakeholder trust are paramount. SOM, introduced by (Kohonen, 2001), are powerful for extracting patterns from multidimensional data sets such as EGS data. SOM performs unsupervised dimensionality reduction by mapping high-dimensional input data onto a low-dimensional (typically 2D) topological grid while preserving neighborhood relationships in feature space. This topology-preserving property enables visualization of multivariate proxy relationships through component planes, cluster analysis, and spatial back-projection, making SOM particularly suitable for subsurface characterization where physical interpretability is essential.

While SOM can identify patterns, quantitative confirmation that these patterns correspond to meaningful distinctions between EGS, and conventional geothermal systems requires independent statistical validation using frameworks designed for heteroscedastic, non-Gaussian geoscientific data.

This study addresses these gaps by presenting an integrated methodology that combines SOM-based multivariate clustering with depth-resolved dual statistical testing to characterize EGS and hydrothermal systems in a fully 3D framework. We focus on two complementary questions: (1) Can SOM identify coherent multivariate subsurface patterns that correspond to geothermal reservoir conditions? and (2) Do these patterns exhibit statistically significant and physically meaningful differences between EGS and conventional hydrothermal systems when validated through rigorous hypothesis testing?

To answer these questions, we developed comprehensive 3D block models of proxies namely, seismic velocity attributes (Vp, Vs, Vp/Vs), temperature, magnetic gravity, fault density, and geology for Utah FORGE (EGS reference site) and RHS (conventional geothermal system). Statistical analyses were conducted using depth-binned (500 m intervals) parametric and nonparametric tests to investigate the separability of SOM clusters at the UTAH and RHS sites. The presented approach quantifies both central-tendency differences and distributional dominance while accounting for the heteroscedastic, non-Gaussian nature of subsurface geothermal data.

2. METHODOLOGY

This study integrates multimodal proxy data from Utah FORGE and RHS into a unified 3D block model comprising 4.5 million blocks at 100 m resolution. Eleven attributes were incorporated: P-wave velocity, S-wave velocity, Vp/Vs ratio, density derived from gravity inversion, temperature, fault density, and five binary lithology indicators. We first incorporated a 3×3 SOM architecture trained on representative block samples and then applied it to the full 4.5 million blocks, producing nine volumetric clusters treated directly as SOM clusters. Conventional hydrothermal and EGS footprints were used to label blocks intersecting known reservoir domains, enabling a quantitative assessment of which multivariate patterns correspond to productive geothermal conditions. Statistical validation employed a dual parametric (Welch's t-test, Hedges' g) and non-parametric (Mann–Whitney U, Cliff's δ) framework applied within 500 m depth bins from surface to 7500 m, with Holm–Bonferroni correction controlling family-wise error rate.

Utah FORGE (Milford, Utah) serves as the EGS reference where permeability is engineered through stimulation in Precambrian crystalline basement at 2000–3000 m depth (Moore et al., 2019). RHS, 15 km north, represents a hydrothermal system with 40+ years of production, characterized by natural convection, fault-controlled permeability, and sustained high-temperature fluid flow Allis et al (2015). Although both sites share Basin and Range extensional tectonics, their reservoir characteristics and depth-dependent thermal behaviors differ fundamentally.

We approach the distinction between these systems as a 3D multivariate pattern recognition problem. We integrate geological, geophysical, structural, and thermal attributes into a unified block domain and employ SOM to identify subsurface patterns based on multivariate similarity in standardized feature space. This data-driven approach enables the discovery of proxy relationships and system signatures that may not be apparent through traditional interpretation methods.

2.1 Data Integration and 3D Block Model Construction

All datasets were combined into a common 3D Cartesian block grid with 100 × 100 × 100 m spacing in UTM Zone 12N (EPSG:26912), producing a unified volumetric representation comprising more than 4.5 million blocks. This resolution was selected to match all models while ensuring computational manageability and compatibility with subsequent analyses. The 3D block model was constructed in Leapfrog Geo by integrating multiple data sources (Table 1).

Table 1. Integrated attributes, data sources, and preprocessing for 3D block model.

Attribute	Data Source and Characteristics	Preprocessing
Lithology (5 binary indicators)	Mud logs, cuttings descriptions, and well logs from Utah FORGE wells (16A(78)-32, 78B-32) and RHS wells (OH-1, OH-4, OH-5, OH-7). U.S. DOE Geothermal Data Repository (2018-2024).	Derived from 3D geological framework model. Encoded as binary indicators (0/1) for alluvium, basement granitoid, basin-fill, granite, and volcanics. Retained in {0,1} form during standardization.

Seismic Velocity (Vp, Vs, Vp/Vs)	3D tomographic velocity model from DAS and borehole geophone data (2024 stimulation), complemented by composite seismic model (Gritto and Jarpe, 2025; Finger et al., 2024)	Resampled to block grid. Standardized using z-score normalization (μ=0, σ=1). No spatial smoothing applied.
Temperature	3D model using wellbore measurements and depth contours (200 m, 1-4 km intervals). RHS constrained by regional gradients (Utah FORGE Data Repository, 2019; (Allis et al., 2015).	Interpolated in Leapfrog Geo. Values capped at 99th percentile to mitigate extrapolation artifacts. Z-score normalized.
Fault Density	Compiled fault maps (Opal Mound Fault, Negro Mag Fault) from Utah Geological Survey and DOE frameworks, 2018).	Calculated as fault intersections within 500 m spherical radius per block. Z-score normalized.
Magnetic Gravity (Density)	Gravity-based geophysical models resolving basement heterogeneity (Utah FORGE 3D Gravity Data, 2019).	Volumetric density estimates harmonized to grid. Z-score normalized.

2.2 SOM Architecture

SOM was implemented as a 3 × 3 two-dimensional neuron lattices following the classical Kohonen formulation to balance geological interpretability and pattern resolution (Kohonen, 2001). Preliminary testing with larger grids resulted in fragmentation of geologically coherent volumetric patterns, whereas smaller grids led to excessive generalization of distinct thermal–structural regimes. While the SOM topology itself is two-dimensional, it was applied to classify a fully three-dimensional subsurface block domain. Each block is represented by an 11-dimensional standardized feature vector, and SOM learning operates in this high-dimensional feature space rather than physical coordinates (Table 1). Competitive learning iteratively adjusted neuron prototype vectors so that blocks with similar multivariate signatures were mapped to the same or neighboring neurons, preserving topological relationships in feature space. After convergence, the trained SOM was applied to the entire 3D block volume, assigning each block to its best-matching neuron (Fig 1). Importantly, SOM training was conducted without knowledge of geothermal footprint locations clustering proceeded solely on multivariate proxy similarity in standardized feature space. Footprints were introduced only after cluster assignment for system-level statistical comparison. This workflow ensures that observed patterns reflect intrinsic data structure rather than label-induced artifacts.

In this study, the nine SOM neurons were treated directly as volumetric patterns (Clusters C1–C9). To summarize the trained map, the nine neurons were treated as clusters following standard SOM clustering practice (Vesanto and Alhoniemi, 2000). The resulting cluster assignments were back-projected into physical space, producing spatially continuous three-dimensional data points. Cluster-wise summaries of key attributes (temperature, Vp/Vs ratio, fault density, and density inferred from gravity inversion) were computed from block memberships to support component-plane interpretation and depth-resolved statistical analysis.

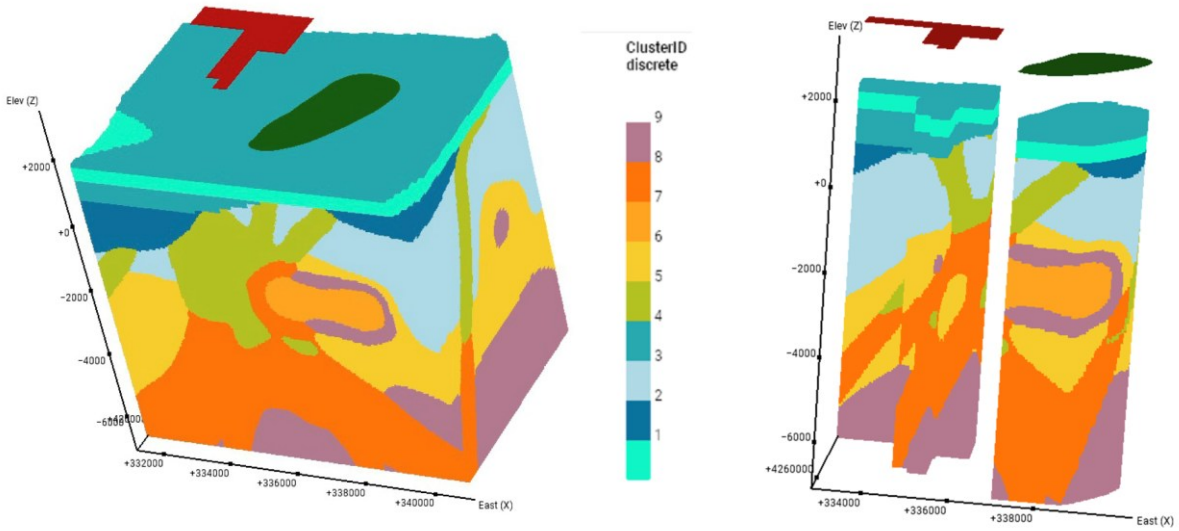

Figure 1: 3D block model showing 9 SOM clusters and study areas, at right (red colored area is FORGE and green colored area is RHS), 3D visualization of nine SOM clusters showing spatial continuity and vertical persistence at left.

To enable statistically grounded system comparisons, blocks were labeled as FORGE or RHS based on spatial intersection with geothermal system footprints representing known reservoir domains. Importantly, this labeling procedure did not modify or override the SOM-derived cluster assignments; it solely identified which pre-existing clusters spatially intersect the mapped geothermal systems. To assess whether footprint-based labeling introduces systematic sampling bias within clusters, we quantified the difference between (i) the mean standardized attribute values computed from all blocks within a given cluster and (ii) the mean values computed from the footprint-labeled subset of that cluster:

$$\Delta\bar{x}_k = \bar{x}_{k,\text{labeled}} - \bar{x}_{k,\text{all}} \quad (1)$$

where k indexes the 11 input attributes. Near-zero values of $\Delta\bar{x}_k$ indicate that the footprint-labeled blocks are representative of the parent cluster, supporting unbiased system-level inference (Table 2).

Table 2. Mean values of four key geoscientific attributes for each SOM cluster computed from the full 3D block model. Values are reported in physical units.

Cluster	Block count	Vp/Vs	Fault Density	Temperature (°C)	Magnetic Gravity Density (g/cm³)
C1	180,355	2.14	37.35	53.64	2.50
C2	197,793	1.72	16.58	67.45	2.43
C3	643,993	1.71	4.46	111.13	2.63
C4	310,272	1.78	29.22	45.79	2.46
C5	422,004	1.71	80.73	136.79	2.64
C6	814,516	1.71	5.73	231.92	2.65
C7	194,269	1.71	29.24	337.49	2.64
C8	634,073	1.71	71.09	302.42	2.67
C9	804,757	1.71	3.04	337.39	2.68

2.3 Depth-Resolved Statistical Analysis

Geothermal systems exhibit strong vertical gradients in temperature, stress, lithology, and fluid circulation regime. Statistical comparison between EGS and RHS must therefore account for depth-dependent variability while accommodating the heteroscedastic, non-Gaussian distributions typical of subsurface geophysical data. We implemented a statistical analysis that combines parametric and nonparametric approaches, applied independently within 500 m depth bins from the surface to 7500 m.

Block-based attributes (temperature, Vp/Vs, density, fault density) were first assigned to geothermal system domains using spatial footprint intersection. Within each depth bin d, we conducted parallel parametric and non-parametric tests to evaluate the null hypothesis that the two systems exhibit equivalent distributions. For parametric analysis, Welch's two-sample t-test (Welch, 1947) was employed to compare means while allowing unequal variances and sample sizes. Welch's t-test is preferred under unequal variances and sample sizes (Ruxton, 2006; Delacre et al., 2017)**.** The null hypothesis for each depth bin was:

$$H0: \mu_{FORGE,d} = \mu_{RHS,d} \quad (2)$$

To quantify effect magnitude beyond statistical significance, we computed Hedges' g (Hedges, 1981), a bias-corrected standardized mean difference appropriate for unequal sample sizes:

$$g = J \cdot (\bar{x}_{RHS} - \bar{x}_{FORGE}) / sp \quad (3)$$

where sp is the pooled standard deviation and J is a small-sample correction factor. Effect sizes were interpreted using conventional thresholds: $|g| < 0.2$ (negligible), $0.2 \le |g| < 0.5$ (small), $0.5 \le |g| < 0.8$ (moderate), $|g| \ge 0.8$ (large).

For non-parametric validation, we applied the Mann–Whitney U test (Mann and Whitney, 1947)**.** We report Cliff's δ as the corresponding non-parametric effect size (Cliff, 1993)**.**The null hypothesis was reformulated as:

$$H0: P(X > Y) = 0.5 \quad (4)$$

where X and Y represent randomly drawn values from both areas' distributions. We used Cliff's δ (Cliff, 1993), as the corresponding non-parametric effect size metric:

$$\delta = P(X > Y) - P(Y > X) = 2U / (n1n2) - 1 \quad (5)$$

Cliff's δ ranges from −1 (complete FORGE dominance) to +1 (complete RHS dominance), with interpretation thresholds: $|\delta| < 0.147$ (negligible), $0.147 \leq |\delta| < 0.33$ (small), $0.33 \leq |\delta| < 0.474$ (medium), $|\delta| \geq 0.474$ (large). Given repeated testing across depth bins, all p-values were adjusted using Holm–Bonferroni correction to control family-wise error rate at $\alpha = 0.05$. Because rank-based comparisons address non-Gaussian and ordinal-like behavior, they are widely recommended when parametric assumptions are questionable by Romano et al., 2006.

3. RESULTS

3.1 Analysis of SOM Patterns

The SOM analysis resolved the 11-dimensional feature space into nine volumetrically coherent subsurface patterns (Clusters C1–C9) exhibiting strong lateral continuity and vertical persistence (Fig. 1). These clusters form spatially continuous bodies rather than scattered artifacts, indicating that the SOM effectively captures geologically meaningful multivariate structure. Transitions between clusters coincide with coupled proxy changes; for example, the shift from C3 to C7 is marked by an increase in density from ~2.63 to 2.64 g cm^{-3}, a rise in mean temperature from 111.13 °C to 337.49 °C, and an increase in fault density from ~4.5 to ~29 (Table 2). Component plane analysis indicates that temperature, Vp/Vs ratio, fault density, and gravity-inferred density dominate cluster separation, reflecting physically consistent coupling between thermal state, elastic properties, and structural controls.

Cluster C7 emerges as the most diagnostic geothermal regime within the RHS footprint (Fig. 2). Across the full 3D block model, C7 exhibits a mean temperature of 337.49 °C, with Vp/Vs = 1.71, gravity-inferred density ≈ 2.64 g cm^{-3}, and moderate fault density (~29) (Table 2). Within the footprint-labeled RHS subset, the mean temperature increases to 421.65 °C (Table 3), indicating preferential intersection of the hydrothermal upflow core with the hottest portions of the SOM-defined cluster. The depth distribution of C7 (Fig. 2) demonstrates vertical persistence of this multivariate signature across the principal RHS reservoir interval. Footprint-based labeling introduces negligible within-cluster bias (Δmean ≈ 0), confirming that the labeled RHS blocks are representative of the underlying SOM prototype.

In contrast, clusters C1, C2, and C4 represent cooler background patterns with limited geothermal footprint overlap. These clusters exhibit mean temperatures below 150 °C across the full model (Table 2), with C1 distinguished by anomalously high Vp/Vs (> 2.1), consistent with low-velocity sedimentary or altered units, and C2–C4 characterized by more moderate Vp/Vs values (~1.7–1.8) indicative of mechanically intact basement or basin-fill sequences. Clusters C6–C9 show elevated temperatures and densities in both systems (Table 3), while clusters with high fault density (notably C5 and C8) delineate structurally controlled permeability zones, within which RHS consistently exhibits higher thermal signatures than FORGE. Cluster C7 does not spatially intersect the FORGE geothermal footprint and is therefore not represented for FORGE-labeled blocks in Table 3.

Table 3. Mean values of key proxies of SOM clusters for FORGE and RHS.

	FORGE					RHS				
Cluster	**Blocks**	**Temperature (°C)**	**Fault Density**	**Magnetic Gravity Density (g/cm³)**	**Vp/Vs**	**Blocks**	**Temperature (°C)**	**Fault Density**	**Magnetic Gravity Density (g/cm³)**	**Vp/Vs**
C1	14,158	43.54	41.16	2.42	2.12	16,696	96.23	34.60	2.54	2.15
C2	4,056	85.12	2.74	2.46	1.73	4,937	81.27	26.04	2.47	1.75
C3	70,844	98.92	4.66	2.63	1.71	47,056	133.69	4.64	2.63	1.71
C4	34,289	46.25	31.45	2.45	1.79	19,148	73.38	35.31	2.56	1.80
C5	38,591	154.10	74.60	2.63	1.71	22,514	171.02	66.52	2.63	1.71
C6	63,913	240.93	11.52	2.65	1.71	62,674	226.30	7.34	2.65	1.71
C7	—	—	—	—	—	51,942	421.65	30.69	2.64	1.71

FORGE						RHS				
Cluster	**Blocks**	**Temperature (°C)**	**Fault Density**	**Magnetic Gravity Density (g/cm³)**	**Vp/Vs**	**Blocks**	**Temperature (°C)**	**Fault Density**	**Magnetic Gravity Density (g/cm³)**	**Vp/Vs**
C8	82,324	301.57	67.76	2.66	1.71	53,160	305.12	68.25	2.67	1.71
C9	47,663	319.52	5.28	2.68	1.71	94,859	341.07	2.72	2.68	1.71

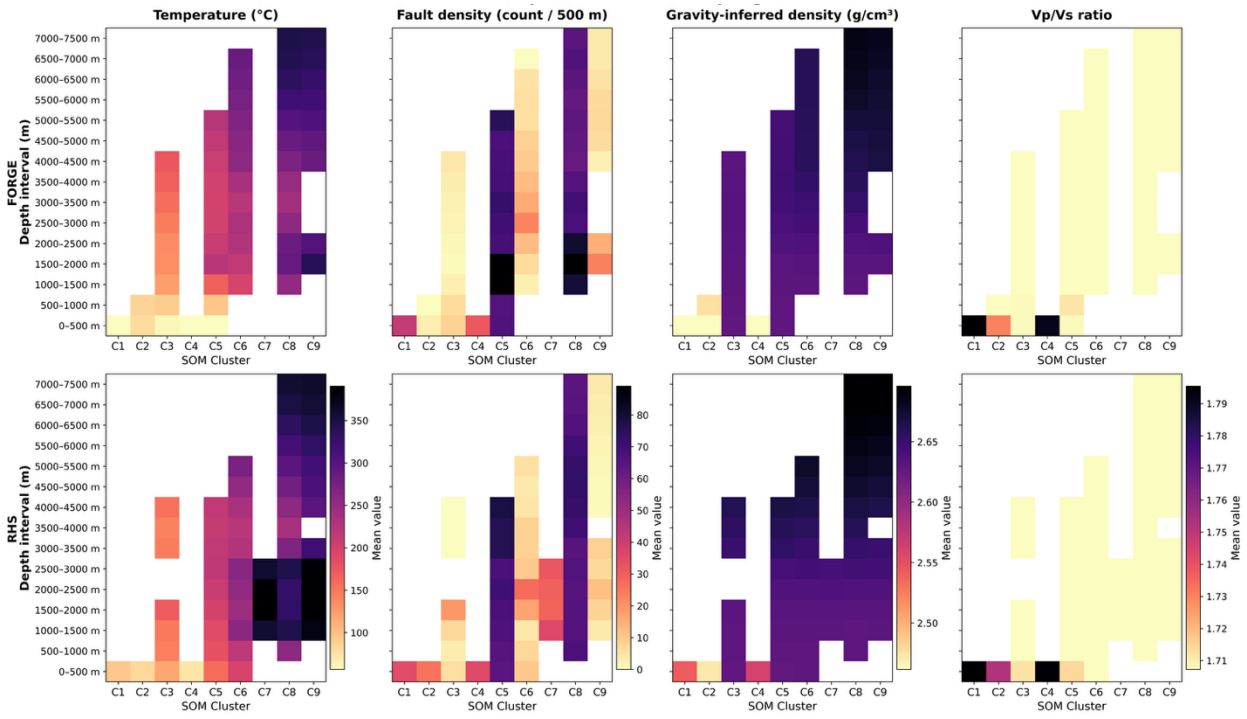

Figure 2. Depth-resolved heatmap of mean proxy values across SOM clusters (C1–C9) for the Forge and RHS illustrating vertical cluster persistence and system-level thermal contrasts.

System-wise analysis reveals pronounced asymmetry between RHS and FORGE. RHS blocks preferentially populate high-temperature clusters at shallow depths (notably C5 and C7 above ~2500 m), reflecting active convective upflow facilitated by natural fault-controlled permeability. In contrast, FORGE blocks increasingly occupy deeper portions of clusters C7–C9, where elevated temperatures persist at depth but are associated with comparatively uniform Vp/Vs ratios and subdued fault-density signatures. This pattern is consistent with a conduction-dominated thermal regime in mechanically competent basement rocks, characteristic of EGS targets where permeability must be engineered rather than naturally sustained.

Depth-resolved cluster statistics further emphasize this distinction. At shallow depths (0–2000 m), RHS exhibits systematically higher temperatures than FORGE, consistent with vigorous hydrothermal convection. Below ~3000 m, this contrast diminishes or locally reverses, with FORGE clusters showing comparable or higher temperatures, indicating efficient conductive heat retention at depth. The persistence of high-temperature clusters in FORGE at greater depths is largely absent or weakly developed in RHS, highlighting a fundamental structural and thermal divergence between EGS and conventional systems. Importantly, this differentiation emerges directly from SOM clustering, demonstrating that patterns alone encode system-specific behavior prior to any formal statistical testing. To evaluate whether footprint-based system labeling introduces systematic intra-cluster bias, we quantified Δmean statistics between footprint-labeled blocks and their parent SOM clusters (Table 3). Across all clusters, Δmean values remain close to zero, indicating that spatial intersection with known geothermal footprints does not distort the intrinsic SOM clusters. C7 exhibits near-zero Δmean across all attributes (−0.05 to +0.03 in standardized units), confirming that this cluster represents a robust and internally consistent geothermal regime rather than a footprint-induced artifact.

For Clusters C3, C5, and C9, modest positive Δmean values in temperature (+0.15 to +0.25) suggest that geothermal footprints preferentially intersect warmer sub-volumes within broader SOM patterns. This behavior is physically expected, as known geothermal reservoirs are defined by elevated temperatures relative to background conditions and therefore sample the upper tail of the cluster's temperature distribution without redefining the cluster itself. Depth distribution analysis further demonstrates that the labeling strategy captures physically meaningful reservoir-scale architecture rather than shallow sampling bias. FORGE-labeled blocks are concentrated between 3000 and 5000 m depth, corresponding to the engineered EGS target interval within crystalline basement. In contrast, RHS-labeled blocks peak between 1000 and 3000 m depth, consistent with documented hydrothermal upflow zones and production intervals. The persistence of specific SOM clusters across depth (e.g., continuous occurrence of Cluster 7 from ~2000 to ~4500 m) confirms that the SOM captures depth-dependent combinations of thermal, elastic, and structural attributes rather than isolated depth slices.

3.2 Statistical Analysis of Clusters in Terms of Temperature

The temperature patterns for FORGE and RHS were evaluated using complementary parametric (Welch's *t*-test) and non-parametric (Mann–Whitney *U*) frameworks. After Holm–Bonferroni correction, statistically significant differences are observed across all 500 m depth bins ($p_{(Holm)} < 0.001$). Given the very large blocks counts per bin, interpretation focuses on effect sizes and their depth dependence rather than statistical significance alone (Table 4; Fig. 3). At greater depths, temperature estimates are increasingly influenced by interpolation due to sparse well control; accordingly, contrasts below ~5 km are interpreted in a relative, system-comparative sense rather than as absolute values.

At shallow depths (0–2500 m), RHS exhibits substantially higher median temperatures than FORGE (Fig 3b), with very large effect sizes (Hedges' g = +1.80 to +4.04; Cliff's δ = +0.82 to +0.97). These values indicate near-complete separation of temperature distributions and are consistent with vigorous convective heat transport in the RHS. Between approximately 3000 and 3500 m depth, effect sizes decrease rapidly and approach zero, defining a transitional thermal regime where median temperatures between FORGE and RHS converge (ΔMedian ≈ 0–20°C; g ≈ +0.1 to +0.6). Below ~3500 m depth, effect sizes reverse sign, with FORGE exhibiting slightly higher median temperatures than RHS (e.g., g ≈ −0.6; δ ≈ −0.3). This reversal is compatible with a transition toward conduction-dominated thermal structure at depth and reflects FORGE's targeting of deeper, EGS conditions relative to the shallower hydrothermal circulation characteristic of RHS.

Mean-based contrasts derived from Welch's t-test (ΔMean) show the same depth intervals of divergence and convergence as median-based contrasts (Fig 3a). The strong correlation between Hedges' g and Cliff's δ (r = 0.96) demonstrates that the observed depth trends are robust and not artifacts of distributional assumptions. Collectively, these results document a systematic thermal divergence between shallow convective hydrothermal behavior at RHS and deeper conductive patterns relevant to FORGE-style EGS systems.

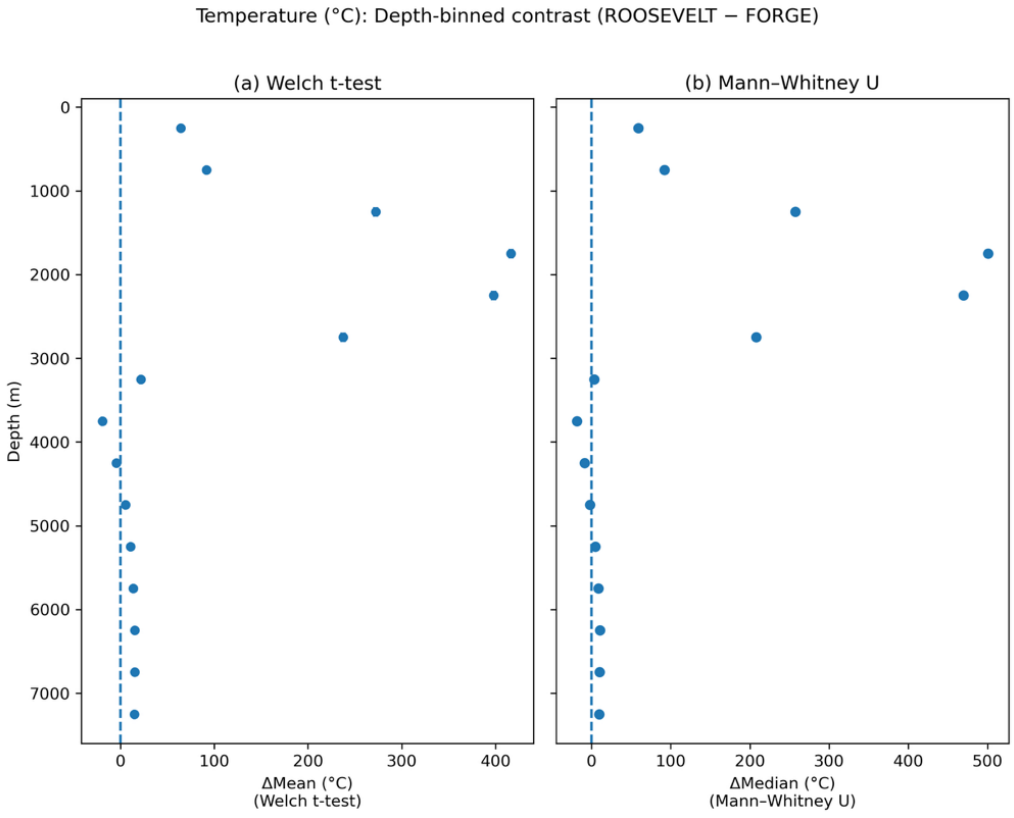

Figure 3. Depth-resolved temperature contrasts between FORGE and RHS derived from a, parametric (ΔMean; Welch's t) and b, non-parametric (ΔMedian; Mann–Whitney U) analyses, showing consistent depth-dependent divergence and convergence patterns.

Table 4. Depth-binned median temperature contrasts between FORGE and RHS.

Depth bin (m)	Median T FORGE (°C)	Median T RHS (°C)	ΔMedian (°C)	Hedges' g	Cliff's δ
0–500	52.21	111.41	59.20	1.80	0.82
500–1000	91.74	183.73	92.00	2.76	0.93
1000–1500	136.17	341.22	205.04	3.19	0.97
1500–2000	194.23	421.65	227.42	3.00	0.94
2000–2500	225.38	421.65	196.27	2.73	0.92
2500–3000	230.21	362.09	131.88	2.34	0.91
3000–3500	225.95	229.56	3.62	0.55	0.24
3500–4000	238.87	220.57	−18.30	−0.58	−0.32
4000–4500	265.52	256.71	−8.81	−0.12	−0.02
4500–5000	287.67	285.93	−1.74	0.19	0.11
5000–5500	305.03	309.76	4.73	0.47	0.22
5500–6000	320.39	329.14	8.75	0.73	0.34
6000–6500	333.67	344.27	10.60	0.93	0.45
6500–7000	344.88	355.40	10.52	1.10	0.55
7000–7500	351.21	360.87	9.66	1.26	0.66

3.3 Statistical Analysis of Clusters in Terms of other Proxies

Extension of the statistical framework to proxies reveals proxy depth dependencies (Fig 4). Temperature provides the strongest shallow-depth discrimination between EGS and hydrothermal systems ($|g| > 2.0$ at 0–2500 m); however, it weakens substantially below ~3000 m as both systems transition toward conduction-dominated patterns.

In contrast, Vp/Vs ratio exhibits moderate yet persistent effect sizes ($g \approx 0.4–0.6$) across the full depth range, reflecting sustained differences in fluid content and fracture architecture between engineered permeability at FORGE and natural fracture networks at RHS. Magnetic gravity contrasts are generally smaller ($|g| < 0.4$) but show localized enhancement at 1500–2500 m and 4000–4500 m depth intervals, coinciding with major lithological transitions from basin fill to basement granitoid and with heterogeneity within basement units. Fault density displays the most complex depth dependence. Shallow intervals (0–1500 m) show higher fault proximity at RHS, consistent with surface-connected fault zones that channel hydrothermal fluids. At greater depths (>3500 m), this pattern weakens or reverses as stimulation-induced fracture networks at FORGE locally exceed natural fault densities. Together, these results demonstrate that no single proxy adequately characterizes geothermal potential across all depths; rather, discriminatory power shifts with depth as different physical processes dominate. 100 m blocks are spatially autocorrelated and therefore do not represent strictly independent samples. Given the very large number of blocks analyzed, statistical significance is expected even for small contrasts; accordingly, interpretation emphasizes effect sizes rather than p-values alone to assess physical relevance. Accordingly, future applications will explore effective sample size estimation based on spatial decorrelation lengths and subsampling strategies to further assess robustness.

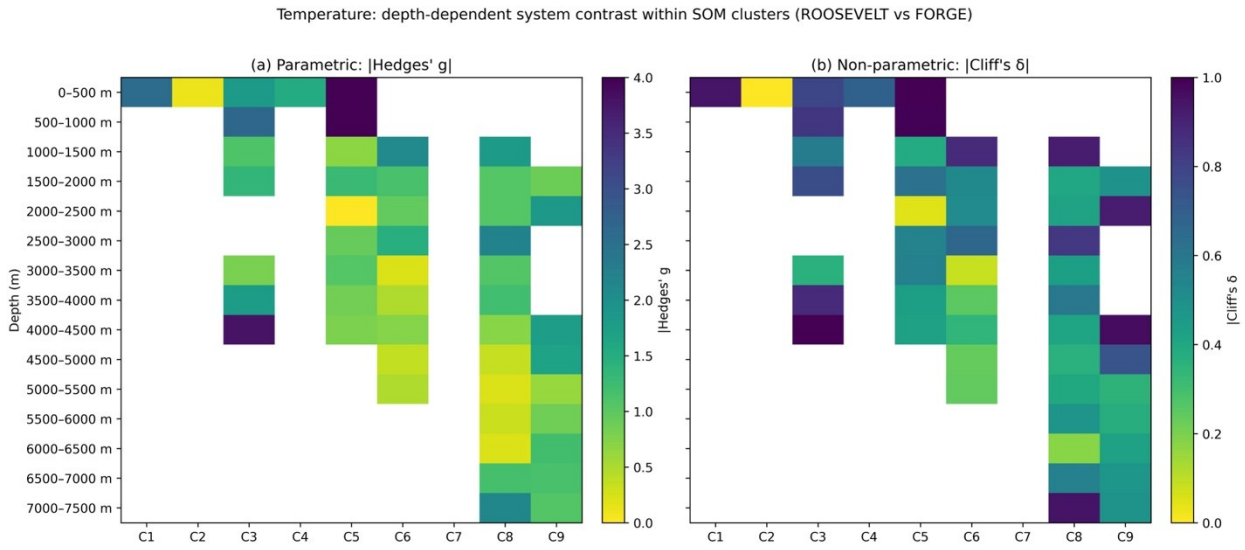

Figure 4. Heatmap of effect sizes (Hedges' g and Cliff's δ) for all proxies as a function of depth (m).

4. CONCLUSIONS

This study demonstrates that integrating SOM with depth-resolved statistical analysis provides a rigorous, explainable framework for distinguishing EGS and conventional systems in 3D subsurface volumes. The principal findings and their implications are:

The 3×3 SOM with 11 attributes produced nine spatially continuous volumetric clusters that correspond to physically meaningful combinations of thermal, elastic, structural, and lithological properties. Cluster boundaries coincide with multivariate proxy transitions rather than single-attribute thresholds, demonstrating that unsupervised learning can capture complex subsurface heterogeneity without imposing predetermined classification rules. Component-plane analysis revealed systematic correlations among temperature, Vp/Vs ratio, and density that confirm physical coupling among thermal anomalies, fluid saturation, and permeability enhancement.

Spatial intersection of known geothermal footprints with SOM clusters revealed distinctive patterns: conventional geothermal conditions occupy only 6 of 9 clusters, with Cluster 7 representing the dominant high-temperature signature within the RHS footprint. This cluster exhibits mean temperatures of 337.49°C across all blocks and 421.65°C within the RHS geothermal domain, a Vp/Vs ratio of ~1.71, a density of ~2.64 g/cm³, and moderate fault density (~29). Only the RHS blocks of C7 highlight fundamental thermal and structural contrasts between shallow hydrothermal convection and deep EGS conduction patterns. Clusters lacking geothermal labels (C1, C2, C4) exhibit substantially cooler temperatures (<150°C), with C1 displaying anomalously high Vp/Vs ratios (>2.1) indicative of low-velocity sedimentary cover or altered zones, while background basement clusters show greater fault distances and lower thermal gradients, consistent with non-geothermal conditions.

Parametric and non-parametric statistical analyses revealed pronounced depth variability in EGS vs. hydrothermal discrimination. At shallow depths (0–2500 m), RHS exhibits very large thermal effect sizes (Hedges' $g > 2.0$, Cliff's $\delta > 0.7$) reflecting strong convective upflow and near-surface thermal manifestations. Between 3000–4000 m, effect sizes decrease rapidly and reverse sign, as FORGE's deep EGS targeting strategy produces temperatures comparable to or higher than those of RHS, where convective heat transport weakens with depth.

Temperature provides maximum discrimination at shallow depths but weakens below 3000 m where both systems transition toward conduction-dominated patterns. Vp/Vs ratio maintains moderate effect sizes ($g = 0.4–0.6$) throughout the depth range, reflecting persistent differences in fluid content and fracture architecture between engineered permeability at FORGE and natural fracture networks at RHS. Density contrasts are smaller ($|g| < 0.4$) but show localized enhancement at lithological transitions (1500–2500 m and 4000–4500 m), coinciding with basin-to-basement interfaces and intra-basement heterogeneity. Fault density exhibits depth-dependent reversal: shallow intervals (0–1500 m) show higher fault proximity at RHS consistent with surface-connected fault zones channeling hydrothermal fluids, while at greater depths (>3500 m) this pattern weakens or reverses as stimulation-induced fracture networks at FORGE locally exceed natural fault densities. These findings demonstrate that no single proxy adequately characterizes geothermal potential across all depths; integrated multivariate analysis is essential. Both Utah FORGE and RHS are located within the Basin and Range extensional province, enabling a controlled comparison between engineered EGS and naturally convecting hydrothermal systems. While this shared tectonic setting strengthens internal consistency, application of the framework to other geological environments will require site-specific calibration.

Unlike black-box ML approaches, the presented approach provides interpretability through component plane visualization, cluster attribute profiling, and 3D spatial back-projection. Statistical validation using effect sizes emphasizes physical relevance over p-values, avoiding overinterpretation in large-sample settings. The framework is fully reproducible: SOM training parameters, clustering criteria, and statistical test specifications are explicitly documented, enabling application to other geothermal sites and facilitating cross-site transferability assessment.

REFERENCES

Allis, R., Moore, J., McClure, M., Podgorney, R., Simmons, S., and Brown, D.: Insights on the formation of the Roosevelt Hot Springs geothermal system from integrated geologic and geophysical studies, *Geothermics*, 54, (2015), 39–50.

Ankut, Z., and Duzgun, S.: Integrated remote sensing and machine learning for copper exploration in the Copper Triangle, Arizona, in Anderson, E.D., and Graham, G.E. (Eds.), *Mineral Resources for Our Ever-Changing World*, Proceedings, 18th SGA Biennial Meeting, Golden, Colorado, USA, Vol. 3, (2025), 1204–1204.

Cliff, N.: Dominance statistics: Ordinal analyses to answer ordinal questions, *Psychological Bulletin*, 114(3), (1993), 494–509.

Delacre, M., Lakens, D., and Leys, C.: Why psychologists should by default use Welch's t-test instead of Student's t-test, *International Review of Social Psychology*, 30, (2017), 92–101.

Finger, C., Niemz, P., Ermert, L., and Lanza, F.: Utah FORGE: Composite 3D seismic velocity model, Dataset, U.S. Department of Energy Geothermal Data Repository, (2024), https://doi.org/10.15121/2305384.

Gritto, R., and Jarpe, S.P.: Utah FORGE: Seismic DAS and geophone borehole data processing and 3D imaging of the Vp/Vs ratio in the 2024 stimulated reservoir, Dataset, U.S. Department of Energy Geothermal Data Repository, (2025), https://doi.org/10.15121/2541598.

Hedges, L.V.: Distribution theory for Glass's estimator of effect size and related estimators, *Journal of Educational Statistics*, 6(2), (1981), 107–128.

Kohonen, T.: Self-organizing maps, 3rd ed., Springer, (2001).

Mann, H.B., and Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other, *Annals of Mathematical Statistics*, 18(1), (1947), 50–60.

Moore, J., McLennan, J., Pankow, K., Wannamaker, P., McLing, T., Podgorney, R., Simmons, S., Allis, R., Gwynn, M., Jones, C., and Nash, G.: The Utah Frontier Observatory for Research in Geothermal Energy (FORGE): Phase 1 technical report, Energy & Geoscience Institute, University of Utah, (2019).

Romano, J., Kromrey, J.D., Coraggio, J., and Skowronek, J.: Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen's d for evaluating group differences on the NSSE and other surveys?, *Journal of Modern Applied Statistical Methods*, 5, (2006), 209–234.

Ruxton, G.D.: The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test, *Behavioral Ecology*, 17, (2006), 688–690.

Utah FORGE: 3D gravity data, U.S. Department of Energy Geothermal Data Repository, (2019), https://gdr.openei.org/submissions/1144.

Utah FORGE: Data for 3-D model development—Lithology, temperature, pressure, and stress, U.S. Department of Energy Geothermal Data Repository, (2019), https://gdr.openei.org/submissions/1205.

Utah FORGE: Geology map and GIS data, U.S. Department of Energy Geothermal Data Repository, (2018), https://doi.org/10.15121/1452761.

Utah FORGE: Observation well data, U.S. Department of Energy Geothermal Data Repository, (2018), https://gdr.openei.org/submissions/1008.

Utah FORGE: Well 16A(78)-32 logs, U.S. Department of Energy Geothermal Data Repository, (2024), https://gdr.openei.org/submissions/1292.

Utah FORGE: Well 78B-32 daily drilling reports and logs, U.S. Department of Energy Geothermal Data Repository, (2021), https://gdr.openei.org/submissions/1330.

Vesanto, J., and Alhoniemi, E.: Clustering of the self-organizing map, *IEEE Transactions on Neural Networks*, 11(3), (2000), 586–600.

Welch, B.L.: The generalization of Student's problem when several different population variances are involved, *Biometrika*, 34(1–2), (1947), 28–35.