

Article

Quantum-Kernel Benchmark for Isotopic Provenance Clustering in the Andes Region

Anibal Alviz-Meza ^{1,*} , Alejandro Valencia-Arias ², Félix Díaz ³  and Segundo Rojas-Flores ⁴ ¹ Escuela de Posgrado, Universidad Continental, Lima 15046, Peru² Vicerrectoría de Investigación y Posgrado, Universidad de Los Lagos, Osorno 5290000, Chile; alejandro.valenciaarias@ulagos.cl³ Facultad de Ingeniería y Arquitectura, Escuela Profesional de Ingeniería de Sistemas, Universidad Autónoma del Perú, Lima 15842, Peru; felix.diaz@autonoma.pe⁴ Institutos y Centros de Investigación, Universidad César Vallejo, Trujillo 13001, Peru; srojasf@ucv.edu.pe

* Correspondence: aalviz@continental.edu.pe

Abstract

Lead isotope ratios are frequently used in archaeometric provenance analysis; however, the overlap of isotopic fields within the Andean metallogenic belt complicates reliable provenance determination. This study presents a reproducible fidelity-based kernel method for the unsupervised clustering of Andean lead-isotope data and investigates whether a quantum-mechanical similarity space can reveal geologically significant structures beyond the classical Euclidean partition. A dataset of 1522 measurements of $^{206}\text{Pb}/^{204}\text{Pb}$, $^{207}\text{Pb}/^{204}\text{Pb}$, and $^{208}\text{Pb}/^{204}\text{Pb}$ was analyzed using a fidelity-based quantum kernel based on a three-qubit Pauli feature map and compared with classical K-means clustering, Gaussian mixture models, and Ward's agglomerative clustering under various preprocessing strategies and cluster counts. The optimal quantum kernel setup achieved the highest silhouette score at $k = 2$. However, because analytical uncertainties were not consistently reported across all the compiled sources, an uncertainty-weighted similarity could not be applied. Geological insights indicate that this binary division separates less radiogenic, arc-related compositions from more radiogenic and thorogenic crustal signatures, a contrast that broadly follows the west-to-east crustal-contamination gradient across the Andes. Conversely, the traditional four-cluster approach provides more detailed subdivisions that align with the previously identified isotopic provinces. The reported separation reflects the geometry of the quantum feature space rather than any hardware-level speed-up, as this work represents only a simulation approach. Overall, these findings support a hierarchical and complementary approach to analyzing Pb isotope origins, in which quantum kernel clustering provides robust large-scale separation and classical clustering enhances regional understanding.



Academic Editors: Fahad Ahmad and Vincenzo Tamma

Received: 8 May 2026

Revised: 20 June 2026

Accepted: 22 June 2026

Published: 27 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: Andes; lead isotopes; archaeometric provenance; unsupervised clustering; quantum kernel methods

1. Introduction

Pb isotope ratios are among the most widely used geochemical fingerprinting tools in archaeometry and economic geology, providing a robust link between ores, artifacts, and crustal materials and their geological origins [1,2]. In the Andean metallogenic belt—which extends across Peru, Bolivia, Chile, Argentina, and Ecuador—varied isotopic signatures from tectonic events, diverse basement regions, and layered magmatic

cycles create overlapping isotopic fields, making provenance determination more challenging [3,4]. Conventional methods using bivariate discrimination plots and isotopic province frameworks [2,5] have established a foundational classification for the region; yet they rely on predefined geological boundaries and are limited in their capacity to objectively handle the multivariate structure of large, heterogeneous isotope compilations.

Unsupervised machine learning clustering has been increasingly adopted in geochemical studies to partition isotope datasets in a data-driven manner, without requiring prior labeled information [6]. Algorithms such as *k*-means, Gaussian mixture models, and hierarchical agglomerative clustering have been applied to delineate geochemical domains, group ore deposits, and explore regional-scale isotopic variability in settings ranging from till geochemistry surveys [7] to Pb isotope databases [8]. A recent study applied *k*-means clustering to a compiled Andean Pb-isotope database and demonstrated a meaningful correspondence between algorithmic partitions and previously established isotopic provinces [8]. However, since these methods rely on Euclidean geometry, they may not effectively capture the nonlinear similarity structure in multivariate isotope-ratio space, especially when isotopic overlap indicates ongoing geological mixing rather than distinct groups.

Quantum machine learning (QML) is a rapidly developing field that uses principles of quantum mechanics, such as superposition and entanglement, to improve or rethink machine learning algorithms [9,10]. The current Noisy Intermediate-Scale Quantum (NISQ) era is characterized by devices operating tens to hundreds of qubits under nonnegligible gate error rates [11], making hybrid quantum–classical architectures the dominant practical paradigm—quantum processors handle computationally intensive subroutines while classical hardware manages data preprocessing, optimization, and post-processing [12]. Within this framework, QML algorithms have been explored across domains including computational chemistry, financial optimization, and earth observation [13,14], with growing interest in identifying application regimes where quantum-induced data representations offer advantages over purely classical methods.

Quantum Kernel Methods (QKMs) are a particularly manageable subset of QML algorithms within NISQ constraints, as they avoid the trainability issues of variational quantum circuits while still exploiting high-dimensional quantum feature spaces [15,16]. This method encodes classical data points into quantum states using parameterized feature-map circuits then estimates pairwise similarities based on the overlap of these states. The resulting similarity matrix can be used with standard kernel-based algorithms such as support vector machines and spectral clustering, enabling a modular workflow that separates quantum encoding from classical learning. Large-scale benchmarking studies have shown that QKMs' performance depends on design choices such as the structure of the encoding circuit, feature scaling, and circuit depth, making systematic hyperparameter tuning crucial for achieving reliable results [17,18]. The Andean metallogenic belt is a particularly demanding testbed for unsupervised and quantum-kernel clustering. It combines a monotonic west-to-east crustal-contamination gradient with overlapping isotopic fields and publicly available Pb-isotope compilations, so it offers both a clear geological expectation and a genuinely nontrivial, high-overlap similarity structure against which algorithms can be benchmarked.

Applications of QML in geoscience are still limited. Most published research focuses on remote sensing and quantum-enhanced classification of satellite images and hyperspectral data [19,20]. Moreover, a recent review concluded that near-term quantum advantage in earth observation is most plausible in resource-bounded, task-specific regimes [21]. To our knowledge, no research in isotopic geochemistry and archaeometric provenance analysis has utilized kernel-based quantum techniques for clustering. This highlights a notable

methodological gap: the potential for quantum-encoded feature spaces to reveal nonlinear similarity structures has not been explored with isotopic data, where traditional clustering methods produce understandable yet limited partitions. It is still unclear whether quantum representations of isotopic ratios can produce geologically meaningful clusters that either complement or improve upon classical Euclidean divisions.

This study addresses a research gap by, to our knowledge, being the first to compare classical clustering methods with a simulation-based quantum kernel pipeline for isotopic geochemical data. It uses a dataset of 1522 lead isotope measurements from the Andes. The source metadata and Pb-isotope data are provided in the Supplementary Materials. The quantum method encodes the three lead-isotope variables into quantum states via Pauli feature-mapping circuits in Qiskit, computes pairwise fidelities to measure similarity, and then applies classical K-means to the resulting quantum kernel matrix. This pipeline is evaluated against K-means, Gaussian mixture models, and Ward's agglomerative clustering, utilizing various preprocessing techniques and cluster counts. The primary criterion for selection is the silhouette coefficient, with additional support from the Davies–Bouldin and Calinski–Harabász indices. This study is presented as a reproducible protocol, conducted entirely via classical simulation, and offers a computational benchmark for integrating quantum kernel techniques into isotope-based origin analysis.

2. Methodology

This study used an unsupervised comparative learning method to determine whether a quantum-kernel representation, based on the fidelity of Andean lead-isotope data, produces clustering patterns that are both internally consistent and geologically meaningful. Two analytical approaches were applied in parallel: a traditional baseline using Euclidean clustering techniques and a quantum kernel process in which isotope measurements were encoded into quantum states and converted into a predetermined kernel matrix. Both approaches were tested on the same cleaned dataset, with consistent preprocessing and model selection criteria, and interpreted in relation to previously identified Andean isotopic provinces.

2.1. Pb Isotope Ratios from the Andes

A database of Andean Pb isotope measurements was compiled from published geological and archaeometric research covering Peru, Bolivia, Chile, Argentina, and Ecuador. This source included ore deposits, mineral samples, igneous rocks, crustal rocks, sediments, ashes, and related environmental materials, as summarized in Table 1. Each entry represented an individual sample with available Pb isotope ratios expressed numerically. This study focused on three variables: $^{206}\text{Pb}/^{204}\text{Pb}$, $^{207}\text{Pb}/^{204}\text{Pb}$, and $^{208}\text{Pb}/^{204}\text{Pb}$, because these were the most consistently reported across the literature, allowing for the broadest comparison. Samples without complete numerical data for these ratios were excluded from the clustering process. The dataset described here is at the source-compile stage; subsequent steps, such as harmonization, duplicate removal, outlier filtering, and data derivation, are detailed in Section 2.3.

Table 1. Number of analyses (N) and percentage of the 1522-analysis corpus contributed by each source. This summary is provided to allow assessment of data heterogeneity across studies.

Ref.	Region	Type of Samples	N Analyses	% of Corpus
[22]	Castillo de Huarney, Peru	Environmental samples	4	0.3
[1]	Peru, Bolivia, northern Chile, and northwest Argentina	Ore and mineral samples	170	11.2
[23]	Chimborazo, Ecuador	Marine cores and ashes	79	5.2
[3]	Peru, Bolivia, Chile, and Argentina	Igneous rocks	504	33.1
[24]	Cerro de Pasco, Peru	Ores and sediment samples	39	2.6
[25]	Cajamarca, Peru	Ore deposits	14	0.9
[26]	Northern domain of southern Peru and northern Chile	Whole-rock samples	48	3.2
[5]	Southwest to south-central Bolivia	Ores, metal ores, crustal rocks, and leachates of crustal rocks	34	2.2
[4]	Basement domains of the Altiplano, central Andes	Ores, volcanic rocks, and their host rocks	334	21.9
[27]	Central Andes/Peru	Ores	81	5.3
[28]	Northern Peru	Ores and crustal rocks	41	2.7
[2]	Central Andes	Ores and crustal rocks	20	1.3
[29]	Southeastern Peru	Ore deposits	31	2.0
[30]	Central and southern Peru	Sulfide ore deposits	22	1.4
[31]	Central Peruvian Andes	Ore and host-rock sample	15	1.0
[32]	Central Coast of Peru	Ore deposits	2	0.1
[33]	Galenas of the Chilean Andes	Galena ores	76	5.0
[34]	Southern Peru	Igneous rocks	8	0.5

2.2. Exploration of the Database of Pb Isotopes

As shown in Figure 1, an initial exploratory analysis was performed on the compiled Pb isotope dataset prior to clustering. This step aimed to understand the distribution of the three analytical variables, detect general correlation patterns, and serve as a reference for later cleaning and clustering processes.

Analysis of the isotope distributions in Figure 1 suggests that the $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios may be slightly skewed toward higher values, potentially due to radiogenic lead enrichment from extended uranium and thorium decay in older crustal rocks. The $^{207}\text{Pb}/^{204}\text{Pb}$ distribution appears a bit more symmetrical. All combinations exhibit a positive correlation, though dispersions differ. For instance, the correlation between $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$ forms a more tightly clustered pattern than the others, suggesting a more direct relationship or a common process impacting these two isotopes. Quantitatively, the sample skewness is 0.11 for $^{206}\text{Pb}/^{204}\text{Pb}$, 0.99 for $^{207}\text{Pb}/^{204}\text{Pb}$, and 1.56 for $^{208}\text{Pb}/^{204}\text{Pb}$, confirming the positive skew of the two thorogenic/radiogenic ratios.

Table 2's descriptive statistics reveal notable variability in the $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ isotopic ratios, likely due to greater heterogeneity among mineral sources in the dataset. The fluctuation ranges for each isotope remain quite broad even after removing outliers. Nonetheless, the median closely aligns with the mean for all three isotopes, suggesting their distributions are fairly symmetrical. Table 2 and Figure 1 refer to the cleaned (post-filtering) dataset of 1522 analyses.

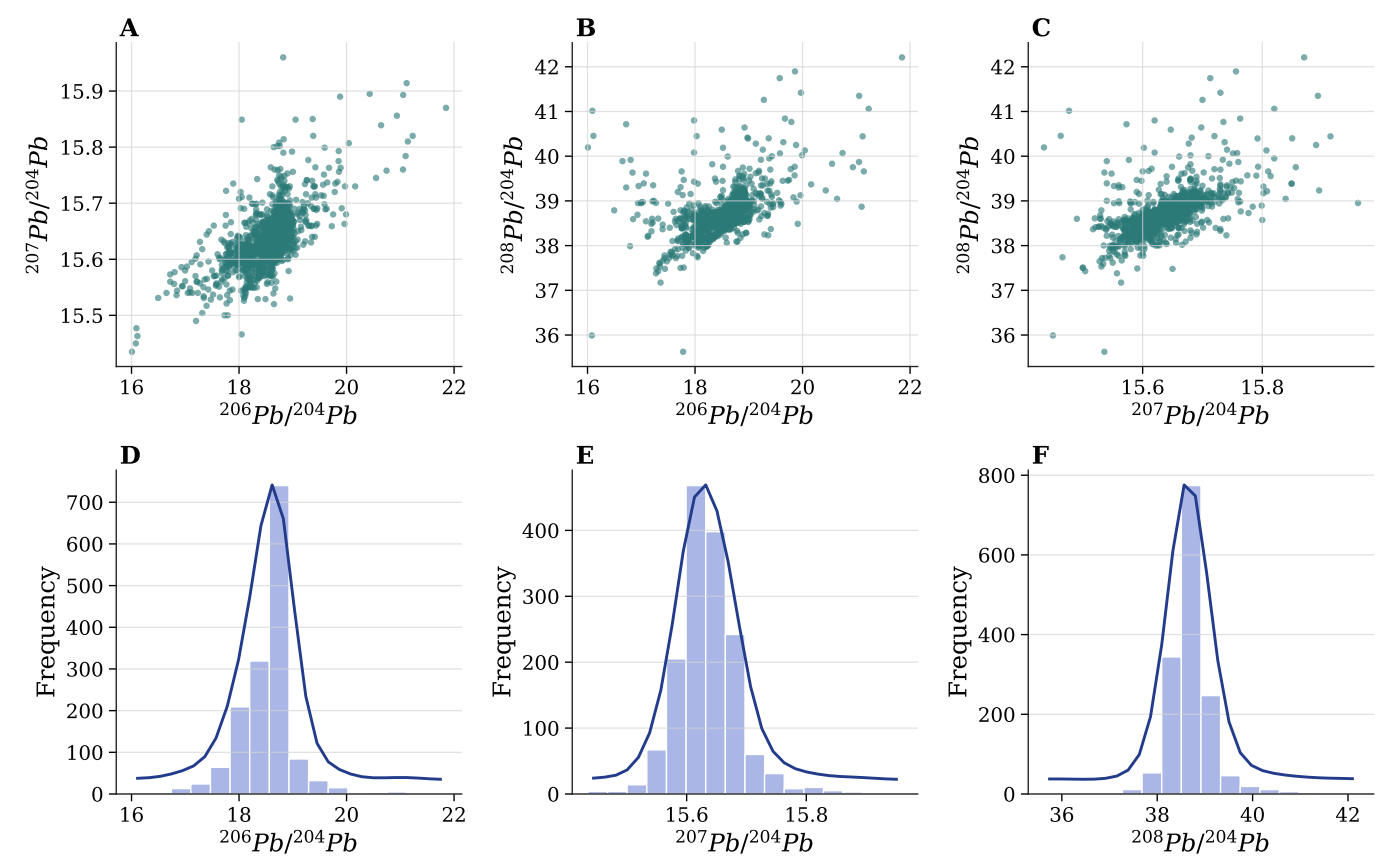


Figure 1. Exploratory structure of the Pb isotope dataset. Bivariate projections of the three original Pb isotope ratios are shown in the upper row: (A) $^{206}\text{Pb}/^{204}\text{Pb}$ versus $^{207}\text{Pb}/^{204}\text{Pb}$, (B) $^{206}\text{Pb}/^{204}\text{Pb}$ versus $^{208}\text{Pb}/^{204}\text{Pb}$, and (C) $^{207}\text{Pb}/^{204}\text{Pb}$ versus $^{208}\text{Pb}/^{204}\text{Pb}$. Teal points represent individual cleaned analyses. The lower row shows the corresponding univariate distributions: (D) $^{206}\text{Pb}/^{204}\text{Pb}$, (E) $^{207}\text{Pb}/^{204}\text{Pb}$, and (F) $^{208}\text{Pb}/^{204}\text{Pb}$; light-blue bars indicate histogram frequencies, and dark-blue curves indicate kernel-density estimates.

Table 2. Descriptive statistics of the Pb isotope data.

	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{204}\text{Pb}$
Count	1522	1522	1522
Mean	18.540	15.636	38.703
Std	0.512	0.052	0.467
Min	16.008	15.435	36.625
25%	18.305	15.604	38.484
50%	18.630	15.631	38.653
75%	18.780	15.663	38.879
Max	21.850	15.960	42.210

The variability in isotope ratios mirrors the complex geological history of the Andes, marked by multiple tectonic events, magmatic activity, and mineralization episodes [3]. Understanding these processes is crucial for accurate provenance analysis, as they influence clustering and predictive modeling.

2.3. Clustering Analysis: Inclusion, Duplicate Control, and Outlier Handling

Three isotope ratios were chosen as the main variables for the clustering analysis: $^{206}\text{Pb}/^{204}\text{Pb}$, $^{207}\text{Pb}/^{204}\text{Pb}$, and $^{208}\text{Pb}/^{204}\text{Pb}$. These ratios were selected because they are the most consistently reported Pb-isotope descriptors in the Andean literature, thereby

maximizing sample retention and ensuring comparability across studies. Coordinates (latitude and longitude) were included only for spatial visualization and interpretive mapping, not as clustering variables.

The working database was cleaned prior to unsupervised analysis. Duplicate observations were removed based on identical triples of $^{206}\text{Pb}/^{204}\text{Pb}$, $^{207}\text{Pb}/^{204}\text{Pb}$, and $^{208}\text{Pb}/^{204}\text{Pb}$ values after numerical harmonization. Floating-point values were uniformly rounded prior to duplicate detection to avoid artificial duplicate entries due to formatting concerns. A small number of extreme outliers previously identified in the exploratory analysis were excluded by direct value filtering. This procedure was carried over from the workflow of the previous manuscript to ensure continuity between classical and QML analyses. The final cleaned dataset comprised 1522 observations.

2.4. Classical Clustering Baseline

The classical branch served as the nonquantum reference framework for comparative clustering analysis. K-means was chosen as the primary classical baseline because it had previously been applied to the same Andean Pb isotope database and offered an interpretable clustering structure for regional isotopic analysis [8]. Its cluster solution demonstrated notable geological agreement with the isotopic provinces analyzed by Macfarlane's related works [1,2].

In this work, additional classical baselines were added to expand the comparison with the quantum-kernel pipeline. These included Gaussian Mixture Models, a distribution-based alternative, and Ward's agglomerative clustering, a deterministic hierarchical method. The models used as benchmarks in this study are summarized in Table 3.

Table 3. Classical benchmark models.

Model	Input Scaling	Key Parameters	Role in Study
K-means	StandardScaler	k = 2–4 n_init = 20	Primary classical baseline
K-means	MinMaxScaler(0, π)	k = 2–4 n_init = 20	Scaling-control baseline
Gaussian Mixture Model	MinMaxScaler(0, π)	k = 2–4 covariance_type = full n_init = 5	Distribution-based baseline
Agglomerative	StandardScaler	k = 2–4 linkage = Ward; Euclidean distance	Deterministic hierarchical baseline
RBF Kernel K-means	StandardScaler	k = 2–4;	Classical nonlinear-kernel control

2.4.1. Feature Scaling for the Classical Baseline

After a rigorous data cleaning phase, scaling was applied to normalize the ranges of the isotope data, preparing them for the ML algorithms (see Equation (1)). Standardization was used because the isotopic ratios have different ranges and variances; equalizing the scale reduces dominance by the variable with the widest range during Euclidean clustering.

$$z = \frac{x - \mu}{\sigma}. \quad (1)$$

where z is the standardized value, x is the original value, μ is the mean, and σ is the standard deviation.

For the benchmark design, StandardScaler served as the primary preprocessing method for the classical K-means and Ward baselines. Additionally, a second K-means baseline employed MinMaxScaler(0, π) as a control comparison for the quantum branch, since that angular scaling was also evaluated in the QML workflow.

2.4.2. K-Means Objective Function

K-means is an unsupervised clustering algorithm that partitions a dataset into k clusters by minimizing the within-cluster sum of squared distances between each observation and its assigned cluster centroid, as shown in Equation (2).

$$J = \sum_{i=1}^n \sum_{j=1}^k r_{ij} \|x_i - \mu_j\|^2. \quad (2)$$

Here, n is the total number of samples, k is the number of clusters, x_i is the i -th sample, μ_j is the centroid of cluster j , and r_{ij} is a binary assignment variable equal to 1 when sample i belongs to cluster j , and 0 otherwise. In this study, classical K-means was evaluated across $k \in \{2, 3, 4\}$ as part of the benchmark design. The 4-cluster classical solution was retained as the primary interpretive reference for comparison with the geological literature and the previous Andean Pb isotope study [8].

2.4.3. Gaussian Mixture Model Formulation

Gaussian Mixture Models (GMMs) have served as a classical baseline for distribution modeling. In this approach, the data distribution is represented as a weighted sum of k Gaussian components, as shown in Equation (3), where k is the number of mixture components, π_j is the mixing weight of component j , and $N(x|\mu_j, \Sigma_j)$ is the multivariate Gaussian density with mean vector μ_j and covariance matrix Σ_j . The mixing weights satisfy Equation (4).

$$p(x) = \sum_{j=1}^k \pi_j \mathcal{N}(x | \mu_j, \Sigma_j). \quad (3)$$

$$\sum_{j=1}^k \pi_j = 1, \quad \pi_j \geq 0 \text{ for all } j. \quad (4)$$

Model parameters were estimated through maximum likelihood using the Expectation–Maximization (EM) algorithm. In this study, GMM was implemented with full-covariance matrices and evaluated for $k \in \{2, 3, 4\}$ as part of the classical benchmark design.

2.4.4. Ward’s Agglomerative Clustering Criterion

Ward’s agglomerative clustering served as a deterministic hierarchical baseline. It starts with each observation as its own cluster and iteratively merges the two clusters that result in the smallest increase in within-cluster variance. This merging process seeks to minimize the increase in the error sum of squares, as shown in Equation (5), where A and B are two candidate clusters, n_A and n_B are their respective sizes, and μ_A and μ_B are their centroids. At each agglomeration step, the pair of clusters with the smallest $\Delta(A, B)$ is merged.

$$\Delta(A, B) = \frac{n_A n_B}{n_A + n_B} \|\mu_A - \mu_B\|^2. \quad (5)$$

Ward’s agglomerative clustering was tested for $k \in \{2, 3, 4\}$ within the classic benchmark to enable a deterministic hierarchical comparison with both K-means and the QML pipeline.

2.4.5. Classical Nonlinear-Kernel Control: RBF Kernel

To address whether the binary separation obtained by the quantum kernel was specific to the quantum feature map or could be reproduced by a conventional nonlinear similarity function, a radial basis function (RBF) kernel K-means baseline was added to the classical benchmark. This model was included as a nonlinear-kernel control because, like the

QML branch, it replaces the original Euclidean input representation with a precomputed similarity matrix. The three Pb isotope ratios were first standardized using the same standard-scaling procedure applied to the main classical baselines. Pairwise similarities between standardized observations were then computed with the RBF kernel, as shown in Equation (6):

$$K_{\text{RBF}}(z_i, z_j) = \exp\left(-\gamma \|z_i - z_j\|^2\right). \quad (6)$$

where z_i and z_j are standardized isotope-ratio vectors, and γ is the bandwidth parameter controlling the locality of the similarity function. The bandwidth was evaluated over five multipliers of the median-heuristic value: 0.1, 0.3, 1, 3, and 10. This median-heuristic value was computed as the inverse median of the nonzero pairwise squared Euclidean distances. For each bandwidth, the resulting RBF kernel matrix was clustered using K-means on the row-wise kernel representation for $k \in \{2, 3, 4\}$, following the same validation protocol used for the QML kernel benchmark. The RBF kernel K-means model therefore served as a classical nonlinear-kernel control to test whether the QML partition could be matched by a standard nonlinear similarity measure. The complete results are reported in Tables A1 and A2.

2.5. Quantum-Kernel Clustering Pipeline

The quantum branch was developed as an unsupervised clustering pipeline that uses quantum-state encoding and fidelity-based similarity. Instead of clustering samples directly in the original Euclidean feature space, each observation was transformed into a quantum state via a parameterized feature-map circuit. Pairwise similarities were then calculated using a fidelity-based quantum kernel. This branch aimed to determine if a quantum-derived representation of the Pb isotope dataset could uncover isotopic groupings that are either complementary to or more distinctly separated than those found using classical methods.

2.5.1. Input Encoding Alternatives and Scaling Strategies

The QML pipeline was tested with various preprocessing techniques on the input data to identify the encoding method that yielded the best clustering performance for the Pb isotope dataset. The tested options included using the original three isotope ratios directly, applying PCA to reduce them to two components, and three scaling methods: StandardScaler, MinMaxScaler($-\pi, \pi$), and MinMaxScaler($0, \pi$). These methods were evaluated within the same clustering procedure to assess how preprocessing influenced clustering performance. When PCA was applied, the transformed variables came from the standardized isotope matrix. For MinMax scaling, each feature x is independently scaled into an angular range, as shown in Equation (7).

$$x' = \pi \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (7)$$

For the $[0, \pi]$ transformation, x is the original value of the feature, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of that feature in the dataset, and x' is the scaled value used for quantum encoding. The $[-\pi, \pi]$ transformation followed the analogous MinMax mapping to that interval. All preprocessing alternatives were evaluated under the same downstream clustering protocol.

2.5.2. Quantum Feature Map

Various quantum feature-map configurations were examined to understand how circuit depth and encoding structure influence clustering performance. These circuits were based on the Pauli feature-map family, using Pauli operators $\{X, Y, Z\}$. The main

hyperparameters analyzed were the number of input features and the number of repetition layers, with feature dimensions $\in \{2, 3\}$ and repetitions $\in \{1, 2, 3, 4\}$. The two-feature configurations corresponded to PCA-reduced inputs, whereas the three-feature configurations corresponded to direct encoding of the original Pb isotope ratios. All circuits were assessed using the same fidelity kernel and clustering framework. Figure 2 is a schematic visualization of the circuit family employed to encode the Pb isotope variables in the final QML branch, redrawn for readability rather than presented as an experimental image.

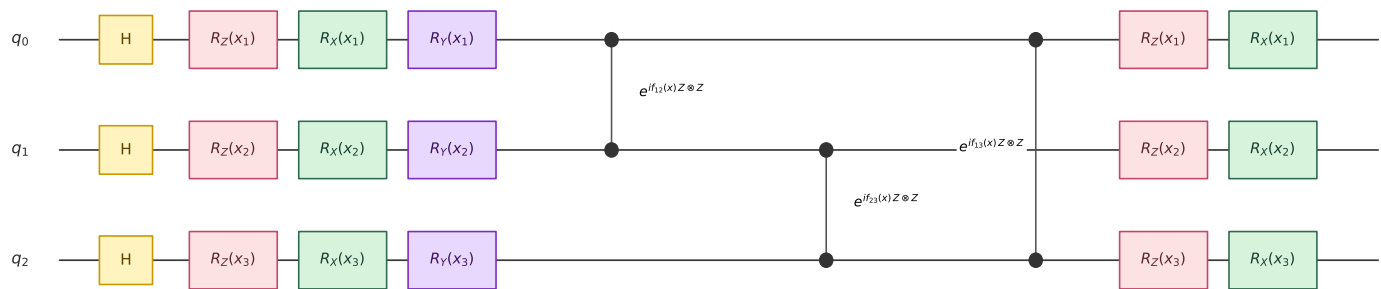


Figure 2. Schematic visualization of the Pauli feature-map circuit used in the final QML branch to encode the three scaled Pb isotope ratios into three qubits and construct the fidelity-based quantum kernel. The wires labeled q_0 – q_2 denote the three qubits, colored boxes denote the single-qubit Hadamard and rotation gates, and the black connected dots denote pairwise ZZ-entangling interactions.

The circuit shown in Figure 2 corresponds to the selected QML configuration: a Qiskit PauliFeatureMap with feature_dimension = 3, reps = 1, paulis = ['Z', 'ZZ'], and entanglement = 'full', applied after MinMaxScaler(0, π) transformation of the three original isotope ratios. In this configuration, each isotope ratio is encoded into one qubit, while the ZZ terms introduce pairwise interactions between qubits. The Hadamard gates and the rotation/entangling blocks shown in the expanded schematic correspond to the Qiskit gate-level decomposition of this feature map and should not be interpreted as additional preprocessing steps or independent encoding layers.

Among the configurations examined, the three-qubit direct-encoding map with a single repetition layer was retained for the final branch because it attained the highest silhouette score while minimizing circuit depth. Deeper circuits—two to four repetitions—and PCA-reduced two-qubit encodings did not improve the internal clustering metrics and increased the number of entangling operations, which is undesirable under NISQ noise constraints. Richer designs, such as data-reuploading maps or problem-tailored entanglement topologies, are promising alternatives but lie beyond the scope of this benchmark and are identified as future work.

2.5.3. Quantum Kernel Optimization

For two observations x_i and x_j , the fidelity-based quantum kernel is defined by Equation (8), where $\phi(x)$ is the quantum state prepared by the feature map for the input x . $K(x_i, x_j)$ measures the similarity between both observations in the induced Hilbert space. The complete kernel matrix K is therefore a symmetric matrix of size $n \times n$. Since each statevector is normalized, the diagonal entries satisfy $K(x_i, x_i) = 1$ for every observation i , whereas the off-diagonal entries $K(x_i, x_j) \neq 1$.

$$K(x_i, x_j) = |\langle \phi(x_i) | \phi(x_j) \rangle|^2. \quad (8)$$

The initial brute-force implementation evaluated this kernel by repeatedly generating two circuits and two statevectors for each sample pair. To make the benchmark compu-

tationally feasible, kernel evaluation was reformulated in a vectorized way. Each sample was converted into a statevector once, and all statevectors were stacked row-wise into the state-vector matrix S . For n observations and q qubits, S has dimensions $n \times 2^q$. The overlap matrix was then computed as $SS^\dagger$, where $S^\dagger$ denotes the conjugate transpose of S . The full kernel matrix was obtained as the element-wise squared modulus of the overlap matrix, as defined in Equation (9).

$$K = |SS^\dagger|^{\circ 2}. \quad (9)$$

Here, $|\cdot|^{\circ 2}$ denotes the element-wise squared modulus. Equivalently, each entry of the kernel matrix is given by $K_{ij} = |(SS^\dagger)_{ij}|^2$.

2.5.4. Clustering over the Quantum Kernel

For each QML configuration, the pre-computed kernel matrix was used as the similarity representation for subsequent clustering. In the main implementation, K-means was applied to the row-based kernel representation, treating each observation as a similarity profile with respect to all other samples. This strategy was chosen as a practical, reproducible comparison framework for classical reference methods. Consequently, the resulting partitions should be interpreted as clusters in the kernel similarity space, rather than direct Euclidean partitions of the original feature matrix.

2.5.5. Experimental Design, Hyperparameter Search, and Evaluation Metrics

The exploration of the QML workflow was conducted using a staged search design rather than a single exhaustive factorial grid. Initially, different input preprocessing options were evaluated. Next, comparisons were made between PCA-based and nonPCA input structures. Then, the impact of quantum circuit depth was assessed by varying the number of reps. Finally, the chosen QML pathway was benchmarked against classical baselines across multiple values of k . Table 4 outlines the exploratory search grid. Model selection was primarily based on internal clustering validation metrics, with the silhouette coefficient as the primary measure, as it balances within-cluster cohesion with between-cluster separation. Two additional metrics supported this choice: the Davies–Bouldin index, where lower values indicate better compactness and separation, and the Calinski–Harabasz index, where higher values suggest improved between-cluster discrimination.

Table 4. Exploratory search grid.

Component	Values Tested	Decision Rule
Input preprocessing	StandardScaler MinMaxScaler(0, π) MinMaxScaler($-\pi$, π)	Select the highest silhouette under the same clustering protocol
Dimensionality reduction	PCA 2 components vs. no PCA	Retain the variant with the best clustering quality
QML circuit depth	reps = 1, 2, 3, 4	Select the smallest depth with the best performance
Number of clusters	$k = 2, 3, 4$	Report metric optimum; keep interpretive alternatives when scientifically useful

The silhouette coefficient for sample i was calculated using Equation (10), where $a(i)$ represents the average dissimilarity of sample i to other points within its cluster, and $b(i)$ indicates the minimum average dissimilarity between sample i and points in all other clusters.

The Davies–Bouldin index served as a supplementary internal validation metric. It measures the average similarity between each cluster and its most similar neighboring cluster.

ter by integrating within-cluster dispersion with the separation between cluster centroids, as illustrated in Equation (11), where k is the number of clusters, S_i and S_j are the average within-cluster dispersions of clusters i and j , respectively, and M_{ij} is the distance between the centroids of clusters i and j . Lower values indicate better clustering quality.

The Calinski–Harabasz index was used as an additional variance-ratio criterion, as defined in Equation (12), where n is the total number of samples, k is the number of clusters, $\text{Tr}(B_k)$ is the trace of the between-cluster dispersion matrix, and $\text{Tr}(W_k)$ is the trace of the within-cluster dispersion matrix. Higher values indicate better defined cluster structure.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}. \quad (10)$$

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right). \quad (11)$$

$$CH = \frac{\text{Tr}(B_k)}{\text{Tr}(W_k)} \cdot \frac{n - k}{k - 1}. \quad (12)$$

Model selection focused mainly on the silhouette coefficient, with support from the Davies–Bouldin and Calinski–Harabasz indices. For stochastic clustering models, setting the random state to 42 ensured reproducibility. Deterministic methods produced the same partitions across different runs.

Stochastic baselines were additionally checked across independent random seeds to visualize initialization sensitivity. K-means and GMM runs were repeated under the same preprocessing and k values, while Ward clustering remained deterministic and the retained QML kernel partition was treated as a fixed reference.

2.6. Computational Environment

All experiments were performed on Google Colab using a CPU-based virtual machine with approximately 12.7 GB of RAM and 107.7 GB of available disk space. The work utilized Python 3.10.12. Classical preprocessing, clustering, and validation metrics were implemented with scikit-learn 1.2.2, alongside NumPy, pandas, and matplotlib for numerical analysis, data handling, and visualization. The quantum components were developed within the Qiskit ecosystem, employing parameterized feature-map circuits, statevector simulation, and custom vectorized kernel evaluation routines. No GPU acceleration or quantum hardware was used. All quantum-kernel experiments were conducted through simulation in a fully classical environment. To mitigate the stochastic effects of non-deterministic clustering models, all random runs used a seed of 42, while deterministic procedures were repeated to ensure consistency across the tables.

2.7. Limitations

The present study has several limitations that should be considered when interpreting the results. First, the database consolidates Pb isotope data from multiple Andean studies, laboratories, and analytical periods, which increases spatial and geological coverage but introduces heterogeneity associated with interlaboratory bias, mass-fractionation correction procedures, and uneven reporting of analytical uncertainty. Because uncertainties were not consistently available, they could not be propagated into the similarity calculations; therefore, cluster boundaries in regions of strong isotopic overlap should be interpreted as analytical partitions rather than sharp geological discontinuities, and an uncertainty-weighted quantum kernel was not implemented.

Second, the quantum-kernel workflow was entirely classically simulated and therefore provides a computational benchmark rather than evidence of quantum advantage.

The analysis was also restricted to the tested preprocessing schemes, feature-map families, and circuit depths, so alternative quantum encodings may lead to different partitions. In addition, the best-performing QML solution was binary ($k = 2$), capturing a coarse first-order isotopic separation rather than fine geographic or provincial resolution. Third, the clustering was unsupervised and evaluated with internal metrics of cohesion and separation; these metrics do not independently validate geological meaning. Consequently, the partitions were interpreted in relation to previously established isotopic provinces and should not be regarded as independent geographic delimitations.

Additional methodological constraints arise from the clustering strategy and scalability of the kernel formulation. The kernel-based workflow used row-wise kernel representations with K-means as a practical comparison, but this does not exhaust methods specifically designed for precomputed similarity matrices. Moreover, although the three-qubit encoding used here does not reach the exponential-concentration regime and retains discriminative power, scaling the same construction to larger feature sets and qubit counts may cause kernel values to concentrate around their mean. On real hardware, this would require a rapidly increasing number of measurement shots to resolve small kernel differences, making the bandwidth and scaling of the encoding map critical design choices. Since all results were obtained from exact state-vector simulation, shot noise and hardware errors were not included. Future work should therefore address multi-cluster quantum formulations, NISQ hardware execution, uncertainty-aware kernels, and broader comparisons with kernel K-means, spectral clustering, and additional classical kernels.

3. Results and Discussion

3.1. QML Model-Selection Screening

The QML branch was first explored through a step-by-step analysis to find the best input representation and circuit depth before benchmarking against classical methods. This analysis used a fixed set of 500 randomly selected observations to compare preprocessing and circuit design options under the same conditions. Different scaling strategies were tested while keeping the same QML clustering protocol. Among these, MinMaxScaler($0, \pi$) delivered the best clustering quality and outperformed StandardScaler and MinMaxScaler($-\pi, \pi$) in silhouette score. The optimal QML setup was with $k = 2$, suggesting that angular scaling to $[0, \pi]$ is the most appropriate for the quantum gates used (see Table 5 and Figure 3).

Table 5. Preliminary results to select the best QML model configuration.

Configuration	Reps	k	Silhouette	Davies_Bouldin	Calinski_Harabasz	Decision
StandardScaler, PCA(2)	1	2	0.151	4.280	41.824	discard
StandardScaler, PCA(2)	1	3	0.059	5.719	34.481	discard
MinMaxScaler($0, \pi$), PCA(2)	1	2	0.245	1.807	215.631	Kept for the PCA analysis
MinMaxScaler($0, \pi$), PCA(2)	1	3	0.216	3.109	156.956	discard
MinMaxScaler($-\pi, \pi$), PCA(2)	1	2	0.203	2.163	148.347	discard
MinMaxScaler($-\pi, \pi$), PCA(2)	1	3	0.165	3.536	103.126	discard
MinMaxScaler($0, \pi$), PCA(2)	1	2	0.368	1.056	670.335	discard

Table 5. Cont.

Configuration	Reps	k	Silhouette	Davies_Bouldin	Calinski_Harabasz	Decision
MinMaxScaler(0, π), PCA(2)	2	2	0.248	1.995	174.348	discard
MinMaxScaler(0, π), PCA(2)	3	2	0.245	1.807	215.631	discard
MinMaxScaler(0, π), PCA(2)	4	2	0.174	2.097	157.261	discard
MinMaxScaler(0, π), No PCA	1	2	0.484	2.045	225.098	selected
MinMaxScaler(0, π), No PCA	2	2	0.428	3.306	97.350	discard
MinMaxScaler(0, π), No PCA	3	2	0.396	4.289	62.825	discard
MinMaxScaler(0, π), No PCA	4	2	0.271	2.626	135.348	discard

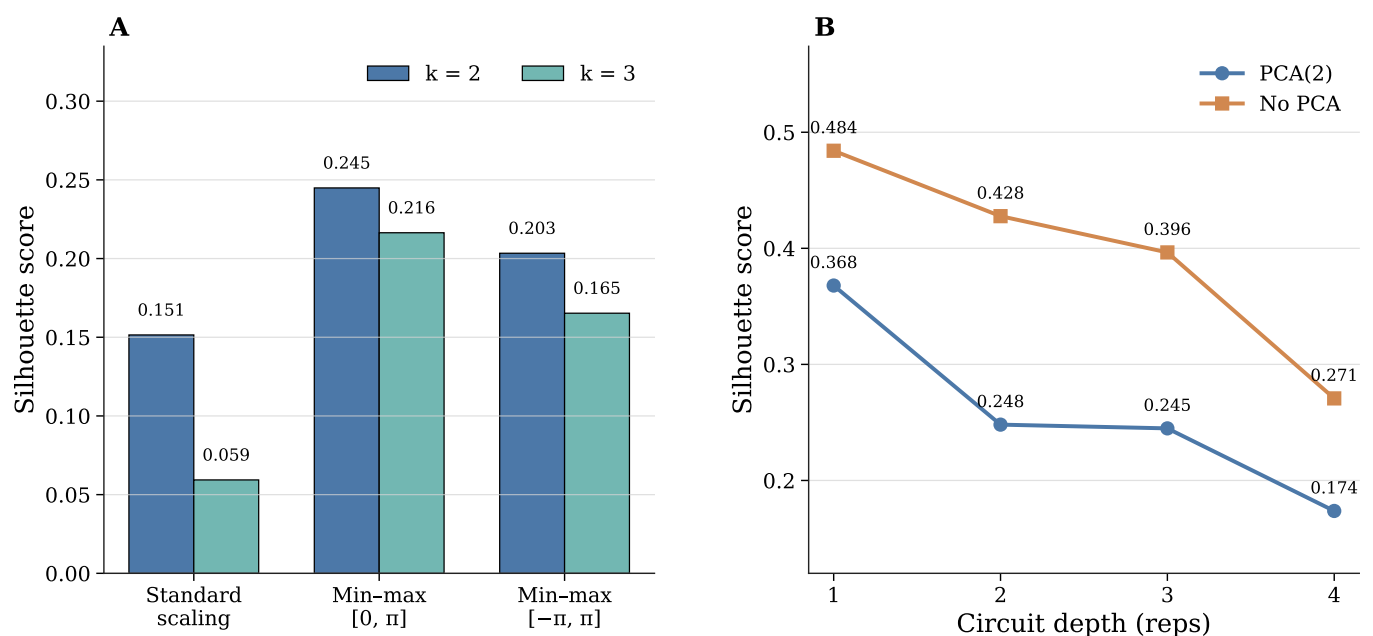


Figure 3. Exploratory QML screening of input scaling, dimensionality reduction, and circuit depth. (A) Effect of the scaling strategy on QML clustering quality during the first screening stage using PCA-reduced inputs and $k = 2$ or $k = 3$. (B) Effect of dimensionality reduction and PauliFeatureMap depth on clustering performance using the selected angular scaling.

In the next phase, the impact of dimensionality reduction and circuit depth was examined by comparing PCA-based inputs to direct encoding of the three original isotopic ratios with different repeat depths ($\text{reps} \in \{1, 2, 3, 4\}$). Results indicated that directly encoding the ratios, without PCA, consistently yielded better cluster quality (see Figure 3). These findings suggest that, for the Andean Pb isotope dataset, the QML branch benefits more from keeping the three original geochemical variables intact than from applying linear dimensionality reduction. They also show that increasing circuit depth does not improve isotopic partitioning; instead, the simplest encoding tested yielded the clearest separation. Therefore, the final QML setup for the benchmark used MinMaxScaler(0, π), direct encoding of the three isotope ratios without PCA, and a Pauli feature map with $\text{reps} = 1$.

3.2. QML Against Classical Baselines

Following the exploratory phase, the chosen QML configuration was benchmarked against traditional reference methods in a controlled, rigorous environment. The complete 1522 observations were used to ensure that all methods were tested on identical data and with consistent validation criteria. The comparison included the final QML pipeline, classical K-means with standardized inputs, K-means with minimum-maximum angle scaling, Gaussian mixture models, and Ward’s agglomerative clustering. Results showed that the QML pipeline achieved the highest cluster quality according to the primary metric. The highest silhouette score—the primary metric, applied uniformly to all methods—was obtained with the quantum kernel configuration combined with K-means ($k = 2$), achieving a silhouette score of about 0.490, which indicates a moderate rather than a strong cluster structure (silhouette values above ~ 0.5 are typically considered good), consistent with the well-documented isotopic overlap of Andean ore fields; surpassing the top Gaussian mixture model at $k = 2$, the best classical K-means with standardized inputs at $k = 3$, the best classical K-means with min-max angular scaling at $k = 3$, and Ward’s method at $k = 2$ (see Figure 4). However, additional validation metrics revealed a more nuanced picture. While the QML approach had the best cluster separation according to silhouette scores, some traditional methods outperformed it on Davies–Bouldin and Calinski–Harabasz scores across different values of k (see Figure 5). It is worth mentioning here that while the Davies–Bouldin index measures how similar each cluster is to its nearest neighbor—lower scores mean better separation—the Calinski–Harabasz index compares between-cluster and within-cluster variance, with higher scores indicating clearer cluster structure. Therefore, the quantum method favored broader but more distinct clusters, whereas classical approaches often favored more detailed structures with varying degrees of compactness and variance ratios.

To verify that the binary separation is not an artifact of the quantum feature map alone, the same fidelity-based pipeline was repeated using a classical radial-basis-function (RBF) kernel, optimizing its bandwidth (γ) by grid search under the identical K-means procedure. A nonlinear classical kernel also recovers a comparable two-group structure, confirming that the macro-partition is a robust property of the isotopic similarity space rather than a method-specific artifact. Nevertheless, under the primary metric adopted in this study—silhouette—the quantum kernel attains the highest score—0.490 versus 0.369 for the best RBF kernel—whereas the RBF kernel performs better on Davies–Bouldin and Calinski–Harabasz (Table 6). The quantum kernel is therefore competitive with—and, by the silhouette criterion, slightly superior to—an optimized classical nonlinear kernel for this binary task, without claiming a general advantage (see the complete set of results in Tables A1 and A2).

Table 6. Binary ($k = 2$) clustering quality of the fidelity-based quantum kernel versus an optimized classical RBF kernel under the same K-means pipeline. Silhouette and Calinski–Harabasz: higher is better; Davies–Bouldin: lower is better.

Method ($k = 2$)	Silhouette	Davies–Bouldin	Calinski–Harabasz
QML kernel + K-means	0.490	1.986	238.4
RBF kernel + K-means ($\gamma \approx 0.042$)	0.369	0.974	871.6

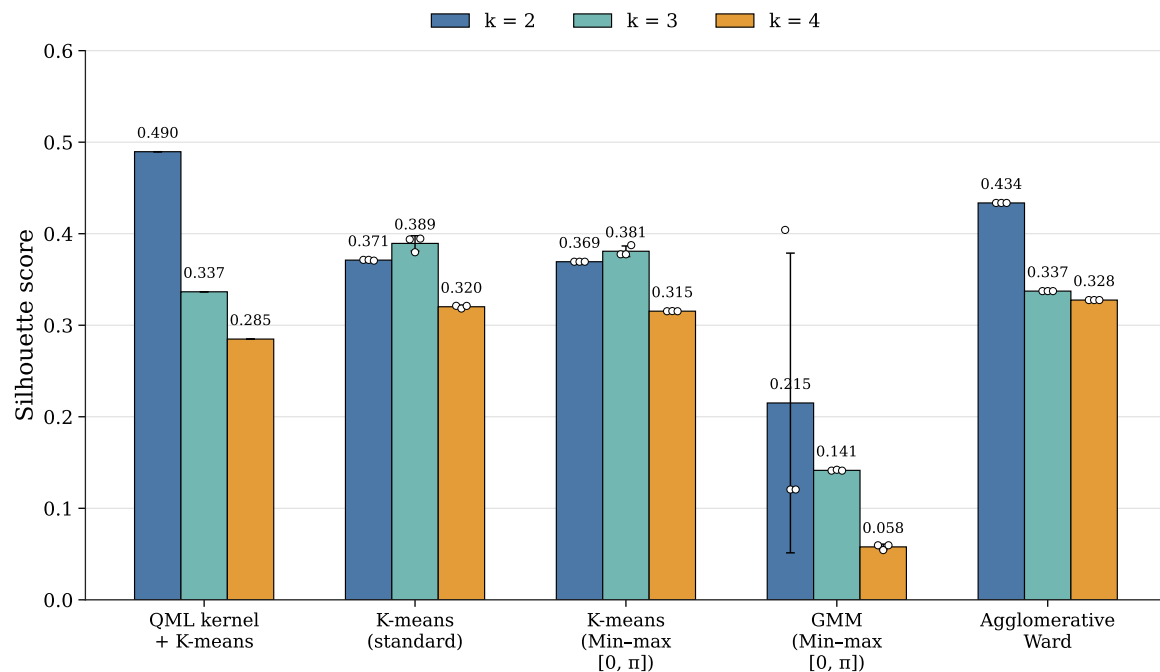


Figure 4. Full-dataset benchmark of clustering performance using the silhouette score. Bar colors distinguish the tested cluster counts ($k = 2$, $k = 3$, and $k = 4$); white circles indicate individual random-seed runs for stochastic models, and vertical whiskers indicate variability where applicable.

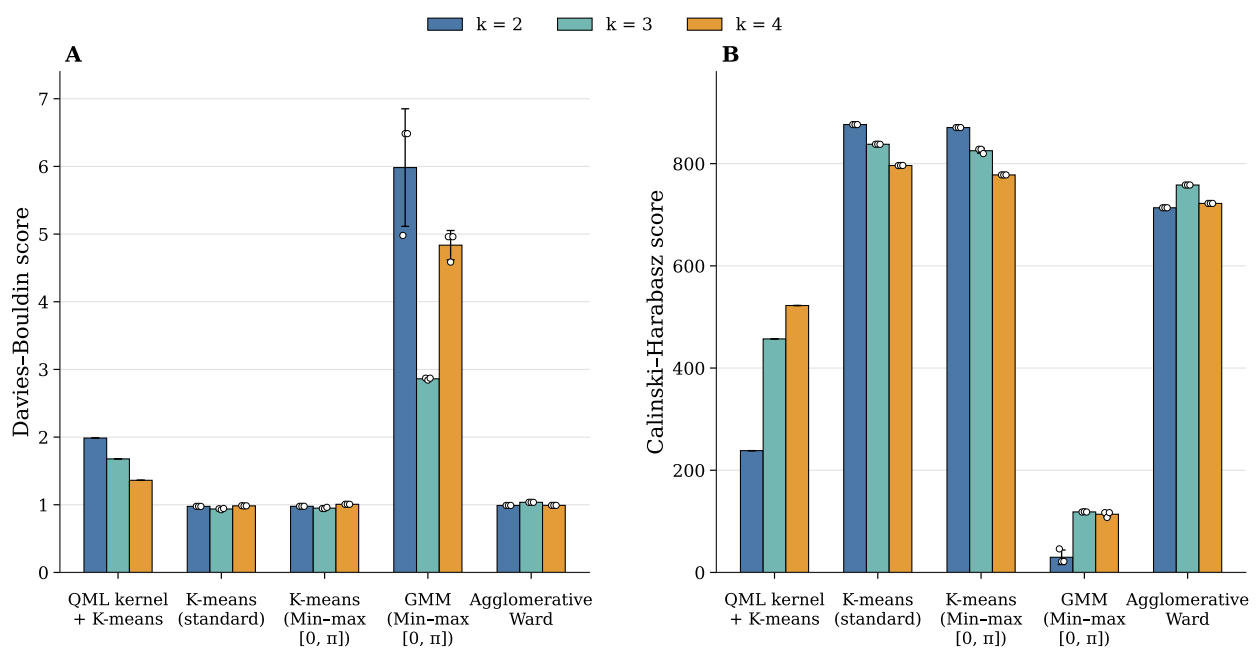


Figure 5. (A) Davies–Bouldin scores and (B) Calinski–Harabasz scores for the retained QML kernel and the classical benchmark models across different k values.

Collectively, these findings justified selecting the $k = 2$ QML solution as the primary quantum partition for the rest of this study. Meanwhile, in the previous classical analysis of the same Andean database, the 4-cluster solution was retained because it provided the best geological interpretability relative to the province framework discussed by Macfarlane and coworkers [1,2], even though alternative unsupervised tools could favor coarser solutions in Figures 5 and 6. The current QML benchmark does not disprove the multi-province interpretation. Instead, it indicates that the quantum-kernel approach tends to identify a

broader isotopic partition before recognizing more detailed geological subdivisions via traditional clustering.

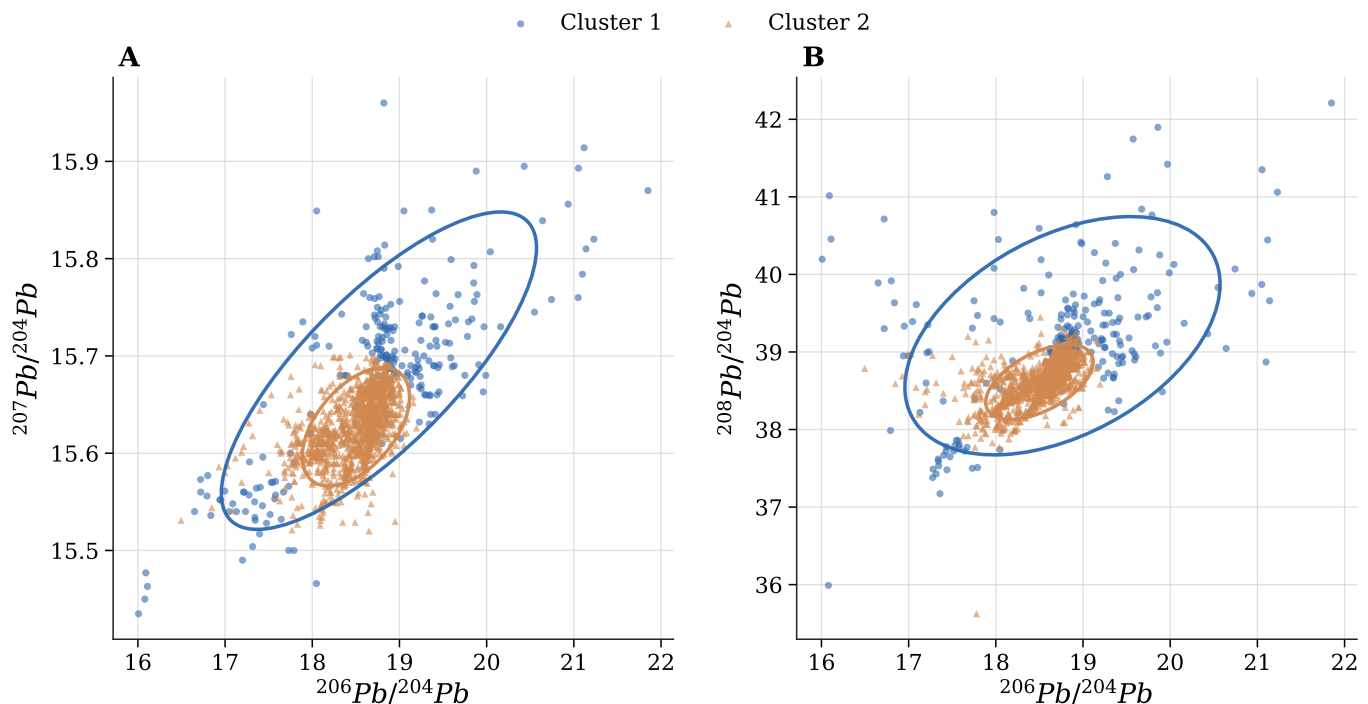


Figure 6. QML cluster structure in two-dimensional Pb isotope projections. Bivariate isotope projections for the retained QML solution at $k = 2$: (A) $^{206}\text{Pb}/^{204}\text{Pb}$ versus $^{207}\text{Pb}/^{204}\text{Pb}$, and (B) $^{206}\text{Pb}/^{204}\text{Pb}$ versus $^{208}\text{Pb}/^{204}\text{Pb}$. Blue circles and orange triangles denote samples assigned to Clusters 1 and 2, respectively; the colored ellipses are visual guides summarizing the dispersion of each cluster in the plotted projection.

Because K-means and GMM are stochastic algorithms, Figures 4 and 5 visualize initialization sensitivity. The large spread observed for GMM at $k = 2$ reflects genuine sensitivity to EM initialization in the MinMax-scaled space rather than a plotting artifact. To avoid overstating this behavior, the figure represents individual seed outcomes as points instead of using a visually dominant standard-deviation bar. The best RBF kernel K-means configuration was obtained at $k = 2$ with $\gamma = \gamma_{\text{median}} \times 0.1$, reaching a silhouette score of 0.369, as seen in Table A2. This value remains below the retained QML partition at $k = 2$ (0.490), indicating that the binary QML separation is not reproduced merely by replacing the Euclidean feature space with an arbitrary nonlinear RBF similarity. At the same time, the RBF control produced stronger Davies–Bouldin and Calinski–Harabasz values than the QML partition, consistent with the broader pattern that classical and RBF-based approaches favor compactness and variance-ratio structure, whereas the QML kernel emphasizes a more separated first-order partition.

3.3. Relationship Between QML Partition and the Classical Framework

After finalizing the QML configuration and validating it using the full dataset, the resulting quantum partition was compared with the four-cluster classical reference solution. This comparison aimed to see if both methods identify different levels of isotopic structure. It is not about whether the quantum method simply reproduces the classical solution at a lower resolution. Instead, it assesses whether the two approaches distinguish the Andean lead isotope space along fundamentally different axes of variation, with different implications for archaeometric provenance research. As shown below, the results indicate a hierarchical, complementary relationship between the two partitions.

3.3.1. A First-Order Binary Organization of the Isotopic Space

The QML partition with $k = 2$ is not entirely independent of the units, as it exhibits a more compact structure and a broader radiogenic distribution. In Figure 6, Cluster 2 (orange) mainly concentrates observations in the isotopic space core, roughly up to $^{206}\text{Pb}/^{204}\text{Pb}$ values of 18–19, $^{207}\text{Pb}/^{204}\text{Pb}$ of 15.58–15.68, and $^{208}\text{Pb}/^{204}\text{Pb}$ of 38.0–39.0. Meanwhile, Cluster 1 (blue), which is the most prevalent, extends toward the radiogenic extremes of all three ratios, with a particularly marked spread in $^{208}\text{Pb}/^{204}\text{Pb}$ (up to 42.21). The silhouette coefficient for this binary partition (≈ 0.490) is the highest among all models tested in this benchmark, indicating that within the similarity space defined by the Pauli feature map, the two-cluster solution provides the most distinct separation.

The geological significance of this first-order binary split becomes clearer when placed in the context of the Andean lead isotope literature. Macfarlane et al. [1,2] established that the main axis of isotopic variation in the central Andes extends from west to east, showing a gradual increase in crustal input as the distance from the Peru–Chile trench increases. Orogenic-arc sources in Province I (coastal belt) exhibit relatively unradiogenic compositions, while Paleozoic upper-crustal and high-grade metamorphic basement sources in the Altiplano and Eastern Cordillera produce increasingly radiogenic and thorogenic signatures. This prominent west-to-east gradient, rather than a north-to-south continuity, serves as the key organizing principle of the Andean isotopic landscape. The QML binary partition precisely reflects this gradient: Cluster 2 (orange) generally represents the lower-crustal or orogenic end of the spectrum, whereas Cluster 1 (blue) encompasses the more radiogenic and thorogenically enriched compositions from deeper or older crustal sources.

To support this interpretation quantitatively rather than qualitatively, we examined the composition and spatial distribution of the two clusters directly as seen in Table A3. The two groups differ strongly and significantly in all three isotope ratios (Welch t-tests, $p < 10^{-111}$ in every case): Cluster 1 is systematically more radiogenic (mean $^{207}\text{Pb}/^{204}\text{Pb} = 15.676$ vs. 15.605) and more thorogenic ($^{208}\text{Pb}/^{204}\text{Pb} = 39.00$ vs. 38.48) than Cluster 2. In contrast, the binary partition is not a geographic east–west split: the two clusters do not differ significantly in longitude (means -70.6° vs. -70.8° ; $t = 0.89$, $p = 0.38$; point-biserial $r = -0.02$). The $k = 2$ partition therefore separates the data by isotopic radiogenicity, which correlates only loosely with the west-to-east gradient because of the strong overlap of Andean ore fields, rather than by raw geographic position. The agreement between the binary and the four-cluster partitions was quantified with the adjusted Rand index (ARI = 0.41) and normalized mutual information (NMI = 0.50), confirming that the two solutions are related but operate at different resolutions: the $k = 2$ partition captures the first-order compositional contrast, while the $k = 4$ solution refines it into provincial-scale subdivisions. Cluster sizes are 857 and 665 samples for $k = 2$.

Killick and coworkers [35] highlighted that provenance analysis using Pb isotopes in the Andes is inherently complex because deposits from different regions, formed within the same convergent plate boundary, often share similar isotopic signatures. This overlap causes provincial boundaries to coincide, complicating clear point-to-source identification. The QML model effectively captures this challenge by not enforcing strict provincial boundaries, where overlaps are common. Instead, it identifies the most significant isotopic discontinuity—distinguishing a dense, compositionally uniform domain from a larger, more varied region characterized by higher radiogenic and thorogenic signatures. This approach provides a meaningful initial division, rather than an arbitrary statistical split.

3.3.2. Classical Clustering as a Finer-Resolution Subdivision

Figure 7 shows the four-cluster classical solution, maintained as the main geological reference in this study with the cluster–province correspondence summarized in Table A4. Compared to the QML partition, the classical method uncovers internal structures within each of the two quantum domains: subclusters with distinct isotopic centroids form, matching the province boundaries initially defined by Macfarlane et al. [2] and later detailed by Aitchison et al. [4] and Kamenov et al. [5]. These subclusters correspond with the geological division of the Andes into units mainly influenced by Jurassic–Cenozoic arc magmatism (Province I), mixing zones between orogenic and crustal sources (Province II), Paleozoic upper-crustal contributions (Province III), and high-grade metamorphic basement (Province IV).

The contingency heat map (Figure 8) measures the relationship between the two partitions and is essential for understanding their connection. If the QML solution were just a coarser version of traditional clustering, each quantum cluster would roughly split the classical clusters into two equal halves—a nearly symmetric redistribution. However, the contingency matrix displays a clear asymmetric pattern: some classical clusters are almost entirely captured by one QML cluster, while others are divided unevenly between both quantum groups. This structural asymmetry is consistent with the kernel emphasizing broad compositional contrasts; we note, however, that a comparable asymmetry can also arise from other nonlinear kernels (see the RBF comparison in Section 3.2), so we interpret it as a property of nonlinear similarity in general rather than as evidence of a uniquely quantum organizing dimension. In machine learning terms, the quantum feature map emphasizes a different—and complementary—direction of maximum separation in the similarity space, distinct from the Euclidean distances used by K-means, GMM, and Ward’s agglomerative algorithm.

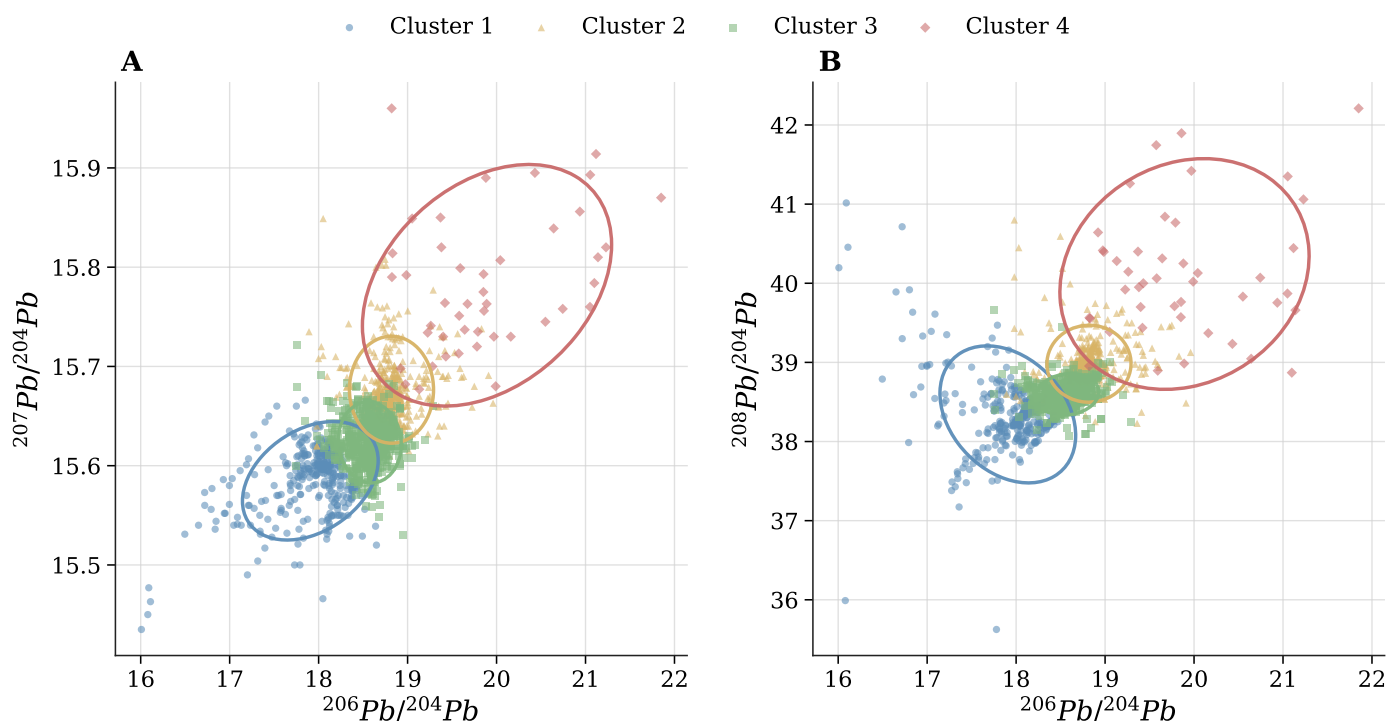


Figure 7. Classical four-cluster structure in two-dimensional Pb isotope projections. Bivariate isotope projections for the retained classical clustering solution at $k = 4$: (A) $^{206}\text{Pb}/^{204}\text{Pb}$ versus $^{207}\text{Pb}/^{204}\text{Pb}$, and (B) $^{206}\text{Pb}/^{204}\text{Pb}$ versus $^{208}\text{Pb}/^{204}\text{Pb}$. Colored markers identify the four cluster assignments, and the matching ellipses are visual guides summarizing within-cluster dispersion in each projection.

This interpretation should be read together with the nonlinear RBF control reported in Section 3.2 and Table A2. Since the optimized RBF kernel did not match the QML silhouette performance, the observed binary split cannot be attributed solely to the generic use of a nonlinear similarity matrix. Within the controls tested here, the Pauli feature-map kernel induced a distinct first-order organization of the Pb isotope dataset, while classical and RBF-based models remained more favorable under compactness-oriented and variance-ratio metrics.

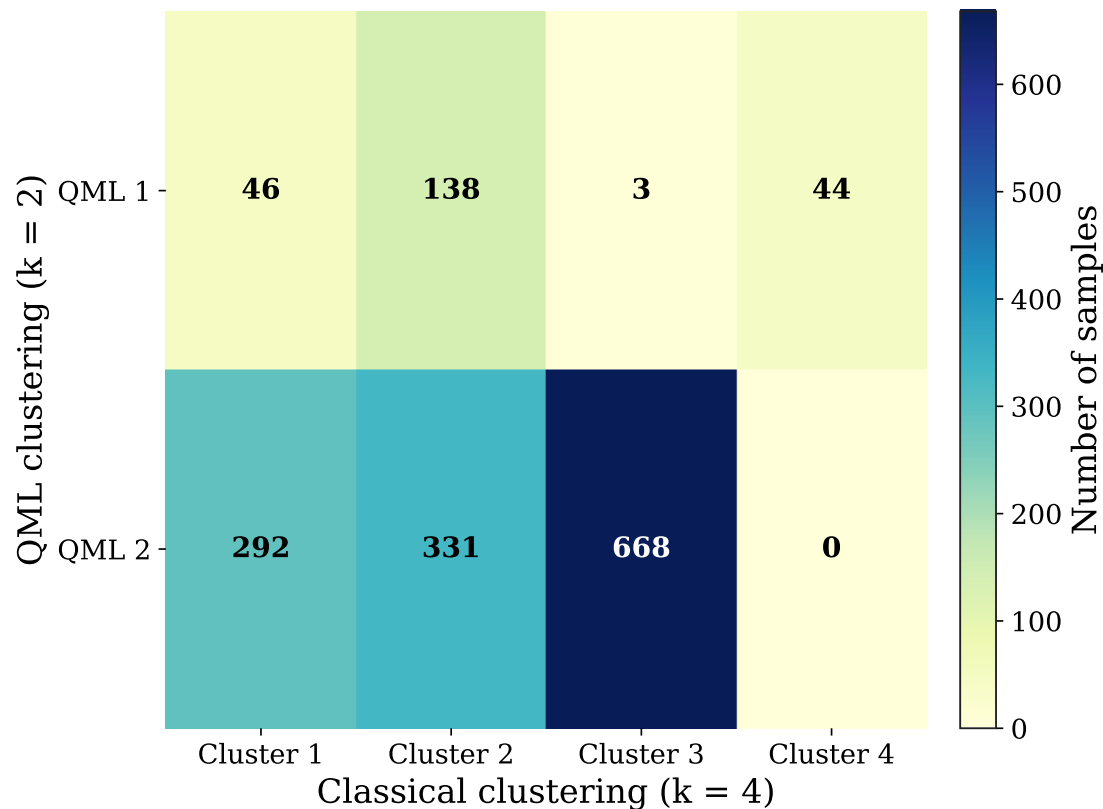


Figure 8. Contingency matrix between the QML and classical clustering solutions.

3.3.3. Quantum Kernel Geometry and Its Archaeometric Implications

The documented behavior aligns with current views on quantum kernel methods within NISQ-era limitations. As highlighted by Thanasilp et al. [17] and empirically shown by Schnabel & Roth [18], QKMs tend to focus on broad differences in the feature space and are influenced by the design of the encoding circuit. The Pauli feature map with one repetition layer—chosen in Section 3.1 as optimal—encodes the three isotope ratios into quantum states using single-qubit rotations and pairwise ZZ entanglement. This entanglement creates cross-ratio interaction terms absent in classical Euclidean distances, possibly explaining why the quantum similarity matrix detects a large-scale separation that classical metrics do not consider the main partition.

The RBF control strengthens this interpretation by showing that the QML result is not simply equivalent to applying a conventional nonlinear radial kernel. Instead, the retained Pauli feature map appears to impose a different similarity geometry on the isotope ratios, although this statement remains limited to the specific classical kernels and quantum circuit family tested in this work.

From an archaeometric perspective, this hierarchical complementarity is highly practical. In the Andes, provenance studies often face the challenge of overlapping isotopic fields: a single bivariate plot of $^{207}\text{Pb}/^{204}\text{Pb}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$ frequently cannot distinguish source candidates due to overlapping regions in that projection [35]. Kernel density estimation methods proposed by De Ceuster & Degryse [36] have explicitly measured this overlap and demonstrated that pairwise source discrimination often fails when comparing different Andean ore fields. In this setting, a data-driven approach that first defines a robust macro-partition—without relying on predefined geographic borders—serves as an effective pre-screening step before detailed province-level clustering. An artifact with a lead-isotope signature within QML Cluster 1 (the radiogenic/thorogenic domain) can be immediately directed to focus on Province III or IV ore bodies, thereby reducing the candidate list prior to performing province-specific bivariate analyses. Conversely, artifacts in Cluster 2 suggest an origin from orogenic/arc-related sources (Provinces I and II), aligning with metal sourcing from the coastal batholithic belt.

The two partitions are best understood hierarchically: the quantum-kernel solution provides a robust first-order, macroscopic separation of the isotopic space along the west-to-east crustal-contamination gradient, whereas the classical four-cluster solution refines this structure into provincial-scale subdivisions. In this sense, the approaches are complementary rather than competing, with the quantum kernel supplying large-scale separation and the classical methods adding regional detail. We emphasize that these results are obtained by exact state-vector simulation; alternative quantum encodings (for example, amplitude encoding or data re-uploading) and richer entanglement topologies may alter the induced similarity geometry and constitute clear directions for future work, ideally on NISQ hardware.

3.4. Isotopic and Geographic Interpretation and Final Partitions

This study examines two partitionings—the traditional four-group solution and the two-group QML solution—within the established isotopic province framework for the central Andes. It also considers published archaeometric case studies that have used lead isotopes to link metal artifacts to their mineral sources. This dual interpretive approach enables us to evaluate the geological validity of the grouping results and to explore their significance for provenance analysis in pre-Columbian metallurgy.

3.4.1. Geographic Distribution and Geological Correspondence of the Classical Clusters

Figure 9 illustrates both partitioning solutions across the Andean region, with the cluster–province correspondence summarized in Table A4. Figure 10 provides an author-redrawn, simplified morphotectonic visualization of the established geological provinces and the west-to-east crustal-contamination trend; it is included as an interpretive aid and schematic guide rather than as an experimental image. The four traditional clusters largely match the four isotopic provinces initially identified by Macfarlane et al. [2] and later refined by Aitchison et al. [4], Kamenov et al. [5], and Macfarlane & Lechtman [1]. Province I, characterized by coastal arc magmatism from the Jurassic to early Tertiary periods and distinguished by orogenic lead compositions, corresponds to the green cluster (cluster 3 in Figure 9). This cluster is primarily aligned along the western Cordillera from the Peruvian coast north to Ecuador. Notably, our results indicate that this cluster extends to around 0° latitude and reaches Ecuador, a geographical continuity already proposed by Macfarlane & Lechtman [1]. This extension aligns with geological expectations, given the ongoing continuity of the Andean magmatic arc, and is supported by the most comprehensive multisource database assembled for this research.

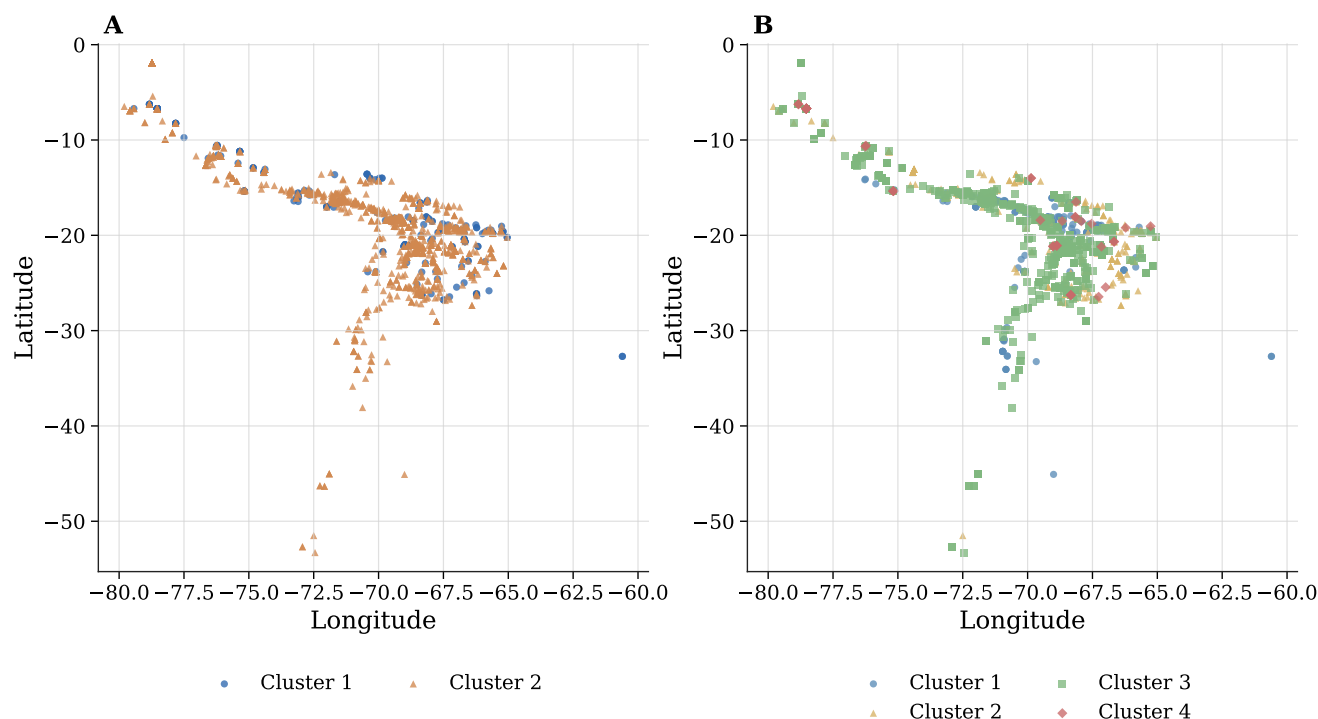


Figure 9. Geographic distribution of QML and classical cluster assignments. Spatial distribution of cluster assignments based on sample longitude and latitude. (A) QML clustering solution at $k = 2$. (B) Classical clustering solution at $k = 4$.

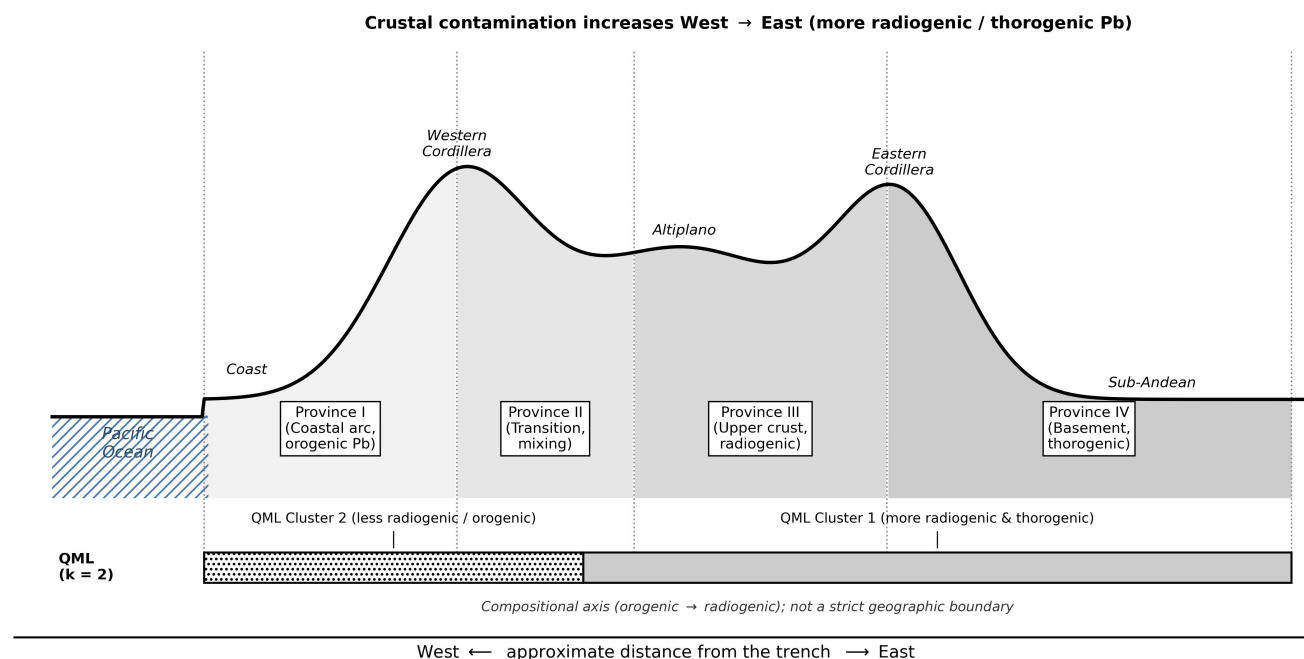


Figure 10. Author-redrawn, simplified morphotectonic visualization and interpretive aid of the central Andes, illustrating the west-to-east increase in crustal contamination of ore lead. The two-cluster quantum-kernel partition aligns with this first-order gradient, whereas the four-cluster partition approximates the isotopic provinces of Macfarlane & Lechtman [1].

Province II, seen as a transitional zone between the orogenic sources (Province I) and the upper-crustal sources (Province III), is classified within the green group. This aligns with Macfarlane and Lechtman's [1] view, who considered Province II as a mixing area rather than an isolated isotopic region. The blue group spans from Province IV, extending

southward from the Pisco coast through northern Chile, western Bolivia, and northwestern Argentina. Geologically, Province IV is characterized by highly metamorphosed Proterozoic and Paleozoic basement rocks involved in deposit formation [4,5], resulting in higher $^{208}\text{Pb}/^{204}\text{Pb}$ ratios relative to orogenic lead, indicating enrichment and increased Th/U ratios in the source rocks. The yellow and red groups together form Province III, starting in Lambayeque (northern Peru), crossing western Bolivia and northern Chile, and reaching northwestern Argentina. Province III features two distinct geochemical subgroups: one with high $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios from Paleozoic upper crust sediments, and another with unusually high $^{206}\text{Pb}/^{204}\text{Pb}$ ratios caused by mixing of orogenic lead with a terminal ^{238}U -rich upper crust component [2]. The proximity of the yellow and red groups in Figure 7 indicates a continuum, supporting the idea that they belong to the same primary geological region.

3.4.2. Role of the $^{208}\text{Pb}/^{204}\text{Pb}$ Ratio as the Primary QML Discriminator

A key insight in interpreting the QML binary partition involves understanding the roles of the three isotope ratios. In Figure 6, the $^{208}\text{Pb}/^{204}\text{Pb}$ ratio most clearly differentiates QML Cluster 1 (blue) from Cluster 2 (orange): the dispersed cluster extends significantly into the high- $^{208}\text{Pb}/^{204}\text{Pb}$ range, while the compact cluster remains within lower thorogenic values. This distinction is geochemically meaningful. The $^{208}\text{Pb}/^{204}\text{Pb}$ ratio indicates the thorogenic component of lead—the decay product of ^{232}Th —and is thus most responsive to crustal sources with high Th/U ratios, such as high-grade metamorphic basement and Paleozoic sedimentary sequences [37]. In the plumbotectonic model by Zartman & Doe [37], the upper crustal growth curve is characterized by high $^{208}\text{Pb}/^{204}\text{Pb}$ ratios compared to the orogen and mantle paths, making this ratio a key factor in distinguishing arc-related from continent-related ore leads. Province IV, in particular—whose blue cluster is located in the southern Altiplano—is characterized by low $^{206}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ ratios but high $^{208}\text{Pb}/^{204}\text{Pb}$, consistent with metamorphic basement sources with high Th/U ratios and old crustal ages [4,5].

The quantum kernel, by encoding all three ratios through a Pauli feature map with ZZ entanglement, effectively captures cross-ratio interaction terms—especially the joint variation of $^{208}\text{Pb}/^{204}\text{Pb}$ with $^{206}\text{Pb}/^{204}\text{Pb}$ —that are linked to the gradual crustal contamination pattern along the west-to-east Andean transect. The binary separation it creates is well aligned with geological observations: the broad difference between arc- or orogenic-dominated lead (shown as a compact cluster) and basement- or crustal-dominated lead (shown as a dispersed cluster) reflects the fundamental plumbotectonic structure of the Andes. The noticeable overlap between clusters in some bivariate projections (Figure 6) is not a flaw in the QML method; rather, it thoughtfully represents the continuous, mixing-rich nature of the Andean isotopic landscape [35,36].

3.4.3. Archaeometric Implications

The archaeometric usefulness of this clustering model is best demonstrated by its ability to classify previously studied artifact assemblages. Macfarlane and Lechtman [1] examined the lead isotopic makeup of Cu-As-Ni bronze artifacts from Tiwanaku, an Altiplano capital, and San Pedro de Atacama, a desert oasis trading post. They found that the metal in both assemblages originated from ore deposits in the Altiplano and High Sierra, a region between Provinces II and IV in our scheme (blue/green boundary in Figure 9), matching their isotopic signatures of mixed orogenic and deep-crustal sources. Their study also indicated that arsenic-bronze artifacts from the Wari site of Conchopata have isotopic signatures aligning with the Julcani copper sulfarsenide deposit in Huancavelica. The literature places this deposit within Province III (the yellow-red cluster in our results),

which matches the high $^{206}\text{Pb}/^{204}\text{Pb}$ ratios seen in ^{238}U -rich crustal springs in central Peru. The spatial correlation between our clustering outcomes and these independent archaeometric findings supports the geological validity of the provincial boundaries defined by the traditional clustering approach.

These agreements demonstrate the practical value of a hierarchical clustering workflow for future provenance research. An artifact from the Wari region assigned to QML Cluster 1 (the radiogenic/thorogenic area) would directly indicate Province III or upper crust ore bodies in the central Peruvian Andes. Conversely, an artifact assigned to Cluster 2 would suggest sources from the coastal arc and Altiplano, corresponding to Provinces I and II. This initial QML classification effectively narrows the search area before applying the province-level distinction, which can be complicated by overlaps. For regional archaeometric databases that combine artifacts from diverse cultural and geographic contexts, this two-stage method could significantly enhance the efficiency of provenance determination. Paleolimnological evidence supporting the geographic interpretation of our clusters includes reconstructions of pre-Columbian metallurgical history. Studies of trace metal deposits in Lake Titicaca sediment cores used lead isotopic signatures to link smelting emissions with ore deposits in the southern Altiplano during the Tiwanaku period (800–1150 CE) and reveal diverse mining activities during the Late Intermediate Period (1150–1450 CE), with the most intense mining beginning under Inca rule [38]. These spatiotemporal patterns largely align with the geographic range of the blue cluster (Province IV) shown in Figure 9, which covers the mining regions of the southern Altiplano in Bolivia, including Potosí, whose minerals were heavily exploited during the Inca era.

3.4.4. Geographic Scope of the QML Partition and Limitations

The two-cluster QML approach (shown in Figures 6 and 9) identifies an isotopic polarity aligned roughly parallel to the main west–east crustal gradient. However, it does not effectively resolve the north–south continuity, which limits differentiation within individual provinces. As Macfarlane and Lechtman [1] explicitly noted, lead isotope ratios vary significantly from west to east across the Andes, but they are less useful for distinguishing deposits that run north–south along the Cordillera's strike. This directional limitation stems from the tectonic structure of the orogen itself, not from the modeling process. Consequently, neither traditional partitioning methods nor QML can reliably differentiate, for example, ore bodies in northern versus southern Peru that are equidistant from the deep-ocean trench and share a similar crustal history. This distinction is crucial in archaeometric work: while group classification can help define the probable geological origin (such as orogenic versus upper-crustal/metamorphic), additional isotopic data—such as Cu, Ag, or Pb–Pb model ages—and archaeological context are necessary to specify the mining districts accurately.

Despite this limitation, our database's extensive geographic coverage—from Ecuador to northwestern Argentina—and the data-driven structure of the QML core offer a significant contribution to the field. Most provincial models currently rely on smaller, more limited collections [1,2,4,5]. The fact that our classical approach with $k = 4$ yields a configuration largely aligned with Macfarlane's provincial model, but extends northward to Ecuador and enhances Altiplano coverage, indicates that our database is sufficiently representative for regional provenance analysis. Additionally, QML's binary partitioning provides a robust, data-driven foundation that does not rely on fixed geographic boundaries, enabling seamless updates with new data from understudied Andean regions, such as northern Ecuador or the eastern sub-Andean ranges, without manual reclassification.

4. Conclusions

This study introduced a reproducible fidelity-based kernel method for unsupervised clustering of lead-isotope data from the Andes and compared it with traditional methods based on 1522 measurements. The optimal quantum setup, utilizing a three-qubit Pauli feature map with MinMaxScaler (0, π) and direct encoding of isotopic ratios, achieved the highest silhouette score among all tested models, indicating that the quantum similarity space captured a moderate—rather than strong—isotopic separation (silhouette $\approx$ 0.49). Geologically, the two-cluster quantum partition revealed a first-order isotopic organization aligning with the west–east crust gradient in the Andes. It distinguished radiogenic compositions associated with volcanic arcs from crustal signatures, whereas the four-cluster classical solution showed finer subdivisions consistent with known isotopic provinces. These results support a hierarchical and complementary interpretation of lead isotope origins in the Andes.

From an archaeometric view, this complementarity suggests a two-step approach: quantum partitioning provides robust macroscopic clustering of isotopic data, which can then be refined with classical clustering and provincial models. The quantum method is not meant to replace geological frameworks but to add a useful layer for analyzing complex datasets. Limitations include reliance on classical simulation to obtain quantum results and on clustering as a similarity benchmark rather than on demonstrating quantum advantage. Future research should examine different clustering methods, incorporate geological validation, and expand datasets. Nonetheless, these findings highlight that quantum representations can enhance isotope origin studies by uncovering large-scale structures that complement classical clustering.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/quantum8030058/s1>, The supplementary material submitted with this article is available with the paper through the journal and contains the source metadata and Pb-isotope data used for the analyses.

Author Contributions: Conceptualization, A.A.-M.; methodology, A.A.-M. and F.D.; software, A.A.-M. and F.D.; validation, A.A.-M., A.V.-A., and S.R.-F.; formal analysis, A.A.-M.; investigation, A.A.-M.; resources, A.A.-M.; data curation, A.A.-M.; writing—original draft preparation, A.A.-M.; writing—review and editing, A.V.-A., F.D., and S.R.-F.; visualization, A.A.-M.; supervision, A.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript, the authors used Claude 4.8 for final language polishing, including typographical, wording, and grammatical revisions. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Additional Benchmark Tables

The following tables summarize the initialization-sensitivity analysis for stochastic baselines and the RBF-kernel control included in the extended benchmark. These supplementary tables support Figures 4 and 5 and the nonlinear-kernel comparison discussed in Section 3.2.

Table A1. Initialization-sensitivity summary for benchmark models. Values are reported as mean $\pm$ SD across independent random-seed runs where applicable. The QML and Ward entries are fixed references.

Method	k	Silhouette Mean $\pm$ SD	Davies–Bouldin Mean $\pm$ SD	Calinski–Harabasz Mean $\pm$ SD
QML kernel + K-means	2	0.490 $\pm$ 0.000	1.986 $\pm$ 0.000	238.4 $\pm$ 0.0
QML kernel + K-means	3	0.337 $\pm$ 0.000	1.678 $\pm$ 0.000	457.0 $\pm$ 0.0
QML kernel + K-means	4	0.285 $\pm$ 0.000	1.362 $\pm$ 0.000	522.4 $\pm$ 0.0
K-means (standard)	2	0.371 $\pm$ 0.000	0.977 $\pm$ 0.001	876.5 $\pm$ 0.0
K-means (standard)	3	0.389 $\pm$ 0.008	0.938 $\pm$ 0.009	838.0 $\pm$ 0.2
K-means (standard)	4	0.320 $\pm$ 0.002	0.984 $\pm$ 0.002	796.4 $\pm$ 0.1
K-means (MinMax 0– π)	2	0.369 $\pm$ 0.000	0.977 $\pm$ 0.000	870.6 $\pm$ 0.0
K-means (MinMax 0– π)	3	0.381 $\pm$ 0.006	0.950 $\pm$ 0.010	825.3 $\pm$ 4.8
K-means (MinMax 0– π)	4	0.315 $\pm$ 0.000	1.007 $\pm$ 0.000	777.9 $\pm$ 0.0
GMM (MinMax 0– π)	2	0.215 $\pm$ 0.164	5.983 $\pm$ 0.869	29.7 $\pm$ 14.2
GMM (MinMax 0– π)	3	0.141 $\pm$ 0.001	2.861 $\pm$ 0.016	118.5 $\pm$ 0.3
GMM (MinMax 0– π)	4	0.058 $\pm$ 0.003	4.836 $\pm$ 0.216	113.9 $\pm$ 5.5
Agglomerative Ward	2	0.434 $\pm$ 0.000	0.990 $\pm$ 0.000	713.7 $\pm$ 0.0
Agglomerative Ward	3	0.337 $\pm$ 0.000	1.036 $\pm$ 0.000	758.2 $\pm$ 0.0
Agglomerative Ward	4	0.328 $\pm$ 0.000	0.991 $\pm$ 0.000	722.2 $\pm$ 0.0

Table A2. RBF-kernel control benchmark compared with the retained QML reference.

Method	k	Gamma Relative to Median	Gamma	Silhouette	Davies–Bouldin/Calinski–Harabasz
QML kernel + K-means	2			0.490	1.986/238.4
RBF kernel K-means	2	0.1	0.042367	0.369	0.974/871.6
RBF kernel K-means	3	0.1	0.042367	0.336	1.030/777.3
RBF kernel K-means	4	0.1	0.042367	0.308	1.121/690.8

Table A3. Per-cluster summary statistics (mean $\pm$ standard deviation) of the three Pb isotope ratios and mean longitude for the binary (k = 2) quantum-kernel partition.

Cluster (k = 2)	n	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{204}\text{Pb}$	Mean Longitude
Cluster 2 (orange)	857	18.29 $\pm$ 0.43	15.605 $\pm$ 0.031	38.48 $\pm$ 0.34	−70.6°
Cluster 1 (blue)	665	18.87 $\pm$ 0.41	15.676 $\pm$ 0.044	39.00 $\pm$ 0.45	−70.8°

Table A4. Explicit correspondence between the four classical clusters (k = 4), their colors in Figure 7, and the Macfarlane–Lechtman provinces to resolve the cluster–colour–province mapping. The same colors are used consistently in Figures 7 and 9.

Cluster (k = 4, Figure 7)	Color	Macfarlane–Lechtman Province
Cluster 1	Blue	Provinces I–II (orogenic + transitional)
Cluster 2	Yellow	Province III (upper crust)
Cluster 3	Green	Province III (upper crust)
Cluster 4	Red	Province IV (basement)

References

- Macfarlane, A.W.; Lechtman, H.N. Andean Ores, Bronze Artifacts, and Lead Isotopes: Constraints on Metal Sources in Their Geological Context. *J. Archaeol. Method Theory* **2014**, *23*, 1–72. [\[CrossRef\]](#)
- Macfarlane, A.W.; Marcet, P.; LeHuray, A.P.; Petersen, U. Lead Isotope Provinces of the Central Andes Inferred from Ores and Crustal Rocks. *Econ. Geol.* **1990**, *85*, 1857–1880. [\[CrossRef\]](#)
- Mamani, M.; Wörner, G.; Sempere, T. Geochemical Variations in Igneous Rocks of the Central Andean Orocline (13°S to 18°S): Tracing Crustal Thickening and Magma Generation through Time and Space. *GSA Bull.* **2010**, *122*, 162–182. [\[CrossRef\]](#)

4. Aitchison, S.J.; Harmon, R.S.; Moorbath, S.; Schneider, A.; Soler, P.; Soria-Escalante, E.; Steele, G.; Swainbank, I.; Wörner, G. Pb Isotopes Define Basement Domains of the Altiplano, Central Andes. *Geology* **1995**, *23*, 555–558. [\[CrossRef\]](#)
5. Kamenov, G.; MacFarlane, A.W.; Riciputi, L. Sources of Lead in the San Cristobal, Pulacayo, and Potosí Mining Districts, Bolivia, and a Reevaluation of Regional Ore Lead Isotope Provinces. *Econ. Geol.* **2002**, *97*, 573–592. [\[CrossRef\]](#)
6. Xu, D.; Tian, Y. A Comprehensive Survey of Clustering Algorithms. *Ann. Data Sci.* **2015**, *2*, 165–193. [\[CrossRef\]](#)
7. Sadeghi, M.; Casey, P.; Carranza, E.J.M.; Lynch, E.P. Principal Components Analysis and K-Means Clustering of till Geochemical Data: Mapping and Targeting of Prospective Areas for Lithium Exploration in Västernorrland Region, Sweden. *Ore Geol. Rev.* **2024**, *167*, 106002. [\[CrossRef\]](#)
8. Alviz-Meza, A. Tracing Andean Origins: A Machine Learning Framework for Lead Isotopes. *EarthArXiv* **2024**. [\[CrossRef\]](#)
9. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum Machine Learning. *Nature* **2017**, *549*, 195–202. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Schuld, M.; Petruccione, F. *Supervised Learning with Quantum Computers*; Springer Nature: Cham, Switzerland, 2018. [\[CrossRef\]](#)
11. Preskill, J. Quantum Computing in the NISQ Era and Beyond. *Quantum* **2018**, *2*, 79. [\[CrossRef\]](#)
12. Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S.C.; Endo, S.; Fujii, K.; McClean, J.R.; Mitarai, K.; Yuan, X.; Cincio, L.; et al. Variational Quantum Algorithms. *Nat. Rev. Phys.* **2021**, *3*, 625–644. [\[CrossRef\]](#)
13. Zaidenberg, D.A.; Sebastianelli, A.; Spiller, D.; Le Saux, B.; Ullo, S.L. Advantages and Bottlenecks of Quantum Machine Learning for Remote Sensing. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 5680–5683. [\[CrossRef\]](#)
14. Bravyi, S.; Hastings, M. On Complexity of the Quantum Ising Model. *Commun. Math. Phys.* **2016**, *349*, 1–45. [\[CrossRef\]](#)
15. Havlíček, V.; Córcoles, A.D.; Temme, K.; Harrow, A.W.; Kandala, A.; Chow, J.M.; Gambetta, J.M. Supervised Learning with Quantum-Enhanced Feature Spaces. *Nature* **2019**, *567*, 209–212. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Quantum Kernel Machine Learning—Qiskit Machine Learning 0.9.0. Available online: https://qiskit-community.github.io/qiskit-machine-learning/tutorials/03_quantum_kernel.html (accessed on 20 April 2026).
17. Thanasilp, S.; Wang, S.; Cerezo, M.; Holmes, Z. Exponential Concentration in Quantum Kernel Methods. *Nat. Commun.* **2024**, *15*, 5200. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Schnabel, J.; Roth, M. Quantum Kernel Methods under Scrutiny: A Benchmarking Study. *Quantum Mach. Intell.* **2025**, *7*, 58. [\[CrossRef\]](#)
19. Fan, F.; Shi, Y.; Zhu, X.X. Earth Observation Data Classification with Quantum-Classical Convolutional Neural Network. In Proceedings of the IGARSS 2022—2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 191–194. [\[CrossRef\]](#)
20. Liu, Y.; Wang, W.; Wang, H.; Alidaee, B.; Florence, A. Quantum Machine Learning on Remote Sensing Data Classification. *J. Eng. Res. Sci.* **2023**, *2*, 23–33. [\[CrossRef\]](#)
21. Miroszewski, A.; Nalepa, J.; Wijata, A.M.; Mielczarek, J.; Le Saux, B.; Sebastianelli, A. Quo Vadis, Quantum Machine Learning?: Quantum Kernel Methods Meet Earth Observation. *IEEE Geosci. Remote Sens. Mag.* **2026**, *14*, 306–334. [\[CrossRef\]](#)
22. Tomczyk, W.; Giersz, M.; Sołtysiak, A.; Kamenov, G.; Krigbaum, J. Patterns of Camelid Management in Wari Empire Reconstructed Using Multiple Stable Isotope Analysis: Evidence from Castillo de Huarmey, Northern Coast of Peru. *Archaeol. Anthropol. Sci.* **2019**, *11*, 1307–1324. [\[CrossRef\]](#)
23. Pichat, S.; Abouchami, W.; Galer, S.J.G. Lead Isotopes in the Eastern Equatorial Pacific Record Quaternary Migration of the South Westerlies. *Earth Planet. Sci. Lett.* **2014**, *388*, 293–305. [\[CrossRef\]](#)
24. Cooke, C.A.; Wolfe, A.P.; Hobbs, W.O. Lake-Sediment Geochemistry Reveals 1400 Years of Evolving Extractive Metallurgy at Cerro de Pasco, Peruvian Andes. *Geology* **2009**, *37*, 1019–1022. [\[CrossRef\]](#)
25. Rivera-Cornejo, R. *Participación en Los Trabajos Finales de Interpretación de Isótopos de Plomo de Los Pórfidos de Oro, Cobre y Molibdeno del Norte del Perú—Región Cajamarca*; Universidad de Sao Paulo: São Paulo, Brazil, 2008.
26. Loewy, S.L.; Connelly, J.N.; Dalziel, I.W. An Orphaned Basement Block: The Arequipa-Antofalla Basement of the Central Andean Margin of South America. *Geol. Soc. Am. Bull.* **2004**, *116*, 171–187. [\[CrossRef\]](#)
27. Lechtman, H. The Production of Copper-Arsenic Alloys in the Central Andes: Highland Ores and Coastal Smelters? *J. F. Archaeol.* **1991**, *18*, 43–76. [\[CrossRef\]](#)
28. Macfarlane; Andrew W. Petersen, U. Pb Isotopes of the Hualgayoc Area, Northern Peru: Implications for Metal Provenance and Genesis of a Cordilleran Polymetallic Mining District. *Econ. Geol.* **1990**, *85*, 1303–1327. [\[CrossRef\]](#)
29. Kontak, D.J.; Cumming, G.L.; Krstic, D.; Clark, A.H.; Farrar, E. Isotopic Composition of Lead in Ore Deposits of the Cordillera Oriental, Southeastern Peru. *Econ. Geol.* **1990**, *85*, 1584–1603. [\[CrossRef\]](#)
30. Mukasa, S.B.; Vidal C, C.E.; Injoque-Espinoza, J. Pb Isotope Bearing on the Metallogenesis of Sulfide Ore Deposits in Central and Southern Peru. *Econ. Geol.* **1990**, *85*, 1438–1446. [\[CrossRef\]](#)
31. Gunnesch, K.A.; Baumann, A.; Gunnesch, M. Lead Isotope Variations across the Central Peruvian Andes. *Econ. Geol.* **1990**, *85*, 1384–1401. [\[CrossRef\]](#)

32. Vidal C, C.E.; Injoque-Espinoza, J.; Sidder, G.B.; Mukasa, S.B. Amphibolitic Cu-Fe Skarn Deposits in the Central Coast of Peru. *Econ. Geol.* **1990**, *85*, 1447–1461. [[CrossRef](#)]
33. Puig, A. Geologic and Metallogenic Significance of the Isotopic Composition of Lead in Galenas of the Chilean Andes. *Econ. Geol.* **1988**, *83*, 843–858. [[CrossRef](#)]
34. Tilton, G.R.; Barreiro, B.A. Origin of Lead in Andean Calc-Alkaline Lavas, Southern Peru. *Science* **1980**, *210*, 1245–1247. [[CrossRef](#)] [[PubMed](#)]
35. Killick, D.J.; Stephens, J.A.; Fenn, T.R. Geological Constraints on the Use of Lead Isotopes for Provenance in Archaeometallurgy. *Archaeometry* **2020**, *62*, 86–105. [[CrossRef](#)]
36. De Ceuster, S.; Hoogewerff, J.; Degryse, P. Provenancing Ancient Materials with Lead Isotopes: Overlap Uncovered. *Sci. Rep.* **2025**, *15*, 4628. [[CrossRef](#)] [[PubMed](#)]
37. Zartman, R.E.; Doe, B.R. Plumbotectonics—the Model. *Tectonophysics* **1981**, *75*, 135–162. [[CrossRef](#)]
38. Guédron, S.; Tolu, J.; Delaere, C.; Sabatier, P.; Barre, J.; Heredia, C.; Brisset, E.; Campillo, S.; Bindler, R.; Fritz, S.C.; et al. Reconstructing Two Millennia of Copper and Silver Metallurgy in the Lake Titicaca Region (Bolivia/Peru) Using Trace Metals and Lead Isotopic Composition. *Anthropocene* **2021**, *34*, 100288. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.