A Robust GDF-ML Framework for Dynamic Grade Modeling: Adaptive Resource Estimation in Complex Porphyry Systems

Liwei Yan *

Posted Date: 25 March 2026

doi: 10.20944/preprints202603.1955.v1

Keywords: porphyry copper deposit; geological distance field (GDF); machine learning; spatial fingerprinting; grade estimation; industrial automation

 doi:10.20944/preprints202603.1955.v1

Article

A Robust GDF-ML Framework for Dynamic Grade Modeling: Adaptive Resource Estimation in Complex Porphyry Systems

Liwei Yan

MCC Tongsin Resources Ltd., Beijing 100028, China; lwyan.cugb@gmail.com

Abstract

Porphyry copper deposits are strategic resources characterized by complex mineralization and high spatial non-stationarity. Traditional linear estimation methods like Ordinary Kriging often fail due to "smoothing effects" and a reliance on manual partitioning. This study proposes the GDF-ML framework, which integrates Geological Distance Fields (GDF) with machine learning to achieve high-fidelity grade modeling. By utilizing cKDTree for efficient spatial indexing, absolute coordinates are translated into Signed Distance Fields (SDF), creating "Spatial Fingerprints" that encode topological relationships between samples and geological entities. Results demonstrate that GDF-ML significantly outperforms Ordinary Kriging, increasing the R^2 score from 0.3080 to 0.7696 on an independent test set while accurately reproducing the "barren core" and irregular high-grade zones. Furthermore, SHAP analysis validates the model's decision-making logic, aligning "intra-domain gain" effects with established metallogenic theories. This framework provides a scalable, automated methodology for dynamic resource evaluation, eliminating the inefficiencies of traditional workflows and supporting the digital evolution of modern smart mines.

Keywords: porphyry copper deposit; geological distance field (GDF); machine learning; spatial fingerprinting; grade estimation; industrial automation

1. Introduction

As the backbone of the global copper and molybdenum supply chains, porphyry deposits hold an irreplaceable strategic position due to their massive resource scale[1]. However, driven by the interaction between multi-stage magmatism and hydrothermal pulsations, these deposits often undergo highly complex mineralization, resulting in severe non-stationary fluctuations of ore grades in 3D space[2]. This complexity, characterized by grade gradients surrounding intrusions interwoven with evolving alteration sequences, poses significant challenges to traditional static models aiming to accurately characterize spatial heterogeneity[3–5].

In the face of such pronounced non-stationarity, linear estimation methods like Ordinary Kriging[6], which are anchored in the "second-order stationarity" hypothesis, reveal significant limitations. Because these algorithms assume spatial homogeneity in both mean and variogram structures, they inevitably fall into the "smoothing effect"[7] trap when dealing with ore bodies characterized by extreme grade gradients and local anisotropy. This not only distorts the spatial morphology of high-grade enrichment zones but also fails—at a fundamental algorithmic level—to reconstruct the local geological textures dictated by metallogenic evolution.

In conventional geostatistical practice, addressing such complexities often necessitates a resort to cumbersome "domaining strategies."[8,9] By forcibly partitioning the deposit into numerous independent geological domains and characterizing parameters individually, one may marginally capture local heterogeneity; however, the entire workflow is not only protracted but also heavily tethered to the subjective expertise of the interpreter. This "labor-intensive" modeling paradigm

 doi:10.20944/preprints202603.1955.v1

struggles under severe computational pressure and human interference, ultimately manifesting as a bottleneck that stifles the objective and rapid evaluation of mineral resources.

To address these deficiencies, this study develops a robust GDF-ML modeling framework. The pivotal breakthrough lies in the introduction of the Geological Distance Field (GDF), which essentially "translates" qualitative geological priors into gradient feature vectors directly interpretable by machine learning algorithms. By extracting the Signed Distance Function (SDF) from arbitrary spatial coordinates to key geological interfaces (e.g., intrusive centers or ore-controlling structures), this approach quantitatively identifies complex topological nesting features while ensuring global consistency of geological logic. This transformation from "qualitative patterns" to "quantitative variables" endows the algorithm with a spatial intuition akin to that of an experienced geologist, enabling dynamic grade inference deeply informed by the topological constraints of geological bodies.

Compared to traditional workflows, the GDF-ML framework significantly mitigates the path dependency on manual intervention, elevating resource evaluation to a new dimension of objectivity. Through deep coupling with implicit modeling mechanisms, this approach autonomously extracts the anisotropic characteristics of mineralization directly from global geological patterns. This advancement entirely bypasses the cumbersome steps of conventional variogram fitting, leading to a quantum leap in modeling efficiency. As GDF feature vectors can be reconstructed synchronously with the evolution of geological bodies, the framework inherently supports real-time iterative updates of resource models, providing a robust technological foundation for the digital evolution and high-precision exploration of modern mines.

2. Geological Characteristics and Data Sources

2.1. Geological Characteristics

The deposit is located within the Chagai magmatic arc of Pakistan. Its metallogenic process was strictly governed by a centrally symmetric hydrothermal fluid system, resulting in a classic upright hollow cylindrical morphology[10,11]. Mineralization is predominantly localized around a central barren porphyry stock; as fluids migrated outward, a series of regular concentric zones developed around the core. In terms of spatial alteration sequences, the deposit exhibits pronounced centripetal zonation: the core consists of a high-temperature, intensely mineralized potassic zone characterized by biotite and K-feldspar assemblages, which progressively grades outward into meso-to-low temperature phyllic and propylitic zones[12]. The mineralization intensity is highly coupled with the alteration facies, with ore grades exhibiting a stepwise decrease from the center to the periphery as the alteration intensity wanes.

On a local scale, the primordial zonation often deviates from perfect geometric symmetry. Driven by distinct episodic metallogenic events, successive pulses of hydrothermal fluids superimposed onto pre-existing mineralized domains, creating complex enrichment patterns through sustained fluid-rock interaction. This superposition not only disrupted the continuity of the original grade field but also facilitated the evolution of multiple asymmetric enrichment centers with diverse orientations and scales. The interplay between macro-zonation, localized clustering, and multi-phase overprinting creates a highly complex architecture. Consequently, the internal grade distribution exhibits extreme non-stationarity, inducing intricate local anisotropic gradients and erratic evolution trends throughout the 3D space.

2.2. Data Sources and Processing

To support the grade modeling of this complex ore body, this study integrated 157 high-reliability exploration drill holes(Figure 1). The layout of these existing engineering works objectively aligns with the upright cylindrical morphology of the deposit: inclined boreholes effectively captured the anisotropic gradients at the ore margins, while the remaining vertical holes focused on characterizing the vertical zonation from the potassic to the propylitic zones. Drill-hole depths range

 doi:10.20944/preprints202603.1955.v1

from 57 to 450 meters, ensuring comprehensive coverage of the shallow oxidized zone and the deep primary sulfide zone. Through systematic sampling of approximately 32,000 meters of core intervals and applying a 3-meter compositing interval, a grade database comprising 10,488 sample points was constructed. This sample size is sufficient to facilitate the deep mining of localized clustering features by machine learning algorithms, ensuring the robustness of the model's generalization capability under extremely non-stationary conditions.

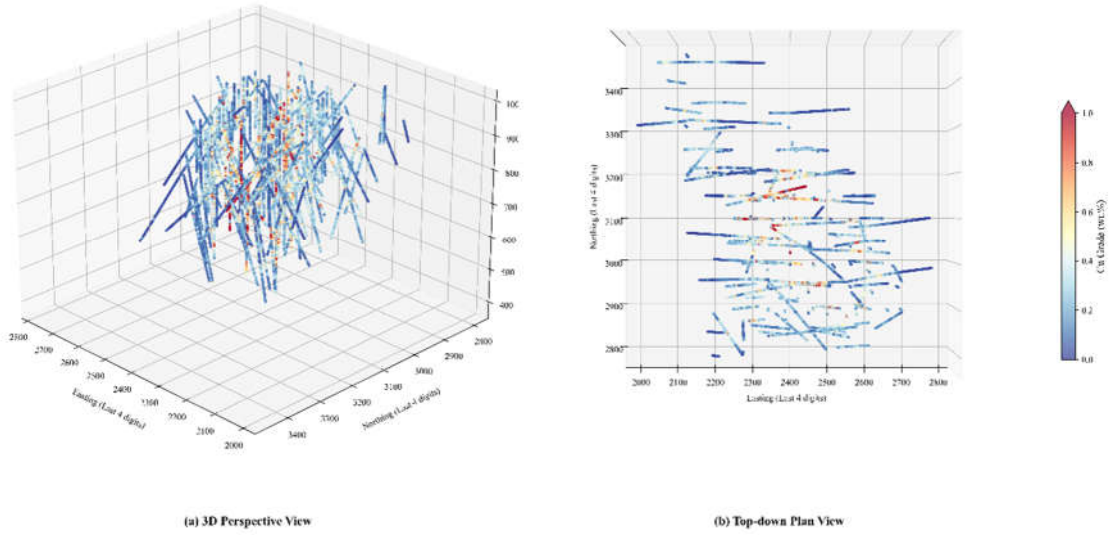


Figure 1. Spatial distribution of drill hole data and Copper (Cu) grades within the study area. (a) 3D perspective view showing the orientation and depth of the drill holes. (b) Top-down plan view illustrating the spatial coverage across Easting and Northing coordinates. The color scale indicates the Cu grade (wt.%), highlighting the high-grade mineralization zones (red) concentrated in the central portion of the drilling campaign.

3. Methods

3.1. Implicit Modeling

The cornerstone of this work lies in the integration of porphyry bodies and mineralized domains into a unified dynamic mathematical framework via implicit modeling techniques[13,14]. Rather than employing traditional sectional tie-lining, this study directly extracts boundary constraints from drill-hole records, treating the primary geological entities and their paragenetic sequences as a topological whole. This enables implicit function fitting to accurately anchor the deposit's macro-skeleton, providing a robust spatial foundation for subsequent grade modeling. Building upon this foundation, and informed by grade evolution trends and alteration zonation logic, the concentric mineralized domains within the porphyry system are further decoupled. These domains are no longer treated as static boundaries; instead, their generated geometric surfaces directly inject directional vector field constraints into the GDF. This mechanism ensures that the interpolation logic remains sensitive to the porphyry's morphology and the mineralization trends, enabling the dynamic calibration of spatial grade distributions.

By defining geological interfaces as continuous analytical functions rather than rigid manual geometric entities, implicit modeling endows the model with exceptional real-time update capabilities. In production environments, as infill drilling data are continuously integrated, the algorithm bypasses the cumbersome manual editing process, automatically refining function parameters to achieve instant boundary adjustments and topological reconstruction. This self-adaptive evolution logic not only provides a stable, continually updated geological foundation for the GDF-ML framework but also demonstrates superior responsiveness when addressing complex

 doi:10.20944/preprints202603.1955.v1

anisotropy and multi-phase mineralization overprinting. Ultimately, it offers an uninterrupted spatial guidance for the refined production of digital mines.

3.2. cKDTree Algorithm

To tackle the massive computational challenges inherent in feature extraction, this study integrates a cKDTree[15] spatial indexing mechanism to resolve high-dimensional retrieval hurdles within complex geological environments. By implementing recursive partitioning of 3D space, the cKDTree constructs a dynamically balanced binary tree architecture, successfully compressing the computational overhead of nearest-neighbor queries from linear growth to logarithmic complexity (O(logN)). This algorithmic advancement enables the system to instantaneously lock onto topological correlations within milliseconds, even when processing tens of thousands of sample points and millions of estimation grid nodes. Such an efficient indexing foundation provides robust computational support for the real-time calculation of GDF features and large-scale grade predictions.

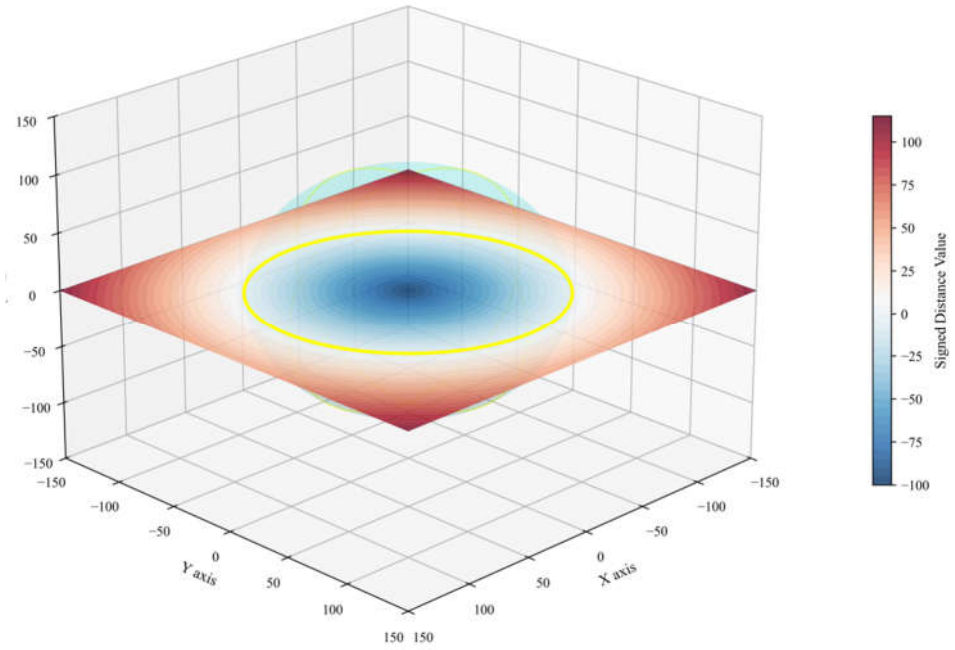
During the GDF-ML model training phase, the core objective of cKDTree is to construct the Signed Distance Field (SDF), which is defined as follows:

$$d_{SDF}(P)=\mathrm{sgn}(P)\cdot \min_{P_{\mathrm{bound}}\in\partial\Omega} \|P-P_{\mathrm{bound}}\| \tag{1}$$

where $\mathrm{sgn}(P)$ is the sign function (or orientation sign function), defined as follows:

$$\mathrm{sgn}(P)=\begin{cases}1 & P\in\Omega_{ext}\ \text{(exterior)}\\ 0 & P\in\partial\Omega\ \text{(boundary)}\\ -1 & P\in\Omega_{int}\ \text{(interior)}\end{cases} \tag{2}$$

The construction of such a Signed Distance Field (SDF) not only quantifies the spatial positioning between sample points and geological entities but also introduces the pivotal geological concept of gradients through its implicit attributes(Figure 2). Here, the gradient does not function as an isolated computational operator; rather, it exists as an inherent directional characteristic of the SDF scalar field. It naturally aligns with the normal direction—either toward or away from the porphyry interface—representing the vector path of the most intense grade attenuation. By learning these implicit gradient attributes, the GDF-ML framework autonomously perceives the anisotropic trends evolving from the mineralization center to its periphery. This enables the machine learning algorithm to precisely capture the grade evolution logic constrained by the morphology of geological domains, without the necessity of explicitly predefined variograms.



 doi:10.20944/preprints202603.1955.v1

Figure 2. 3D visualization of the Signed Distance Field (SDF) and its implicit gradients. The scalar field quantifies the spatial distance from the porphyry interface, with the color scale representing the Signed Distance Value.

3.3. Methods of Machine Learning

3.3.1. Random Forest

As a potent non-linear regression tool within the Ensemble Learning framework, Random Forest[16] constructs a vast array of independent decision trees in parallel through Bootstrap Aggregating, utilizing the ensemble mean as the final output. When confronted with the high-heterogeneity data environments characteristic of porphyry deposits, the algorithm demonstrates exceptional robustness and noise immunity. It adeptly deconstructs the grade-zonation logic constrained by geological domains from the spatial dimensions provided by the Signed Distance Field (SDF). Inherently resilient to high-dimensional features and conducive to parallel acceleration, Random Forest circumvents the mathematical constraints of traditional variograms. This enables the handling of throughput for million-scale estimation grids, providing an efficient computational pathway to reconstruct the anisotropic distributions steered by complex geological entities.

3.3.2. XGBoost

XGBoost stands at the forefront of gradient boosting algorithms[17,18]. Its core logic involves iteratively constructing new trees to precisely capture the residuals of preceding models, thereby achieving a progressive approximation of the target values. By incorporating second-order Taylor expansion information into the objective function and embedding regularization terms, XGBoost effectively mitigates the risk of overfitting while maintaining superior predictive accuracy. Compared to Random Forest, this algorithm exhibits higher sensitivity in scenarios with distinct spatial trends, such as sharp grade attenuation along SDF gradients. Furthermore, its intrinsic sparsity-aware mechanism adeptly handles feature gaps caused by the uneven distribution of drill holes, making it a formidable tool for detecting and characterizing subtle pulsatory features within geological zonation.

3.3.3. CatBoost

CatBoost, a high-performance gradient boosting framework developed by Yandex[19], is distinguished by its ability to natively process categorical features, effectively bypassing the high-dimensional sparsity issues typically associated with one-hot encoding[20]. In geological modeling, lithological codes or geological domain IDs can be directly integrated as input variables. Leveraging its unique Ordered Boosting mechanism, CatBoost mitigates gradient bias and significantly enhances the model's generalization robustness. When coupled with the Signed Distance Field (SDF), CatBoost excels at capturing non-linear grade discontinuities at geological interfaces while maintaining exceptional computational efficiency, making it particularly well-suited for industrial-scale modeling constrained by complex geological domains.

3.3.4. Hyperparameter Optimization via Optuna

To fully unlock the predictive potential of the models, this study employs the Optuna framework based on Bayesian Optimization principles[21]. Distinct from the exhaustive and blind grid search, Optuna "learns" from previous trial trajectories, utilizing the TPE (Tree-structured Parzen Estimator) sampler[22] to continuously contract the search space and locate the optimal parameters within a significantly condensed timeframe. During the optimization process, the system monitors key indicators such as decision tree depth and learning rate in real-time, leveraging a pruning mechanism to decisively terminate underperforming trial branches. This entirely data-driven, closed-loop tuning ensures that the model precisely aligns with the intricate non-linear features of geological domains

 doi:10.20944/preprints202603.1955.v1

while effectively mitigating overfitting risks, thereby enhancing the reliability of mineral resource estimation.

3.4. SHAP Analysis

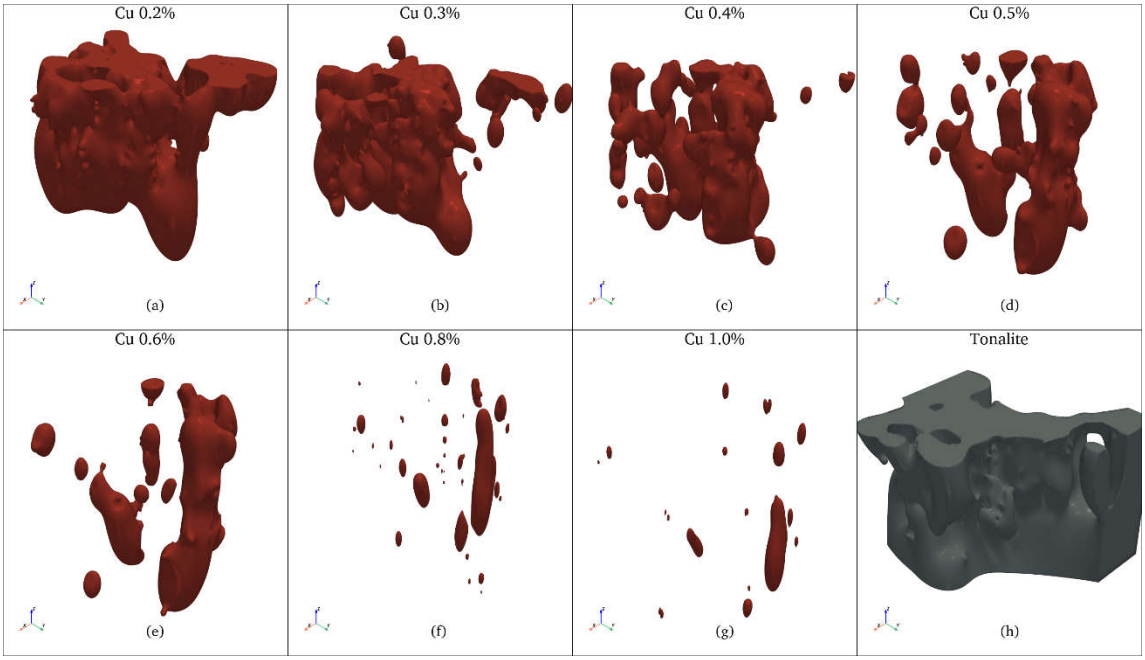
To quantify the marginal contribution of spatial features to grade prediction and enhance model interpretability, this study introduces the SHAP (SHapley Additive exPlanations) framework based on cooperative game theory. By calculating the Shapley Value for each input feature, SHAP transforms the "black-box" predictive process of machine learning into a quantifiable attribution analysis[23–25].

In the context of geological modeling, SHAP is primarily utilized to verify the consistency between model inference logic and established geological principles. By observing the contribution trends of grade predictions as they vary with GDF, it is possible to intuitively determine whether the model has captured the natural law of mineralization attenuation as it moves away from the porphyry interface. This feature-response analysis provides robust support for the geological plausibility of the model. Furthermore, SHAP identifies the priority ranking of various GDF features in their contribution to the final grade, thereby revealing the spatial influence weights of dominant ore-controlling factors.

4. Model Training and Evaluation

4.1. Implicit Modeling and GDF Construction

Given the wide grade range and complex spatial evolution characteristic of the porphyry copper deposit in the study area, a multi-threshold constraint strategy was employed. This involved the simultaneous construction of an implicit model for the porphyry intrusion(Figure 3h) and seven nested mineralization domains (Figure 3a-g, with grade thresholds of 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.8%, and 1.0%, respectively). On this basis, the cKDTree algorithm was utilized to efficiently generate eight types of Geological Distance Fields (GDF) as primary input features. The feature naming convention strictly adheres to geological semantics: mineralization domain-related features are designated as NDTS (Nearest Distance to Shell) with the grade threshold as a suffix (e.g., NDTS02), while the intrusion-related feature is defined as NDTI (Nearest Distance to Intrusion).



 doi:10.20944/preprints202603.1955.v1

Figure 3. 3D mineralization envelope models of the deposit. (a–g) 3D envelopes of Copper (Cu) mineralization at increasing cut-off grades from 0.2% to 1.0%, showing the internal distribution of high-grade cores. (h) 3D solid model of the Tonalite host rock providing the geological context for the mineralized envelopes.

Statistical analysis and feature sensitivity testing indicate that following the introduction of the Geological Distance Field (GDF), the marginal contribution of original spatial coordinates (X, Y, Z) to model decisions becomes negligible. From a mechanistic perspective, raw coordinates often induce a "Coordinate Memorization" effect, where the algorithm mechanically fits the absolute positions of sampling points rather than deciphering the underlying genesis of grade distribution[9,26–30]. Such fitting lacks intrinsic correlation with the spatial structural characteristics of mineralization evolution and constitutes a classic redundant feature.

Furthermore, the weights of categorical parameters—such as lithology and alteration—remain insignificant within the model. This is attributed to the fact that categorical features possess only binary "yes/no" attributes, providing step-wise discrete information that fails to represent the continuous spatial evolution of mineralization intensity. In contrast, GDF, as a numerical parameter, meticulously characterizes grade gradients through a continuous distance vector field. Eliminating these low-weight redundant parameters not only mitigates bias caused by multicollinearity but also compels the algorithm to focus on excavating the core driving logic of geological entities on grade evolution.

4.2. Model Construction and Validation Strategy

Leveraging the high-dimensional geological feature fields described above, this study developed grade prediction models based on Random Forest and two high-performance gradient boosting frameworks: XGBoost and CatBoost. The experimental dataset was partitioned into a training set and an independent test set at a ratio of 8:2. During the training phase, 5-fold cross-validation was employed for model training and hyperparameter optimization. The goodness-of-fit and interpolation accuracy were assessed using the coefficient of determination (R^2) of the full training set and the mean R^2 of the validation folds. The independent test set, which remained entirely unseen during the training process, was utilized for the final evaluation to objectively assess the generalization performance of the optimized models in unknown spatial regions.

Table 1. Regression Performance Metrics and Optuna Optimization Results for Each Model.

Model	Test_R^2_CV	Full_Data_R2	Best_Params
Random Forest	0.749	0.844	"n_estimators": 462, "max_depth": 9, "min_samples_leaf": 4
CatBoost	0.744	0.796	"iterations": 1221, "depth": 4, "learning_rate": 0.0130, "l2_leaf_reg": 2.761
XGBoost	0.731	0.846	"n_estimators": 573, "max_depth": 3, "learning_rate": 0.0581, "subsample": 0.740

Based on a comprehensive comparison of the experimental results across the three algorithms (Table 1), this study selects Random Forest (RF) as the core algorithm for grade estimation in this deposit, for the following reasons:

First, Random Forest demonstrates superior and stable generalization performance. The experimental results show that the cross-validation score for RF (Test_R^2_CV = 0.749) ranks highest among all algorithms. In contrast, while XGBoost achieves a slightly higher training accuracy (0.846), its validation score (0.731) exhibits a notable decline, suggesting that gradient boosting algorithms are more prone to local overfitting when processing high-dimensional GDF features. RF, through

 doi:10.20944/preprints202603.1955.v1

stochastic sampling of both feature and sample spaces, exhibits enhanced noise immunity and robustness.

Second, the Random Forest algorithm is highly compatible with the continuous Geological Distance Fields (GDF) constructed in this study. Since GDF features (e.g., NDTS, NDTI) are smoothly evolving numerical sequences, the Bagging-based ensemble strategy of RF effectively smoothes the predictive fluctuations of individual decision trees, generating continuous grade zonation that aligns with geological common sense. Furthermore, hyperparameter optimization revealed that RF reaches an optimum at a tree depth of 9, indicating that the algorithm can capture the non-linear logic of grade evolution with a relatively parsimonious structure, ensuring higher reliability in complex porphyry environments.

Crucially, despite significant differences in learning mechanisms, all algorithms demonstrate a consistent high-precision convergence when driven by GDF features. This provides objective evidence that GDF has successfully replaced traditional spatial coordinates and discrete categories, accurately capturing the geometric topological logic underlying mineralization. This geology-constrained feature representation elevates the model from mere "data fitting" to a profound "characterization of metallogenic patterns," ensuring that the modeling results possess both mathematical precision and geological rigor.

To further evaluate the model's robustness in unknown spatial regions, this study conducted an independent validation of the optimized model using a hold-out test set. The results (Figure 4) reveal a high correlation between observed and predicted copper grades, with an independent test set R^2 of 0.7696, a Mean Absolute Error (MAE) of only 0.0788, and a remarkably low Bias of 1.33%.

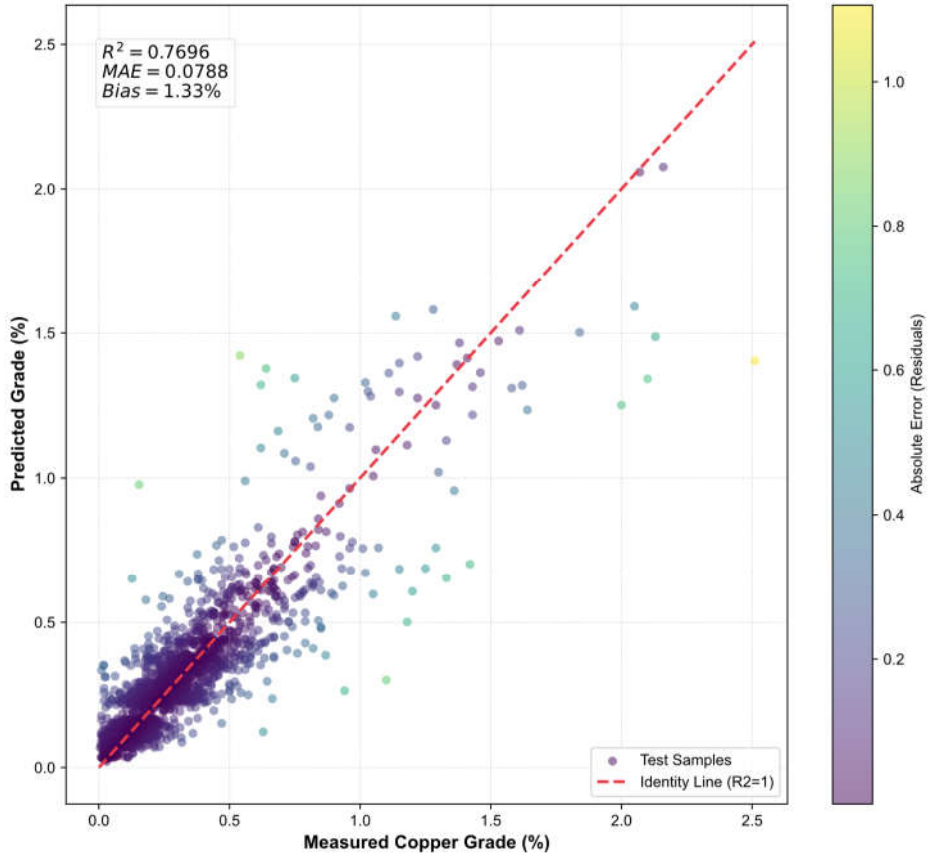


Figure 4. Performance evaluation of the copper grade prediction model. The scatter plot compares Measured Copper Grade (%) against Predicted Grade (%) for the test dataset. The red dashed line represents the Identity Line (R^2 = 1), indicating perfect prediction. Data points are color-coded by their Absolute Error (Residuals), demonstrating high model precision (R^2 = 0.7696, MAE = 0.0788) particularly in the low-to-mid grade ranges.

 doi:10.20944/preprints202603.1955.v1

The distribution characteristics of the scatter plot show that data points are densely clustered around the y=x line, indicating excellent predictive consistency across diverse grade intervals. Notably, despite the presence of a few high-grade outliers, the model exhibits no significant systematic bias. This underscores the superior generalization robustness achieved by integrating the Geological Distance Field (GDF) as a primary driver. Such high-precision independent validation proves that this method effectively overcomes the limitations of spatial sampling density, providing a reliable computational foundation for grade prediction in the deep and peripheral zones of the deposit.

4.3. SHAP Interpretation

SHAP feature contribution analysis (Figure 5) reveals that the mineralization domain distance fields (NDTS) exert an absolute dominance in the model's decision-making process, with NDTS04 and NDTS02 emerging as the pivotal variables driving grade prediction. This attribution result demonstrates clear geological plausibility: the NDTS, through "multi-threshold nesting," precisely quantifies the geometric topological relationships between sample points and various grade shells. By prioritizing features associated with mid-to-low grade boundaries, the model successfully captures the spatial pattern of concentric zonation typical of porphyry copper deposits, where grades evolve outward from the mineralization center. This alignment between data-driven feature importance and geological principles underscores the scientific validity of the model's inferential path.

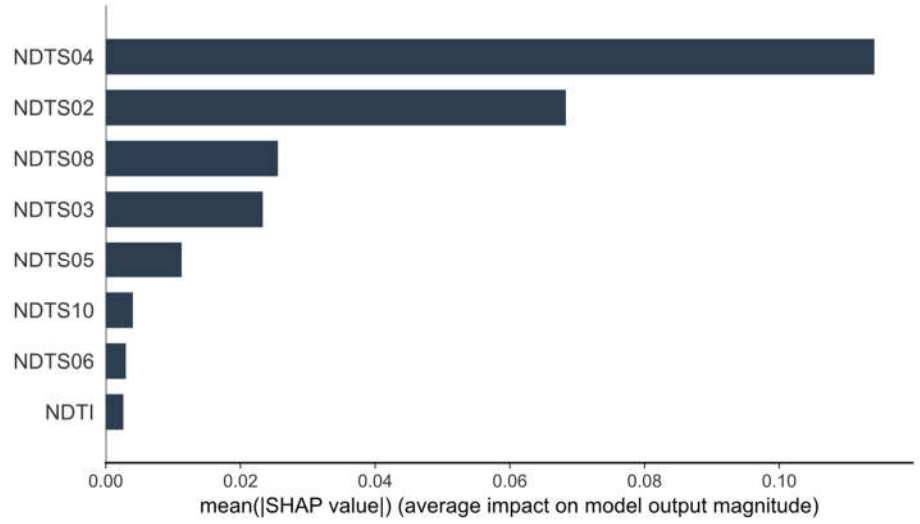


Figure 5. Feature importance analysis based on SHAP values. The bar chart ranks the input features according to their mean(|SHAP value|), representing their average impact on the model output magnitude. Features such as NDTS04 and NDTS02 exhibit the highest global importance, indicating they are the primary drivers in the prediction of copper mineralization within the GDF-ML framework.

The SHAP Beeswarm plot clarifies the non-linear response mechanism between GDF features and grade predictions(Figure 6). NDTS04 (the 0.4% grade domain distance field) exhibits the strongest positive driving effect: a large cluster of blue points (low feature values, representing positions inside the 0.4% mineralization domain) is concentrated in the positive SHAP zone, showing the largest positive displacement (range). This indicates that the model has successfully identified the interior of the 0.4% shell as the core contribution zone for grade gain; once a spatial position "enters the domain," the predicted value undergoes a significant uplift.

 doi:10.20944/preprints202603.1955.v1

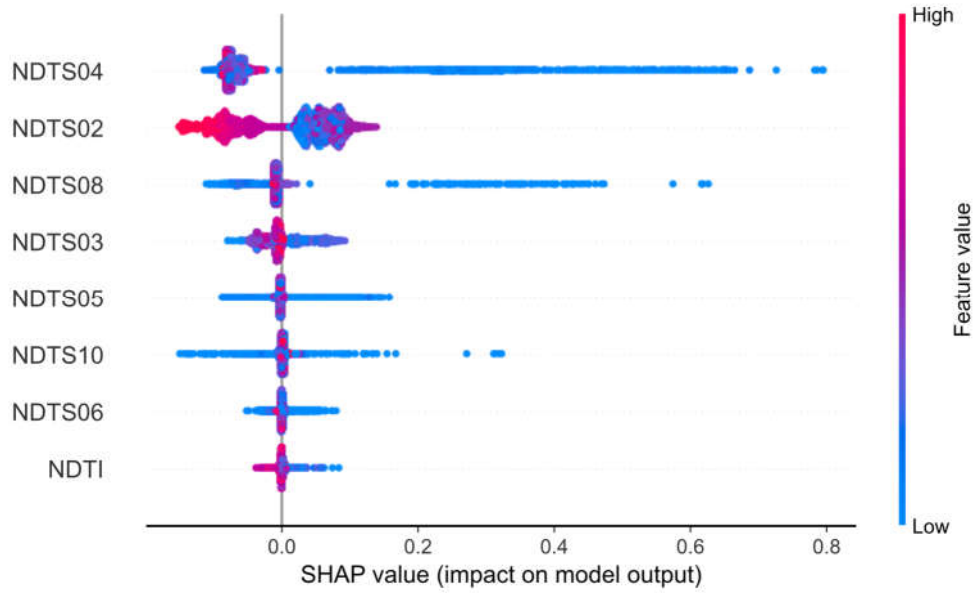


Figure 6. SHAP beeswarm plot of feature contributions. The plot displays the impact of each feature on the model output, with colors representing feature values (blue to red). Beyond the primary NDTS04, features like NDTS08 and NDTS03 show a "high-resolution" character, where the wide dispersion of blue clusters (intra-domain samples) indicates their critical role in capturing complex grade variations within the mineralization envelopes.

Complementing this is the "boundary constraint" effect of NDTS02 (the 0.2% grade domain distance field): its red points (high feature values, representing positions far from or outside the 0.2% boundary) are densely distributed in the negative SHAP zone, exerting a clear inhibitory effect on the grade. The logical combination of "intra-domain gain (blue points of NDTS04)" and "extra-domain reduction (red points of NDTS02)" precisely simulates the gradient characteristics of porphyry copper deposits, where grades rapidly attenuate from the core enrichment zone to the peripheral low-grade zones. Furthermore, the long-tail distribution observed in high-threshold features such as NDTS08 at minimum values (blue points) further proves the model's acute ability to capture localized high-grade ore bodies (enrichment centers).

The SHAP Beeswarm plot further elucidates the differentiated functions of various mineralization domain features in grade refinement. Beyond the dominant NDTS04, features such as NDTS08 and NDTS03 exhibit a distinctive "intra-domain high-resolution" character. The blue clusters representing intra-domain samples (low feature values) are widely dispersed across both sides of the SHAP axis, forming long tails or broad-spectrum oscillations.

This indicates that these features serve not only as basic boundary constraints but, more importantly, possess a robust capacity to distinguish intra-domain grade heterogeneities. By performing refined attribute decomposition on internal samples, the model can acutely capture complex grade fluctuations within the mineralization centers and transition zones. This combination of "coarse-frame constraint" and "fine-grained identification" provides the logical foundation for the model's high-precision estimation.

In contrast, red points representing extra-domain samples (high feature values) are highly concentrated near the zero SHAP baseline, exhibiting minimal contribution weights. This asymmetric distribution pattern objectively validates the design intent of the GDF feature fields: once the model determines a sample is outside a specific mineralization domain, it is immediately categorized as background; however, once a sample enters the domain, the model activates a gradient network formed by multi-threshold distance fields to perform refined characterization of internal mineralization intensity. This "outward-exclusion, inward-refinement" decision mechanism ensures that the model precisely filters out non-mineralized backgrounds while deeply extracting detailed

 doi:10.20944/preprints202603.1955.v1

information from high-grade zones, thereby enhancing the geological interpretability of the modeling results.

Feature sensitivity analysis indicates that the influence of the intrusion distance field (NDTI) is significantly lower than that of the mineralization domain distance fields (NDTS). This phenomenon stems from the spatial containment relationship between the ore body and the intrusion: since most high-grade ore is hosted within the intrusion, NDTI provides only a coarse macroscopic positioning and struggles to capture the complex internal grade fluctuations. Near the mineralization center, the NDTI values of samples are highly homogeneous, leading to insufficient spatial resolution for distinguishing grade variations, causing its contribution to be overshadowed by the more granular NDTS features.

Furthermore, the diminished weight of NDTI reveals the limitations of geometric control exerted by the intrusive body. While the intrusion provides the spatial host, the actual grade evolution is driven by deep dynamic factors such as hydrothermal fluid flow and structural fracturing, leading to a decoupling between mineralization morphology and intrusion boundaries. By directly delineating the topological structure of grades, the NDTS feature fields more acutely reflect the internal spatial textures of mineralization. The model's preferential selection of NDTS over NDTI provides objective evidence that "mineralization domain constraints" possess higher geological resolution and predictive accuracy than simple "petrological constraints" when characterizing the complex zonation patterns of porphyry systems.

The SHAP dependence plot matrix(Figure 7) reveals a striking "step-wise" transition across the zero-point interface of each mineralization domain feature (NDTS). As the distance field value shifts from positive to negative—marking the sample's transition into the domain—the contribution to the predicted grade peaks instantaneously, followed by non-linear fluctuations or decay toward the domain's interior. This precise lock-on at the interface demonstrates the model's exceptional sensitivity to geometric boundaries, effectively capturing the truncation effects of mineralization at spatial limits.

The essence of this interfacial sensitivity lies in the algorithm's ability to capture "gradient alignment" within the feature space. In porphyry systems, the spatial evolution of grade is a complex scalar field constrained by geological structures. The multi-threshold GDF fields constructed in this study predefine a set of geometric gradients aligned with the ore body's topology. When the geometric vectors of the GDF achieve spatial coupling with the physical gradients of grade evolution, the model rapidly captures the most intense mineralization fluctuations along the normal direction. This transforms qualitative geological interface identification into a feature-driven gradient response.

From a higher theoretical perspective, this identification of gradient alignment mathematically deconstructs the singularity, non-linearity, and non-stationarity of mineralization. Mineralization is essentially the abrupt accumulation of elements at phase-change interfaces (singularity), and its patterns vary drastically across space (non-stationarity), mechanisms that traditional linear models struggle to address. Through non-linear mapping of multi-threshold features, machine learning precisely captures the logic of energy release and material deposition of mineralizing fluids at specific physicochemical boundaries. This signifies the model's elevation from simple data fitting to knowledge-embedded pattern representation, ensuring that the final modeling results possess high mathematical precision while profoundly restoring the intrinsic dynamical characteristics of the metallogenic system.

 doi:10.20944/preprints202603.1955.v1

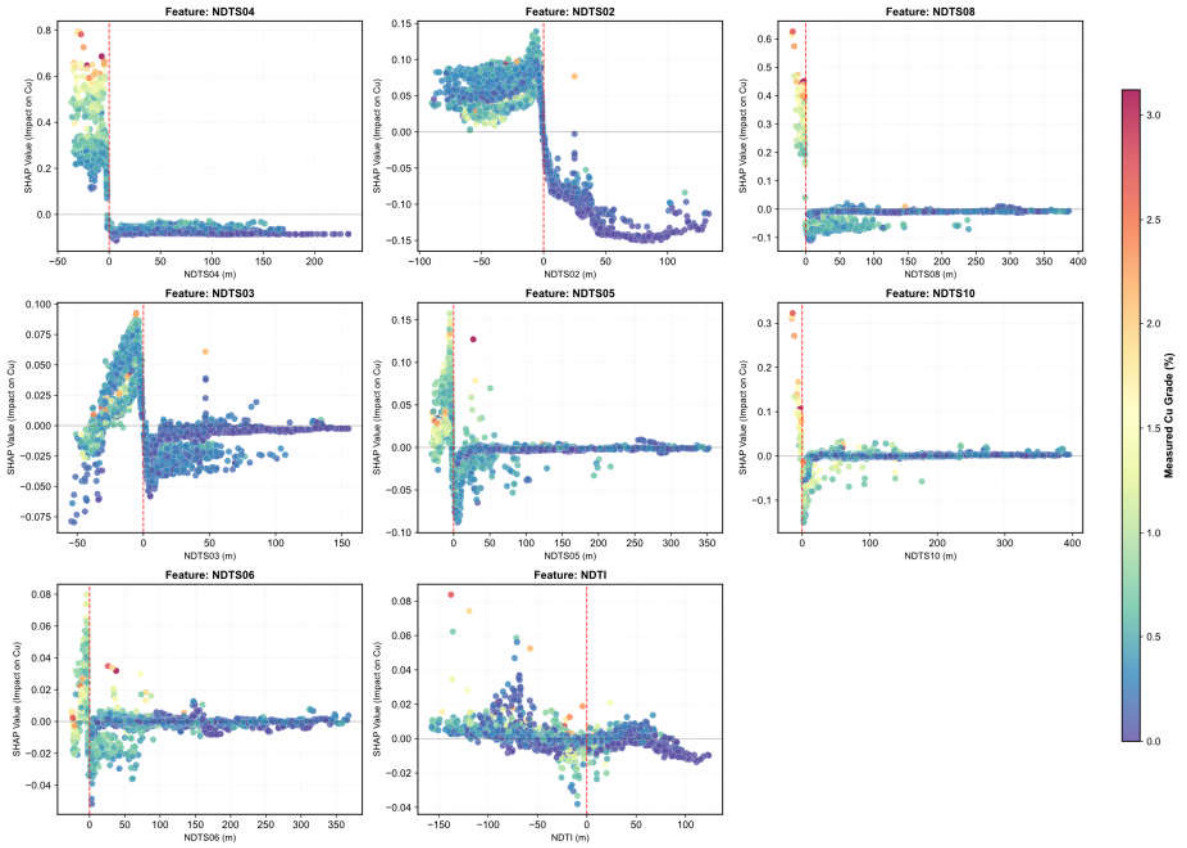


Figure 7. SHAP dependence plots for various mineralization domain features. The matrix illustrates the relationship between feature values (distance in meters) and their impact on Cu grade (SHAP value). A prominent "step-wise" transition is observed at the zero-point interface (red dashed lines), where the model's contribution peaks as samples enter the domain. This indicates the model's high sensitivity to geometric boundaries and its ability to capture non-linear grade evolution relative to geological interfaces.

To further dissect the spatial response of key features to grade prediction, this study selected a representative high-grade sample primarily driven by the NDTS04 feature for detailed attribution analysis (observed grade: 0.630%).

Within the SHAP decomposition path (Waterfall Plot) for this sample, NDTS04 exerts a dominant explanatory weight(Figure 8). With a feature value of -7.06 (indicating a position within the mineralization domain), its marginal contribution to the prediction reaches +0.38, effectively elevating the predicted value from the global mean into the high-grade range. From a geospatial perspective, this phenomenon clearly captures the geometric characteristics of the central mineralization zone: the negative value of NDTS04 signifies a tight spatial coupling with the core ore-forming interface. The model has learned to identify this specific location as a "favorable window" for high-intensity copper enrichment..

Although features such as NDTS08 provide a minor negative correction (-0.07), the absolute dominance of NDTS04 ensures the final predicted value (0.695%) accurately locks onto the high-grade anomaly. This explanatory outcome—characterized by single-feature dominance and multi-feature synergy—not only quantifies the contribution of spatial location to mineralization but also confirms that the model has internalized the geological logic of "increased grade with proximity to the ore body center." This case provides robust evidence of the model's exceptional discriminative power and interpretability in identifying high-grade spatial targets.

 doi:10.20944/preprints202603.1955.v1

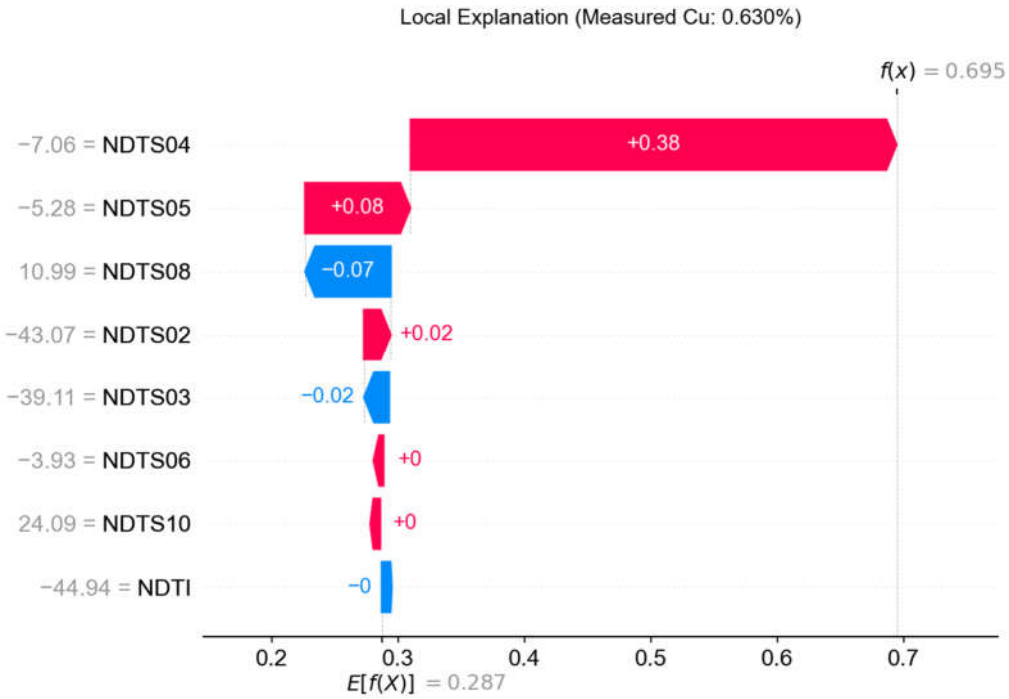


Figure 8. SHAP waterfall plot for a local sample explanation. The plot decomposes the prediction process for a single sample (Measured Cu: 0.630%). Starting from the base value E[f(X)] = 0.287, each horizontal bar shows how individual features either increase (red) or decrease (blue) the final predicted grade f(x) = 0.695. In this instance, the proximity to the NDTS04 interface (value: -7.06) provides the most significant positive contribution (+0.38) to the estimated grade.

4.4. Grade Model Reconstruction and 3D Spatial Analysis

A 3D block model with a unit cell size of 10m × 10m × 10m was constructed across the deposit as the baseline for estimation. Leveraging the cKDTree algorithm, the Geological Distance Field (GDF) values relative to geological entities were efficiently mapped for each block center. These GDF features were then processed through the optimized Random Forest (RF) model to achieve fine-grained 3D grade interpolation and reconstruction.

Deep Restoration of Macroscopic Spatial Zonation: By visualizing voxels with predicted grades above 0.2% (Figure 9a), the model demonstrates its exceptional capacity to restore the spatial distribution characteristics of the porphyry copper deposit. The 3D visualization clearly outlines the macroscopic contours and microscopic details of the ore body, successfully reproducing the classic "barren core" phenomenon typical of porphyry systems[1,31] while illustrating systematic concentric mineralization zones through delicate color gradients. This reproduction of spatial topology proves the model's ability to transcend complex non-linear geological variables and accurately capture the diffusion and precipitation patterns of mineralizing fluids.

Refined Decomposition of High-Grade Core Zones: When the visualization threshold is increased to above 0.35% (Figure 9b), the model resolves the intricate structures of high-grade core zones. Unlike the smoothing effects common in traditional interpolation algorithms, this model accurately restores irregular geometric morphologies with high geological authenticity, delineating tectonic-controlled meandering, branching, and discrete mineralization centers. This capacity for "in-situ reconstruction" of high-value zones ensures that spatial heterogeneity is preserved, reflecting the accumulation trajectories of fluids within preferential pathways.

 doi:10.20944/preprints202603.1955.v1

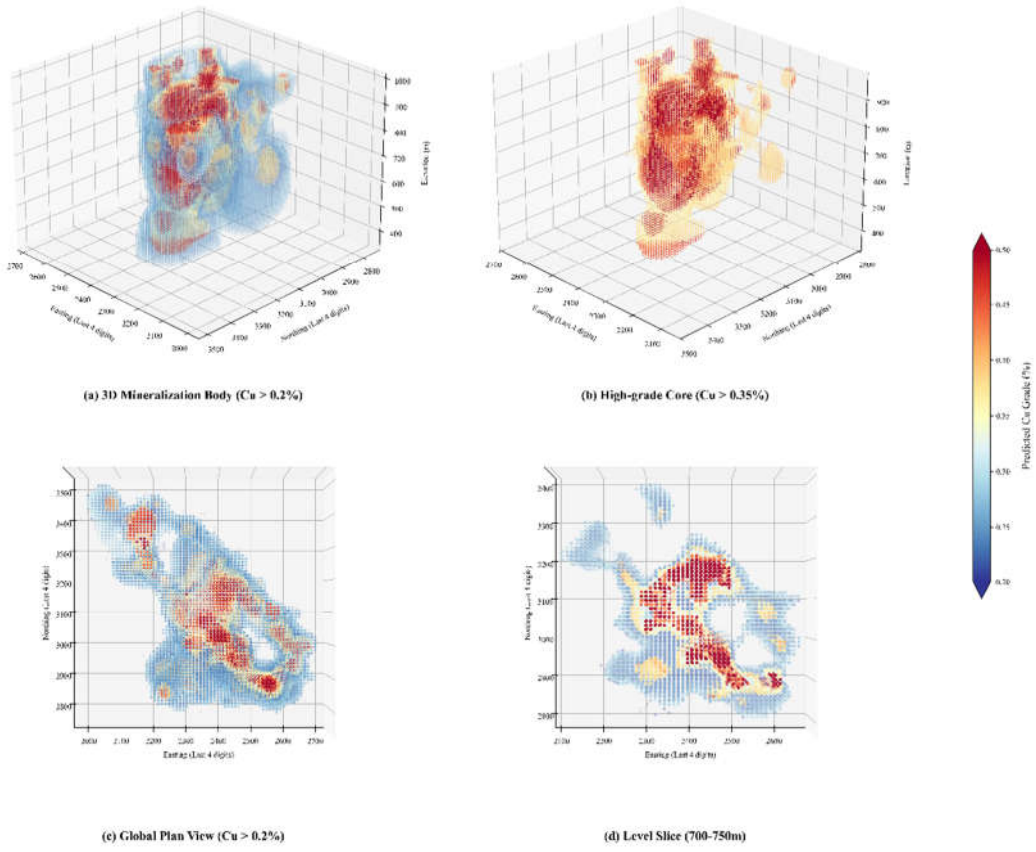


Figure 9. 3D block model and spatial distribution of predicted Cu grades. (a) 3D view of the estimated mineralization body at a 0.2% Cu cut-off grade. (b) Visualization of the high-grade core (>0.35% Cu), showing the concentration of mineralization within the center of the domain. (c) Global plan view projected across Easting and Northing coordinates. (d) Horizontal level slice (700–750 m elevation) revealing the internal grade distribution and continuity. The color bar indicates the predicted Cu grade (wt.%), ranging from low (blue) to high (red).

Multi-scale Consistency Validation: Through the integration of plan views (Figure 9c) and elevation slices (Figure 9d), the model further validates the strike continuity and local resolution of the ore body. The global perspective reveals the broad-scale continuity of mineralization in horizontal space, while local dissections exhibit the refined characteristics of mineralization centers at the 700–750m elevation. Such multi-scale performance—from global trends to local details—demonstrates that even with limited deep sampling points, the model provides high-resolution, geologically consistent predictions, achieving high-fidelity reconstruction of complex geological entities.

5. Discussion

5.1. Semantic Translation of Spatial Features and Principle-Driven Modeling

The core logic of the GDF-ML framework lies in utilizing Signed Distance Fields (SDF) to deconstruct and reconstruct absolute Euclidean spatial coordinates into topological relationships relative to geological entities. Essentially, this achieves the quantitative representation of a geological expert's a priori spatial logic. This process facilitates a dimensional leap from "absolute geographic location" to "geoscientific semantic features": it discards raw coordinate values that lack relevance to metallogenic principles, replacing them with distance vector fields constrained by topology. Consequently, the model transcends the non-stationary noise inherent in spatial sampling, precisely

 doi:10.20944/preprints202603.1955.v1

capturing the field gradients of grade evolution along geological boundaries.Because this topological logic possesses spatial translation invariance, the model achieves a predictive accuracy of $R^2 = 0.75$ based solely on quantified mineralization-controlling laws, while strictly avoiding spatial data leakage. Unlike traditional models that rely on the mechanical memorization of specific geographic points, the GDF-ML framework mitigates the interference of the "coordinate memorization effect" on generalization through logical mapping. This modeling paradigm shift from "data-driven" to "principle-driven" not only mathematically bolsters the model's robustness during geological extrapolation but also provides interpretable geoscientific logic for the intelligent estimation of complex porphyry deposits.

5.2. Adaptive Anisotropic Characterization via Spatial Fingerprinting

The GDF-ML framework demonstrates a remarkable capacity for adaptive anisotropic characterization, rooted in the "Spatial Fingerprints" assigned to each sample by the SDF. Unlike traditional interpolation algorithms that rely on pre-defined static search parameters (e.g., variogram search ellipsoids), this framework encodes the 3D geometric constraints—such as a sample's proximity to the intrusion center, contact zones, and mineralization boundaries—directly into unique semantic features through the nonlinear superposition of multi-dimensional SDF fields. This "spatial fingerprint" is essentially a unique identifier within the metallogenic topological system; it not only meticulously defines the 3D position but also implies the geometric evolution patterns radiating from the mineralization center at that specific location.

This fingerprint-driven mechanism facilitates an implicit adaptive fit for complex mineralization anisotropy. By identifying these "spatial fingerprints," the machine learning model automatically perceives the local geological environment—such as high-gradient zones near the mineralization core or stable zones along the domain peripheries—and dynamically adjusts the local field gradients accordingly. In mineralization domains with distorted morphologies or intricate zonation, the model is no longer restricted by global stationary anisotropy assumptions. Instead, it adjusts its weight distribution in real-time according to the geometric fluctuations of geological entities. This paradigm shift from "static geometric search" to "dynamic fingerprint recognition" ensures that estimation results remain highly coupled with 3D mineralization zonation, significantly enhancing the spatial resolution for complex heterogeneous deposits.

5.3. Global Consistency Modeling and Workflow Optimization Analysis

The GDF-ML framework achieves true Global Integrated Modeling while maintaining consistency with established geological principles. Traditional modeling workflows often necessitate tedious "Spatial Partitioning" by geologists, requiring separate variogram fitting and manual configuration of local search parameters for different lithologies or grade zones[8]. This human-heavy process is not only inefficient but also prone to producing discontinuous "edge effects" at the boundaries of different domains, thereby disrupting the inherent logic of the metallogenic system.

Logic-based Adaptation Replacing Manual Partitioning: The "spatial fingerprints" generated via GDF mapping provide automated constraints across the entire domain, significantly reducing the need for manual intervention. Based on the topological logic provided by the GDF, the model automatically identifies and processes the spatial correlation patterns and anisotropic characteristics of different geological zones within a single unified framework. Consequently, fragmented fitting for individual mineralization zones is no longer required; the algorithm spontaneously comprehends the evolutionary variances between the mineralization core and its periphery.

Mitigating Human Bias and Computational Redundancy: Global integrated modeling avoids statistical biases typically generated at the junctions of multiple sub-models, ensuring consistency in the final estimation results. This "end-to-end, automated, and globalized" training paradigm not only streamlines the processing of complex mine-scale datasets but also eliminates uncertainties introduced by the subjective experience of different geologists. This transition from "localized fragmented fitting" to "global consistent inference" highlights the logical advantages of the GDF-ML

 doi:10.20944/preprints202603.1955.v1

framework in handling complex porphyry systems and establishes a methodological foundation for future automated 3D mineral resource estimation.

5.4. Potential for Industrial Integration and Future Outlook

The industrial advantage of the GDF-ML framework lies in its seamless alignment with existing digital mine production workflows. The framework's frontend leverages mature Implicit Modeling systems, utilizing existing 3D geological geometric models as topological constraints to facilitate the lossless transfer of geological knowledge into algorithmic inputs. At the computational level, the integration of the cKDTree algorithm constructs an efficient spatial index structure, reducing the complexity of feature translation from large-scale sample points to complex geometric boundaries from $O(N^2)$ to $O(N/\log N)$. This ensures millisecond-level response times for millions of voxels. This combination of "model-driven logic + high-efficiency field-based algorithms" not only minimizes R&D costs but also provides the high-throughput performance necessary to support real-time dynamic reserve updates during active mining operations.Furthermore, the GDF-ML framework establishes an "End-to-End" automated pipeline, providing an engineering pathway for the dynamic evaluation of mineral resources. Given its global integrated modeling nature, the entire process—from multi-dimensional GDF feature extraction to non-linear grade mapping—can be encapsulated into standardized automated operators. This paradigm significantly compresses the tedious manual partitioning and variogram fitting cycles traditional in classical estimation, drastically shortening the reserve verification period. By incorporating Explainable AI (XAI) tools like SHAP, the framework successfully transforms the machine learning "black box" into traceable, quantifiable geoscientific evidence. This technical architecture—balancing logical robustness, computational efficiency, and decision transparency—establishes a solid methodological foundation for dynamic resource supervision and precision decision-making in the era of Intelligent Mining[32].

5.5. Modeling Paradigm Efficiency Comparison: Spatial Fingerprinting Inference vs. Linear Interpolation

A profound comparison of performance metrics between the GDF-ML framework and traditional Ordinary Kriging (OK)[6,7,30] clearly highlights the technical bottlenecks of classical geostatistical methods in handling complex deposits with high spatial heterogeneity(Table 2). The coefficient of determination (R^2) for Ordinary Kriging reached only 0.3080, a stark contrast to the 0.7696 achieved by the GDF-ML framework. This significant performance gap is not merely a numerical lead; it reveals a fundamental divergence in the spatial cognitive logic between the two modeling paradigms.

Firstly, the failure of Ordinary Kriging stems from its over-reliance on the assumption of "spatial stationarity." As a distance-weighted linear smoother, the OK method is intrinsically limited to characterizing spatial autocorrelation through variograms. It struggles to decipher grade mutations and complex non-stationary zonation characteristics triggered by multiple intrusive phases in porphyry systems. Consequently, when confronted with the intense gradient fluctuations evolving from the mineralization core to its periphery, the OK method produces a severe "smoothing effect." Its RMSE (0.2137) is considerably higher than that of our framework (0.1233), demonstrating its inherent inability to accurately restore local high-frequency details and the extreme-value representations of the ore body.

Table 2. Comparison of Estimation Results between the GDF-ML Framework and Ordinary Kriging on the Independent Test Set.

Item	Random Forest	Ordinary Kriging
R^2 Score	0.7696	0.3080
RMSE	0.1233	0.2137
Mean Cu Predicted	0.2916	0.2927
Mean Cu Observed	0.2877	0.2827

 doi:10.20944/preprints202603.1955.v1

Secondly, the success of GDF-ML proves the necessity of transitioning from 'physical distance calculation' to 'topological semantic inference.' During the estimation process, Kriging lacks the perception of geometric constraints from diverse geological entities, reducing complex metallogenic logic to simple Euclidean distance weighting. In contrast, this study, by constructing 'spatial fingerprints,' precisely anchors sampling points within a logical coordinate system governed by the overall geological framework. This enables the model to identify and adapt to anisotropic evolution trends more effectively. Comparative data indicates that when the model interprets the topological position of samples relative to the mineralized domains through GDF fields, predictive accuracy achieves a leapfrog improvement of nearly 150%.

In conclusion, the statistical failure of Ordinary Kriging further corroborates the necessity of introducing spatial logical translation into complex ore body estimation. The GDF-ML framework maintains exceptional robustness and spatial resolution even while strictly excluding the interference of "coordinate memorization." This signifies that the intelligent estimation paradigm centered on "spatial fingerprints" possesses unparalleled logical advantages and engineering value over traditional linear models when tackling high-heterogeneity tasks, such as deep-seated resource evaluation.

6. Conclusions

This study addressed the challenges of high spatial heterogeneity, pronounced non-stationarity, and the severe "smoothing effect" inherent in traditional grade estimation for complex porphyry copper deposits by innovatively constructing an intelligent estimation framework integrating Geological Distance Fields (GDF) and Machine Learning (ML). Based on the in-depth data mining and empirical analysis of the study area, the following core conclusions are drawn:

Paradigm Shift in Feature Engineering: The research demonstrates that "Spatial Fingerprints" constructed via Geological Distance Fields (GDF) successfully translate absolute Euclidean coordinates into topological logical relationships with geoscientific semantics. This approach effectively circumvents the mechanical memorization of coordinate values by machine learning models, achieving a paradigm shift from "data-driven" to "principle-driven" modeling. This significantly enhances the model's extrapolation capability and robustness under complex geological constraints.

Dual Breakthrough in Accuracy and Fidelity: Comparative experiments indicate that the GDF-ML framework achieved an R^2 of 0.7696 on the independent test set, vastly outperforming the 0.3080 of Ordinary Kriging. The model not only accurately restored macroscopic zonation features, such as the "barren core" and concentric mineralization typical of porphyry systems, but also captured high-frequency signals to delineate irregular high-grade core zones. This effectively overcomes the inherent "smoothing effect" of traditional linear interpolation algorithms.

Validation of Model Interpretability: Utilizing SHAP (Explainable AI) tools, this study quantitatively revealed the contribution weights of different mineralization domain boundaries to model decisions. The "intra-domain gain" and "boundary constraint" effects of NDTS04 and NDTS02 accurately mirror the geological attenuation of mineralizing fluids from the core to the periphery, confirming that the model's decision path possesses substantial geoscientific plausibility.

Broad Industrial Application Prospects: Combined with the cKDTree spatial indexing algorithm, the framework achieved millisecond-level feature mapping for millions of data points. This architecture, which integrates computational efficiency, logical transparency, and global consistency modeling, is deeply compatible with existing digital mine production workflows. It provides a solid methodological foundation for dynamic resource evaluation and precision decision-making in the context of Intelligent Mining.

In summary, the GDF-ML framework not only provides a high-precision technical tool for digital modeling of complex deposits but also explores a new path for the deep coupling of Artificial Intelligence with geoscientific knowledge, offering significant academic value and engineering utility.

 doi:10.20944/preprints202603.1955.v1

Funding: This research received no external funding.

Data Availability Statement: The data is available on request.

Acknowledgments: We thank the anonymous reviewers for their constructive comments.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

GDF-ML	Geological Distance Field-Machine Learning
SDF	Signed Distance Field

References

1. Sillitoe, R.H. Porphyry Copper Systems. *Economic Geology* **2010**, *105*, 3–41.
2. Afzal, P.; Alghalandis, Y.F.; Khakzad, A.; Moarefvand, P.; Omran, N.R. Delineation of mineralization zones in porphyry Cu deposits by fractal concentration–volume modeling. *Journal of Geochemical Exploration* **2011**, *108*, 220–232, doi:10.1016/j.gexplo.2011.03.005.
3. James, C.; David, R.C.; John, L.W.; Holly, S. Geology, Mineralization, Alteration, and Structural Evolution of the El Teniente Porphyry Cu-Mo Deposit. *Economic Geology* **2005**, *100*, 979–1003.
4. Lewis, B.G.; Jorge, Q.G. Patterns of Mineralization and Alteration below the Porphyry Copper Orebody at El Salvador, Chile. *Economic Geology* **1995**, *90*, 2–16.
5. Liu, H.; Wang, Q.; Zhang, C.; Lou, D.; Zhou, Y.; He, Z. Spatial pattern and dynamic control for mineralization in the Pulang porphyry copper deposit, Yunnan, SW China: Perspective from fractal analysis. *Journal of Geochemical Exploration* **2016**, *164*, 42–53, doi:10.1016/j.gexplo.2015.07.001.
6. Cressie, N. Spatial Prediction and Ordinary Kriging. *Mathematical Geology* **1988**, *20*, 405–421.
7. Yamamoto, J.K. Correcting the Smoothing Effect of Ordinary Kriging Estimates. *Mathematical Geology* **2005**, *37*, 69–94, doi:10.1007/s11004-005-8748-7.
8. Emery, X.; Ortiz, J.M. Estimation of Mineral Resources Using Grade Domains: Critical Analysis and a Suggested Methodology. *The Journal of The South African Institute of Mining and Metallurgy* **2005**, *105*, 247–256.
9. Maleki, M.; Mery, N.; Soltani-Mohammadi, S.; Plaza-Carvajal, J.; Varouchakis, E.A. Integrating Geological Domains into Machine Learning for Ore Grade Prediction: A Case Study from a Porphyry Copper Deposit. *Minerals* **2025**, *15*, 1175, doi:10.3390/min15111175.
10. Hong, J.; Khalil, Y.S.; Narejo, A.A.; Yang, X.; Khan, T.; Wang, Z.; Tang, H.; Zhang, H.; Yang, B.; Li, W. Magmatic Evolution at the Saindak Cu-Au Deposit: Implications for the Formation of Giant Porphyry Deposits. *Minerals* **2025**, *15*, 768, doi:10.3390/min15080768.
11. Wang, L.; Zheng, Y.; Hou, Z.; Xue, C.; Yang, Z.; Shen, Y.; Li, X.; Ghaffar, A. The subduction-related Saindak porphyry Cu-Au deposit formed by remelting of a thickened juvenile lower crust underneath the Chagai belt, Pakistan. *Ore Geology Reviews* **2022**, *149*, 105062, doi:10.1016/j.oregeorev.2022.105062.
12. Rose, A.W. Zonal Relations of Wallrock Alteration and Sulfide Distribution at Porphyry Copper Deposits. *Economic Geology* **1970**, *65*, 920–936.
13. Basson, I.J.; Anthonissen, C.J.; McCall, M.J.; Stoch, B.; Britz, J.; Deacon, J.; Strydom, M.; Cloete, E.; Botha, J.; Bester, M.; et al. Ore-structure relationships at Sishen Mine, Northern Cape, Republic of South Africa, based on fully-constrained implicit 3D modelling. *Ore Geology Reviews* **2017**, *86*, 825–838, doi:10.1016/j.oregeorev.2017.04.007.
14. Wang, J.; Zhao, H.; Bi, L.; Wang, L. Implicit 3D Modeling of Ore Body from Geological Boreholes Data Using Hermite Radial Basis Functions. *Minerals* **2018**, *8*, doi:10.3390/min8100443.
15. Gilberto, G.; Rodrigo, T.A.; Monica, C. cKdtree: a Compact Kdtree for Spatial Data. In Proceedings of the AMW 2024: 16th Alberto Mendelzon International Workshop on Foundations of Data Management, Mexico City, Mexico, 2024.
16. Leo, B. Random Forests. *Machine Learning* **2001**, *45*, 5–32.

 doi:10.20944/preprints202603.1955.v1

17. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016; pp. 785–794.
18. Zhang, P.; Jia, Y.; Shang, Y. Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks* **2022**, *18*, doi:10.1177/15501329221106935.
19. Liudmila, P.; Gleb, G.; Aleksandr, V.; Anna Veronika, D.; Andrey, G. CatBoost: unbiased boosting with categorical features. *32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.* **2018**.
20. Rodríguez, P.; Bautista, M.A.; Gonzàlez, J.; Escalera, S. Beyond one-hot encoding: Lower dimensional target embedding. *Image and Vision Computing* **2018**, *75*, 21–31, doi:10.1016/j.imavis.2018.04.004.
21. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter OptimizationFramework. In Proceedings of the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019; pp. 2623–2631.
22. Watanabe, S. Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance. *arXiv preprint* **2023**, arXiv:2304.11127, doi:10.48550/arXiv.2304.11127.
23. Chen, Y.; Chen, B.; Shayilan, A. Combining categorical boosting and Shapley additive explanations for building an interpretable ensemble classifier for identifying mineralization-related geochemical anomalies. *Ore Geology Reviews* **2024**, *173*, 106263, doi:10.1016/j.oregeorev.2024.106263.
24. Eyal, W. The Shapley Value. In *Handbook of Garne Theory, Volume 3*; 2002; pp. 2027–2054.
25. Zhang, M.; Wang, X.; Chen, C.; Ding, J.; Zhou, X.; Qu, J. Interpretable ore classification using SHAP-enhanced LightGBM: A case study from the Qiaomaishan deposit, China. *Applied Computing and Geosciences* **2025**, *28*, 100295, doi:10.1016/j.acags.2025.100295.
26. Jairo, M.-A.; Marco, C.-T.; Jose, M.-Q.; Eduardo, N.-V.; Juan, V.-G.; Juan, C.-G. Copper Ore Grade Prediction using Machine Learing Techniques in a Copper Deposit. *Journal of Mining and Environment* **2024**, *15*, 1011–1027, doi:10.22044/jme.2024.14032.2617.
27. Jafrasteh, B.; Fathianpour, N.; Suárez, A. Comparison of machine learning methods for copper ore grade estimation. *Computational Geosciences* **2018**, *22*, 1371–1388, doi:10.1007/s10596-018-9758-0.
28. Kaplan, U.E.; Dagasan, Y.; Topal, E. Mineral grade estimation using gradient boosting regression trees. *International Journal of Mining, Reclamation and Environment* **2021**, *35*, 728–742, doi:10.1080/17480930.2021.1949863.
29. Kaplan, U.E.; Topal, E. A New Ore Grade Estimation Using Combine Machine Learning Algorithms. *Minerals* **2020**, *10*, 847, doi:10.3390/min10100847.
30. Maniteja, M.; Samanta, G.; Gebretsadik, A.; Tsae, N.B.; Rai, S.S.; Fissha, Y.; Okada, N.; Kawamura, Y. Advancing Iron Ore Grade Estimation: A Comparative Study of Machine Learning and Ordinary Kriging. *Minerals* **2025**, *15*, 131, doi:10.3390/min15020131.
31. Sillitoe, R.H. The Tops and Bottoms of Porphyry Copper Deposits. *Economic Geology* **1973**, *68*, 799–815.
32. Li, J.-g.; Zhan, K. Intelligent Mining Technology for an Underground Metal Mine Based on Unmanned Equipment. *Engineering* **2018**, *4*, 381–391, doi:10.1016/j.eng.2018.05.013.