

ARTICLE

A Probabilistic Framework for Sedimentary Stratigraphic Modeling Using Anisotropic Indicator Kriging and Entropy-Based Uncertainty Quantification

Liyang Wang¹, Lianjiang Yang¹, Guang He¹, Dong Bai¹, Xingxing Wei^{2,*}, Wen Luo³ and Min Peng³

¹China Railway No.18 Engineering Group Company Limited, Tianjin, China

²School of Smart Construction and Energy Engineering, Hunan Institute of Engineering, Xiangtan, China

³Science and Technology Information Center, Department of Transportation of Hunan Province, Changsha, China

*Corresponding Author: Xingxing Wei. Email: xingxingwei15@163.com

Received: 04 May 2026; Accepted: 21 July 2026; Published: 21 October 2026

ABSTRACT: Sedimentary stratigraphic modeling plays a vital role in geoscience and engineering applications, particularly in oil and gas exploration, mineral resource evaluation, and the planning and construction of subsurface infrastructure. However, accurately reconstructing stratigraphic architecture and quantifying prediction uncertainty remain challenging due to the limited number and sparse distribution of boreholes. This study aims to develop a robust probabilistic modeling framework for accurately characterizing the geometry and spatial distribution of sedimentary lithological categories while quantifying prediction uncertainty. To address this issue, an integrated probabilistic stratigraphic modeling framework based on anisotropic Indicator Kriging (IK) is developed. Multi-class lithological data are transformed into binary indicator variables, and exponential variogram models are fitted to characterize spatial correlations. Ordinary kriging is then applied to estimate the conditional probability distribution of each lithological class. To capture the anisotropic characteristics of sedimentary strata, a coordinate scaling method is introduced to enhance lateral continuity while reducing vertical correlation. Furthermore, probability normalization ensures valid probabilistic interpretation, and information entropy is employed to quantify prediction uncertainty. The proposed method is evaluated using both synthetic and real-world benchmark datasets. Results demonstrate that the IK-based framework can accurately reconstruct layered structures under limited borehole constraints, achieving high predictive accuracy (over 90% in synthetic cases and up to 95.26% in real cases). Compared with Markov Random Field (MRF) and IC-XGBoost methods, the proposed framework shows advantages in boundary delineation, thin-layer preservation, spatial continuity, and probabilistic uncertainty representation, although the improvement in overall accuracy may be moderate in some cases. The entropy-based uncertainty maps effectively identify high-risk and data-sparse regions. In conclusion, the proposed IK-based probabilistic modeling framework is both theoretically sound and practically adaptable. It provides reliable probabilistic predictions and uncertainty quantification, offering an effective tool for geological modeling, drilling optimization, and risk assessment in sedimentary stratigraphic settings.

KEYWORDS: Sedimentary stratigraphic modeling; indicator kriging; uncertainty quantification; anisotropic modeling; exponential variogram model; probabilistic modeling; information entropy

1 Introduction

Sedimentary stratigraphic modeling is a cornerstone of geological and geotechnical research, playing a pivotal role in applications such as oil and gas exploration, mineral resource evaluation, groundwater development, and the planning and construction of subsurface infrastructure [1–3]. Stratigraphic models not only reveal subsurface lithological distributions but also provide critical support for reservoir evaluation,

fluid flow simulation, and engineering risk prediction [4–6]. However, the limited number and sparse distribution of boreholes, combined with complex geological conditions, pose significant challenges in reconstructing accurate subsurface structures from limited observational data. This challenge is particularly pronounced in sedimentary stratigraphic successions characterized by significant layer undulations and thinly interbedded units, where conventional interpolation or modeling methods often struggle to balance accuracy and geological realism [7,8].

Previous studies have shown that integrating stratigraphic concepts with geostatistical modeling can improve the geological realism of facies reconstruction, lithofacies modeling, and uncertainty assessment [9–13]. However, for sparse-borehole sedimentary lithological/lithofacies reconstruction, the joint incorporation of anisotropic continuity, cross-validation-based parameter selection, probability normalization, and entropy-based boundary uncertainty quantification within a unified Indicator Kriging (IK) workflow remains insufficiently addressed. Therefore, this study develops an anisotropic IK-based probabilistic framework for sparse-borehole lithological/lithofacies reconstruction and stratigraphic interpretation.

Although these studies provide important foundations for facies modeling, lithofacies reconstruction, and geological property modeling, several issues remain insufficiently addressed for sparse-borehole sedimentary lithological reconstruction. In particular, the integration of anisotropic sedimentary continuity, cross-validation-based anisotropic parameter selection, multi-class IK probability normalization, and entropy-based boundary uncertainty quantification within a unified probabilistic workflow still requires further investigation. Therefore, the present study builds upon previous stratigraphic and geostatistical modeling studies but focuses on developing an anisotropic IK-based probabilistic framework tailored to sedimentary lithological reconstruction under sparse borehole constraints.

Beyond reconstructing spatial distributions, quantifying prediction uncertainty is equally critical [14–16]. In practice, large distances between boreholes and rapid local changes in lithological boundaries introduce unavoidable uncertainties in model predictions [17,18]. Relying solely on deterministic “hard classification” results often masks potential prediction risks, limiting the model’s utility in optimizing drilling strategies and engineering decision-making [19]. Therefore, incorporating probabilistic predictions and uncertainty quantification into stratigraphic modeling not only highlights variations in prediction reliability across borehole-constrained zones, layer boundaries, and data-sparse areas but also provides a more scientific basis for resource evaluation and risk management [20–23]. Recent lithofacies modeling studies have also emphasized that integrating stratigraphic constraints with geostatistical methods helps improve geological consistency and uncertainty interpretation in data-sparse settings [9].

To address these challenges, various methods have been proposed in academia. For instance, early Markov Random Field (MRF) models [24–26] offer a flexible and intuitive approach to reconstructing stratigraphic structures but require subjective, empirically defined parameters that remain fixed during inference, limiting their ability to account for model bias or uncertainty. Improved MRF models [8] overcome some of these limitations and can efficiently predict stratigraphic patterns in stationary environments; however, their performance may deteriorate in non-stationary geological settings. Multi-point statistics (MPS) algorithms [27–29] excel in pattern reproduction and geological realism but are computationally intensive. Other machine learning methods, such as conditional Gaussian random fields, Markov chain models, XGBoost, and deep learning, have been widely applied to stratigraphic modeling and lithological classification [30–34]. These methods effectively capture nonlinear relationships and perform well in localized areas but fall short in boundary delineation and probabilistic interpretation. In contrast, geostatistical methods remain popular due to their rigorous theoretical foundation and ability to model spatial correlations. Among these, IK stands out as a probabilistic modeling approach that transforms multi-class lithological predictions into multiple binary field estimations [35,36]. Its strengths lie in providing probability distributions for different classes and

enabling uncertainty quantification through statistical measures like entropy, enhancing both interpretability and practical utility.

Previous studies have demonstrated that geostatistical facies modeling benefits from the integration of sedimentological and stratigraphic concepts. For example, Falivene et al. [11] developed hierarchical and grid-based facies reconstruction approaches for heterogeneous sedimentary systems, and Tentori et al. [9] integrated sequence stratigraphy with geostatistical methods for 3D lithofacies modeling. These studies provide important foundations for stratigraphically constrained facies modeling, but they did not specifically focus on the integrated use of anisotropic IK, cross-validation-based parameter selection, probability normalization, and entropy-based boundary uncertainty quantification for sparse-borehole sedimentary lithological reconstruction.

The stratigraphic systems considered in this study are typical sedimentary sequences characterized by pronounced lateral continuity, relatively rapid vertical facies changes, and locally developed thin interbeds or undulating interfaces. These geological patterns are commonly associated with repeated depositional processes, in which lithological units tend to extend continuously in the horizontal direction, whereas their thicknesses and boundaries may vary significantly in the vertical direction. In engineering geological investigations, such strata are usually constrained by sparsely distributed boreholes, making it difficult to reconstruct continuous layer boundaries, identify thin or transitional units between boreholes, and quantify the uncertainty associated with boundary zones and data-sparse regions.

Indicator Kriging has been widely used in geological modeling and categorical variable prediction, including lithological interpolation, mineral deposit boundary modeling, and probabilistic resource assessment [35,36]. Therefore, the contribution of this study does not lie in proposing IK as a completely new geostatistical algorithm. The novelty of this work lies in the problem-oriented integration of anisotropic IK, cross-validation-based parameter selection, probability normalization, and entropy-based uncertainty interpretation into a unified workflow for sparse-borehole sedimentary lithological reconstruction. This study develops an integrated anisotropic IK-based probabilistic stratigraphic modeling framework tailored to sedimentary lithological reconstruction from sparse borehole data. Unlike previous studies that primarily used IK to correct the smoothing effects of ordinary kriging [35] or applied it to mineral deposit boundary modeling with uncertainty maps to identify undersampled areas [36], the proposed framework further incorporates geological anisotropy, cross-validation-based parameter selection, probability normalization, entropy-based uncertainty quantification, and multi-model comparison into a unified workflow.

Specifically, an exponential variogram model is employed to characterize the spatial correlation of lithological indicator variables, and coordinate scaling is introduced to represent the anisotropic continuity of sedimentary strata, i.e., strong lateral continuity and rapid vertical variation. The anisotropic parameters are selected through cross-validation using both Overall Accuracy (OA) and Negative Log-Likelihood (NLL), so that both classification accuracy and probabilistic reliability are considered. In addition, independently estimated class probabilities are normalized to ensure valid probabilistic interpretation, and information entropy is used to identify uncertainty in boundary-related and data-sparse zones. The effectiveness of the proposed framework is validated through synthetic and real-world stratigraphic case studies and compared with MRF and IC-XGBoost models [8,37] in terms of accuracy, boundary delineation, thin-layer preservation, uncertainty representation, and computational efficiency. The results demonstrate that the anisotropic IK-based framework provides a practical and interpretable probabilistic workflow for sedimentary lithological reconstruction, drilling optimization, and geological risk assessment.

2 Stratigraphic Modeling Framework

2.1 Data Representation

To facilitate stratigraphic modeling, the study area is discretized into a two-dimensional regular grid, with each cell corresponding to a unique spatial location. Borehole data provide lithological class labels at a limited number of locations. The set of lithological classes is defined as $\mathbf{L} = \{0, 1, 2, \dots, l\}$, representing distinct lithological categories such as sandstone, mudstone, and silt. At borehole locations, observed lithological labels are available, while non-borehole locations are treated as missing values. In this study, the term “lithological class” refers to the categorical lithology or lithofacies label observed at borehole cells, whereas the term “stratigraphic structure” refers to the spatial arrangement, boundaries, and geometry reconstructed from these class labels. As shown in Fig. 1a, borehole locations are marked with class labels, while non-borehole locations remain unlabeled.

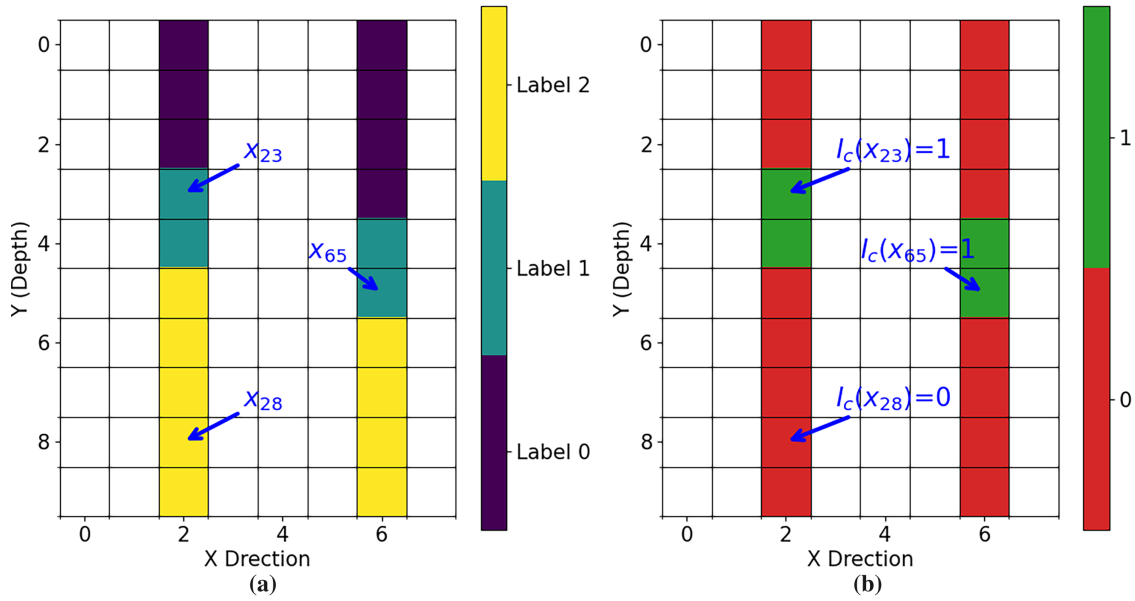


Figure 1: (a) Borehole data labeling; (b) indicator variable values for class 1.

The primary objective of modeling is to estimate the conditional probability of each lithological class at all unknown locations. These probabilities not only enable direct geological interpretation (e.g., identifying the most likely lithological distribution) but also provide a foundation for uncertainty analysis.

To enable probabilistic modeling, discrete lithological classes are transformed into indicator variables. Let $Z(x)$ denote the lithological label at location x . For each class $c \in \mathbf{L} = \{0, 1, 2, \dots, l\}$, an indicator variable is defined as shown in Eq. (1) [38,39]:

$$I_c(x) = \begin{cases} 1, & Z(x) = c \\ 0, & Z(x) \neq c \end{cases} \quad (1)$$

Fig. 1b illustrates the indicator variable map for Class 1 ($Z(x) = 1$) corresponding to Fig. 1a. In Fig. 1a, locations x_{23} and x_{65} have a label of 1, while x_{28} has a label of 2. Consequently, in the indicator variable map, $I_c(x_{23}) = 1$, $I_c(x_{65}) = 1$, and $I_c(x_{28}) = 0$.

The advantage of indicator variables is that they transform a multi-class classification problem into multiple binary field estimations, allowing classical geostatistical methods to be applied directly to model the spatial distribution of each class. Thus, the modeling task reduces to estimating the conditional expectation of the indicator variable field, as shown in Eq. (2):

$$p_c(x) = E[I_c(x) | \text{borehole data}] \quad (2)$$

Here, $p_c(x)$ represents the probability that location x belongs to class c . Since $I_c(x)$ takes values in $[0, 1]$, its expectation directly corresponds to a probabilistic interpretation. For instance, if $E[I_c(x)] = 0.7$, there is a 70% probability that location x belongs to class c .

2.2 Indicator Kriging Framework

IK is a geostatistical method specifically designed for categorical variable modeling. Its core principle involves transforming categorical data into multiple binary indicator variable fields and applying ordinary kriging to each field to estimate spatial probability distributions [38,40]. This approach ensures linear unbiasedness and minimum variance, enabling robust interpolation of lithological class probabilities across space.

For a given class c , the indicator variable estimate at an unknown location x_0 is expressed as Eq. (3) [41]:

$$\hat{I}_c(x_0) = \sum_{i=1}^n \omega_i I_c(x_i) \quad (3)$$

where $I_c(x_i)$ is the indicator value at the i -th known point, ω_i is the corresponding kriging weight, and n is the number of neighboring sample points. This estimate represents a weighted average of known indicator values, with the weights determined by the ordinary kriging system. The weight vector ω is obtained by solving the following linear system:

$$\begin{bmatrix} \Gamma & 1 \\ 1^T & 0 \end{bmatrix} \begin{bmatrix} \omega \\ \mu \end{bmatrix} = \begin{bmatrix} \gamma_0 \\ 1 \end{bmatrix} \quad (4)$$

where Γ is the $n \times n$ semi-variance matrix among the neighboring sample points, and its element is defined as $\Gamma_{ij} = \gamma(\|x_i - x_j\|)$. γ_0 is the $n \times 1$ semi-variance vector between the neighboring sample points and the prediction point, and its i -th element is defined as $\gamma_{0,i} = \gamma(\|x_i - x_0\|)$. μ is the Lagrange multiplier introduced to satisfy the unbiasedness constraint, and 1 is a vector of ones. The last row of Eq. (4) enforces the ordinary kriging constraint $\sum_i \omega_i = 1$. By solving Eq. (4), the kriging weights are obtained and then used to estimate the indicator value at the prediction location.

The weights ω ensure that the predicted value $\hat{I}_c(x_0)$ satisfies the best linear unbiased estimator. For binary indicator variables, the expected value of the prediction directly corresponds to the class probability (see Eq. (5)):

$$p_c(x_0) \approx \hat{I}_c(x_0) \quad (5)$$

Since each class's indicator variable field is modeled independently, the sum of probabilities across all classes at a given location, $\sum_c p_c(x_0)$, may not strictly equal 1. To ensure valid probabilistic interpretation, normalization is applied using Eq. (6):

$$\tilde{p}_c(x_0) = \frac{p_c(x_0)}{\sum_c p_c(x_0)} \quad (6)$$

The normalized probability vector provides a complete probabilistic representation, reflecting the uncertainty distribution of the strata. At borehole locations, where the class is known, probabilities are set to one-hot encoding, with the observed class assigned a probability of 1 and all others set to 0.

2.3 Variogram Estimation and Model Fitting

The spatial correlation of indicator variables is characterized by the semi-variogram, which describes how attribute differences between sample points vary with spatial distance, serving as the core of kriging estimation. For the indicator variable of class c , the empirical semi-variogram is estimated as Eq. (7) [42]:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} (I_c(x_i) - I_c(x_j))^2 \quad (7)$$

where $N(h)$ is the set of point pairs with distances falling within the lag interval $[h - \Delta h/2, h + \Delta h/2]$, and $|N(h)|$ is the number of such pairs. Intuitively, if closely spaced point pairs exhibit small differences, the class shows strong spatial continuity over short distances. Conversely, rapidly increasing differences with distance indicate weaker spatial correlation.

2.3.1 Adaptive Lag Binning Strategy

In practice, the large number of point pairs makes direct computation of semi-variograms computationally expensive and unstable. To address this, we adopt a random sampling and adaptive binning strategy [43,44]: (a) randomly sample a large subset of point pairs; (b) calculate their distance distribution and set the 95th percentile as the maximum lag distance to avoid extreme value effects; (c) divide the interval $[0, h_{max}]$ into N_{lags} equal bins and compute the empirical semi-variogram for each bin. This approach enhances computational efficiency while improving the smoothness and stability of the empirical semi-variogram curve.

2.3.2 Theoretical Variogram Model

For theoretical variogram modeling, we adopt an exponential model, expressed as Eq. (8) [43,44]:

$$\gamma(h) = nugget + sill(1 - e^{-h/a}) \quad (8)$$

where a is the correlation length, reflecting the scale of spatial correlation decay; $sill$ approximates the variance of the indicator variable, representing overall variability; and $nugget$ accounts for micro-scale fluctuations or measurement errors.

The exponential model exhibits monotonic decay, with the semi-variogram increasing rapidly with distance and approaching the $sill$ asymptotically within a finite range. This behavior effectively captures the spatial characteristics of lithological data, where correlations are strong over short distances (within typical borehole spacing) and weaken over longer distances, stabilizing at the global mean. The exponential model is thus both statistically robust and consistent with the spatial continuity observed in sedimentary strata.

2.3.3 Parameter Fitting Method

The empirical semi-variogram is fitted to the exponential model using least-squares optimization to determine the optimal a , $sill$, and $nugget$ parameters. The objective function is shown as Eq. (9) [43,44]:

$$\min_{a, sill, nugget} \sum_h [\hat{\gamma}(h) - \gamma(h; a, sill, nugget)]^2 \quad (9)$$

where $\hat{\gamma}(h)$ is the empirical semi-variogram value, and $\gamma(h)$ is the theoretical value from the exponential model.

The fitted variogram ensures smoothness and positive definiteness while capturing differences in spatial continuity among lithological classes. Specifically, the correlation length a governs the spatial extent of class continuity, while *sill* and *nugget* reflect global variance and micro-scale variability, respectively.

2.3.4 Geological Interpretation and Limitations

By fitting exponential variogram models to indicator variables for each class, the spatial structural characteristics of different lithological categories can be quantitatively revealed. A larger correlation length for a given lithological class indicates stronger spatial continuity, whereas a smaller correlation length suggests a more localized or heterogeneous distribution. These model parameters provide a quantitative basis for subsequent IK prediction and probabilistic lithological reconstruction.

It should be noted that the proposed IK framework treats lithological or lithofacies classes as categorical indicator fields. Therefore, the model does not explicitly impose stratigraphic order rules, vertical succession constraints, or depositional bounding surfaces in the same way as a process-based or sequence-stratigraphic model. Instead, stratigraphic information is incorporated indirectly through borehole-derived lithological labels, spatial correlation structures, and anisotropic coordinate scaling. The reconstructed result should therefore be interpreted as a probabilistic lithological/lithofacies distribution that supports stratigraphic interpretation, rather than as a fully process-based stratigraphic architecture model.

In this study, the term “stratigraphic reconstruction” refers to the reconstruction of the spatial arrangement, boundaries, and geometry of lithological classes from sparse borehole constraints. The anisotropic IK framework is designed to preserve the dominant lateral continuity and rapid vertical variation commonly observed in sedimentary successions. However, explicit incorporation of stratigraphic order, vertical transition rules, and depositional boundary constraints would require additional geological prior information or rule-based constraints, which is beyond the scope of the present work and will be considered in future research.

2.4 Anisotropic Modeling

Sedimentary strata typically exhibit strong lateral continuity and relatively weak vertical correlation [45], driven by sedimentological dynamics where depositional processes promote lateral expansion but are constrained vertically by layering rhythms. To account for this anisotropy, isotropic models must be adjusted to avoid excessive vertical continuity, which would produce geologically unrealistic results.

We adopt a coordinate scaling method to introduce anisotropy. This approach involves linearly scaling the coordinate system before variogram estimation and kriging prediction to adjust the “effective distance” in different directions. Let the original coordinates be (O_x, O_y) , and the scaled coordinates be (O'_x, O'_y) . The transformation is shown as Eq. (10):

$$O'_x = \frac{O_x}{S_x}, O'_y = \frac{O_y}{S_y} \quad (10)$$

where S_x and S_y are the scaling factors for the horizontal and vertical directions, respectively. The purpose of this transformation is to modify the effective distance used in variogram estimation and kriging prediction. When $S_x > S_y$, the horizontal coordinate is compressed more strongly than the vertical coordinate in the scaled space, so horizontal separation contributes less to the effective distance than vertical separation. This leads to stronger lateral correlation and relatively weaker vertical correlation, which is consistent with the

geological feature of sedimentary strata, where lithological units commonly exhibit strong lateral continuity and relatively rapid vertical variation.

For two points $x_i = (x_i, y_i)$ and $x_j = (x_j, y_j)$ in the original coordinate system, the scaled distance used for variogram estimation and kriging prediction is defined as Eq. (11):

$$h'_{ij} = \sqrt{\left(\frac{x_i - x_j}{S_x}\right)^2 + \left(\frac{y_i - y_j}{S_y}\right)^2} \quad (11)$$

where h'_{ij} is the effective distance between points x_i and x_j in the scaled coordinate system. Equivalently, if $\Delta x = x_i - x_j$ and $\Delta y = y_i - y_j$, Eq. (11) can be written as $h'_{ij} = \sqrt{(\Delta x/S_x)^2 + (\Delta y/S_y)^2}$. This scaled distance is then used in the variogram model instead of the original Euclidean distance.

The semi-variogram is then computed based on this scaled distance h'_{ij} . Although an isotropic variogram model (e.g., exponential) is used, this approach effectively introduces an *elliptical anisotropic covariance structure* in the original coordinate system [43,44]. If the correlation length of the exponential model is β , the effective correlation lengths in the original coordinates are shown as Eq. (12):

$$L_x = \beta \cdot S_x, \quad L_y = \beta \cdot S_y \quad (12)$$

By adjusting the scaling factors (S_x, S_y), the ratio of horizontal to vertical continuity can be flexibly controlled. For instance, when $S_x = 8$ and $S_y = 1$, the horizontal correlation length is approximately eight times the vertical, consistent with the “strong lateral continuity, thin vertical thickness” characteristic of sedimentary strata.

In real borehole data, vertical sampling within boreholes is typically high-resolution, while horizontal spacing between boreholes is much larger. Using an isotropic model would overestimate vertical correlation, resulting in unrealistic “vertical streaks” in predictions. The coordinate scaling method promotes lateral extension and rapid vertical decorrelation, yielding predictions that align with the layered nature of sedimentary strata.

2.5 Prediction and Post-Processing

After fitting variogram models for each class's indicator variables, kriging is used to predict probabilities at all unknown locations. For each class c , the conditional expectation is computed by solving the kriging system, and probabilities are interpolated across the entire two-dimensional grid, yielding the spatial probability distribution for class c .

2.5.1 Batch Prediction and Neighborhood Restriction

For large-scale modeling, using global borehole data for kriging at all prediction locations would result in excessively large matrices and computational demands. To address this, we employ a batch prediction strategy, dividing the prediction grid into smaller blocks and performing interpolation block by block to reduce memory usage.

Additionally, to ensure computational efficiency and geological realism, a neighborhood search is introduced. A search radius is defined around each prediction point, and only nearby borehole data are used in the kriging system. As shown in Fig. 2, the black pentagon represents the prediction location, the cyan translucent cell indicates the prediction cell, red-bordered cells denote nearby borehole data identified via KDTree search, the dashed circle represents the search radius, and dashed lines illustrate relationships between the prediction point and neighboring cells. This approach ensures that only local

borehole samples contribute to the kriging system, significantly reducing computational costs while aligning with the geological assumption of stronger spatial correlation over short distances. Similar batch prediction and neighborhood search strategies are widely used in geostatistics [43,46,47].

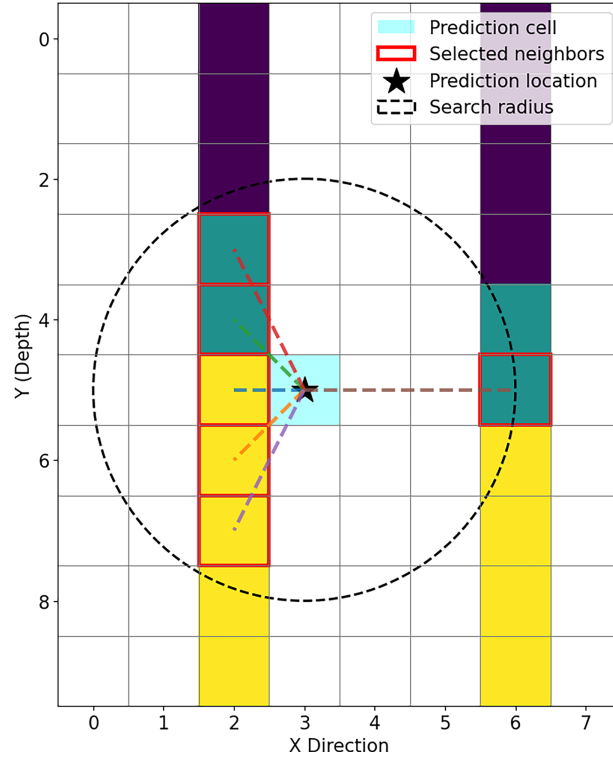


Figure 2: KDTree neighborhood on the borehole grid.

2.5.2 Hard Classification Map

In practical applications, a single-class lithological class distribution is often required. We apply the Maximum A Posteriori (MAP) criterion (see Eq. (13)) for hard classification [8,48]:

$$Z(x) = \arg \max_c (\tilde{p}_c(x)) \quad (13)$$

This result provides the most likely lithological class distribution, suitable as input for geological modeling and numerical simulations.

2.5.3 Uncertainty Quantification

To quantify uncertainty in model predictions beyond hard classification, we introduce probabilistic metrics. The maximum probability value serves as a reliability indicator: values close to 1 indicate high confidence, while values approaching a uniform distribution (e.g., ~ 0.33 for three classes) suggest significant uncertainty.

Additionally, information entropy is used to measure the dispersion of class probabilities. Lower entropy indicates more reliable predictions, while higher entropy suggests that class probabilities are nearly equal, making differentiation challenging [8]. The entropy is defined as Eq. (14):

$$H(x) = - \sum_c \tilde{p}_c(x) \widetilde{\log p_c(x)} \quad (14)$$

By generating heatmaps of maximum probability and entropy distributions, the model's spatial confidence can be visualized, aiding in the identification of high-risk areas.

2.5.4 Accuracy Evaluation

When reference lithological or stratigraphic data are available, prediction accuracy can be quantitatively assessed. We use Overall Accuracy (OA), defined as the proportion of correctly predicted sample points relative to the total number of samples, to evaluate the model's classification performance. The OA is calculated using Eq. (15):

$$OA = \frac{1}{N} \sum_x 1\{Z(x) = Z_T(x)\} \quad (15)$$

in which, N represents the total number of validation sample cells; $Z(x)$ denotes the predicted class at location x ; $Z_T(x)$ indicates the true class at location x ; and $1\{\cdot\}$ is the indicator function, taking a value of 1 when the predicted class matches the true class and 0 otherwise. This metric enables a quantitative assessment of the overall classification accuracy of the predictive model for lithological class distributions.

2.5.5 Negative Log-Likelihood Evaluation

In addition to OA and information entropy, this study further employs Negative Log-Likelihood (NLL) to evaluate the quality of probabilistic predictions. The NLL is defined as Eq. (16):

$$NLL = -\frac{1}{n} \sum_{i=1}^n p_{Z(x_i)}(x_i) \quad (16)$$

where n represents the number of validation points, $Z(x_i)$ denotes the true class at the i -th validation point, and $p_{Z(x_i)}(x_i)$ is the predicted probability of the true class at that point. Intuitively, when the model assigns a higher probability to the true class, the NLL value is lower, indicating more reliable predictions. Conversely, if the model assigns a lower probability to the true class, the NLL increases.

Unlike OA, which solely measures whether classifications are correct, NLL focuses on the rationality of the predicted probability distribution, sensitively reflecting the model's performance in uncertainty quantification. Therefore, the combined use of OA and NLL provides a more comprehensive assessment of model performance, evaluating both classification accuracy and probabilistic reliability.

3 Synthetic Case Study

3.1 Case Description

This case was designed to represent a simplified sedimentary stratigraphic system characterized by laterally continuous layers, gently undulating interfaces, and locally developed thin units. These features are commonly observed in sedimentary environments, where lithological units tend to extend continuously in the horizontal direction, whereas vertical variations in lithology and layer thickness may occur over relatively short distances. Therefore, the synthetic profile provides a controlled benchmark for evaluating whether the proposed anisotropic IK framework can preserve lateral continuity, reproduce vertical lithological variations, and identify uncertainty near layer boundaries under sparse borehole constraints.

In this synthetic case, stratigraphic boundaries are generated using four gently undulating curves. Each curve consists of a sinusoidal or cosinusoidal component, a mean depth term, a regional dipping trend, and a smoothed random perturbation. The sinusoidal terms are used to represent local interface undulations, whereas the regional trend term is introduced to simulate the overall gently dipping structure of sedimentary strata. The four initial boundary curves are defined as Eq. (17):

$$\begin{aligned} A_1(x) &= 6.0 \sin\left(\frac{\pi}{45}x\right) + 22.0 + 0.5dip(x) \\ A_2(x) &= 5.5 \cos\left(\frac{\pi}{35}x + \frac{\pi}{3}\right) + 33.0 + 0.6dip(x) \\ A_3(x) &= 5.5 \sin\left(\frac{\pi}{30}x + \frac{\pi}{9}\right) + 50.0 + 0.7dip(x) \\ A_4(x) &= 7.0 \sin\left(\frac{\pi}{45}x\right) + 68.0 + 0.5dip(x) \end{aligned} \quad (17)$$

The regional dip angle, denoted as $dip(x)$ (shown as Eq. (18)), is incorporated to simulate the overall gently dipping stratigraphic structure. These curves are sorted column-wise on a two-dimensional grid to delineate five spatial segments, each assigned one of three lithological labels. The layer boundaries defined by the four curves are illustrated in Fig. 3. This approach ensures controlled regularity in layer boundaries while incorporating natural variability through noise and dip, mimicking real sedimentary strata.

$$dip(x) = -0.05(x - 50) \quad (18)$$

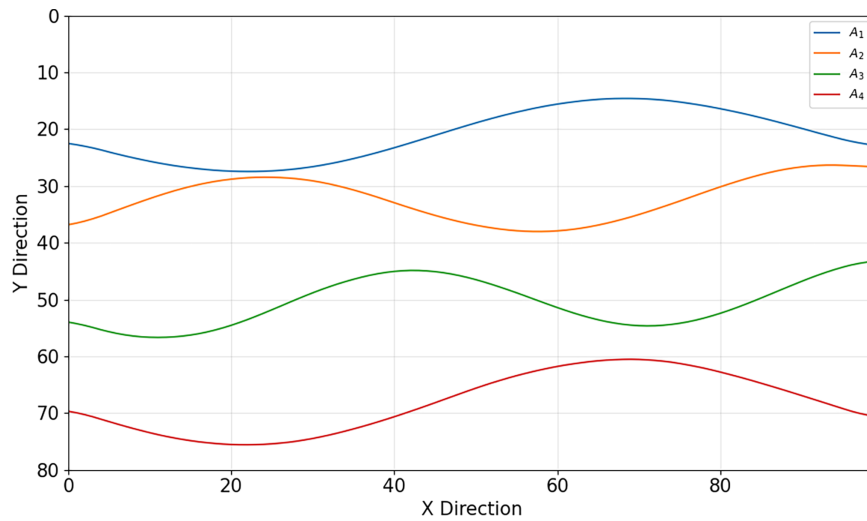


Figure 3: Layer boundaries defined by the four curves.

As shown in Fig. 4, the resulting synthetic stratigraphic profile exhibits prominent undulations, with a thin sedimentary unit identifiable to the left of borehole BH2. To simulate realistic exploration conditions, four boreholes are uniformly placed across the profile (as shown) to provide observational constraints. Using the known lithological data from these boreholes, we perform prediction and interpolation to reconstruct the complete lithological class distribution.

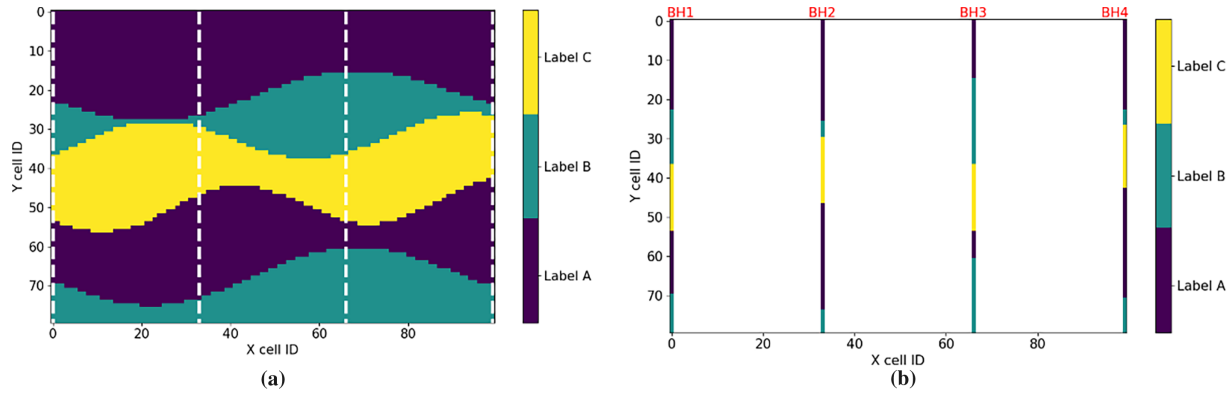


Figure 4: Synthetic case information: (a) synthetic stratigraphic profile; (b) known boreholes.

3.2 Selection of Optimal Anisotropic Parameters

In two-dimensional stratigraphic modeling, horizontal and vertical directions exhibit distinct spatial correlation patterns. To capture this anisotropy, we employ coordinate scaling, fixing the vertical scaling factor at $S_y = 1$ and enumerating horizontal scaling factors S_x to adjust the model's sensitivity to horizontal correlation. A K -fold cross-validation approach with $K = 4$ is used, given the four boreholes. In each fold, one borehole is reserved as the validation set, with the remaining three used for training. The training phase involves constructing indicator variables for three classes and fitting exponential variogram models, while the validation phase performs ordinary kriging to predict probabilities at the excluded borehole locations. Prediction accuracy is evaluated using OA and NLL, averaged over the four folds.

Fig. 5a shows that OA remains high (~ 0.93) for $S_x = 2-6$ but declines significantly for $S_x > 6$, dropping to ~ 0.70 . Fig. 5b illustrates the NLL trend, with a minimum value (0.118) at $S_x = 5$, where OA also peaks (0.938). Thus, $S_x = 5$ and $S_y = 1$ are selected as the optimal anisotropic parameters, indicating that horizontal correlation length is approximately five times the vertical, consistent with the strong lateral continuity and rapid vertical variation of sedimentary strata.

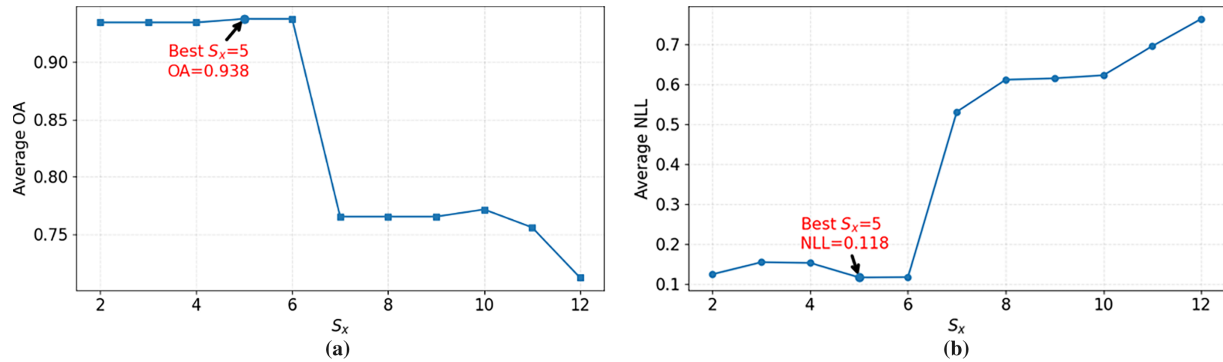


Figure 5: Synthetic case cross-validation results: (a) OA; (b) NLL as a function of S_x .

3.3 Prediction Results

Using the optimal anisotropic parameters ($S_x = 5$, $S_y = 1$), we applied Indicator Kriging to the synthetic case. Variogram models were fitted for three lithological labels (Label A, Label B, Label C), with results shown in Fig. 6. The empirical semi-variogram points (blue dots) exhibit a monotonic increase with lag distance (scaled coordinates), and the exponential model (orange curve) provides a smooth fit. The fitted

variograms exhibit clear class-dependent differences. For Label A, the semivariogram has a nugget of approximately 0.025 and a sill of approximately 0.225, and it gradually approaches a relatively stable level at lag distances of approximately 8–10 units, indicating comparatively strong spatial continuity. For Label B, the nugget is also approximately 0.025, while the sill is slightly lower, at approximately 0.200. Its semivariogram becomes relatively stable at lag distances of approximately 6–8 units, suggesting moderate spatial continuity. Label C has a nugget of approximately 0.02–0.03 and the lowest sill, approximately 0.14. Its semivariogram approaches a stable level at lag distances of approximately 3–4 units, indicating a more localized spatial distribution and weaker continuity than Labels A and B. These class-dependent variogram characteristics reflect differences in the spatial distribution and continuity of the lithological classes. The fitted variogram models are subsequently used in the anisotropic IK framework to estimate class probabilities across the prediction domain.

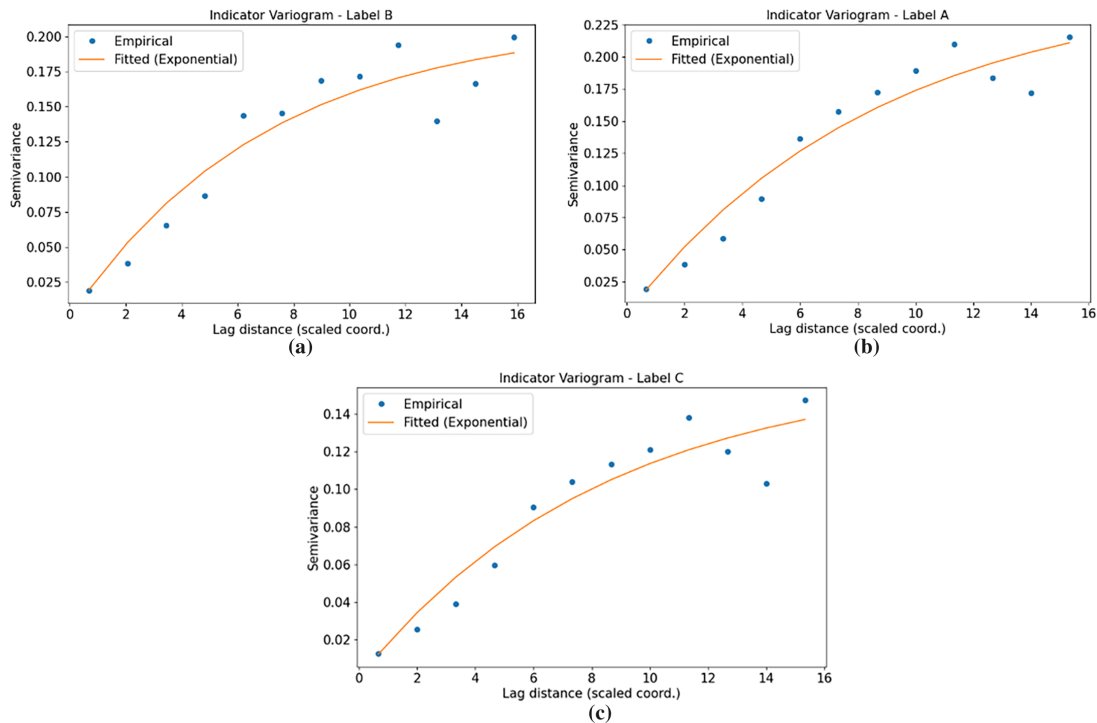


Figure 6: Fitted exponential variogram models for the indicator variables of the three lithological classes: (a) label A; (b) label B; (c) label C.

Probability predictions were performed across the two-dimensional grid using the fitted variograms and four borehole constraints. The probability distribution maps for the three labels are shown in Fig. 7. Label A's probability heatmap highlights high probabilities (>0.8) in the upper and lower profile regions, corresponding to stable layer segments, with one-hot probabilities ($=1$) near boreholes BH1 and BH3. Label B's probabilities are concentrated in the middle and lower profile, while Label C's appear in the middle transition zone. These maps successfully reconstruct the layered structure of the synthetic strata, providing smooth interpolations in borehole gaps and demonstrating IK's ability to handle categorical uncertainty.

The hard classification map, generated using the MAP criterion, is shown in Fig. 8a. It accurately reconstructs the synthetic strata's overall distribution, with clear boundaries for the five segments and high consistency with the true profile, including the thin layer to the left of BH2. The OA is 90.45%, reflecting high predictive accuracy under limited constraints. The entropy map in Fig. 8b quantifies prediction

uncertainty, with near-zero entropy at borehole locations and their vicinity, indicating high confidence, and higher entropy (0.5–1.0) in borehole gaps and near layer boundaries, particularly in areas with significant undulations. This entropy distribution aids in identifying high-risk exploration zones and guiding further drilling.

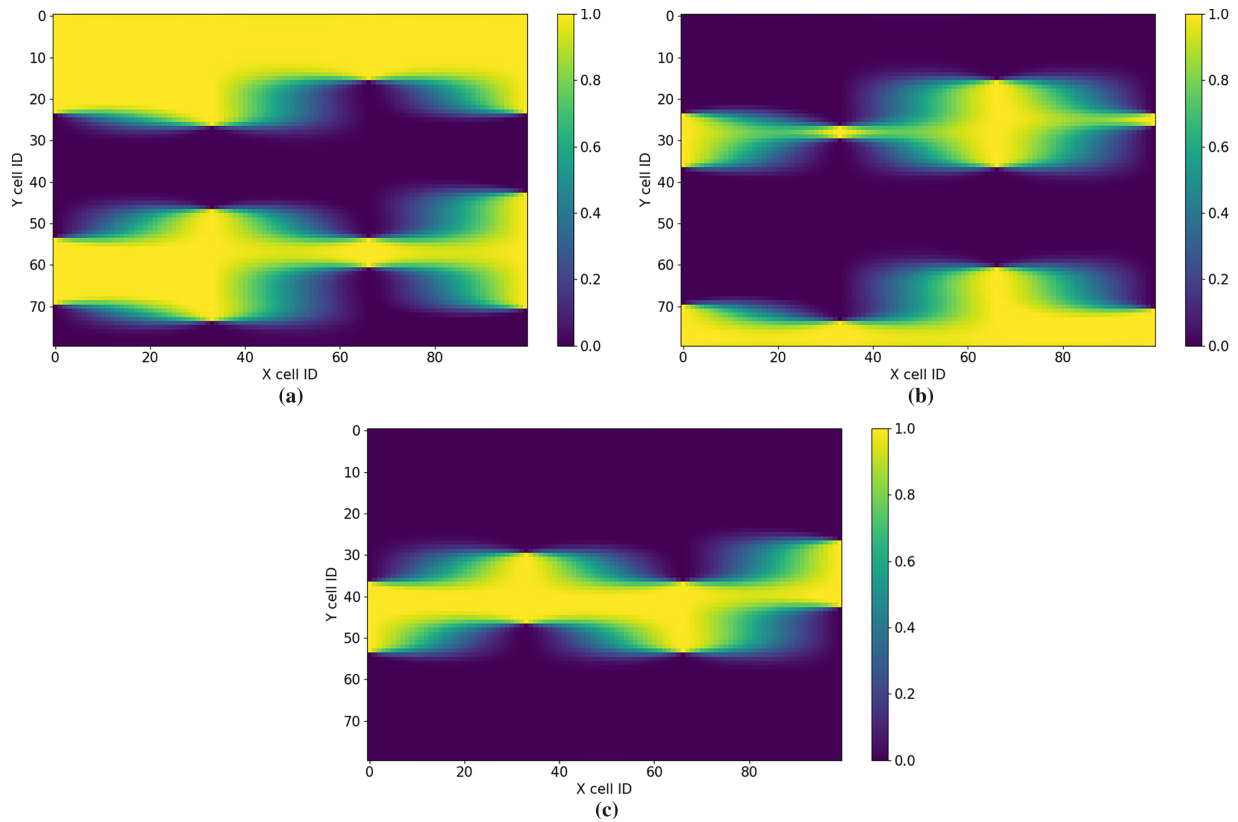


Figure 7: Predicted probability distribution maps for: (a) label A; (b) label B; (c) label C.

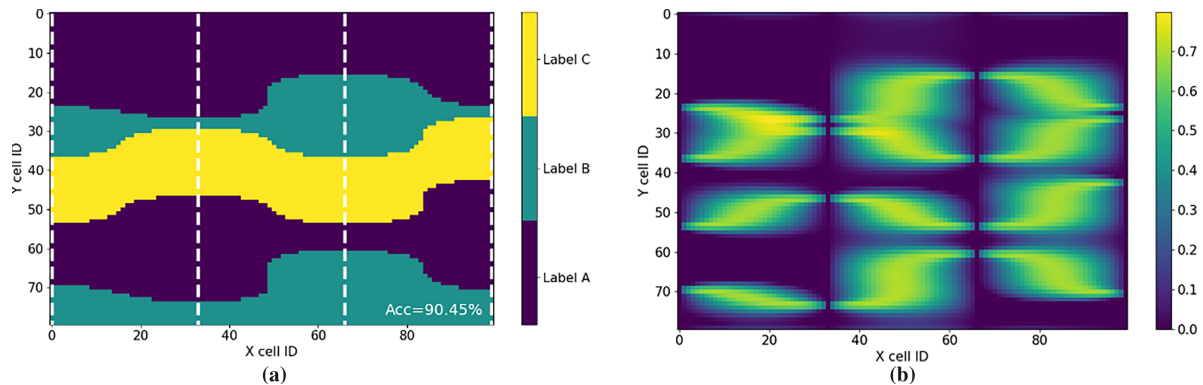


Figure 8: Synthetic case prediction results: (a) MAP; (b) corresponding entropy.

For comparison, a Markov Random Field (MRF) model [8] is applied to the same case, with results shown in Fig. 9. The MRF's hard classification map shows smoother boundary transitions but exhibits localized overfitting, such as overextending the thin layer to the left of BH2, resulting in a slightly lower OA

(~87.92%). Additionally, MRF's uncertainty quantification is less intuitive than IK's entropy map, which more clearly highlights spatial variability. Overall, the IK framework's integration of anisotropy and probabilistic interpretation makes it more suitable for sedimentary stratigraphic modeling and validation.

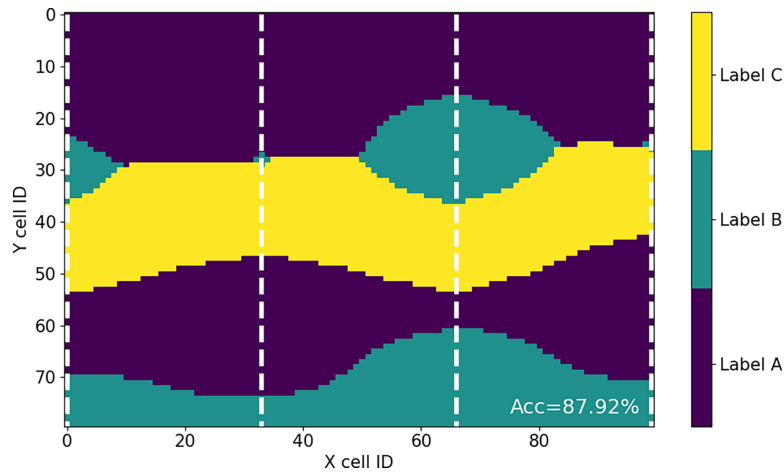


Figure 9: MRF model (Wei and Wang, 2022) [8] prediction results.

3.4 Complex Synthetic Case

To further evaluate the robustness of the proposed framework under more challenging geological conditions, a complex synthetic stratigraphic case was constructed. The original synthetic case in [Section 3.1](#) mainly consists of smooth and laterally continuous layer boundaries, which provides a controlled benchmark for validating the basic performance of the proposed IK framework. However, real sedimentary systems may contain more complex local features, such as pinching-out thin layers, lenticular units, and channel-like erosional fills. These features may weaken local spatial continuity and introduce more complex boundary geometries. Therefore, an additional complex synthetic case was designed to examine whether the proposed anisotropic IK framework can still reconstruct the dominant stratigraphic architecture and identify uncertain regions under sparse borehole constraints.

The model domain has a horizontal length of 100 cells and a depth of 80 cells. Five boreholes were uniformly distributed along the horizontal direction, including two boreholes located at the left and right boundaries. The borehole positions were set at $x = 0$, $x = 25$, $x = 50$, $x = 75$, and $x = 99$, respectively. As shown in [Fig. 10a](#), the background stratigraphic framework consists of three main lithological classes, namely Label A, Label B, and Label C. These three classes form the dominant layered sedimentary structure. To explicitly represent local complex geological features, three additional feature-specific classes were introduced: a pinching-out thin layer, a lenticular unit, and a channel-like erosional fill. Therefore, the complex synthetic case contains six labels in total, including three main lithological classes and three local complex geological features.

The pinching-out thin layer is located in the upper-middle part of the profile and gradually decreases in thickness along the lateral direction until it disappears. This feature represents a typical local thinning process in sedimentary environments. The lenticular unit is embedded in the lower-middle part of the profile and has a limited horizontal extent, reflecting localized sedimentary deposition. The channel-like erosional fill is developed in the middle part of the profile and cuts into the surrounding lithological categories with a curved geometry. Compared with the basic synthetic case, these features introduce stronger local heterogeneity,

discontinuous boundaries, and abrupt changes in lithological distribution, making the reconstruction task more challenging.

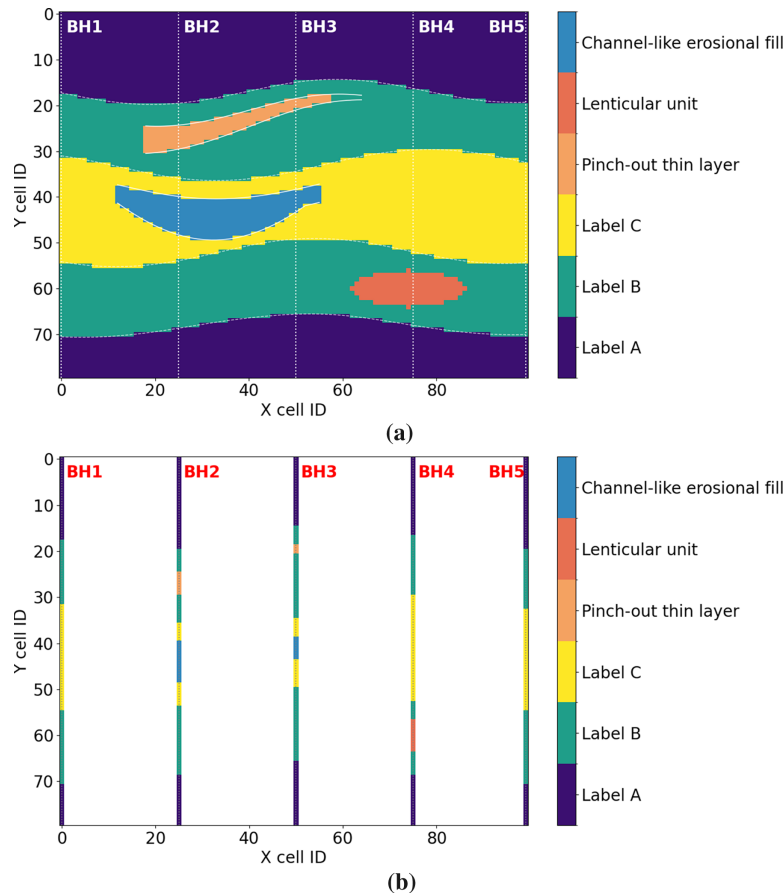


Figure 10: Complex synthetic case: (a) true stratigraphic profile; (b) sparse borehole constraints.

Fig. 10b shows the sparse borehole constraints extracted from the true stratigraphic profile. Although five boreholes are used, the borehole spacing remains relatively large compared with the scale of local geological features. Some complex structures are only partially intersected by the boreholes. For example, the pinching-out thin layer is captured by the left and middle boreholes, while its lateral termination remains weakly constrained. The channel-like erosional fill is mainly constrained by the middle boreholes, and its lateral boundary must be inferred from sparse observations. The lenticular unit is only locally sampled by one borehole, which makes its horizontal extent difficult to determine. Therefore, this case provides a stricter test for evaluating the ability of the proposed method to infer complex geological structures from limited borehole information.

Using the same anisotropic IK modeling procedure, the reconstructed stratigraphic profile is shown in Fig. 11a. The proposed framework successfully recovers the overall layered architecture of the complex synthetic case. The major distributions of Label A, Label B, and Label C are reconstructed with good lateral continuity, and the main stratigraphic boundaries are consistent with the true profile. In addition, the local complex features are generally identified. The pinching-out thin layer in the upper-middle part is reproduced with a clear thinning trend. The channel-like erosional fill in the middle part is also reconstructed, although its boundary is slightly smoothed compared with the true profile. The lenticular unit in the lower-middle part

is well captured in terms of its approximate location and finite horizontal extent. The overall accuracy reaches 95.3%, indicating that the proposed framework maintains high prediction performance even in the presence of local complex geological features. The overall accuracy reaches 95.3%, indicating that the proposed framework maintains high prediction performance even in the presence of local complex geological features. The higher OA than that of the basic case is mainly attributable to the smoother background interfaces, the additional borehole constraint, and the limited spatial extent of the local complex features; therefore, the two OA values should not be interpreted as a direct comparison of geological difficulty.

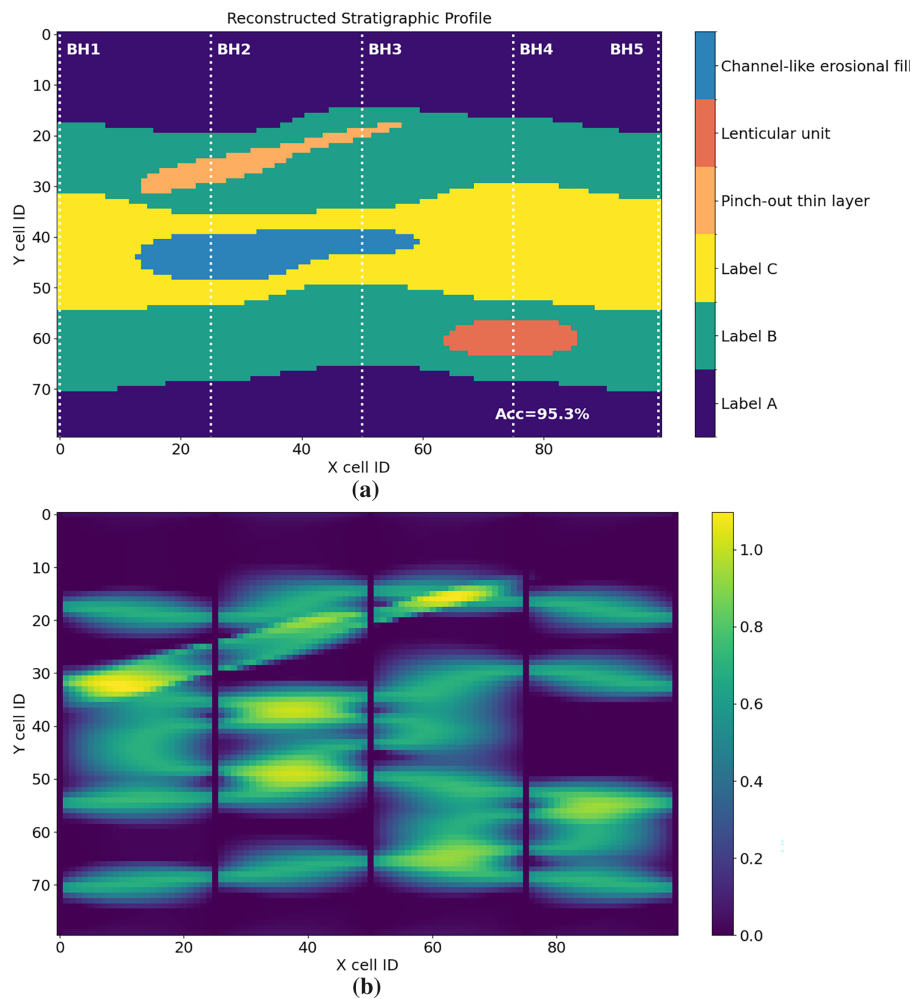


Figure 11: Prediction results for the complex synthetic case: (a) reconstructed stratigraphic profile; (b) entropy-based uncertainty map.

Nevertheless, some local deviations can still be observed. The reconstructed channel-like body is smoother than the true one, especially near its lateral margin and lower erosional boundary. This is mainly because IK relies on spatial correlation and tends to produce smooth probability transitions between sparsely sampled boreholes. Similarly, the detailed termination of the pinching-out thin layer is difficult to reconstruct exactly because its thickness gradually decreases and only limited borehole information is available near its pinch-out end. These local differences are reasonable because such features represent small-scale heterogeneity relative to the borehole spacing.

[Fig. 11b](#) presents the corresponding entropy-based uncertainty map. The uncertainty distribution shows clear geological consistency. Low-entropy regions mainly occur around borehole locations and within thick, laterally continuous lithological classes, indicating high prediction confidence. In contrast, high-entropy regions are concentrated near stratigraphic boundaries, borehole gaps, the pinch-out zone, the channel-like erosional boundary, and the margin of the lenticular unit. These are precisely the locations where multiple lithological classes compete and where the sparse borehole data provide insufficient constraints. The entropy map therefore provides useful information beyond the hard classification result by identifying zones where the reconstructed stratigraphy is less reliable.

Overall, the complex synthetic case demonstrates that the proposed anisotropic IK framework is not limited to smooth and continuous stratigraphic systems. Although local smoothing occurs near abrupt or small-scale geological features, the method can still recover the dominant stratigraphic architecture and identify the approximate locations of complex features. More importantly, the entropy-based uncertainty map effectively highlights regions with increased prediction uncertainty, such as pinch-out zones, lenticular boundaries, and channel-like erosional interfaces. This confirms the practical value of the proposed framework for probabilistic sedimentary stratigraphic modeling under sparse borehole constraints.

4 Real-World Stratigraphic Benchmark Case Study

4.1 Prediction Results and Model Comparison

To further validate the proposed anisotropic IK modeling framework, we applied it to a real-world stratigraphic benchmark profile from [\[37\]](#), referred to as the Australia case. This case consists of sparse borehole constraints and an interpreted reference stratigraphic profile. The reference profile was not generated by the proposed method, but was adopted from previously interpreted geological information. Therefore, in this study, it is used as a benchmark reference model for quantitative comparison rather than as an independently verified ground truth obtained from additional field investigations. Accordingly, the reported accuracy values should be interpreted as the consistency between the reconstructed result and the interpreted reference profile under sparse borehole constraints.

The Australia case represents a sedimentary stratigraphic system composed of several lithological classes, including relatively continuous upper layers, transitional middle units, and deeper basal strata (see [Fig. 12](#)). The profile exhibits clear sub-horizontal layering, local interface undulations, and thickness variations of individual lithological classes, which are common features of sedimentary successions. From a geological perspective, the main challenge in this case is to reconstruct laterally continuous but vertically variable lithological boundaries from only a few boreholes. In particular, the boundaries between adjacent lithological classes are not strictly planar, and local undulations may introduce uncertainty in borehole gaps.

It should also be noted that the selected Australia benchmark profile is mainly characterized by laterally continuous layering and moderate boundary undulations. Such conditions are suitable for evaluating whether the proposed anisotropic IK framework can preserve lateral continuity, capture vertical variation, and quantify prediction uncertainty. However, these conditions may also be relatively favorable for kriging-based methods because the dominant lateral continuity is consistent with the spatial correlation assumption of IK. Therefore, the results from this real-world benchmark case should not be overgeneralized to all sedimentary depositional architectures, especially those involving strongly discontinuous, channelized, lenticular, erosional, or abrupt facies geometries. To complement this limitation, a more complex synthetic case involving a pinching-out thin layer, a lenticular unit, and a channel-like erosional fill was introduced in [Section 3.4](#).

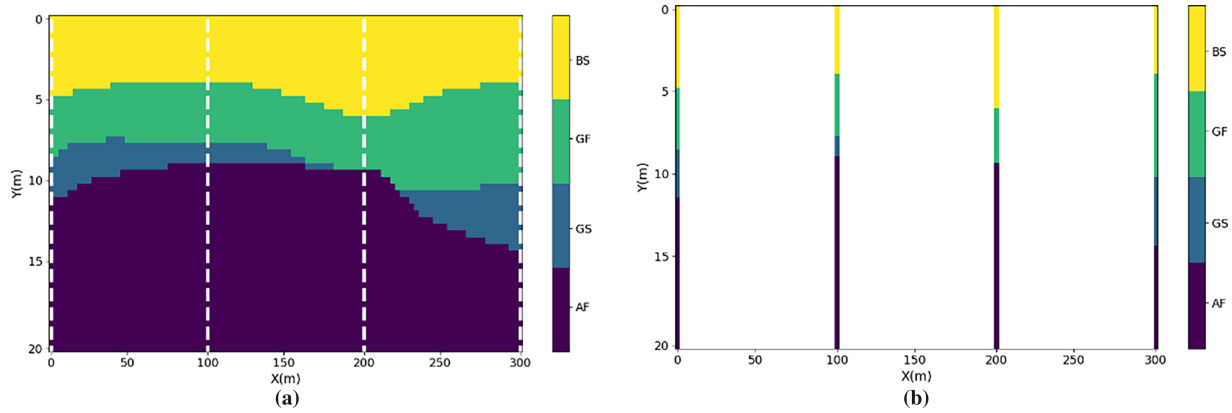


Figure 12: Australia case information: (a) original strata; (b) known boreholes.

With the vertical scaling factor fixed at $S_y = 1$, we evaluated candidate horizontal scaling factors $S_x \in [2, 12]$ using 4-fold cross-validation, selecting parameters that minimize NLL and maximize OA. The results, shown in Fig. 13, indicate that OA remains high for $S_x = 2-6$, with NLL minimized (0.104) at $S_x = 3$, where OA is also high (0.935). Thus, $S_x = 3$ and $S_y = 1$ were chosen as the optimal anisotropic parameters for subsequent predictions.

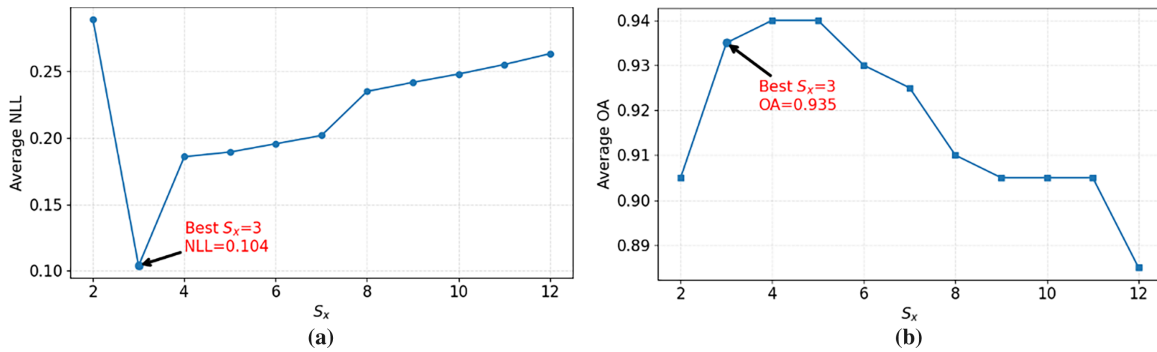


Figure 13: Australia case cross-validation results: (a) OA; (b) NLL as a function of S_x .

Using these parameters, IK modeling was performed, with results shown in Fig. 14. The hard classification map effectively recovers the layered distribution of the strata, with clear and continuous lithological boundaries that closely match the true profile. Minor deviations occur in localized thin layers and boundary transitions, but the OA reaches 95.26%, demonstrating the model's ability to capture primary structural features under limited borehole constraints.

The entropy distribution map reveals prediction uncertainty patterns. Entropy is near zero at borehole locations and their vicinity, indicating high reliability, while it increases significantly (0.5–1.0) in borehole gaps and near layer boundaries, reflecting near-uniform class probabilities and higher uncertainty. This spatial uncertainty pattern aligns with geological expectations, as boundary zones and data-sparse areas are inherently harder to predict accurately, providing valuable guidance for drilling planning and risk analysis.

To comprehensively evaluate the proposed method, we compared IK predictions with those from the MRF model [8] and the IC-XGBoost model [37], as shown in Fig. 15. The MRF model produces smoother boundary transitions but achieves a slightly lower OA (~94.8%) compared to IK. The IC-XGBoost model captures the main stratigraphic distribution but exhibits deviations in boundary positions and slightly lower

continuity and accuracy, with an OA of 91.2%. Overall, IK shows complementary advantages over MRF and IC-XGBoost, particularly in probabilistic interpretation and uncertainty identification, achieving higher accuracy and smoother, geologically realistic boundary transitions, demonstrating its superior applicability and potential for real-world stratigraphic modeling and risk assessment.

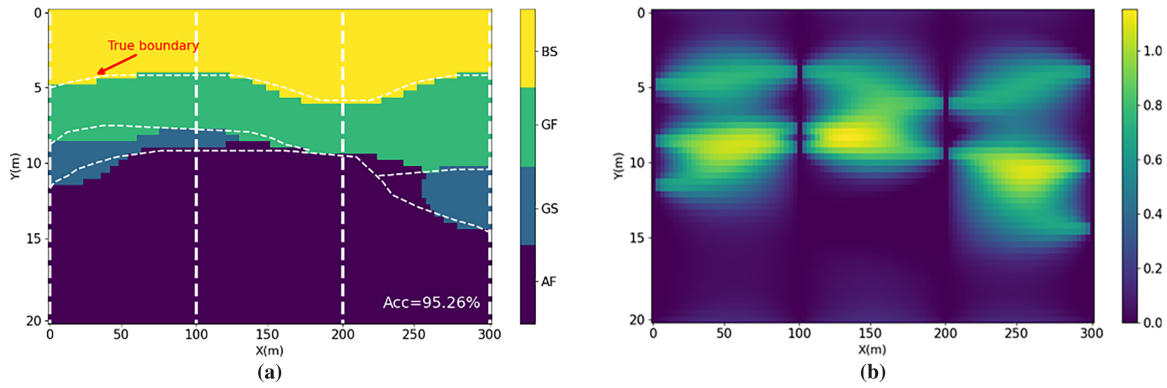


Figure 14: Australia case prediction results: (a) MAP; (b) corresponding entropy.

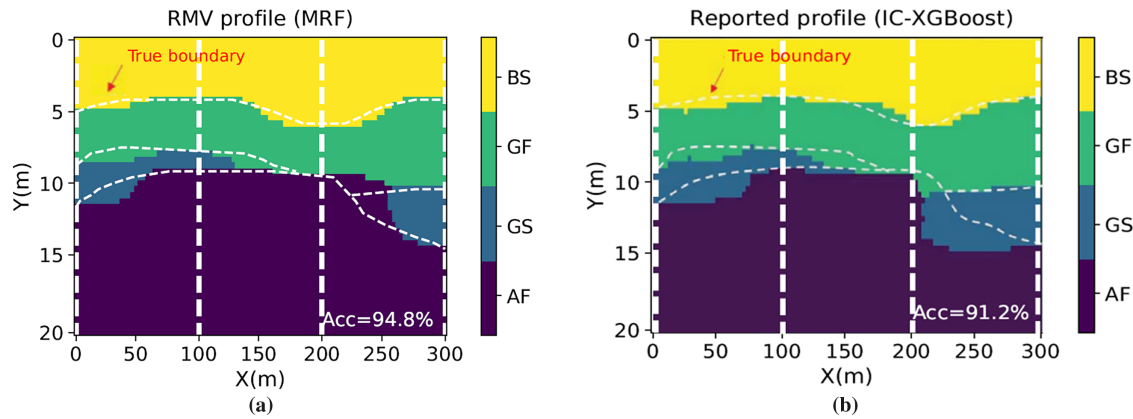


Figure 15: Australia case results: (a) MRF model; (b) IC-XGBoost model.

It should be noted that the accuracy difference between IK and MRF in the Australia case is relatively small. Therefore, the advantage of the proposed framework should not be interpreted solely from the perspective of OA. Instead, the comparison should be understood from several complementary aspects, including boundary delineation, thin-layer preservation, spatial continuity, probabilistic interpretation, and uncertainty quantification.

The MRF model can effectively enforce spatial smoothness and therefore produces a generally continuous stratigraphic pattern. However, its performance is strongly influenced by the predefined neighborhood system and smoothing parameters. When the smoothing constraint is relatively strong, local thin layers and rapidly varying boundaries may be over-smoothed. This explains why the MRF result shows smoother transitions but may weaken local boundary details. In contrast, the proposed IK framework estimates class-wise conditional probabilities based on spatial correlation and anisotropic distance scaling, which helps preserve the dominant lateral continuity of sedimentary strata while maintaining probabilistic information near uncertain boundaries.

The IC-XGBoost model can capture nonlinear relationships between input features and lithological classes and performs well in identifying the main stratigraphic distribution. However, as a data-driven classification model, it does not explicitly impose geostatistical spatial correlation during prediction. As a result, the predicted stratigraphic boundaries may show local deviations or reduced geological continuity, especially in borehole gaps where direct constraints are limited. Moreover, although machine learning classifiers may output class probabilities, these probabilities are not necessarily spatially calibrated for geological uncertainty interpretation.

Therefore, the proposed IK framework does not simply improve the OA by a large margin. Its main advantage lies in providing a geologically interpretable probabilistic reconstruction. The normalized class probabilities and entropy maps allow the model to identify uncertain regions near stratigraphic boundaries, borehole gaps, thin layers, and local complex structures. These uncertainty results provide additional information for drilling optimization and geological risk assessment, which cannot be fully reflected by OA alone.

4.2 Computational Efficiency Comparison

To further evaluate the practical applicability of the proposed framework, the computing time and resource requirements of IK, MRF, and IC-XGBoost were compared under the same computational environment. All experiments were conducted on a workstation equipped with Intel i7, 32 GB RAM. For each method, the reported runtime includes model fitting or training and full-grid prediction. The same borehole constraints and prediction grid were used for all models to ensure a fair comparison.

The computational cost of the IK framework mainly consists of variogram fitting, neighborhood search, and kriging-based probability prediction. The MRF model requires iterative energy minimization, and its computational time is affected by the number of iterations and convergence criteria. The IC-XGBoost model requires classifier training and subsequent prediction. Although IC-XGBoost is efficient after training, it does not explicitly impose geostatistical spatial correlation during prediction.

As shown in [Table 1](#), the proposed IK framework required 46 s to complete the full modeling process, which was shorter than MRF and IC-XGBoost. The MRF model required 1.2 min, mainly due to iterative spatial regularization, while IC-XGBoost required 1.3 min, including classifier training and prediction. These results indicate that the proposed IK framework achieves a favorable balance between computational efficiency, geological interpretability, and probabilistic uncertainty quantification. In addition to its relatively low computational cost, IK directly provides class-wise probability distributions and entropy-based uncertainty maps, which are useful for identifying boundary-related and data-sparse uncertainty zones.

Table 1: Computing time and resource requirements of different models.

Method	Main Computational Process	Total Computing Time	Relative Efficiency
IK	Variogram fitting, neighborhood search, and kriging prediction	46 s	Fast
MRF	Iterative spatial regularization and energy minimization	1.2 min/72 s	Moderate
IC-XGBoost	Classifier training and full-grid prediction	1.3 min/78 s	Moderate

5 Conclusion

This study proposes a two-dimensional stratigraphic modeling framework based on IK. By transforming multi-class lithological data into binary indicator variables and fitting exponential variogram models, the method enables joint probability predictions for multiple lithological classes. Coordinate scaling is introduced to capture the anisotropic characteristics of sedimentary strata, effectively reflecting their strong lateral continuity and rapid vertical variation. Probability normalization ensures the validity of predictions, while information entropy quantifies uncertainty. Validation through synthetic and real-world case studies demonstrates that the method accurately reconstructs layered structures under limited borehole constraints, achieving high predictive accuracy, particularly in maintaining boundary continuity and geological realism.

Comparative analyses highlight the advantages of the IK framework. Compared to the MRF model, IK avoids overlooking thin layers and more accurately reproduces undulating layer boundaries. Compared with MRF and IC-XGBoost, the IK framework shows complementary advantages in boundary delineation, thin-layer preservation, spatial continuity, and probabilistic uncertainty representation, although the improvement in overall accuracy may be moderate in some cases. In summary, the proposed IK framework is both theoretically robust and practically adaptable, offering an effective tool for sedimentary stratigraphic modeling, drilling optimization, and geological risk assessment.

However, the proposed framework still has some limitations. The present IK-based workflow treats lithological or lithofacies classes as categorical indicator fields and does not explicitly impose stratigraphic order rules, vertical succession constraints, or depositional bounding surfaces. In addition, the real-world benchmark profile used in this study is mainly characterized by laterally continuous layering, and further validation is needed for highly discontinuous, channelized, lenticular, or erosional depositional architectures using independent field data.

Acknowledgement: The authors gratefully acknowledge the support provided by the relevant research institutions and project teams.

Funding Statement: This research was funded by the Project of the Science and Technology Development Plan of China Railway No. 18 Engineering Group Co., Ltd. (No. 185G25-01), the Natural Science Foundation of Hunan Province (No. 2026JJ60411), and the Scientific Research Foundation of Hunan Provincial Education Department (No. 25B0655).

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, Liying Wang; methodology, Xingxing Wei; software, Liying Wang; validation, Liying Wang and Lianjiang Yang; formal analysis, Xingxing Wei; investigation, Guang He and Dong Bai; resources, Guang He and Dong Bai; data curation, Wen Luo and Min Peng; writing—original draft preparation, Liying Wang; writing—review and editing, Xingxing Wei; visualization, Liying Wang; supervision, Xingxing Wei; project administration, Liying Wang; funding acquisition, Liying Wang. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Zhao C, Gong W, Juang CH, Tang H, Liu Z, Hu X. Probabilistic characterization of subsurface stratigraphic configuration by merging borehole and geophysical data. *Acta Geotech.* 2025;20(5):2421–39. doi:10.1007/s11440-025-02550-w.

2. Jiang QH, Zhang JZ, Zhang DM, Huang HW. Simulation of geological uncertainty based on improved three-dimensional coupled Markov chain model. *Eng Geol.* 2024;340(5):107647. doi:10.1016/j.enggeo.2024.107647.
3. Wu J, Zhuo L, Pei J, Li Y, Xie H, Wu J, et al. Modularized and parametric modeling technology for finite element simulations of underground engineering under complicated geological conditions. *Comput Model Eng Sci.* 2024;140(1):621–45. doi:10.32604/cmesci.2024.046398.
4. Zhang JZ, Jiang QH, Zhang DM, Huang HW, Liu ZQ. Influence of geological uncertainty and soil spatial variability on tunnel deformation and their importance evaluation. *Tunn Undergr Space Technol.* 2024;152(9):105930. doi:10.1016/j.tust.2024.105930.
5. Wang J, Sun P, Chen L, Yang J, Liu Z, Lian H. Recent advances of deep learning in geological hazard forecasting. *Comput Model Eng Sci.* 2023;137(2):1381–418. doi:10.32604/cmesci.2023.023693.
6. Zinas O, Papaioannou I, Schneider R, Cuéllar P. Multivariate Gaussian process regression for 3D site characterization from CPT and categorical borehole data. *Eng Geol.* 2025;352(41):108052. doi:10.1016/j.enggeo.2025.108052.
7. Qian ZH, Shi C, Wang Y, Cao ZJ. Nonparametric and continuous variable-based stratigraphic modelling from sparse boreholes using signed distance function and Bayesian compressive sensing. *Can Geotech J.* 2025;62(6):1–23. doi:10.1139/cgj-2024-0295.
8. Wei X, Wang H. Stochastic stratigraphic modeling using Bayesian machine learning. *Eng Geol.* 2022;307:106789. doi:10.1016/j.enggeo.2022.106789.
9. Tentori D, Mancini M, Stigliano F, Milli S, Simionato M, Livani M, et al. Integrating sequence stratigraphy and geostatistical methods for 3D lithofacies modelling of the Tiber alluvial plain, Rome, Italy. *Basin Res.* 2025;37(2):e70024. doi:10.1111/bre.70024.
10. Liu Y, Zhang X, Guo W, Kang L, Gao J, Yu R, et al. Research status of and trends in 3D geological property modeling methods: a review. *Appl Sci.* 2022;12(11):5648. doi:10.3390/app12115648.
11. Falivene O, Arbués P, Howell J, Muñoz JA, Fernández O, Marzo M. Hierarchical geocellular facies modelling of a turbidite reservoir analogue from the Eocene of the Ainsa basin, NE Spain. *Mar Petrol Geol.* 2006;23(6):679–701. doi:10.1016/j.marpetgeo.2006.05.004.
12. Falivene O, Cabrera L, Muñoz JA, Arbués P, Fernández Ó, Sáez A. Statistical grid-based facies reconstruction and modelling for sedimentary bodies. Alluvial-palustrine and turbiditic examples. *Geol Acta Int Earth Sci J.* 2007;5(3):199–230. doi:10.1344/105.000000295.
13. Falivene O, Cabrera L, Sáez A. Optimum and robust 3D facies interpolation strategies in a heterogeneous coal zone (Tertiary as Pontes basin, NW Spain). *Int J Coal Geol.* 2007;71(2–3):185–208. doi:10.1016/j.coal.2006.08.008.
14. Zhang Y, Ren X, Wang X, Zhu H, Jiang D. Simulation of geological uncertainty using coupled Markov chain: a case study at a manually filled loess site. *ASCE-ASME J Risk Uncertain Eng Syst Part A Civ Eng.* 2024;10(2):05024002. doi:10.1061/ajrua6.rueng-1263.
15. Wang H, Wei X. Three-dimensional stochastic model for stratigraphic uncertainty quantification using Bayesian machine learning. *IOP Conf Ser Earth Environ Sci.* 2024;1337(1):012012. doi:10.1088/1755-1315/1337/1/012012.
16. Wei X, Wang H. Stochastic stratigraphic simulation using image warping from sparse data. *Front Built Environ.* 2025;11:1651919. doi:10.3389/fbuil.2025.1651919.
17. Qian Z, Shi C. Prior geological knowledge enhanced Markov random field for development of geological cross-sections from sparse data. *Comput Geotech.* 2024;173(1):106587. doi:10.1016/j.compgeo.2024.106587.
18. Cardenas IC. A two-dimensional approach to quantify stratigraphic uncertainty from borehole data using non-homogeneous random fields. *Eng Geol.* 2023;314(6):107001. doi:10.1016/j.enggeo.2023.107001.
19. Jiang QH, Zhang JZ, Zhang DM, Huang HW, Shi JK, Li ZL. Influence of geological uncertainty on longitudinal deformation of tunnel based on improved coupled Markov chain. *Eng Geol.* 2024;337(6):107564. doi:10.1016/j.enggeo.2024.107564.
20. Zhen J, Lai F, Shiao JS, Huang M, Lu Y, Lin J. An unsupervised incremental learning model to predict geological conditions for earth pressure balance shield tunneling. *J Rock Mech Geotech Eng.* 2025;17(11):6993–7006. doi:10.1016/j.jrmge.2024.12.018.

21. Sun H, Zhu M, Dai Y, Li X, Rui Y, Zhu W, et al. Digital twin of advanced geological conditions in rock tunnel construction. *Autom Constr.* 2025;178(1):106407. doi:10.1016/j.autcon.2025.106407.
22. Gong W, Zhao C, Juang CH, Zhang Y, Tang H, Lu Y. Coupled characterization of stratigraphic and geo-properties uncertainties—a conditional random field approach. *Eng Geol.* 2021;294(8):106348. doi:10.1016/j.enggeo.2021.106348.
23. Jiang SH, Papaioannou I, Straub D. Optimization of site-exploration programs for slope-reliability assessment. *ASCE-ASME J Risk Uncertain Eng Syst Part A Civ Eng.* 2020;6(1):04020004. doi:10.1061/ajrua6.0001042.
24. Li XY, Zhang LM, Li JH. Using conditioned random field to characterize the variability of geologic profiles. *J Geotech Geoenviron Eng.* 2016;142(4):04015096. doi:10.1061/(asce)gt.1943-5606.0001428.
25. Wang X, Li Z, Wang H, Rong Q, Liang RY. Probabilistic analysis of shield-driven tunnel in multiple strata considering stratigraphic uncertainty. *Struct Saf.* 2016;62:88–100. doi:10.1016/j.strusafe.2016.06.007.
26. Wang X, Wang H, Liang RY, Liu Y. A semi-supervised clustering-based approach for stratification identification using borehole and cone penetration test data. *Eng Geol.* 2019;248(6):102–16. doi:10.1016/j.enggeo.2018.11.014.
27. Fadlilmula MM, Akin S, Duzgun S. Parameterization of channelized training images: a novel approach for multiple-point simulations of fluvial reservoirs. In: *Mathematics of planet earth*. Berlin/Heidelberg, Germany: Springer; 2014. p. 557–60. doi:10.1007/978-3-642-32408-6_122.
28. Hu LY, Chugunova T. Multiple-point geostatistics for modeling subsurface heterogeneity: a comprehensive review. *Water Resour Res.* 2008;44(11):2276–83. doi:10.1029/2008wr006993.
29. Shi C, Wang Y. Nonparametric and data-driven interpolation of subsurface soil stratigraphy from limited data using multiple point statistics. *Can Geotech J.* 2021;58(2):261–80. doi:10.1139/cgj-2019-0843.
30. Li Z, Wang X, Wang H, Liang RY. Quantifying stratigraphic uncertainties by stochastic simulation techniques based on Markov random field. *Eng Geol.* 2016;201:106–22. doi:10.1016/j.enggeo.2015.12.017.
31. Chessa AG. A Markov chain model for subsurface characterization: theory and applications. *Math Geol.* 2006;38(4):503–5. doi:10.1007/s11004-006-9037-9.
32. Qi XH, Li DQ, Phoon KK, Cao ZJ, Tang XS. Simulation of geologic uncertainty using coupled Markov chain. *Eng Geol.* 2016;207:129–40. doi:10.1016/j.enggeo.2016.04.017.
33. Qiu Y, Zhang N, Yin Z, Wang Y, Xu C, Zhang P. Novel multi-spatial receptive field (MSRF) XGBoost method for predicting geological cross-section based on sparse borehole data. *Eng Geol.* 2024;338(11):107604. doi:10.1016/j.enggeo.2024.107604.
34. Liu HC, Zhang N, Yin ZY. Probabilistic stratigraphic modelling from sparse boreholes based on deep learning. *Géotechnique.* 2025;75(11):1457–69. doi:10.1680/jgeot.24.00998.
35. Marinoni O. Improving geological models using a combined ordinary–indicator kriging approach. *Eng Geol.* 2003;69(1–2):37–45. doi:10.1016/S0013-7952(02)00246-6.
36. Pereira PEC, Rabelo MN, Ribeiro CC, Diniz-Pinto HS. Geological modeling by an indicator kriging approach applied to a limestone deposit in Indiará city—Goiás. *REM Int Eng J.* 2017;70(3):331–7. doi:10.1590/0370-44672016700113.
37. Shi C, Wang Y. Development of subsurface geological cross-section from limited site-specific boreholes and prior geological knowledge using iterative convolution XGBoost. *J Geotech Geoenviron Eng.* 2021;147(9):04021082. doi:10.1061/(asce)gt.1943-5606.0002583.
38. Smith JL, Halvorson JJ, Papendick RI. Using multiple-variable indicator kriging for evaluating soil quality. *Soil Sci Soc Amer J.* 1993;57(3):743–9. doi:10.2136/sssaj1993.03615995005700030020x.
39. Li Y, Wu Y, Huang Z. An incremental kriging method for sequential optimal experimental design. *Comput Model Eng Sci.* 2014;97(4):323–57.
40. Bastante FG, Taboada J, Alejano LR, Ordóñez C. Evaluation of the resources of a slate deposit using indicator kriging. *Eng Geol.* 2005;81(4):407–18. doi:10.1016/j.enggeo.2005.08.003.
41. Guan Y, Zhi P, Wang Z. A novel variable-fidelity kriging surrogate model based on global optimization for black-box problems. *Comput Model Eng Sci.* 2025;144(3):3343–68. doi:10.32604/cmesci.2025.069515.
42. Lloyd CD, Atkinson PM. Assessing uncertainty in estimates with ordinary and indicator kriging. *Comput Geosci.* 2001;27(8):929–37. doi:10.1016/S0098-3004(00)00132-1.

43. Wackernagel H. Multivariate geostatistics. Berlin/Heidelberg, Germany: Springer; 2003. doi:10.1007/978-3-662-05294-5.
44. Goovaerts P. Geostatistical software. In: Handbook of applied spatial analysis: software tools, methods and applications. Berlin/Heidelberg, Germany: Springer; 2010. p. 125–34. doi:10.1007/978-3-642-03647-7_8.
45. Dominici S. A man with a master plan: steno's observations on earth's history. *Substantia*. 2021;59–75. doi:10.36253/substantia-1278.
46. Deutsch CV, Journel AG. Geostatistical software library and user's guide. New York, NY, USA: Oxford University Press; 1992.
47. Chiles JP, Delfiner P. Geostatistics: modeling spatial uncertainty. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2012.
48. Shi C, Wang Y. Machine learning of three-dimensional subsurface geological model for a reclamation site in Hong Kong. *Bull Eng Geol Environ*. 2022;81(12):504. doi:10.1007/s10064-022-03009-y.