Stochastic stratigraphic modeling using Bayesian machine learning

Article *in* Engineering Geology · July 2022

DOI: 10.1016/j.enggeo.2022.106789

2 authors:

Xingxing Wei
Central South University
12 PUBLICATIONS **16** CITATIONS
SEE PROFILE

Hui Wang
University of Dayton
55 PUBLICATIONS **906** CITATIONS
SEE PROFILE

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

Stochastic Stratigraphic Modeling Using Bayesian Machine Learning

Xingxing Wei[1,2], Hui Wang[2]*

[1] School of Civil Engineering, Central South University, Changsha, Hunan 410075, China

[2] Department of Civil and Environmental Engineering, The University of Dayton, Dayton, Ohio 45469-0243, USA

*Corresponding author: hwang12@udayton.edu

Abstract

Stratigraphic modeling with quantified uncertainty is an open question in engineering geology. In this study, a novel stratigraphic stochastic simulation approach is developed by integrating a Markov random field (MRF) model and a discriminant adaptive nearest neighbor-based *k*-harmonic mean distance (DANN-KHMD) classifier into a Bayesian framework. The DANN-KHMD classifier is effective for extracting anisotropic patterns from sparse and heterogeneous spatial categorical data such as borehole logs. The MRF parameters can be initially estimated roughly or customized (if site-specific knowledge is available). Later these parameters can be updated and regularized in an unsupervised manner with constraints from site exploration results in a Bayesian manner. Throughout the learning process, both the soil profile and the MRF parameters are updated in a probabilistic manner. The advantages of the proposed approach can be summarized into four points: 1) inferring stratigraphic profile and associated uncertainty in an automatic and fully unsupervised manner; 2) reasonable

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

initial stratigraphic configurations can be sampled and hence lower the computational cost; 3) both stratigraphic uncertainty and model uncertainty are taken into consideration throughout the inferential process; 4) relying on no training stratigraphy images. To illustrate the effectiveness of the developed approach, two synthetic cases and three real-world cases are studied and the advantages of the new approach over existing approaches are demonstrated.

Keywords: Markov random field, stratigraphic uncertainty, discriminant adaptive nearest neighbor, Bayesian machine learning, uncertainty quantification

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

1. Introduction

Obtaining accurate stratigraphic information with quantified uncertainty is essential in planning and designing geosystems. The complete stratigraphic setting at a specific site is typically impossible to acquire due to the limited site exploration data (Gong et al. 2019; Wang et al. 2016; Zhang et al. 2021). The stratigraphic layers are associated with inherent heterogeneity and randomness (Juang et al. 2019). Therefore, inferring subsurface stratigraphy inevitably involves various uncertainties. Such uncertainties cast significant effects on the design and/or construction of geosystems (Wang et al. 2018).

The uncertainty of geo-materials arises mainly from the property uncertainty (or geotechnical uncertainty) and the stratigraphic uncertainty (or geological uncertainty) (Li et al. 2016c). Substantial work has been done on the former type, including conditional Gaussian random field (Jiang et al. 2020; Jiang et al. 2018; Li et al. 2016b), Kriging (Li et al. 2016a) and Bayesian compressive sampling (BCS) (Wang and Zhao 2017; Zhao et al. 2018). In order to provide estimates of stratigraphic uncertainty, several stochastic modeling frameworks have been developed and applied in engineering geology. If cone penetration testing (CPT) logs are available, CPT-based stratigraphic modeling approach using both Robertson chart (Robertson 1990) and Bayesian compressive sampling (BCS) is proposed (Wang et al. 2019c) and further improved in Hu and Wang (2020). If geotechnical borehole logs are available, coupled Markov chain models (CMC) (Elfeki and Dekking 2001; Qi et al. 2016), multiple-point

geostatistics (Fadlelmula et al. 2014; Hu and Chugunova 2008; Shi and Wang 2021c) and iterative convolutional XGBoost (IC-XGBoost) (Shi and Wang 2021b, d) have been investigated. For classical CMC models, there are two descriptors of spatial constraints: the vertical transition probability matrix (VTPM) and horizontal transition probability matrix (HTPM). As this model setting only applies contextual constraints in two directions, it has been observed that the simulated soil profiles may deviate from the reality with notable artifacts and hence additional mitigations approaches are needed (Li et al. 2019). The multiple point geostatistics approach and IC-XGBoost approach require training stratigraphy images reflecting engineering experience on similar ground conditions, which may be difficult to implement in some projects with limited prior information. To date, developing stochastic stratigraphic modeling approaches is still an active research filed in the community of geotechnical engineering and engineering geology.

In recent years, Markov random field (MRF)-based stochastic simulation approaches have been adopted in engineering geology. The study of MRF theory has a long history that can be traced back to Ising's 1925 thesis (Ising 1925). Since then, several pioneers (Besag 1974; Cross and Jain 1983) contributed to a critical extension of the texture modeling, which paves the way to their recent applications in engineering geology, geosciences, and remote sensing (Daly 2005; Norberg et al. 2002; Toftaker and Tjelmeland 2013; Wang et al. 2017; Xie et al. 2002). The MRF models can provide a flexible and intuitive way to describe the spatial Markovian contextual constraints,

which enables the characterization and reproduction of the anisotropy and heterogeneity of stratigraphic structures. Note that the existing MRF models (Gong et al. 2021; Li et al. 2016c; Wang et al. 2019a; Wang et al. 2019b; Zhao et al. 2021) have a certain drawback: the model parameters need to be defined in priori using subjective engineering judgments from local experience, besides they are fixed during the entire inferential process without consideration of model bias/uncertainty. This strategy is not robust as the simulated subsurface profiles may deviate from reality due to the subjective guess of model parameters and the associated uncertainty could be underestimated. In addition, the current MRF-based approach may generate unrealistic soil profiles if there is no proper regularization (which is non-trivial) on model parameters. This drawback can significantly affect the robustness, slow down the convergence, and affect the computational efficiency. Gong et al. (2020) modified the MRF approach regarding model regularization and improved the performance, however, this method is computationally expensive and hence lacks scalability.

In this study, a novel stratigraphic uncertainty quantification approach is developed by integrating the MRF theory and the *discriminant adaptive nearest neighbor-based k-harmonic mean distance* (DANN-KHMD) classifier into a Bayesian framework. The proposed MRF-based approach has an advanced and flexible spatial contextual model, whose parameters can be continuously updated along the inferential process using borehole information. The novel DANN-KHMD classifier can sample reasonable initial guesses of the stratigraphic profile (aka. initial field) from known sparse borehole

93 logs and estimate local label (i.e., soil/rock type) preferences at un-known places. We
94 observe that these initial fields and local label preferences can improve the robustness
95 and accuracy of the MRF-based approaches, help to better quantify the stratigraphic
96 uncertainty, and reduce the computational cost. In this work, we will demonstrate the
97 developed new approach has the following advantages: 1) inferring stratigraphic profile
98 and associated uncertainty in an automatic and fully unsupervised manner; 2)
99 reasonable initial stratigraphic configurations can be sampled and hence lower the
100 computational cost; 3) both stratigraphic uncertainty and model uncertainty are taken
101 into consideration throughout the inferential process; 4) relying on no training
102 stratigraphy images.

103 This paper is organized in the following manner. In *Section 2*, we briefly review
104 the MRF theory and the spatial contextual model in terms of potential functions for
105 modeling the stratigraphic uncertainty, then introduce the development of the DANN-
106 KHMD classifier. In *Section 3*, two synthetic cases are studied to demonstrate the
107 validity and effectiveness of the developed approach and a benchmark comparison
108 study with the CMC model. *Section 4* presents the analyzing results of three case
109 histories and a fair comparison to three existing techniques (i.e., CMC, multiple point
110 geostatistics, and IC-XGBoost) using their original published datasets. In *Section 5*, the
111 computational costs of the synthetic cases and real-world cases are analyzed. Finally,
112 concluding remarks are provided in *Section 6*.

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

2. Model Framework

The flowchart of the developed approach is shown in Fig. 1. Detailed explanations regarding each component are provided blow.

2.1. Review of Markov random field

A stratigraphic profile can be discretized into pixels using a two-dimensional uniform square lattice. $\mathbf{S}=\{i|i=1,2,3,...,s\}$ is the set of all pixels (or lattice vertices) and i is the pixel/vertex index. For a concise narrative, the words "pixel" and "vertex" are interchangeable hereafter. Adjacent pixels that share at least one corner are defined as neighbors to each other and linked together using an edge $\mathbf{E}=\{e_{i,j}=(i,j)\,|\,i\in\mathbf{S},j\in\mathbf{S}\}$ to form an undirected graph $\boldsymbol{G}(\mathbf{S},\mathbf{E})$. On this graph, ∂_i denotes a local neighborhood system consisting of all neighbors of a given pixel i. Clearly, for i, j of a given edge $e_{i,j}$, $j\in\partial_i$ and $i\in\partial_j$.

Let $\mathbf{X}=\{X_i\,|\,i\in\mathbf{S}\}$ be a set of random variables indexed by $\mathbf{S}$. Each random variable X_i takes a label x_i (i.e., a certain soil/rock type) from the state space $\mathbf{L}=\{1,2,3,...,l\}$ consisting of all possible labels representing different soil/rock types. A possible stratigraphic profile (i.e., a realization of $\mathbf{X}$) is denoted as a label configuration of all pixels $\boldsymbol{x}=\{x_i\,|\,i\in\mathbf{S},x_i\in\mathbf{L}\}$.

A random field $P(\mathbf{X})$ defined on $\boldsymbol{G}$ is an MRF if its local characteristics $P(x_i|\boldsymbol{x}_{\mathbf{S}-\{i\}})$ depend only on the neighboring pixels on the graph $\boldsymbol{G}$ (Besag 1986).

$$P(x_i\,|\,\boldsymbol{x}_{\mathbf{S}-\{i\}})=P(x_i\,|\,\boldsymbol{x}_{\partial_i}) \tag{1}$$

Eq. (1) states the Markovianity of an MRF.

According to the Hammersley–Clifford theorem (Besag 1974; Li et al. 2016c), a random field defined on a graph $\boldsymbol{G}$ is an MRF if and only if the random field is a Gibbs field. Detailed proofs can be found in Besag (1974) and Li (2009). Then $P(\mathbf{X})$ follows a Gibbs distribution defined on $\boldsymbol{G}$ with respect to a neighborhood system ∂. Accordingly, the joint probability $P(\boldsymbol{x})$ can be expressed by Eq. (2):

$$P(\boldsymbol{x})=Z^{-1}\exp(-U(\boldsymbol{x})/T) \tag{2}$$

where $U(\boldsymbol{x})$ is the energy of the configuration $\boldsymbol{x}$ and Z is a normalizing constant called the partition function and has the following form:

$$Z=\sum_{\boldsymbol{x}\in\Omega}\exp(-U(\boldsymbol{x})/T) \tag{3}$$

where T stands for "temperature" in the simulated annealing algorithm, which has been discussed in detail in Geman and Geman (1984) and could be simply set as unit. $\Omega=\left\{\boldsymbol{x}=\{x_i\}\middle| i\in\mathbf{S},x_i\in\mathbf{L}\right\}$ is a configuration space that contains all possible soil profiles. Locally, the conditional probability can be expressed as:

$$P(x_i\mid\boldsymbol{x}_{\partial_i})=\frac{P(x_i,\boldsymbol{x}_{\partial_i})}{\sum_{x_i'\in L}P(x_i',\boldsymbol{x}_{\partial_i})}=\frac{\exp[-U(x_i,\boldsymbol{x}_{\partial_i})]}{\sum_{x_i'\in L}\exp[-U(x_i',\boldsymbol{x}_{\partial_i})]} \tag{4}$$

in which $U(x_i,\boldsymbol{x}_{\partial_i})$ is the local energy of the local neighborhood system ∂_i, and x_i' indicates any possible label at pixel *i*, which loops over **L**. The widely used Potts model (Koller and Friedman 2009) is adopted here to characterize the local energy and has the following expression (Li 2009):

$$U(x_i,\boldsymbol{x}_{\partial_i})=V_i(x_i)+\sum_{j\in\partial_i}V_{i,j}(x_i,x_j) \tag{5}$$

where $V_i(x_i)$ is the potential function defined solely on pixel *i* indicating the preference of choosing different labels at pixel *i* and $V_{i,j}(x_i,x_j)$ is the potential

function reflecting the local contextual interaction between neighboring pixels and defined as:

$$V_{i,j}(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j \\ \beta_d & \text{if } x_i \neq x_j \end{cases} \quad (6)$$

where $\beta_d \in \boldsymbol{\beta} = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ indicates the contextual constraint, corresponding to one of the four independent directions (i.e., 0, π/2, π/4, 3π/4), within a two-dimensional lattice grid, and is referred to as a *granularity coefficient*. The illustration is shown in Fig. 2. Note that the adopted anisotropic Potts model is axisymmetric and hence can apply axisymmetric contextual constraints via granularity coefficients to neighboring pixels. It is possible to define a non-axisymmetric (central-asymmetric) Potts model and use it to simulate MRF realizations. However, as reported by Song et al. (2016) in a comprehensive study, converged realizations generated by central-asymmetric parameters do not compliant to the original central-asymmetric parameters. Instead, the contextual interaction of pixels in an MRF eventually will approach to a mutually equivalent (stable) axisymmetric status regardless the Potts model is central-asymmetric or not. Therefore, central-symmetry is a feasible and reasonable assumption for most Gibbs fields. It should be highlighted that, although the Potts model is axisymmetric, it does not indicate nor require the stratigraphic profile of interest to be axisymmetric since even simple local axisymmetric contextual constraints can generate numerous complex soil spatial configurations, which usually are not axisymmetric. As indicated by Eq. (4), the contextual constraints within a local neighborhood system on a graph can be characterized using local energy $U(x_i, \boldsymbol{x}_{\partial_i})$,

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

which has the following two behaviors: 1) a higher preference to a specific label x_i corresponds to a lower potential $V_i(x_i)$ and 2) pixels within a local neighborhood system tend to have a label configuration that can minimize $\sum_{j\in\partial_i} V_{i,j}(x_i, x_j)$, as the entire system prefers a lower energy level. Note that the first behavior can be homogeneous (i.e., the preference setting does not depend on the pixel location, and the simplest setting is $V_i(x_i) = 0, i \in \mathbf{S}, x_i \in \mathbf{L}$), or non-homogeneous (i.e., depends on the pixel location) (Li 2009). In this work, we adopt the non-homogeneous version and characterize $V_i(x_i)$ via the DANN-KHMD classifier and details will be shown in Section 2.3. We will show that this strategy can regularize the estimation of $\boldsymbol{\beta}$ and make it compatible with known pixels (boreholes) with only less- or even non-informative prior. The second behavior is controlled by $V_{i,j}(x_i, x_j)$ and assumed to be homogeneous (otherwise it is intractable in implementation) and intimately related to the granularity coefficients $\boldsymbol{\beta}$. For an anisotropic Potts model, positive elements in $\boldsymbol{\beta}$ cause attraction of neighboring pixels, or encourage clustering effects along a certain direction, while negative elements result in repulsion, or prevent clustering (Cross and Jain 1983).

2.2. Bayesian machine learning

Two different sets of labels, $\boldsymbol{x}_{BH}$ and $\boldsymbol{x}_{unknown}$, are categorized from all labels in the modeling domain. $\boldsymbol{x}_{BH}$ is the set of labels at observed pixels indicating sparse borehole information and $\boldsymbol{x}_{unknown}$ is the set of labels at unobserved pixels elsewhere. The known information $\boldsymbol{x}_{BH}$ is used to infer both $\boldsymbol{x}_{unknown}$ and $\boldsymbol{\beta}$. In this study, a

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

Markov Chain Monte Carlo (MCMC) technique is employed to implement Bayesian machine learning and sample $\boldsymbol{x}_{unknown}$ and $\boldsymbol{\beta}$ iteratively via two conditional posterior distributions $P(\boldsymbol{x}_{unknown}|\boldsymbol{x}_{BH},\boldsymbol{\beta})$ and $P(\boldsymbol{\beta}|\boldsymbol{x}_{unknown},\boldsymbol{x}_{BH})$ iteratively.

1) *Simulation of conditional Markov random field* $P(\boldsymbol{x}_{unknown}|\boldsymbol{x}_{BH},\boldsymbol{\beta})$

$P(\boldsymbol{x}_{unknown}|\boldsymbol{x}_{BH},\boldsymbol{\beta})$ is a Gibbs field with fixed soil labels only at the borehole locations. Given the current label filed (initialized via the DANN-KHMD classifier), the local energy at unknown pixels can be calculated using Eq. (5), and then the corresponding probability of choosing each label at unknown pixels can be evaluated via Eq. (4). Then, realizations of the conditional random field $P(\boldsymbol{x}_{unknown}|\boldsymbol{x}_{BH},\boldsymbol{\beta})$ can be iteratively simulated via a parallel strategy named chromatic sampler, which first assigns all pixels different code names (usually described in terms of different "colors") following the graph coloring rule (a way of coloring the vertices of a graph such that no two neighboring vertices are of the same color using the minimum number of colors), and then samples all pixels having the same code name (i.e., same "color") independently and in parallel according to their local current label configuration and local conditional probability $P(x_i \mid \boldsymbol{x}_{\partial_i})$. The process loops over all code names in order to update the entire label field. The chromatic sampler is much faster than the conventional Gibbs sampler as it vectorizes the sampling process by only looping over a small number of "colors" rather than all pixels. Detailed information can be found in the corresponding author's previous published work (Wang et al. 2017).

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

2) *Simulation of the model parameters from* $P(\boldsymbol{\beta}|\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH})$

During the iterative process, every time after a realization of $\boldsymbol{x}_{unknown}$ is simulated, $\boldsymbol{\beta}$ is sampled following the conditional posterior distribution:

$$\text{Post}(\boldsymbol{\beta}) \propto \text{Prior}(\boldsymbol{\beta}) L(\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH} \mid \boldsymbol{\beta}) \tag{7}$$

where $\text{Post}(\boldsymbol{\beta})$ and $\text{Prior}(\boldsymbol{\beta})$ are the posterior distribution and prior distribution of $\boldsymbol{\beta}$, respectively; $L(\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH}|\boldsymbol{\beta})$ is the likelihood function indicating the possibility of observing the simulated soil configuration together with the borehole information. In this work, a multivariate Gaussian distribution is adopted as $\text{Prior}(\boldsymbol{\beta})$ with 1) a mean vector $\boldsymbol{\mu}$ indicating the rough estimates of the granularity coefficients, and 2) a diagonal covariate matrix $\Sigma_\beta = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2)$, where σ_i is the standard deviation of the corresponding granularity coefficient. The likelihood function $L(\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH}|\boldsymbol{\beta})$ can be evaluated via the following equation:

$$L(\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH} \mid \boldsymbol{\beta}) = \prod_{x_i \in \{\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH}\}} P(x_i \mid \boldsymbol{x}_{\partial_i}; \boldsymbol{\beta}) = \prod_{x_i \in \{\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH}\}} \frac{\exp[-U(x_i, \boldsymbol{x}_{\partial_i})]}{\sum_{x_i' \in L} \exp[-U(x_i', \boldsymbol{x}_{\partial_i})]} \tag{8}$$

To update $\boldsymbol{\beta}$, the Metropolis-Hastings (M-H) algorithm (Hastings 1970) is employed to implement the MCMC sampling process. The $\log(\text{target})$ function is expressed as:

$$\log(\text{target}) = \log(\text{Prior}(\boldsymbol{\beta})) + \log(L(\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH} \mid \boldsymbol{\beta})). \tag{9}$$

The $\log(\text{target})$ measures the log scale of the joint probability of the simulated granularity coefficients and the simulated soil profile. Higher values of the $\log(\text{target})$ indicate higher possibility that the simulated soil profile is plausible and the corresponding granularity coefficients is compatible with the simulated field. Through

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

the M-H algorithm, the log(target) function is being optimized in a probabilistic sense and the posterior of $\boldsymbol{\beta}$ can be estimated from generated samples.

It is worth noting that the Bayesian learning implemented by MCMC is not just for smoothing (or removing noise from) the soil stratigraphic profile $\boldsymbol{x}_{unknown}$. More importantly, it is designed for extracting the spatial pattern of the soil stratigraphic profile characterized by the granularity coefficients $\boldsymbol{\beta}$, which serves as digitized geological knowledge of a site and can be used to quantitatively measure stratigraphic similarity across different sites or to simulate soil profile realizations at a similar site. This advantage shows the added value of the developed Bayesian learning process.

2.3. Discriminant adaptive nearest neighbor-based k-harmonic mean distance (DANN-KHMD) classifier

For generating reasonable initial fields, the DANN-KHMD classifier is developed to label $\boldsymbol{x}_{unknown}$ using long-range spatial pattern learned from $\boldsymbol{x}_{BH}$. It is essentially an approach to roughly "guess" possible labels of $\boldsymbol{x}_{unknown}$ given $\boldsymbol{x}_{BH}$ in a probabilistic manner. Accordingly, the initial fields can be sampled independently on each pixel solely via the DANN-KHMD classifier.

Briefly, DANN algorithm utilizes a suitable pixel-centered learning domain (defined in the physical space, i.e., framed by width and depth in the current problem) to compute a local metric Ψ_i of unknown pixel i using coordinates and labels of known pixels within the learning domain. Ψ_i can locally adapt to the optimal decision boundary and behaves similar to but more robust than a linear discriminant classifier

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

(Hastie and Tibshirani 1996). The nearest neighboring known pixels around pixel *i* can then be identified in terms of DANN distance using the metric Ψ_i. The harmonic mean of the k nearest DANN distances about a certain label is then used to robustly assess the possibility of choosing that label. The technical details are provided below.

1) *Estimate the DANN distance*

The metric Ψ_i at an unknown pixel i is evaluated over a rectangular $2h \times 2v$ learning domain $\aleph_i$, which is centered at i and includes all pixels in the rectangular region defined by $2h$ and $2v$ (see Fig. 3a), where h and v are measured in the number of pixels. The label configuration of the known pixels in $\aleph_i$ is used to compute the metric Ψ_i. Note that this learning domain is defined in the physical space rather than on the graph and hence it has nothing to do with MRF.

The DANN (squared) distance $D(j,i)$ between pixels j and the center unknown pixel i can be calculated using the local metric Ψ_i (Friedman et al. 2001; Hastie and Tibshirani 1996):

$$D(j,i) = (\boldsymbol{v}_j - \boldsymbol{v}_i)^{\mathrm{T}} \Psi_i (\boldsymbol{v}_j - \boldsymbol{v}_i) \tag{10}$$

in which $\boldsymbol{v}_j = [c_j, r_j]^{\mathrm{T}}$ and $\boldsymbol{v}_i = [c_i, r_i]^{\mathrm{T}}$ are the coordinates of pixels in the original physical space, where c and r are the column and row indices, respectively. The metric Ψ_i is defined as follow (Friedman et al. 2001; Hastie and Tibshirani 1996):

$$\Psi_i = \mathbf{W}^{-1/2} \left[\mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2} + \varepsilon \mathbf{I} \right] \mathbf{W}^{-1/2} \tag{11}$$

where ε is a tuning parameter. The term $\varepsilon \mathbf{I}$ is used to round the metric Ψ_i from an infinite strip shape to an ellipse and hence avoids using known pixels far away from the

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

center pixel. Ψ_i approximately behaves like a linear discriminant metric when $\varepsilon = 0$. $\mathbf{W}$ and $\mathbf{B}$ are the within-label covariance matrix and the between-label covariance matrix, respectively. They are calculated using known pixels $j, j \in \aleph_i^*$, where $\aleph_i^*$ is a subset of $\aleph_i$ only consisting of all known pixels, by following the equations below (Friedman et al. 2001):

$$\mathbf{W} = \sum_{m=1}^{l} f^{(m)} \mathbf{C}^{(m)} \tag{12}$$

$$\mathbf{B} = \sum_{m=1}^{l} f^{(m)} (\bar{\boldsymbol{v}}^{(m)} - \bar{\boldsymbol{v}})(\bar{\boldsymbol{v}}^{(m)} - \bar{\boldsymbol{v}})^T \tag{13}$$

in which $\mathbf{C}^{(m)}$ is the covariance of pixel coordinates in $\aleph_i^{*(m)}$, where $\aleph_i^{*(m)}$ is a subset of $\aleph_i^*$ consisting of known pixels having the same label *m*. $f^{(m)} = \left|\aleph_i^{*(m)}\right| / \left|\aleph_i^*\right|$ is the proportion of $\aleph_i^{*(m)}$ in $\aleph_i^*$. $\bar{\boldsymbol{v}}$ is the centroid of $\aleph_i^*$ and $\bar{\boldsymbol{v}}^{(m)}$ is the centroid of $\aleph_i^{*(m)}$.

Fig. 3b illustrates the contour map of $D(j,i)$ about a given pixel i in physical space calculated from the learning domain $\aleph_i$ (Fig. 3a). It can be noticed that the elliptical iso-distance lines are stretched along the optimal decision boundary (Fig. 3b). For a better understanding, Fig. 3c shows the corresponding contour map in the $\Psi_i^{1/2}$-transformed space, in which $D(j,i)$ can be simply considered as the squared Euclidean distance. We can see that, in this transformed space, the known pixels are compressed horizontally (more significantly) and vertically (less significantly) so that the similarity (in terms of Euclidean distance) along the optimal decision boundary can be pronounced. One-step further, Fig. 3d shows the local metric shape (i.e., the ellipse

representing the shape of iso-distance lines) of four pixels at different locations in physical space. As can be seen, for Points 1, 2, 3 near the stratigraphic boundary, the sizes, shapes, and rotations of their local metrics are different and controlled by their local boundary conditions. In the pure regions with only one soil type, such as Point 4, the metric shape remains circular as its between-matrix $\mathbf{B}=0$ and Ψ_i is the identify matrix.

2) *The KHMD rule based on DANN distances*

Given a dataset with t elements $\{y_1, y_2, ..., y_t\}$, its harmonic average is defined as:

$$HA = \frac{t}{\sum_{i=1}^{t} \frac{1}{y_i}} \tag{14}$$

Similarly, the harmonic mean distance (HMD) can be calculated with the formulation below (Pan et al. 2017):

$$HMD_i^{(m)} = \frac{k}{\sum_{j=1}^{k} \frac{1}{D(j,i)}}, x_j = m \tag{15}$$

where $HMD_i^{(m)}$ is the HMD of the k nearest known pixels having label m to the unknown pixel i. Note that the k nearest known pixels are selected based on $D(j,i)$ rather than the Euclidean distance. $HMD_i^{(m)}$ can be considered as a robust measure of the similarity between pixel i and nearby known pixels with label *m*. In other words, $HMD_i^{(m)}$ indicates the local preference of label *m*. Recall $V_i(x_i)$ in Eq. (5) is the potential function defined solely on pixel *i* and indicates the preference of choosing different labels at pixel *i*, it is reasonable to set $V_i(m) = HMD_i^{(m)}$. If no contextual

constraints is considered, the local probabilities $Pro_i^{(m)}$ of choosing label m for pixel i can be calculated with the form below following Besag (1986) and Geman and Geman (1984):

$$Pro_i^{(m)} = \frac{\exp(-HMD_i^{(m)})}{\sum_{m' \in \mathbf{L}} \exp(-HMD_i^{(m')})} \tag{16}$$

The initial stratigraphic configuration can be sampled directly and independently via Eq. (16) at each unknown pixel. In order to hold the local label preference characterized by $HMD_i^{(m)}$ throughout the stochastic simulation, the HMD is integrated into the local energy $U(x_i, \boldsymbol{x}_{\partial_i}) = HMD_i^{(x_i)} + \sum_{j \in \partial_i} V_{i,j}(x_i, x_j)$ and the conditional probability Eq. (4) for a local neighborhood system is updated as:

$$P(x_i \mid \boldsymbol{x}_{\partial_i}) = \frac{\exp[-(HMD_i^{(x_i)} + \sum_{j \in \partial_i} V_{i,j}(x_i, x_j))]}{\sum_{x_i' \in \mathbf{L}} \exp[-(HMD_i^{(x_i')} + \sum_{j \in \partial_i} V_{i,j}(x_i', x_j))]} \tag{17}$$

Accordingly, the likelihood function $L(\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH} \mid \boldsymbol{\beta})$ Eq. (8) is updated as:

$$L(\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH} \mid \boldsymbol{\beta}) = \prod_{x_i \in \{\boldsymbol{x}_{unknown}, \boldsymbol{x}_{BH}\}} \frac{\exp[-(HMD_i^{(x_i)} + \sum_{j \in \partial_i} V_{i,j}(x_i, x_j))]}{\sum_{x_i' \in \mathbf{L}} \exp[-(HMD_i^{(x_i')} + \sum_{j \in \partial_i} V_{i,j}(x_i', x_j))]} \tag{18}$$

3) *Procedure of parameter sensitivity analysis*

The DANN-KHMD procedure has a few tuning parameters: a) h, v, i.e., the size of the learning domain $\aleph_i$; b) k, i.e., the number of nearest known pixels having the same label in the KHMD rule; c) ε, the roundness parameter. The experiments in Hastie and Tibshirani (1996) suggest that ε can be fixed for different datasets, and $\varepsilon = 1$ is adopted by following Hastie and Tibshirani (1996). The model behavior corresponding to different h, v, and k can be analyzed as presented below.

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

The sensitivity of these parameters is analyzed through the error rates of the inferred stratigraphic profiles acquired using different parameter values. The error rate (ER) of an inferred stratigraphic profile is defined as:

$$ER = \frac{\sum_{i=1}^{s} IN(x_i^{(R)} \neq x_i^{(T)})}{s} \tag{19}$$

in which $IN(\cdot)$ is an indication function and equals to 1 when the label of pixel i in the inferred profile (i.e., $x_i^{(R)}$) has a different label as that in the original profile (i.e., $x_i^{(T)}$). s is the total number of pixels in the modeling domain. For simplicity, the parameters h, v and k will be discussed separately with a few testing cases.

The parameter k can be set to 5 first following Pan et al. (2017). Parameters h and v are increased from small to large, which results in an increasing number of known pixels in the learning domain $\aleph_i$. In the first step, the parameter v can be set to an appropriate value (neither too big nor too small based on the size of the modeling domain), and parameter h increases from the maximum distance between two neighboring boreholes in the modeling domain to a full width of the modeling domain. The value of h corresponding to the lowest error rate can be selected as the optimal. Then, parameter v is increased from 1 to an appropriate number with the optimal h. The value corresponding to the lowest error rate is the optimal value for v. Then, along with the optimal h and v, the parameter k increases from 1 to an appropriate number. The optimal value of k corresponds to the lowest error rate. Detailed results of this procedure using a synthetic example are discussed in Section 3. And we will

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

demonstrate that these three tuning parameters are case invariant and hence can be left as default in implementation for most cases.

2.4. Uncertainty quantification

The estimated stratigraphic uncertainty using an MRF model includes *local uncertainty* and *global uncertainty* in *configuration space*. Local uncertainty exists in a single MCMC simulation, and each realization in a single simulation is the result of the local uncertainty. While the global uncertainty refers to the variation of the realizations simulated in different Markov chains. The global uncertainty is caused by the variation of the entire initial field in each simulation. In contrast, given a specific initial field, the driving force of local uncertainty only comes from the spatial constraints within the local MRF neighborhood system.

The convergence of global uncertainty in a statistical sense is an important requirement to quantify the stratigraphic uncertainty. Thus, a batch simulation needs to be performed. As the granularity coefficients $\boldsymbol{\beta}$ and its prior are only roughly defined initially, the burn-in period of $\boldsymbol{\beta}$ could be fairly long (say, ~1000 iterations). Hence it seriously hampers the computational efficiency. It is therefore recommended to do a pre-test using a single simulation. Utilizing the posterior information of $\boldsymbol{\beta}$ acquired from this pre-test as the prior information of $\boldsymbol{\beta}$ for the batch simulation can significantly improve the computational efficiency.

Given a single simulation, each unknown pixel i has a probability $P_i^{(m)}$ for choosing label m through the following formulation:

$$P_i^{(m)} = \frac{N_i^{(m)}}{N_{ite}} \tag{20}$$

where N_{ite} is the total number of realizations in a single simulation, and $N_i^{(m)}$ is the number of realizations in which label m is assigned at pixel i. Note that the probability $P_i^{(m)}$ is quantified for a single simulation. Then we can have a mean value of $P_i^{(m)}$ across the batch simulation, represented by $MP_i^{(m)}$.

The robust majority vote (RMV) soil profile can be determined using the majority vote principle (i.e., the label having the highest MP_i is chosen as the RMV label of pixel i). The following equation shows the implementation of the majority vote principle for pixel i (Wang et al. 2021):

$$\mathrm{RMV}(i) = \arg\max_m \left(MP_i^{(m)}; m \in \mathbf{L} \right) \tag{21}$$

In order to quantify the uncertainty of the estimated soil labels at each pixel in $\boldsymbol{x}_{unknown}$, the concept of information entropy is adopted here. The visual inspection of information entropy can give a direct overview of uncertainties associated with each unknown pixel in the modeling domain. More details in the context of geological modeling can be found in Wellmann and Regenauer-Lieb (2012). Accordingly, the robust information entropy (RIE) quantifying the global uncertainty at a given pixel i can be expressed as the formulation below (Li et al. 2016c; Shi and Wang 2021c):

$$RIE_i = -\sum_{m \in L} \left[MP_i^{(m)} \log MP_i^{(m)} \right] \tag{22}$$

It is worth noting that the MP of each soil label derived from the realizations has already taken the uncertainty of granularity coefficients $\boldsymbol{\beta}$ into consideration, and thus differs itself from other non-Bayesian methods with fixed model parameters. Hence

this work thoroughly assesses model bias/uncertainty and incorporate it into uncertainty quantification. According to Eq. (22), the RIE is 0 when no uncertainty exists (i.e., one soil label at pixel i has the probability 1) and the RIE will be the highest value when all soil labels are equally probable across multiple simulations. In other words, the higher RIE_i is, the higher uncertainty level it is at the given pixel i, and hence it will be more difficult to determine the soil label given the nearby borehole information.

In addition, for validation purpose only, in this paper, the following formula is used to measure the deviation of the simulation results from the ground truth (if known) (Shi and Wang 2021c):

$$Acc = \frac{\sum_{i=1}^{s} IN(x_i^{(R)} = x_i^{(T)})}{s} \tag{23}$$

in which $IN(\cdot)$ is an indication function and equals to 1 when the label of pixel i in the simulation profile (i.e., $x_i^{(R)}$) has the same label as that in the original profile (i.e., $x_i^{(T)}$). s is the total number of pixels in the modeling domain.

2.5. Implementation

We summarize the proposed approach as *Algorithm A* and it has been implemented in Python 3.7. Interested audiences may contact the corresponding author for the in-house developed Python package "PyMRF".

Algorithm A:

Step 1. Sample an initial field using the DANN-KHMD classifier.

1) Spread out a rectangular $2h \times 2v$ learning domain $\aleph_i$ around an unknown pixel i.

2) Calculate the within-label covariance and between-label covariance matrices $\mathbf{W}$ and $\mathbf{B}$ using Eq. (12) and Eq. (13), respectively.

3) Calculate the local metric Ψ_i using Eq. (11).

4) Compute $D(j,i)$ to all known pixels from unknown pixel i using the local metric Ψ_i via Eq. (10).

5) For label $m \in \mathbf{L}$, find the k nearest known pixels having label m of pixel i in terms of $D(j,i)$, then compute the $HMD_i^{(m)}$ using Eq. (15).

6) Compute the probabilities of choosing label m for pixel i : $Pro_i^{(m)}$ using Eq. (16).

7) Loop over all unknown pixels and generate an initial filed $\boldsymbol{x}_0$ according to the probability $Pro_i^{(m)}$ at each unknown pixel. Note that multiple $\boldsymbol{x}_0$ can be generated for a batch simulation.

Step 2. Conduct Gibbs sampling and update $\boldsymbol{\beta}$.

1) Define a prior distribution $\text{Prior}(\boldsymbol{\beta})$ via a multivariate Gaussian distribution with a mean vector $\boldsymbol{\mu}$ and a diagonal covariate matrix $\boldsymbol{\Sigma}_\beta$. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_\beta$ .can be simply left as default or estimated roughly (less-informative prior) or customized if site-specific knowledge is available as $\text{Prior}(\boldsymbol{\beta})$ is not sensitive to the final estimation results.

2) Provide an initial guess of $\boldsymbol{\beta}$, e.g., $\boldsymbol{\beta}_0 = \boldsymbol{\mu}$, and an initial field $\boldsymbol{x}_0$.

3) Given the current $\boldsymbol{\beta}$ and the current stratigraphic configuration, calculate the conditional probability of all unknown pixels using Eq. (17).

4) Generate an updated stratigraphic configuration $\boldsymbol{x}$ via the chromatic sampler according to the conditional probability acquired in 3) of step 2.

5) Perform a single M-H sampling step for $\boldsymbol{\beta}$ according to Eq. (9) using *Prior*($\boldsymbol{\beta}$) and the likelihood from Eq. (18).

6) Iterate 3)-5) of Step 2 N_{ite} times. This is called a single simulation (or a single Markov chain) hereafter.

Step 3. Perform a batch simulation: Generate N_{sim} initial fields from Step 1 and execute Step 2 N_{sim} times. Then perform uncertainty quantification.

3. Synthetic Case Study

In this section, we validate the proposed approach using two synthetic cases named 1) auto-synthetic case and 2) manual-synthetic case. The auto-synthetic case is an automatically generated "soil profile" using a Gibbs sampler with known MRF granularity coefficients while the manual-synthetic case is a manual drawing of multiple sinusoidal and cosinusoidal curves as synthetic interfaces of different "soil layers". The auto-synthetic case is designed for validating the proposed approach when the homogeneous assumption of $V_{i,j}(x_i, x_j)$ can be fully satisfied, whereas the manual-synthetic case is designed for testing the approach when the homogeneous assumption of $V_{i,j}(x_i, x_j)$ may not hold. Note that in both scenarios, $V_i(x_i)$ is assumed to be non-homogeneous as mentioned in Section 2 above.

More specific, the stratigraphic profile of the auto-synthetic case having size 100×100 pixels is generated using the granularity coefficients $\boldsymbol{\beta}$ =[4.50, 0.15, 0.15,

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

0.15] and is shown in Fig. 4a. Four evenly distributed virtual boreholes annotated by the dashed lines are extracted for inferring the "unknown" pixels as shown in Fig. 4a. The soil profile in Fig. 4a is considered as the "ground truth" throughout the auto-synthetic case. While the manual-synthetic profile is generated in this way: a) the modeling domain is first divided into six areas using five sinusoid and cosinusoid boundary lines (see Fig. 4b), and b) fill the six areas with three different labels (see Fig. 4c). Five virtual boreholes evenly distributed are extracted at the dashed lines shown in Fig. 4c. Then the labels of the five boreholes are used to infer the "unknown" portion and the soil profile in Fig. 4c is considered as the "ground truth" throughout the manual-synthetic case.

3.1. Results of parameters sensitivity analysis

The performance of the proposed approach may be affected by three factors: 1) the quality of the initial field and the estimated $HMD_i^{(m)}$ (i.e., the behavior of the DANN-KHMD classifier); 2) The extent that the underlying stratigraphic profile is compliant with the homogeneous assumption of the potential function $V_{i,j}(x_i, x_j)$ (i.e., the behavior of the MRF model); 3) The amount of known pixels.

Following the procedure discussed in Subsection 2.3, the error rate curves of the parameters h, v, and k in the two synthetic cases are shown in Fig. 5. It should be noted that, in order to eliminate the effects of the third factor, five evenly distributed boreholes are extracted from both two synthetic cases and used for producing Fig. 5. Hence the differences of the two curves only reflect the effects of the first two factors.

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

462 The first finding is that, for each of the three parameters, the trend of the error rate is
463 similar for both cases and the optimal parameters corresponding to their lowest error
464 rate is exactly same for both two cases. This observation means that the optimal choice
465 of the three parameters seems to be case invariant. The second finding is that, in general,
466 the error rate is lower for auto-synthetic case since the homogeneous model assumption
467 is fully satisfied, however, the difference is very small to negligible (~0.2%) if each
468 parameter is set within a reasonable range (say, $h \geq 80$, $4 \leq v \leq 6$, and $k \geq 5$). This
469 observation means that, when the DANN-KHMD parameters are not appropriate and
470 hence the classifier is not effective, possible violation of the homogeneous assumption
471 may jeopardize the performance of the proposed approach while this issue can be
472 significantly mitigated if the DANN-KHMD classifier can function well with
473 appropriate settings. The third finding is that the horizontal (i.e., approximately along
474 the direction of the optimal decision boundary) size h of $\aleph_i$ needs to be sufficiently
475 large in order to incorporate enough known pixels in the learning domain and seems to
476 be the larger the better, while for the vertical (i.e., perpendicular to the optimal decision
477 boundary) size v, once the optimal setting is reached, the DANN-KHMD classifier
478 will perform worse for larger vertical size. The fourth finding is that the parameter k
479 indicating the size of the nearest known pixels with a given label will not further benefit
480 the performance once it is beyond a suitable number.

481 Therefore, based on these new findings, two advantages of the proposed approach
482 can be reflected: 1) The optimal DANN-KHMD parameters seems to be case invariant

and hence can be fixed, say $h=100$, $v=5$, and $k=5$. This default setting is adopted for analyzing all synthetic examples and case histories in this paper. 2) When the optimal DANN-KHMD parameters are used, possible violation of the homogeneous assumption can be well handled thanks to the DANN-KHMD classifier.

3.2. Auto-synthetic case

1) *Pre-test*

As mentioned in *Subsection 2.4*, a pre-test is required before performing a batch simulation. The posterior information of $\boldsymbol{\beta}$ after the burn-in period acquired from a few pre-tests using different prior distribution of granularity coefficients $\boldsymbol{\beta}$ is shown in Table 1. These prior information of $\boldsymbol{\beta}$ can ensure the interval $[\boldsymbol{\mu}-3\boldsymbol{\sigma}, \boldsymbol{\mu}+3\boldsymbol{\sigma}]$ is large enough for covering the reasonable values of $\boldsymbol{\beta}$. As can be noticed, the posteriors are not sensitive to the priors and the inferred $\boldsymbol{\mu}$ is close to the original $\boldsymbol{\beta}$ ($\boldsymbol{\beta}$=[4.50, 0.15, 0.15, 0.15]). Thus, reasonable estimation results can be eventually acquired regardless of prior information of $\boldsymbol{\beta}$ given sufficient known pixels. The reason is that the DANN-KHMD is integrated into the local energy as the non-homogeneous $V_i(x_i)$. This strategy can regularize the estimation of $\boldsymbol{\beta}$ and make it compatible with known pixels (boreholes) with only less- or even non-informative prior. This is an advantage compared with previous similar works using MRF methods (Li et al. 2016c; Wang et al. 2016; Zhao et al. 2021). The simulated MCMC trace of $\boldsymbol{\beta}$ and total energy from one ($\boldsymbol{\mu}$=[1,1,1,1] and $\boldsymbol{\sigma}$=[10,10,10,10]) of the pre-tests are shown in Fig. 6, from which we can see the traces of $\boldsymbol{\beta}$ and total energy converge after around 750

iterations. This is due to a reasonable initial field sampled by and the long-range spatial constraints applied by the DANN-KHMD classifier. Then the posterior of $\boldsymbol{\beta}$ is used as the prior in the following batch simulation in order to eliminate the burn-in period. In this way, it can render the M-H sampler achieve burn-in at the very beginning so that only a small number of iterations (say, 100 iterations) is enough for each simulation. Accordingly, the computational efficiency is significantly improved due to a large number of iterations for burn-in are waived.

2) *Sampling of initial stratigraphic configuration*

DANN-KHMD classifier is employed first to sample the initial field given the known borehole information. A sampled initial field is shown in Fig. 7b. As can be noticed, it is noisy yet similar to the original profile (shown in Fig 7a). This is because no local spatial constraints are applied at this stage. Significant dispersions can be found at the boundaries between soil layers. The reason is that pixels inside each soil layer have a high probability of getting the same label, while pixels near the boundaries have comparable probabilities of getting different labels. Though the initial profile is noisy, the DANN-KHMD classifier can already achieve a good and reasonable approximate of the ground truth via very sparse known information, and the initial field provides the following MCMC algorithm with a good starting point in configuration space and therefore reduces the computational time of the Bayesian optimization process and increase the accuracy and robustness of the simulated profiles.

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

3) *Simulations results*

The following batch simulation contains 100 Markov chains. It can be seen that, from Fig. 7f, the total RIE converges after about 40 simulations. Thus, 100 simulations are sufficient for the convergence of global uncertainty quantification. To better illustrate both local and global uncertainties, the probability-triangle is adopted (Fig. 7c) as there are three labels in this case. A few representative pixels in the original stratigraphic profile (Fig. 7a) are analyzed to illustrate the quantified uncertainty. As can be seen, pixels *a* and *d* are in the interior of formations C and B, respectively; pixels *c*, *e*, *f* are around the interface between formations A and C, formations A and B, formations B and C, respectively; pixels *b*, *g* are around the interface of formations A, B and C. The probability P of these representative pixels choosing different labels and MP (mean of P) are depicted in the probability-triangle. Each point of *P* measures local uncertianty and the scatter illustrates global uncertainty. The triangle is divided into three portions stamped by three labels according to the majority vote rule. Accordingly, the label of the portion that MP_i locates is the RMV label of pixel i. Besides, the probability-triangle can intuitively reflect the position type of pixel i. For instance, it can be seen that, from Fig. 7c, all P_a (red dots) from 100 simulations and MP_a (red pentagram) are 1 for label C, and all the P_d (yellow dots) and MP_d (yellow pentagram) are 1 for label B. Note that these dots are completely covered by the pentagrams. It demonstrates that pixels *a* and *d* are in the interior of formations C and B, respectively. While the probability P of pixels *c*, *e*, *f* varies on the boundaries of

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

the triangle (between labels A and C, labels A and B, labels B and C, respectively), meaning pixels *c*, *e*, *f* are around the interface between two formations. P of pixels *b* and *g* varies in the interior of the triangle, meaning the two pixels are around the interface among multiple (>2) formations.

The RMV profile of the auto-synthetic case is shown in Fig. 7d. It can be seen that the position types of the representative pixels in the RMV profile are similar to their counterparts in the original profile (see Fig. 7a). Comparison between RMV and original profile shows that there are a few differences between them, while these differences are basically concentrated at the junction of multiple formations. To our surprise, based on only 4 boreholes (4% information), the accuracy of the RMV profile is 93.28% against the original profile, which is very impressive considering the complexity of this synthetic soil configuration.

The RIE image is shown in Fig. 7e. It provides us with a clear view of the uncertainties associated with each pixel within the modeling domain. The pixels located in interior region of each formation have extremely low uncertainty level, while the regions with high uncertainty levels are concentrated at the layer interfaces, which agrees well with our intuitions. The RIE of the representative pixels are also marked in Fig. 7e. It again certifies that the uncertainty of pixels is related to their position types.

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

3.3. Manual-synthetic case

In this example, an investigation on a manual-synthetic case is presented for testing the approach when the homogeneous assumption of $V_{i,j}(x_i, x_j)$ may not hold.

Fig. 8b shows an initial field of the manual-synthetic case. Fig. 8f indicates that 100 simulations are enough for the convergence of the global uncertainty as the total RIE converges after about 50 simulations. The prior information of $\boldsymbol{\beta}$ from a pre-test is shown in Table 2. The original profile, the probability-triangle, and the RMV profile are shown in Fig. 8a, Fig. 8c and Fig. 8d, respectively. It can be seen that, from Fig. 8d, pixel *a* is in interior of formation C, pixels *d* and *c* are near the interface between two different formations, and pixels *b*, *e* are near the interface of three different formations. Similar as the auto-synthetic case, Fig. 8c shows that the probability-triangle can properly reflect the position types of different pixels through the quantified local and global uncertainties. The accuracy of RMV profile is 93.96%, which is, again, a remarkable result for this case with very sparse borehole information and incompliance with the homogeneous assumption of $V_{i,j}(x_i, x_j)$. The RIE image is shown in Fig. 8e. It shows the high uncertainty level areas are concentrated at the boundaries and the pixels located in interior of each formation have extremely low uncertainty level. The RIE of the representative pixels marked in Fig. 8e again reflects that the uncertainty of pixels is related to their position types.

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

3.4. Comparison with CMC

The two synthetic cases are also analyzed using the most recently improved CMC approach developed by Zhang et al. (2021). Fig. 9a and Fig. 9c show the RMV profile and *RIE* image of the auto-synthetic case estimated using the proposed MRF model, respectively. Fig. 9b and Fig. 9d show the most likely (ML) profile and the information entropy (IE) image of the auto-synthetic case estimated using the CMC approach, respectively, based on one simulation including 500 realizations as there is no "global uncertainty" concept in a CMC model. When using CMC approach, the most frequently occurred label for each pixel is chosen to generate the ML profile. The overview of Fig. 9a achieves a good estimate of the original profile (see Fig. 4a). The boundaries in Fig. 9b are less smooth. Quantitatively, the accuracy of Fig. 9a is 93.28% and the accuracy of Fig. 9b is 89.13%. It can be seen that, from Fig. 9c and Fig. 9d, the uncertainty in both is mainly concentrated at the layer boundaries. The uncertainty level in Fig. 9d is generally much higher than that in Fig. 9c, especially near the boundaries. Though the uncertainty level in Fig. 9c is lower, it is generally not underestimated since the *local uncertainty* and the *global uncertainty* are thoroughly considered in the process and also can be partially supported by the high accuracy of the RMV profile. Besides, the *RIE* in Fig. 9c is changing smoother than the *IE* from the CMC approach. The uncertainty of three validation boreholes (see red dash lines in Fig. 9a and Fig. 9b) is plotted in Fig. 9e for a more detailed comparison. It can be noted that the uncertainty level estimated via the CMC model is generally higher than that via the proposed MRF

model at multiple boundaries of three verification boreholes, while the overall accuracy of the CMC model is slightly lower than that of the MRF model.

The manual-synthetic case is also studied by using the CMC approach. Fig. 10a and Fig. 10b show the RMV profile inferred using the proposed MRF model and the ML profile inferred using the CMC model, respectively. From the two profiles, we can see the boundaries in Fig. 10a change smoothly, while the boundaries in Fig. 10b change very steeply. Besides, the accuracy of Fig. 10a and Fig. 10b are 93.96% and 88.23%, respectively, showing the proposed approach can achieve a better estimate of subsurface profile. Fig. 10c and Fig. 10d are the RIE image and IE image acquired via the proposed MRF model and the CMC model, respectively. It can be seen that the layer boundaries in Fig. 10c can be easily delineated by the RIE pattern, while that in Fig. 10d is seriously blurred due to high uncertainty level yet the overall accuracy is slightly lower. The uncertainties corresponding to four validation boreholes (Fig. 10e, and see red dash lines in Fig. 10a and Fig. 10b) also support that the uncertainty level from the CMC model is higher than that from the proposed MRF model at boundaries.

The results demonstrate that the proposed MRF model can achieve a more accurate estimate of subsurface profile with lower uncertainty compared with the CMC model. Besides, the profiles acquired via the CMC model have unnatural boundaries. The reason is that there are four different granularity coefficients $\boldsymbol{\beta}$ can represent the spatial constraints in four different directions, while in the CMC model, the HTPM and

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

VTPM can only represent the spatial constraints in two directions. In addition, the HTPM and VTPM are fixed throughout the process of stochastic simulation.

4. Real-world Case Histories

In this section, dataset A (Shi and Wang 2021c) acquired from a reclamation project in Hong Kong (referred to as Hong Kong case), dataset B (Zhang et al. 2021) collected from a construction site of Norway (referred to as Norway case) and dataset C (Shi and Wang 2021b) extracted from a tunnel project in Australia (referred to as Australia case) are studied to illustrate the performance of the proposed approach and compare the proposed approach with existing approaches. The complete soil profile of the Hong Kong case is published in Shi and Wang (2021c) and shown in Fig. 11a. In order to demonstrate the effectiveness of the developed approach, five evenly distributed virtual boreholes are extracted at the dashed lines as shown in Fig. 11a. Fig. 11b shows the collected sparse borehole data (no complete soil profile) of the Norway case. Fig. 11c shows the complete soil profile and the locations (dashed lines) of four evenly distributed virtual boreholes of the Australia case published in Shi and Wang (2021b).

Fig. 12b shows a simulated initial stratigraphic configuration of the Hong Kong case. Though the initial profile is noisy, the DANN-KHMD classifier can achieve a good approximate of the ground truth (see Fig. 12a) via very sparse prior information from real-world data. The posteriors $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ of a pre-test are shown in Table 3 for interested audiences. Then these parameters are used in a batch simulation (test of total

645 RIE versus number of simulations shows that 100 simulations are sufficient) in order
646 to improve computational efficiency. The RMV profile (Fig. 12c), having an accuracy
647 of 91.97%. Fig. 12d shows the RIE image of the Hong Kong case. It can be seen that
648 the high uncertainty areas are mainly distributed around boundaries. Fig. 13 shows the
649 simulation profiles inferred via the proposed approach and the multiple point statistics
650 (MPS) approach developed by Shi and Wang (2021c) with 3, 6 and 11 known
651 boreholes. All inferred profiles agree well with the ground truth. Notably, the profiles
652 estimated via the proposed MRF model all achieve better accuracies (86.41%, 93.44%
653 and 96.25% with 3, 6 and 11 known boreholes, respectively) than the reported profiles
654 (79.9%, 90.2% and 92.8% with 3, 6 and 11 known boreholes, respectively) estimated
655 via MPS in Shi and Wang (2021c). Moreover, the MPS approach requires a training
656 soil profile, while the proposed MRF approach does not. As shown above, the proposed
657 approach demonstrates a decent performance in the Hong Kong case.

658 In the Norway case, 9 schemes using different boreholes (see white dash lines in
659 Fig. 14a and Fig. 14b) are studied using the proposed approach and the improved CMC
660 approach developed by Zhang et al. (2021) and the rest boreholes are reserved for
661 testing. The posterior $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ of pre-tests for 9 schemes are shown in Table 4. It can
662 be seen that the posterior $\boldsymbol{\beta}$ slightly vary in different schemes indicating that the
663 information of $\boldsymbol{\beta}$ can be well estimated with very sparse known boreholes (say, only
664 3) using the proposed approach. Then these parameters are used in a batch simulation
665 (tests show that 100 simulations are sufficient) for corresponding schemes in order to

improve the computational efficiency. Fig. 14a and Fig. 14b show the estimates of schemes 1-9 using the MRF model and the reported estimates of schemes 1-9 using the CMC model in Zhang et al. (2021), respectively. It can be seen that the boundaries in Fig. 14a change relatively smooth, while the boundaries in Fig. 14b are step-like, which may not be observed in real stratum. The accuracies at the testing boreholes in different schemes are evaluated using the following formulation:

$$Acc_{BH} = \frac{\sum_{i=1}^{s_{BH}} IN(x_i^{(R)} = x_i^{(T)})}{s_{BH}} \tag{24}$$

in which $IN(\cdot)$ is an indication function and equals to 1 when the label of pixel i in the simulation profile (i.e., $x_i^{(R)}$) has the same label as that in the original borehole (i.e., $x_i^{(T)}$). s_{BH} is the total number of pixels along the testing borehole. Hence Eq. (24) simply measures the proportion of pixels correctly inferred along a single testing borehole. The statistics of accuracy in different schemes are shown in Table 5. The maximum and minimum accuracy of the realizations, including 100 realizations acquired using the majority vote principle from 100 simulations implemented using the proposed approach, and 500 realizations from a single simulation implemented using the CMC model, at testing boreholes are reported in the table, along with the accuracy of the RMV profile of proposed approach and ML profile of the CMC approach at testing boreholes. It can be seen that the accuracy range (maximum to minimum accuracy) for all boreholes is much narrower from the proposed approach, which is the direct indicator that the quantified uncertainty using the proposed approach is lower (but without underestimating) than that from CMC approach. It can be also reflected

686 from Fig. 14c and Fig. 14d. For a quantitative comparison, the borehole accuracy of the
687 proposed approach higher or lower than that of the CMC model by more than 0.05 is
688 marked by blue and red in Table 5, respectively. As can be seen, in general, the
689 proposed approach outperforms the CMC model. Again, the proposed approach
690 demonstrates a good performance in the Norway case.

691 Fig. 15b shows a sampled initial stratigraphic configuration of the Australia case.
692 Again, the DANN-KHMD classifier can produce a reasonably good initial guess of the
693 ground truth (see Fig. 15a) via very sparse prior information. Table 6 shows the
694 posterior $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ of the granularity coefficients from a pre-test for interested
695 audiences. Then these parameters are used in a batch simulation consisting of 100
696 chains. Fig. 15c and Fig. 15d show the RIE image and RMV profile of the Australia
697 case, respectively. As we can see, the RMV profile has an accuracy of 94.8% and the
698 high uncertainty areas are mainly around boundaries. Fig. 15e shows the stratigraphic
699 profile inferred via the IC-XGBoost and reported in Shi and Wang (2021b) and the
700 accuracy is 91.2%. Furthermore, the IC-XGBoost approach also requires a training soil
701 profile, while the proposed MRF approach does not. As demonstrated in this example,
702 compared with the IC-XGBoost approach, the proposed approach can achieve slightly
703 higher accurate stratigraphic profile based on less prior knowledge of the site.

704 Based on the analyzing results of the three real case histories, the stratigraphic
705 profiles estimated using the proposed approach generally have higher accuracy than
706 that using the MPS approach, CMC approach, or IC-XGBoost approach. Moreover, the

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

inferred layer boundaries using the proposed approach does not require additional training profiles and more consistent with the morphology of natural sedimentary strata than that using CMC model.

5. Computational Cost Analysis

The procedure of the proposed approach is divided into four stages to analyze the computational cost: 1) prediction (i.e., get the local probabilities $Pro_i^{(m)}$ using DANN-KHMD classifier); 2) sampling (i.e., sample 100 initial fields using the local probabilities $Pro_i^{(m)}$), 3) pre-test (i.e., a single simulation containing 2000 iterations); 3) batch simulation (i.e., 100 simulations and 200 iterations in each simulation). The computational cost is evaluated using a laptop with an Intel CORE i7 CPU and 8G memory. The testing results are shown in Table 7. As can be seen, conducting Stage 1 is generally less than 30s and the cost of Stage 2 is negligible. However, the batch simulation stage is relatively computational expensive and the cost is positively related to the stratigraphic complexity (amount of layer boundaries). Luckily, due to the perfectly parallel nature of the batch simulation stage, the computational time can be significantly reduced by using a multi-core cluster and therefore the scalability of the proposed approach is promising. Future experiments in this regard will be performed.

6. Concluding Remarks

In the present work, we proposed a novel stratigraphic uncertainty quantification approach by integrating the Markov random field (MRF) model and the discriminant adaptive nearest neighbor-based k-harmonic mean distance (DANN-KHMD) classifier

into a Bayesian framework. The proposed approach aims at improving the interpretation performance of stratigraphic uncertainty in terms of accuracy and computational efficiency, especially in the situation that only sparse boreholes are available. The model parameters can be initially defined in terms of prior distributions based on prior geological knowledge if available or leave as default. Later these parameters are further updated with constraints from the site exploration results through Bayesian machine learning. In the proposed framework, reasonable initial stratigraphic configurations can be sampled using DANN-KHMD classifier and the initial configuration provides the following MCMC algorithm with a good initial point in the configuration space and therefore reduces the computational time and increase the accuracy of the simulated profiles.

Two synthetic cases and three real-world case histories are studied to demonstrate the performance of and further validate the proposed approach under real-world conditions. The new approach is also compared with other existing methods. The results show that the proposed MRF model can achieve more accurate results compared with the recent modified CMC approach, MPS approach, or the emerging IC-XGBoost approach. Moreover, the proposed approach does not require any training image in the workflow.

The proposed approach also has certain limitations. First, the spatial variation of geomaterials properties is not incorporated into the scope of the present work. It is still an open question in engineering geology on how to effectively and efficiently quantify

749 the stratigraphic uncertainty and the property uncertainty of geomaterials
750 simultaneously. A brutal simulation taking both uncertainties into consideration is
751 known to be computationally expensive if not intractable. Shi and Wang (2021a) have
752 proposed an approach integrating IC-XGBoost and BCS into a same framework that
753 takes both stratigraphic uncertainty and spatial variability of soil properties into
754 consideration to analyze consolidation settlement. However, a training image is
755 essential for this approach. We see there might be an alternative strategy to combine
756 MRF model with BCS to address this challenge and we have already launched this
757 effort. Second, the real-world sites are three-dimensional (3-D) geological bodies.
758 Applying the proposed approach to 3-D subsurface modeling with quantified
759 uncertainty is the ultimate goal though it is challenging. Actually, we are currently
760 working on extending the current approach, including constructing a 3-D neighborhood
761 system by adding additional granularity coefficients and developing fast simulation
762 algorithms using enhanced chromatic sampler for 3-D sampling. It is expected that the
763 number of the granularity coefficients in 3-D space will increase since there are more
764 independent directions of spatial constraints compared with the 2-D scenario. Strictly
765 speaking, in 3-D space, a full neighborhood system has 13 independent directions,
766 however, this setting can be relaxed to a certain extent if some symmetric assumptions
767 are deemed to be reasonable under certain site conditions. Hence the number of model
768 parameters can be potentially reduced. To have more insights in this regard, the work

for applying the proposed approach to 3-D scenarios is currently on-going. We will report more detailed results on this track in future submissions.

Acknowledgement

This work is partially supported by Central South University under the Project Number 1053320192341. This work is also partially supported by the Ohio Department of Transportation under the Agreement Number 31795, and by the STEM Catalyst grant from the University of Dayton. The financial supports from all sponsors are gratefully acknowledged. Meanwhile, the authors greatly appreciate Dr. Yu Wang for providing the dataset of the Hong Kong case and Australia case, and Dr. Jinzhang Zhang for providing the analyzing results of the Norway case using his recently published CMC simulation approach in order to facilitate the comparison study.

Reference

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological), **36**, 192-225, doi: https://doi.org/10.1111/j.2517-6161.1974.tb00999.x.

Besag, J. 1986. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society: Series B (Methodological), **48**, 259-279, doi: https://doi.org/10.1111/j.2517-6161.1986.tb01412.x.

Cross, G.R. & Jain, A.K. 1983. Markov random field texture models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25-39, doi: https://doi.org/10.1109/TPAMI.1983.4767341.

Daly, C. 2005. Higher order models using entropy, Markov random fields and sequential simulation. *Geostatistics Banff 2004*. Springer, 215-224.

Elfeki, A. & Dekking, M. 2001. A Markov chain model for subsurface characterization: theory and applications. Mathematical geology, **33**, 569-589, doi: https://doi.org/10.1023/A:1011044812133.

Fadlelmula, M.M., Akin, S. & Duzgun, S. 2014. Parameterization of Channelized Training Images: A Novel Approach for Multiple-Point Simulations of Fluvial Reservoirs. *Mathematics of Planet Earth*. Springer, 557-560.

Friedman, J., Hastie, T. & Tibshirani, R. 2001. The elements of statistical learning. Springer series in statistics New York.

Geman, S. & Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 721-741, doi: https://doi.org/10.1109/TPAMI.1984.4767596.

Gong, W., Tang, H., Wang, H., Wang, X. & Juang, C.H. 2019. Probabilistic analysis and design of stabilizing piles in slope considering stratigraphic uncertainty. Engineering Geology, **259**, 105162, doi: https://doi.org/10.1016/j.enggeo.2019.105162.

Gong, W., Zhao, C., Juang, C.H., Tang, H., Wang, H. & Hu, X. 2020. Stratigraphic uncertainty modelling with random field approach. Computers and Geotechnics, **125**, 103681, doi: https://doi.org/10.1016/j.compgeo.2020.103681.

Gong, W., Zhao, C., Juang, C.H., Zhang, Y., Tang, H. & Lu, Y. 2021. Coupled characterization of stratigraphic and geo-properties uncertainties–A conditional random field approach. Engineering Geology, **294**, 106348, doi: https://doi.org/10.1016/j.enggeo.2021.106348.

Hastie, T. & Tibshirani, R. 1996. Discriminant adaptive nearest neighbor classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, **18**, 607-616, doi: https://doi.org/10.1109/34.506411.

Hastings, W.K. 1970. Monte Carlo sampling methods using Markov chains and their applications. doi: https://doi.org/10.1093/biomet/57.1.97.

Hu, L. & Chugunova, T. 2008. Multiple-point geostatistics for modeling subsurface heterogeneity: A comprehensive review. Water Resources Research, **44**, doi: https://doi.org/10.1029/2008WR006993.

Hu, Y. & Wang, Y. 2020. Probabilistic soil classification and stratification in a vertical cross-section from limited cone penetration tests using random field and Monte Carlo simulation. Computers and Geotechnics, **124**, 103634, doi: https://doi.org/10.1016/j.compgeo.2020.103634.

Ising, E. 1925. Beitrag zur theorie des ferromagnetismus. Zeitschrift für Physik, **31**, 253-258, doi: https://doi.org/10.1007/BF02980577.

Jiang, S.-H., Huang, J., Qi, X.-H. & Zhou, C.-B. 2020. Efficient probabilistic back analysis of spatially varying soil parameters for slope reliability assessment. Engineering Geology, **271**, 105597, doi: https://doi.org/10.1016/j.enggeo.2020.105597.

Jiang, S.-H., Papaioannou, I. & Straub, D. 2018. Bayesian updating of slope reliability in spatially variable soils with in-situ measurements. Engineering Geology, **239**, 310-320, doi: https://doi.org/10.1016/j.enggeo.2018.03.021.

Juang, C.H., Zhang, J., Shen, M. & Hu, J. 2019. Probabilistic methods for unified treatment of geotechnical and geological uncertainties in a geotechnical analysis. Engineering Geology, **249**, 148-161, doi: https://doi.org/10.1016/j.enggeo.2018.12.010.

Koller, D. & Friedman, N. 2009. Probabilistic graphical models: principles and techniques. MIT press.

Li, J., Cai, Y., Li, X. & Zhang, L. 2019. Simulating realistic geological stratigraphy using direction-dependent coupled Markov chain model. Computers and Geotechnics, **115**, 103147, doi: https://doi.org/10.1016/j.compgeo.2019.103147.

Li, J., Cassidy, M.J., Huang, J., Zhang, L. & Kelly, R. 2016a. Probabilistic identification of soil stratification. Géotechnique, **66**, 16-26, doi: https://doi.org/10.1680/jgeot.14.P.242.

Li, S.Z. 2009. Markov random field modeling in image analysis. Springer Science & Business Media.

Li, X., Zhang, L.M. & Li, J. 2016b. Using conditioned random field to characterize the variability of geologic profiles. Journal of Geotechnical and Geoenvironmental Engineering, **142**, 04015096, doi: https://doi.org/10.1061/(ASCE)GT.1943-5606.0001428.

Li, Z., Wang, X., Wang, H. & Liang, R.Y. 2016c. Quantifying stratigraphic uncertainties by stochastic simulation techniques based on Markov random field. Engineering Geology, **201**, 106-122, doi: https://doi.org/10.1016/j.enggeo.2015.12.017.

Norberg, T., Rosén, L., Baran, A. & Baran, S. 2002. On modelling discrete geological structures as Markov random fields. Mathematical geology, **34**, 63-77, doi: https://doi.org/10.1023/A:1014079411253.

Pan, Z., Wang, Y. & Ku, W. 2017. A new k-harmonic nearest neighbor classifier based on the multi-local means. Expert Systems with Applications, **67**, 115-125, doi: https://doi.org/10.1016/j.eswa.2016.09.031.

Qi, X.-H., Li, D.-Q., Phoon, K.-K., Cao, Z.-J. & Tang, X.-S. 2016. Simulation of geologic uncertainty using coupled Markov chain. Engineering Geology, **207**, 129-140, doi: https://doi.org/10.1016/j.enggeo.2016.04.017.

Robertson, P.K. 1990. Soil classification using the cone penetration test. Canadian Geotechnical Journal, **27**, 151-158, doi: https://doi.org/10.1139/t90-014.

Shi, C. & Wang, Y. 2021a. Assessment of Reclamation-induced Consolidation Settlement Considering Stratigraphic Uncertainty and Spatial Variability of Soil Properties. Canadian Geotechnical Journal, doi: https://doi.org/10.1139/cgj-2021-0349.

Shi, C. & Wang, Y. 2021b. Development of Subsurface Geological Cross-Section from Limited Site-Specific Boreholes and Prior Geological Knowledge Using Iterative Convolution XGBoost. Journal of Geotechnical and Geoenvironmental Engineering, **147**, 04021082, doi: https://doi.org/10.1061/(ASCE)GT.1943-5606.0002583.

Shi, C. & Wang, Y. 2021c. Nonparametric and data-driven interpolation of subsurface soil stratigraphy from limited data using multiple point statistics. Canadian Geotechnical Journal, **58**, 261-280, doi: https://doi.org/10.1139/cgj-2019-0843.

Shi, C. & Wang, Y. 2021d. Training image selection for development of subsurface geological cross-section by conditional simulations. Engineering Geology, 106415, doi: https://doi.org/10.1016/j.enggeo.2021.106415.

Song, S., Si, B., Herrmann, J.M. & Feng, X. 2016. Local autoencoding for parameter estimation in a hidden Potts-Markov random field. IEEE Transactions on Image Processing, **25**, 2324-2336, doi: https://doi.org/10.1109/TIP.2016.2545299.

Toftaker, H. & Tjelmeland, H. 2013. Construction of binary multi-grid Markov random field prior models from training images. Mathematical Geosciences, **45**, 383-409, doi: https://doi.org/10.1007/s11004-013-9456-3.

Wang, H., Wang, X., Wellmann, J.F. & Liang, R.Y. 2019a. A Bayesian unsupervised learning approach for identifying soil stratification using cone penetration data. Canadian Geotechnical Journal, **56**, 1184-1205, doi: https://doi.org/10.1139/cgj-2017-0709.

Wang, H., Wellmann, J.F., Li, Z., Wang, X. & Liang, R.Y. 2017. A segmentation approach for stochastic geological modeling using hidden Markov random fields. Mathematical Geosciences, **49**, 145-177, doi: https://doi.org/10.1007/s11004-016-9663-9.

Wang, X., Li, Z., Wang, H., Rong, Q. & Liang, R.Y. 2016. Probabilistic analysis of shield-driven tunnel in multiple strata considering stratigraphic uncertainty. Structural Safety, **62**, 88-100, doi: https://doi.org/10.1016/j.strusafe.2016.06.007.

Wang, X., Wang, H. & Liang, R.Y. 2018. A method for slope stability analysis considering subsurface stratigraphic uncertainty. Landslides, **15**, 925-936, doi: https://doi.org/10.1007/s10346-017-0925-5.

Wang, X., Wang, H., Liang, R.Y. & Liu, Y. 2019b. A semi-supervised clustering-based approach for stratification identification using borehole and cone penetration test data. Engineering Geology, **248**, 102-116, doi: https://doi.org/10.1016/j.enggeo.2018.11.014.

Wang, Y., Hu, Y. & Zhao, T. 2019c. CPT-based subsurface soil classification and zonation in a 2D vertical cross-section using Bayesian compressive sampling. doi: http://www.nrcresearchpress.com/doi/abs/10.1139/cgj-2019-0131.

Wang, Y., Shi, C. & Li, X. 2021. Machine learning of geological details from borehole logs for development of high-resolution subsurface geological cross-section and geotechnical analysis. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 1-19, doi: https://doi.org/10.1080/17499518.2021.1971254.

Wang, Y. & Zhao, T. 2017. Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling. Géotechnique, **67**, 523-536, doi: https://doi.org/10.1680/jgeot.16.P.143.

Wellmann, J.F. & Regenauer-Lieb, K. 2012. Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models. Tectonophysics, **526**, 207-216, doi: https://doi.org/10.1016/j.tecto.2011.05.001.

Xie, H., Pierce, L.E. & Ulaby, F.T. 2002. SAR speckle reduction using wavelet denoising and Markov random field modeling. IEEE Transactions on geoscience and remote sensing, **40**, 2196-2212, doi: https://doi.org/10.1109/TGRS.2002.802473.

Zhang, J.-Z., Huang, H.-W., Zhang, D.-M., Phoon, K.K., Liu, Z.-Q. & Tang, C. 2021. Quantitative evaluation of geological uncertainty and its influence on tunnel structural performance using improved coupled Markov chain. Acta Geotechnica, 1-16, doi: https://doi.org/10.1007/s11440-021-01287-6.

Zhao, C., Gong, W., Li, T., Juang, C.H., Tang, H. & Wang, H. 2021. Probabilistic characterization of subsurface stratigraphic configuration with modified random field approach. Engineering Geology, **288**, 106138, doi: https://doi.org/10.1016/j.enggeo.2021.106138.

Zhao, T., Hu, Y. & Wang, Y. 2018. Statistical interpretation of spatially varying 2D geo-data from sparse measurements using Bayesian compressive sampling. Engineering Geology, **246**, 162-175, doi: https://doi.org/10.1016/j.enggeo.2018.09.022.

List of symbols

i pixel index

x_i the label of pixel i

∂_i a local neighborhood system of pixel i (MRF)

$\mathbf{S}$ the set of all pixels

s the total number of pixels in the modeling domain

s_{BH} the total number of pixels along testing borehole

930 $\mathbf{L}$ a set of all possible labels
931 Ω the configuration space
932 $\boldsymbol{x}$ a label configuration of all pixels
933 $\boldsymbol{x}_{\mathbf{S}\text{-}\{i\}}$ the labels of pixels in $\mathbf{S}$ but pixel i
934 $\boldsymbol{x}_{\partial_i}$ the labels of pixels in ∂_i
935 $\boldsymbol{x}_{BH}$ the set of pixels with known labels indicating sparse borehole information
936 $\boldsymbol{x}_{unknown}$ the set of pixels with unknown labels
937 $\boldsymbol{\beta}$ the granularity coefficients including $\beta_1, \beta_2, \beta_3, \beta_4$
938 Σ_β the diagonal covariate matrix of $\boldsymbol{\beta}$
939 σ_i the standard deviation of β_i
940 $\boldsymbol{\sigma}$ the standard deviation vector of $\boldsymbol{\beta}$
941 $\boldsymbol{\mu}$ the mean vector of $\boldsymbol{\beta}$
942 Ψ_i the local metric of pixel i
943 $\aleph_i$ a rectangular learning domain of i (DANN)
944 $\aleph_i^*$ the set of pixels with known labels in $\aleph_i$
945 $\aleph_i^{*(m)}$ a subset of $\aleph_i^*$ indicating known pixels j satisfying $x_j = m, j \in \aleph_i^*$
946 h half of the horizontal length of $\aleph_i$
947 v half of the vertical length of $\aleph_i$
948 $\mathbf{B}$ the between-label covariance matrix
949 $\mathbf{W}$ the within-label covariance matrix
950 $\boldsymbol{v}_i$ the coordinates of pixels i
951 $\bar{\boldsymbol{v}}$ the center of $\aleph_i^*$
952 $\bar{\boldsymbol{v}}^{(m)}$ the center of $\aleph_i^{*(m)}$
953 c_i the column index of $\boldsymbol{v}_i$
954 r_i the row index of $\boldsymbol{v}_i$
955 $\mathbf{C}^{(m)}$ the covariance of pixel coordinates in $\aleph_i^{*(m)}$
956 $f^{(m)}$ the proportion of $\aleph_i^{*(m)}$ in $\aleph_i^*$
957 k the number of nearest known pixels in terms of DANN distance
958 $HMD_i^{(m)}$ the harmonic mean distance
959 $Pro_i^{(m)}$ the local probability of choosing label m for pixel i
960 N_{ite} the total number of realizations in the single simulation
961 $N_i^{(m)}$ the number of realizations in which label m is assigned at site i
962 $P_i^{(m)}$ the probability of choosing label m for the unknow pixel i
963 $MP_i^{(m)}$ the mean value of $P_i^{(m)}$
964 RMV(i) the robust majority vote label of pixel i
965 RIE_i the robust information entropy at a given pixel i
966 $IN(\cdot)$ an indication function
967 $x_i^{(R)}$ the label of pixel i in a simulated profile

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

968 $x_i^{(T)}$ the label of pixel i in the original profile

969

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

970 **Table and Figure Captions**

971 **List of table captions**

972 **Table 1.** Posterior $\boldsymbol{\beta}$ of a few pre-tests with different prior $\boldsymbol{\beta}$ in the auto-synthetic

973 case.

974 **Table 2.** Posterior $\boldsymbol{\beta}$ of a pre-test in the manual-synthetic case.

975 **Table 3.** Posterior $\boldsymbol{\beta}$ of a pre-test in the Hong Kong case.

976 **Table 4.** Posterior $\boldsymbol{\beta}$ of a pre-test in the Norway case.

977 **Table 5.** Validation results corresponding to different borehole schemes in the Norway
978 case.

979 **Table 6.** Posterior $\boldsymbol{\beta}$ of a pre-test in the Australia case.

980 **Table 7.** Computational cost of the five cases.
981

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

982 Table 1. Posterior $\boldsymbol{\beta}$ of a few pre-tests with different prior $\boldsymbol{\beta}$ in the auto-synthetic case.

Prior $\boldsymbol{\beta}$	Posterior $\boldsymbol{\mu}$				Posterior $\boldsymbol{\sigma}$			
	μ_1	μ_2	μ_3	μ_4	σ_1	σ_2	σ_3	σ_4
$\boldsymbol{\mu}$=1 $\boldsymbol{\sigma}$=10	4.6	0.24	0.16	0.18	0.31	0.33	0.32	0.37
$\boldsymbol{\mu}$=5 $\boldsymbol{\sigma}$=10	4.65	0.20	0.18	0.15	0.36	0.32	0.39	0.37
$\boldsymbol{\mu}$=10 $\boldsymbol{\sigma}$=10	4.52	0.21	0.20	0.17	0.34	0.40	0.35	0.41
$\boldsymbol{\mu}$=20 $\boldsymbol{\sigma}$=10	4.68	0.26	0.22	0.19	0.41	0.37	0.42	0.44

983

984 Table 2. Posterior $\boldsymbol{\beta}$ of a pre-test in the manual-synthetic case.

Posterior $\boldsymbol{\mu}$				Posterior $\boldsymbol{\sigma}$			
μ_1	μ_2	μ_3	μ_4	σ_1	σ_2	σ_3	σ_4
3.34	1.89	0.19	0.21	0.34	0.28	0.32	0.46

985

986 Table 3. Posterior $\boldsymbol{\beta}$ of a pre-test in the Hong Kong case.

Posterior $\boldsymbol{\mu}$				Posterior $\boldsymbol{\sigma}$			
μ_1	μ_2	μ_3	μ_4	σ_1	σ_2	σ_3	σ_4
4.78	0.37	0.11	0.10	0.33	0.31	0.36	0.38

987

988

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

989 Table 4. Posterior $\boldsymbol{\beta}$ of pre-tests in the Norway case.

Scheme	Posterior $\boldsymbol{\mu}$				Posterior $\boldsymbol{\sigma}$			
	μ_1	μ_2	μ_3	μ_4	σ_1	σ_2	σ_3	σ_4
1	4.10	0.32	0.15	0.07	0.34	0.35	0.32	0.37
2	4.06	0.34	0.13	0.06	0.31	0.38	0.40	0.34
3	4.04	0.35	0.16	0.07	0.35	0.37	0.34	0.40
4	4.09	0.33	0.14	0.08	0.33	0.34	0.32	0.35
5	4.04	0.31	0.15	0.10	0.34	0.36	0.34	0.34
6	4.13	0.36	0.12	0.09	0.32	0.40	0.36	0.31
7	4.07	0.35	0.16	0.11	0.37	0.38	0.34	0.36
8	4.11	0.32	0.17	0.08	0.33	0.36	0.32	0.37
9	4.06	0.34	0.14	0.10	0.34	0.40	0.33	0.31

990

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

991 Table 5. Validation results corresponding to different borehole schemes in the Norway case.

BH	Model	Scheme 1	Scheme 2	Scheme 3	Scheme 4	Scheme 5	Scheme 6	Scheme 7	Scheme 8
BH2	MRF	0.82(0.75-0.85)	0.73(0.7-0.73)	0.8(0.75-0.85)	0.73(0.7-0.73)	/	0.82(0.75-0.9)	/	/
	CMC	0.72(0.65-0.85)	0.63(0.4-0.82)	0.72(0.63-0.85)	0.65(0.5-0.75)		0.72(0.6-0.88)		
BH3	MRF	/	0.72(0.5-0.78)	/	0.64(0.6-0.65)	0.76(0.7-0.82)	/		
	CMC		0.53(0.4-0.75)		0.65(0.4-0.73)	0.68(0.55-0.8)			
BH4	MRF	0.65(0.63-0.67)	0.83(0.8-0.85)	0.85(0.77-0.9)	/	/	0.85(0.75-0. 9)	0.85(0.78-0.9)	
	CMC	0.62(0.45-0.72)	0.85(0.4-0.92)	0.72(0.6-0.87)			0.72(0.6-0.88)	0.72(0.6-0.87)	
BH5	MRF	0.64(0.6-0.65)	/	/	0.75(0.74-0.8)	0.75(0.74-0.8)	/	/	
	CMC	0.48(0.4-0.65)			0.7(0.67-0.85)	0.7(0.63-0.82)			
BH6	MRF	0.72(0.68-0.75)	0.78(0.73-0.8)	0.77(0.73-0.8)	/	/	0.87(0.7-0.93)	0.87(0.8-0.93)	
	CMC	0.63(0.53-0.75)	0.85(0.62-0.9)	0.85(0.57-0.9)			0.83(0.6-0.95)	0.78(0.7-0.95)	
BH7	MRF	0.7(0.68-0.77)	0.73(0.7-0.77)	0.73(0.72-0.77)	0.88(0.78-0.9)	0.88(0.88-0.9)	/	/	0.87(0.8-0.89)
	CMC	0.82(0.65-0.85)	0.75(0.5-0.85)	0.75(0.6-0.87)	0.8(0.72-0.9)	0.8(0.7-0.92)			0.8(0.68-0.9)

992 Note: The values inside and outside the parentheses represent the accuracy range at validation boreholes in 100 realizations (MRF) or 500 realizations (CMC)
993 and the accuracy at validation boreholes in RMV profiles (MRF) or ML profiles (CMC), respectively; the symbol '/' indicates known boreholes; the blue and
994 red parts mean the accuracy of the proposed approach is higher and lower than the accuracy of the CMC approach by more than 0.05, respectively.

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

Table 6. Posterior $\boldsymbol{\beta}$ of a pre-test in the Australia case.

Posterior $\boldsymbol{\mu}$				Posterior $\boldsymbol{\sigma}$			
μ_1	μ_2	μ_3	μ_4	σ_1	σ_2	σ_3	σ_4
3.72	0.29	0.14	0.16	0.43	0.37	0.31	0.41

Table 7. Computational cost of the five cases.

Case	Stage			
	Prediction	Sampling initial field	Pre-test	Batch simulation
Auto-synthetic case	29.35s	0.001s	44.78s	408.47s
Manual-synthetic case	10.62s	0.001s	22.65s	213.68s
Hong Kong case	27.12s	0.001s	117.50s	1078.22s
Norway case (Scheme 3)	7.85s	0.001s	24.59s	230.90s
Australia case	7.25s	0.001s	26.52s	255.73s

Published in Engineering Geology Volume 307, 20 September 2022, 106789
https://doi.org/10.1016/j.enggeo.2022.106789

List of figure captions

Fig. 1. Flowchart of the proposed approach.

Fig. 2. Second-order neighborhood system.

Fig. 3. (a) The rectangular learning domain $\aleph_i$ of pixel *i*; (b) the DANN distance contour map of pixel *i*; (c) the $\Psi_i^{1/2}$-transformed distance contour map of pixel *i*; (d) the local iso-distance ellipse of four pixels.

Fig. 4. (a) Simulated profile of the auto-synthetic case; (b) boundary lines and (c) stratigraphic profile of the manual-synthetic case.

Fig. 5. Error rate versus (a) parameter *h*, (b) parameter *v* and (c) parameter *k* in two synthetic cases.

Fig. 6. Pre-test results of the auto-synthetic case: (a) MCMC traces of $\boldsymbol{\beta}$; (b) total energy along iterations.

Fig. 7. Simulations results of the auto-synthetic case: (a) original profile; (b) an initial field; (c) the probability *P* of these representative pixels in the probability-triangle (Pentagrams represent their *MP*); (d) RMV profile; (e) *RIE* image; (f) total *RIE* versus number of simulations.

Fig. 8. Simulations results of the manual-synthetic case: (a) original profile; (b) an initial field; (c) the probability *P* of these representative pixels in the probability-triangle (Pentagrams represent their *MP*); (d) RMV profile; (e) *RIE* image; (f) total *RIE* versus number of simulations.

Fig. 9. Comparison results of the auto-synthetic case: (a) RMV profile estimated via the proposed approach; (b) ML profile estimated via CMC; (c) *RIE* image acquired via the proposed approach; (d) IE image acquired via CMC; (e) validation results.

Fig. 10. Comparison results of the manual-synthetic case: (a) RMV profile estimated via the proposed approach; (b) ML profile estimated via CMC; (c) *RIE* image acquired via the proposed approach; (d) IE image acquired via CMC; (e) validation results

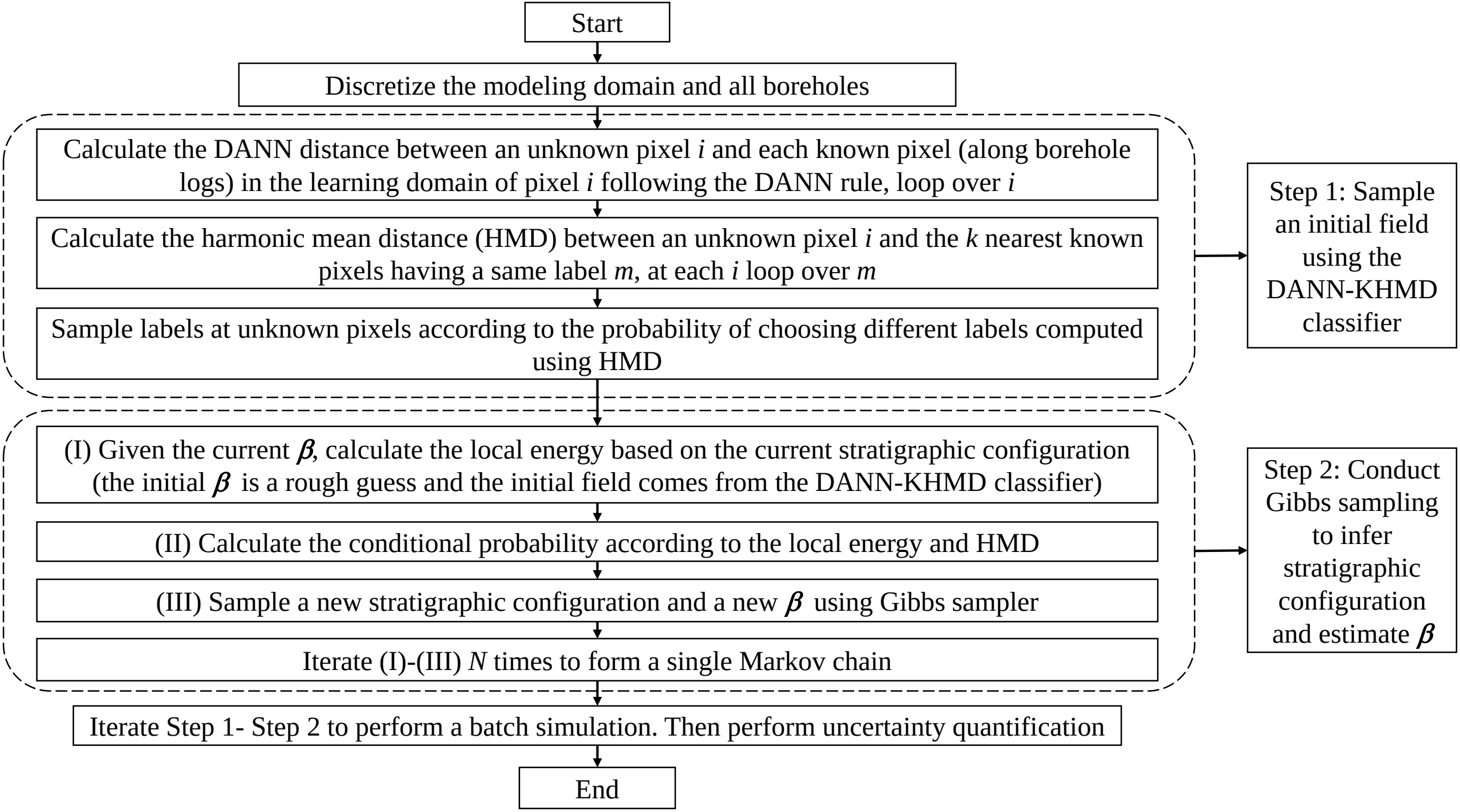
Fig. 11. (a) Real complete stratigraphic profile of the Hong Kong case; (b) collected boreholes of the Norway case; (d) complete stratigraphic profile of the Australia case.

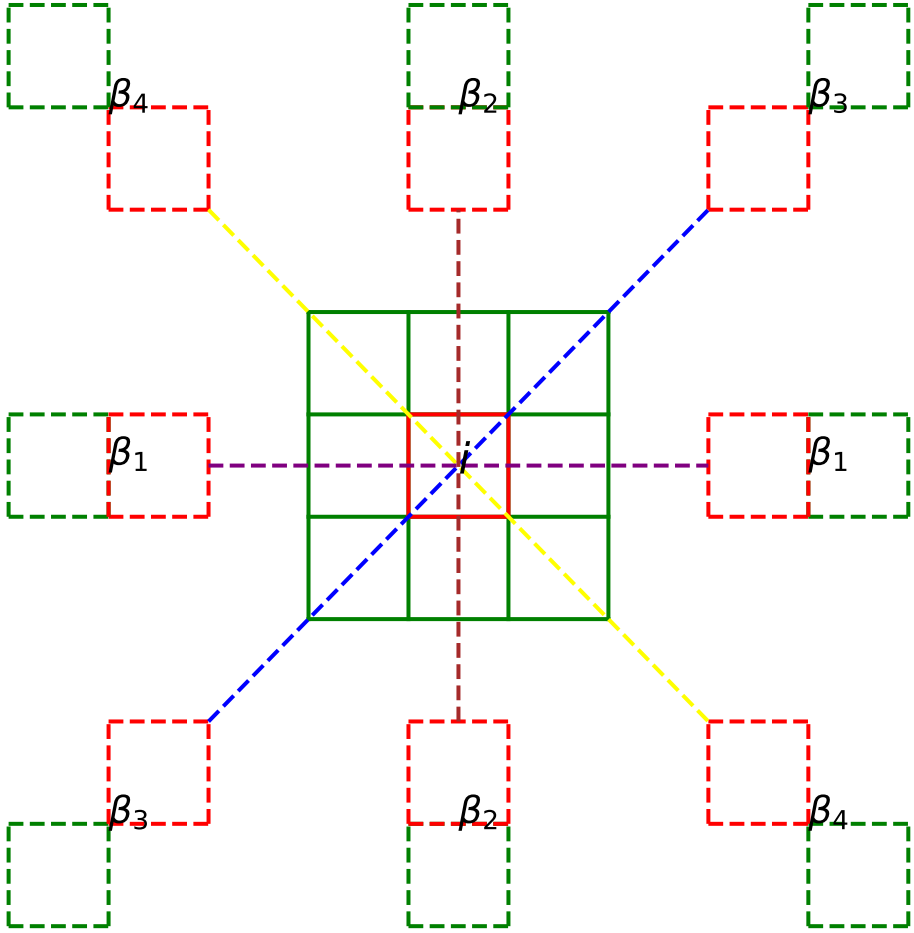
Fig. 12. Simulations results of the Hong Kong case: (a) Original profile; (b) An initial field; (c) RMV profile; (d) *RIE* image.

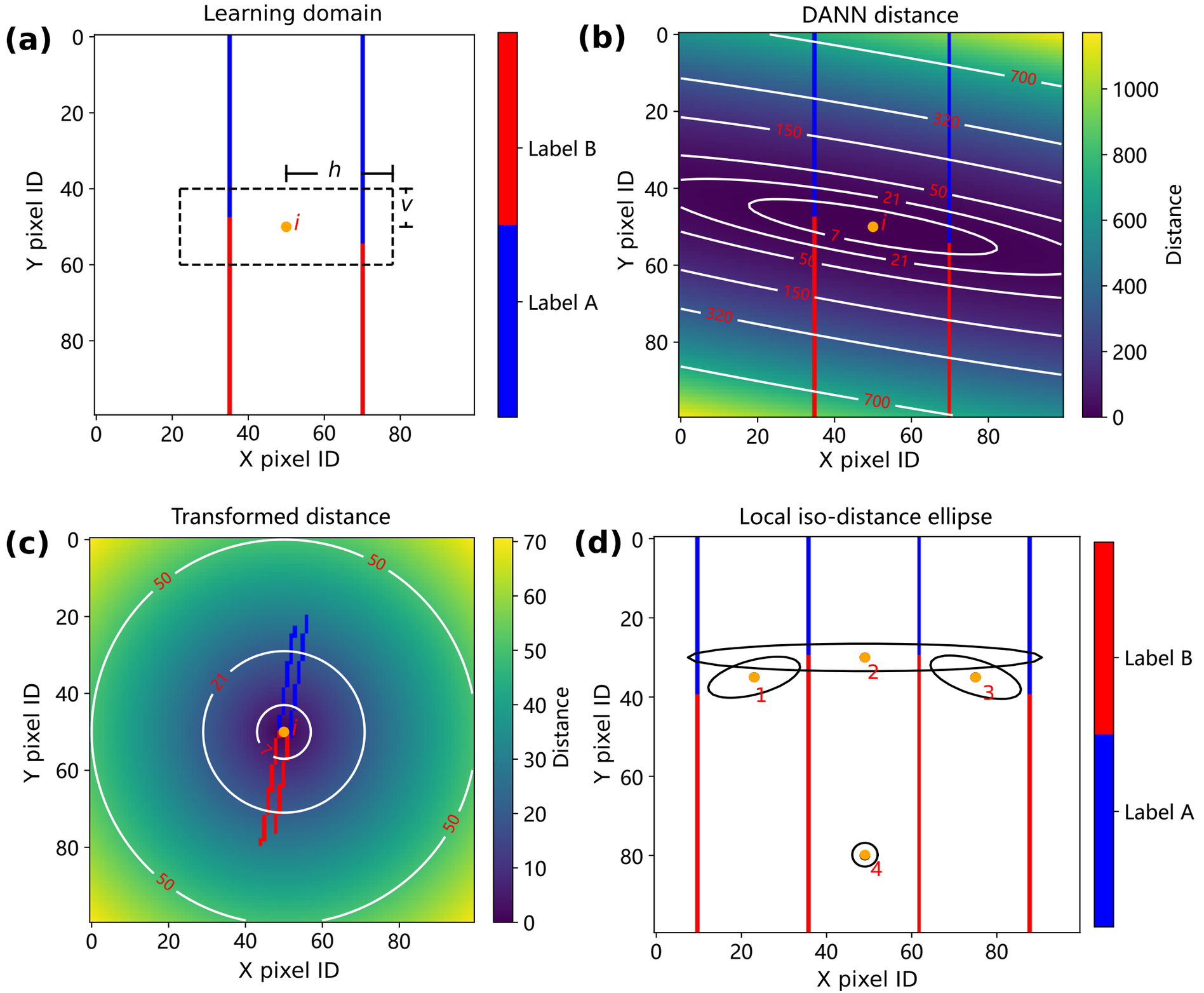
Fig. 13. RMV profiles of the Hong Kong case estimated with (a) 3 boreholes, (b) 6 boreholes and (c) 11 boreholes via the proposed approach; reported profiles estimated with (d) 3 boreholes, (e) 6 boreholes and (f) 11 boreholes via MPS.

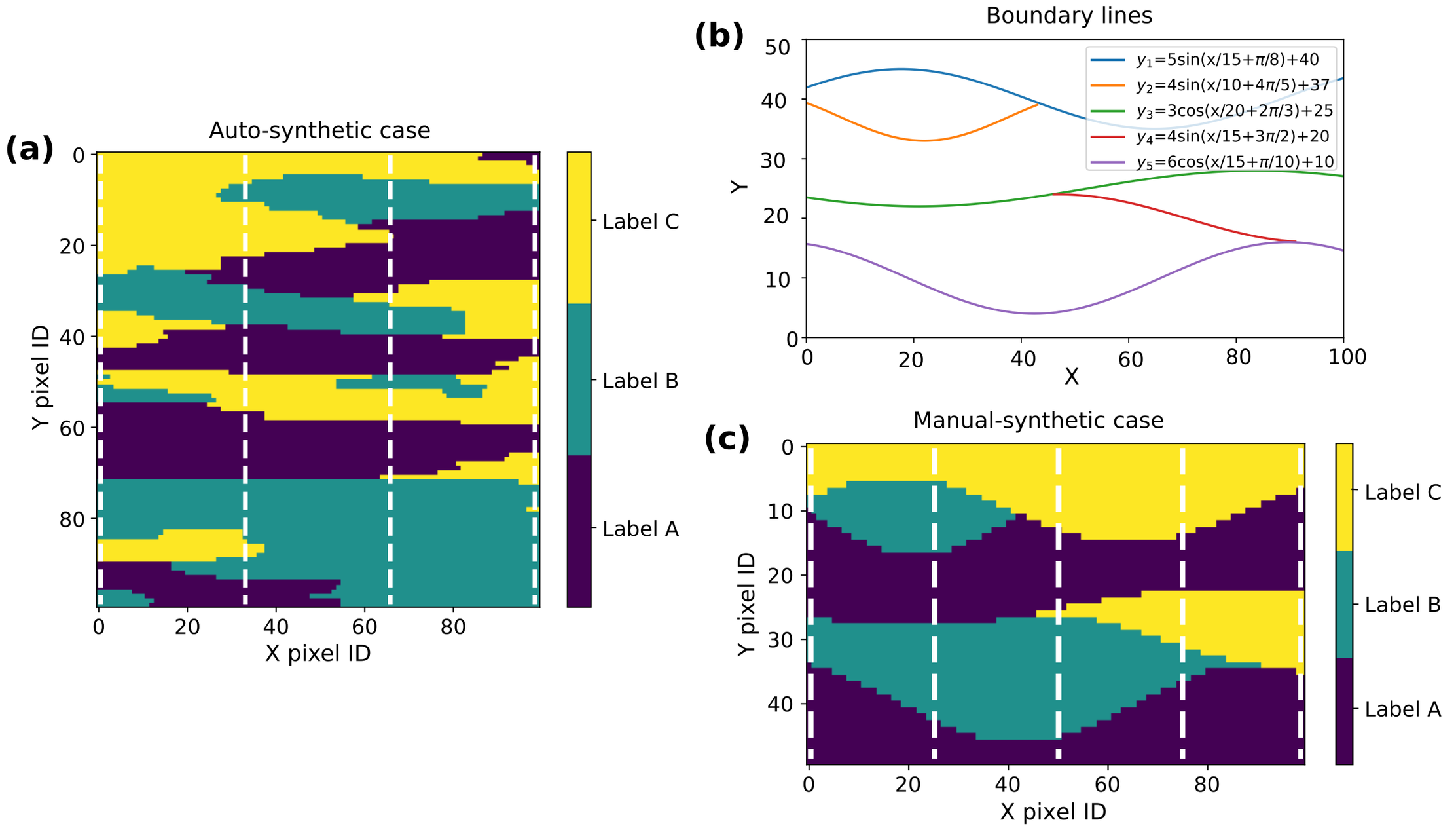
Fig. 14. Simulations results of the Norway case: (a) RMV profiles of scheme 1-9 estimated via the proposed approach; (b) ML profiles of scheme 1-9 estimated via CMC; (c) *RIE* images of scheme 1-9 estimated via the proposed approach; (d) IE images of scheme 1-9 estimated via CMC.

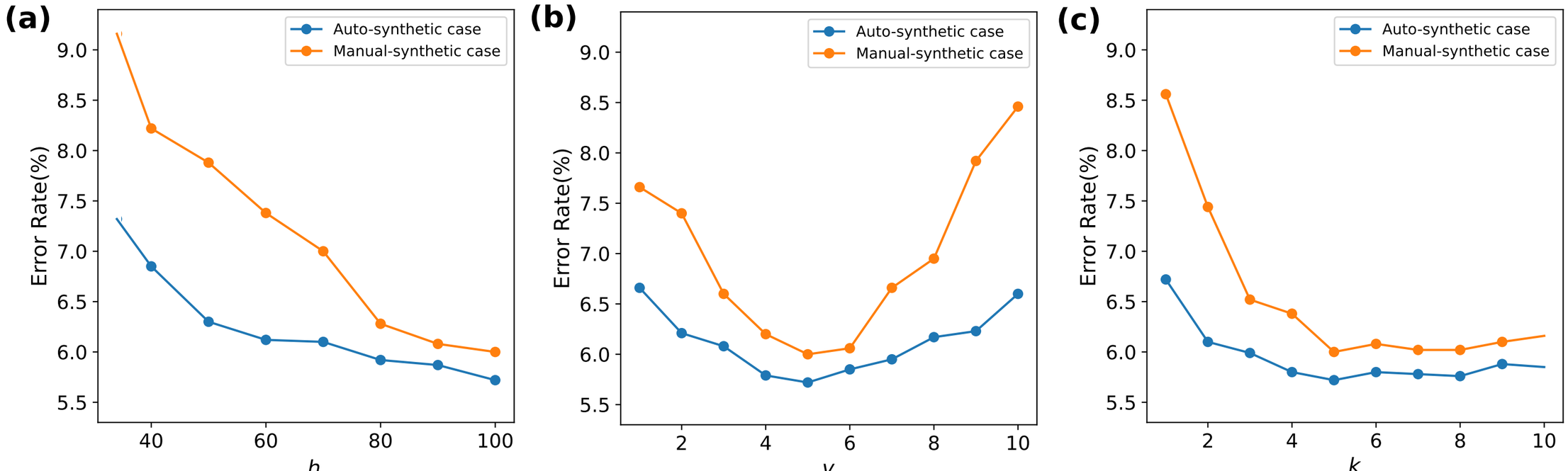
Fig. 15. Simulations results of the Australia case: (a) original profile; (b) an initial field; (c) *RIE* image; (d) RMV profile; (e) reported profiles estimated via the IC-XGBoost approach.

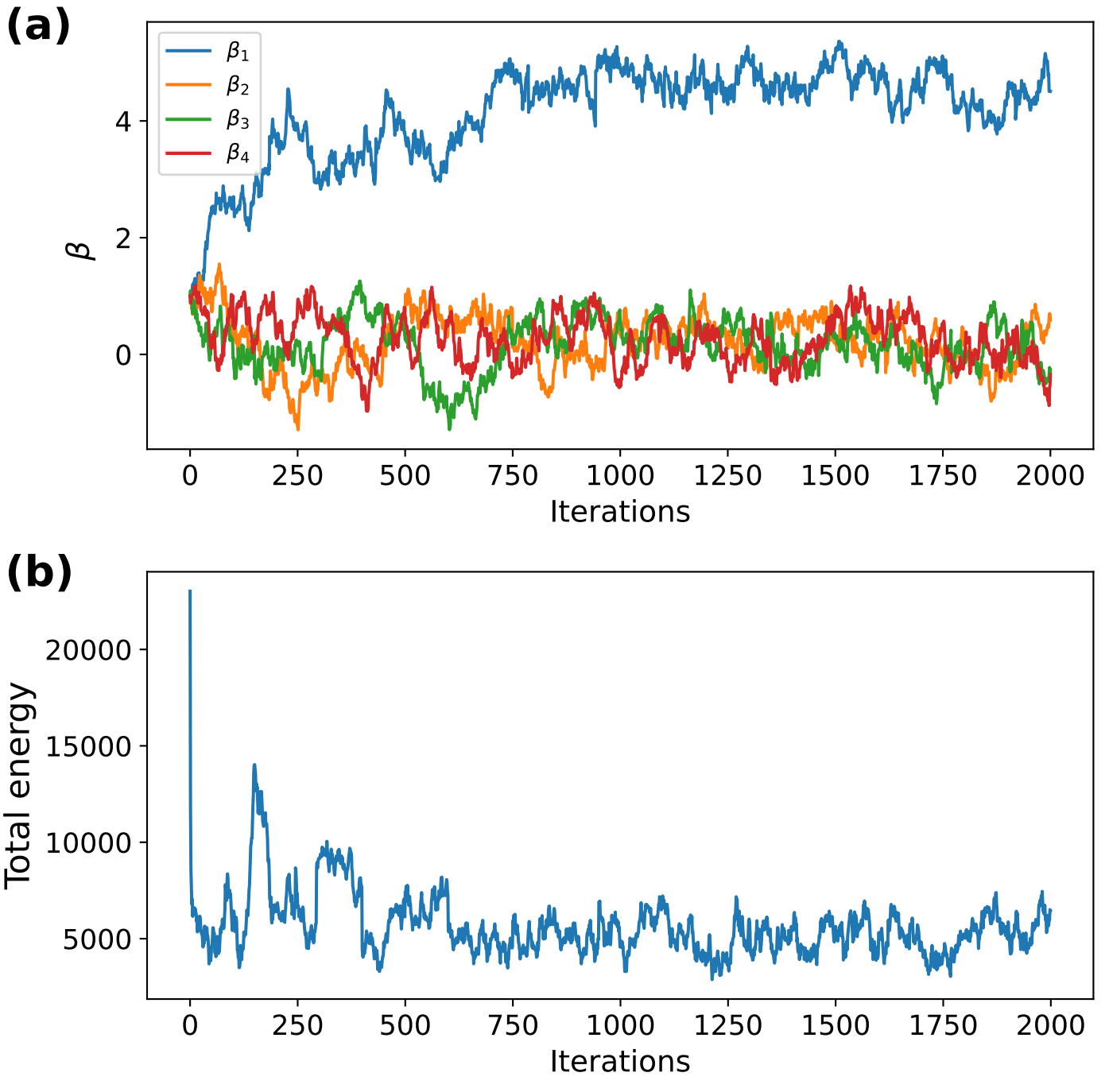

Start
Discretize the modeling domain and all boreholes
Calculate the DANN distance between an unknown pixel *i* and each known pixel (along borehole logs) in the learning domain of pixel *i* following the DANN rule, loop over *i*
Calculate the harmonic mean distance (HMD) between an unknown pixel *i* and the *k* nearest known pixels having a same label *m*, at each *i* loop over *m*
Sample labels at unknown pixels according to the probability of choosing different labels computed using HMD
Step 1: Sample an initial field using the DANN-KHMD classifier
(I) Given the current β, calculate the local energy based on the current stratigraphic configuration (the initial β is a rough guess and the initial field comes from the DANN-KHMD classifier)
(II) Calculate the conditional probability according to the local energy and HMD
(III) Sample a new stratigraphic configuration and a new β using Gibbs sampler
Iterate (I)-(III) *N* times to form a single Markov chain
Step 2: Conduct Gibbs sampling to infer stratigraphic configuration and estimate β
Iterate Step 1- Step 2 to perform a batch simulation. Then perform uncertainty quantification
End

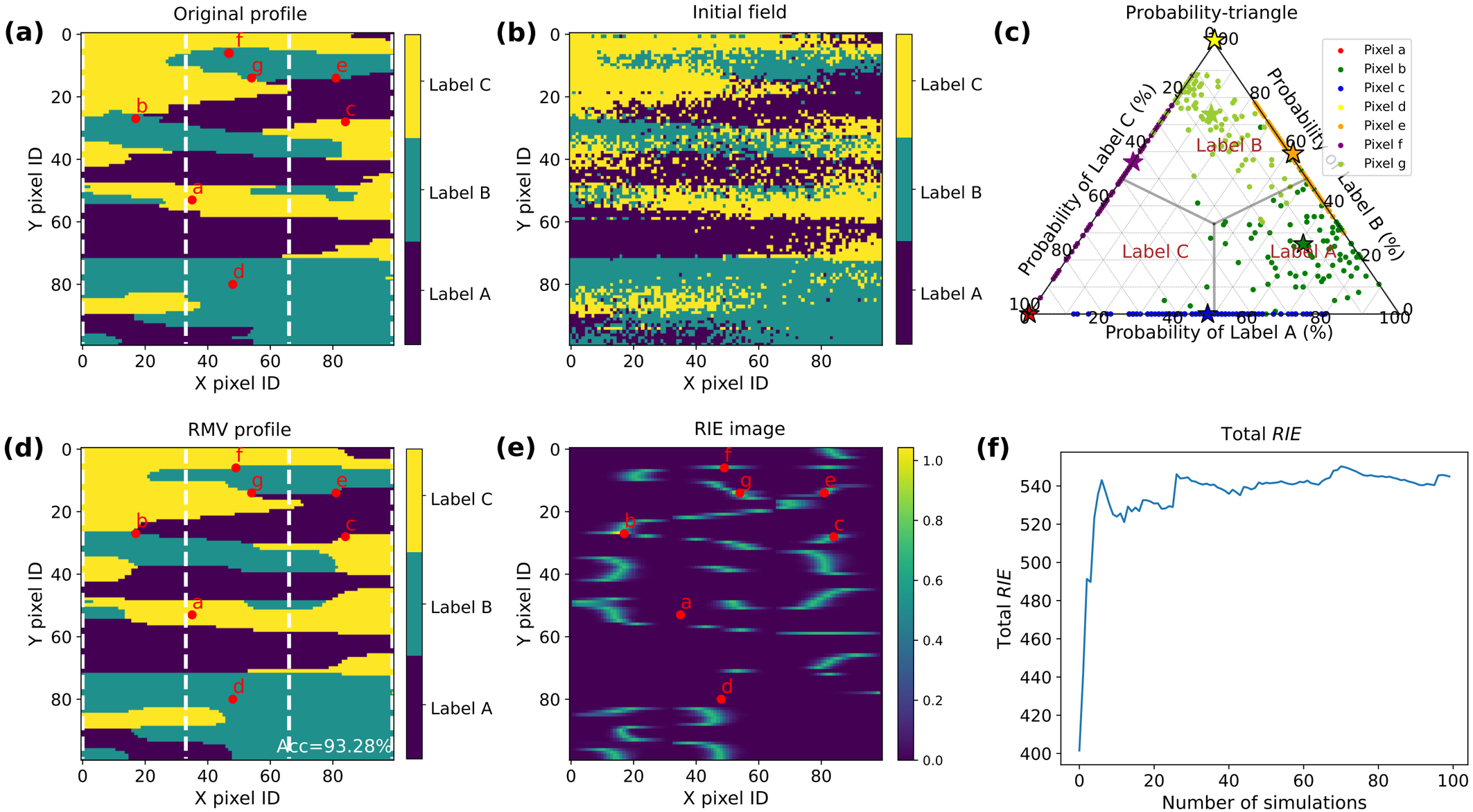

β_4
β_2
β_3
β_1
i
β_1
β_3
β_2
β_4

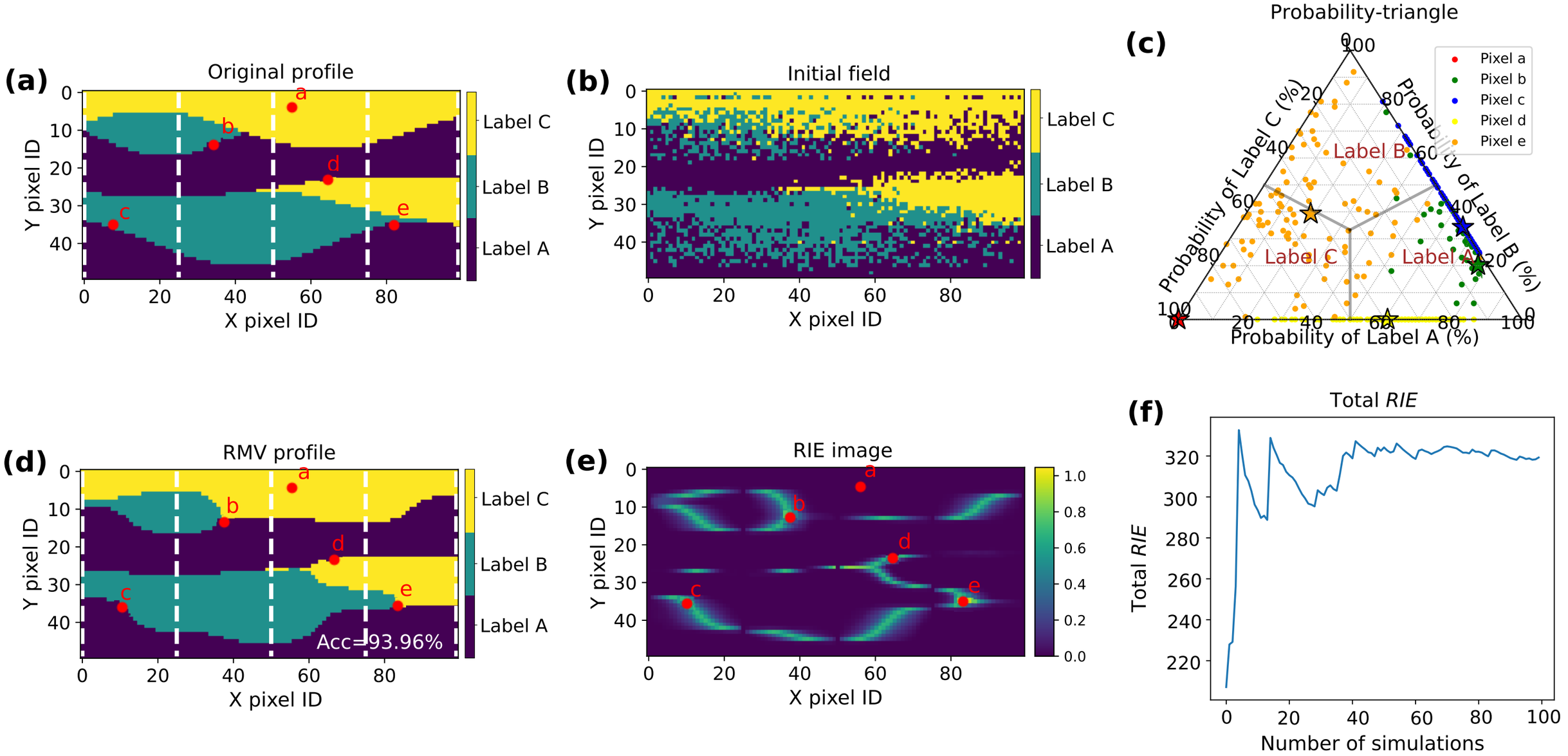

(a)
Learning domain
h
v
i
Label B
Label A
Y pixel ID
X pixel ID
(b)
DANN distance
700
320
150
50
21
7
Distance
(c)
Transformed distance
Distance
(d)
Local iso-distance ellipse
1
2
3
4

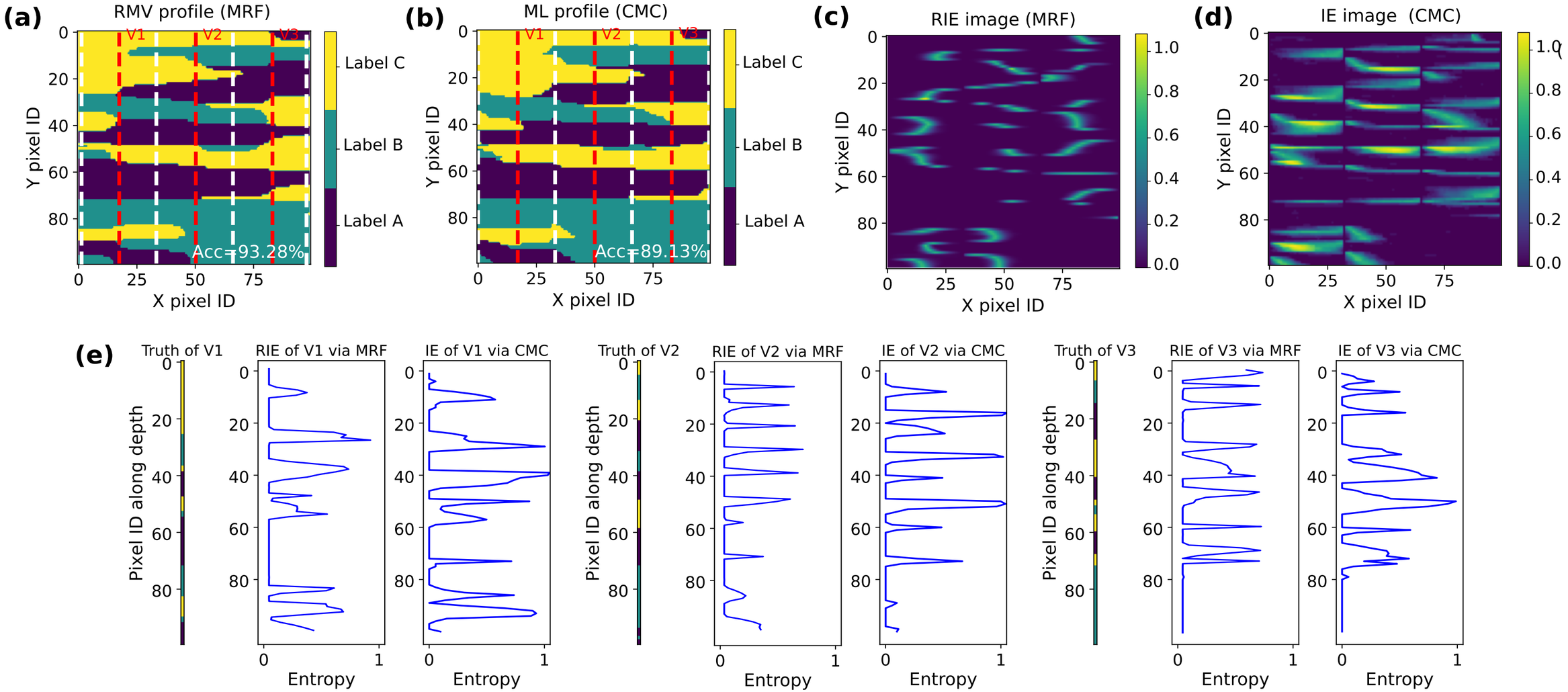

(a)
Auto-synthetic case
Y pixel ID
X pixel ID
Label C
Label B
Label A
(b)
Boundary lines
y_1=5sin(x/15+π/8)+40
y_2=4sin(x/10+4π/5)+37
y_3=3cos(x/20+2π/3)+25
y_4=4sin(x/15+3π/2)+20
y_5=6cos(x/15+π/10)+10
Y
X
(c)
Manual-synthetic case
Y pixel ID
X pixel ID
Label C
Label B
Label A

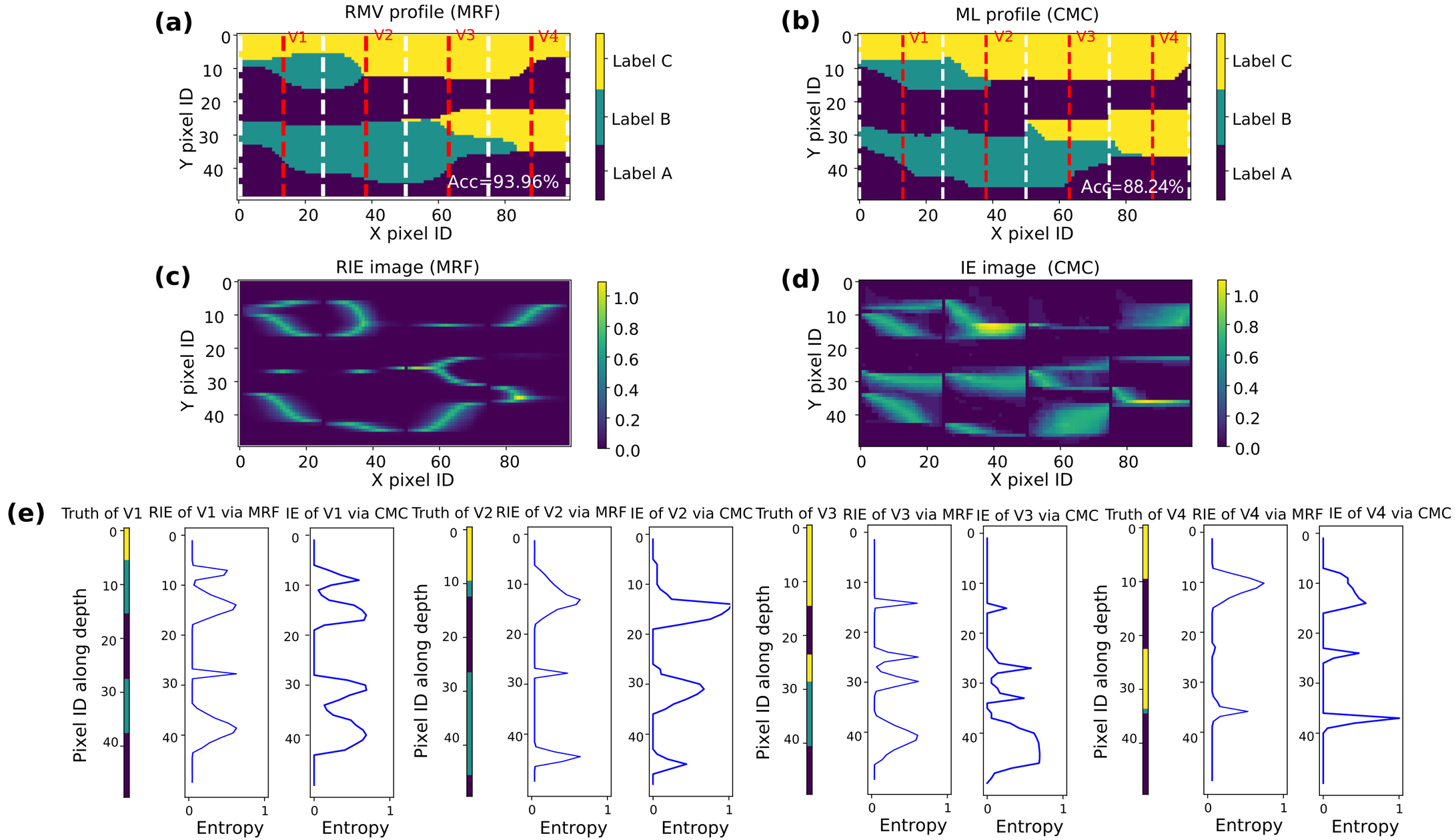

(a)
Auto-synthetic case
Manual-synthetic case
Error Rate(%)
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
40
60
80
100
h
(b)
Auto-synthetic case
Manual-synthetic case
Error Rate(%)
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
2
4
6
8
10
v
(c)
Auto-synthetic case
Manual-synthetic case
Error Rate(%)
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
2
4
6
8
10
k

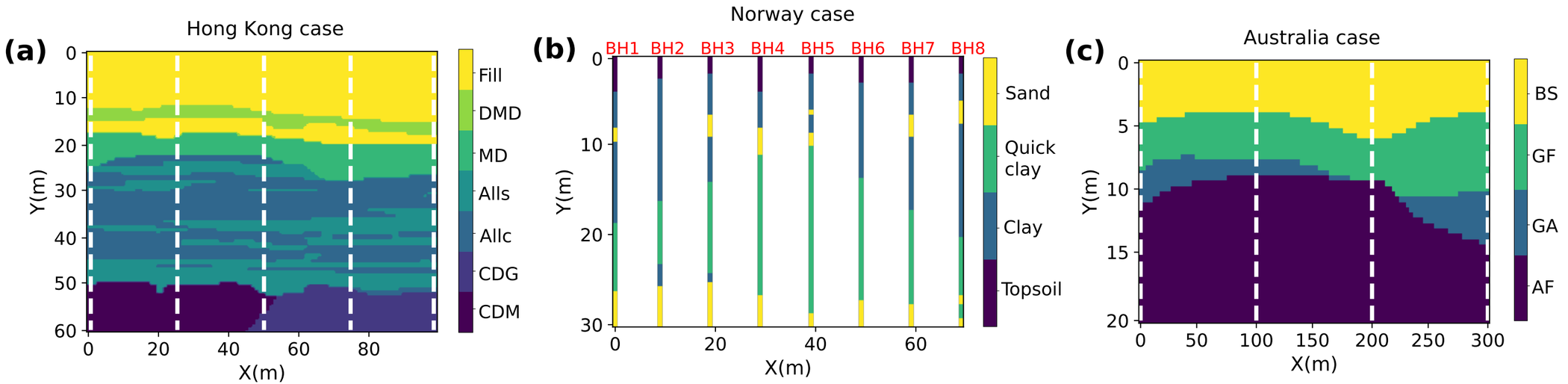

(a)
β_1
β_2
β_3
β_4
β
4
2
0
0
250
500
750
1000
1250
1500
1750
2000
Iterations
(b)
Total energy
20000
15000
10000
5000
0
250
500
750
1000
1250
1500
1750
2000
Iterations

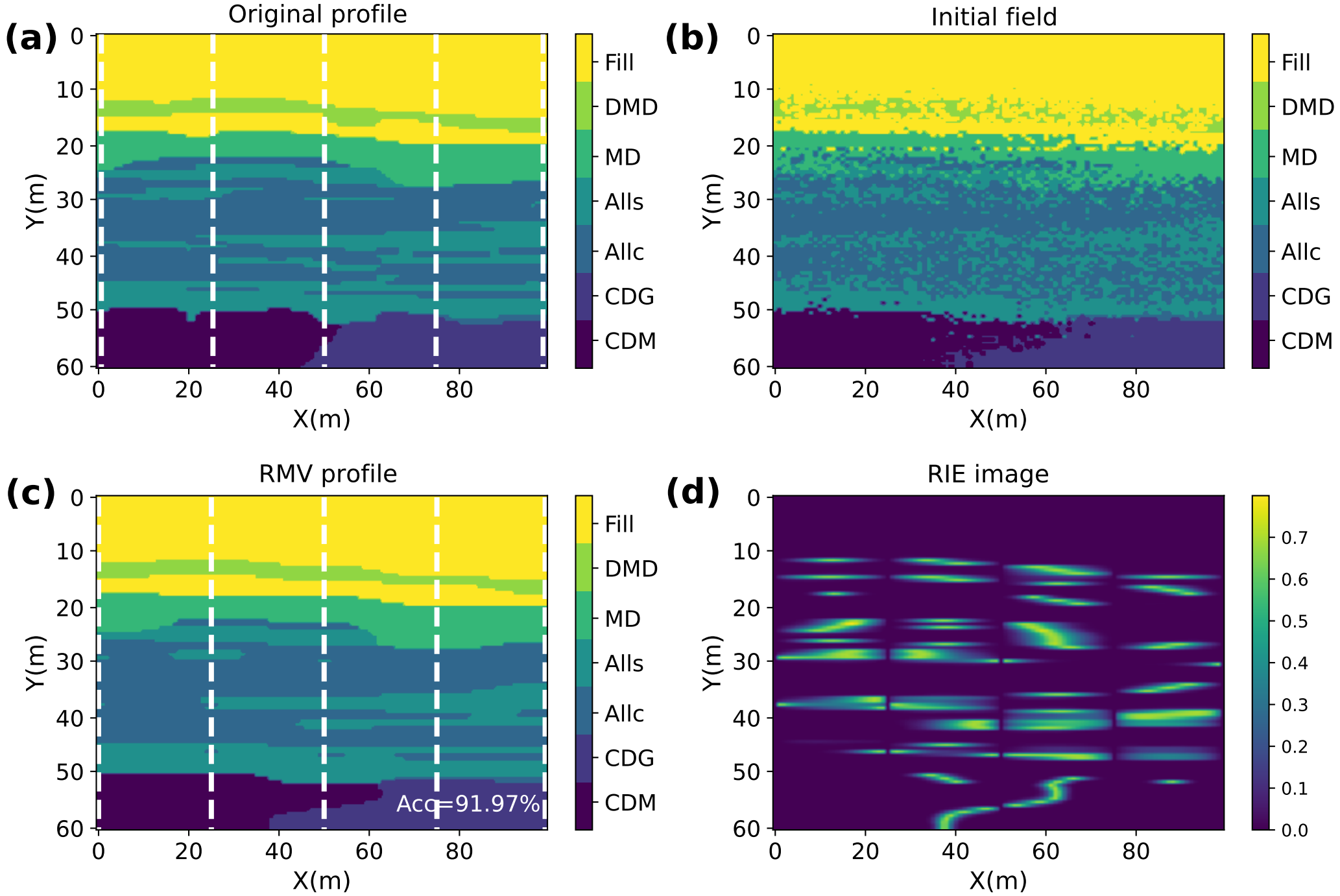

(a)
Original profile
Y pixel ID
X pixel ID
Label C
Label B
Label A
a
b
c
d
e
f
g
(b)
Initial field
(c)
Probability-triangle
Probability of Label C (%)
Probability of Label B (%)
Probability of Label A (%)
Label B
Label C
Label A
Pixel a
Pixel b
Pixel c
Pixel d
Pixel e
Pixel f
Pixel g
(d)
RMV profile
Acc=93.28%
(e)
RIE image
(f)
Total RIE
Number of simulations

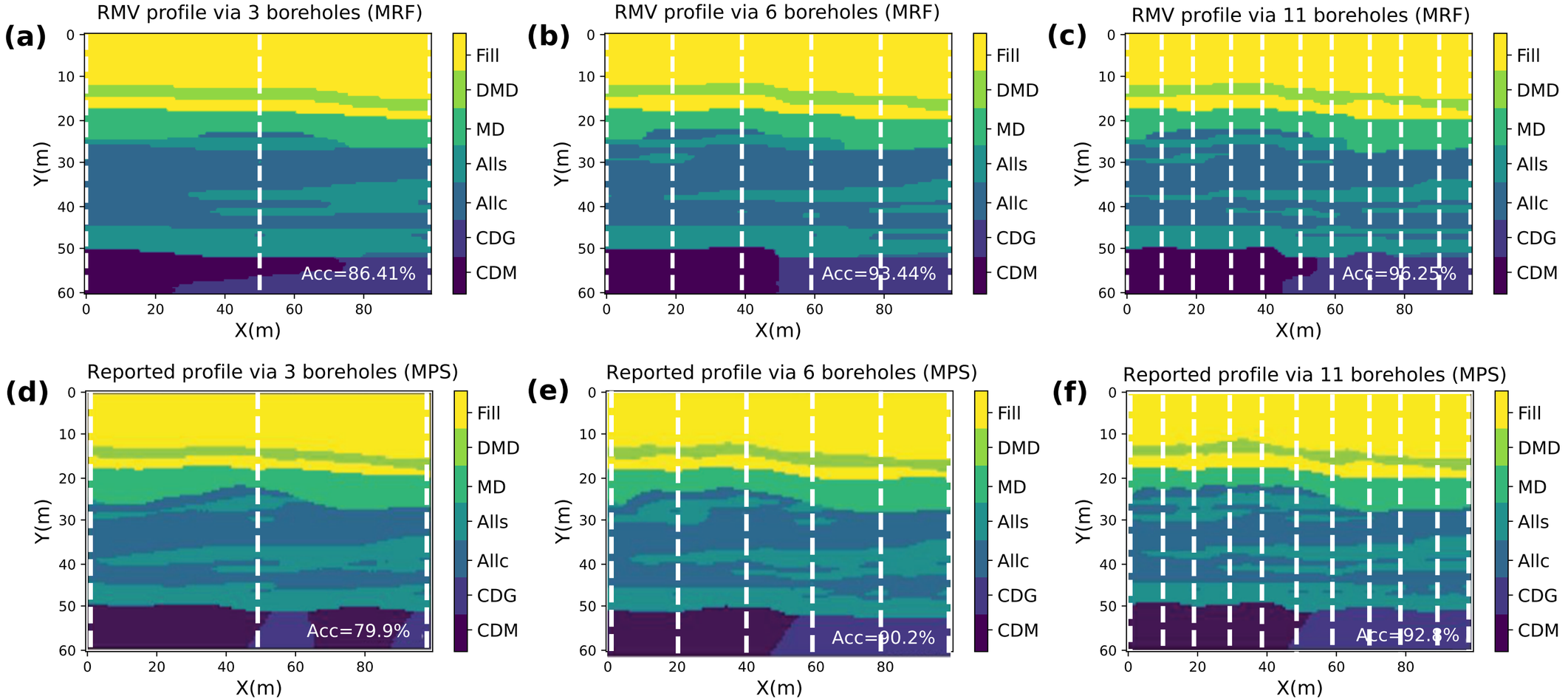

(a)
Original profile
Y pixel ID
X pixel ID
Label C
Label B
Label A
a
b
c
d
e
(b)
Initial field
(c)
Probability-triangle
Pixel a
Pixel b
Pixel c
Pixel d
Pixel e
Label B
Label C
Label A
Probability of Label C (%)
Probability of Label B (%)
Probability of Label A (%)
(d)
RMV profile
Acc=93.96%
(e)
RIE image
(f)
Total RIE
Number of simulations

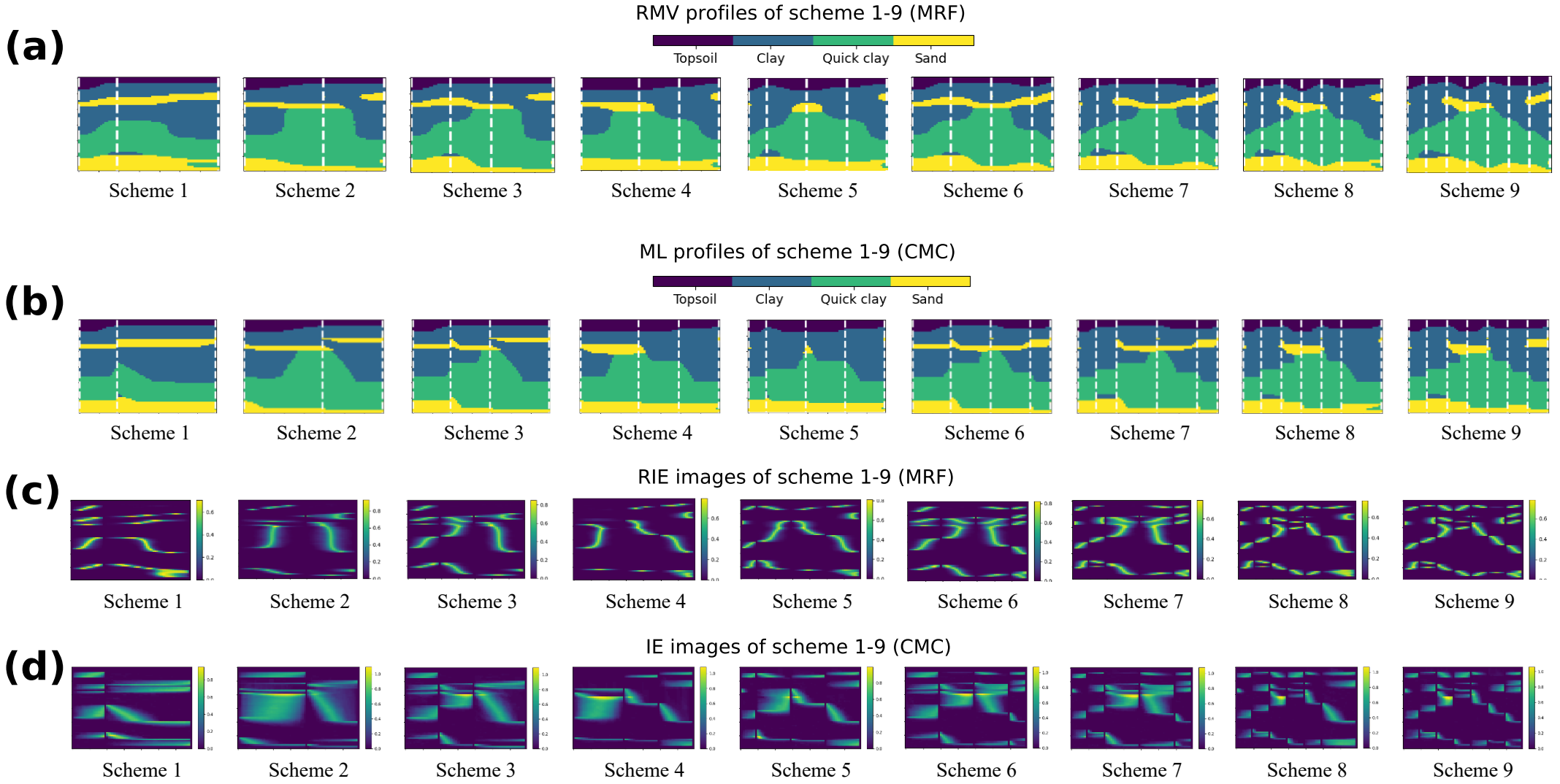

(a)
RMV profile (MRF)
V1
V2
V3
Label C
Label B
Label A
Acc=93.28%
Y pixel ID
X pixel ID
(b)
ML profile (CMC)
Acc=89.13%
(c)
RIE image (MRF)
(d)
IE image (CMC)
(e)
Truth of V1
RIE of V1 via MRF
IE of V1 via CMC
Truth of V2
RIE of V2 via MRF
IE of V2 via CMC
Truth of V3
RIE of V3 via MRF
IE of V3 via CMC
Pixel ID along depth
Entropy


(a)
RMV profile (MRF)
V1
V2
V3
V4
Acc=93.96%
Y pixel ID
X pixel ID
Label C
Label B
Label A
(b)
ML profile (CMC)
Acc=88.24%
(c)
RIE image (MRF)
(d)
IE image (CMC)
(e)
Truth of V1
RIE of V1 via MRF
IE of V1 via CMC
Truth of V2
RIE of V2 via MRF
IE of V2 via CMC
Truth of V3
RIE of V3 via MRF
IE of V3 via CMC
Truth of V4
RIE of V4 via MRF
IE of V4 via CMC
Pixel ID along depth
Entropy


(a)
Hong Kong case
Y(m)
X(m)
Fill
DMD
MD
Alls
Allc
CDG
CDM
(b)
Norway case
BH1 BH2 BH3 BH4 BH5 BH6 BH7 BH8
Y(m)
X(m)
Sand
Quick clay
Clay
Topsoil
(c)
Australia case
Y(m)
X(m)
BS
GF
GA
AF


(a)
Original profile
(b)
Initial field
(c)
RMV profile
Acc=91.97%
(d)
RIE image
Y(m)
X(m)
Fill
DMD
MD
Alls
Allc
CDG
CDM
0
10
20
30
40
50
60
0
20
40
60
80
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7


(a)
RMV profile via 3 boreholes (MRF)
Acc=86.41%
(b)
RMV profile via 6 boreholes (MRF)
Acc=93.44%
(c)
RMV profile via 11 boreholes (MRF)
Acc=96.25%
(d)
Reported profile via 3 boreholes (MPS)
Acc=79.9%
(e)
Reported profile via 6 boreholes (MPS)
Acc=90.2%
(f)
Reported profile via 11 boreholes (MPS)
Acc=92.8%
Y(m)
X(m)
Fill
DMD
MD
Alls
Allc
CDG
CDM


(a)
RMV profiles of scheme 1-9 (MRF)
Topsoil
Clay
Quick clay
Sand
Scheme 1
Scheme 2
Scheme 3
Scheme 4
Scheme 5
Scheme 6
Scheme 7
Scheme 8
Scheme 9
(b)
ML profiles of scheme 1-9 (CMC)
Topsoil
Clay
Quick clay
Sand
Scheme 1
Scheme 2
Scheme 3
Scheme 4
Scheme 5
Scheme 6
Scheme 7
Scheme 8
Scheme 9
(c)
RIE images of scheme 1-9 (MRF)
Scheme 1
Scheme 2
Scheme 3
Scheme 4
Scheme 5
Scheme 6
Scheme 7
Scheme 8
Scheme 9
(d)
IE images of scheme 1-9 (CMC)
Scheme 1
Scheme 2
Scheme 3
Scheme 4
Scheme 5
Scheme 6
Scheme 7
Scheme 8
Scheme 9