HOKKAIDO UNIVERSITY

Title	Integrated Stratigraphic and Soil Property Modeling Using Potts Model and Gaussian Process Regression
Author(s)	Kawata, Kosei
Relation	Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026), June 28 - July 1, 2026, Sapporo, Japan
Citation	Proceedings of the 10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026)
Issue Date	2026-07-01
Doc URL	https://hdl.handle.net/2115/100476
Type	conference paper
File Information	C000261_1.pdf



Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

10th International Symposium on Reliability Engineering and Risk Management (ISRERM 2026),
28 June-01 July 2026, Sapporo, Japan
Kitahara, T., Miyamori, Y. & Beer, M. (editors)

Integrated Stratigraphic and Soil Property Modeling Using Potts Model and Gaussian Process Regression

Kosei Kawata[1, *]

[1]*Technical Research Institute, Toyo construction co., ltd, 1-25-1, Naruohama, Nishinomiya, Hyogo, Japan.*
E-mail: kawata-kosei@toyo-const.co.jp
**Corresponding Author*

This study examines the applicability of a Potts-model-based layered Gaussian process regression (GPR) framework for three-dimensional subsurface characterization. Ground properties are notoriously challenging to map due to their sparse, uncertain, and spatially variable nature (the MUSIC-3X paradigm). The proposed framework first utilizes an anisotropic Potts-model segmentation to assign discrete stratigraphic labels to grid cells, then estimates continuous geotechnical properties within each assigned label using GPR. Its performance is benchmarked against three reference approaches: a naive layer-mean model, a fifth-order sparse polynomial model, and an unlayered GPR model. Two datasets are considered: the S-VG5 benchmark dataset based on synthetic cone penetration test data with complex heterogeneity, and the Naruohama boring dataset, evaluated via leave-one-source-out cross-validation utilizing SPT N-values. In the data-rich S-VG5 benchmark problem, the layered GPR achieved the best prediction accuracy (RMSE = 1.340), substantially outperforming the unlayered GPR (2.338) and the polynomial regression (2.787). However, in the highly sparse Naruohama field dataset, the naive model yielded the smallest mean cross-validation error. This phenomenon highlights a fundamental bias-variance tradeoff under extreme data starvation: while layered GPR strictly honors geological boundaries, severe intra-layer data sparsity drastically inflates predictive variance, whereas simple mean-based models remain robust. These findings indicate that Potts-model-based layered GPR represents a theoretical optimum when sufficient observations are available within stratigraphic units, yet simpler, high-bias estimators remain highly competitive under the severe sparsity typical of practical engineering conditions.

Keywords: Potts model, Gaussian Process Regression, stratigraphic modeling, spatial variability, ground investigation

1 Introduction

Three-dimensional subsurface characterization is required for rational geotechnical design, but field data are usually sparse, spatially heterogeneous, and affected by stratigraphic discontinuity. In data-driven site characterization, it is therefore important to compare interpolation strategies under both controlled benchmark conditions and sparse field conditions (Phoon et al. 2022). Potts-model-based approaches have been used to represent discrete spatial structure even from extremely limited borehole data (Shuku and Phoon 2023), whereas Gaussian process regression (GPR) provides a flexible probabilistic model for continuous property interpolation with predictive uncertainty.
The present study combines these ideas. A Potts-model segmentation is used to define discrete geological labels, and continuous property estimation is performed by GPR within each label. The objective is to test whether this layered GPR framework improves prediction accuracy relative to simpler baselines and to clarify the data conditions under which such an advantage is observed. To this end, the method is evaluated using both a robust synthetic benchmark dataset (S-VG5) (Shuku and Phoon 2023) and a real-field dataset with only six boreholes.

2 Methodology

2.1 *Problem setting*

The target workflow consists of two steps. First, the ground domain is discretized into three-dimensional cells, and each cell is assigned a Potts-model label representing a stratigraphic class. Second, a continuous target property is estimated at each cell. In this study, cone tip resistance q_t is predicted for the benchmark problem and SPT N-value is predicted for the field case.
The Potts model based segmentation (Shuku and Phoon 2023) assigns a stratigraphic label to each cell by evaluating an anisotropic energy function $E(x)$. For a three-dimensional domain, the total energy is defined as:

$$E(\mathbf{x}) = -\sum_{i,j,k} [\theta_h d(x_{i,j,k}, x_{i+1,j,k}) + \theta_h d(x_{i,j,k}, x_{i,j+1,k}) + \theta_v d(x_{i,j,k}, x_{i,j,k+1})] \quad (1)$$

where $x_{i,j,k}$ is the categorical label at grid cell (i,j,k) , and the indicator function $d(a,b) = 1$ if $a = b$ and 0 otherwise (Shuku and Phoon 2023). The hyperparameters θ_h and θ_v explicitly control the spatial smoothness in the horizontal and vertical directions, respectively, effectively capturing the inherent anisotropy of soil strata. To handle data sparsity, the available borehole observations y are incorporated as strict hard constraints ($Ax = y$, where A is a design matrix) rather than a soft data-fitting term (Shuku and Phoon 2023).

2.2 *Compared methods*

Four methods are compared.

- Naive model: the target value is represented by a label-wise mean estimated from the available observations.
- Sparse polynomial model: a fifth-order polynomial basis is fitted with L1-based sparsification.
- Unlayered GPR: all observations are modeled by a single GPR without stratigraphic separation.
- Layered GPR: observations are divided by Potts-model label, and an independent GPR is fitted within each label.

The polynomial model relies on Sparse Polynomial Chaos Expansions (SPCE) (Blatman and Sudret 2010), which approximates the response as

$$Y \approx \sum_{\boldsymbol{\alpha} \in \mathcal{A}} y_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \tag{2}$$

where $\psi_\alpha(x)$ denotes a polynomial basis and y_α are coefficients. Sparse coefficient estimation is obtained by solving

$$\hat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \left\{ \frac{1}{N} \sum_{i=1}^{N} \left(Y^{(i)} - \sum_{\boldsymbol{\alpha} \in \mathcal{A}} y_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}^{(i)}) \right)^2 + \lambda \|\boldsymbol{y}\|_1 \right\} \tag{3}$$

For GPR, the joint prior distribution of the observed values y at training inputs X and the latent values f^* at new inputs X^* is

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}^* \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} K(\boldsymbol{X},\boldsymbol{X}) + \sigma_n^2 \boldsymbol{I} & K(\boldsymbol{X},\boldsymbol{X}^*) \\ K(\boldsymbol{X}^*,\boldsymbol{X}) & K(\boldsymbol{X}^*,\boldsymbol{X}^*) \end{bmatrix}\right) \tag{4}$$

where $K(\cdot,\cdot)$ is the covariance matrix and σ_n^2 is the observation-noise term. In the unlayered model, all observations share a single GPR. In the layered model, separate GPRs are fitted within each Potts label.
For the Naruohama case, a practical fallback was additionally examined for visualization and final-field generation: cells that could not be predicted by the pure layered model because of zero or insufficient observations within a label were filled by the unlayered GPR prediction. This fallback was not used to reinterpret the pure layered cross-validation behavior; instead, it was used to visualize a gap free final field. Workflow of the proposed subsurface Fig 1 is characterization framework.

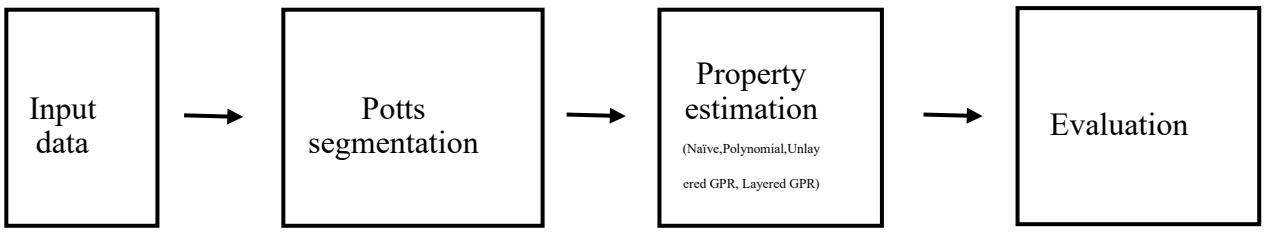


Figure 1. Overview of the proposed workflow.

2.3 *Benchmark dataset and evaluation*

The S-VG5 benchmark dataset provides an objective ground truth for evaluating DDSC algorithms (Shuku and Phoon 2023). It contains seven training soundings and an independent validation set for prediction. The layered GPR used the tuned anisotropy setting. Prediction accuracy was evaluated by root-mean-square error (RMSE) on and mean absolute error (MAE) 1,200 validation points.

2.4 *Field dataset and evaluation*

The Naruohama dataset consists of six boreholes with SPT N-values and an existing three-dimensional VTK grid containing Potts model labels. Naruohama is a reclaimed land area in Japan that suffered severe liquefaction and ground settlement during the 1995 Hyogo-ken Nambu earthquake, resulting in highly complex SPT N-value distributions (Tokimatsu et al. 1996). Because no independent validation set is available, leave-one-source-out cross-validation was performed. In each fold, five boreholes were used for training and the remaining one was predicted. Mean RMSE and mean MAE over the six folds were used for comparison. In addition, final-fit profiles were produced by training each model on all six boreholes and predicting values back at the observed coordinates.

2.5 *Uncertainty interpretation*

The GPR models provide predictive mean and predictive standard deviation. In this study, the predictive standard deviation is interpreted as model-based interpolation uncertainty. It does not yet fully separate inherent spatial variability, measurement error, and transformation uncertainty in the sense discussed by Phoon and Kulhawy (1999) 18. More precisely, the present formulation uses a spatial covariance term and a noise term within the regression model, but it does not explicitly decompose geotechnical uncertainty into physically identifiable components. Therefore, the current uncertainty evaluation should be interpreted as conditional predictive uncertainty within the adopted regression framework.

3 Results and discussion

3.1 *Benchmark results*

Table 1 summarizes the benchmark validation results. The layered GPR outperformed the three reference methods. Its RMSE was 1.340, whereas the RMSE values of unlayered GPR, naive, and sparse polynomial regression were 2.338, 2.485, and 2.787, respectively. The corresponding MAE values were 0.966, 1.647, 1.854, and 2.167. These results indicate that, when an adequate number of observations are available within each stratigraphic class, separating the domain by Potts-model labels can improve continuous property prediction substantially. Figure 2 is Benchmark comparison (S-VG5).

Table 1. The caption must be centered. Benchmark S-VG5 validation metrics.

Method	RMSE	MAE
Naive	2.485	1.854
Polynomial (5th-order sparse)	2.787	2.167
Unlayered GPR	2.338	1.647
Layered GPR	1.340	0.966

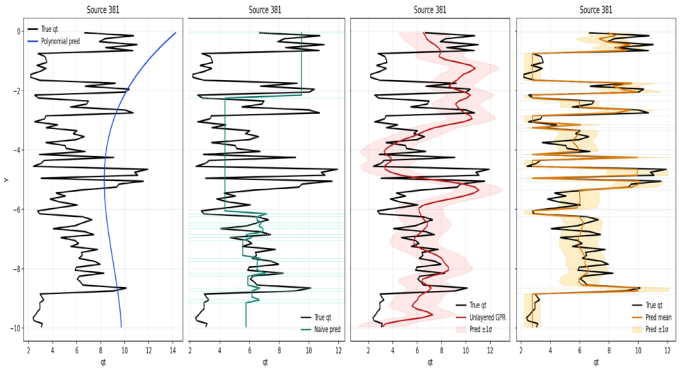


Figure 2. Benchmark comparison (S-VG5)

3.2 *Naruohama results*

Table 2 summarizes the leave-one-source-out results for the Naruohama case. In contrast to the benchmark problem, the naive model gave the smallest mean RMSE (6.557) and mean MAE (3.934). Unlayered GPR produced a nearly identical mean RMSE (6.565) but a slightly larger mean MAE (4.119). The pure layered GPR showed mean RMSE = 6.799 and mean MAE = 4.103, while the sparse polynomial model performed worst (mean RMSE = 7.916, mean MAE = 6.112). Figure 3 is Naruohama field comparison (LOSO). The field case is evaluated by leave-one-source-out cross-validation. Each fold uses five boreholes for training and one borehole for testing.

This contrast with the benchmark results is important. The field dataset is much smaller and more unevenly distributed across labels. Several labels had zero or insufficient observations even when all six boreholes were used, which means that the pure layered GPR could not predict some cells without fallback. Therefore, the potential benefit of stratigraphic separation was offset by severe data sparsity. Under such conditions, a simple mean-based model can remain competitive.

Table 2. Naruohama leave-one-source-out cross-validation metrics.

Method	Mean RMSE	Mean MAE	Std. RMSE	Std. MAE
Naive	6.557	3.934	3.123	1.255
Unlayered GPR	6.565	4.119	3.917	1.927
Layered GPR (pure)	6.799	4.103	3.582	1.690
Polynomial	7.916	6.112	3.549	3.229

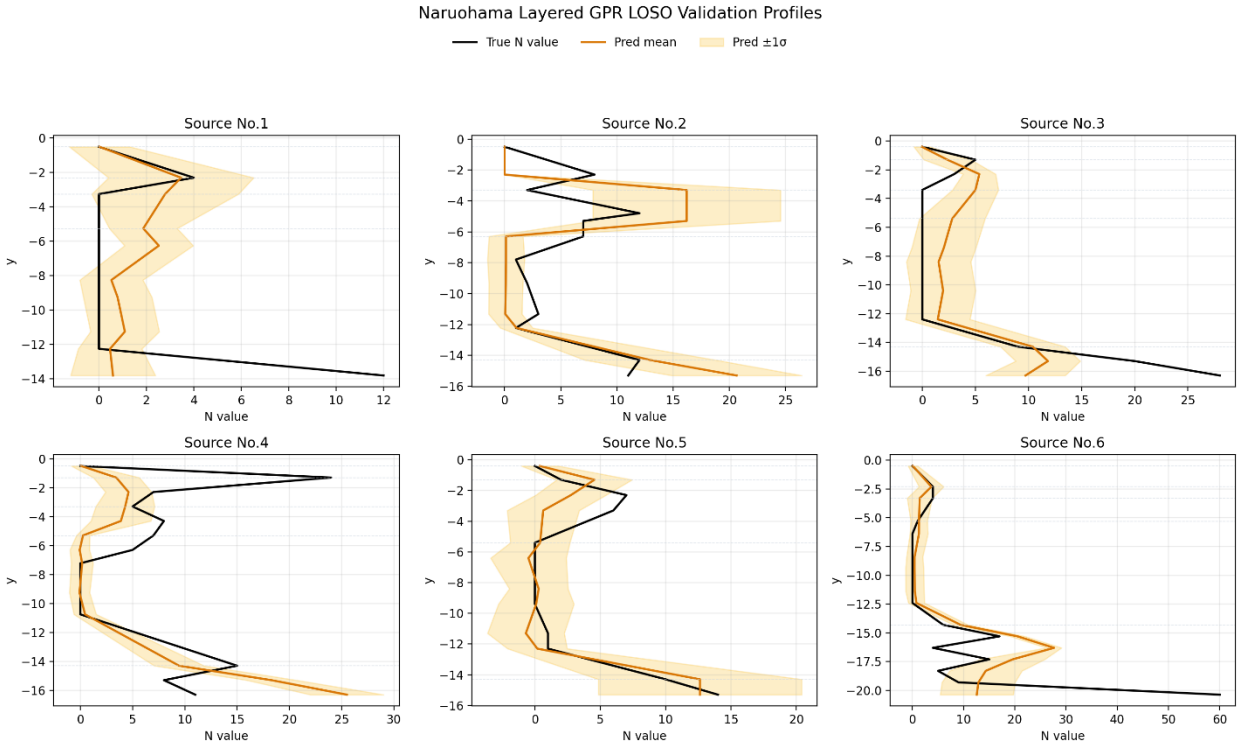


Figure 3. Naruohama field comparison (LOSO)

3.3 *Implications for practical use*

The present results suggest that layered GPR should not be assumed to dominate simpler approaches in all practical cases. Its advantage depends on whether each stratigraphic class is supported by a sufficient number of representative observations. If label-wise observations are sparse, the method becomes sensitive to missing labels, unstable local fits, and discontinuous final fields. In contrast, the naive model is robust because it uses only low-order summary information.

The study nevertheless supports the usefulness of Potts-model-based layering. In the benchmark problem, the layered model provided the best predictive performance. In the field problem, the Potts labels remain valuable as a structural representation of the ground, even though the continuous regression step may require denser observations, coarser label grouping, or a hybrid strategy such as fallback to an unlayered model.

4 Conclusions

This study compared four methods for three-dimensional geotechnical property prediction: a naive layer-mean model, a sparse polynomial model, an unlayered GPR model, and a Potts-model-based layered GPR model. The following conclusions were obtained.

1.In the S-VG5 benchmark problem, layered GPR achieved the best validation accuracy and outperformed the three reference methods.

2.In the Naruohama field dataset, the naive model provided the smallest mean cross-validation error, while layered and unlayered GPR showed similar but slightly inferior performance because of extreme data sparsity.

3.The predictive standard deviation produced by the current GPR framework is useful as a model-based interpolation uncertainty measure, but it does not yet represent a full decomposition of geotechnical variability.

4.Potts-model-based layered GPR is promising when observations are sufficiently representative within each layer, but simpler models remain important baselines in sparse practical settings.

Future work should focus on improving the relation between Potts-model labels and observation density, introducing coarser or hierarchical label definitions, and extending the framework to practical decision-making problems such as optimal additional boring placement and pile embedment design support.

Acknowledgments

During a part of this study, ChatGPT, provided by OpenAI, was utilized to obtain useful advice for code improvement during Python-based data processing and analysis. It was also used for grammar checking of the manuscript. The author would like to express their gratitude for this assistance.

References

Blatman, G. and Sudret, B., An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, Probab. Eng. Mech., 25(2), 183-197, 2010.

Phoon, K. K. and Kulhawy, F. H., Characterization of geotechnical variability, Can. Geotech. J., 36(4), 612-624, 1999.

Phoon, K. K., Ching, J., and Shuku, T., Benchmark Examples for Data-Driven Site Characterization, Georisk, 16(4), 599-621, 2022.

Shuku, T. and Phoon, K. K., Data-driven subsurface modelling using a Markov random field model, Georisk, 17(1), 41-63, 2023a.

Shuku, T. and Phoon, K. K., Comparison of Data-Driven Site Characterization Methods through Benchmarking: Methodological and Application Aspects, ASCE-ASME J. Risk Uncertainty Eng. Syst. Part A Civ. Eng., 9(2), 04023006, 2023b.

Tokimatsu, K., Mizuno, H., and Kakurai, M., Geotechnical site characterization and liquefaction of soil during the 1995 Hyogo-ken Nambu earthquake, Soils Found., 36(1), 217-234, 1996.