World Journal of Engineering and Technology, 2025, 13(1), 1-11
https://www.scirp.org/journal/wjet
ISSN Online: 2331-4249
ISSN Print: 2331-4222

Probabilistic Site Investigation Optimization of Gassy Soils Based on Conditional Random Field and Monte Carlo Simulation

Shaolin Ding[1,2]

[1]Key Laboratory of Geotechnical Mechanics and Engineering of the Ministry of Water 7 Resources, Changjiang River Scientific Research Institute, Wuhan, China
[2]National Dam Safety Research Center, Wuhan, China
Email: shaolin_ding@whu.edu.cn

How to cite this paper: Ding, S. L. (2025) Probabilistic Site Investigation Optimization of Gassy Soils Based on Conditional Random Field and Monte Carlo Simulation. *World Journal of Engineering and Technology*, **13**, 1-11.
https://doi.org/10.4236/wjet.2025.131001

Received: September 24, 2024
Accepted: November 25, 2024
Published: November 28, 2024

Abstract

Gassy soils are distributed in relatively shallow layers the Quaternary deposit in Hangzhou Bay area. The shallow gassy soils significantly affect the construction of underground projects. Proper characterization of spatial distribution of shallow gassy soils is indispensable prior to construction of underground projects in the area. Due to the costly conditions required in the site investigation for gassy soils, only a limited number of gas pressure data can be obtained in engineering practice, which leads to the uncertainty in characterizing spatial distribution of gassy soils. Determining the number of boreholes for investigating gassy soils and their corresponding locations is pivotal to reducing construction risk induced by gassy soils. However, this primarily relies on the engineering experience in the current site investigation practice. This study develops a probabilistic site investigation optimization method for planning investigation schemes (including the number and locations of boreholes) of gassy soils based on the conditional random field and Monte Carlo simulation. The proposed method aims to provide an optimal investigation scheme before the site investigation based on prior knowledge. Finally, the proposed approach is illustrated using a case study.

Keywords

Gassy Soils, Site Investigation, Uncertainty, Conditional Random Field, Monte Carlo Simulation

1. Introduction

Gassy soils have been widely discovered in shallow soil layers in eastern costal area of China, such as Hangzhou Bay area, Zhejiang province. It was generated by the anaerobic decomposition process of organic materials [1]. Since main compositions of biogenic gas is flammable gas (e.g., methane CH_4), the presence of shallow gassy soils may cause significant risk for infrastructure construction, such as fire outbreaks and blasting during underground construction [2]. Therefore, proper characterization of spatial distribution of gassy soils is indispensable prior to the construction of underground projects. Due to the significant cost and human commitments required in the site investigation for gassy soils, only a limited number of gas pressure data can be obtained in engineering practice, which leads to the uncertainty in characterizing spatial distribution of gassy soils. Determining the site investigation scheme (including the number and locations of boreholes) is pivotal to reducing construction risk induced by gassy soils.

Since only a limited number of borehole data can be obtained during geotechnical site investigation, it is prudent to carefully determine the number and locations of boreholes so as to maximize the value of data. However, this is, in general, fundamentally challenging task due to the spatial variability of geotechnical materials and a lack of handy methods and tools for this purpose. Some studies have been performed to investigate the gassy soils in the Hangzhou Bay area. However, most of them focused on the formation and composition of biogenic gas [2], features and distributions of gas pools [3], and the exploration methods ([4] [5]). Research is rare that focused on how to, efficiently and properly, characterize the spatial distribution of gassy soils with a limited number of borehole data. Determination of the optimal scheme for investigating gassy soils prior to site investigation remains an open question.

This study presents a probabilistic site investigation optimization method to determine the optimal scheme of investigating gassy soils. The proposed method makes uses of the prior knowledge in a quantitative and transparent manner, and explicitly models the spatial variability of gassy soils by the conditional random field. The paper starts with introduction of the proposed method, followed by illustration and verification by a case study in Hangzhou Bay area.

2. Conditional Random Field Modelling of Horizontal Spatial Variability of Gas Pressure

Site investigation of gassy soils are usually carried out to measure and estimate the gas pressure of gassy soils at different locations with a limited number of boreholes. In this research, the number and locations of measuring boreholes are of interest. The main problem solved in this study is to obtain the optimal investigation scheme. Let L and ΔL denote the length of site investigation field concerned and the interval between adjacent boreholes. The investigation schemes (including the number and locations of boreholes) can be represented by a set, $S = \{ i\Delta L, i = 1, 2, 3, \cdots, N_B \}$, of borehole locations in the horizontal direction, where $N_B = \text{INT}[L/\Delta L]$ denotes the number of locations and INT[·] is a round function that returns the integer part of $L/\Delta L$. Note that this study only accounts for one-dimensional spatial variability of gas pressure in the horizontal direction, and the vertical spatial variability is ignored, which can be pursued in future study.

Since no data is available prior to site investigation, the gas pressure of the N_B locations can be arbitrary values. For planning investigation schemes, simulated gas pressure data at some investigation locations are generated using the prior knowledge, which will be introduced in the next subsection. Let $Z_{br} = \{PR_i, i = 1, 2, 3, \cdots, N_B\}$ be a set of simulated data for a given investigation scheme S, and it consists of gas pressure data PR_i at the N_B borehole locations. Using Z_{br}, a Kriging-based conditional random field model is applied to modelling the horizontal spatial variability of gas pressure, by which the gas pressure in horizontal direction can be written as [6] [7]:

$$Z_c(x) = Z_k(x) + \left(Z_{uc}(x) - Z_{sk}(x)\right) \tag{1}$$

where x is the horizontal coordinate; $Z_c(x)$ is the conditional random field; $Z_k(x)$ is the Kriging interpolation of gas pressure over the domain of interest based on the Z_{br} [8]; $Z_{uc}(x)$ is the unconditional random field of gas pressure; $Z_{sk}(x)$ is the Kriging interpolation of gas pressure over the domain of interest based on the values of gas pressure at borehole locations simulated via the unconditional random fields. Equation (1) ensures that the realizations of random fields exactly match the Z_{br}, for which the ordinary Kriging method is employed herein to perform the interpolation over the domain of interest.

3. Prior Knowledge for Simulating Data Given Candidate Design Schemes

According to Section 2.1, the gas pressures (*i.e.*, Z_{br}) at the borehole locations are needed to simulate the spatial variability using Equation (1). Unfortunately, there is no measuring data prior to site investigation. For planning the investigation scheme, the simulated data can be generated using the prior knowledge (e.g., engineering experience and judgments) on the gas pressure, such as typical ranges of statistics of gas pressures. Consider, for example, that the mean values μ, standard deviations σ and scale of fluctuation λ vary within their respective typical ranges $[\mu_{min}, \mu_{max}]$, $[\sigma_{min}, \sigma_{max}]$, and $[\lambda_{min}, \lambda_{max}]$. Then, the μ, σ and λ can be treated as uniform random variables defined by their respective typical ranges and their random samples can be generated. Let $\mu_{s,i}$, $\sigma_{s,i}$, and $\lambda_{s,i}\,(i = 1, 2, 3, \cdots, N_e)$ denote a number, N_e, sets of random samples of μ, σ and λ simulated from the prior knowledge. For each set of $\mu_{s,i}$, $\sigma_{s,i}$ and $\lambda_{s,i}$, a set of the simulated data $Z_{s,i}\,(i = 1, 2, 3, \cdots, N_e)$ is simulated using Karhunen-Loeve (KL) expansion ([9] [10]) in this study, which is written as:

$$Z_{s,i}(x) = \mu_{s,i} + \sum_{j=1}^{\infty} \sigma_{s,i} \sqrt{v_j} f_j(x) \zeta(\theta) \tag{2}$$

where $Z_{s,i}$ is the gas pressure simulated using the sample $\mu_{s,i}$, $\sigma_{s,i}$ and $\lambda_{s,i}$; $\zeta(\theta)$ is independent standard normal random variable; v_j and $f_j(x)$ are the eigenvalues and eigenfunctions of the covariance function, which is taken as a squared exponential correlation function in this study:

$$\rho(\tau) = \exp\left[-\pi\left(\tau/\lambda_{s,i}\right)^2\right] \quad (3)$$

where τ is the separate distance between two locations in the horizontal direction; $\rho(\tau)$ is the autocorrelation coefficient between the gas pressures at the two locations. For the sake of conciseness, details of the random field simulation based on KL expansion are not provided here. Interested reader may refer to Huang *et al.* (2001) [9] and Phoon *et al.* (2002) [10] for more details.

Using KL expansion, the $Z_{s,i}\left(i = 1,2,3,\cdots,N_e\right)$ can be simulated using each set of samples of μ, σ and λ, *i.e.*, $\mu_{s,i}$, $\sigma_{s,i}$ and $\lambda_{s,i}$. Then, the gas pressures at borehole locations in an investigation scheme S in $Z_{s,i}$ are taken as a simulated data at these locations, denoted by $Z_{br,i}$.

4. Monte Carlo Simulation for Calculating Occurrence Probability of Gassy Soils

For a given set of $\mu_{s,i}$, $\sigma_{s,i}$ and $\lambda_{s,i}$ and its corresponding $Z_{br,i}$ simulated from KL expansion in the preceding subsection, Monte Carlo simulation is performed with Equation (1) to determine the probability distribution of the gas pressure at each location. For this purpose, the $Z_{br,i}$ is used in Equation (1) to obtain $Z_k(x)$ using Kriging interpolation and $\mu_{s,i}$, $\sigma_{s,i}$ and $\lambda_{s,i}$ are used to simulate $Z_{uc}(x)$ using KL expansion again. By this means, for each set of $Z_{br,i}$, a number, N_a, of realizations of conditional random field of the gas pressure can be generated using Equation (1). The N_a realizations, $Z_{c,i}$, of the conditional random field consist of the N_a values of the gas pressure at each location. Consider, for example, a threshold value, R, of the gas pressure. As the gas pressure of some location x is greater than R, it is considered risky; otherwise, it is ignorable. These two situations are represented by a pair of complementary events E_0 (Ignorable event) and E_r (Risky event), with respective probabilities of p_r and p_0, which are calculated as

$$p_0 = \frac{1}{N_a}\sum_{k=1}^{N_a} I\left[Z_{c,i}^k\left(x\right) < R\right] \quad (4)$$

$$p_r = 1 - p_0 \quad (5)$$

where $Z_{c,i}^k\left(x\right)$ is the gas pressure at position x in the k-th realization of the conditional random field based on $Z_{br,i}$; N_a is the total number of realizations of the conditional random field; $I[\cdot]$ is the indicative function, and it is equal to 1 if $Z_{c,i}^k\left(x\right) < R$ or 0 otherwise.

In order to quantify the uncertainty in the presence of gassy soils, Monte Carlo simulation is adopted to repeatedly simulate gas pressure from the prior knowledge. Based on each set of simulated data $Z_{br,i}\left(i = 1,2,\cdots,N_e\right)$, the probabilities of E_0 and E_r are calculated as $p_{0,i}$ and $p_{r,i}\left(i = 1,2,\cdots,N_e\right)$ using Equations (4)-(5). After that, the mean values (*i.e.*, $p_{0,e}$ and $p_{r,e}$) of p_0 and p_r corresponding to the N_e sets of simulated data $Z_{br,i}\left(i = 1,2,\cdots,N_e\right)$ are obtained:

$$p_{0,e} = \frac{1}{N_e}\sum_{i=1}^{N_e} p_{0,i} \quad (6)$$

$$p_{r,e} = 1 - p_{0,e} \tag{7}$$

where N_e is the total number of simulated data $Z_{br,i}\left(i = 1, 2, \cdots, N_e\right)$. Then, the larger value between $p_{0,a}$ and $p_{r,a}$ is used to determine whether gassy soils present at some location or not, *i.e.*,

$$p_a = \max\left\{p_{0,e}, p_{r,e}\right\} \tag{8}$$

If $p_a = p_{0,e}$ and is greater than a certain value (e.g., 0.9) selected for decision-making, no gassy soils present at the location; otherwise, if $p_a = p_{r,e}$ and is greater than the selected threshold of probability value, gassy soils occur at the location. The investigation scheme shall guarantee that the p_a values in the whole domain of interest are greater than the prescribed threshold probability value.

5. Illustrative Example

Gassy soils were found to be distributed in a crossing section of Hangzhou Metro Line 1. The proposed approach is applied to determining the optimal scheme for investigating gassy soils in the cross section of the Hangzhou Metro Line 1. Although the gassy soils may be distributed within depths ranging from 25m to 32m, only the horizontal spatial variability of gas pressure is taken into account in this study. The horizontal length of cross section concerned in this study is 3000m along the Hangzhou Metro Line 1. The influence of gassy soils on the construction of underground projects depends on the gas pressure. Generally speaking, gassy soils are considered to be dangerous if the gas pressure is greater than or equal to 50 kPa; otherwise, the risk induced by gassy soils is ignorable. Therefore, the R is set as 50 kPa in this study. In other words, E_0 (Ignorable event) and E_r (Risky event) are defined as the event with gas pressure less than 50 kPa and its complementary event (see **Table 1**), respectively.

Table 1. Definition of ignorable and risky events.

Events	Gas pressure (kPa)	Notes
E_0	(0, 50)	Ignorable event
E_r	[50, +∞)	Risky event

5.1. Candidate Investigation Schemes

As mentioned in Section 2.3, the proposed approach uses the p_a value to determine whether gassy soils present at some location. Generally speaking, the larger the p_a value is, the less the uncertainty in the presence of gassy soils. Based on the verbal descriptors of probabilities (see **Table 2**) [11], the threshold value for judging the presence of gassy soils is taken as 0.9 (very likely) in this example. When the p_a is greater than 0.9, E_0 or E_r is very likely to occur depending on $p_a = p_{0,e}$ or $p_{0,r}$. The selected investigation scheme shall ensure that the p_a values of different borehole locations at the site of interest are all greater than 0.9 no matter $p_a = p_{0,e}$ or $p_{0,r}$.

For consideration of different investigation schemes, let ΔL range from 10 m to 60 m at an interval of 10 m, *i.e.*, 10, 20, ..., 50, 60 m, which correspond to 300, 150, …, 50 boreholes along the cross section, as shown in Table 3. The optimal scheme is then determined among them. For this purpose, the simulated data is generated using the prior knowledge. Consider, for example, that the prior knowledge of gas pressure statistics is taken as $\mu \in (0\ \text{kPa}, 250\ \text{kPa}]$, $\sigma \in (0\ \text{kPa}, 125\ \text{kPa}]$, and $\lambda \in (0\ \text{m}, 100\ \text{m}]$ in this study, and the random field of gas pressure in the horizontal direction is discretized as a grid with an interval of 10m. Using the prior knowledge, the proposed approach is applied to determining the optimal investigation scheme among the six candidate ones, as presented in the next subsection.

Table 2. Verbal descriptors and their probability equivalents.

Verbal Descriptor	Virtually impossible	Very unlikely	Equally likely	Very likely	Virtually certain
Probability equivalent	0.01	0.10	0.50	0.90	0.99

Table 3. Candidate investigation schemes considered in this example.

ID	S1	S2	S3	S4	S5	S6
Borehole Interval, ΔL (m)	60	50	40	30	20	10
Borehole Number, N_B	50	60	75	100	150	300

5.2. Occurrence Probability of Gassy Soils Given Different Investigation Schemes

The p_a values of different locations given an investigation scheme need to be calculated for determining the optimal scheme among the six scheme shown in Table 3. For example, given S3 with $\Delta L = 30$ m and 100 boreholes at the locations. Firstly, a number, N_e, of random field parameters μ, σ and λ were generated from the prior knowledge (*i.e.*, uniform distribution with typical ranges of μ, σ and λ). Based on each set of samples of μ, σ and λ, $Z_{s,i}$ is simulated using Equation (2)-(3). According to $Z_{s,i}$, the values of gas pressures at borehole locations in S3 are taken as a set of simulated data $Z_{b,i}$. With $Z_{b,i}$, a number, N_a, of realizations of the conditional random field of the gas pressure are generated using Equation (1) (Figure 1), which are then used to calculate the p_a at each location using Equations (6)-(8). Both N_a and N_e may affect the accuracy of p_a values calculated from the proposed approach. Figure 2(a) demonstrates the effect of N_a on the averaged gas pressure at different locations simulated from the conditional random field given S3. It is shown that the estimated mean values of gas pressures at different locations converge as N_a is greater than 500. Similarly, Figure 2(b) shows the effect of N_e on the p_a calculated from the proposed approach for S3. It is found that N_e = 500 ensures the convergence of p_a values estimated from the proposed approach for S3.

Using $N_e = 500$ and $N_a = 500$, the p_a values at different locations given different investigation schemes are calculated using the proposed approach. Figures 3(a)-(f) show the p_a values along the cross section concerned in this example given S1

- S6, respectively. It is shown that as ΔL reduces to 20 m (*i.e.*, S5), all the p_a values along the cross section are greater than 0.9, some of which correspond to $p_a = p_{0,e}$ (see green circles showing ignorable gas pressures) while the others correspond to $p_a = p_{D,e}$ (see red squares showing risky gas pressures). For those locations with $p_a = p_{0,e}$ and $p_a \geq 0.9$, it is very likely that the gas pressure is too small to be risky. On the other hand, for those locations with $p_a = p_{D,e}$ and $p_a \geq 0.9$, it is very likely that the gas pressure is risky. For S6, similar results are obtained from the proposed approach (see Figure 3(f)) with all p_a values greater than 0.9. However, S6 costs more investigation efforts since the borehole interval is smaller and the number of boreholes is greater than those of S5. Hence, S5 is taken as the optimal investigation scheme among the six candidate schemes.

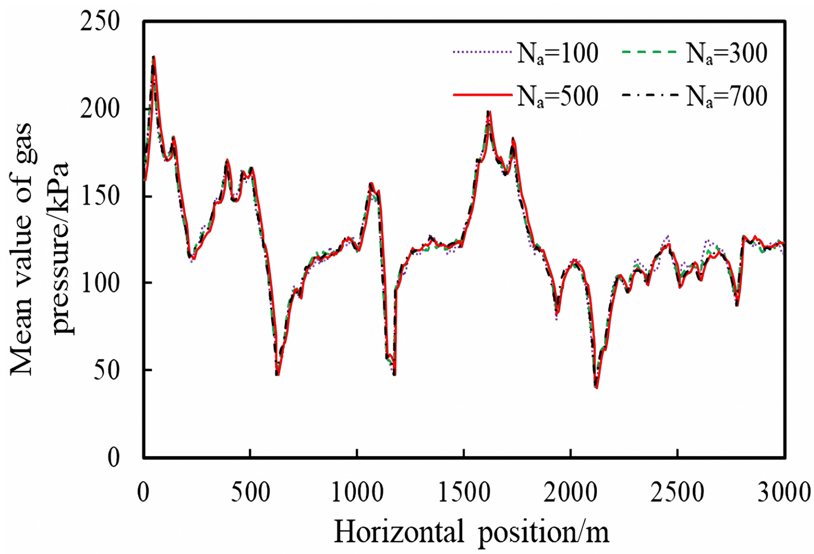

Figure 1. Effect of N_a on the mean gas pressure at different locations.

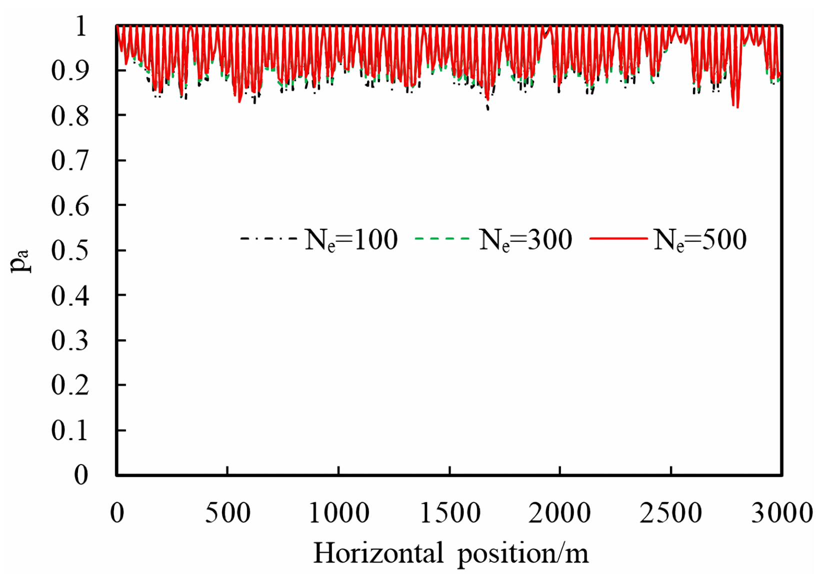

Figure 2. Effect of N_e on the p_a values calculated from the proposed approach.

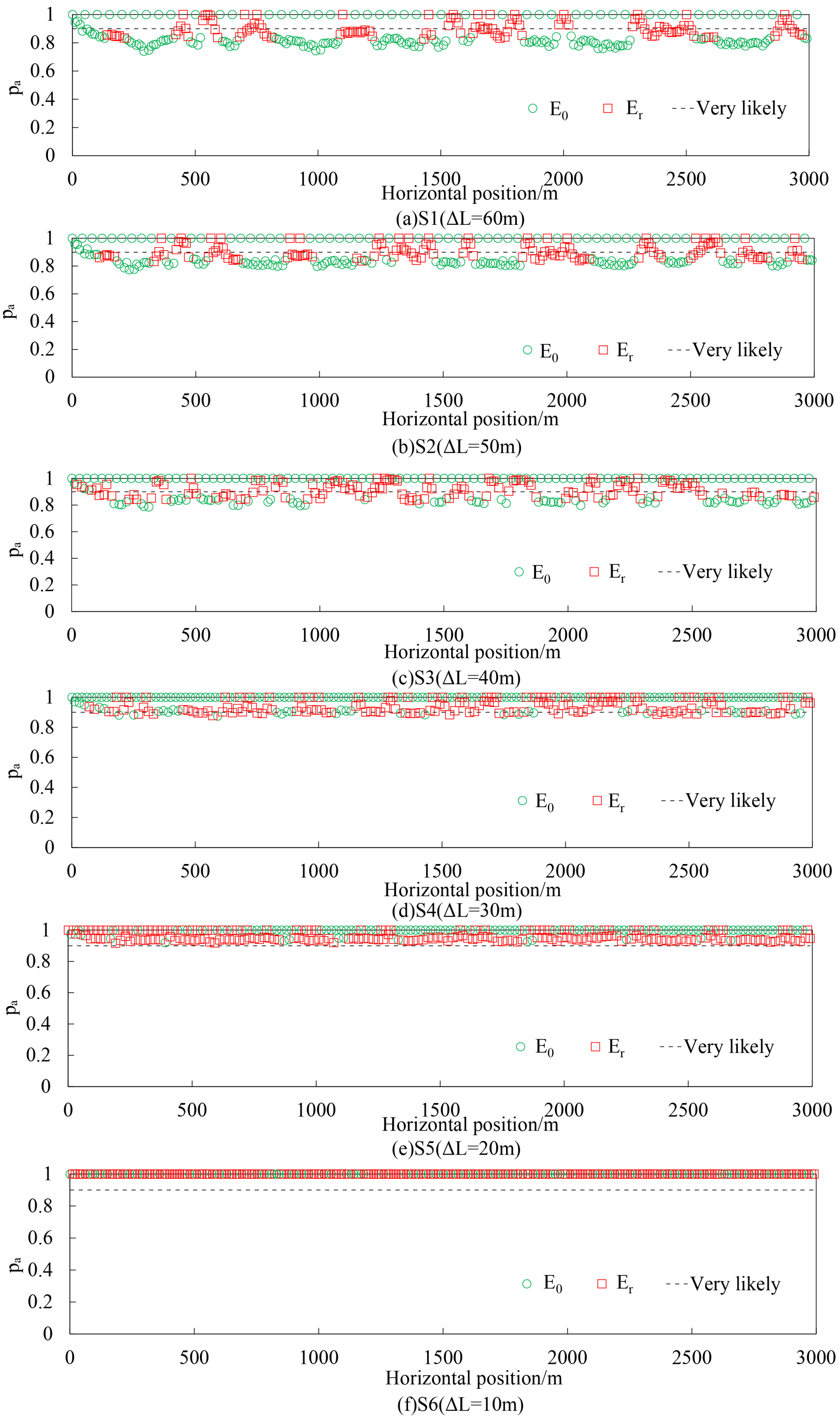

Figure 3. Predictions of the presence of gassy soils at different locations for each candidate scheme.

6. Verification with Measured Data

For a given investigation scheme, the proposed approach gives the locations where the gas pressures are very likely to be risky (*i.e.*, greater than 50 kPa in this example), which are shown in **Figure 3** by open squares. To verify the results obtained from the proposed approach, gas pressure data were obtained from 59 locations along the cross section, as shown in **Figure 4** by open circles. All the gas pressures are greater than 50 kPa, which indicates the occurrence of the risky event at the 59 locations. The results are compared with those shown in **Figure 3**. For a given investigation scheme, the ratio of correct predictions is calculated as $R_{cp} = N_r/59$, where N_r is the number of locations with $p_a = p_{D,e}$ (*i.e.*, E_r has a larger probability

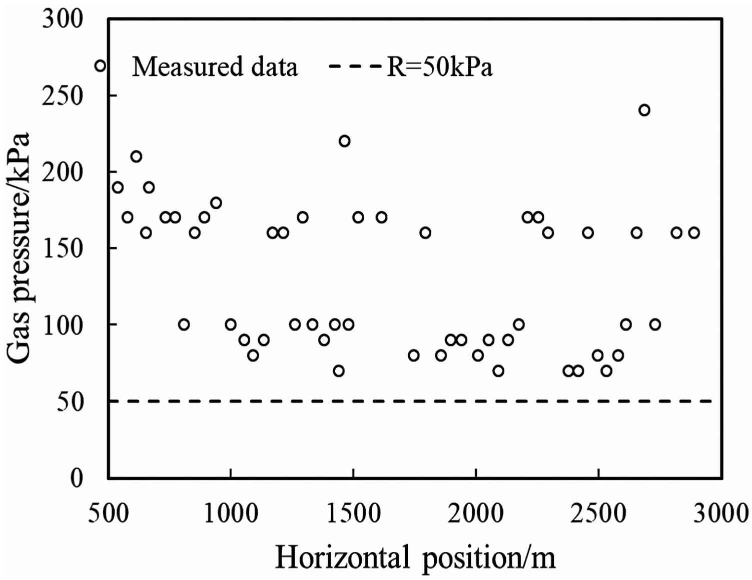

Figure 4. Gas pressure data at 59 locations along the crossing section of Hangzhou Metro Line 1.

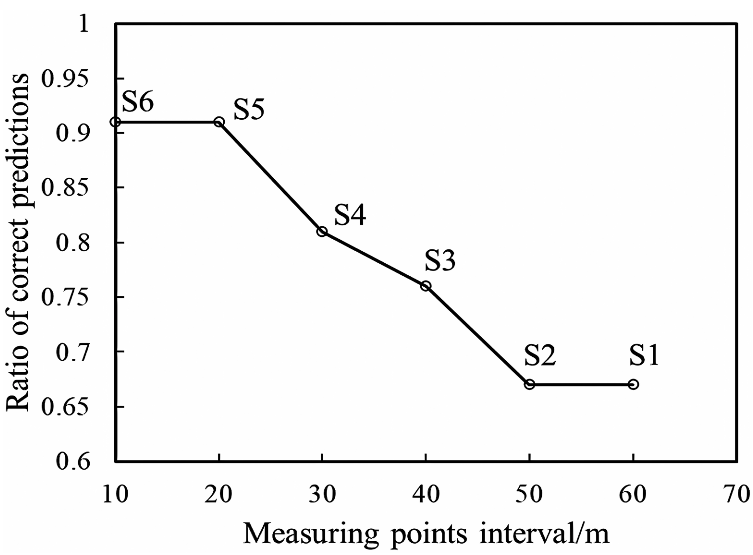

Figure 5. Ratio of correct predictions given different candidate investigation schemes.

than E_0) among the 59 locations. Figure 5 shows the variation of R_{cp} as the borehole interval ΔL decreases from 60 to 10 m, corresponding to S1 - S6 shown in Table 3. The R_{cp} increases considerably as ΔL decreases. For S5 with $\Delta L = 20$ m, the R_{cp} is greater than 0.9. As ΔL decreases to 10 m for S6, the R_{cp} increases slightly. These observations verify the proposed approach and support that S5 with $\Delta L = 20$ m is the optimal investigation scheme among the six candidate ones with a trade-off between investigation efforts and predication accuracy.

7. Summary and Conclusions

This study developed a site investigation optimization method to characterize the spatial distribution of shallow gassy soils along the horizontal direction based on prior knowledge. The horizontal spatial variability of gas pressures is modeled using Kriging-based conditional random field. Then, Monte Carlo simulation is performed to simulate the horizontal spatial distribution of gassy soils with considering the uncertainty in random field parameters given the prior knowledge. Based on the simulated results, the presence of gassy soils was determined based on a prescribed probability threshold value, *i.e.*, 0.9 ("very likely"). Proper investigation schemes shall provide sufficient information to predict the presence of gassy soils and to judge whether the gas pressures are ignorable or risky at different locations of the site concerned. The optimal scheme is then determined as a trade-off between investigation efforts and prediction accuracy. For illustration and validation, the proposed approach was applied to characterize gassy soils along a cross section of Hangzhou Metro Line 1. Results were compared with gas pressure data measured at site. It was shown that the optimal investigation scheme obtained from the proposed method allows characterizing the horizontal spatial distribution of gassy soils reasonably. Last but not the least, it is worthy pointing out that future developments of the proposed approach shall be devoted to 2D/3D spatial variability modeling, efficient optimization algorithms, and decision-making tools and criterion to facilitate site investigation optimization of gassy soils.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Li, Y. and Lin, C. (2010) Exploration Methods for Late Quaternary Shallow Biogenic Gas Reservoirs in the Hangzhou Bay Area, Eastern China. *AAPG Bulletin*, **94**, 1741-1759. https://doi.org/10.1306/06301009184

[2] Lin, C.M., Gu, L.X., Li, G.Y., Zhao, Y.Y. and Jiang, W.S. (2004) Geology and Formation Mechanism of Late Quaternary Shallow Biogenic Gas Reservoirs in the Hangzhou Bay Area, Eastern China. *AAPG Bulletin*, **88**, 613-625. https://doi.org/10.1306/01070403038

[3] Li, L., Zhao, Y. and Yu, L. (2009) Exploration for Quaternary Shallow Biogenic Gas by Sealed Core Drilling and Modified CPT. *Coal Geology and Exploration*, **37**, 72-76. (In Chinese)

[4] Jiang, W., Ye, Z., Zheng, H. and Yong, Z. (1997) Quaternary Shallow Gas Characteristics in Hangzhou Bay and Exploration Method. *Natural Gas Industry*, **17**, 20-23. (In Chinese)

[5] Yan, J., Sun, G., Ru, G. and Rui, S. (1998) Reservoir Formation Conditions and Exploration Prospects of the Shallow Gas in Quaternary along the Coastal Plain of Zhejiang Province. *Natural Gas Industry*, **18**, 16-19. (In Chinese)

[6] Griffiths, D.V. and Fenton, G.A. (2007) Probabilistic Methods in Geotechnical Engineering. Springer.

[7] Liu, L., Cheng, Y. and Zhang, S. (2017) Conditional Random Field Reliability Analysis of a Cohesion-Frictional Slope. *Computers and Geotechnics*, **82**, 173-186. https://doi.org/10.1016/j.compgeo.2016.10.014

[8] Henderson, C.R. (1975) Best Linear Unbiased Estimation and Prediction under a Selection Model. *Biometrics*, **31**, 423-447. https://doi.org/10.2307/2529430

[9] Huang, S.P., Quek, S.T. and Phoon, K.K. (2001) Convergence Study of the Truncated Karhunen-Loeve Expansion for Simulation of Stochastic Processes. *International Journal for Numerical Methods in Engineering*, **52**, 1029-1043. https://doi.org/10.1002/nme.255

[10] Phoon, K.K., Huang, S.P. and Quek, S.T. (2002) Implementation of Karhunen-Loeve Expansion for Simulation Using a Wavelet-Galerkin Scheme. *Probabilistic Engineering Mechanics*, **17**, 293-303. https://doi.org/10.1016/s0266-8920(02)00013-9

[11] Vick, S.G. (2002) Degrees of Belief: Subjective Probability and Engineering Judgment. ASCE Press.