Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly

Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5182-7_08-006-cd

ISGSR 2022

Optimization of Borehole Location for Site Investigation Based on Coupled Markov Chain

Qihao Jiang[1], Hua Shao[2], Jin-Zhang Zhang[3], Cong Nie[4] and Dongming Zhang[5]

[1]Dept. of Geotechnical Engineering, Tongji Univ, Shanghai 200092, China.
E-mail: 2132585@tongji.edu.cn
[2]Shanghai Shentong Metro Co.,Ltd. Shanghai 201103, China.
E-mail: shkypy79@sina.com
[3]Dept. of Geotechnical Engineering, Tongji Univ, Shanghai 200092, China.
E-mail: zhangjz@tongji.edu.cn
[4]Dept. of Geotechnical Engineering, Tongji Univ, Shanghai 200092, China.
E-mail: 2132586@tongji.edu.cn
[5]Dept. of Geotechnical Engineering, Tongji Univ, Shanghai 200092, China.
E-mail: 09zhang@tongji.edu.cn

Abstract: Based on coupled Markov chain and information entropy theory, an optimization method for geological borehole is proposed to obtain the location, depth and quantity of boreholes by considering the tradeoff between the accuracy of ground information and cost of boreholes. The probabilistic distribution of soil information is predicted and identified based on coupled Markov chain theory. To evaluate both the information entropy and exploration cost, genetic algorithm is used to calculate the optimal solution of borehole location and depth under the condition of dual objectives. Based on orthogonal boundary crossing method, the optimal number of boreholes is calculated. Finally, the feasibility of the proposed method is verified by taking the data of Shenzhen Mawan tunnel cross section 4 borehole as an example. The proposed method can provide the basis for the optimization of the location, depth and number of boreholes in tunnel geological prospecting. According to considering the uncertainty of borehole cost and site information comprehensively, it provides a feasible means for predicting engineering geological information based on limited exposed soil data.

Keywords: Coupled Markov chain; Information entropy; Genetic algorithm; Orthogonal boundary crossing method

1 Introduction

The exploratory boreholes are the most effective means to observe the stratigraphic distribution, which is widely used in the engineering field. The cost of geological exploration is directly proportional to the number and depth of exploration points. In addition, the exploration point information should reflect as fully as possible the geological changes in the construction area. Therefore, under the condition of limited exploration cost, the number, location and depth of exploration points are particularly critical.

Fortunately, some scholars have introduced Shannon's information entropy theory into geological modeling to measure the overall and local uncertainties of the model (Wellmann and Regenauer, 2012). Some scholars combined information entropy theory with stratigraphic simulation to explore the borehole layout method. This method could optimize the placement of boreholes and the stratigraphic results were re-simulated based on the optimized scheme, The results showed that the total information entropy of the optimized stratigraphic distribution was significantly lower than the original scheme (Xu et al., 2017; Dai et al., 2020).

It is difficult to discern the geological information of underground engineering. This paper focuses on the optimization of the location, depth and number of new exploration points to achieve the purpose of obtaining more and more accurate geological information by using lower cost. Based on the method of coupled Markov chain and the basic theory of information entropy, considering the exploration cost of the borehole and combining with the existing exploration information data in the site. This paper proposes a method of optimizing the survey scheme based on bi-objective optimization to find the optimal scheme at the inflection point of the Pareto front surface by maximizing the certainty of the scheme and minimizing the survey cost.

2 Methodology

2.1 Implementation procedure

Stochastic simulation of the two-dimensional stratigraphic profile is carried out using the coupled Markov chain theory proposed by Elfeki (Elfeki and Dekking, 2001). Based on the two-dimensional coupled Markov chain model, using the existing borehole and geological a priori data, the transfer probability matrix of the regional stratigraphic soil state in the vertical and horizontal directions is estimated. And combined with the Monte-Carlo method, the stratigraphic distribution state of the unknown area is simulated (Li et al., 2016; Qi et al., 2017). The borehole cost is calculated by referring to the base price table of borehole charges in the Engineering Survey and Design Charges (2002 Revision), and together with the previous information entropy map, an information entropy-borehole cost map is drawn; after that, the Pareto surface is calculated based on NSGA-II, and then the location and depth of the next borehole are determined; finally, the optimized number of boreholes is determined by the orthogonal boundary crossing method. The flow of the detailed stratigraphic simulation method is as follows Figure 1.

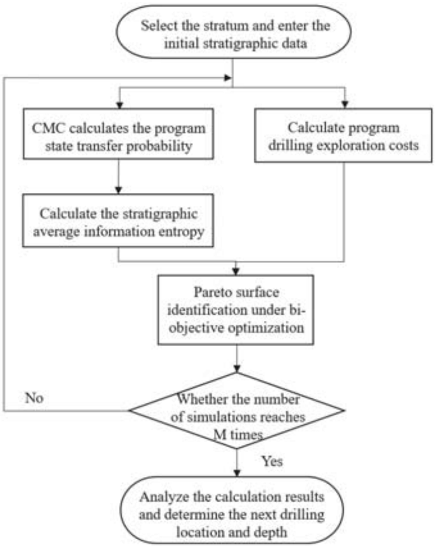

Figure 1. Stochastic simulation process for stratigraphy based on coupled Markov chains.

2.2 The information entropy

Entropy is an important concept in thermodynamics, characterizing the degree of chaos within an isolated system; the greater the degree of chaos in the system, the higher the entropy of the system. Shannon introduced the concept of entropy into information theory and proposed to use information entropy as a measure of the amount of information of a random event and to reflect the uncertainty of the occurrence of a random event; the greater the information entropy indicates the greater the amount of information required and the greater the uncertainty. For a random variable X with the value space $\{x_1,x_2,...,x_n\}$, each of its values in the value space corresponds to a non-negative probability of taking the value. Satisfying condition

$$P(X = x_i) = p_i, i = 1,2, \dots, n, 0 \leq p_i < 1, \sum_{i=1}^{n} p_i = 1 \tag{1}$$

where information entropy is defined as:

$$H(X) = H(p_1, p_2, \dots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{2}$$

The relationship between the information entropy $H(x)$ of this random variable and the probability p obeys the entropy function above, where $H(x)$ is a strictly upper convex function of the probability p. $H(x)$ reaches its maximum value when the two possibilities of the variable taking values are equal, i.e., $p = 1/2$, while $H(x)$ decreases rapidly to 0 when p tends to 0 or 1.

2.3 Couple Markov chain

In the stratigraphic simulation, different soil types are distributed in different spatial locations in the stratum, and they correspond to different states in the Markov chain. Meanwhile, based on the cell division of the stratigraphic space, the distance between soil units is equivalent to the step difference between states in the Markov chain, and thus the Markov chain can be applied to describe the transfer of soil states. Since the simulation of one-dimensional Markov chain is relatively limited, it is more appropriate to use two-dimensional coupled Markov chain to simulate the stratum. The two-dimensional coupled Markov chain describes the coupled behavior of two independent Markov chains X and Y defined on the same state space $\{S_1, S_2, ..., S_n\}$, in horizontal and vertical directions, as shown in Figure 2 below, whose common conditional probability distribution can be expressed as:

$$P\left(Z_{i,j}=S_k \mid Z_{i-1,j}=S_l, Z_{i,j-1}=S_m\right)=C\cdot P(X_i=S_k \mid X_{i-1}=S_l)P\left(Y_j=S_k \mid Y_{i-1}=S_m\right)\Big) \tag{3}$$

where C is the normalization constant that serves as a constraint on the transfer of $(X_i)^h, (Y_i)^h$ to different states, with the expression

$$C=\left(\sum_{f=1}^{n} p_{lf}^{h}\cdot p_{mf}^{v}\right)^{-1} \tag{4}$$

The expression for the transfer probability $p_{lm,k}$of the coupled Markov chain can be obtained from the above equation as:

$$p_{lm,k}=Cp_{lk}^{h}p_{mk}^{v}=\frac{p_{lk}^{h}p_{mk}^{v}}{\sum_{f=1}^{n} p_{lf}^{h}\cdot p_{mf}^{v}} \tag{5}$$

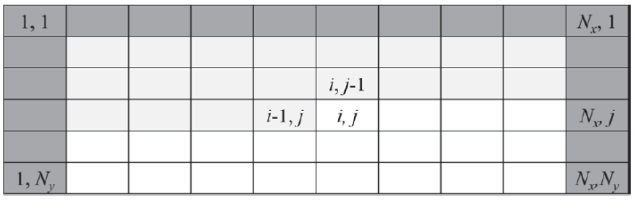

Figure 2. Two-dimensional coupled Markov chain lattice system.

3 Case Study

3.1 Site information

The stratigraphic section to be optimized comes from the shield sea section of Shenzhen Ma Wan cross-sea channel, the stratigraphic section given in the geological survey report is shown in Figure 3, the soil types from top to bottom of the stratigraphy are mainly ③ silt, ⑤ clay, ⑧ sandy chalk and ⑩ mixed weathering rocks with different degrees of weathering. The distance between the two initial survey boreholes ZK30 and ZK31 is 122 m. Two initial survey boreholes and two virtual survey boreholes were used as the actual borehole program, and the location, number and depth of the boreholes were optimized on the basis of this program. In addition, the minimum spacing of the survey points in the tunnel project is 10 m, so a maximum of 9 additional boreholes were drilled on top of the existing ones.

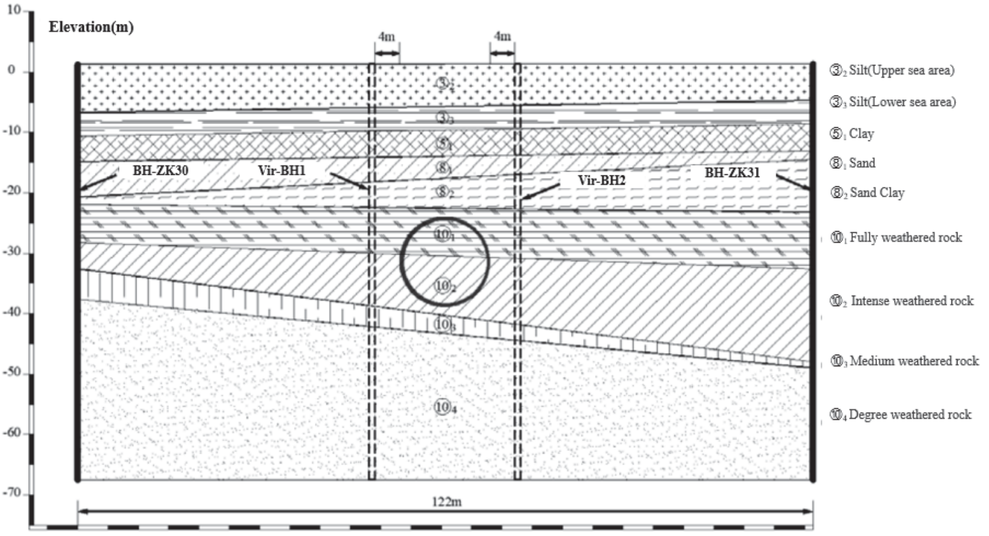

Figure 3. Geological profile of the tunnel cross-section stratigraphy.

Before optimizing the borehole, the borehole cost is first calculated based on the geotechnical type revealed by the borehole and the corresponding fee base price. For each borehole scenario with equal depth, the cost of individual borehole w_0 and the total cost w of the borehole scenario can be calculated separately as:

$$w_0 = \sum_i q_i \cdot h_i \; ; \quad w = n \cdot w_0 \tag{6}$$

Where, q_i is the base price of the fee corresponding to the section *i* of borehole, h_i is the depth of the section *i* of borehole, and *n* is the number of boreholes of the drilling scheme. Combined with the relevant specifications and tariffs, the geotechnical categories and base prices of boreholes through the strata in the region are established as shown in Table 1.

Table 1. The table of base price for drilling through strata.

Depth (m)	Soil type	Geotechnical category	Fee base price (Yuan/m)
h≤10	Silt	I	46
10<h≤15	Clay	II	89
15<h≤20	Sand Clay	III	147
20<h≤25	Sand Clay	III	176
25<h≤30	Fully weathered rock	IV	311
30<h≤35	Fully weathered rock	IV	368
35<h≤40	Intense weathered rock	IV	368
40<h≤45	Intense weathered rock	IV	439
45<h≤50	Medium/Degree weathered rock	V	639
50<h≤60	Medium/Degree weathered rock	V	711
60<h≤70	Medium/Degree weathered rock	V	789

3.2 Optimization Process

Following the method given in 2.2, all possible next borehole location and depth program are first searched with a search step of 2m for the borehole location and depth. Then the borehole allocation scheme is combined with the actual borehole scheme into an additional borehole scheme, which is input to the stratigraphic stochastic simulation algorithm to obtain the simulated average information entropy value and calculate the borehole cost for this borehole optimization scheme. Figure 4 establishes the relationship between the average information entropy and the borehole cost, and each scheme point in the figure corresponds to one borehole scheme. NSGA-II is used to identify the Pareto surface of all solution points, which is the frontier surface composed of red dots in figure 4. The solution points on the Pareto surface are all the optimal drilling solutions at the corresponding cost levels. Finally, the inflection point on the Pareto surface is found to be the optimal solution point, which is the red solid circle.

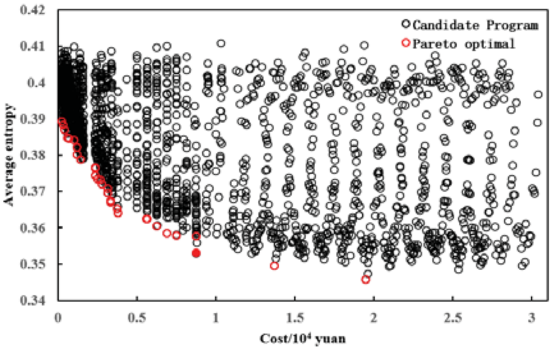

Figure 4. The relationship between average information entropy and drilling costs.

Figure 5 shows the information entropy maps of each grid point, where the larger the value and the darker the color, the higher the information entropy at that location and the higher the uncertainty of the formation. Among them, Figure 5(a) shows the information entropy map of original 4 boreholes, and Figure 5(b) shows the information entropy map of additional 1 borehole. It can be seen that there is a significant decrease in the information entropy value in the additional borehole attachment.

(a) 4 Drilling (b) 5 Drilling

Figure 5. Stratigraphic stochastic simulation results for drilling program.

Based on the results of the stratigraphic stochastic simulation of the original 4 borehole program, the distribution curve of the sum of vertical information entropy in the horizontal direction is plotted, as shown in Figure 7(a). Combining with figure 7(b), the distribution curve of the 5 borehole program, it can be seen that the horizontal coordinate corresponding to the position with the largest information entropy value corresponds to the next optimal drilling arrangement point.

(a) 4 Drilling (b) 5 Drilling

Figure 6. The distribution of information entropy along the horizontal direction.

Next, the number of boreholes was determined. Based on the information entropy calculation results of different number of boreholes, the relationship between the average information entropy and the number of boreholes is established as shown in Figure 7, and the inflection point of the curve is determined as the red solid circle in the figure by using the orthogonal boundary crossing method, so that the optimal number of boreholes is 7. Meanwhile, combining the results of borehole location and depth, the optimal location and depth of 7 boreholes can be determined.

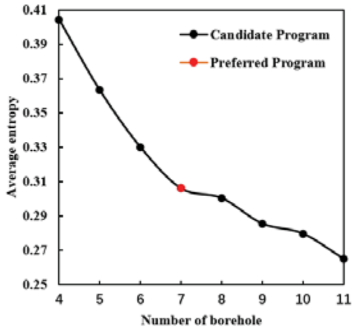

Figure 7. The relationship between average information entropy and number of boreholes.

To reflect the effect of optimizing the drilling depth, the comparison of the results of the optimized and unoptimized drilling solutions is given in Table 2. As can be seen from the table, the average information entropy values of the two are quite close, but the total drilling cost of the unoptimized depth drilling solution is significantly higher than that of the optimized depth drilling solution.

The schematic diagram of the boreholes corresponding to the final 7 drilling scheme is shown in Figure 8. The results of the scheme show that the location of the additional boreholes is at the maximum of the information entropy distribution curve of the original drilling scheme, and the depth of the boreholes is near the stratigraphic junction. So far, the location, number and depth of boreholes have been determined.

Table 2. Average information entropy values and drilling costs for different drilling program.

	Actual 4 borehole	Isometric 7 borehole	Optimization 7 drilling
Average entropy	0.4031	0.3043	0.3059
Total cost/yuan	115415	201742	140888
Total depth/m	280	490	404

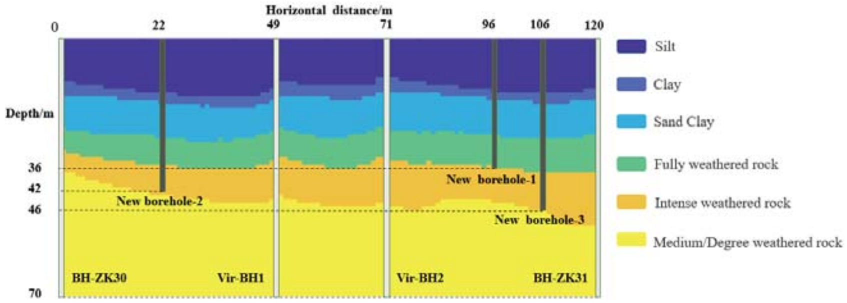

Figure 8. 7 Schematic diagram of the drilling program.

4 Conclusion

With the information entropy results as the criterion for judging the stratigraphic simulation effect, a more complete drilling scheme determination method is proposed in combination with the stratigraphic simulation method of coupled Markov chain model, which aims at minimizing the drilling cost and optimizing the information entropy distribution, and can achieve the optimization of drilling location, quantity and depth. The proposed method is used to optimize the drilling scheme for the cross-sectional stratigraphy of the Mawan tunnel, and the stratigraphic simulation results of the optimized drilling scheme and the non-optimized scheme are compared. The results show that the optimized drilling scheme significantly reduces the uncertainty of the stratigraphic simulation by improving the high information entropy distribution, while reducing the drilling cost by a considerable percentage, thus proving that the optimized drilling scheme is better than the actual engineering drilling scheme or the equally spaced drilling scheme at the stratigraphic simulation level, and preliminarily verifying the effectiveness of the proposed drilling scheme optimization method. There are two main conclusions as follows.

(1) By comparing the information entropy before and after, we can see that the location of the new borehole corresponds to the maximum information entropy distribution curve of the original borehole scheme. The location of the new borehole can be determined on this basis.

(2) The stratigraphic uncertainty at the soil junction is large and needs to be controlled. In the case of known borehole location, considering the exploration cost, the depth of borehole should be near the junction of the deeply buried strata.

Acknowledgments

This study is supported by National Key Research and Development Program for Young Scientists (No. SQ2021YFFO500150) and Scientific research project of Shanghai Shentong Metro Group Program (No. JS-KY20R001). Hereby, the authors are grateful to these programs.

References

Dongming Zhang, Hongfeng Dai, Hui Wang, Hongwei Huang, Dejun Liu. (2021) Investigating the Effect of Geological Heterogeneity of Strata on the Bearing Capacity of Shallow Foundations Using Markov Random Field [D]. *J ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*.

Elfeki A, Dekking M. (2001) A Markov chain model for subsurface characterization: theory and applications[J]. *Mathematical geology*, 33(5): 569-589.

Jin Zhang Zhang, Hong-Wei Huang, Dong-Ming Zhang et al. (2021) Effect of ground surface surcharge on deformational performance of tunnel in spatially variable soil. *Computers and Geotechnics*.

J.Z. Zhang, H.W. Huang, D.M. Zhang et al. (2021) Quantitative evaluation of geological uncertainty and its influence on tunnel structural performance using improved coupled Markov chain [J]. *Acta Geotechnica*.

Wellmann J F, Regenauer Lieb K. (2012) Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models[J]. *Tectonophysics*, 526: 207-216.