Procds. of the 16th Asian Regional Conference on Soil Mechanics and Geotechnical Engineering,
October 14-18, 2019, Taipei, Taiwan

Evaluation of the spatial variability of cone penetration resistance inside an earth-fill dam composed of materials with different particle sizes with use of geostatistics

Kazunari Imaide, S. Nishimura, T. Shibata, and T. Shuku

Graduate School of Environmental and Life Science, Okayama University, 3-1-1 Tsushimanaka, Kita-ku, Okayama 700-8530, Japan

ABSTRACT

In this study, an evaluation method is proposed for the spatial variability of the soil strength derived from the results of cone penetration tests (CPTs) conducted inside an earth-fill dam composed of mixed materials with different particle size distributions. First, in order to reduce the influence of outliers, which have high values or low values, on the modeling of the random field, the measured values are divided into outliers and the others, namely, high, middle, and low groups, respectively. Second, a statistical model is determined for each of the three groups, and the spatial variability of the soil strength considering the outliers is evaluated based on the simulation results of the three groups. The novelty of the approach exists in the re-composition of the simulated values of the three groups. In the proposed method, the measured values are statistically modeled, including the spike-like distribution affected by the outliers. Finally, the estimated values obtained by the proposed method and the measured values are compared at the same location to confirm their correspondence. As a result, it is verified that the proposed method can be used to reasonably simulate the spatial variability of the soil strength inside an earth-fill dam considering outliers.

Keywords: geostatistical method; outliers; spatial variability; earth-fill dam; semi-variogram; cone penetration test

1 INTRODUCTION

The spatial variability of the soil parameters has a great influence on the failure of soil structures, such as levees, earth-fill dams, and slopes. Therefore, the spatial variability of the soil strength should be evaluated appropriately. To cope with the spatial variability inside soil structures, the geostatistical method and the random field theory have been widely used in past studies. However, earth-fill dams in Japan, for instance, sometimes partially contain gravel, as gravel is commonly used to reinforce soil structures. Due to the outliers caused by the gravel, it becomes difficult to apply the geostatistical method for the evaluation of the spatial variability of the soil strength. Therefore, an evaluation method is proposed in this paper for the spatial variability of the cone tip resistance inside an earth-fill dam composed of materials with different particle sizes.

In the present work, first, the outliers are separated from the other data in order to identify the geostatistical parameters of the measured CPT values. To reduce the influence of the outliers on the modeling of the random field, the measured values are then divided into three groups, namely, high, middle, and low, respectively, as shown in Fig. 1. Second, a statistical model is determined for each of the three groups, and a geostatistical simulation is applied for each group. The unique feature is the re-composition of the three groups to incorporate the effect of the outliers into the spatial variability of the soil strength. In the process, based on

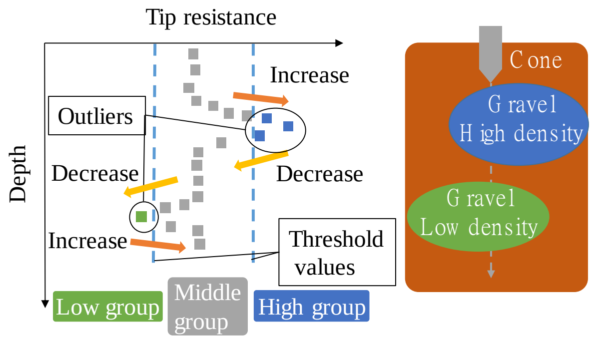


Fig. 1. Distribution of tip resistance in cone penetration tests (CPTs).

the spike-like distribution of the tip resistance affected by the outliers shown in Fig. 1, the locations where the outliers appear are evaluated. Finally, the simulated values for the three groups are re-composed. Furthermore, the measured values are compared with the simulated values at the same location, and the accuracy of the proposed method is validated.

2 STATISTICAL MODELS

For evaluating the soil strength inside a dam, the N-values calculated from the CPTs, N_c, were employed here. The conversion formula used to derive N_c was presented by Suzuki et al. (2003), while Nishimura et al.

(2017) used N_c to identify the weak areas of a river dyke.

At the studied site, CPTs were conducted at 15 points at the top of the dam at intervals of 2 m along the embankment axis, as shown in Fig. 2. The geological cross section of the dam is given in Fig. 3. The height of the dam is 6.6 m, and the soil profile is classified into four layers, namely, backfill sand (Bs), alluvial clay (Ac), alluvial gravel (Ag), and weathered slate (Pl-w).

The soil parameters for the points where test results do not exist can be estimated using a statistical model. To determine the statistical model properly, the maximum likelihood method (MLE) and the semi-variogram, which is one of the geostatistical methods, are employed here, namely, the mean function and the standard deviation are determined by the MLE, and the covariance function, the horizontal correlation distance l_x, and that of depth direction l_z are identified by the semi-variogram, respectively. The detailed procedure for determining the statistical model is described in Nishimura et al. (2016).

Since the measured values contain outliers which can cause the mis-estimation of the statistical parameters, the outliers should be separated from the original data. Thus, outliers which have values close to the maximum or minimum values of the measured values are defined, as shown in Fig. 1. In order to reduce the effects of outliers when modeling the random field inside a dam, the N_c values are divided into three groups, namely, high, $\mathbf{Y}_H$, middle, $\mathbf{Y}_M$, and low, $\mathbf{Y}_L$.

Measured data categorized into $\mathbf{Y}_M$ are used to determine the geostatistical parameters. In Fig. 4, the semi-variogram values are modeled by the regression function. The root mean squared error (RMSE) in $\mathbf{Y}_M$ is calculated from the residual error between the semi-variogram values and the regression function and depends on the threshold values for separating the N-values into the three groups. The threshold values between the high and low groups are determined so that the RMSE is minimized. From among the many candidates for the threshold values, a pair of values that corresponds to the minimum case of the RMSE is finally determined as the optimal case, and the statistical models for the three groups are identified, as shown in Table 1. The correlation distances of $\mathbf{Y}_M$ are reasonable compared with the published values (Phoon and Kulhawy, 1999). In addition, the semi-variograms derived from all the data and those calculated from the optimal case of $\mathbf{Y}_M$ are presented in Fig. 4. The figure shows that the correspondence of the regression function is improved in the optimal case of $\mathbf{Y}_M$.

3 RE-COMPOSITION OF SIMULATION RESULTS

To evaluate the safety of soil structures, information on the outliers inside the soil structures is important. To simulate each spatial variability of the CPT N-value, N_c,

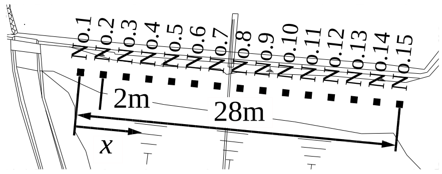


Fig. 2. Plan view of dam and testing points of CPTs.

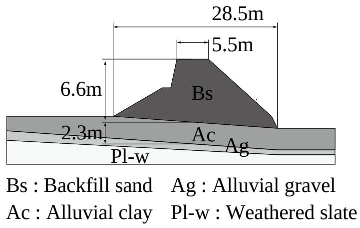
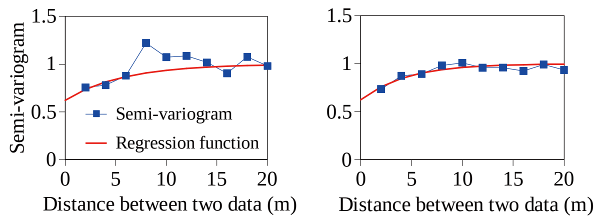


Fig. 3. Geological cross section of dam.

Table 1. Statistical models of $\mathbf{Y}_H$, $\mathbf{Y}_L$, and $\mathbf{Y}_M$.

Mean function	Standard deviation
$\mu = 0.321$	0.371
Covariance function $(i, j = 1, 2, \ldots, M)$	
$\mathbf{Y}_M$: $\mathbf{C} = 0.371^2 \cdot N_e \exp\left(-\lvert x_i - x_j\rvert / 4.45 - \lvert z_i - z_j\rvert / 0.41\right)$ $(i \neq j)$	
$N_e = 0.298$ $\quad (\lvert x_i - x_j\rvert \neq 0, \lvert z_i - z_j\rvert \neq 0)$	
$N_e = 0.377$ $\quad (\lvert x_i - x_j\rvert \neq 0, \lvert z_i - z_j\rvert = 0)$	
$N_e = 0.790$ $\quad (\lvert x_i - x_j\rvert = 0, \lvert z_i - z_j\rvert \neq 0)$	
$\mathbf{C} = 0.371^2$ $(i = j)$	
$\mathbf{Y}_L, \mathbf{Y}_H$: $\mathbf{C} = 0$ $(i \neq j)$	
$\mathbf{C} = 0.371^2$ $(i = j)$	

M: Number of measurements, N_e: Nugget effect parameter



(a) All data. (b) Optimal case of $\mathbf{Y}_M$ data.

Fig. 4. Semi-variograms in horizontal direction.

based on the statistical model for each of the three groups, the geostatistical software library GSLIB (Deutsch and Journel, 1992) is used as a conditional simulation tool. By combining the simulation results of the three groups, namely, $\mathbf{R}_H$, $\mathbf{R}_M$, and $\mathbf{R}_L$, the proper spatial distribution of N_c, $\mathbf{R}_G$, can be evaluated.

In the re-composition of the simulation results, the spike-like distribution of the tip resistance affected by the outliers, presented in Fig. 1, is utilized to evaluate the locations of the outliers. In Fig. 5, difference values D_H and D_L, between the simulation value from middle group R_M and the threshold values between the high and low groups, are defined, namely, $D_H(x, z) = T_H - R_M(x, z)$ and $D_L(x, z) = R_M(x, z) - T_L$, respectively. It is assumed that spatially, around the points where the values of D_H or D_L are small, the probability of the existence of outliers is high.

In addition, the rate of outliers of high strength, included in the simulated results, and the rate of outliers of low strength are assumed to be same as the classification rate for the high range taken from the measured values, P_H, and for the low range, P_L. P_H and P_L are defined in Table 2. The flowchart given in Fig. 6 is used to derive the re-composed results of the simulation, $\mathbf{R}_G$. The procedure shown in the figure is repeated to obtain a large number of realizations for the random field.

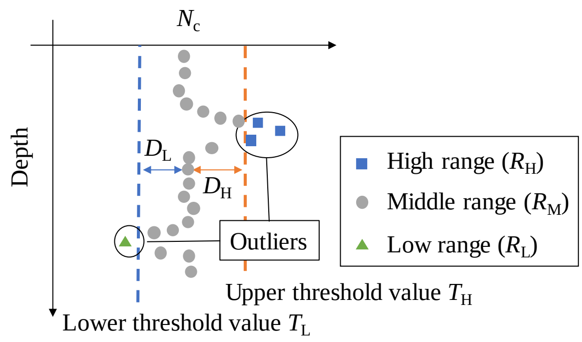


Fig. 5. Definition of D_H and D_L.

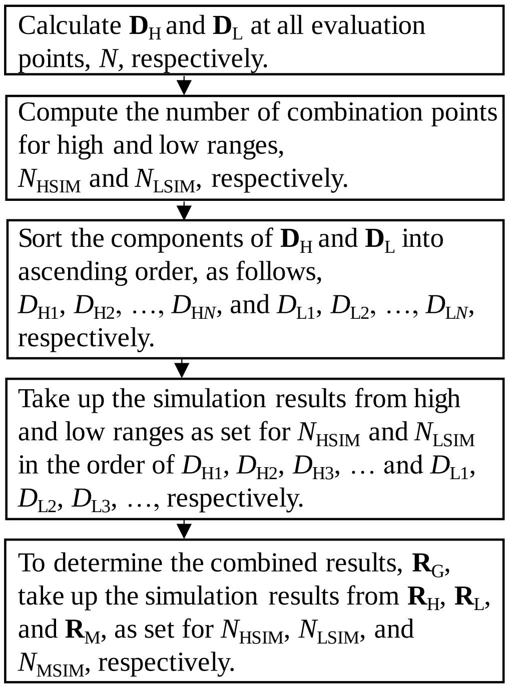


Fig. 6. Flowchart of re-composition of simulation results.

Table 2. Variables for re-composition of simulation results.

N	4831	(Number of all evaluation points)
N_{OB}	1955	(Number of all in-situ data) $N_{OB} = N_{HOB} + N_{MOB} + N_{LOB}$
N_{SIM}	2876	$N_{SIM} = N - N_{OB} = N_{HSIM} + N_{MSIM} + N_{LSIM}$
N_{HOB}	224	(In-situ data of high range)
N_{MOB}	1619	(In-situ data of middle range)
N_{LOB}	112	(In-situ data of low range)
P_H	11.5%	$P_H = N_{HOB} / N_{OB}$
P_L	5.7%	$P_L = N_{LOB} / N_{OB}$
N_{HSIM}	329	$N_{HSIM} = P_H N - N_{HOB}$
N_{MSIM}	2382	$N_{MSIM} = N_{SIM} - (N_{HSIM} + N_{LSIM})$
N_{LSIM}	165	$N_{LSIM} = P_L N - N_{LOB}$

4 EVALUATION OF SPATIAL DISTRIBUTION OF SOIL STRENGTH INSIDE DAM AND VALIDATION

The spatial distribution of the expected values for the N-value from the CPTs, N_c, and the spatial distribution of the probability of N_c>6.25 are given in Figs. 7 and 8, respectively. N_c = 6.25 corresponds to the threshold value, T_H, between the high and middle groups. According to Fig. 7, the expected value for N_c in the dominant space is N_c<4, while there are particularly weak areas around x = 10~17 m and z = 1~9 m. On the other hand, the spatial distribution of the strong areas inside the dam is given in Fig. 8. The probability of the occurrence of outliers of high strength is more than 0.5 within x = 5~25 m and z = 0.5~3 m.

To validate the proposed method, part of the data is intentionally removed from the measured data, and the remaining data are applied to the simulation. The simulated values and the removed data are compared at the same location to verify their correspondence. In Figs. 9 and 10, a comparison of the soil strength distribution at x = 6 m, between the expected values and the in-situ data, and a comparison of the probability density function at x = 6 m, between the simulated values and the in-situ data, are presented, respectively.

According to Fig. 9, the expected values roughly follow the measured values, except for around z = 6 m and z = 8 m. It seems that the difference is caused by the great variability in the soil strength of the measured values. The expected values simulate the trend of the measured values well, but the expected values occasionally yield mis-estimations of the measured values. In Fig. 10, since the shapes of the probability density functions correspond well, namely, both of the distributions have the peak around N_c=2, it is confirmed that the proposed method can simulate the measured data.

5 CONCLUSION

In this paper, an evaluation method has been proposed for the spatial variability of the soil strength derived from the results of cone penetration tests (CPTs) performed inside an earth-fill dam composed of mixed materials with different particle size distributions. The concluding remarks are summarized below.

1. The measured values were classified into three groups, namely, high, middle, and low, by the threshold values between the high and low groups. The threshold values were determined so that the RMSE would be minimized. Since the influence of outliers on the random field modeling of the $\mathbf{Y}_M$ group was reduced by their removal, the correlation distances could be properly estimated. As a result, the horizontal value was about 10 times that of the vertical one.

2. It is seen in the CPTs that the spatial distributions of the weak areas and the strong areas originated from the amount of gravel mixed into the soil and affect the soil strength. In other words, the weak areas contain a smaller amount of gravel, while the strong areas contain a larger amount of gravel. The novelty of the proposed method is in the re-composition of the simulation values of the three groups, namely, high, $\mathbf{Y}_H$, middle, $\mathbf{Y}_M$, and low, $\mathbf{Y}_L$. The $\mathbf{Y}_H$ and $\mathbf{Y}_L$ groups model the outliers of the high strength and the low strength, respectively. In the proposed method, the rate of outliers is determined from the measured values. The simulated values for each of the three groups are re-composed so as to follow the determined rate, and the locations of the outliers are determined based on the simulated values of the middle range, $\mathbf{Y}_M$.

3. As a result of a comparison between the simulated values and the measured values, the distribution of these values at the same place roughly corresponded, and the shapes of the probability density functions were also similar. Thus, it has been verified that the proposed method can be used to reasonably simulate the spatial variability of the soil strength considering outliers.

ACKNOWLEDGEMENTS

This research was partly supported by JSPS KAKENHI Grant Numbers 26252040 and 16H02577.

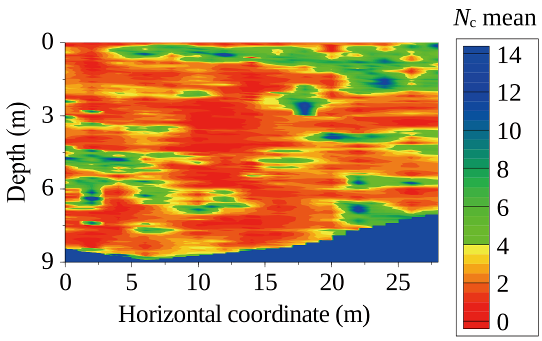


Fig. 7. Spatial distribution of expected values for N_c.

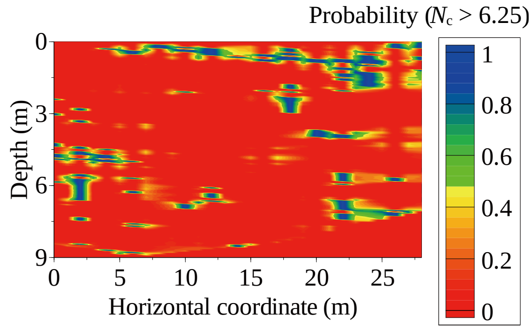


Fig. 8. Spatial distribution of probability of N_c> 6.25.

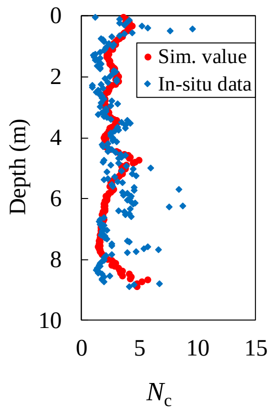


Fig. 9. Comparison of soil strength distribution at x =6 m between expected values and in-situ data.

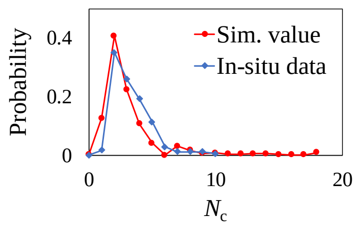


Fig. 10. Comparison of probability density function at x= 6 m between simulated values and in-situ data.

REFERENCES

Journel, A.G., Huijbregts, C.J., 1976. Mining Geostatistics. Academic Press, London, United Kingdom.

Nishimura, S., Shibata, T., Shuku, T., 2016. Diagnosis of earth-fill dams by synthesised approach of sounding and surface wave method. Geo-Risk 10, 312–319.

Nishimura, S., Shibata, T., Shuku, T., Imaide, K., 2017. Geostatistical Analysis for Identifying Weak Soil Layers in Dikes, Geo-Risk 2017: Geotechnical Risk Assessment and Management, GSP 285, 529-538.

Phoon, K. K., Kulhawy, F.H., 1999. Characterization of geotechnical variability. Can. Geotech. J. 36, 612–624.

Suzuki, Y., Tokimatsu, K., Sanematsu, T., 2003. Correlations between CPT data and soil characteristics obtained from CPT, J. Struct. Constr. Eng., AIJ, 566, 73-80 (in Japanese)