Estimation of horizontal transition probability matrix for coupled Markov chain

Xiao-hui Qi\textsuperscript{i)}, Dian-qing Li\textsuperscript{ii)} and Kok-kwang Phoon\textsuperscript{iii)}

\textsuperscript{i}) Ph.D, State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, 8 Donghu South Road, Wuhan 430072, P. R. China.
\textsuperscript{ii}) Professor, State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, 8 Donghu South Road, Wuhan 430072, P. R. China.
\textsuperscript{iii}) Professor, Department of Civil and Environmental Engineering, National University of Singapore, Blk E1A, #07-03, 1 Engineering Drive 2, Singapore 117576.

ABSTRACT

Geologic uncertainty appears in the form of one soil layer embedded in another or the inclusion of pockets of different soil types within a more uniform soil mass. An efficient coupled Markov chain (CMC) model has been proposed to simulate geologic uncertainty in the literature. This model, however, cannot be directly applied to geotechnical engineering. The primary problem lies in the estimation of horizontal transition probability matrix (HTPM), one key input of the CMC model. The HTPM is difficult to estimate due to the wide spacing between boreholes in the horizontal direction. Hence, a practical method for estimating the HTPM is verified using artificial borehole data. The effectiveness of this method is evaluated using the approach as follows. Several virtual boreholes are created using a set of prescribed HTPM and vertical transition probability matrix (VTPM). The HTPM estimated from the virtual boreholes is compared with the prescribed (or actual) HTPM. The evaluation results show that the estimated HTPM agrees well with the prescribed HTPM if the prescribed HTPM and VTPM are both highly diagonally dominant (diagonal element is larger than the sum of off-diagonal elements).

Key words: geologic uncertainty, coupled Markov chain, horizontal transition probability matrix

1 INTRODUCTION

The performance of a geotechnical system such as a slope is affected by soil heterogeneity (e.g. Li et al. 2014). One type of soil heterogeneity is geologic uncertainty. It appears in the form of one soil layer embedded in another or inclusion of pockets of different soil types within a more uniform soil mass (Elkateb et al. 2003). This form of soil heterogeneity can be found in landslide cases (e.g. Evans 1982; Hansen et al. 2007). The role of geologic uncertainty has been studied in geotechnical engineering. For example, Eibacher (1971) pointed out that geologic uncertainty may control the distribution and direction of rockslides after surveying twenty-five rockslides in British Columbia. However, geologic uncertainty is not well studied in the geotechnical engineering literature. Although efforts have been made to consider geologic uncertainty in some geotechnical problems (e.g. Halim, 1991; Kohno et al., 1992), the random process used to model geologic uncertainty is not practical. For example, Kohno et al. (1992) adopted Poisson process to study the system reliability of a tunnel running through rocks. Yet, the Poisson process can only consider two types of soil or rocks. The more general form, namely layers with more than two soil types was not studied.

In reality, geologic uncertainty typically involves more than two soils in a layered profile (e.g. Evans 1982; Hansen et al. 2007). This kind of geologic uncertainty can be modeled quite realistically using a coupled Markov chain (CMC). However, it is not easy to apply CMC in practice. The main problem lies in the estimation of one input parameter for CMC. The main inputs of CMC are the vertical transition probability matrix (VTPM) and the horizontal transition probability matrix (HTPM). The VTPM can be estimated directly from borehole data. It is more difficult to determine the HTPM, because of the typical wide spacings between boreholes. Qi et al. (2015) proposed a practical method to estimate HTPM using real borehole data. This paper further verifies that the proposed method is
reasonable using simulated borehole data.

2 COUPLED MARKOV CHAIN MODEL

2.1 One dimensional first-order Markov chain

One-dimensional Markov chain is a probabilistic model based on the idea that the state of a system in current step depends only on the state in previous step. Given a sequence of random variables \( Z_1, Z_2, ..., Z_n \) taking values in the state space \( \{S_1, S_2, ..., S_m\} \) (\( m \) is the total number of the state), this sequence is a one-dimensional Markov chain if

\[
P(Z_k = S_j | Z_{k-1} = S_i, Z_{k-2} = S_i, Z_{k-3} = S_i, ..., Z_1 = S_i) = P(Z_k = S_j | Z_{k-1} = S_i)
\]  
(1)

where \( P(Z_k = S_j | Z_{k-1} = S_i, Z_{k-2} = S_i, Z_{k-3} = S_i, ..., Z_1 = S_i) \) is the probability of \( Z_k = S_j \) given that \( Z_{k-1} = S_i, Z_{k-2} = S_i, Z_{k-3} = S_i, ..., Z_1 = S_i \). The one-step probability for all the transitions can be denoted as a matrix \( P \) (also called transition probability matrix, TPM), whose elements are \( p_{ij} \) (\( i=1,...,m; j=1,...,m \)). The \( N \)-step transition probabilities can be obtained by multiplying the single-step transition probability matrix by itself \( N \) times.

2.2 Two dimensional Coupled Markov chain

Higher dimensional Markov chain can characterize soil transitions more realistically than a one-dimensional Markov chain. Elfeki and Dekking (2001) proposed a two dimensional Markov chain by coupling two one-dimensional Markov chains, i.e., horizontal Markov chain and vertical Markov chain. The transition probability in horizontal and vertical direction are respectively described by horizontal transition probability matrix (HTPM, \( P^h \)) and vertical transition probability matrix (VTPM, \( P^v \)). The basic idea of the coupling is as follows. As shown in Fig. 1, the domain to be modeled is discretized into a number of cells with the same size. The state of cell \((i, j)\) (\( i \)-column number; \( j \)-row number) depends on the states of the cells on the top \([cell \ (i-1, j)]\), left \([cell \ (i, j-1)]\) of the current cell and the rightmost cell \([cell \ (N_i, j)]\), \( N_i \) is the column number for the rightmost cells in the same row. The conditioning probability is given by Eq. 2.

\[
p_{k,ijk} = P(Z_{i,j} = S_k | Z_{i+1,j} = S_i, Z_{i,j+1} = S_S, Z_{N_i,j} = S_q) = \frac{P^n_{ji} \cdot P^{h(N_i,i)}_{ijk} \cdot P^v_{kq}}{\sum_{l=1}^{m} P^n_{ji} \cdot P^{h(N_i,i)}_{ijk} \cdot P^v_{lk}}
\]  
(2)

where \( p_{k,ijk} \) is the probability that cell \((i, j)\) is in state \(S_k\), given that cell \((i-1, j), (i, j-1)\) and \((N_i, j)\) is in state \(S_S\) and \(S_q\); \( P^n_{ji} \) and \( P^v_{kq} \) are the corresponding elements of the horizontal and vertical transition probability matrices, \( P^h \) and \( P^v \); \( P^{h(N_i,i)}_{ijk} \) \( (P^{v(N,i)}_{ij}) \) is the probability of transition from \(S_i \ (S_j)\) to \(S_q \) in \((N_i,i)\) steps in the horizontal direction. Details are given elsewhere (Elfeki and Dekking 2001).

![Fig. 1 Numbering system in a two-dimensional domain for the coupled Markov chain](image)

2.3 Estimation of transition probability matrices

The transition probability matrices (HTPM and VTPM) are the key inputs of the coupled Markov chain. The VTPM can be estimated directly from borehole data. The HTPM is more difficult to obtain due to the wide spacing between boreholes in the horizontal direction. Hence, a practical method is proposed to estimate the HTPM in this section.

To obtain the vertical transition probability matrix, the boreholes are firstly divided into equidistance intervals in vertical direction. Each interval has a soil state. Thereafter, \( p_{ij}^v \) is calculated by counting the number of state \(S_i\) followed by states \(S_j\) and dividing the number by the total number of transitions from state \(S_i\) in the vertical direction, i.e.

\[ p_{ij}^v = T_{ij}^v / (\sum_{l=1}^{m} T_{il}^v) \]

where \( T_{ij}^v \) (\( T_{il}^v \)) is the number of observed transitions from \(S_i\) to \(S_j\) (\(S_i\)) in the vertical direction. Note that the vertical sampling interval is also referred to as the cell height in Fig. 1.

Theoretically, the horizontal transition probability matrix can be obtained using the same procedure as that described for VTPM. However, it must be note that geology maps are customarily of limited resolution in practice. Geology maps are usually produced at the scale of kilometers. It cannot provide the details of stratigraphy for a geotechnical problem (e.g. slope), which is typically at the scale of meters. Hence, a new method for estimating HTPM based on borehole data is proposed below.

Borehole data are used to estimate the HTPM. The basic idea of the estimation method is similar to the maximum likelihood estimation method. If one has \(N_b\) (\( N_b > 2 \)) boreholes in hand, the two outmost boreholes are used to provide conditional information. The remaining \(N_b-2\) boreholes are viewed as observed scenario (see Fig. 2). Different HTPMs can be assumed. Based on the two conditional boreholes and the assumed HTPM, the likelihood (probability) for the
occurrence of observed scenario can be calculated using Monte Carlo simulation. The HTPM possessing the maximum likelihood of the observed scenario is considered to be the most probable ratio of \( T^n / T^m \). This method is realistic, because it depends on actual borehole data. Details of the estimation method are illustrated by Qi et al. (2015) using real borehole data. This paper further verifies that the proposed method is reasonable using simulated borehole data.

2.4 Evaluation of the proposed HTPM method

The effectiveness of the proposed method is verified using artificial borehole data. The “actual” HTPM and VTPM are obviously known and they are equal to the prescribed matrices. Two virtual boreholes are adopted to conditionally simulate the CMC. Several vertical lines are drawn in one arbitrary realization of CMC. Each line overlays a cell column. These lines are viewed as virtual boreholes. New VTPM and HTPM can be estimated from all the virtual boreholes (boreholes A, B, and C for the example shown in Fig. 2) using the proposed method. The estimated TPMs are compared with the prescribed TPMs (“actual” or correct answers). If the prescribed TPMs and the estimated TPMs are similar, the proposed method can be considered as effective.

![Fig. 2 Boreholes used for HTPM estimation](image)

3 ILLUSTRATIVE EXAMPLES

The approach described in part 2.4 is adopted to evaluate the effectiveness of the proposed method for HTPM estimation. Four artificial sets of VTPM and HTPM [see Table 1(a1, a2, b1, b2, c1, c2, d3, d4)] are prescribed as inputs for CMC to study the range of possible geologic uncertainty. The four sets of TPMs represent four typical types of combination of HTPM and VTPM, i.e. set 1: lowly diagonally dominant HTPM, lowly diagonally dominant VTPM; set 2: highly diagonally dominant HTPM, lowly diagonally dominant VTPM; set 3: lowly diagonally dominant HTPM, highly diagonally dominant VTPM; set 4: highly diagonally dominant HTPM, highly diagonally dominant VTPM. Herein the diagonal dominancy means that the value of diagonal element is larger than the sum of the off-diagonal element for a matrix. The influence of diagonal dominancy on the distribution of soil state can be well observed in Fig. 3(a1, b1, c1, d1). Fig. 3(a1, b1, c1, d1) plots one arbitrary realization of CMC for the four sets of TPMs. As seen from the figures, the extension length (thickness) of soil states increases with an increasing degree of diagonal dominancy of HTPM (VTPM). This phenomenon actually can be easily expected. Imagine one extreme case of the transition probability matrix. If all the values of diagonal elements for the HTPM are 1, one soil state will transit only to itself in the horizontal direction. In this case, one type of soil will run through all the space in a given elevation and no uncertainties exist in the horizontal soil transitions. In another extreme case, if the diagonal elements of HTPM are 0, all the soil states cannot transit to itself in horizontal direction, and the extension length of each state is one cell length. The uncertainty in horizontal soil transition is huge. Hence, the diagonal dominancy controls the length and thickness of each soil state, and uncertainty in soil transition. In other words, the uncertainty in soil transition decreases with an increasing degree of diagonal dominancy of HTPM or VTPM.

| Table 1 Artificial TPMs and estimated TPMs using virtual boreholes |
|---|---|---|
| (a1) artificial VTPM (set 1) | state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.70 | 0.15 | 0.15 |
| 2(sand) | 0.15 | 0.70 | 0.15 |
| 3(silt) | 0.15 | 0.15 | 0.70 |
| (a2) artificial HTPM (set 1) | state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.82 | 0.09 | 0.09 |
| 2(sand) | 0.09 | 0.82 | 0.09 |
| 3(silt) | 0.09 | 0.09 | 0.82 |
| (a3) estimated VTPM (set 1) | state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.68 | 0.11 | 0.22 |
| 2(sand) | 0.14 | 0.67 | 0.19 |
| 3(silt) | 0.12 | 0.14 | 0.74 |
| (a4) estimated HTPM (set 1) | state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.68 | 0.11 | 0.22 |
| 2(sand) | 0.14 | 0.67 | 0.19 |
| 3(silt) | 0.12 | 0.14 | 0.74 |
| (b1) artificial VTPM (set 2) | state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.70 | 0.15 | 0.15 |
| 2(sand) | 0.15 | 0.70 | 0.15 |
| 3(silt) | 0.15 | 0.15 | 0.70 |
| (b2) artificial HTPM (set 2) | state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.92 | 0.04 | 0.04 |
The realizations in Fig. (a1, b1, c1, d1) are used to evaluate the effectiveness of the proposed method. These realizations are viewed as the real conditions of soil profile. Four vertical lines locating in x intervals of [0m, 2m], [18m, 20m], [38m, 40m] and [58m, 60m] are assumed to be four virtual boreholes in these realizations. The soil states in these four virtual boreholes are used to estimate the VTPM and HTPM using the method in part 2.3. The four sets of estimated VTPM and HTPM are summarized in Table 1 (a3, a4, b3, b4, c3, c4, d3, d4). As shown in Table 1, the estimation error (i.e., difference between the artificial TPMs and estimated TPMs) decreases with increasing diagonal dominancy of TPM. For lowly diagonally dominant TPMs (set 1), the estimation error is large. As shown by Table 1(a2, a4), the difference between estimated and artificial (correct) is as large as 0.14. However, as the diagonal dominancy of TPM increases, the estimation error decreases significantly. As shown by Table 1(d2, d4), the estimation errors for set 4 are zero. There is some error for , but it is quite small. The proposed method appears to be effective under the condition of highly diagonally dominant TPM (diagonal element of VTPM>0.8 or diagonal element of HTPM>0.9). This phenomenon can be explained as follows. If one simulates a series of (say 10000) Monte Carlo realizations of CMC, one averaged realization can be obtained from the 10000 realizations. The soil state in each cell of the averaged realization is the most likely soil state in the cell. For example, if the number of realizations for the occurrence of clay, sand and silt in cell (i, j) are 1000, 2000 and 7000, respectively, the most likely soil state in cell (i, j) is silt. As the degree of diagonal dominancy of HTPM or VTPM is high, the uncertainty in distribution of soil state is low (see the discussion in the last paragraph). The averaged realization is similar to each individual realization (if the diagonal element is 1, the averaged realization is the same as each individual realization). In this case, the diagonal dominancy for one realization could be accurately estimated because of the high consistency between individual realization and the averaged realization. On the contrary, if the HTPM or VTPM is lowly diagonally dominant, huge disparity exists between each individual realization and the averaged realization. One individual realization with one degree of diagonal dominancy could be mostly consistent with the averaged realization with another degree of diagonal dominancy. The estimated diagonal dominancy could be easily biased.

Fig. 3(a2, b2, c2, d2) also plots the simulated CMC using the estimated TPMs in Table 1 and the four virtual boreholes in Fig. 3(a1, b1, c1, d1) as conditional information. As shown in Fig. 3, the "simulated CMCs" are very similar with the "real conditions" except set 1 (i.e., lowly diagonally dominant HTPM, lowly diagonally dominant VTPM). This further indicates that the proposed HTPM estimation method are effective when HTPM and VTPM are both highly diagonally dominant.

| state | 1(clay) | 2(sand) | 3(silt) |
|-------|---------|---------|---------|
| (b3) estimated VTPM (set 2) | 0.04 | 0.92 | 0.04 |
| state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.53 | 0.24 | 0.24 |
| 2(sand) | 0.09 | 0.75 | 0.15 |
| 3(silt) | 0.02 | 0.26 | 0.72 |
| (c1) artificial VTPM (set 3) | 0.80 | 0.10 | 0.10 |
| state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.27 | 0.27 | 0.47 |
| 2(sand) | 0.12 | 0.83 | 0.05 |
| 3(silt) | 0.10 | 0.10 | 0.80 |
| (d1) artificial VTPM (set 4) | 0.80 | 0.10 | 0.10 |
| state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.95 | 0.02 | 0.02 |
| 2(sand) | 0.02 | 0.95 | 0.02 |
| 3(silt) | 0.02 | 0.95 | 0.95 |
| (d3) estimated VTPM (set 4) | 0.75 | 0.20 | 0.05 |
| state | 1(clay) | 2(sand) | 3(silt) |
| 1(clay) | 0.08 | 0.10 | 0.81 |
| 2(sand) | 0.02 | 0.95 | 0.04 |
| 3(silt) | 0.08 | 0.10 | 0.81 |
Fig. 3 Artificial and simulated realizations of CMC for different combinations of VTPM and HTPM.
4 CONCLUSIONS

Qi et al. (2015) proposed a practical method to estimate the horizontal transition probability matrix of CMC (HTPM) using real borehole data. This paper further verifies that the proposed method is reasonable using simulated borehole data. The results show that the estimation error of HTPMs decreases with increasing diagonal dominancy of HTPM and VTPM. Significant error may exist in the estimated HTPM if both HTPM and VTPM is lowly diagonally dominant. However, the method is very effective if the HTPM and VTPM are highly diagonally dominant (diagonal element of VTPM>0.8 and diagonal element of HTPM>0.9).

ACKNOWLEDGEMENTS

This work was supported by the National Science Fund for Distinguished Young Scholars (Project No. 51225903), the National Basic Research Program of China (973 Program) (Project No. 2011CB013506) and the National Natural Science Foundation of China (Project No. 51329901).

REFERENCES

1) Eisbacher, G.H. (1971): Natural slope failure, Northeastern Skeena Mountains, Canadian Geotechnical Journal, 8(3), 384-90.
2) Elfeki, A. and Dekking, M. (2001): A Markov chain model for subsurface characterization: theory and applications, Mathematical Geology, 33(5), 569-89.
3) Elfeki, A. and Dekking, F. (2005): Modelling subsurface heterogeneity by coupled Markov chains: directional dependency, Walther’s law and entropy, Geotechnical and Geological Engineering, 23(6), 721-56.
4) Elkateb, T., Chalaturnyk, R. and Robertson, P.K. (2003): An overview of soil heterogeneity: quantification and implications on geotechnical field problems, Canadian Geotechnical Journal, 40(1), 1-15.
5) Evans, S.G. (1982): Landslides and surficial deposits in urban areas of British-Columbia - a review, Canadian Geotechnical Journal, 19(3), 269-88.
6) Halim, I.S. (1991): Reliability of geotechnical systems considering geological anomaly [dissertation], Urbana–Champaign, University of Illinois at Urbana-Champaign.
7) Hansen, L., Eilertsen, R., Solberg, I.L. and Rokoengen, K. (2007): Stratigraphic evaluation of a holocene clay-slide in Northern Norway, Landslides, 4(3), 233-44.
8) Kohno, S., Ang, A.H.S. and Tang, W.H. (1992): Reliability evaluation of idealized tunnel systems, Structural Safety, 11(2), 81-93.
9) Li, D.Q., Qi, X.H., Phoon, K.K., Zhang, L.M. and Zhou, C.B. (2014): Effect of spatial variability of shear strength parameters that increase linearly with depth on reliability of infinite slopes, Structural Safety, 49, 45-55.
10) Qi, X.H., Li, D.Q., Phoon, K.K. (2015): Simulation of geologic uncertainty using coupled Markov chain, Structural Safety (under review).