Predicting Subsurface Stratigraphy using an Improved Coupled Markov Chain Method

Divya Varkey
Dept. of Geoscience and Engineering, Delft University of Technology, Delft, Netherlands

Abraham P. van den Eijnden
Dept. of Geoscience and Engineering, Delft University of Technology, Delft, Netherlands

Michael A. Hicks
Dept. of Geoscience and Engineering, Delft University of Technology, Delft, Netherlands

ABSTRACT: Geological uncertainty can significantly influence the computed response of a geotechnical structure. For example, ignoring the presence of a weak soil layer embedded within a stronger layer and assuming a deterministic stratigraphic boundary can significantly underestimate the probability of failure. In this paper, the coupled Markov chain method has been used for modelling this form of uncertainty. A strategy for estimating the horizontal transition probability matrix with limited data has been proposed, which is one of the biggest challenges with using this method. In particular, different sampling intervals in the vertical and horizontal directions have been considered in estimating the matrix for simulating realistic field situations. The applicability of the proposed method has been demonstrated using a set of CPTs in the Netherlands. The results highlight a problem that arises due to the coupling algorithm used in this method.

1. INTRODUCTION

One of the first steps in the analysis of a geotechnical structure is the characterisation of the site. Uncertainty is inevitable in the characterisation and analysis of structures, due to the inherently heterogeneous nature of soils. For example, the subsurface is composed of different soil units that were formed due to a combination of various geological, environmental and physico-chemical processes (Phoon and Kulhawy, 1999), resulting in the spatial variability of the units and of the properties within the units.

Research on the characterisation of the subsurface is limited, although it is now receiving increasing attention. Wang et al. (2020) and Hu and Wang (2020) have proposed methods for interpolating cone penetration test (CPT) data in a 2D domain and using the interpolated data with the soil behaviour type index to classify the subsurface. Qi et al. (2016), Xiao et al. (2017), Li et al. (2019) and Zhang et al. (2022) have used the coupled Markov chain (CMC) method (Elfeki and Dekking, 2001) for predicting soil types between the measurement locations. The main parameters of the CMC model are the transition probabilities in the vertical and horizontal directions. The former can be readily estimated from the measurements, whereas estimating the latter from the usually limited number of measurements is rather difficult.

In this paper, a methodology for estimating the horizontal transition probabilities has been proposed that considers different sampling intervals in the two directions. The applicability of the proposed method has been demonstrated for a domain comprising five CPTs in the Netherlands.

2. THE COUPLED MARKOV CHAIN METHOD

In the CMC model, two independent 1D Markov chains in the vertical and horizontal directions are coupled together to predict the same state at unsampled locations. In this paper, a state in the CMC model represents a soil type, and a first-

order Markov chain has been used to describe the transition from one state to the other. The first-order Markov chain is based on the principle that the state of a system in the present step is dependent only on its state in the previous step. For example, a 1D Markov chain consisting of a sequence of *n* random variables Z_1, Z_2, …, Z_n taking values from *m* mutually exclusive states is a first-order Markov process if:

$$P\left(Z_i = S_k | Z_{i-1} = S_q, Z_{i-2} = S_t, \dots, Z_1 = S_l\right) = P\left(Z_i = S_k | Z_{i-1} = S_q\right) \quad (1)$$

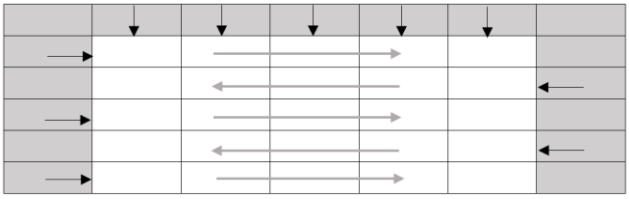
where $P\left(Z_i = S_k | Z_{i-1} = S_q, Z_{i-2} = S_t, \dots, Z_1 = S_l\right)$ is the probability that Z_i is in state S_k given that $Z_{i-1} = S_q, Z_{i-2} = S_t$, ... and $Z_1 = S_q$.

The probabilities of the transitions can be represented in a transition probability matrix (***P***), where each entry p_{ij} of the matrix is the one-step probability of transitioning from state *i* (row) to state *j* (column). To obtain an r-step transition probability from one state to the other, the matrix can be multiplied *r* times. Under the Markovian property, as *r* increases, the probability of being in certain states becomes more likely than being in others. When *r* becomes very large, i.e., when the Markov chain tends to infinity, the distribution will reach an equilibrium with an associated probability of being in each state, known as the stationary probability. This stationary distribution of a Markov chain is a vector ***w*** such that:

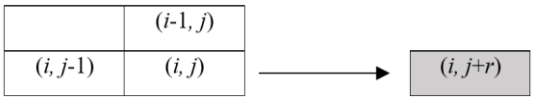
$$\boldsymbol{wP} = \boldsymbol{w} \quad (2)$$

The stationary probability can be calculated by the eigen decomposition of ***P*** and, based on Eq. (2), it is the eigenvector corresponding to an eigenvalue of unity.

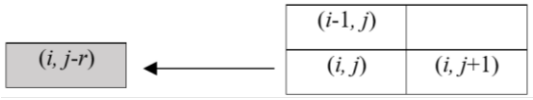
An illustration of the simulation process using the CMC model is shown in Figure 1. In this figure, each cell represents a Markovian step. The shaded cells at the two ends and at the top represent the cells with known states. As shown by the arrows in Figure 1(a), the direction of simulation in the vertical direction is from top to bottom, and alternating between left to right and from right to left in the horizontal direction, thus resulting in forward and backward CMC processes, respectively. Because of this alternating arrangement of the direction of simulation, the soil state in the current step is dependent on the soil state in a previous step generated by the other CMC process, and it is therefore likely to be less biased than a fully forward or a fully backward CMC process. Furthermore, to make full use of the available information and to reduce uncertainty in the simulation, the CMC model in each direction is conditioned on the future states, as was proposed by Elfeki and Dekking (2001).



(a) alternating forward and backward simulation



(b) forward simulation



(c) backward simulation

Figure 1: An illustration of the simulation process used in the paper

With reference to Figure 1(b), using forward CMC simulation, the probability that a cell (i, j) is in the state S_k is given by:

$$P\left(Z_{i,j} = S_k \middle| Z_{i-1,j} = S_l, Z_{i,j-1} = S_q, Z_{i,j+r} = S_t\right) = \frac{{p^{\mathrm{v}}}_{lk} {p^{\mathrm{h}}}_{qk} {p^{\mathrm{h}(r)}}_{kt}}{\sum_{f=1}^{m} {p^{\mathrm{v}}}_{lf} {p^{\mathrm{h}}}_{qf} {p^{\mathrm{h}(r)}}_{ft}} \quad (3)$$

where S_l and S_q are the states in cell $(i\text{-}1, j)$ on the top and in cell $(i, j\text{-}1)$ on the left, respectively, of the cell (i, j), S_t is the state in the conditioning cell $(i, j+r)$ on the right, ${p^{\mathrm{v}}}_{lk}$ is the one-step vertical transition probability from S_l to S_k, ${p^{\mathrm{h}}}_{qk}$ is the

one-step horizontal transition probability from S_q to S_k, and ${p^{\mathrm{h}(r)}}_{kt}$ is the *r*-step horizontal transition probability from S_k to S_t.

Similarly, using backward CMC simulation (see Figure 1(c)), the probability that the cell (i,j) is in the state S_k is given by:

$$P\left(Z_{i,j}=S_k \middle| Z_{i-1,j}=S_l, Z_{i,j+1}=S_q, Z_{i,j-r}=S_t\right) = \frac{{p^{\mathrm{v}}}_{lk}{p^{\mathrm{h}'}}_{qk}{p^{\mathrm{h}(r)'}}_{kt}}{\sum_{f=1}^{m}{p^{\mathrm{v}}}_{lf}{p^{\mathrm{h}'}}_{qf}{p^{\mathrm{h}(r)'}}_{ft}} \quad (4)$$

where S_l and S_q are the states in cell $(i\text{-}1,j)$ on the top and in cell $(i,j\text{+}1)$ on the right, respectively, of the cell (i,j), S_t is the state in the conditioning cell $(i,j\text{-}r)$ on the left, and ${p^{\mathrm{h}'}}_{qk}$ and ${p^{\mathrm{h}(r)'}}_{kt}$ are the one-step and *r*-step horizontal transition probabilities in the backward direction, respectively.

Although the horizontal transition probabilities in the forward and backward directions are not the same, one can be calculated from the other by using the horizontal stationary probabilities $\boldsymbol{w^{\mathrm{h}}}$ by the following equation (Elfeki and Dekking, 2005):

$$\begin{aligned} {p^{\mathrm{h}'}}_{qk} &= {p^{\mathrm{h}}}_{kq}\times\frac{{w^{\mathrm{h}}}_{k}}{{w^{\mathrm{h}}}_{q}} \\ {p^{\mathrm{h(r)}'}}_{kt} &= {p^{\mathrm{h(r)}}}_{tk}\times\frac{{w^{\mathrm{h}}}_{t}}{{w^{\mathrm{h}}}_{k}} \end{aligned} \quad (5)$$

Thus, Eq. (4) can be simplified to:

$$P\left(Z_{i,j}=S_k \middle| Z_{i-1,j}=S_l, Z_{i,j+1}=S_q, Z_{i,j-r}=S_t\right) = \frac{{p^{\mathrm{v}}}_{lk}{p^{\mathrm{h}}}_{kq}{p^{\mathrm{h}(r)}}_{tk}}{\sum_{f=1}^{m}{p^{\mathrm{v}}}_{lf}{p^{\mathrm{h}}}_{fq}{p^{\mathrm{h}(r)}}_{tf}} \quad (6)$$

The transition probabilities in the vertical and horizontal directions can also be represented via matrices, namely the vertical transition probability matrix and the horizontal transition probability matrix.

2.1. Estimating the transition probability matrices

The vertical transition probability matrix ($\boldsymbol{P^{\mathrm{v}}}$) can be directly estimated from the distribution of soil types obtained from the measurements. In order to do that, the 1D soil stratification at the sampled location can be divided into equidistant intervals (cells). Following the division, ${p^{\mathrm{v}}}_{lk}$ is calculated using:

$${p^{\mathrm{v}}}_{lk} = \frac{{T^{\mathrm{v}}}_{lk}}{\sum_{f=1}^{m}{T^{\mathrm{v}}}_{lf}} \quad (7)$$

where ${T^{\mathrm{v}}}_{lf}$ is the total number of cells that have soil type S_l and are followed by cells with soil type S_f in the vertical direction.

In contrast, following the above procedure for estimating the horizontal transition probability matrix ($\boldsymbol{P^{\mathrm{h}}}$) is more complex because of the usually limited number of measurements. A few recent studies have tried to simplify the problem by using either prior geological information or by combining $\boldsymbol{P^{\mathrm{v}}}$ with Walther's law, which states that lithologies that are observed in the vertical depositional sequences must also be deposited in adjacent transects at another scale (Elfeki and Dekking, 2001 & 2005). For example, Qi et al. (2016) and Zhang et al. (2022) proposed a simple method to estimate $\boldsymbol{P^{\mathrm{h}}}$ from $\boldsymbol{P^{\mathrm{v}}}$ by assuming ${T^{\mathrm{h}}}_{lf} = {T^{\mathrm{v}}}_{lf}$ for $l \neq f$ and ${T^{\mathrm{h}}}_{ll} = K \times {T^{\mathrm{v}}}_{ll}$, thus simplifying the problem to finding the value of K. However, it may be noted that, on using this assumption, an *r*-step transition probability in the horizontal direction can no longer be obtained by multiplying the one-step $\boldsymbol{P^{\mathrm{h}}}$ r times.

Considering the above, the following equation has been proposed to estimate $\boldsymbol{P^{\mathrm{h}}}$ by combining the prior knowledge obtained from $\boldsymbol{P^{\mathrm{v}}}$ with Walther's law:

$$\boldsymbol{P^{\mathrm{h}}} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\alpha/\xi}\boldsymbol{Q}^{-1} \quad (8)$$

where $\boldsymbol{Q}$ and $\boldsymbol{\Lambda}$ are the matrices of eigenvectors and eigenvalues, respectively, of $\boldsymbol{P^{\mathrm{v}}}$, α is the ratio of cell (step) size in the horizontal to vertical direction, and ξ represents the ratio of the scale of deposition of a geological unit in the horizontal direction to that in the vertical. The unknown ξ can be estimated by maximising the likelihood (L) of observing the measurements.

Note that the proposed method of estimating $\boldsymbol{P^{\mathrm{h}}}$ using Eq. (8), in addition to incorporating the effect of different step sizes, ensures that the

distributions of soil types in the horizontal direction are also governed by the same marginal distributions as are obtained from the measurements.

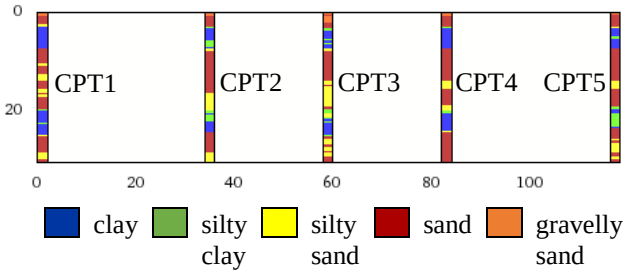


Figure 2: Soil types identified at the locations of the five selected CPTs (scale in m)

3. RESULTS AND DISCUSSIONS

The applicability of the proposed model has been demonstrated for a domain comprising five CPTs (CPT1–CPT5) in the Netherlands. As shown in Figure 2, the CPTs are spread over a domain of width 118 m and were carried out to a depth of 30 m below the ground surface. Also shown in the figure are the soil-behaviour type classifications at the CPT locations. Although any method of CPT-based soil classification can be used, these have here been obtained by plotting the normalised CPT measurements on the Robertson chart (Robertson, 1990) and assuming a minimum layer thickness of 0.4 m. The stratification between the CPTs has been determined using the CMC model described in Section 2, by dividing the domain into cells of height 0.4 m and of different widths based on the value of α.

The vertical transition probability matrix obtained using the spatial distribution of soil types at the CPTs is listed in Table 1(a). The stationary probabilities calculated using $\boldsymbol{P}^{\mathbf{v}}$ represents the spatial distribution of soil types over the domain. However, due to the limited number of measurements there can be uncertainties in the estimated $\boldsymbol{P}^{\mathbf{v}}$. Therefore, inconsistencies in the form of negative probabilities could appear in the matrix on using smaller sampling intervals. Although, in this paper, this inconsistency has been avoided by nullifying all the negative probabilities, they can also be avoided by defining the transition probabilities based on the rate of change of states in a continuous Markov process.

Table 1: Estimated transition probability matrices

(a) $\boldsymbol{P^v}$

	clay	silty clay	silty sand	sand	gravelly sand
clay	0.778	0.111	0.049	0.062	0
silty clay	0.551	0.379	0.035	0.035	0
silty sand	0.029	0.044	0.652	0.275	0
sand	0	0.033	0.111	0.850	0.006
gravelly sand	0	0	0	0.546	0.454

(b) $\boldsymbol{P^h}$

	clay	silty clay	silty sand	sand	gravelly sand
clay	0.948	0.031	0.010	0.011	0
silty clay	0.159	0.833	0.005	0.003	0
silty sand	0.003	0.011	0.927	0.059	0
sand	0	0.006	0.024	0.969	0.001
gravelly sand	0	0	0	0.124	0.876

For a certain value of α, i.e., for a discretisation with a cell width of 0.4 times α, $\boldsymbol{P}^{\mathbf{h}}$ has been calculated using Eq. (8) by choosing the value of ξ that maximises the likelihood of independently observing CPT2 and CPT4 simultaneously and conditioning on CPT1 and CPT3 and on CPT3 and CPT5, respectively. For different trial values of ξ, the CMC simulation is performed repeatedly in a Monte Carlo simulation to compute the likelihood using:

$$L = \prod_{i=1}^{N_{\mathrm{v}}} \left(\frac{\sum_{R=1}^{N_R} \mathrm{I}(Z^R{}_{i,\mathrm{CPT2}} = S_{i\mathrm{CPT2}})}{N_R} \times \frac{\sum_{R=1}^{N_R} \mathrm{I}(Z^R{}_{i,\mathrm{CPT4}} = S_{i\mathrm{CPT4}})}{N_R} \right) \quad (9)$$

where I(.) is an indicator function which is equal to 1 if the condition within the brackets is true and is 0 otherwise, N_v is the total number of cells in the vertical direction, N_R is the total number of Monte Carlo realisations, $Z^R{}_{i,\mathrm{CPT}j}$ is the soil type predicted in realisation R for a cell at depth i in CPTj, and $S_{i\mathrm{CPT}j}$ is the identified soil type in that cell based on the CPT measurements (i.e., the correct solution).

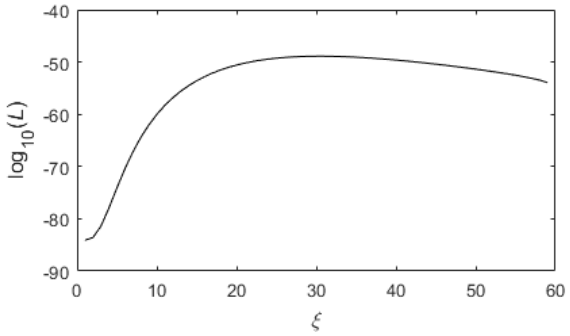


Figure 3: Likelihoods estimated as a function of ξ

Table 2: Stationary probabilities

soil type	1D		2D
	vertical	horizontal	
clay	0.226	0.226	0
silty clay	0.081	0.081	0
silty sand	0.197	0.197	0.004
sand	0.491	0.491	0.996
gravelly sand	0.005	0.005	0

The logarithm of likelihoods computed using 1000 realisations are shown in Figure 3. As can be seen from the figure, the maximum likelihood is obtained for $\xi \approx 30$. The corresponding $\boldsymbol{P}^{\mathbf{h}}$ computed for $\alpha = 5$ is listed in Table 1(b).

Table 2 shows the stationary probabilities computed in 1D by using the two transition probability matrices individually and those in 2D obtained by combining the matrices to predict the same soil type, i.e., by following the coupling principle of CMC. As expected, identical stationary probabilities are computed in 1D using $\boldsymbol{P}^{\mathbf{v}}$ and $\boldsymbol{P}^{\mathbf{h}}$. However, because of the coupling in CMC, the stationary probabilities computed in 2D are quite different from those in 1D. This highlights a problem with the theory behind the CMC model itself. That is, the stationary probabilities of the coupled model do not reflect the marginal distributions of the soil types (estimated here from the CPTs), although the individual 1D chains do satisfy this criterion. This problem requires a more detailed investigation and correction of the CMC theory and is outside the scope of this paper.

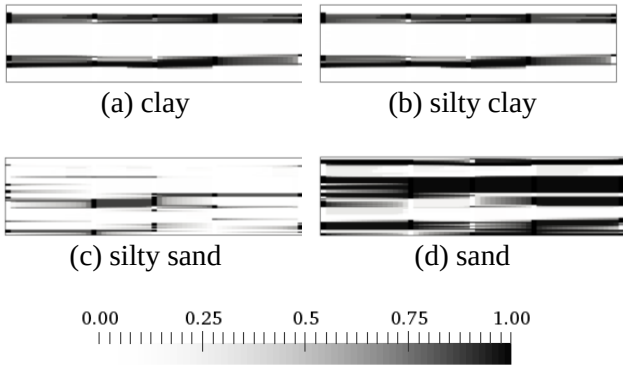
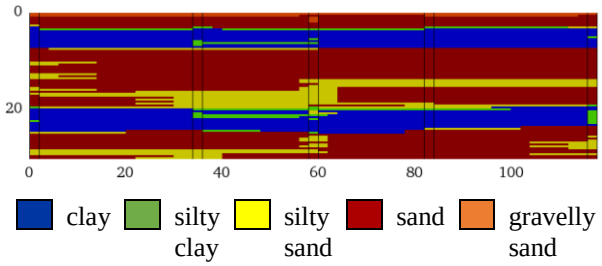
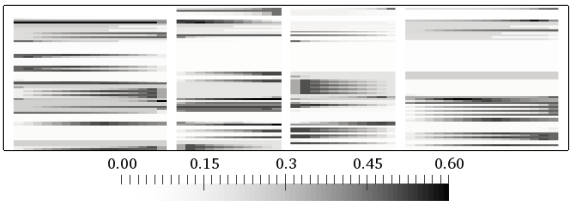


Figure 4: Occurrence probability of the soil types



(a) most likely stratigraphy between the CPT locations (scale in m)



(b) uncertainty in the most likely prediction

Figure 5: The most likely scenario of subsurface stratification and its uncertainty

Following the estimation of the two matrices, CMC simulation was carried out using $\alpha = 5$ in the four zones delineated by the five CPTs and

conditioned on them. A total of 1000 realisations were carried out to predict soil types between the CPTs. To demonstrate the distributions of the predicted soil types in the domain, Figure 4 shows their occurrence probabilities calculated using:

$$O_{i,j,k} = \frac{\sum_{R=1}^{N_R} \mathrm{I}(Z^R{}_{i,j}=S_k)}{N_R} \qquad (10)$$

where $O_{i,j,k}$ is the occurrence probability of soil type S_k in cell (i,j), and $Z^R{}_{i,j}$ is the soil type predicted in a realisation for that cell.

The generated stratigraphic realisations can also be summarised via a single representative scenario, for example, the most likely scenario amongst all the realisations, as is shown in Figure 5(a). The most likely soil type in a cell has been determined by choosing the soil type with the highest occurrence probability in that cell. Figure 5(b) shows significant deviations of the average predictions from this most likely scenario.

4. CONCLUSIONS

The coupled Markov chain method has been used in this paper to predict the stratification between the measurement locations. This method is based on coupling first-order Markov chains in the vertical and horizontal directions to predict the same soil type at any location. In this method, the transition probabilities in each direction are represented in a transition probability matrix. The vertical transition probability matrix can directly be estimated from the distribution of soil types obtained from the measurements, whereas estimating the horizontal transition probability matrix can be more complex because of the usually limited number of measurements.

A methodology for estimating the horizontal transition probability matrix has been proposed in this paper that considers different scales of deposition of soil types in the vertical and horizontal directions. The proposed methodology ensures that, irrespective of different sampling intervals, the distributions of soil types in the two directions are identical and reflect their marginal distributions.

The applicability of the proposed method has been demonstrated via coupled Markov chain simulation for a domain comprising five CPTs in the Netherlands. The results highlight a problem arising from the coupling of the 1D Markov chains that needs correction.

5. ACKNOWLEDGEMENT

This work is a part of the research programme DeepNL/SOFTTOP with project number DEEP.NL.2018.006, financed by the Netherlands Organisation for Scientific Research (NWO).

6. REFERENCES

Elfeki, A.M.M, and Dekking, F.M. (2001). "A Markov chain model for subsurface characterization: Theory and applications." *Mathematical Geology*, 33(5), 569–589.

Elfeki, A.M.M., and Dekking, F.M. (2005). "Modelling subsurface heterogeneity by coupled Markov chains: Directional dependency, Walther's law and entropy." *Geotechnical and Geological Engineering*, 23(6), 721–756.

Hu, Y., and Wang, Y. (2020). "Probabilistic soil classification and stratification in a vertical cross-section from limited cone penetration tests using random field and Monte Carlo simulation." *Computers and Geotechnics*, 124, 103634.

Li, J.H., Cai, Y.M., Li, X.Y., and Zhang, L.M. (2019). "Simulating realistic geological stratigraphy using direction-dependent coupled Markov chain model." *Computers and Geotechnics*, 115, 103147.

Phoon, K.K., and Kulhawy, F.H. (1999). "Characterization of geotechnical variability." *Canadian Geotechnical Journal* 36(4), 612–624.

Qi, X.H., Li, D.Q., Phoon, K.K., Ciao, Z.J., and Tang, X.S. (2016). "Simulation of geologic uncertainty using coupled Markov chain." *Engineering Geology*, 207, 129–140.

Robertson, P.K. (1990). "Soil classification using the cone penetration test." *Canadian Geotechnical Journal*, 27(1), 151–158.

Wang, Y., Hu, Y., and Zhao, T. (2020). "Cone penetration test (CPT)-based subsurface soil classification and zonation in two-dimensional vertical cross section using Bayesian

compressive sampling." *Canadian Geotechnical Journal*, 57(7), 947–958.

Xiao, T., Zhang, L.M., Li, X.Y., and Li, D.Q. (2017). "Probabilistic stratification modelling in geotechnical site characterization." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(4), 04017019.

Zhang, J.Z., Liu, Z.Q., Zhang, D.M., Huang, H.W., Phoon, K.K., and Xue, Y.D. (2022). "Improved coupled Markov chain method for simulating geological uncertainty." *Engineering Geology*, 298, 106539.