

文章编号: 1000-7598 (2009) 08-2348-05

模糊聚类在地层分析中的应用

张 鹏^{1,2}, 张在明^{1,2}, 杨宇友¹, 胡鹏飞¹

(1. 北京工业大学 建筑工程学院, 北京 100124; 2. 北京市勘察设计研究院有限公司, 北京 100038)

摘 要: 传统上, 对于依据原始资料进行地层的划分与合并往往依靠工程师个人的经验与习惯。引入了模糊聚类概念, 提出了一种根据基本物理力学指标进行土层归并分类的合理方法, 物理意义明确, 计算简捷, 具有普遍意义。还提出了一种用参数的变异系数来检验模糊聚类结果的方法, 使模糊聚类的结果与统计参数对应起来, 从而使验证结果满足规范的规定, 保证了地层归并的合理性。

关 键 词: 地层分析; 工程勘察; 模糊聚类

中图分类号: TU 12

文献标识码: A

Application of fuzzy cluster to geotechnical stratum analysis

ZHANG Peng^{1,2}, ZHANG Zai-ming^{1,2}, YANG Yu-you¹, HU Peng-fei¹

(1. College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China;

2. BGI Engineering Consultants Ltd., Beijing 100038, China)

Abstract: Traditionally, it is usually depended on engineers' personal experience and habits to establish the strata layers on the basis of the original data when they doing the desk studies of a site investigation work. The concept of fuzzy cluster is introduced to propose a more reasonable way to merge and classify the soil based on the physico-mechanical indexes of the soil. It can be seen that this method is more simple and universal. In addition, a method used to check the results of fuzzy cluster based on the variation coefficients study of the soil parameters is proposed. The results of fuzzy cluster are corresponding with the statistical parameters; meanwhile, the method also can be used to guarantee the results to meet the standard requirements; and the rationality of the way about merging of soil is proved subsequently.

Key words: stratum analysis; geotechnical investigation; fuzzy cluster

1 引 言

土是自然地质历史的产物, 充满了独特性和变异性, 这为其分类与归并的研究提供了广泛的空间。工程勘察中的土层划分通常都是参照土的工程分类以及其他力学性质进行的。我国的规范基本上是根据塑性指数 I_p 进行土层划分的。但是在许多情况下, 如果严格按照规范中的塑性指数范围一一划分土层, 往往出现工程场地土层过多、过于繁杂的情况, 这不仅造成剖面图杂乱无章、工程分析的困难, 也掩盖了土层分布最基本的力学规律。以北京市区为代表的冲洪积扇地层尤其如此, 如国家体育场的工程场地, 如果不进行适当的合并, 在对工程有影响的深度内, 土层可达近 40 层。对土层进行适当的

归并是工程勘察的基本做法, 数十年来的经验证明, 这样做是必不可少的, 已经形成了工程勘察中工程主持者的一道基本工序。但过去地层归并的原则是塑性指数相近, 沉积循环相邻。具体操作时则是依据各人的工程经验。这种模式的问题是: ①由于缺乏一定的规则, 不同工程师的习惯和经验做法常常出现差别; ②对归并后的地层与各分层之间的土工参数在统计意义上的合理性缺乏认真的研究, 就直接提供给设计方采用。

在岩土工程领域, 一方面, 人们强调依据经验的工程判断^[1], 这反映了岩土工程的基本工作规律, 是完全正确的; 另一方面, 随着各种工程分析根据的不断发展与扩充, 也应当运用这些工具, 改进我们的工作。Baecher 和 Christian 曾指出^[2-3]: “岩

收稿日期: 2007-10-12

基金项目: 北京市自然科学基金重点项目(No. KZ200710005004); 教育部博士点基金项目(No. 20060005012)。

第一作者简介: 张鹏, 男, 1982 年生, 硕士研究生, 主要从事岩土工程勘察方面的学习和研究。E-mail: zhangpeng0217@126.com

土工程在 21 世纪中要成为一个能有效解决问题并且受人尊重的学问, 就是要正视这种不确定问题, 通过量化的运算和处理, 使工程师们得到可靠的数据, 得出可靠的工程判断”, 这说明了科学的“工程判断”离不开定量的计算分析。

数学上一种按照样本的彼此相近的程度进行分类的理论称为模糊聚类。人们利用它的基本原理在岩土工程中进行了膨胀土的分类、砂土的地震液化判别等^[4]。运用的例子还包括樊哲超等^[5]对堤坝渗漏进行了研究分析, 曲永玲等利用扁铲侧胀试验数据对土层进行了分析^[6]等等。本文利用多参数因子的模糊聚类分析, 从物理力学的相近性考虑, 对工程勘察实测得到的原始剖面中分布与性质相近的地层进行合理的模糊聚类合并。在文献[7]的基础上, 对聚类分析过程首次提出了一种聚类合理性的验证方法, 进行了深入的讨论, 得到的剖面图不仅更能显示土层的基本规律, 在参数统计上会更加合理, 而且也使得土层的归并划分更有依据。

2 模糊聚类的方法

所谓聚类分析就是根据事物间的不同特征、亲疏程度和相似性等关系, 对它们进行分类的一种数学方法, 如果伴随有模糊性的分类, 就是模糊聚类。如前所述, 在通常的分类学中, 人们主要靠经验和专业知识, 很少利用数学工具。随着生产技术和科学的发展, 分类越来越细, 以致有时候光凭经验和专业知识还不能确切分类, 而近几十年来模糊聚类这个分支很受人们的重视, 并解决了越来越多的实际问题^[4-5]。

2.1 模糊矩阵

一般地, 若 $X=\{x_1, x_2, \dots, x_m\}$ 和 $Y=\{y_1, y_2, \dots, y_n\}$ 均为有限集, 则它们的直积集 $X \times Y = \{(x, y) \mid x \in X, y \in Y\}$ 中的模糊关系可用隶属函数 $\mu_{\tilde{R}}(x, y)$ ($x \in X, y \in Y$) 表示为一个 $m \times n$ 阶的矩阵叫做模糊矩阵, 记为

$$\tilde{R} = (r_{ij})_{m \times n} = [\mu_{\tilde{R}}(x_i, y_j)]_{m \times n} = \begin{pmatrix} \mu_{\tilde{R}}(x_1, y_1) & \mu_{\tilde{R}}(x_1, y_2) & \cdots & \mu_{\tilde{R}}(x_1, y_n) \\ \mu_{\tilde{R}}(x_2, y_1) & \mu_{\tilde{R}}(x_2, y_2) & \cdots & \mu_{\tilde{R}}(x_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{\tilde{R}}(x_m, y_1) & \mu_{\tilde{R}}(x_m, y_2) & \cdots & \mu_{\tilde{R}}(x_m, y_n) \end{pmatrix} \quad (1)$$

式中: $0 \leq \mu_{\tilde{R}}(x_i, y_j) \leq 1, 1 \leq i \leq m, 1 \leq j \leq n$ 。

可见, 模糊矩阵与普通矩阵形式相同, 但模糊矩阵中的元素都是 $[0, 1]$ 中的数, 且模糊矩阵的运算和普通矩阵也不同。例如, 对于模糊矩阵 $\tilde{R}$ 和 $\tilde{Q}$,

它们的乘法遵循“最大-最小原则”^[8], 公式如下 (模糊矩阵的乘法符号用“ $\cdot$ ”表示):

$$\tilde{R} \cdot \tilde{Q} = \left\{ \max_{k=1}^n \min [r_{ik}, q_{kj}] \right\} = \left\{ \bigvee_{k=1}^n [r_{ik} \wedge q_{kj}] \right\} \quad (2)$$

2.2 模糊等价关系

事物都是普遍联系的, 它们因共性而聚类, 因个性而区别。在经典集合论中, 通常利用“等价性”来区分它们。

设论域 X 上的一个模糊矩阵 $\tilde{R} = (r_{ij})_{n \times n} = [\mu_{\tilde{R}}(x_i, x_j)]_{n \times n}$, 若满足:

(1) 自反性 $\mu_{\tilde{R}}(x_i, x_i) = 1 (i = 1, 2, \dots, n)$, 即任何一个对象必和自己同在一类;

(2) 对称性 $\mu_{\tilde{R}}(x_i, x_j) = \mu_{\tilde{R}}(x_j, x_i)$, 即若对象甲与对象乙同类, 则乙和甲也应同类;

(3) 传递性 $\mu_{\tilde{R}}(x_i, x_j) \geq \lambda, \mu_{\tilde{R}}(x_j, x_k) \geq \lambda \Rightarrow \mu_{\tilde{R}}(x_i, x_k) \geq \lambda, (\lambda \in [0, 1]; i, j, k = 1, 2, \dots, n)$, 即如果对象甲与对象乙同类, 而乙又与对象丙同类, 则甲与丙也应同类; $\tilde{R}$ 为模糊等价关系^[8]。

2.3 传递闭包

设 $\tilde{R}$ 为一个 n 阶模糊矩阵, 包含 $\tilde{R}$ 的最小的模糊传递矩阵称为 $\tilde{R}$ 的传递闭包, 记做 $t(\tilde{R})$, 公式为

$$t(\tilde{R}) = \bigcup_{k=1}^{\infty} \tilde{R}^k \quad (3)$$

一般情况下, 所建立的模糊关系仅满足自反性和对称性, 不满足传递性, 因此, 不是模糊等价关系, 但可以证明, 按照所谓“传递闭包”理论改造后的 $t(\tilde{R})$ 必是模糊等价关系^[8]。

3 地层模糊聚类的思路及程序化实现

作为模糊矩阵的构成基础, 需要先将数据构成一个初始矩阵。对工程勘察所得的土工参数进行初步统计后, 取未经处理的各原始地层每一维指标的平均值作为其代表值。设未经合并前, 场地共有 n 个原始地层 $x'_i (i = 1, \dots, n)$ 。每个地层各有 m 个需要考虑的岩土参数, 即 m 维聚类特性指标, 也就是说, x'_i 可由 m 维特性指标向量 $\mathbf{X}_i = \{x'_{i1}, x'_{i2}, \dots, x'_{im}\}$ 来表达, 其中 x'_{ij} 表示第 i 个原始地层的第 j 项特性指标数据值, 这样 n 个原始地层的所有特性指标构成的初始矩阵记作:

$$\mathbf{X}^* = \begin{pmatrix} x'_{11} & x'_{12} & \cdots & x'_{1m} \\ x'_{21} & x'_{22} & \cdots & x'_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x'_{n1} & x'_{n2} & \cdots & x'_{nm} \end{pmatrix} \quad (4)$$

3.1 数据标准化

由于量纲和绝对值意义均不同的各项模糊聚类指标不具有可比性。为了平衡各指标的作用, 消除它们之间的差异, 可以按下式对样本的各项原始数据指标进行无量纲标准化处理:

$$x_{ij} = \frac{x'_{ij} - \min x'_j}{\max x'_j - \min x'_j} \quad (5)$$

式中: $\max(x'_j)$ 、 $\min(x'_j)$ 分别是初始矩阵中第 j 项指标的最大值和最小值。经过归一化处理后, 消除了量纲的影响, 得 $0 \leq x_{ij} \leq 1$ 。

3.2 构造模糊相似关系

相似系数是描述变量之间相似程度的一个量, 即 $\tilde{R}$ 中的元素 r_{ij} 。用标准化的数据标定计算各个样本间的相似系数的方法很多, 一般有相似系数法(数量积、夹角余弦、相关系数、指数相似系数)、距离法(Hamming 距离、Chebyshev 距离、Euclid 距离、Minkowski 距离等)、贴近度法(最大最小法、几何平均最小法等)和主观评定法^[9]。但在衡量土性参数的相似度中, 目前还没有十分确切的原则或理论作为参考。在这些方法中, 夹角余弦有自己独特的优势: 它是由相似形演变而来的, 它把土工样本当作一个忽略绝对长度的向量, 当两个样本向量的方向相近时, 夹角余弦值较大, 相似度越高; 反之, 则较低。特殊情况下, 如果两个样本向量正交, 则夹角余弦为 0, 表示两样本完全不相似。在分析过程中, 我们曾编制程序对各种方法进行了试算比较, 结果也说明采用夹角余弦法时聚类结果比较理想, 所以在后续编程计算中, 均采用了夹角余弦来计算土工试样的相似系数(见公式(6))。

$$r_{ij} = \frac{\sum_{k=1}^m x_{ik} x_{jk}}{\sqrt{\left(\sum_{k=1}^m x_{ik}^2\right)} \sqrt{\left(\sum_{k=1}^m x_{jk}^2\right)}} \quad (6)$$

求得各样本的相似系数 r_{ij} 后, 模糊相似矩阵 $\tilde{R}$ 便确定了。

3.3 确定模糊等价关系

正如前面所述, 矩阵 $\tilde{R}$ 一般不具有模糊等价关系, 仅仅具有模糊相似关系, 需要按照传递闭包原理进行改造: 逐步计算 $\tilde{R}^2$, $\tilde{R}^4$, ..., 直至 $\tilde{R}^k = \tilde{R}^{2k} = R^*$ 。此时, R^* 满足 2.2 节中的各项条件要求, 得到了模糊等价矩阵。

3.4 聚类

选取不同的置信水平值 λ , 前面得到的模糊等

价矩阵 R^* 就逐渐归并得到不同的聚类结果。当 λ 由 1 下降到 0 时, 聚类的要求由细变粗, 逐层归并, 形成一个动态的聚类图。

因为模糊聚类各参数权重的确定并没有贴切合理的方法, 而人为的给一些权重也并不能很好地代表自然的规律。考虑到勘察规范对地层划分的要求, 本文创新性的在聚类中增加了两项模糊聚类检验:

①在场区地层归并时, 岩性差别不可太大, 一般情况是塑性指数指标相邻的地层才进行合并。②按照岩土工程勘察规范^[10-11], 进行土层归并后形成的各“大层”的土性指标应满足变异系数的要求, 因此模糊聚类后, 还做了变异系数的检验, 以确定合理的置信水平 λ , 最后确认地层合并的合理性。Ingles 和 Meyerhof 等^[12-13]都对同类土的各项指标的变异性进行过系统的研究, 并提出了同一层土的各物理力学指标变异系数的最大允许值区间。北京地方勘察规范借鉴他们的研究成果, 规定了北京地区同一层土的各类指标的变异系数上限。如压缩模量 E_s 的变异系数不大于 0.35, 孔隙比 e 的变异系数不大于 0.1 等等。多年来的实践证明, 这些规定符合北京市土工参数的实际规律。因此, 模糊聚类也理应进行变异系数检验。

程序化实现的总流程如图 1 所示。

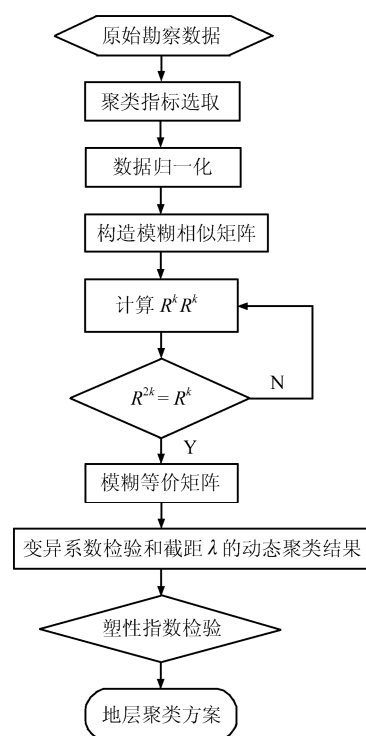


图 1 程序流程图

Fig.1 Process flow chart

4 实例分析与聚类合理性的检验

4.1 聚类分析过程

利用地层模糊聚类的程序对北京泛海国际居住区 2 期工程勘察资料^[14]的上部黏性土层进行了分析。该场区位于永定河冲洪积扇的中下部, 性质相似的黏性土层较多, 且考虑到北京规范对工程性状规定的主要依据是沉降控制原理, 所以在本例中采用了塑性指数、孔隙比和压缩模量这 3 个土性指标。分析的目标黏性土层大致位于杂填土和第 1 层细砂之间, 深度范围约为 1.5~7 m。各项原始数据指标和按规范^[10]统计后的资料见表 1。

表 1 勘察取样的试验参数和各层的基本统计量
Table 1 The test parameters of soil samples and basic statistics of strata

按规范分类定名	塑性指数 I_p	孔隙比 e	压缩模量 E_s/MPa	各土性指标平均值		
				$\bar{I}_p$	$\bar{e}$	$\bar{E}_s$
黏土 $I_p > 17$	20.2	0.97	4.6	18.88	0.93	4.98
	19.2	1.01	3.4			
	17.8	0.82	5.1			
	17.3	0.87	4.7			
	19.9	0.96	7.1			
重粉质黏土 $14 < I_p \leq 17$	14.5	0.69	8.0	14.70	0.71	7.88
	15.0	0.7	7.5			
	15.3	0.74	6.7			
	14.0	0.71	9.3			
	12.1	0.73	7.6			
粉质黏土 $10 < I_p \leq 14$	10.7	0.68	6.8	11.94	0.65	8.50
	12.8	0.62	8.0			
	10.5	0.63	10.2			
	11.1	0.56	6.4			
	13.5	0.64	9.9			
	12.6	0.76	8.7			
	12.2	0.61	10.4			
黏质粉土 $7 < I_p \leq 10$	8.9	0.64	12.9	8.27	0.57	15.8
	8.1	0.49	16.7			
	9.3	0.52	21.3			
	8.4	0.55	18.8			
	7.5	0.61	15.1			
	7.4	0.56	10.1			
砂质粉土 $3 < I_p \leq 7$	6.9	0.58	21.4	5.96	0.56	19.5
	5.5	0.62	21.5			
	6.0	0.51	19.3			
	6.4	0.53	16.8			
	5.0	0.54	18.3			

为了说明问题, 例子阐述的是一个比较简单的条件, 分类指标也只选用 3 个, 但可看出分类的原理应该是适用于更加复杂的条件中的。利用地层模糊聚类的程序计算得到的模糊等价矩阵为

$$\tilde{R} = \begin{bmatrix} 1 & 0.946 & 0.946 & 0.597 & 0.597 \\ 0.946 & 1 & 0.983 & 0.597 & 0.597 \\ 0.946 & 0.983 & 1 & 0.597 & 0.597 \\ 0.597 & 0.597 & 0.597 & 1 & 0.974 \\ 0.597 & 0.597 & 0.597 & 0.974 & 1 \end{bmatrix}$$

从式 (6) 可知, 模糊等价矩阵的各元素值反映了各土层指标之间的相似程度, 取不同的置信水平截距 λ 就可以聚类看出各土层的“相似性”。模糊聚类结果见图 2。

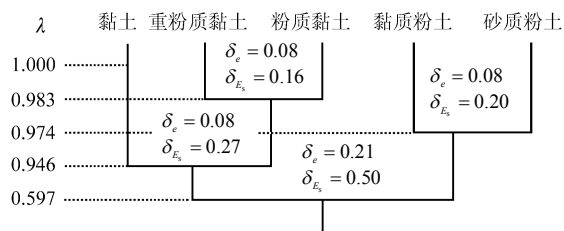


图 2 动态聚类图

Fig.2 The graph of dynamical clustering

4.2 聚类合理性的检验方法

对图 2 进行分析可以得出, 当 λ 逐步缩小时土层相似程度越来越差, 相应地, 变异系数也随之增加, 如将这些土层合并, 将不能满足规范要求。从参数的变异系数 δ 与置信水平 λ 之间的关系(图 3)可以看出: 为了满足规范中要求的变异系数, 置信水平取值是很高的。在本例中, $\lambda = 0.974$, 对重粉质黏土和粉质黏土进行合并的同时, 也可对黏质粉土和砂质粉土地层进行合并。合并后塑性指数变化范围合理, 主要参数压缩模量的变异系数 δ_{E_s} 和孔隙比的变异系数 δ_e 均在规范要求的范围之内($\delta_{E_s} = 0.08$, $\delta_e = 0.20$), 而且经过合并的地层显示出很好的沉积规律性。而如果置信水平 λ 在 0.946 时, 孔隙比大大超标, 反映了土层在某些本质意义上的差别。聚类前后地层对比结果见图 4。通过这样处理, 综合考虑了各参数, 为勘察分层及地层合并提供了参考依据。

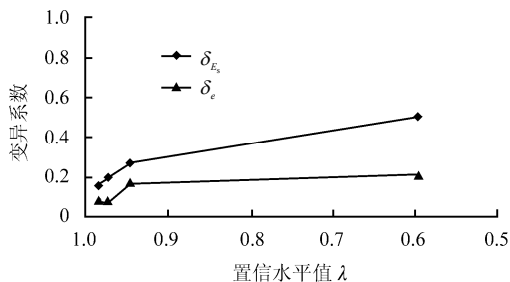


图 3 变异系数与置信水平值 λ 关系图

Fig.3 The graph of relationship between coefficient of variance and confidence measure

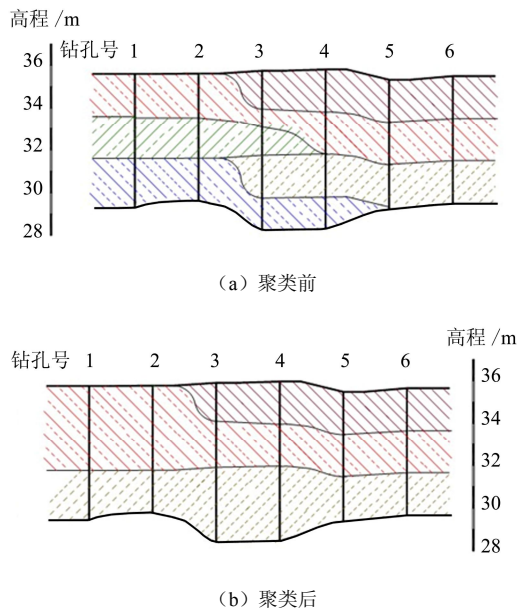


图4 聚类对比示意图

Fig.4 The sketch map of contrast in clustering

5 结 论

(1) 引入模糊聚类思想方法, 提出了一种根据基本物理力学指标进行土层归并分类的方法, 具有普遍意义。本项研究也提供了改变传统上仅仅依靠工程师们的个人经验进行该项工作的可能性, 意义明确, 计算简捷。

(2) 提出了用参数的变异系数来检验模糊聚类结果的方法, 使模糊聚类的结果与统计参数能够对应起来, 并验证了结果满足规范的规定, 从而也证明大量地层归并的合理性。

(3) 合理地合并具有相似性的地层, 无疑增加了每一大层土样本的容量, 在变异系数基本保持不变的前提下, 这样做在工程参数统计意义上的优点将在另文讨论。

参 考 文 献

- [1] 张在明. 岩土工程的工作方法[C]//第二届全国岩土与工程学术大会论文集(上册). 北京: 科学出版社, 2006.
- [2] BAECHER G B, CHRISTIAN J T. Geotechnical Reliability: Playing Cards with the Universe[C]// Proceedings of 12th Pan-American Conference on Soil Mechanics and Geotechnical Engineering. Cambridge: Massachusetts, America, 2003.
- [3] BAECHER G B, CHRISTIAN J T. Reliability and statistics in geotechnical engineering [M]. West Sussex: John Wiley and Sons Ltd., 2003.
- [4] 何广讷, 朱范宇, 张业民. 场地砂土地震液化的模糊聚类分析[J]. 地震工程与工程振动, 1989, 9(4): 83—91.
- [5] HE Guang-na, ZHU Fan-yu, ZHANG Ye-min. Fuzzy clustering analysis for predicting seismic liquefaction potential of sandy deposits[J]. **Earthquake Engineering and Engineering Vibration**, 1989, 9(4): 83—91.
- [6] 樊哲超, 陈建生, 董海洲, 等. 模糊聚类方法研究堤坝渗漏[J]. 岩土力学, 2006, 27(7): 1214—1218.
- [7] FAN Zhe-chao, CHEN Jian-sheng, DONG Hai-zhou, et al. Applying fuzzy clustering method to study leakage of dyke[J]. **Rock and Soil Mechanics**, 2006, 27(7): 1214—1218.
- [8] 曲永玲, 陈新民. 聚类分析法在土层划分中的应用[J]. 南京工业大学学报, 2005, 27(5): 85—88.
- [9] QU Yong-ling, CHEN Xin-min. Application of cluster analysis method in soil division[J]. **Journal of Nanjing University of Technology**, 2005, 27(5): 85—88.
- [10] 张鹏, 张在明, 杨宇友. 工程勘察数据的传递控制与可视化[J]. 岩土工程技术, 2008, 22(2): 55—59.
- [11] ZHANG Peng, ZHANG Zai-ming, YANG Yu-you. Transfer control and visualization of geotechnical engineering investigation data[J]. **Geotechnical Engineering Technique**, 2008, 22(2): 55—59.
- [12] 张尧庭, 方开泰. 多元统计分析引论[M]. 北京: 科学出版社, 1982.
- [13] 陈水利, 李敬功, 王向公. 模糊集理论及其应用[M]. 北京: 科学出版社, 2005.
- [14] 北京市勘察院, 北京市建筑设计研究院. DBJ01—501—92 北京地区建筑地基基础勘察设计规范[S]. 北京: [出版社不详], 1992.
- [15] 中华人民共和国建设部. GB50021—2001 岩土工程勘察规范[S]. 北京: 中国建筑工业出版社, 2002.
- [16] INGLES O G. Statistics and probability: the engineer-client interaction problem[C]//Reliability and Risk Analysis in Civil Engineering, Proceedings of ICASP5, the Fifth International Conference on Applications of Statistics and Probability in Soil and Structural Engineering. Vancouver: Inst for Risk Research, Univ. of Waterloo, 1987.
- [17] MEYERHOF G G. Partial and total safety factors[C]// Proceedings of Special Session 6 of 9th International Conference on Soil Mechanics and Foundations Engineering. Tokyo: JSSMFE, 1977.
- [18] 北京市勘察设计研究院. 泛海国际居住区二期工程岩土工程勘察报告[R]. 北京: 北京市勘察设计研究院有限公司, 2007.