

基于大数据融合技术的稀疏测量场地 三维概率化重构

杨智勇^{1,2}, 丁宇超¹, 冷振东³, 刘志军^{4,5}, 黎学优^{1,2}

(1. 中山大学 土木工程学院, 广东 珠海 519000; 2. 中山大学 南方海洋科学与工程广东省实验室(珠海), 广东 珠海 519082;
3. 中国葛洲坝集团易普力股份有限公司, 重庆 401121; 4. 中交四航工程研究院有限公司, 广东 广州 510230;
5. 中交集团交通基础工程环保与安全重点实验室, 广东 广州 510230)

摘 要: 受工程预算影响, 特定场地的勘探数据一般十分有限, 有限的勘测数据往往会导致估计的土体参数存在较大的统计不确定性。与此同时, 随着岩土勘探设备与信息化技术的发展, 岩土体参数数据库却日益庞大, 如何有效地利用岩土体大数据去量化三维稀疏测量场地土体参数不确定性是一个富有挑战性的问题。为此, 提出了基于大数据融合技术的三维稀疏测量场地概率化重构方法。首先, 根据吉布斯抽样技术估计有限勘探数据条件下三维场地土体参数的概率分布模型, 该模型同时考虑不同钻孔勘测数据在水平和垂直方向的空间自相关性。其次, 同样采用吉布斯抽样构建基于数据库数据的土体参数概率分布模型。在此基础上, 通过变种贝叶斯理论融合三维场地土体参数概率模型和基于数据库数据的土体参数概率模型, 推导了融合数据库信息后的稀疏测量场地土体参数变种概率分布模型, 通过变种概率分布模型模拟不完备钻孔数据, 得到了具有格栅结构的完备钻孔数据, 利用 Kronecker 内积分解大尺度自相关矩阵, 实现高效的三维场地概率化重构方法。最后, 以一个模拟虚拟场地和美国德克萨斯州某勘探场地为例, 阐明了所提方法的有效性。结果表明: 所提方法能有效地融合岩土体大数据, 融合岩土体大数据能够显著地降低三维稀疏测量场地土体参数的不确定性, 为稀疏测量场地的三维概率化重构提供一种有效分析工具。

关 键 词: 场地勘测; 不确定性; 岩土体大数据; 吉布斯抽样; 数据融合

中图分类号: TU412

文献标识码: A

文章编号: 1000-7598(2025)06-1881-16

Three-dimensional probabilistic reconstruction of sparse measurement site based on a big data assimilation technique

YANG Zhi-yong^{1,2}, DING Yu-chao¹, LENG Zhen-dong³, LIU Zhi-jun^{4,5}, LI Xue-you^{1,2}

(1. School of Civil Engineering, Sun Yat-Sen University, Zhuhai, Guangdong 519000, China; 2. Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-Sen University, Zhuhai, Guangdong 519082, China; 3. China Gezhouba Group Explosive Co., Ltd., Chongqing 401121, China;
4. CCCC Fourth Harbor Engineering Institute Co., Ltd., Guangzhou, Guangdong 510230, China; 5. CCCC Key Laboratory of Environmental Protection & Safety in Foundation Engineering of Transportation, Guangzhou, Guangdong 510230, China)

Abstract: Site investigation data are often sparse due to limited project budgets. Using sparse site investigation data to quantify soil parameter inevitably results in significant statistical uncertainties. Meanwhile, the soil parameter database is expanding rapidly due to advancements in geotechnical investigation equipment and information technology. Effectively utilizing the big data in the soil parameter database to quantify soil parameter uncertainties at three-dimensional sparse measurement sites remains a challenging problem. This study proposes a big data assimilation technique to effectively quantify soil parameter uncertainties at three-dimensional sparse measurement sites. The proposed method first constructs a probability distribution model of soil parameters using limited sparse site investigation data. This model considers the autocorrelation of soil data in both horizontal and vertical directions across different soundings. Secondly, a probability distribution model based on the soil parameter database is constructed using Gibbs sampler. Based on the two probability distribution models, the hybrid Bayesian theory is employed to integrate the

收稿日期: 2024-08-30

录用日期: 2024-11-13

基金项目: 国家自然科学基金(No.52109144); 广东省基础与应用基础研究基金(No.2025A151011207); 武汉大学水资源工程与调度全国重点实验室开放研究基金项目(No.2023SGG04)。

This work was supported by the National Natural Science Foundation of China (52109144), the Guangdong Basic and Applied Basic Research Foundation (2025A151011207) and the Open Research Fund of State Key Laboratory of Water Resources Engineering and Management (Wuhan University) (2023SGG04).

第一作者简介: 杨智勇, 男, 1989 年生, 博士, 副教授, 主要从事岩土工程可靠性与风险分析方面的研究。E-mail: yangzhy85@mail.sysu.edu.cn

statistical information of the big data into the site-specific data. A hybrid probability model of soil parameters for the sparse measurement site is derived after assimilating the big data. The incomplete borehole data are simulated to be complete data that satisfy the lattice structure using the hybrid probability density function. The proposed method uses Kronecker product to decompose the large autocorrelation matrix and achieves the efficient reconstruction of the 3D site. Finally, the proposed method is demonstrated using a simulated virtual site and a site in Texas, USA. Results indicate that the proposed method effectively utilizes geotechnical big data. Assimilating geotechnical big data significantly reduces soil parameter uncertainties at three-dimensional sparse measurement sites. The proposed method provides a valuable tool for quantifying soil parameter uncertainties at sparse measurement sites.

Keywords: site investigation; uncertainty; geotechnical big data; Gibbs sampler; data assimilation

1 引言

岩土体参数的物理力学性质对岩土结构物的稳定性具有重要影响。岩土结构物的设计与施工首先必须探明特定场地土体物理力学性质参数的大小,然而受工程预算、勘探时间及施工环境等因素的影响,特定场地勘察数据往往十分有限,有限勘察数据仅能揭示土体参数在三维空间上有限几个勘察位置处的物理力学性质,难以明晰未勘察位置处的土体力学参数性质^[1-3]。为了估计未勘察位置处的土体参数,有必要采用概率分析方法,建立考虑土体参数不确定性的三维场地概率化重构模型。在岩土工程可靠度分析领域,三维场地概率化模型的构建也是对该场地岩土结构物进行三维可靠性分析与风险评估的基础。建立三维场地概率化模型一般包括两方面的内容:①估计土体参数的统计量,如分布类型、均值、相关系数、标准差及波动范围等;②根据得到的土体参数统计量与勘察数据重构三维场地,即模拟大尺度三维约束随机场。

在估计土体参数统计量方面,国内外学者开展了大量有益的研究。如黄天朗等^[4]提出了基于正交多项式推断方法的小样本土体参数概率密度函数估计方法;陈旺旺等^[5]提出了基于最大熵估计的抗剪强度参数概率密度函数估计方法;田密等^[6]提出了基于间接测量数据的土体稀疏测量参数不确定性量

化方法;李典庆等^[7]提出了基于多维 Copula 的岩土体参数概率模型构建方法,利用直接测量参数与设计参数之间的相关性来降低设计参数的不确定性。这些方法多针对单一土体参数的不确定性量化,缺乏对多土体参数联合概率密度函数的深入研究,且在处理不完备数据方面存在一定局限性。为解决基于多源勘测数据的土体参数不确定性量化问题,Ching 等^[8]提出了基于 MUSIC (multivariate, uncertain, sparse, incomplete and sometimes corrupted) 数据(假设勘测数据在水平和垂直方向独立)的土体参数概率密度函数估计方法;Ching 等^[9]提出了进一步提出了基于 MUSIC-X (X 代表空间变异性)数据(假设勘测数据在水平方向独立)的土体参数概率密度函数估计方法。

尽管上述 MUSIC 或 MUSIC-X 方法能够有效地量化多土体参数的不确定性,然而当勘探数据特别稀疏时,这些方法得到的土体参数往往仍具有非常大的统计不确定性^[10-11],严重降低了勘察数据在岩土结构物可靠性评估与风险量化方面的价值^[12-13]。与特定场地勘察数据稀疏的情况截然不同,随着岩土勘探设备与信息化技术的发展,岩土体参数的数据库却日益庞大。如表 1 列出了 TC304 总结的部分岩土体参数数据库^[14](表中符号含义详见参考文献^[14]),以 CLAY/10/7490 数据库为例,该数据库共收集了 7 490 条来自全球各地黏性土的 10 个土体

表 1 TC304 总结的部分岩土体参数数据库^[14]
Table 1 Some databases of geotechnical parameters summarized by TC304^[14]

数据库 (土体性质/土体参数个数/数据记录个数)	收集参数	场地个数/文献篇数	覆盖范围
CLAY/5/345	$I_L, S_u, S_u^{ic}, S_v, S_p$	37	全球范围
CLAY/6/535	$S_u/S_v, OCR, q_{t1}, q_{tu}, (u_2-u_0)/S_v, B_q$	40	全球范围
CLAY/10/7490	$w_L, I_P, I_L, S_v/P_a, S_p/P_a, S_u/S_v, S_t, q_{t1}, q_{tu}, B_q$	251	全球范围
F-CLAY/7/216	$w_L, w_P, w, S_v, S_p, S_u, S_t$	24	芬兰
S-CLAY/7/168	$w_L, w_P, w, S_v, S_p, S_u, S_t$	22	挪威、瑞典
J-CLAY/5/124	$M_r, q_c, f_s, w, \gamma_d$	16	中国江苏省
ROCK/9/4069	$n, g, R_L, S_{lt}, S_{bt}, I_{s50}, V_p, S_{ci}, E_i$	184	全球范围
SAND/7/2794	$D_{50}, C_u, D_r, s'_v/P_a, \phi', q_{t1}, (N_1)_{60}$	176	全球范围
SH-CLAY/11/4051	$w_L, I_P, I_L, e, K_0, s'_v/P_a, S_u(UCST)/s'_v, S_u(UCST), S_u(VST)/s'_v, S_u(VST), p_s/s'_v$	51	中国上海市
ROCK/10/4025	$n, \gamma, R_L, BPI, \sigma_{bt}, I_{s50}, V_p, \sigma_{ci}, E_i, m_i$	95	全球范围

参数的数据, 初具大数据的特点。对于稀疏测量场地, 仅依赖该场地的勘探数据去量化土体参数不确定性不仅会导致土体参数具有较大的统计不确定性, 也会造成巨大的信息浪费。有必要充分挖掘数据库中的大数据信息来量化特定场地的土体参数不确定性。

近年来, 针对岩土体数据库信息挖掘的研究得到越来越多的重视。如 Ching 等^[8]提出了基于变种贝叶斯理论的数据库信息挖掘方法, 表明数据库信息在稀疏测量场地土体参数不确定性过程中具有重要作用; Han 等^[11]结合 4 个场地的无侧限抗压强度数据, 通过贝叶斯马尔科夫链蒙特卡洛抽样方法去估计某一场地无侧限抗压强度的后验样本数据; Ching 等^[15]提出了基于层次贝叶斯理论的数据融合方法, 在数据融合过程中未考虑特定场地的土体参数空间变异性; Ching 等^[16]进一步提出了基于层次贝叶斯理论的场地土体参数联合概率密度函数估计方法, 并考虑了特定场地的空间自相关性, 该文同时对基于变种贝叶斯的三维场地不确定性量化进行了简单介绍; Sharma 等^[17]提出了基于层级贝叶斯理论的相似性场地识别方法; 张德等^[18]采用层次贝叶斯分析法研究了场地有限数据条件下的土体不排水强度概率分布估计问题, 研究发现土体强度参数数据较少时, 通过层次贝叶斯融合利用多场地数据能够有效降低估计参数的不确定性; Tabarrokhi 等^[19]通过层次贝叶斯理论分析了多个开挖场地的挡墙变形曲线数据, 并提出了融合数据库信息的特定场地开挖挡墙变形预测模型。上述方法在挖掘数据库信息方面取得了许多有益的成果, 但仍存在以下几点不足: ①层次贝叶斯理论较为复杂, 在处理多参数数据库信息方面计算量较大, 且要求数据库数据具有明确的场地标签, 对于缺乏场地标签的区域数据库, 层次贝叶斯理论难以应用; ②变种贝叶斯方法虽较为简单, 但在构建特定场地土体参数概率模型时未能充分考虑土体参数的空间自相关性; ③上述研究多局限于利用大数据信息来估计土体参数的统计量, 未涉及对三维场地的概率化重构。

为解决上述问题, 本文提出了一种基于大数据融合技术的稀疏测量场地三维概率化重构方法, 通过吉布斯抽样分别估计特定场地考虑土体参数三维空间变异性条件下的土体参数联合概率密度函数和数据库土体参数的联合概率密度函数, 采用变种贝叶斯理论融合特定场地土体参数概率密度函数与数据库土体参数概率密度函数, 根据融合后变种分布模拟不完备钻孔数据, 在此基础上生成三维场地概

率化重构模型。本文所提方法能有效地融合岩土体大数据, 为稀疏测量场地的土体参数不确定性量化提供一种有效分析工具。融合岩土体大数据重构三维场地模型能够显著地降低三维稀疏测量场地的不确定性。

2 大数据融合技术

本章提出一种大数据分析技术, 通过融合岩土数据库的大数据来量化特定稀疏测量场地的土体参数不确定性。为便于理论推导, 本文假设土体参数勘察数据均服从正态分布。在实际工程中, 岩土数据库中的数据和特定场地勘探得到的土体参数数据实际上并不一定严格服从正态分布, 对于非正态分布的岩土体勘察数据, 可通过 Nataf 变换^[20]、三参数正态变换^[21]、Box-Cox 变换^[22]与 Johnson 分布变换^[23]等先转换为正态分布, 分析得到结果后再通过逆变换转换数据为原始数据。相比于传统马尔科夫链抽样方法, 吉布斯抽样的抽样效率高 (接受率为 100%), 且适用于高维分布, 因此, 本文首先通过吉布斯抽样估计三维场地土体参数概率密度函数和数据库数据概率密度函数, 再通过变种贝叶斯构建考虑数据库信息的特定场地概率密度函数, 在此基础上, 通过完备钻孔数据生成大尺度三维约束随机场, 实现对三维场地的概率化重构。下面首先介绍基于吉布斯抽样的三维场地概率密度函数估计方法。

2.1 三维场地土体参数不确定性量化

杨智勇等^[10]提出了一种基于吉布斯抽样技术的三维场地概率密度函数估计方法, 该方法可以有效地考虑土体参数的三维空间自相关性, 下面简要介绍该方法。

设稀疏测量场地包含 n_h 个勘探钻孔, 每个钻孔可能在不同的深度处取样, 将勘探钻孔沿深度方向离散为 n_v 个深度, n_v 个深度应包含所有钻孔的勘探深度, 设该场地测量了 m 个土体参数的数据。设 $\mathbf{x}_i$ ($i = (k-1)n_v + j$, $k=1, 2, \dots, n_h$, $j=1, 2, \dots, n_v$) 代表第 k 个钻孔位置处第 j 个深度的土体数据, 则 $\mathbf{x}_i$ 是一个包含 m 个元素的列向量。 $\mathbf{X} = [\mathbf{x}_1^T \mathbf{x}_2^T \cdots \mathbf{x}_n^T]^T$ 代表该场地在 n_h 个钻孔的 n_v 个深度处的土体数据, $\mathbf{X}$ 通过顺序垂直堆叠 $\mathbf{x}_i$ ($i=1, 2, \dots, n$, $n=n_h \times n_v$) 形成, 是一个尺寸为 $(m \times n_h) \times 1$ 的列向量。注意到某一深度处土体参数的数据可能在一个钻孔位置被观测, 然而在另一钻孔位置却未被观测, 即使某一钻孔位置某个深度的土体参数被观测, 也可能并非 m 个土体参数均被观测。换言之, 上述 $\mathbf{X}$ 既包含了观测到的数据 $\mathbf{X}^o$, 同时也包含了未观测到的数据 $\mathbf{X}^u$ 。场地不

确定量化的首要目标是根据给定的勘测数据 $\mathbf{X}^0$ 去估计土体参数的均值 $\underline{\mu}$ 、协方差矩阵 $\mathbf{C}_s$ 以及 $\mathbf{X}^u$, 即求解 $f(\mathbf{X}^u, \underline{\mu}, \mathbf{C}_s | \mathbf{X}^0)$ 。杨智勇等^[10]假设土体参数空间自相关性在水平和垂直方向可分离, 提出采用如下吉布斯抽样估计 $\mathbf{X}^u, \underline{\mu}, \mathbf{C}_s$ 的条件样本:

(1) 采用随机值初始化 $\underline{\mu}, \mathbf{C}_s, \mathbf{X}^u$ 和 $\underline{a}$, 其中, $\underline{a}$

$$f(\underline{\mu}_s | \mathbf{X}, \mathbf{C}_s, \underline{a}) \sim$$

$$N \left\{ \left[\mathbf{C}_0^{-1} + \left(\mathbf{1}_{n_h}^T \mathbf{R}_h^{-1} \mathbf{1}_{n_h} \right) \left(\mathbf{1}_{n_v}^T \mathbf{R}_v^{-1} \mathbf{1}_{n_v} \right) \mathbf{C}_s^{-1} \right]^{-1} \left\{ \left[\left(\mathbf{1}_{n_h}^T \mathbf{R}_h^{-1} \right) \otimes \left(\mathbf{1}_{n_v}^T \mathbf{R}_v^{-1} \right) \otimes \mathbf{C}_s^{-1} \right] \mathbf{X} + \mathbf{C}_0^{-1} \underline{\mu}_0 \right\}, \left[\mathbf{C}_0^{-1} + \left(\mathbf{1}_{n_h}^T \mathbf{R}_h^{-1} \mathbf{1}_{n_h} \right) \left(\mathbf{1}_{n_v}^T \mathbf{R}_v^{-1} \mathbf{1}_{n_v} \right) \mathbf{C}_s^{-1} \right]^{-1} \right\} \quad (1)$$

式中: $\mathbf{1}_{n_h}$ 和 $\mathbf{1}_{n_v}$ 分别为分量全为 1 的 n_h 维列向量和 n_v 维列向量; $\mathbf{R}_h$ 和 $\mathbf{R}_v$ 分别为土体参数在水平方向和垂直方向的自相关矩阵; $\underline{\mu}_0, \mathbf{C}_0$ 均为 $\underline{\mu}$ 的共轭先验分布函数 $N(\underline{\mu}_0, \mathbf{C}_0)$ 的参数, $N(\underline{\mu}_0, \mathbf{C}_0)$ 代表均值为 $\underline{\mu}_0$ 、协方差为 $\mathbf{C}_0$ 的多变量正态分布。为确保 $\underline{\mu}$ 的共轭先验分布为无信息分布, 计算过程中取 $\underline{\mu}_0 = \underline{0}$, $\mathbf{C}_0$ 为对角线元素等于 10^4 的对角矩阵。

(3) 根据 $\mathbf{C}_s$ 的条件概率密度函数抽取 $\mathbf{C}_s$ 样本 (MATLAB 函数 `iwishrnd`):

$$f(\mathbf{C}_s | \mathbf{X}, \underline{\mu}_s, \underline{a}) \sim \text{IW} \{ \mathbf{S} + \mathbf{\Sigma}, n + m + 1 \} \quad (2)$$

式中: $\mathbf{S} = [\text{mat}(\mathbf{X}) - \underline{\mu}_s \mathbf{1}_n^T] (L_h^{-1} \otimes L_v^{-1})^T (L_h^{-1} \otimes L_v^{-1}) \cdot [\text{mat}(\mathbf{X}) - \underline{\mu}_s \mathbf{1}_n^T]^T$, L_h 和 L_v 分别为 $\mathbf{R}_h$ 和 $\mathbf{R}_v$ 的乔列斯基分解下三角矩阵, $\text{mat}(\mathbf{X}) = [\mathbf{x}_1 \mathbf{x}_2 \mathbf{x}_3 \cdots \mathbf{x}_n]$ 代表水平堆叠列向量 $\mathbf{x}_i$; $\mathbf{\Sigma}$ 为对角元素, 是 $\underline{a}$ 的对角矩阵; $\mathbf{1}_n$ 代表包含 n 个元素为 1 的列向量。

(4) 根据 a_i 的条件概率密度函数 $f(a_i | \mathbf{X}, \underline{\mu}_s, \mathbf{C}_s, \underline{a}_{-i})$ 抽取 a_i 样本 (MATLAB 函数 `gamrnd`):

$$f(a_i | \mathbf{X}, \underline{\mu}_s, \mathbf{C}_s) \sim \text{IG} \left(a_i; \alpha + \frac{m+1}{2}, \beta + 2\mathbf{C}_{s,ii}^{-1} \right) \quad (3)$$

式中: $\mathbf{C}_{s,ii}^{-1}$ 为矩阵 $\mathbf{C}_s^{-1}$ 中第 i 行第 i 列个元素; $\underline{a}_{-i}$ 为 $\underline{a}$ 中剔除 a_i 后的元素向量; 为确保 $\text{IW}(\mathbf{\Sigma}, n_h n_v + m + 1)$ 是关于 $\mathbf{C}_s$ 的无信息先验分布, 设 $\alpha = 0.5$, $\beta = 10^{-4}$ 。

(5) 根据 $\mathbf{X}^u$ 的条件概率密度函数 $f(\mathbf{X}^u | \mathbf{X}^0, \underline{\mu}, \mathbf{C}_s, \underline{a})$ 抽取 $\mathbf{X}^u$ 样本。为了避免模拟 $\mathbf{X}^u$ 中可能遇到的超大矩阵运算问题, 杨智勇等^[10, 24-25]建议采用基于顺序钻孔吉布斯抽样的方法模拟 $\mathbf{X}^u$, 第 k 个钻孔中的未观测数据 $k = 1, 2, \dots, n_h$ 可通过下式模拟 (MATLAB 函数 `mvnrnd`):

$$\left. \begin{aligned} E(\underline{\mathbf{X}}_k^u | \underline{\mathbf{X}}_k^0, \mathbf{X}_{-k}) &= E(\underline{\mathbf{X}}_k^u | \mathbf{X}_{-k}) + \text{Cov}(\underline{\mathbf{X}}_k^u, \underline{\mathbf{X}}_k^0 | \mathbf{X}_{-k}) \cdot \\ &\quad \text{Var}(\underline{\mathbf{X}}_k^0 | \mathbf{X}_{-k})^{-1} [\underline{\mathbf{X}}_k^0 - E(\underline{\mathbf{X}}_k^0 | \mathbf{X}_{-k})] \\ \text{Var}(\underline{\mathbf{X}}_k^u | \underline{\mathbf{X}}_k^0, \mathbf{X}_{-k}) &= \text{Var}(\underline{\mathbf{X}}_k^u | \mathbf{X}_{-k}) - \text{Cov}(\underline{\mathbf{X}}_k^u, \underline{\mathbf{X}}_k^0 | \mathbf{X}_{-k}) \cdot \\ &\quad \text{Var}(\underline{\mathbf{X}}_k^0 | \mathbf{X}_{-k})^{-1} \text{Cov}(\underline{\mathbf{X}}_k^0, \underline{\mathbf{X}}_k^u | \mathbf{X}_{-k})^T \end{aligned} \right\} \quad (4)$$

为 $\mathbf{C}_s$ 共轭先验分布 $\text{IW}(\mathbf{\Sigma}, n+m+1)$ 的尺度矩阵 $\mathbf{\Sigma}$ 的对角线元素, 为一个 $m \times 1$ 列向量; $\text{IW}(\mathbf{\Sigma}, n+m+1)$ 代表自由度系数为 $n+m+1$ 、尺度矩阵为 $\mathbf{\Sigma}$ 的逆 Wishart 分布。

(2) 根据 $\underline{\mu}_s$ 的条件概率密度函数抽取 $\underline{\mu}_s$ 样本 (MATLAB 函数 `mvnrnd`):

式中: $\underline{\mathbf{X}}_k^u$ 和 $\underline{\mathbf{X}}_k^0$ 分别为第 k 个钻孔未观测到的数据和观测到的数据; $\mathbf{X}_{-k}$ 为 $\mathbf{X}$ 中剔除第 k 个钻孔后的其他钻孔的数据; $E(\cdot)$ 代表均值运算; $\text{Var}(\cdot)$ 代表方差运算; $\text{Cov}(\cdot)$ 代表协方差运算。

(6) 步骤 (2) ~ (5) 循环 T 次得到 T 个 $(\mathbf{X}^u, \underline{\mu}_s, \mathbf{C}_s, \underline{a})$ 样本。

注意到吉布斯抽样初始阶段得到的样本可能处于非平稳状态且相邻循环样本间可能存在一定相关性, 根据吉布斯抽样的样本链视觉判断样本是否处于平稳状态^[8, 26], 丢弃初始的 t_b 个非平稳样本, 然后在剩下的 $(T-t_b)$ 个样本中每间隔 Δt 循环步挑选一个样本得到服从目标分布 $f(\underline{\mu}, \mathbf{C}_s, \underline{a}, \mathbf{X}^u | \mathbf{X}^0)$ 的 $(T-t_b)/\Delta t$ 个 $(\mathbf{X}^u, \underline{\mu}, \mathbf{C}_s, \underline{a})$ 样本。 $(\underline{\mu}, \mathbf{C}_s)$ 样本反映了量化土体参数过程中存在的统计不确定性, 当观测到的场地数据 $\mathbf{X}^0$ 较稀疏时, $\mathbf{X}^0$ 携带的关于土体参数均值和协方差的信息非常有限, 因此, 估计的土体参数均值和协方差会具有很大不确定性, 具体表现为 $(\underline{\mu}, \mathbf{C}_s)$ 样本具有很大离散性; 相反地, 当观察到的数据较多时, $(\underline{\mu}, \mathbf{C}_s)$ 离散性较小。根据全概率理论, 在距离勘探钻孔无限远处的某一深度处土体参数的概率密度函数可表示为

$$\begin{aligned} f(\mathbf{x} | \mathbf{X}^0) &= \int f(\mathbf{x} | \underline{\mu}_s, \mathbf{C}_s, \mathbf{X}^0) \cdot f(\underline{\mu}_s, \mathbf{C}_s | \mathbf{X}^0) \cdot \\ &\quad d\underline{\mu}_s d\mathbf{C}_s \approx \frac{1}{T-t_b} \left[\sum_{t=t_b+1}^T f(\mathbf{x} | \underline{\mu}_s, \mathbf{C}_s, \mathbf{X}^0) \right] \quad (5) \end{aligned}$$

2.2 数据库土体参数概率模型估计

本节主要介绍基于吉布斯抽样的岩土体数据库土体参数概率模型的构建方法。岩土数据库中一般包含全球范围内几百甚至上千个场地的勘测数据, 具有大数据的特点。尽管数据库中存在大量的勘测数据, 但由于每个场地测量的土体参数个数并不相同, 且同一场地某个位置处的土体参数也并不是都被观测, 数据库中的数据并不完备。具体表现为某一条目的记录数据仅包含少数几个岩土体参

数,而另一条目的记录则包含其他部分数据。融合数据库信息,首先需要估计关于数据库内数据的统计信息。本文同样采用吉布斯抽样技术估计数据库内数据的概率分布信息,由于数据库内数据来源范围分布较广,不考虑数据内数据的空间自相关性。

(1) 采用随机值初始化 $\underline{\mu}_{db}, C_{db}, \underline{a}_{db}, X_{db}^u$, 其中下标 db 代表这些参数通过数据库数据估计。

(2) 根据 $\underline{\mu}_{db}$ 的条件概率密度函数抽取 $\underline{\mu}_{db}$ 样本:

$$f(\underline{\mu}_{db} | X_{db}, C_{db}, \underline{a}_{db}) \sim N\left\{ (C_0^{-1} + n_{db} C_s^{-1})^{-1} \left(\sum_{j=1}^{n_{db}} C_s^{-1} \underline{x}_{db,j} \right), (C_0^{-1} + n_{db} C_s^{-1})^{-1} \right\} \quad (6)$$

式中: $\underline{x}_{db,j}$ 为数据库中第 j 个条目的数据点。

(3) 根据 C_{db} 的条件概率密度函数 $f(C_{db} | X_{db}, \underline{\mu}_{db}, \underline{a}_{db})$ 抽取 C_{db} 样本:

$$f(C_{db} | X_{db}, \underline{\mu}_{db}, \underline{a}_{db}) \sim IW\left\{ \Sigma + \sum_{j=1}^{n_{db}} (\underline{x}_{db,j} - \underline{\mu}_{db})(\underline{x}_{db,j} - \underline{\mu}_{db})^T, n_{db} + m_{db} + 1 \right\} \quad (7)$$

(4) 根据 $a_{db,i}$ 的条件概率密度函数 $f(a_{db,i} | X, \underline{\mu}_s, C_s, \underline{a}_{db,i})$ 抽取 $a_{db,i}$ 样本:

$$f(a_{db,i} | X, \underline{\mu}_{db}, C_{db}) \sim IG\left(\alpha + \frac{m_{db} + 1}{2}, \beta + 2C_{db,ii}^{-1} \right) \quad (8)$$

(5) 根据 X_{db}^u 的条件概率密度函数 $f(X_{db}^u | X_{db}^o, \underline{\mu}_{db}, C_{db}, \underline{a}_{db})$ 抽取 X^u 样本:

$$f(\underline{x}_{db,j}^u | \underline{x}_{db,j}^o, \underline{\mu}_{db}, C_{db}, \underline{a}_{db}) \sim N\left\{ \underline{\mu}_{db}^u + C_{db}^{uo} (C_{db}^{oo})^{-1} (\underline{x}_{db,j}^o - \underline{\mu}_{db}^o), C_{db}^{uo} (C_{db}^{oo})^{-1} C_{db}^{ou} \right\} \quad (9)$$

式中: C_{db}^{uo} 、 C_{db}^{oo} 和 C_{db}^{ou} 为 C_{db} 的子矩阵, 通过 $\underline{x}_{db,j}^u$ 和 $\underline{x}_{db,j}^o$ 在 $\underline{x}_{db,j}$ 中的位置索引得到。

(6) 步骤(2)~(5)循环 T_{db} 次得到 T_{db} 个 $(X_{db}^u, \underline{\mu}_{db}, C_{db}, \underline{a}_{db})$ 样本。

同样地, 丢弃初始的 t_{db} 个处于非平稳状态样本, 然后在剩下的 $(T_{db} - t_{db})$ 个样本中每间隔 Δt 循环步挑选一个样本得到服从目标分布 $f(\underline{\mu}_{db}, C_{db}, \underline{a}_{db}, X_{db}^u | X_{db}^o)$ 的 $(T_{db} - t_{db}) / \Delta t$ 个 $(\underline{\mu}_{db}, C_{db}, \underline{a}_{db}, X_{db}^u)$ 样本。根据全概率理论, 数据库样本的概率密度函数可表达为

$$f(\mathbf{x} | X_{db}^o) \approx \frac{1}{T_{db} - t_{db}} \left[\sum_{t=t_{db}+1}^{T_{db}} f(\mathbf{x} | \underline{\mu}_{db}, C_{db}, X_{db}^o) \right] \quad (10)$$

在实际应用中, 数据库信息主要扮演先验信息的角色, 通过融合数据库信息, 一方面可以降低特定场地土体参数统计量的统计不确定性; 另一方面也可以确保土体参数的估计样本会保持在合理的取值范围内。下面介绍基于变种贝叶斯理论的数据库数据与特定场地勘察数据融合方法。

2.3 变种贝叶斯

当特定场地勘探数据十分稀疏时, 仅根据该场地勘探数据估计的土体参数的后验分布具有很大不确定性, 后验分布 $f(\mathbf{x} | X^o)$ 趋向于扁平。根据该后验分布场地土体参数在其取值范围内取任意值的可能性是近似相等的, 这意味着岩土工程师几乎没有掌握场地土体强度参数的任何有价值信息, 此时经验丰富的工程师可能更倾向于采用相对较平均或保守的土体参数来进行岩土工程结构设计, 大部分情况下该平均或保守的土体参数取值十分依赖于工程师的个人经验, 不合理的取值可能会导致不合理的工程设计进而造成巨大的安全隐患。在工程师依赖经验取值的过程中, 工程师事实上已经在一定程度上使用了历史数据库土体参数的信息。依赖经验的方式存在很大主观不确定性, 为了避免这种主观不确定性, 可以通过定量的方式使用历史数据库信息, 即构建融合历史数据信息的场地土体参数概率分布模型。融合方式应满足以下目标:

(1) 当场地勘探数据量较稀疏时, 融合后的概率分布模型 $f(\mathbf{x} | X^o, X_{db}^o)$ 更接近于仅依赖数据库数据建立得概率分布模型 $f(\mathbf{x} | X_{db}^o)$ 。

(2) 当场地勘探数据量较充裕时, 融合后的概率分布模型 $f(\mathbf{x} | X^o, X_{db}^o)$ 更接近于仅依赖场地勘探数据建立的概率分布模型 $f(\mathbf{x} | X^o)$ 。

根据两个分布相乘的特性, 即一个扁平形状的分布乘以一个钟形分布得到的新分布, 更接近于钟形分布, Ching 等^[8]提出一种基于变种贝叶斯理论的数据融合方法, 该融合方法通过直接相乘 $f(\mathbf{x} | X_{db}^o)$ 与 $f(\mathbf{x} | X)$ 实现融合数据库数据信息与特定场地勘察数据信息, 能够满足前述两个目标, 实现基于数据驱动的稀疏测量场地土体参数不确定性量化。然而 Ching 等^[8]的研究主要针对单一钻孔数据, 并未考虑特地土体参数的空间自相关性。本文进一步融合多钻孔勘察数据, 同时考虑场地勘察数据的三维空间变异性, 克服原有变种贝叶斯数据融合方法难以考虑空间自相关性的局限性, 实现基于大数据驱动的

稀疏测量场地三维概率化重构。研究表明,波动范围的统计不确定性也会对土体参数均值和标准差的估计产生一定影响,但对土体参数交叉相关系数估计的影响较低^[27]。对于稀疏勘察场地,融合数据库信息后的特定场地概率密度函数与数据库概率密度函数更为接近,因此,受波动范围统计不确定性的影响较小。此外,考虑波动范围的统计不确定性,会导致计算成本急剧上升,因此,本文在估计三维场地概率密度函数的过程中忽视了波动范围的统计不确定性。理论上,变种贝叶斯融合方法需要直接相乘土体参数数据库数据概率密度函数 $f(\mathbf{x}|\mathbf{X}_{\text{db}}^{\circ})$ 和场地土体参数概率密度函数 $f(\mathbf{x}|\mathbf{X}^{\circ})$, 即通过下式估计土体参数融合数据库信息后的变种分布:

$$f(\mathbf{x}|\mathbf{X}^{\circ}, \mathbf{X}_{\text{db}}^{\circ}) \propto f(\mathbf{x}|\mathbf{X}_{\text{db}}^{\circ}) \cdot f(\mathbf{x}|\mathbf{X}^{\circ}) \propto \left[\sum_{i=t_{\text{db}}+1}^{T_{\text{db}}} f(\mathbf{x}|\underline{\mu}_{\text{db}}, \mathbf{C}_{\text{db}}, \mathbf{X}_{\text{db}}^{\circ}) \right] \cdot \left[\sum_{j=t_b+1}^T f(\mathbf{x}|\underline{\mu}_s, \mathbf{C}_s, \mathbf{X}^{\circ}) \right] \propto \sum_{i=t_{\text{db}}+1}^{T_{\text{db}}} \sum_{j=t_b+1}^T w_{ij} N(\mathbf{x}|\underline{\mu}_{ij}, \mathbf{C}_{ij}) \quad (11)$$

式中: w_{ij} 为 $f(\mathbf{x}|\mathbf{X}_{\text{db}}^{\circ})$ 中第 i 个高斯分量与 $f(\mathbf{x}|\mathbf{X}^{\circ})$ 中第 j 个高斯分量相乘得到的权系数; $\underline{\mu}_{ij}$ 、 $\mathbf{C}_{ij}$ 分别为 $f(\mathbf{x}|\mathbf{X}_{\text{db}}^{\circ})$ 中第 i 个高斯分量与 $f(\mathbf{x}|\mathbf{X}^{\circ})$ 中第 j 个高斯分量相乘得到的高斯分布的均值向量与协方差矩阵。 w_{ij} 、 $\underline{\mu}_{ij}$ 和 $\mathbf{C}_{ij}$ 可通过下式估计:

$$\left. \begin{aligned} w_{ij} &= \left| \mathbf{C}_{\text{db},i} + \mathbf{C}_{s,j} \right|^{-0.5} (2\pi)^{-m/2} \cdot \exp \left[-\frac{1}{2} (\underline{\mu}_{\text{db},i} - \underline{\mu}_{s,j})^T (\mathbf{C}_{\text{db},i} + \mathbf{C}_{s,j}) (\underline{\mu}_{\text{db},i} - \underline{\mu}_{s,j}) \right] \\ \underline{\mu}_{ij} &= (\mathbf{C}_{\text{db},i}^{-1} + \mathbf{C}_{s,j}^{-1})^{-1} (\mathbf{C}_{\text{db},i}^{-1} \underline{\mu}_{\text{db},i} + \mathbf{C}_{s,j}^{-1} \underline{\mu}_{s,j}) \\ \mathbf{C}_{ij} &= (\mathbf{C}_{\text{db},i}^{-1} + \mathbf{C}_{s,j}^{-1})^{-1} \end{aligned} \right\} \quad (12)$$

注意到式(11)包含成千上万个高斯分量,过多的高斯分量不仅导致根据式(11)直接抽样十分困难,同时对计算机内存具有较高的要求,导致普通台式计算机难以计算。为了便于计算,本文首先对数据库土体参数概率密度函数 $f(\mathbf{x}|\mathbf{X}_{\text{db}}^{\circ})$ 进行折减,采用一个近似的单高斯概率密度函数来近似 $f(\mathbf{x}|\mathbf{X}_{\text{db}}^{\circ})$, 考虑到数据库信息在特定场地土体参数不确定量化过程中主要扮演先验信息的作用,采用近似的单高斯概率密度函数虽然可以降低数据库土体参数概率密度函数精度,但并不会对估计特定场地土体参数的概率密度函数产生较大影响。根据 Crouse^[28]、Cui^[29]等的研究,高斯混合密度函数可通过如下方式折减为一个单高斯密度函数:

$$\left. \begin{aligned} f(\mathbf{x}|\mathbf{X}_{\text{db}}^{\circ}) &\approx f(\mathbf{x}|\underline{\mu}_{\text{rdb}}, \mathbf{C}_{\text{rdb}}) = (2\pi)^{-\frac{m}{2}} |\mathbf{C}_{\text{rdb}}|^{-\frac{1}{2}} \cdot \exp \left[-\frac{1}{2} (\mathbf{x} - \underline{\mu}_{\text{rdb}})^T \mathbf{C}_{\text{rdb}}^{-1} (\mathbf{x} - \underline{\mu}_{\text{rdb}}) \right] \\ \underline{\mu}_{\text{rdb}} &= \frac{1}{T - t_b} \sum_{i=t_b+1}^T \underline{\mu}_{\text{db},i} \\ \mathbf{C}_{\text{rdb}} &= \frac{1}{T - t_b} \sum_{i=t_b+1}^T \left[\mathbf{C}_{\text{db},i} + (\underline{\mu}_{\text{db},i} - \underline{\mu}_{\text{rdb}})(\underline{\mu}_{\text{db},i} - \underline{\mu}_{\text{rdb}})^T \right] \end{aligned} \right\} \quad (13)$$

式中: $\underline{\mu}_{\text{rdb}}$ 和 $\mathbf{C}_{\text{rdb}}$ 分别为折减后高斯分布的均值向量和协方差矩阵; $\underline{\mu}_{\text{db},i}$ 和 $\mathbf{C}_{\text{db},i}$ 分别为 $f(\mathbf{x}|\mathbf{X}_{\text{db}}^{\circ})$ 的第 i 个高斯分量的均值向量和协方差矩阵。将式(13)代入式(11)得到

$$f(\mathbf{x}|\mathbf{X}^{\circ}, \mathbf{X}_{\text{db}}^{\circ}) \propto f(\mathbf{x}|\underline{\mu}_{\text{rdb}}, \mathbf{C}_{\text{rdb}}) \cdot \left[\sum_{i=1}^N f(\mathbf{x}|\underline{\mu}_i, \mathbf{C}_i, \mathbf{X}^{\circ}) \right] \propto \sum_{i=1}^N w_i N(\mathbf{x}|\underline{\mu}_i, \mathbf{C}_i) \quad (14)$$

式中: w_i 为 $f(\mathbf{x}|\underline{\mu}_{\text{rdb}}, \mathbf{C}_{\text{rdb}})$ 与 $f(\mathbf{x}|\mathbf{X}^{\circ})$ 中第 i 个高斯分量相乘得到的权系数; $\underline{\mu}_i$ 和 $\mathbf{C}_i$ 分别为 $f(\mathbf{x}|\underline{\mu}_{\text{rdb}}, \mathbf{C}_{\text{rdb}})$ 与 $f(\mathbf{x}|\mathbf{X}^{\circ})$ 中第 i 个高斯分量相乘得到的高斯分布的均值向量与协方差矩阵。 w_i 、 $\underline{\mu}_i$ 和 $\mathbf{C}_i$ 可通过下式估计:

$$\left. \begin{aligned} w_i &= \left| \mathbf{C}_{\text{rdb}} + \mathbf{C}_{s,i} \right|^{-0.5} (2\pi)^{-m/2} \cdot \exp \left[-\frac{1}{2} (\underline{\mu}_{\text{rdb}} - \underline{\mu}_{s,i})^T (\mathbf{C}_{\text{rdb}} + \mathbf{C}_{s,i}) (\underline{\mu}_{\text{rdb}} - \underline{\mu}_{s,i}) \right] \\ \underline{\mu}_i &= (\mathbf{C}_{\text{rdb}}^{-1} + \mathbf{C}_{s,i}^{-1})^{-1} (\mathbf{C}_{\text{rdb}}^{-1} \underline{\mu}_{\text{rdb}} + \mathbf{C}_{s,i}^{-1} \underline{\mu}_{s,i}) \\ \mathbf{C}_i &= (\mathbf{C}_{\text{rdb}}^{-1} + \mathbf{C}_{s,i}^{-1})^{-1} \end{aligned} \right\} \quad (15)$$

式中: $\underline{\mu}_i$ 和 $\mathbf{C}_i$ 代表了给定 $\mathbf{X}^{\circ}$ 和 $\mathbf{X}_{\text{db}}^{\circ}$ 的条件均值与条件协方差样本。当特定场地勘察数据稀疏时, $\underline{\mu}_i$ 和 $\mathbf{C}_i$ 受数据库均值和协方差样本影响较大。当土体参数数据库中数据的不确定性较小时,融合数据库信息后特定场地土体参数统计量的不确定性越小。一般情况下,区域数据库数据比全球数据库具有更小的不确定性。因此,有条件的情况下,建议使用区域数据库进行三维场地的重构。

根据式(15)估计的 $\underline{\mu}_i$ 和 $\mathbf{C}_i$ 样本,给定 $\mathbf{X}^{\circ}$ 和 $\mathbf{X}_{\text{db}}^{\circ}$ 的条件 $\mathbf{X}^{\circ}$ 样本可以通过顺序钻孔模拟方法高效模拟,第 k 个钻孔未观测到的数据可通过如下公式模拟:

$$\left. \begin{aligned} E(\underline{\mathbf{X}}_k^u | \underline{\mathbf{X}}_k^{\circ}, \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ}) &= E(\underline{\mathbf{X}}_k^u | \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ}) + \left\{ \begin{aligned} &\text{Cov}(\underline{\mathbf{X}}_k^u, \underline{\mathbf{X}}_k^{\circ} | \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ}) \text{Var}(\underline{\mathbf{X}}_k^{\circ} | \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ})^{-1} \cdot \\ &\left[\underline{\mathbf{X}}_k^{\circ} - E(\underline{\mathbf{X}}_k^{\circ} | \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ}) \right] \\ &\text{Var}(\underline{\mathbf{X}}_k^u | \underline{\mathbf{X}}_k^{\circ}, \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ}) = \text{Var}(\underline{\mathbf{X}}_k^u | \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ}) - \\ &\text{Cov}(\underline{\mathbf{X}}_k^u, \underline{\mathbf{X}}_k^{\circ} | \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ}) \text{Var}(\underline{\mathbf{X}}_k^{\circ} | \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ})^{-1} \cdot \\ &\text{Cov}(\underline{\mathbf{X}}_k^u, \underline{\mathbf{X}}_k^{\circ} | \underline{\mathbf{X}}_{\sim k}, \mathbf{X}_{\text{db}}^{\circ})^T \end{aligned} \right\} \quad (16)$$

式中： $E(\underline{X}_k^u | \underline{X}_{-k}, \underline{X}_{db}^o)$ 和 $E(\underline{X}_k^o | \underline{X}_{-k}, \underline{X}_{db}^o)$ 分别为给定 $\underline{X}_{-k}$ 和 $\underline{X}_{db}^o$ 条件下 $\underline{X}_k^u$ 和 $\underline{X}_k^o$ 的均值； $Cov(\underline{X}_k^u, \underline{X}_k^o | \underline{X}_{-k}, \underline{X}_{db}^o)$ 为给定 $\underline{X}_{-k}$ 和 $\underline{X}_{db}^o$ 条件下 $\underline{X}_k^u$ 和 $\underline{X}_k^o$ 之间的协方差矩阵； $Var(\underline{X}_k^u | \underline{X}_{-k}, \underline{X}_{db}^o)$ 和 $Var(\underline{X}_k^o | \underline{X}_{-k}, \underline{X}_{db}^o)$ 分别为给定 $\underline{X}_{-k}$ 和 $\underline{X}_{db}^o$ 条件下 $\underline{X}_k^u$ 和 $\underline{X}_k^o$ 的方差矩阵。 $E(\underline{X}_k^u | \underline{X}_{-k}, \underline{X}_{db}^o)$ 和 $E(\underline{X}_k^o | \underline{X}_{-k}, \underline{X}_{db}^o)$ 均为 $E(\underline{X}_k | \underline{X}_{-k}, \underline{X}_{db}^o)$ 的子向量， $Cov(\underline{X}_k^u, \underline{X}_k^o | \underline{X}_{-k}, \underline{X}_{db}^o)$ 、 $Var(\underline{X}_k^u | \underline{X}_{-k}, \underline{X}_{db}^o)$ 和 $Var(\underline{X}_k^o | \underline{X}_{-k}, \underline{X}_{db}^o)$ 均为 $Var(\underline{X}_k | \underline{X}_{-k})$ 的子矩阵。 $E(\underline{X}_k | \underline{X}_{-k}, \underline{X}_{db}^o)$ 与 $Var(\underline{X}_k | \underline{X}_{-k})$ 可通过下式估计：

$$\left. \begin{aligned} E(\underline{X}_k | \underline{X}_{-k}, \underline{X}_{db}^o) &= \underline{\mu}_{i,k} + \\ &\quad \text{vec} \left[\text{mat}(\underline{X}_{-k} - \underline{\mu}_{i,-k}) (R_{h,X_{-k}}^{-1} R_{h,\underline{X}_k}^T) \right] \\ Var(\underline{X}_k | \underline{X}_{-k}, \underline{X}_{db}^o) &= \\ &\quad (1 - R_{h,\underline{X}_k} R_{h,X_{-k}}^{-1} R_{h,\underline{X}_k}^T) (R_{v,X} \otimes C_i) \end{aligned} \right\} \quad (17)$$

式中： $\text{vec}(\bullet)$ 代表重排矩阵为一个列向量； $R_{h,\underline{X}_k}$ 为第 k 个钻孔与其他钻孔间的水平自相关矩阵； $R_{h,X_{-k}}$ 为除第 k 个钻孔外其他钻孔间的水平自相关矩阵； $R_{v,X}$ 为钻孔数据的垂直自相关矩阵。通过式(17)可以得到完备的钻孔数据 $\underline{X}$ ，下面介绍基于完备钻孔数据的三维场地概率模型构建方法。

2.4 三维场地概率模型构建

设 $\underline{X}_{3D}$ 为三维场地的数据，一般情况下 $\underline{X}_{3D}$ 的维度远大于钻孔数据 $\underline{X}$ 的维度。若三维场地被划分为 $n_{x,3D} \times n_{y,3D} \times n_{z,3D}$ 三维网格，则 $\underline{X}_{3D}$ 为一个 $m \times n_{x,3D} \times n_{y,3D} \times n_{z,3D}$ 的列向量。在岩土结构物可靠度分析与风险评估中，可以通过映射 $\underline{X}_{3D}$ 到有限元或有限差分网格中模拟土体参数的空间不确定性。如前所述，吉布斯抽样可以模拟钻孔数据中的未观测数据 $\underline{X}^o$ ，得到完备的钻孔数据，此时， $\underline{X}_{3D}$ 与 $\underline{X}$ 均满足三维格栅结构。对于满足三维格栅结构的数据，若其相关函数满足在水平与垂直方向为可分离的，利用Kronecker积可以有效地降低 $\underline{X}_{3D}$ 与 $\underline{X}$ 间自相关矩阵的尺寸，在此基础上， $\underline{X}_{3D}$ 可通过下式高效地模拟^[9, 25]：

$$\underline{X}_{3D} = \underline{\mu}_{3D} + \text{vec} \left[(R_{v,X_{3D}} R_v^{-1} \otimes I_s) \text{mat}(\underline{X} - \underline{\mu}) \cdot R_h^{-1} R_{h,X_{3D}}^T \right] + \text{vec} \left[(L_{v,X_{3D}} \otimes L_c) Z L_{h,X_{3D}}^T \right] \quad (18)$$

式中： $R_{v,X_{3D}}$ 尺寸为 $n_{v,3D} \times n_v$ ，为三维场地数据 $\underline{X}_{3D}$ 与钻孔数据 $\underline{X}$ 在垂直方向的自相关矩阵； $R_{h,X_{3D}}$ 尺寸为 $n_{h,3D} \times n_h$ ，为三维场地数据 $\underline{X}_{3D}$ 与钻孔数据 $\underline{X}$ 在水平方向的自相关矩阵； $\underline{\mu}_{3D}$ 为 $n_{3D} \times m$ 维列向量，通过堆叠向量 $\underline{\mu}$ 形成； $\underline{\mu}$ 为 $n \times m$ 维列向量，也是通过堆叠向量 $\underline{\mu}$ 形成； $L_{v,X_{3D}}$ 、 $L_{h,X_{3D}}$ 和 L_c 分别为

$R_{v,X_{3D}}$ 、 $R_{h,X_{3D}}$ 和 C_i 的乔列斯基分解下三角矩阵； $R_{v,X_{3D}}$ 尺寸为 $n_{v,3D} \times n_{v,3D}$ ，为三维场地数据 $\underline{X}_{3D}$ 在垂直方向的自相关矩阵； $R_{h,X_{3D}}$ 尺寸为 $n_{h,3D} \times n_{h,3D}$ ，为三维场地数据 $\underline{X}_{3D}$ 在水平方向的自相关矩阵； $\underline{Z}$ 为包含 $(n_{v,3D} \times m) \times n_{h,3D}$ 个标准正态独立随机变量的矩阵； $\text{mat}(\underline{X} - \underline{\mu})$ 代表重排向量 $(\underline{X} - \underline{\mu})$ 为一个 $(m \times n_v) \times n_h$ 维的矩阵； I_s 为单位矩阵，尺寸为 $m \times m$ 。

需要注意的是，通过式(18)准确模拟 $\underline{X}_{3D}$ 需满足两个条件：①钻孔数据 $\underline{X}$ 的抽样间距应小于垂直波动范围；②抽样区间应覆盖 $\underline{X}_{3D}$ 的深度范围。在实际执行中，通过在 $\underline{X}$ 中插入额外的未观测深度可以轻易满足上述两个条件。本文主要研究包含单一土层的三维场地，对于多土层复杂场地，本文方法可以结合现有的土层分布边界的识别方法一起使用。在已知三维场地土层分布的情况下，可以通过本文方法估计不同土层对应的土体参数统计量并模拟不同土层对应的三维约束随机场，最后通过土层分界面插值不同土层的随机场实现多土层三维场地重构。

3 算例分析

3.1 数值模拟场地案例

3.1.1 虚拟场地模拟

本节采用一个数值模拟虚拟场地案例来说明本文所提方法。假设该虚拟场地包含了 X_1 、 X_2 和 X_3 共3个土体参数， X_1 、 X_2 和 X_3 服从协方差矩阵为 $C_s = [1 \ -0.5 \ -0.3; -0.5 \ 1 \ 0.7; -0.3 \ 0.7 \ 1]$ 与均值向量 $\underline{\mu} = [0 \ 0 \ 0]$ 的三元正态分布，参数 X_1 与 X_2 和 X_3 之间具有负相关性，参数 X_2 和 X_3 具有正相关性。3个土体参数具有一致的空间自相关性，采用单指数自相关函数 $\rho = \exp[-2(|\Delta x|^2 + |\Delta y|^2)^{0.5} / \delta_h - |\Delta z| / \delta_z]$ 表征3个参数的空间自相关性，其中 Δx 、 Δy 和 Δz 分别为空间上两点之间的水平距离， δ_h 和 δ_z 分别为水平和垂直波动范围，分别取 $\delta_h = 10 \text{ m}$ 和 $\delta_z = 0.5 \text{ m}$ 。虚拟场地尺寸为 51 m （ x 方向） $\times 51 \text{ m}$ （ y 方向） $\times 10.1 \text{ m}$ （ z 方向），为有效模拟虚拟场地土体参数的空间变异性，虚拟场地沿 x 方向和 y 方向分别离散为51个网格，沿 z 方向离散为101个网格，离散后虚拟场地共包含20 604个三维网格。假设土体参数自相关性在水平和垂直方向可分离，采用Li等^[30]所提逐步分解法模拟 X_1 、 X_2 和 X_3 3个土体参数的随机场，图1给出了虚拟场地的3个土体参数在三维空间上的分布情况。生成虚拟场地后，假设土体参数的值在三维空间上是未知的，仅在少数勘探位置获得了土体参数的值，同时土体参数的均值向量和协方差

矩阵也假设为未知量。上述假设与实际场地勘察情况基本相符,实际工程中土体参数的统计量往往并不能预先获得,需要通过勘察数据估计,且勘察数据一般分布在有限几个钻孔位置处。本文假设该虚拟场地仅在图 1 中 4 个钻孔位置(钻孔 1~4)测量土体参数,图 2 给出了钻孔 1~4 的土体参数数据。假设钻孔 1~4 仅在有限的深度进行土体参数测量,为了研究观测数据量对虚拟场地重构的影响,分别

考虑以下几种勘察工况:(a)1 个观测深度($z=3\text{ m}$);(b)2 个观测深度($z=3、7\text{ m}$);(c)3 个观测深度($z=0.1、3、7\text{ m}$);(d)4 个观测深度($z=0.1、3、7、10\text{ m}$);(e)5 个观测深度($z=0.1、3、4.9、7、10\text{ m}$);(f)10 个观测深度($z=0、1.1、2.2、3.3、4.4、\cdots、9.9\text{ m}$);(g)15 个观测深度($z=0、0.7、1.4、2.1、2.8、\cdots、9.8\text{ m}$);(h)21 个观测深度($z=0、0.5、1.0、1.5、2.0、\cdots、10.0\text{ m}$)。

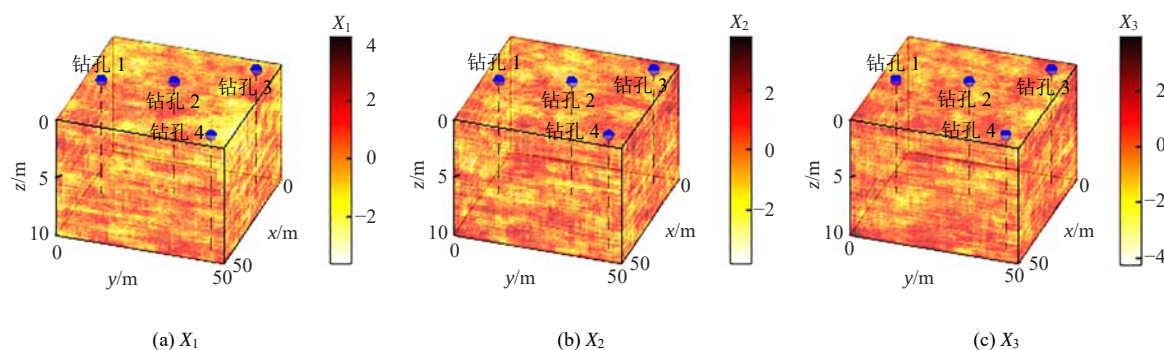


图 1 虚拟场地
Fig.1 The virtual site

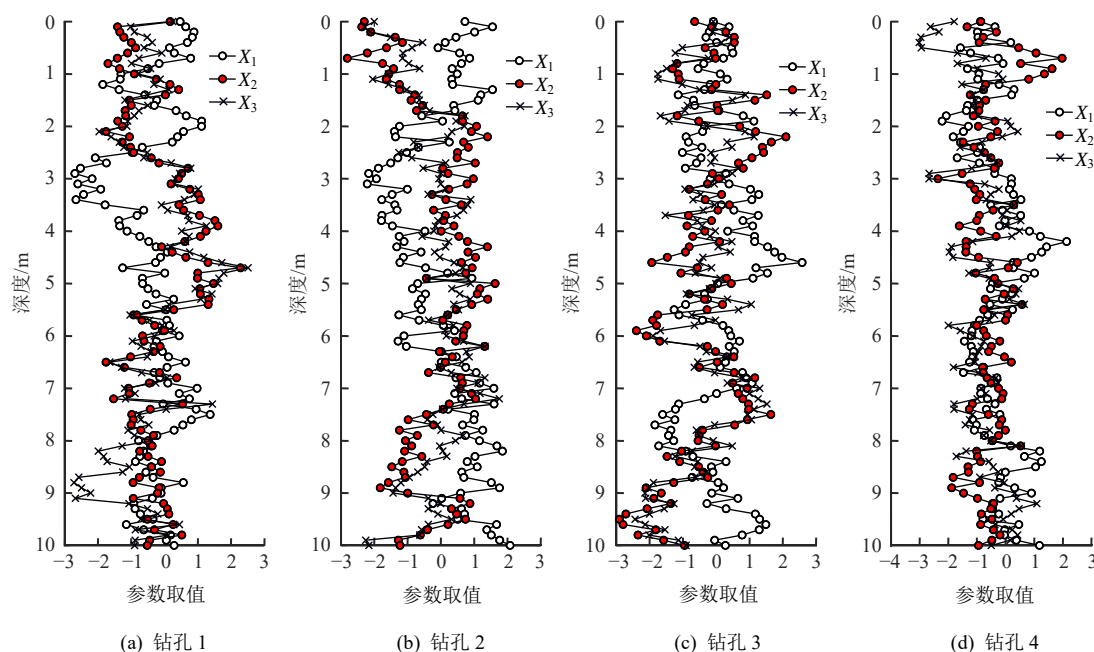


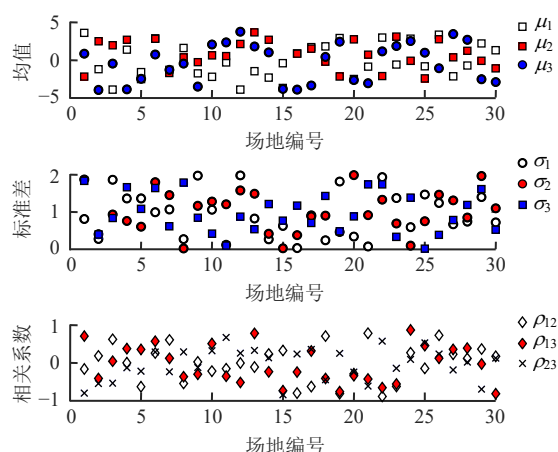
图 2 虚拟场地 4 个钻孔位置处的数据
Fig.2 Soil profiles at the four boreholes for virtual site

3.1.2 数据库概率密度函数估计

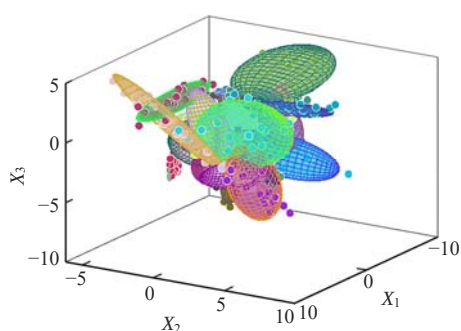
本节通过一个虚拟数据库来为虚拟场地土体参数不确定性量化提供先验信息。注意到,与某一特定场地相比,数据库土体参数数据往往具有更大的取值范围,特定场地可以视为数据库内某一场地的相似场地(均值、标准偏差与相关系数等接近)。通过 30 个不同的虚拟场地来模拟数据库内的土体

参数数据,图 3(a)给出了数据库中 30 个场地的土体参数均值 μ 、标准差 σ 和相关系数 ρ ,根据这些均值、标准差和相关系数,每个场地随机生成不同数目的 X_1 、 X_2 和 X_3 样本。图 3(b)给出了数据库中 30 个虚拟场地的 $2\text{-}\sigma$ 椭球与 398 个模拟的 X_1 、 X_2 和 X_3 样本。

根据模拟的 X_1 、 X_2 和 X_3 样本采用 2.2 节所提吉



(a) 数据库中不同场地土体参数统计量



(b) 数据库中虚拟场地 2-σ 椭球

图3 数据库中的虚拟场地及数据

Fig.3 Virtual sites and data contained in database

布斯抽样方法估计数据库土体参数概率模型。吉布斯抽样得到 21 000 个 (μ_{db}, C_{db}) 样本, 丢弃初始的 1 000 个样本, 在剩下的 20 000 个样本中每 4 个样本取 1 个样本, 得到 5 000 个 (μ_{db}, C_{db}) 样本, 将 (μ_{db}, C_{db}) 样本代入式 (10) 得到数据库土体参数的高斯混合概率密度函数。同样地, 针对 3.1.1 小节生成虚拟场地的不同勘察工况, 采用 2.2 节吉布斯抽样方法估计特定场地土体参数概率模型。由于特定场地概率密度函数估计过程中考虑了土体参数的空间自相关性, 吉布斯抽样在丢弃初始 1 000 个样本之后, 每隔 10 个样本挑选 1 个样本, 共得到 2 000 个样本作为场地土体参数的条件均值样本和条件均值方差, 根据式 (5) 可得到基于有限勘察数据的特定场地土体参数概率密度函数。数据库土体参数概率密度函数与特定场地土体参数概率密度函数均为高斯混合密度函数, 通过变种贝叶斯理论直接融合两个概率密度函数得到的新的高斯混合密度函数具有 $2\,000 \times 5\,000$ 个高斯分量, 不仅计算量巨大, 且抽样困难。为此, 采用式 (13) 先对数据库概率高斯混合密度函数进行折减, 折减后数据库土体参数

概率密度函数为一单高斯分布, 图 4 给出了折减后单高斯分布的 $2\text{-}\sigma$ 椭球。根据式 (14) 将折减后的高斯概率密度函数与特定场地土体概率密度函数进行融合, 得到新的包含 2 000 个高斯分量的高斯混合密度函数, 该概率密度函数则代表了考虑数据库信息与特定场地勘察信息的土体参数概率密度函数。

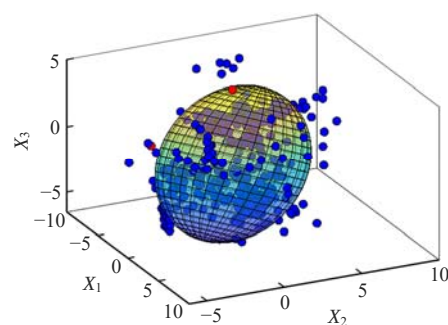


图4 数据库数据分布

Fig.4 Distribution of data contained in database

3.1.3 MUSIC-3X 方法与本文方法对比

图 5 给出了 1、3、15 个观测深度工况下, 杨智勇等^[10]所提 MUSIC-3X 方法与本文所提方法的对比。图 5 中不同颜色的点分别代表基于 $f(\mu, C_s|X^o)$ 和 $f(\mu, C|X^o, X_{db}^o)$ 模拟得到的样本数据, 其中 $f(\mu, C_s|X^o)$ 由 MUSIC-3X 方法估计, 并未考虑数据库信息, $f(\mu, C|X^o, X_{db}^o)$ 通过本文所提方法估计, 考虑了数据库信息。图 5 中蓝色线条的交点为均值、标准差及相关系数的真实解。可以看出, 当勘察数据稀疏时 (1 个观测深度), MUSIC-3X 方法估计的均值样本、标准差样本与相关系数样本离散性较大, 具有更大的不确定性, 这是由于 MUSIC-3X 方法未能考虑数据库信息, 勘察数据稀疏条件下土体参数统计量具有较大统计不确定性。相比之下, 本文所提融合数据库信息的方法得到的均值、标准差与相关系数样本不确定性较小, 表明本文所提方法能极大地降低土体参数统计量的统计不确定性。随着勘察数据量的增加, 两种方法估计的均值、标准差与相关系数样本不确定性逐渐接近。这是当特定场地勘察数据量充裕时, 由于数据库信息对特定场地土体参数不确定性的影响变小, 该场地的土体参数统计量能够根据场地自身的勘察数据有效估计。图 6 给出了基于 $f(x|X^o)$ 和 $f(x|X^o, X_{db}^o)$ 得到的模拟样本, 其中 $f(x|X^o)$ 由 MUSIC-3X 方法估计, 并未考虑数据库信息, $f(x|X^o, X_{db}^o)$ 通过本文所提方法估计, 考虑了数据库信息。注意到, 这些样本可以视为距离勘察

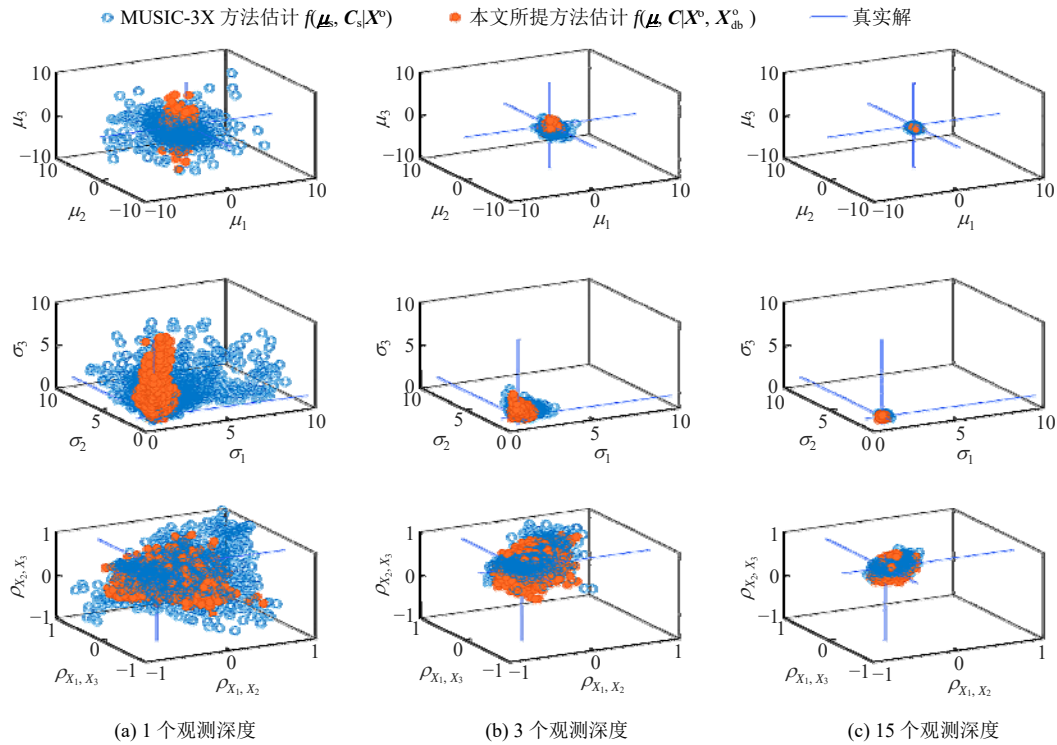
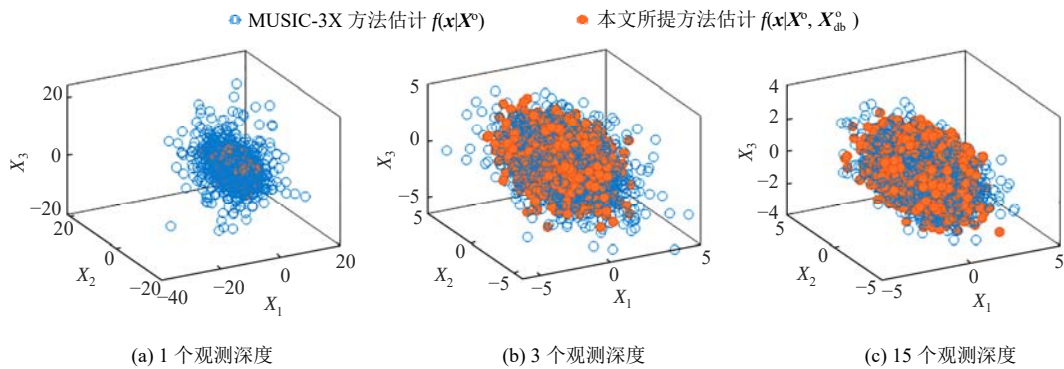


图5 均值、标准差和相关系数样本

Fig.5 Samples of mean, standard deviation and correlation coefficient

图6 基于 $f(x|X^0)$ 和 $f(x|X^0, X_{db}^0)$ 的模拟样本Fig.6 Simulated samples based on $f(x|X^0)$ and $f(x|X^0, X_{db}^0)$

钻孔位置无限远处的土体参数取值, 因此与勘察数据独立。可以看出, 当勘察数据稀疏时, 本文所提方法能有效地降低土体参数估计的不确定性, 当勘察数据增加时, 两种方法模拟的土体参数样本具有相似的分布。图7给出了钻孔1和钻孔2在不同观测深度条件下参数 X_1 沿深度方向的置信区间, 图中空心圆代表实际的土体参数值, 实心圆代表 X_1 的观测值。可以看出, 对于两种方法土体参数真实值基本都落在 95% 置信区间内, 对于仅有 1 个观测深度的工况, 本文所提方法估计的 X_1 置信区间较窄, MUSIC-3X 方法得到的置信区间较大, 表面 X_1 在不同深度取值具有很大不确定性。对于 15 个观测深度的工况, 两种方法估计的置信区间基本一致。

基于 $f(x|X^0)$ 和 $f(x|X^0, X_{db}^0)$ 均可以对钻孔未观测到的数据 X^0 进行模拟, 模拟后钻孔数据 X 完备数据, 根据前述吉布斯抽样, 两种方法分别可以得到 2 000 个 X 完备样本, 然后根据式 (18) 模拟虚拟场地三维约束随机场 X_{3D} 。图8给出了 1 个观测深度下基于两种方法得到的虚拟场地一次典型实现。可以看出, 与虚拟场地的真实土体参数相比, 本文所提方法估计的 X_1 、 X_2 和 X_3 较为合理, MUSIC-3X 方法估计的 X_1 、 X_2 和 X_3 具有非常大的取值范围, 如 X_1 和 X_3 的下界取值为 -10, 已严重偏离实际虚拟场地土体参数的取值 (见图1)。根据 2 000 次约束随机场的典型实现, 可以估计不同参数样本标准偏差在空间上的分布情况。图9给出了 1、3、15 个观

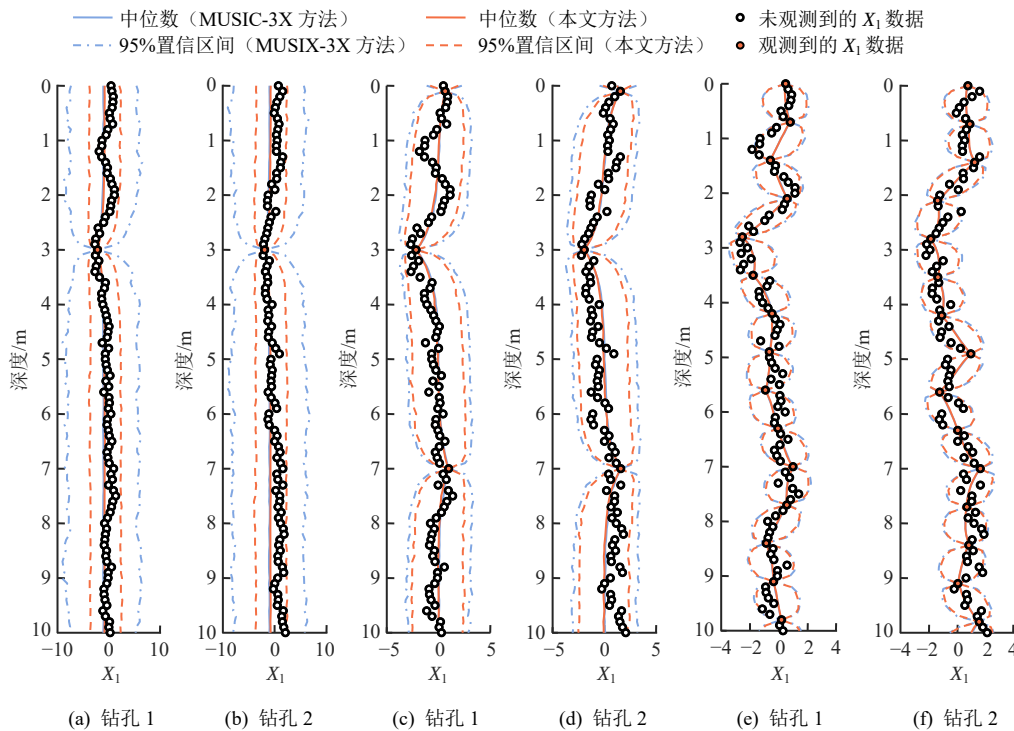


图7 钻孔位置处的 X_1 的模拟样本
Fig.7 Simulated samples of X_1 at selected soundings

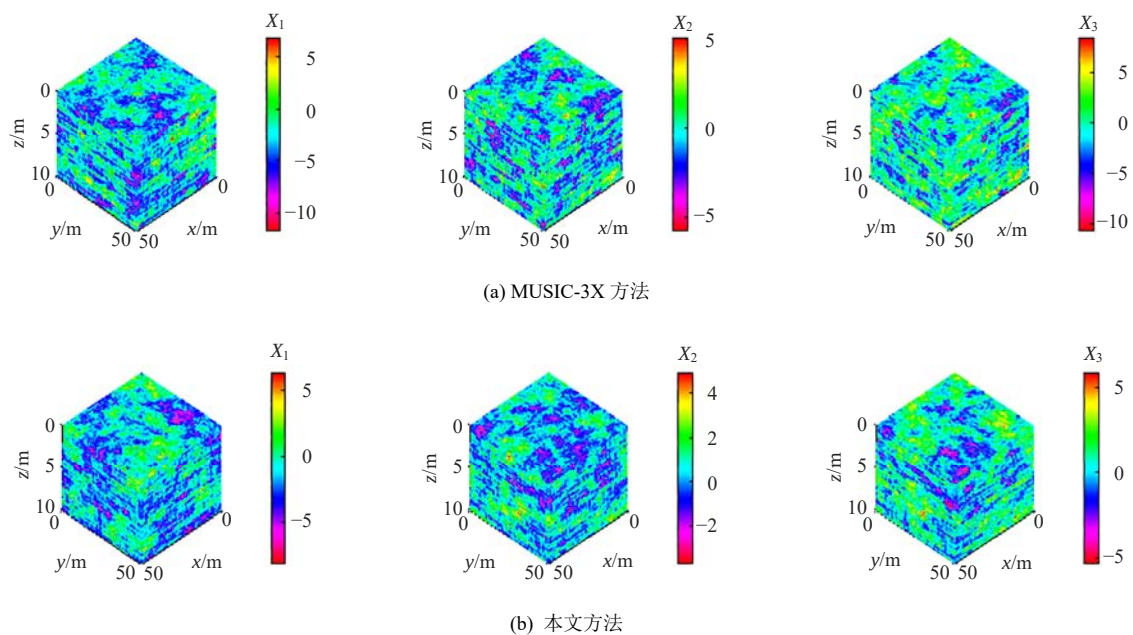


图8 基于 MUSIC-3X 和本文方法的虚拟场地典型实现
Fig.8 One typical realization of virtual site based on MUSIC-3X and the proposed method

测深度工况下 X_1 样本偏差在三维空间上的分布。可以看出,对于1个观测深度和3个观测深度的情况,本文方法得到的样本标准偏差明显小于 MUSIC-3X 方法,越靠近观测深度, X_1 样本标准偏差越小,表明勘测数据可以有效地降低勘测位置附近土体参数的不确定性。这是由于考虑了空间自相关性,观测位置附近处的土体参数受观测数据约束较大,因此在实际工程中有必要土体参数的观测数据,通过约

束随机场模型表征土体参数的空间不确定性。

图10为 X_1 样本标准偏差在三维空间上的均值随不同观测深度个数的变化情况。可以看出,随着观测深度个数的增加, X_1 样本标准偏差均值逐渐降低,本文方法估计的 X_1 样本标准偏差的均值总是低于 MUSIC-3X 方法估计的结果,观测深度个数越少,本文方法相较于 MUSIC-3X 方法的优势越明显,对土体参数不确定性的降低越显著。存在一个临界的

观测深度个数, 超过该观测深度个数, 两种方法对土体参数不确定性的折减均十分有限。根据图 10

可以判断临界观测深度个数为 5, 对应的观测数据个数为 20 个。

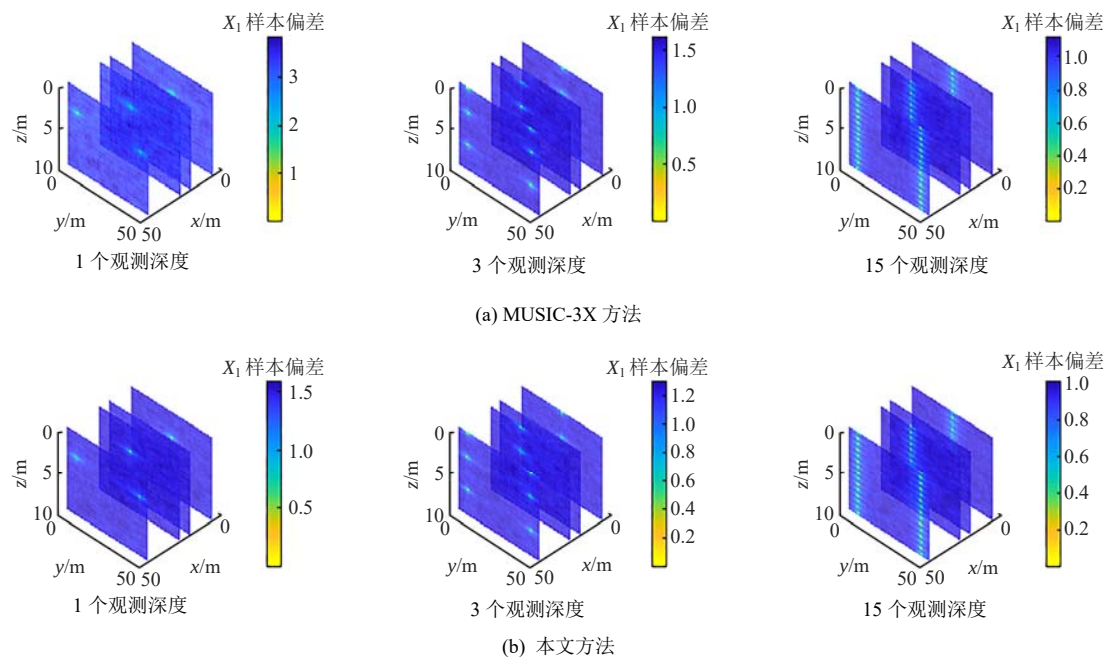


图 9 基于 MUSIC-3X 和本文方法的 X_1 样本标准偏差

Fig.9 Standard deviation of samples of X_1 based on MUSIC-3X and the proposed method

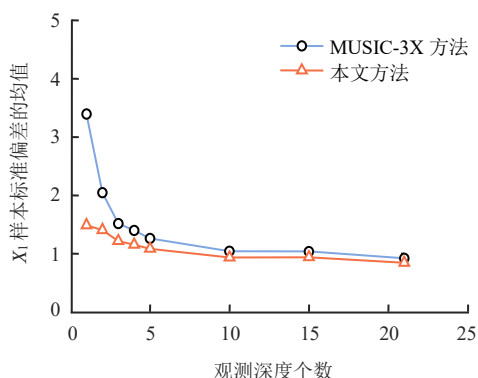


图 10 X_1 的样本标准随观测深度个数的变化

Fig.10 Variation of standard deviation of samples of X_1 with number of observed depth

表 2 对比了两种方法生成虚拟场地 2 000 次实现的计算效率。可以看出, 本文所提方法的计算耗时要大于 MUSIC-3X 方法, 这是因为 MUSIC-3X 方法不需要构建数据库数据的概率密度函数。需要指出的是, 对于某一特定场地, 本文所提方法需要估计场地概率密度函数 $f(\underline{\mu}, C_s, \underline{a}, X^u | X^o)$ 、数据库数据概率密度函数 $f(\underline{\mu}_{db}, C_{db}, \underline{a}_{db}, X_{db}^u | X_{db}^o)$ 和变种概率密度函数 $f(x | X^o, X_{db}^o)$ 1 次, 若要增加虚拟场地的实现次数, 表 2 中仅需要额外增加模拟 X_{3D} 的耗时。此外, $f(\underline{\mu}_{db}, C_{db}, \underline{a}_{db}, X_{db}^u | X_{db}^o)$ 后续也可用于其他场地的三维概率化重构, 不需要重复计算。

表 2 虚拟场地 2 000 次实现两种方法计算效率对比

Table 2 Comparison of computational efficiency of the two methods for 2 000 realizations of virtual site

重构方法	估计不同函数的耗时/s			模拟 X_{3D} 耗时/s	总耗时/s
	$f(\underline{\mu}, C_s, \underline{a}, X^u X^o)$	$f(\underline{\mu}_{db}, C_{db}, \underline{a}_{db}, X_{db}^u X_{db}^o)$	$f(x X^o, X_{db}^o)$		
MUSIC-3X 方法	64	—	—	57	121
本文所提方法	64	185	6	57	312

3.2 美国德克萨斯州场地案例

以美国德克萨斯州的某一场地为例来验证本文所提方法的有效性。该场地位于休斯敦东 50 km 处, 场地主要为黏土, 地下水位深度为 2.9 m。杨智勇等^[10]通过 MUSIC-3X 方法量化了该场地 6 个土体参数 (液限 w_L 、塑性指数 I_P 、液性指数 I_L 、不排水抗剪强度参数 s_u / σ'_v 、孔隙压力比 $B_q = (u_2 - u_0) / (q_t - \sigma'_v)$ 和正则化锥尖阻力 $(q_t - \sigma'_v) / \sigma'_v$) 的不确定性。场地包括 5 个 Borehole 钻孔和 9 个 CPT 钻孔, 详细场地勘察钻孔布置方案和勘探数据分布情况可以参考杨智勇等^[10]的研究。该场地包含的不排水抗剪强度参数数据较为稀疏, 杨智勇等^[10]通过 MUSIC-3X 方法利用不同参数间的交叉相关性来降低不排水抗剪强度的统计不确定性, 且研究仅限于对钻孔位置处未观测数据的估计, 并未涉及模拟该场地在三维空间上非勘探钻孔位置处的数据。通过本文所提方法进一步量化了该场地 6 个土体参数的

不确定性, 通过估计的土体参数条件统计量, 进一步模拟该场地的三维约束随机场。

图 11 给出了数据库 CLAY/10/7490 中 6 个土体参数的数据分布图, 其中蓝色样本点代表了实测的场地勘测数据, 红色样本点代表了通过 $f(\mathbf{x}|\mathbf{X}_{\text{db}}^o)$ 模拟的随机数据点。图 11 位于上三角区域的图为实测

数据与模拟数据在正态空间中的散点图, 其中蓝色椭圆为根据 $f(\mathbf{x}|\mathbf{X}_{\text{db}}^o)$ 绘制的 $2\text{-}\sigma$ 椭圆, 位于下三角区域的图为实测数据与模拟数据在原始空间中的散点图, 位于对角线的图给出了土体参数在正态空间中的边缘概率分布。可以看出, $f(\mathbf{x}|\mathbf{X}_{\text{db}}^o)$ 能够较好地模拟数据库中 6 个土体参数勘察数据的边缘分布及

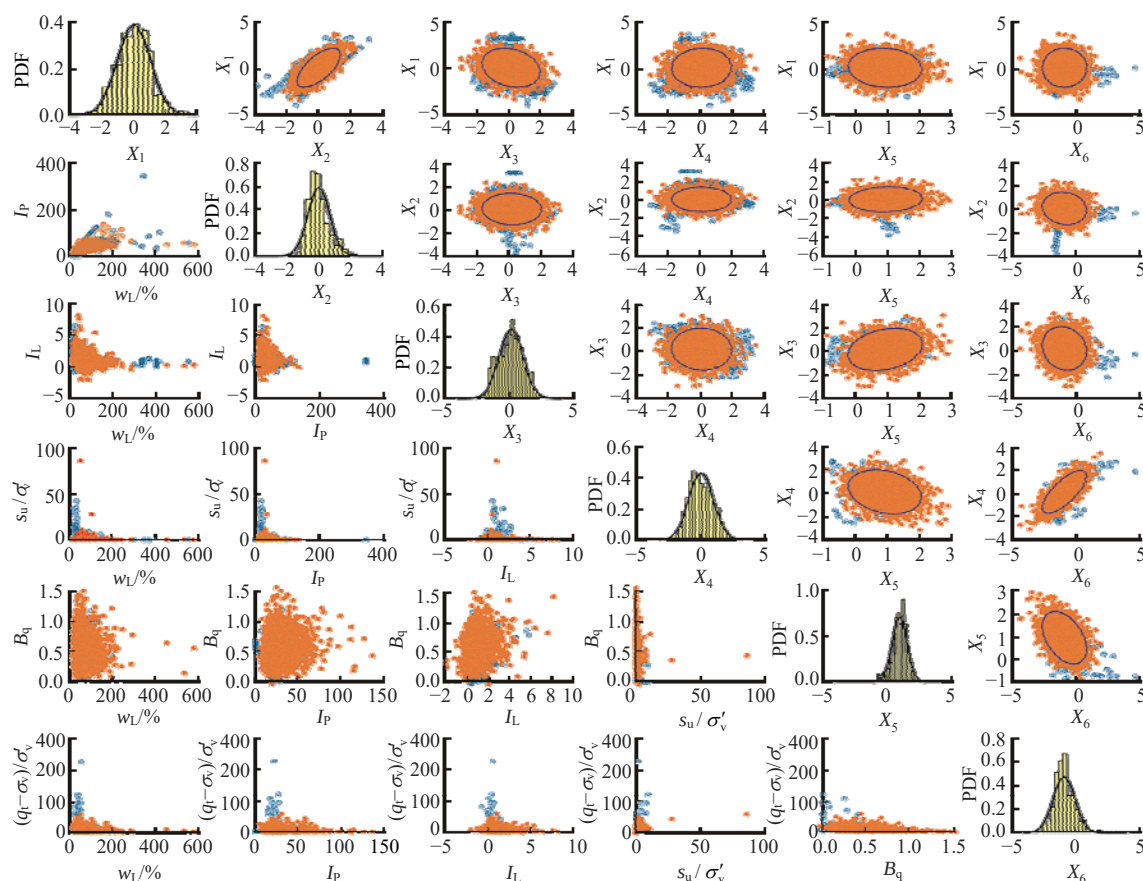


图 11 数据库 CLAY/10/7490 中的土体参数数据及其分布

Fig.11 Soil data at CLAY/10/7490 database and distribution of these soil parameters

其相关性。

图 12 进一步给出了基于本文所提方法和 MUSIC-3X 方法估计的不同钻孔内土体参数的置信区间。可以看出, 在观测深度处土体参数与勘探数据一致, 与观测深度越近, 两种方法估计的土体参数置信区间越窄。这是由于两种方法都可以有效地考虑土体参数空间自相关性, 空间自相关性会降低观测深度附近土体参数的不确定性。在距离观测深度较远处, 空间自相关性的影响逐渐变弱, 本文所提方法得到的土体参数置信区间明显小于 MUSIC-3X 方法, 表明本文所提方法能够有效利用大数据信息降低稀疏测量土体参数的统计不确定性。注意到, 钻孔 1 在 9 m 处及 12 m 处不包含 I_p 观测数据, 但包含 I_L 观测数据, 由于 I_L 与 I_p 较强的

交叉相关性, I_p 的置信区间在 9 m 处和 12 m 处也变窄, 表明本文所提方法也能够利用土体参数间的交叉相关性降低土体参数的不确定性。与 MUSIC-3X 方法相比, 本文所提方法不仅能充分利用土体参数间的交叉相关性降低稀疏测量土体参数的不确定性, 同时也能利用大数据信息降低稀疏测量土体参数的统计不确定性。图 13 给出了基于两种方法估计的德克萨斯州场地一次典型实现。以不排水抗剪强度参数 s_u 为例, 对于该次实现, MUSIC-3X 方法得到的不排水抗剪强度参数 s_u 取值范围为 [8 kPa, 1440 kPa], 本文所提方法为 [9 kPa, 577 kPa]。可以看出, MUSIC-3X 方法得到的土体参数估计值具有较大取值范围, 本文所提方法估计的土体参数取值范围较小, 能够有效地降低土体参数的统计不确定

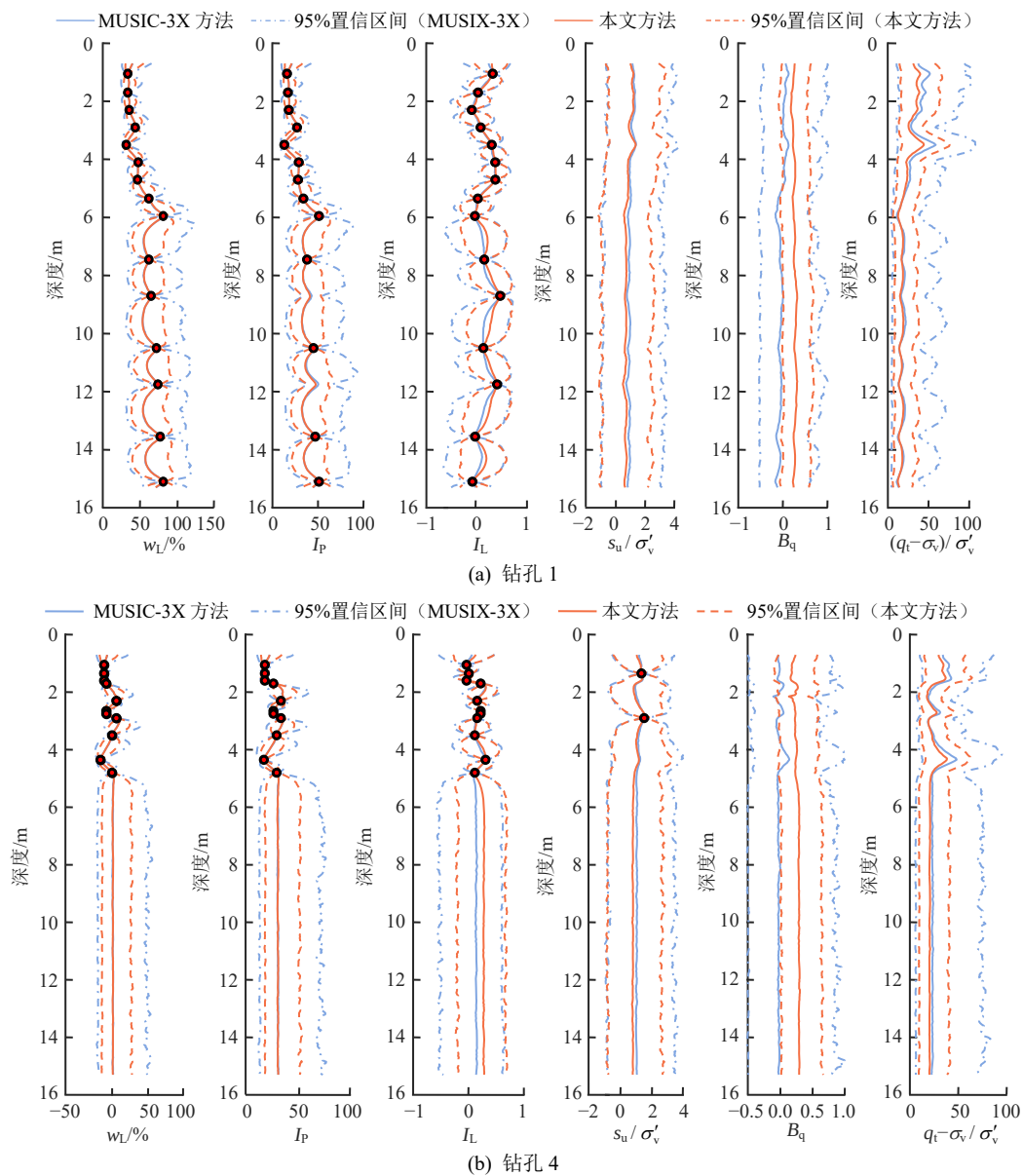


图 12 钻孔 1 与钻孔 4 中土体参数样本中位数及 95%置信区间

Fig.12 Median and 95% confidence intervals of soil parameters' samples at soundings 1 and 4

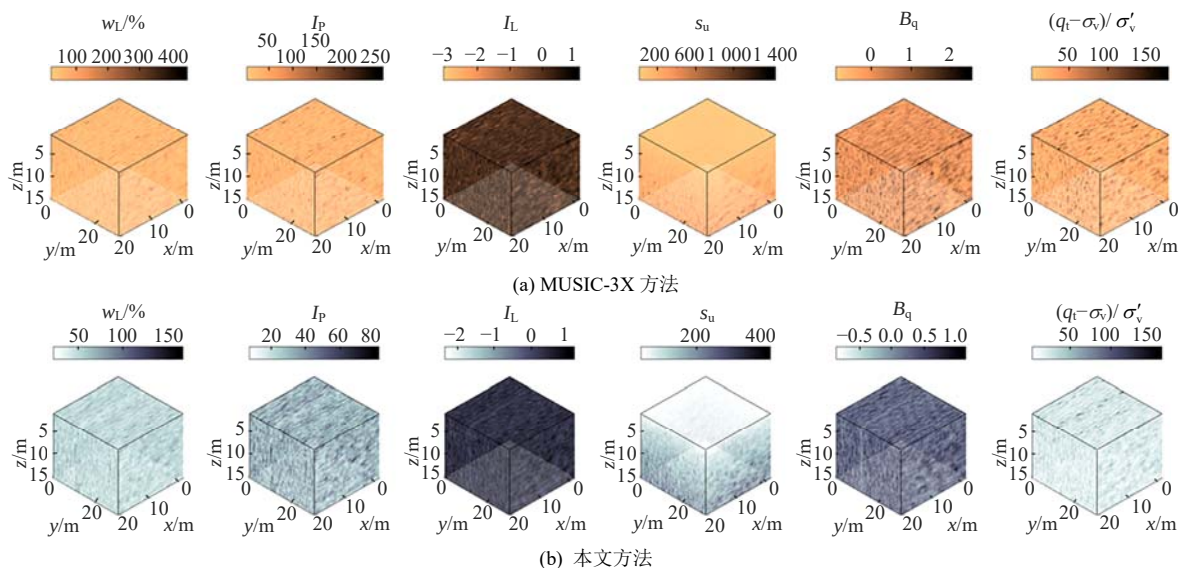


图 13 基于 MUSIC-3X 和本文方法的德克萨斯州场地一次典型实现

Fig.13 One typical realization of the Texas site based on MUSIC-3X and the proposed method

性。表3进一步对比了两种方法生成德克萨斯场地2 000次重构的计算效率。可以看出,本文所提方法的计算耗时要大于MUSIC-3X方法。与虚拟场地对比,本文方法和MUSIC-3X方法计算效率均有一定程度下降,这主要是由于随机变量维度升高时,通过吉布斯抽样估计三维场地概率密度函数 $f(\underline{\mu}_s, C_s, \underline{a}, X^u|X^o)$ 的效率降低所致。

表3 德克萨斯场地2 000次实现两种方法计算效率对比
Table 3 Comparison of computational efficiency of the two methods for 2 000 realizations of Texas site

重构方法	估计不同函数的耗时/s			模拟 X_{3D} 耗 时/s	总耗时 /s
	$f(\underline{\mu}_s, C_s, \underline{a}, X^u X^o)$	$f(\underline{\mu}_{db}, C_{db}, \underline{a}_{db}, X_{db}^u X_{db}^o)$	$f(x X^o, X_{db}^o)$		
MUSIC-3X方法	13 853	—	—	462	14 315
本文所提方法	13 853	3 843	967	462	19 125

4 结 论

本文提出了一种基于大数据融合技术的稀疏测量场地三维概率化重构方法,通过吉布斯抽样分别估计特定场地考虑土体参数三维空间变异性条件的土体参数联合概率密度函数和数据库土体参数联合概率密度函数,采用变种贝叶斯理论融合特定场地土体参数概率密度函数与数据库土体参数概率密度函数,根据融合后分布模拟不完备钻孔数据,在此基础上生成三维场地概率化重构模型。最后,以一个虚拟场地案例和美国德克萨斯州某一实际场地案例阐明了本文所提方法的有效性。主要结论如下:

(1) 与MUSIC-3X方法相比,本文所提方法估计的土体参数不确定性显著降低,且随着勘察数据量的减少,本文所提方法在不确定性折减方面的优势更为明显。

(2) 场地勘察数据量对三维场地不确定性具有较大影响,数值分析案例表明存在一个临界观测数据量,约为20个数据,当勘察数据量大于临界勘察数据量时,勘察数据量的增加对场地不确定性影响逐渐降低。

(3) 本文所提方法能有效地融合岩土体大数据信息,为稀疏测量场地的土体参数不确定性量化提供一种有效分析工具。融合岩土体大数据重构三维场地模型能够显著地降低三维稀疏测量场地的不确定性。

参 考 文 献

- [1] 蒋水华, 曾绍慧, 潘嘉铭, 等. 岩土体参数约束随机场解析模拟方法及边坡可靠度分析[J]. 岩石力学与工程学报, 2018, 37(3): 642-651.
- [2] JIANG Shui-hua, ZENG Shao-hui, PAN Jia-ming, et al. Slope reliability analysis based on analytical simulation of conditional random fields of rock and soil parameters[J]. *Chinese Journal of Rock Mechanics and Engineering*, 2018, 37(3): 642-651.
- [3] XIAO T, LI D Q, CAO Z J, et al. CPT-based probabilistic characterization of three-dimensional spatial variability using MLE[J]. *Journal of Geotechnical and Geoenvironmental Engineering*, 2018, 144(5): 04018023.
- [4] ZHAO T, WANG Y. Statistical interpolation of spatially varying but sparsely measured 3D geo-data using compressive sensing and variational Bayesian inference[J]. *Mathematical Geosciences*, 2021, 53(6): 1171-1199.
- [5] 黄天朗, 宫凤强, 伍婷玉. 小样本岩土参数概率分布的正交多项式推断方法[J]. 工程力学, 2017, 34(8): 85-95.
- [6] HUANG Tian-lang, GONG Feng-qiang, WU Ting-yu. Orthogonal polynomial inference method of the probability distribution function for small samples of geotechnical parameters[J]. *Engineering Mechanics*, 2017, 34(8): 85-95.
- [7] 陈旺旺, 李典庆, 唐小松, 等. 抗剪强度参数概率分布的最大熵估计及边坡可靠度分析[J]. 岩土力学, 2018, 39(4): 1469-1478.
- [8] CHEN Wang-wang, LI Dian-qing, TANG Xiao-song, et al. Probability distribution of shear strength parameters using maximum entropy principle for slope reliability analysis[J]. *Rock and Soil Mechanics*, 2018, 39(4): 1469-1478.
- [9] 田密, 张帆, 李丽华. 间接测量数据条件下岩土参数空间变异性定量分析方法对比研究[J]. 岩土力学, 2018, 39(12): 4673-4680.
- [10] TIAN Mi, ZHANG Fan, LI Li-hua. Comparative study on the quantitative analysis methods of inherent spatial variability of soil properties based on indirect test data[J]. *Rock and Soil Mechanics*, 2018, 39(12): 4673-4680.
- [11] 李典庆, 吕天健, 唐小松. 基于多维Gaussian Copula的岩土体设计参数概率转换模型构建方法[J]. 岩土工程学报, 2021, 43(9): 1592-1601.
- [12] LI Dian-qing, LÜ Tian-jian, TANG Xiao-song. Establishing probabilistic transformation models for geotechnical design parameters using multivariate Gaussian Copula[J]. *Chinese Journal of Geotechnical Engineering*, 2021, 43(9): 1592-1601.
- [13] CHING J Y, PHOON K K. Constructing site-specific multivariate probability distribution model using Bayesian machine learning[J]. *Journal of Engineering Mechanics*, 2019, 145(1): 04018126.
- [14] CHING J Y, PHOON K K. Constructing a site-specific multivariate probability distribution using sparse, incomplete,

- and spatially variable (MUSIC-X) data[J]. **Journal of Engineering Mechanics**, 2020, 146(7): 04020061.
- [10] 杨智勇, 尹城川, 聂家岩, 等. 考虑场地多源勘测数据三维空间相关性的土体参数概率密度函数估计[J]. 岩土力学, 2022, 43(6): 1571-1584.
- YANG Zhi-yong, YIN Cheng-chuan, NIE Jia-yan, et al. PDF estimation of geotechnical parameters considering the three-dimensional spatial variability based on multi-source site investigation data[J]. **Rock and Soil Mechanics**, 2022, 43(6): 1571-1584.
- [11] HAN L, WANG L, ZHANG W, et al. Quantification of statistical uncertainties of unconfined compressive strength of rock using Bayesian learning method[J]. **Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards**, 2022, 16(1): 37-52.
- [12] CHING J, PHOON K K, WU S H. Impact of statistical uncertainty on geotechnical reliability estimation[J]. **Journal of Engineering Mechanics**, 2016, 142(6): 04016027.
- [13] HU J Z, ZHANG J, HUANG H W, et al. Value of information analysis of site investigation program for slope design[J]. **Computers and Geotechnics**, 2021, 131: 103938.
- [14] International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE). TC304 304dB database[DB/OL]. (2021-12-6)[2024-8-30]. http://140.112.12.21/issmge/Database_2010.htm.
- [15] CHING J, WU S, PHOON K K. Constructing quasi-site-specific multivariate probability distribution using hierarchical Bayesian model[J]. **Journal of Engineering Mechanics**, 2021, 147(10): 04021069.
- [16] CHING J Y, PHOON K K, YANG Z Y, et al. Quasi-site-specific multivariate probability distribution model for sparse, incomplete, and three-dimensional spatially varying soil data[J]. **Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards**, 2022, 16(1): 53-76.
- [17] SHARMA A, CHING J, PHOON K K. A hierarchical Bayesian similarity measure for geotechnical site retrieval[J]. **Journal of Engineering Mechanics**, 2022, 148(10): 04022062.
- [18] 张德, 张泽超, 张璐璐, 等. 场地有限数据条件下土体不排水抗剪强度的概率分布的贝叶斯估计研究[J]. 岩土工程学报, 2023, 45(6): 1259-1268.
- ZHANG De, ZHANG Ze-chao, ZHANG Lu-lu, et al. Bayesian estimation of probability distributions of undrained shear strength of soils with limited site data[J]. **Chinese Journal of Geotechnical Engineering**, 2023, 45(6): 1259-1268.
- [19] TABARROKI M, CHING J, YUAN S H, et al. Data-driven hierarchical Bayesian model for predicting wall deflections in deep excavations in clay[J]. **Computers and Geotechnics**, 2024, 168: 106135.
- [20] 吕大刚. 基于线性化 Nataf 变换的一次可靠度方法[J]. 工程力学, 2007, 24(5): 79-86.
- LÜ Da-gang. First order reliability method based on linearized Nataf transformation[J]. **Engineering Mechanics**, 2007, 24(5): 79-86.
- [21] ZHOU Y, ZHANG D, HUANG H, et al. Effect of normal transformation methods on performance of multivariate normal distribution[J]. **ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering**, 2022, 8(1): 04021074.
- [22] BOX G E P, COX D R. An analysis of transformations[J]. **Journal of the Royal Statistical Society: Series B (Methodological)**, 1964, 26(2): 211-252.
- [23] CHING J Y, PHOON K K. Correlations among some clay parameters—the multivariate distribution[J]. **Canadian Geotechnical Journal**, 2014, 51(6): 686-704.
- [24] YANG Z Y, CHING J Y. Simulation of three-dimensional random field conditioning on incomplete site data[J]. **Engineering Geology**, 2021, 281: 105987.
- [25] YANG Z Y, LI X Y, QI X H. Efficient simulation of multivariate three-dimensional cross-correlated conditional random fields[J]. **Computer Methods in Applied Mechanics and Engineering**, 2022, 388: 114208.
- [26] YANG H Q, ZHANG L, PAN Q, et al. Bayesian estimation of spatially varying soil parameters with spatiotemporal monitoring data[J]. **Acta Geotechnica**, 2021, 16: 263-278.
- [27] LI X Y, ZHANG Y X, YANG Z Y, et al. Estimation of site-specific multivariate probability distribution of soil properties using a mixed sampling technique[J]. **Computers and Geotechnics**, 2024, 166: 105956.
- [28] CROUSE D F, WILLETT P, PATTIPATI K, et al. A look at Gaussian mixture reduction algorithms[C]//14th International Conference on Information Fusion. Chicago: IEEE, 2011: 1-8.
- [29] CUI S. Comparison of approximation methods to Kullback-Leibler divergence between Gaussian mixture models for satellite image retrieval[J]. **Remote Sensing Letters**, 2016, 7(7): 651-660.
- [30] LI D Q, XIAO T, ZHANG L M, et al. Stepwise covariance matrix decomposition for efficient simulation of multivariate large-scale three-dimensional random fields[J]. **Applied Mathematical Modelling**, 2019, 68: 169-181.