



建筑结构学报
Journal of Building Structures
ISSN 1000-6869, CN 11-1931/TU

《建筑结构学报》网络首发论文

题目: 基于聚类分析的我国工程场地分类方案优化研究
作者: 任叶飞, 刘也, 张鹏, 冀昆, 王宏伟, 温瑞智, 谢俊举
DOI: 10.14006/j.jzjgxb.2022.0273
收稿日期: 2022-05-05
网络首发日期: 2022-10-10
引用格式: 任叶飞, 刘也, 张鹏, 冀昆, 王宏伟, 温瑞智, 谢俊举. 基于聚类分析的我国工程场地分类方案优化研究[J/OL]. 建筑结构学报.
<https://doi.org/10.14006/j.jzjgxb.2022.0273>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于聚类分析的我国工程场地分类方案优化研究

任叶飞^{1,2}, 刘也^{1,2}, 张鹏^{1,2}, 冀昆³, 王宏伟^{1,2}, 温瑞智^{1,2}, 谢俊举⁴

(1. 中国地震局工程力学研究所, 地震工程与工程振动重点实验室, 黑龙江哈尔滨 150080; 2. 地震灾害防治应急管理
部重点实验室, 黑龙江哈尔滨 150080; 3. 河海大学, 土木与交通学院, 江苏南京 210098;
4. 中国地震局地球物理研究所, 北京 100081)

摘要: 在抗震设计规范中, 场地分类方案是一项重要内容。针对目前广泛关注的我国现行规范中 II 类场地定义范围是否过于宽泛的问题, 利用日本 KiK-net 台网的强震动记录和台站钻孔资料, 基于地震动预测残差分析方法计算台站的真实场地放大因子, 引入 K-Means 聚类算法根据台站真实场地放大因子对 II 类场地进行聚类分析, 给出建议的场地分类优化方案。得到如下结论: (1) II 类场地的地震动放大因子离散性较大, 存在与覆盖层厚度相关的明显分簇现象, 验证了 II 类场地定义范围存在过于宽泛的问题, 也为聚类分析提供了先决条件; (2) 当 II 类场地的台站被划分为 2 个簇类时, 聚类结果最优; 两个簇类的场地放大因子曲线差异明显、覆盖层厚度总体上呈分簇现象, 验证了聚类结果的可靠性。 (3) 建议以覆盖层厚度 20m 作为分界指标将 II 类场地细分为 II_a 和 II_b 两个子类, 两个子类的场地放大因子具有较好的类间差异度与类内聚集度。研究结果对我国现行抗震规范场地分类方案的改进具有一定参考价值。

关键词: 场地分类方案; 场地放大因子; 抗震设计规范; K-Means 聚类算法; 覆盖层厚度
中图分类号: P315.9

Research on optimization of site classification scheme in Chinese seismic code based on clustering analysis

REN Yefei^{1,2}, LIU Ye^{1,2}, ZHANG Peng^{1,2}, JI Kun³, WANG Hongwei^{1,2}, WEN Ruizhi^{1,2}, XIE Junju⁴

(1. Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150080, China; 2. Key Laboratory of Earthquake Disaster Mitigation, Ministry of Emergency Management, Harbin 150080, China; 3. School of Civil and Transportation Engineering, Hohai University, Nanjing 210098, China; 4. Institute of Geophysics, China Earthquake Administration, Beijing 100081, China)

Abstract: It is of great importance that site classification scheme is usually included as part of seismic codes. It is debated widely whether the scope of site class II in current Chinese seismic codes is too broad or not. With this regard, in this study the strong motion records and borehole data from KiK-net seismograph networks are collected to calculate the site amplification factors using the method of residual analysis of ground motion model. Then the K-means clustering algorithm is introduced to implement a clustering analysis on stations of site class II by setting the site amplification factors as explanatory variables. An optimized site classification scheme is suggested finally. The results show that: (1) the site amplification factors of Class II sites appear great variability and clustered obviously with respect to soil covering depth, that implies broad scope of definition of site class II and provides a prerequisite for cluster analysis; (2) it is recognized that the optimum number of clusters is two for the best clustering site class II stations; the site amplification curves of two clusters are significantly different and their soil covering depths are clustered obviously, that validate the reliable results of K-means clustering; (3) it is recommended that the site class II could be subdivided into two sub-classes II_a and II_b considering the 20 m of soil covering depth as a threshold value, that both reflect good performance of inter-class aggregation and intra-class variation of site amplification factors. The results achieved by this study will provide a beneficial suggestion to modify the site classification scheme in the current seismic codes of China.

基金项目: 国家重点研发计划项目 (2018YFE0109800); 国家自然科学基金 (51878632); 黑龙江省自然科学基金优秀青年项目 (YQ2019E036); 黑龙江省头雁行动计划(无项目号)

作者简介: 任叶飞, 出生于 1983 年, 男, 博士, 研究员, 主要从事工程地震相关研究。 E-mail: renyefei@iem.net.cn

通信作者: 温瑞智, 出生于 1968 年, 男, 博士, 研究员, 主要从事工程地震相关研究。 E-mail: ruizhi@iem.ac.cn

收稿日期: 2022-05-05

Keywords: Site classification scheme; Site amplification factor; Seismic code; K-Means clustering algorithm; Soil covering depth

0 引言

强震动观测和理论研究表明,不同的场地对地震动的影响存在一定的差别。各国的抗震规范中均引入了场地类别的概念,通常依据场地基本周期、等效剪切波速等能够表征场地地震效应的参数作为划分依据,工程师进行结构抗震设计时,依据场地类别调整设计地震动参数已成为国际通用的方法^[1]。我国抗震设计规范自颁布以来已经历8次修订更新^[2],在最初的建筑抗震设计规范(59 规范)^[3]中并未给出明确的场地分类方案,随后的几次修订仅通过对岩土特性的宏观描述进行场地类别的划分,在第五次修订(89 规范)^[4]中首次通过定量的方式划分场地类别,提出利用土层平均剪切波速与覆盖层厚度确定场地类别的分类方案,并将场地划分为I、II、III和IV类四个类别。第六次修订(01 规范)^[5]在沿用上一版场地类别划分结果的基础上对剪切波速和覆盖层厚度计算的具体规定上做了改进,提出使用更具有物理意义的等效剪切波速代替平均剪切波速,并将土层计算深度从15 m提高到20 m,形成了沿用至今的采用“覆盖层厚度(H)”和“土层等效剪切波速(V_{se})”作为划分依据的双指标场地分类方案。现行的《建筑抗震设计规范》(GB50011-2010)^[6]与01 规范相比仅将 V_{se} 上限增加至800 m/s, I类场地细分为 I_0 与 I_1 两个子类。在01 规范颁布时考虑到经济与分类方法可靠度等方面的因素,有意识的扩大了II类场地的范围^[7]。通常,规范的制定与实施需要兼顾当前社会经济水平对规范要求的可承受能力。近二十年我国经济水平已显著提升,现行规范的场地分类方案能否适应当前工程应用的需求,尤其是关于II类场地划分范围的合理性已引起了广泛的讨论。例如,周健^[8]等从实际工程钻孔土层信息的角度,指出II类场地涵盖了大量存在很大土层差异的钻孔;陈国兴^[9]等提出我国规范的II类场地覆盖美国的C、D和E类,导致被划分为II类的场地实际场地反应离散性很大。Ji^[10]等通过HVSr方法识别场地类别时指出由于II类宽泛范围,单条平均曲线难以代表II类场地的HVSr特征。虽然大量的研究探讨了规范中II类场地的划分范围是否过于宽泛的问题,但鲜有研究从地震中真实表现的场地反应角度给出科学合理的优化建议。

针对上述问题,本文引入一种基于机器学习的聚类算法根据地震真实场地反应确定场地分类方案。聚类分析是数据挖掘的主要方法之一,通过无监督的机

器学习过程,将未被标记的相似样本划分到同一集群中,使得类内差异度较小,类间差异度较大^[11]。目前聚类分析已在地震工程领域取得了诸多应用,例如,Chen^[12]等基于地震到时前后可作为单独分类的思想提出一种利用聚类算法检测地震波到时的方法; Riahi 和 Gerstoft^[13]等利用图形聚类方法在密集传感器阵列中定位震源位置。Kotha^[14]等利用数据驱动的思想借助聚类算法给出了将工程场地划分为8个类别的方案; Ji^[15]等利用K-Means 聚类算法实现了非线性场地反应的有效识别。规范中场地分类方案的作用是使场地效应相近的场地定义为相同类,并对场地效应差异较大的场地区分为不同类,这恰好与聚类算法的功能相吻合,这种相似性为聚类分析在场地分类方案中的优化研究奠定了坚实的理论基础。

如何合理量化场地反应是开展这项研究工作需要解决的首要问题。以往研究大多采用数值计算的方式求解场地地震反应^[9]。这种方法针对单个场地具有一定优势,但如果是针对大量场地,则将因场地岩性、波速和土动力参数等资料获取困难而受到限制。研究发现,基岩和土层场地的强震动观测记录对比分析是探讨场地效应的最直接方法^[1]。基于不同场地上的强震动记录资料,可以揭示场地条件对地震动的影响特征和规律。刘也^[16]等提出了一种基于观测地震动残差分析的场地放大因子计算方法,其优势是可以获得大量强震动台站在地震中的真实场地反应,为实现场地反应的聚类分析提供了前提条件。

本文利用日本KiK-net 台网的强震动记录和台站钻孔资料,按照我国《抗震设计规范》(GB50011-2010)进行场地分类,基于地震动预测残差分析方法计算台站的真实场地放大因子,引入K-Means 聚类算法针对台站真实场地放大因子、覆盖层厚度和等效剪切波速对II类场地进行聚类分析,最终依据聚类结果给出优化II类场地划分方案的建议。

1 场地放大因子

1.1 强震动数据集

日本KiK-net 台网台站密集,记录了大量高质量强震动数据,且提供了详实的台站场地资料,为计算不同的场地特征参数(覆盖层厚度,等效剪切波速)以及求解地震中真实的场地放大因子提供了便利。本文采用刘也^[16]等选取的2012-2018年KiK-net 台网从131个地震

事件中由 424 个台站收集到的 10731 组强震动记录作为研究数据集；它们对应的矩震级范围为 5.0-7.0，震源距范围为 13-300km。这 131 个地震事件按震源机制划分包含 87 个逆断层地震、30 个正断层地震、14 个走滑断层地震；按构造类别划分包含 47 个浅层地壳地震、21 个俯冲板间地震、63 个俯冲板内地震。图 1(a)和(b)给出了这些记录的矩震级与震源距的对应分布关系及记录对应震级的直方分布，(c)和(d)给出了 PGA 与震源距的对应分布关系及记录数量对应 PGA 的直方分布。需要说明的是，在 10731 组记录中有 148 组记录的 PGA 超过了 100gal，占总数的 1.3%，也就是说仅有少量的强震动记录中包含可能的场地非线性反应。因此，本文计算的场地放大因子主要体现的是场地的线性反应。依据我国场地分类标准，数据集包含的 424 个台站中，5 个台站被划分为 I₀ 类场地，62 个台站被划分为 I₁ 类场地，322 个台站被划分为 II 类场地，32 个台站被划分为 III 类场地，3 个台站被划分为 IV 类场地。其中位

于 II 类场地的台站数量最多，占总数的 75.94%，超过了位于其他四个类别场地的台站数量总和。在本文中将重点关注 II 场地的放大因子分布。

1.2 场地放大因子计算方法

地震动预测模型(GMM)通常综合考虑了震源、路径和场地效应对地震动的影响，给出较为可靠的地震动中位预测值。利用适用于目标场址的 GMM，通过对实际强震动观测数据进行残差分析可分离震源、路径和场地的影响，对场地放大因子进行近似估计^[16]：

$$\ln(A_j) = \frac{1}{N} \sum_{i=1}^N [\ln(Y_{ij}) - \ln(f(M, R, HR)) - \eta_i - \phi_{path}] \quad (1)$$

其中， A_j 为第 j 个台站的近似平均场地放大函数； N 为第 j 个台站收集到的强震动记录总数； Y_{ij} 为第 i 个地震事件在第 j 个台站记录到的地震动强度指标观测值(如 PGA、PGV、不同周期的反应谱值等)；

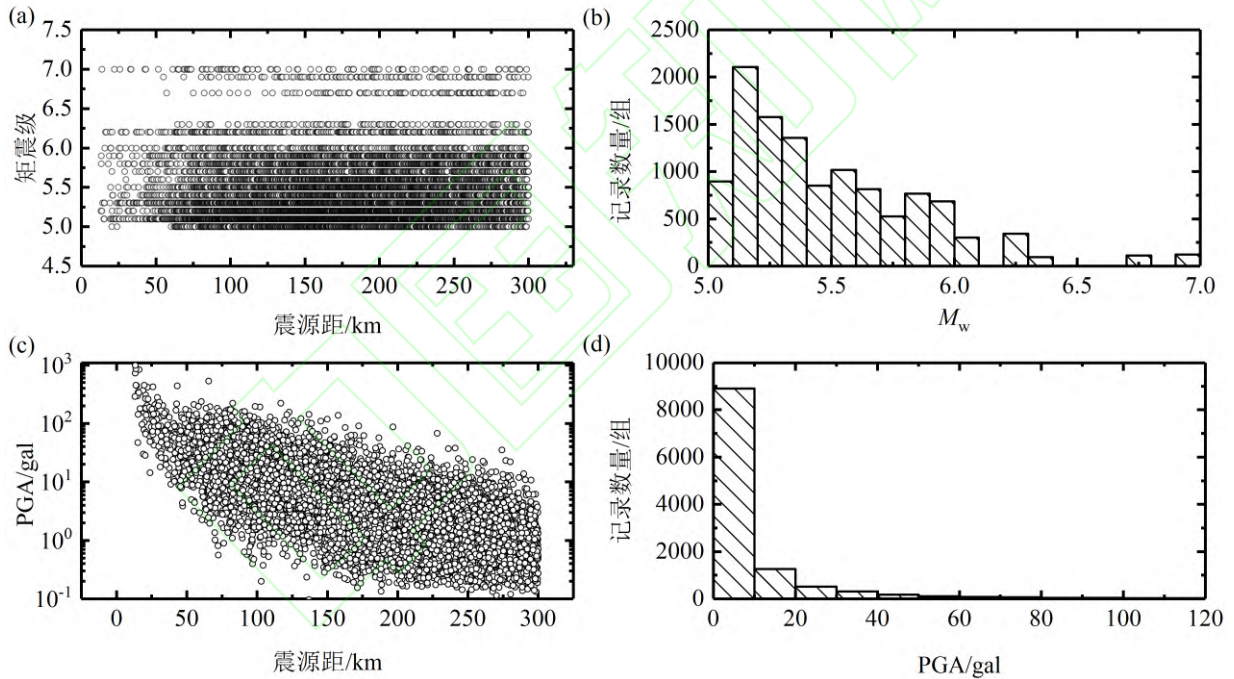


图 1 本研究采用的强震动记录对应的(a)震源距与矩震级对应分布；(b)记录数量的震级直方分布；(c)震源距与 PGA 对应分布；(d)记录数量的 PGA 直方分布

Fig.1 (a) Earthquake magnitude vs. hypocentral distance; (b) histogram of earthquake magnitude; (c) PGA vs. hypocentral distance; (d) histogram of PGA for the selected strong motion records in this study

$f(M, R, HR)$ 为硬基岩场地条件下的预测值，本文选用与强震动数据分布区域相一致的 Zhao2006 地震动预测模型^[17]计算地震动强度指标的预测值； η_i 为第 i 个地震事件的事件间残差； ϕ_{path} 为路径修正函数，当强震记录的震源距大于 80km 时，地震动受区域非弹性衰减差异的影响较大，需消除由路径效应引起的非弹性衰减造成

的预测偏差。参考 Boore^[18]等的研究，路径修正函数可表示为与震源距相关的函数：

$$\phi_{path} = \Delta c_3 \cdot R + \Delta \phi \quad (2)$$

其中， R 为震源距， Δc_3 、 $\Delta \phi$ 为回归系数。

限于篇幅原因，针对 GMM 的输入参数、预测值计

算、以及上述 ϕ_{path} 和 A_f 的计算这里不再赘述，也非本文关注重点，详细过程可参见文献[16]。

1.3 场地放大因子分布

图2以 $S_a(T=0.1)$ 为例展示了 I_1 、II 和 III 类场地台站的放大因子分布；由于数据集中 I_0 类台站与 IV 类台站过少，统计意义有限，因此这里不作展示。图中可见，场地放大因子体现出随场地变软而逐渐增大的趋势。不同于 I_1 类和 III 类场地的放大因子主要分布在 0.6-1.7 和 1.7-3.3（25%分位值-75%分位值）较小的范围内，II 类场地由于其定义的宽泛范围，场地放大因子出现了较大的离散性，放大因子最大值与最小值之差达到 12.4。放大因子为 12.7 的 MYGH07 台站与放大因子为 0.3 的 NIGH17 台站均被划分为 II 类，若对这两个场地上的相同结构进行抗震设计，将采用相同的设计反应谱（考虑相同的设防烈度和设计分组的前提下），显然会引起其中的一个结构设计偏保守、或引起另一结构设计偏不安全。

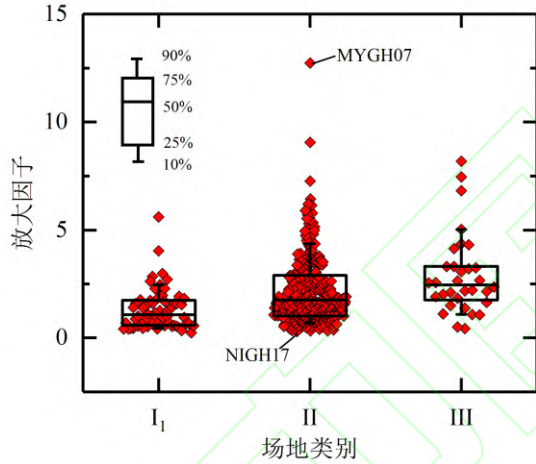


图2 不同场地类别内 $S_a(T=0.1s)$ 的场地放大因子分布
Fig.2 Site amplification factors for $S_a(T=0.1s)$ associated different site class

图3给出了6个典型地震动强度指标的场地放大因子与 H 和 V_{se} 的对应关系分布，可以发现无论哪个地震动强度指标，均体现出整体上场地放大因子随覆盖层厚度增加及等效剪切波速降低而逐渐增大的趋势。图3可见，II类场地的放大因子存在分簇现象，覆盖层

厚度较小部分场地放大因子相对较小与 I 类场地更为接近，覆盖层厚度较大的部分场地放大因子相对较大与 III 类场地更为接近。倘若将 II 类场地进一步细分，可一定程度解决 II 类场地划分范围宽泛，场地放大因子离散性较大的问题。下一节中将借助 K -Means 聚类算法探究 II 类场地细分的数量及分类界限。

2 K -Means 聚类分析

2.1 K -Means 聚类原理

聚类分析在对数据不作任何假设的条件下，用数学方法研究和处理所给样本的分类以及各类之间的亲疏程度，按照某种相似性度量，将样本分成多个类或簇，同一簇中的对象具有较高的相似度，而不同簇中的样本差别较大，属于一种无监督机器学习方法。目前聚类分析已被广泛的应用于各种科学和工程领域^[19]。

K -Means 算法是聚类算法中最流行的算法，相较于其他的聚类算法， K -Means 算法以效果较好、思想简单的优点在聚类算法中得到了广泛的应用^[20]。对于一个给定的包含 n 个 d 维数据点的数据集 $X=\{x_1, x_2, \dots, x_i, \dots, x_n\}$ ，其中 $x_i \in R^d$ (d 维实数集)， K -Means 聚类算法将数据对象组织为 K 个分簇 $C=\{c_k, k=1, 2, \dots, K\}$ 。每个分簇代表一个类 c_k ，每个类 c_k 具有一个聚类中心 μ_k 。算法需要预先指定初始聚类数目 K 以及 K 个初始聚类中心，根据数据对象与聚类中心之间的相似度，不断更新聚类中心的位置，不断降低类的误差平方和 (Sum of Squared Error, SSE)，当 SSE 不再变化或目标函数收敛时，聚类结束，得到最终结果。通常可采用欧氏距离作为相似性的判别准则，计算该类内各点到聚类中心 μ_k 的距离 D ：

$$D(x, \mu_k) = \sqrt{\sum_{j=1}^d (x_j - \mu_{kj})^2} \quad (3)$$

其中， x 为数据对象， d 为数据对象的维度， x_j 和 μ_{kj} 为数据对象和聚类中心的第 j 维的值。整个数据集的 SSE 计算公式为：

$$SSE = \sum_{k=1}^K \sum_{x \in c_k} |D(x, \mu_k)|^2 \quad (4)$$

需要注意的是 K -Means 算法是一个贪心算法，仅

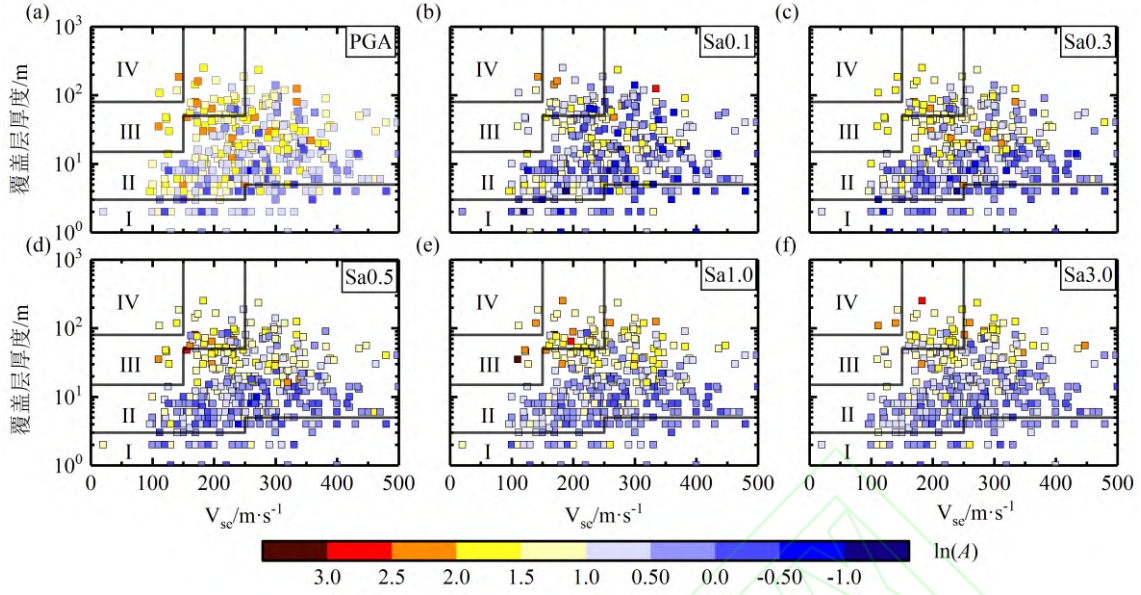


图 3 (a)PGA 和(b-f)Sa(T=0.1s、0.3s、0.5s、1s、3s)场地放大因子（对数刻度下）与覆盖层厚度及 V_{se} 的对应关系分布

Fig.3 Correlations of site amplification factors (in logarithmic scale) with H and V_{se} for (a) PGA and (b-f) Sa(T=0.1s, 0.3s, 0.5s, 1.0s and 3.0s)

能获取局部最优，不同的初始聚类中心得到的最终局部最优结果是不同的，本文采用随机重启的初始聚类中心的选取方法^[21]，通过对 K -means 算法的 100 次重启（即 $N_R=100$ ），得到 100 次不同初始聚类中心的聚类结果，选取 SSE 最小的一次聚类结果，确保得到全局最优解，具体流程如图 4 所示。

上节中已得到 KiK-net 台网中 322 个 II 类场地台站不同强度指标的场地放大因子。为了借助 K -Means 方法进一步将我国抗震规范中 II 类场地的范围进一步细分，本文将每个 II 类台站的等效剪切波速、覆盖层厚度、PGA 以及不同周期反应谱值的放大因子视为数据对象进行 K -Means 聚类分析。

2.2 数据标准化

在进行 K -Means 聚类分析前，数据的归一化处理是一个必要的过程。当样本各维度变量的量级不一致时，在聚类过程中小量级变量的作用可能被大量级变量所淹没，使聚类结果主要受大量级变量主导，从而产生错误的结果。在本文中等效剪切波速、覆盖层厚度、PGA 以及不同周期反应谱值的放大因子组成了样本中各维度的变量，其中等效剪切波速及覆盖层厚度与放大因子间量级相差巨大，因此采用 Z-score 标准化方法，将数据转化为无量纲的数值，并统一数据的量级。

$$x'_i = \frac{x_i - \bar{x}}{\sigma} \quad (5)$$

其中 x'_i 为变量 x_i 归一化后的结果， $\bar{x}$ 为变量均值，

σ 为变量标准差。

2.3 聚类个数 K 最优确定

在 K -Means 算法中，聚类个数 K 需要事先确定，是无监督聚类算法中唯一需要的先验知识^[22]。 K 值的大小将对聚类结果产生较大影响，若 K 值过小，则会导致同一类别内数据对象差异很大，若 K 值过大则失去分类的意义。同时， K 值选取不当也会使最终的聚类结果陷入局部最优。通过评价不同聚类个数下聚类结果的内部有效性是确定最佳聚类个数的常用做法。

聚类结果中同类样本的相似性与不同类样本的差异度统称为聚类结果的内部有效性。若某一聚类个数下，聚类结果同类相似性最高，不同类相似性最低，则该聚类个数即为最佳聚类个数。常用于确定最佳聚类个数的内部有效性评价指标有 CH 指标、DB 指标、BWP 指标和 Sil 指标等^[23]。本节选取最为广泛使用的 CH 指标与 DB 指标通过评价聚类结果的内部有效性确定最佳聚类个数 K ，CH 指标和 DB 指标均是通过样本与聚类中心间的欧式距离表示同一簇类中样本的紧密程度与不同簇类之间样本分散程度的函数，CH 值越大、DB

值越小代表聚类质量越好。

图 5(a)和(b)分别给出了 CH 指标与 DB 指标随聚类个数的变化趋势,当聚类个数为 2 类时,CH 指标最大和 DB 指标最小,两种不同的内部有效性评价指标均表明当聚类个数为 2 时,聚类质量最好。

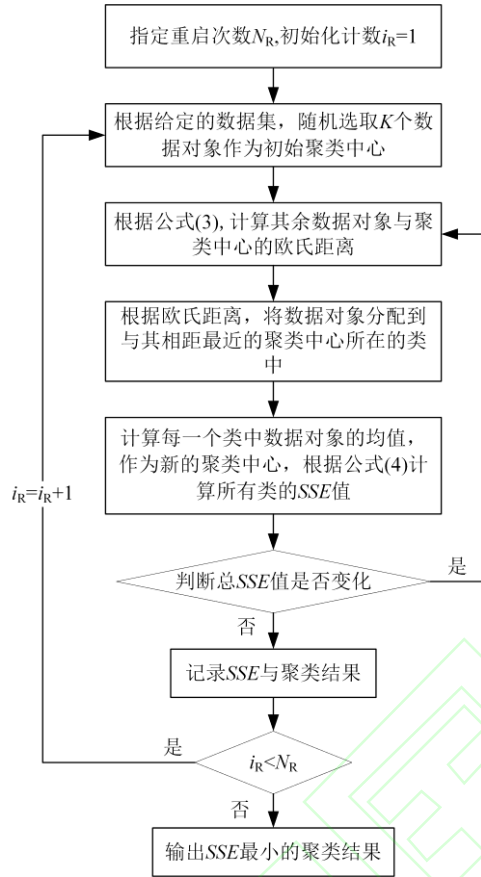


图 4 K-Means 算法聚类流程

Fig.4 Steps of K-Means clustering algorithm

根据不同聚类个数时 SSE 的变化趋势,选择“肘点”对应的聚类个数作为最优聚类个数也是一种常用的确定最优聚值的判断方法^[24]。当聚类个数增加时,SSE 随聚类个数的增加逐渐降低,SSE 下降速率发生突变的点称为“肘点”,在“肘点”后 SSE 的下降速率变得缓慢,意味着此时聚类个数的增加难以显著的提升聚类效果,因此“肘点”处的聚类个数可认为是最优值。图 5(c)给出了 SSE 随聚类个数的变化趋势,当聚类个数为 2 时,SSE 变化曲线出现“肘点”。综合基于内部有效性评价指标和基于 SSE 的评价的结果,我们确定针对本文聚类分析工作,选用聚类个数 2 作为最优 K 值。

2.4 聚类结果分析

数据集中的 322 个 II 类场地台站,通过上述 K-Means 聚类算法被划分为 2 个簇类,其中簇类 1 共包含 147 个台站,簇类 2 共包含 175 个台站,如图 6(a)。由于数据集中每个样本的维度过多,难以通过可视化的方式展示所有维度的数据,在这里通过主成分分析(Principal Component Analysis, PCA)方法,在尽量保证样本信息量不丢失的前提下,将高维数据映射到低维空间中^[25]。PCA 方法通过正交投影的方式,将原始特征往具有最大投影信息量的维度上进行投影,保留原始数据信息最多的维度称为主成分 1,保留原始数据信息第二多的维度称为主成分 2,以此类推。本文用于 K-Means 聚类的数据集中主成分 1 和主成分 2 的贡献率分别为 59.2%和 23.2%,前两个主成分即可反映超过 80%的数据信息。如图 6(b)所示,将原始数据投影到主成分 1 和主成分 2 组成的二维空间中,从视觉上清晰可见 K-Means 算法将 322 个 II 类场地台站划分为两个相对独立的簇群,得到了较为满意的分类结果。

图 6(c)给出了聚类后不同周期 Sa 的场地放大因子分布,图中红色与蓝色线条分别代表聚类结果中的簇类 1 与簇类 2,粗实线代表每个类别对数刻度下的场地放大因子均值,粗划线代表均值加减一倍的标准差。经过 K-Means 聚类算法分类后,反应谱放大因子曲线被划分为具有明显差异的两簇,对于 PGA 和 0.05-0.3s 周期段两个簇类之间 PGA 和加速度反应谱放大因子出现轻微的交叉,但随着反应谱周期的增大,两个簇类间的交叉逐渐减少,尤其是周期为 0.3-1.0s 中长周期段,两类间几乎不存在反应谱放大因子交叉现象。以覆盖着厚度和等效剪切波速作为场地特征参数的分类方法并未对土体的自然周期进行严格的限制,尤其是较为宽泛的 II 类场地,同一类别场地的卓越频率可能相差较大,导致采用 K-means 聚类算法对 II 类场地聚类时,各簇类的场地放大因子在 PGA 或部分周期段不可避免的出现交叉。场地条件对地震动的放大效应应具有较大不确定性,场地分类只能大致的对具有不同放大效应的局部场地进行区分,因此簇类 1 与簇类 2 之间在 PGA 和短周期段出现轻微的交叉是可以接受的。无论是主成分分析结果还是分类后的场地放大因子分布均表明 K-Means 聚类结果是合理可信的。

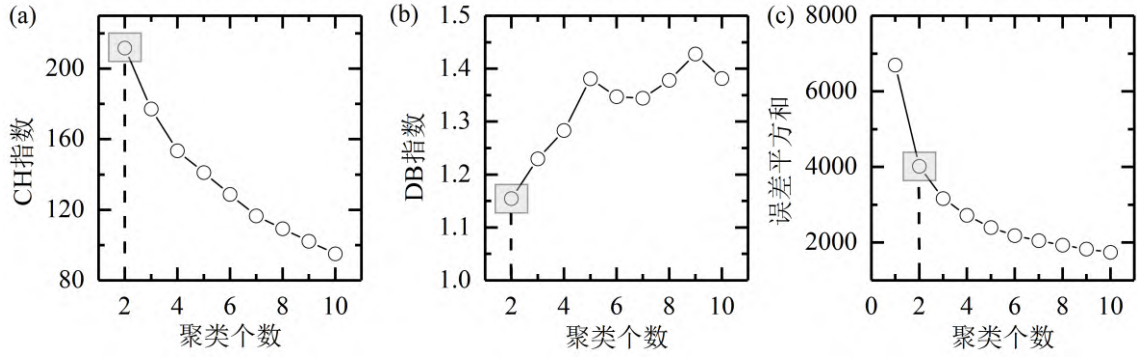


图5 聚类个数与(a)CH 指数、(b)DB 指数和(c)误差平方和的相关关系

Fig.5 Relationships between clustering number and (a) the CH index (b) DB index and (c) sum of squared error

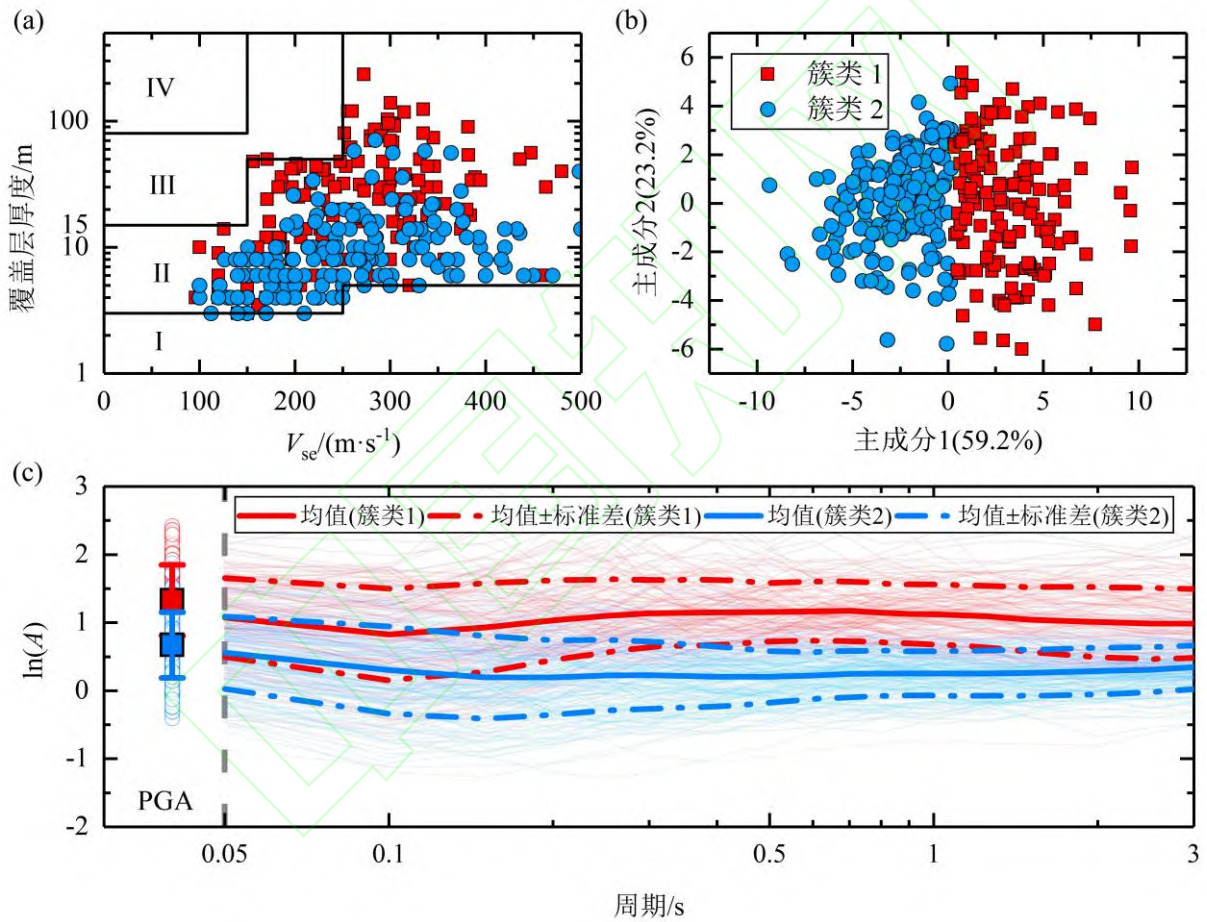


图6 聚类后得到2个分类的(a)场地 v_{se} 和覆盖层厚度的相关关系、(b)主成分分析结果和(c)不同周期反应的场地放大因子（对数刻度下）

Fig.6 (a) Correlation of soil covering depth and v_{se} ; (b) Result of principal component analysis; (c) Site amplification factors (in logarithmic scale) of spectral accelerations with different periods for two clusters of class-II sites respectively.

另外，从图 6(a)给出的 K -Means 聚类后各类别覆盖层厚度与等效剪切波速分布看，聚类结果与覆盖层厚度具有较强的相关性，簇类 1 中的绝大多数场地具有更厚的覆盖层厚度，簇类 2 则出现相反的趋势。

3 II 类场地划分优化及讨论

上文通过 K -Means 聚类算法依据不同周期谱加速度的场地放大因子将 II 类场地台站划分为两个簇类，

两类台站的场地放大效应表现出明显的类间差异,表明我国的 II 类场地具备进一步细分的理论依据。同时,聚类结果与覆盖层厚度间表现出的强相关性为 II 类场地的细分提供了可参考的分界指标。覆盖层厚度不仅能影响地震动的幅值,还能反映地震动的频谱特性。通常情况下,覆盖层越厚,反应谱长周期成份越卓越,相应的场地自振周期越长^[26-27]。覆盖层厚度作为场地分类指标的适用性与合理性已被诸多学者所证实^[28-30]。在我国规范给出的现行场地分类方案基础上,以覆盖层厚度作为 II 类场地细分的分界指标,将 II 类场地划分为 II_a 与 II_b 两个子类,不失为一种可行的优化方案。依据图 6(a)中给出的 K-Means 聚类结果与覆盖层厚度及等

效剪切波速的分布,覆盖层厚度 15m 可近似主观判断为两个子类的区分界限,因此暂定覆盖层厚度小于 15m 的 II 类场地划分为 II_a 类,覆盖层厚度大于 15m 的台站场地划分为 II_b 类。为了降低这种主观性,给出相对客观的 II 类场地细分的分界值,接下来分别选取 10m、15m、20m 分别作为 II 类场地细分的分界值,通过观察不同分界值划分的两个子类中的台站与前述聚类划分的 2 个簇类中的台站吻合程度,以及不同分界值的细分方案下同类别内场地放大因子的聚集程度和类别间的差异程度确定合适的分界值。图 7 中给出了不同分界值下 II_a 类与 II_b 类的两个子类场地数量及占比分布。图中可见,无论选取哪个分界值,II_a 类场地中绝大多数

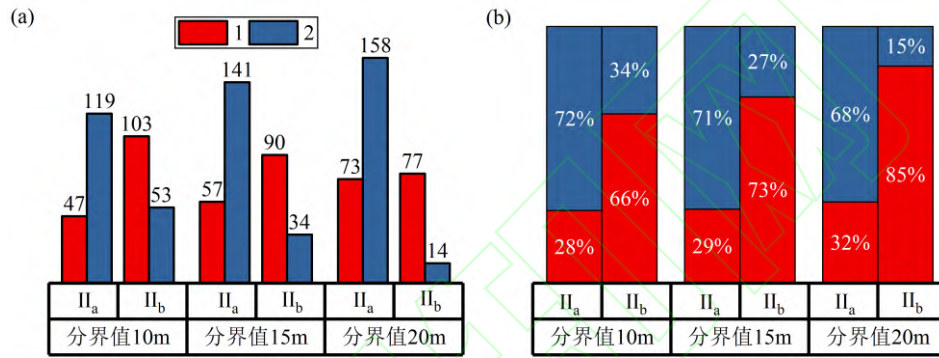


图 7 不同覆盖层厚度分界值的分类结果在(a)数量分布及(b)百分比分布上的对比

Fig.7 Comparison of clustering result with different threshold values in terms of (a) Number distribution and (b) percentage distribution

台站来自聚类结果中的第 2 簇类, II_b 类场地中则绝大多数台站来自第 1 簇类。随着分界值的提高, II_a 类场地的台站数量逐渐增加, II_b 类逐渐减少, II_a 类中占主导的第 2 簇类场地的占比逐渐减少(由 72%减少至 68%), II_b 类中占主导的第 1 簇类场地占比逐渐增大(由 66%显著增加至 85%)。后者增大的速率显著的超过了前者减少的速率。最终可见,以覆盖层厚度 20m 为分

界值的细分方案中, II_a 类场地中来自第 2 簇类的台站占比仅略小于其它两个方案,但 II_b 类场地中来自第 1 簇类的台站占比则显著超过了另两个方案,说明此时 II 类场地细分结果与 K-Means 聚类结果吻合程度最好,建议采用覆盖层厚度 20m 作为 II 类场地细分为 II_a 与 II_b 两个子类的界限值。

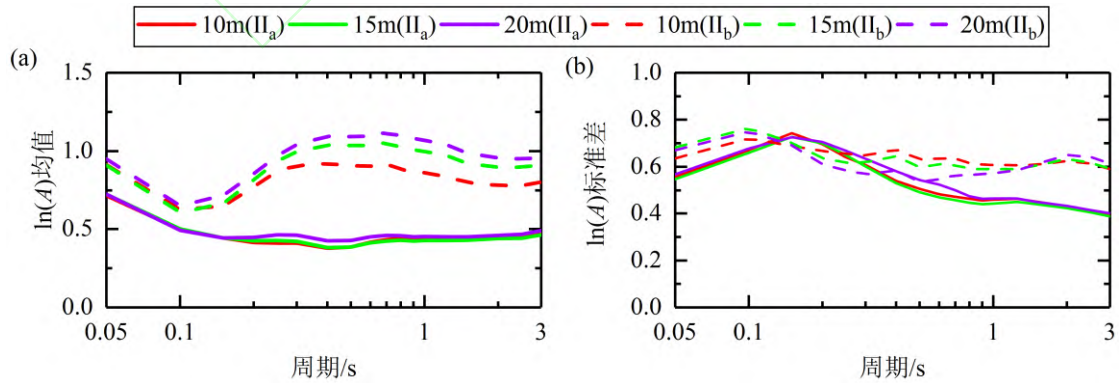


图 8 不同覆盖层厚度分界值的分类结果在(a)放大因子均值及(b)标准差上的对比(对数刻度下)

Fig. 8 Comparison of clustering result with different threshold values in terms of (a) the mean and (b) standard deviation of site amplification factors (in logarithmic scale)

图8中给出了不同分界值下II_a类与II_b类场地的各周期点反应谱的放大因子均值与标准差。从均值角度可见,无论采用哪个分界值,均能较好的体现II_a类与II_b类的场地反应差异性,其中当取分界值为20m时在各周期点两个子类别间的场地放大因子差异度最明显,当取10m时差异度则最小。从标准差角度可见,当分界值为20m时,II_a类场地反应标准差在周期0.2-1.0s范围内略大于其它方案;但II_b类标准差相较于其它方案在这个周期范围明显更小;其他周期段,各方案的标准差值都较为接近。因此,综合II_a类与II_b类的标准差对比结果,当分界值为20m时,同类别内场地放大因子的聚集度最优。

综上所述,以覆盖层厚度20m作为II类场地细分的分界值,既能较好的与聚类结果相吻合,又能表现出良好的类间差异性与类内聚集性。建议的优化后的场地分类方案如表1所示。图9给出了优化后的场地分类方案示意图,以覆盖层厚度20m为分界线,将II类场地细分为两个子类,蓝色阴影为II类中覆盖层厚度相对较浅的II_a类场地,红色阴影为覆盖层厚度相对较厚的II_b类场地。

需要注意,地震动区划和地震安全性评价中对场地条件的划分与我国建筑抗震设计规范一致,本文对场地分类优化的建议同样适用于地震动区划和地震安全性评价。

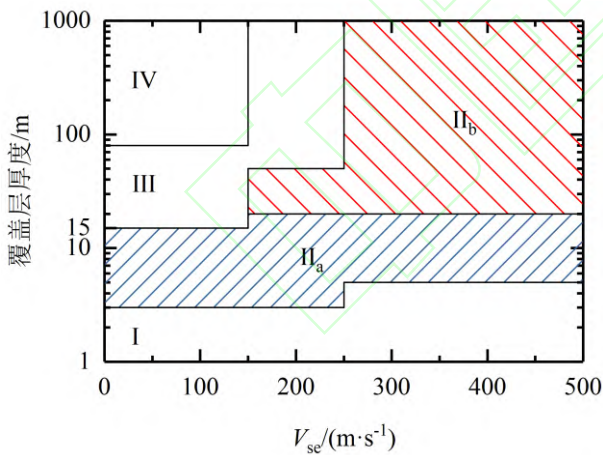


图9 推荐的场地分类优化方案示意图
Fig.9 Schematic drawing of recommended optimized scheme of site classification

表1 推荐的场地分类优化方案

剪切波速	场地类别的覆盖层厚度					
	I ₀	I ₁	II _a	II _b	III	IV
$v_s>800$	0					
$800\geq v_s>500$	0					
$500\geq v_{sc}>250$	<5					
$250\geq v_{sc}>150$	<3					
$v_{sc}\leq 150$	<3					

注： v_s 为岩石剪切波速， v_{sc} 为土的等效剪切波速

4 结论

针对我国规范场地分类方案中II类场地范围过于宽泛的问题,本文利用日本KiK-net台站强震动记录和场地钻孔资料,基于地震动预测残差分析方法计算了台站真实场地放大因子,引入K-Means聚类算法对II类场地进行聚类分析,得到如下结论:

(1) 从计算的场地放大因子分布情况可以发现,不同于I类和III类场地,II类场地的各IM场地放大因子均出现了较大的离散性,并且存在与覆盖层厚度相关的明显分簇现象,为聚类分析提供了先决条件。

(2) 依据CH与DB两种聚类内部有效性评价指标,结合SSE的表现,确定了最优聚类个数为2;将场地覆盖层厚度、等效剪切波速和PGA及不同周期点反应谱的场地放大因子作为数据对象,利用K-Means聚类算法将322个II类场地台站划分为两个独立的簇类;结果显示,两个簇类的场地放大因子曲线差异明显、覆盖层厚度总体上呈分簇现象,验证了聚类结果的可靠性。

(3) 分别以覆盖层厚度10m、15m和20m为界限值,将现行场地分类方案中的II类场地划分为II_a与II_b两个子类,通过分析两个子类中的台站与聚类划分的2个簇类中的台站吻合程度,以及同一子类内场地放大因子的聚集程度和不同子类间的差异程度确定合适的分界值为20m。最后,给出了将我国规范定义的II类场地细分为II_a和II_b两个子类的优化建议。抗震设计理论的修订与完善是一项系统性工作,本文研究仅从场地反应聚类结果的角度给出了场地分类方案的优化建议,期冀能为这方面的继续深入研究提供一些启示。

参 考 文 献

[1] 李小军. 地震动参数区划图场地条件影响调整[J]. 岩

- 土 工 程 学 报, 2013, 35(S2): 21-29.(LI Xiaojun. Adjustment of seismic ground motion parameters considering site effects in seismic zonation map[J]. Chinese Journal of Geotechnical Engineering, 2013, 35(S2): 21-29.(in Chinese))
- [2] 薄景山, 李琪, 孙强强, 等. 场地分类研究现状及有关问题的讨论[J]. 自然灾害学报, 2021, 30(3): 1-13.(BO Jingshan, LI Qi, SUN Qiangqiang, et al. Site classification research status and discussion of related issues[J]. Journal of Natural Disasters, 2021, 30(3): 1-13.(in Chinese))
- [3] 中国科学院土木建筑研究所. 地震区建筑规范(草案)[R]. 哈尔滨: 中国科学院土木建筑研究所, 1959: 3-4.(Institute of Civil Engineering and Architecture, Chinese Academy of Sciences. Building seismic design specification(draft)[R]. Harbin: Institute of Civil Engineering and Architecture, Chinese Academy of Sciences, 1959: 3-4.(in Chinese))
- [4] 建筑抗震设计规范: GBJ 11-89[S]. 北京: 中国建筑工业出版社, 1989.(Code for seismic design of buildings: GBJ11-89[S]. Beijing: China Architecture & Building Press, 1989.(in Chinese))
- [5] 中华人民共和国建设部、国家质量监督检验检疫总局. 建筑抗震设计规范: GB50011-2001[S]. 北京: 中国建筑工业出版社, 2004.(General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. Code for seismic design of buildings: GB50011-2001[S]. Beijing: China Architecture & Building Press, 2004.(in Chinese))
- [6] 中华人民共和国住房和城乡建设部, 国家质量监督检验检疫总局. 建筑抗震设计规范: GB50011-2010[S]. 北京: 中国建筑工业出版社, 2010.(Ministry of Housing and Urban-Rural Development of the People's Republic of China, General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. Code for seismic design of buildings: GB50011-2010[S]. Beijing: China Architecture & Building Press, 2010.(in Chinese))
- [7] 周锡元, 樊水荣, 苏经宇. 场地分类和设计反应谱的特征周期: 《建筑抗震设计规范》修订简介(八)[J]. 工程抗震, 1999, 21(4): 3-8.
- [8] 周健, 李小军, 李亚琦, 等. 中美建筑抗震设计规范中工程场地类别的对比和换算关系[J]. 地震学报, 2021, 43(4): 521-532, 534.(ZHOU Jian, LI Xiaojun, LI Yaqi, et al. Comparative analysis and transformation relations between China and the US site classification systems in building seismic code provisions[J]. Acta Seismologica Sinica, 2021, 43(4): 521-532, 534.(in Chinese))
- [9] 陈国兴, 丁杰发, 方怡, 等. 场地类别分类方案研究[J]. 岩土力学, 2020, 41(11): 3509-3522, 3582. (CHEN Guoxing, DING Jiefa, FANG Yi, et al. Investigation of seismic site classification scheme[J]. Rock and Soil Mechanics, 2020, 41(11): 3509-3522, 3582.(in Chinese))
- [10] JI K, REN Y, WEN R, et al. HVSR-based site classification approach using general regression neural network (GRNN): case study for China strong motion stations[J]. Journal of Earthquake Engineering, 2021: 1-23.
- [11] BAILEY K. Numerical taxonomy and cluster analysis[J]. Typologies and taxonomies, 1994, 34: 24.
- [12] CHEN Y K. Automatic microseismic event picking via unsupervised machine learning[J]. Geophysical Journal International, 2017, 212(1): 88-102.
- [13] RIAHI N, GERSTOFT P. Using graph clustering to locate sources within a dense sensor array[J]. Signal Processing, 2017, 132: 110-120.
- [14] KOTHA S R, COTTON F, BINDI D. A new approach to site classification: mixed-effects Ground Motion Prediction Equation with spectral clustering of site amplification functions[J]. Soil Dynamics and Earthquake Engineering, 2018, 110: 318-329.
- [15] JI K, WEN R, REN Y, et al. Nonlinear seismic site response classification using K-means clustering algorithm: case study of the September 6, 2018 Mw6. 6 Hokkaido Iburi-Tobu earthquake, Japan[J]. Soil Dynamics and Earthquake Engineering, 2020, 128: 105907.
- [16] 刘也, 任叶飞, 王大任等. 基于地震动预测残差分析的工程场地分类标准检验与评价[J]. 工程力学, 2022(Online).(LIU Ye, REN Yefei, WANG Daren et al. Evaluation on the Schemes of Engineering Site Classification Based on Residual Analysis of Ground Motion Prediction[J]. Engineering mechanics, 2022.(in Chinese))
- [17] ZHAO J X, ZHANG J, ASANO A, et al. Attenuation relations of strong ground motion in Japan using site classification based on predominant period[J]. Bulletin of the Seismological Society of America, 2006, 96(3): 898-913.
- [18] BOORE D M, STEWART J P, SEYHAN E, et al. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes[J]. Earthquake Spectra, 2013, 30(3): 1057-1085.
- [19] SCITOVSKI S. A density-based clustering algorithm for earthquake zoning[J]. Computers & Geosciences, 2018, 110: 90-95.
- [20] AHMED M, SERAJ R, ISLAM S M S. The k-means algorithm: a comprehensive survey and performance evaluation[J]. Electronics, 2020, 9(8): 1295.
- [21] HANSEN P, NGAI E, CHEUNG B K, et al. Analysis of global k-means, an incremental heuristic for minimum sum-of-squares clustering[J]. Journal of Classification, 2005, 22(2): 287-310.
- [22] SINAGA K P, YANG M S. Unsupervised K-means

-
- clustering algorithm[J]. IEEE Access, 2020, 8: 80716-80727.
- [23] YI T H, YAO X J, QU C X, et al. Clustering number determination for sparse component analysis during output-only modal identification[J]. Journal of Engineering Mechanics, 2019, 145(1): 04018122.
- [24] ZHOU S, XU Z, LIU F. Method for determining the optimal number of clusters based on agglomerative hierarchical clustering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(12): 3007-3017.
- [25] SASAN K, Shahidan M A, Azizah A M, et al. An overview of principal component analysis[J]. Journal of Signal and Information Processing, 2013, 4: 3B.
- [26] 周锡元, 王广军, 苏经宇. 场地分类和平均反应谱[J]. 岩土工程学报, 1984, 6(5): 59-68.(ZHOU Xiyuan, WANG Guangjun, SU Jingyu. Site classification and average response spectra[J]. Chinese Journal of Geotechnical Engineering, 1984, 6(5): 59-68.(in Chinese))
- [27] 吕悦军, 彭艳菊, 兰景岩, 等. 场地条件对地震动参数影响的关键问题[J]. 震灾防御技术, 2008, 3(2): 126-135.(LU Yuejun, PENG Yanju, LAN Jingyan et al. Some key problem about site effects on seismic ground motion parameters[J]. Technology for Earthquake Disaster Prevention, 2008, 3(2): 126-135.(in Chinese))
- [28] 郭俊平. 模糊综合评判和灰色模式识别在场地分类中的应用[D]. 南宁: 广西大学, 2003.(GUO Junping. Application of fuzzy comprehensive evaluation method and grey pattern recognition in site classification[D]. Nanning: Guangxi University, 2003.(in Chinese))
- [29] 李敏. 考虑场地地震动影响的场地分类方法研究[D]. 北京: 中国地震局地球物理研究所, 2015.(LI Min. Study on the site classification method based on seismic site effects[D]. Beijing: Institute of Geophysics, China Earthquake Administration, 2015.(in Chinese))
- [30] 窦立军, 杨柏坡. 场地分类新方法的研究[J]. 地震工程与工程振动, 2001, 21(4): 10-17.(DOU Lijun, YANG Baipo. Research on new method of site classification[J]. Earthquake Engineering and Engineering Vibration, 2001, 21(4): 10-17.(in Chinese))