

土壤质量评价中少量样本最小数据集的构建 ——以内蒙古杭锦旗黄河南岸灌区典型地块为例

周文涛¹ 兰天^{1,2} 潘岳¹ 公衍丽¹ 高云悦³ 李品芳^{1,4*}

(1. 中国农业大学 土地科学与技术学院,北京 100193;

2. 清华大学 环境学院,北京 100084;

3. 中国农业大学 资源与环境学院,北京 100193;

4. 农业农村部华北耕地保育重点实验室,北京 100193)

摘要 为研究土壤质量评价过程中样本容量较少情况下最小数据集的构建方法。采集并测定内蒙古杭锦旗黄河南岸灌区典型地块的12个土壤理化指标,利用聚类分析、相关分析和主成分分析等探讨了少量样本最小数据集的构建方法。结果表明:1)通过聚类分析、相关分析和主成分分析所构建的最小数据集中的指标包括土壤容重、饱和含水量、土壤电导率、阳离子交换量、硝态氮、速效钾、速效磷;2)全体数据集与最小数据集的土壤质量指数呈显著正相关, R^2 达到了0.735,Nash有效系数为0.917,偏差系数为0.057;3)基于全体数据集和最小数据集计算得出的土壤质量指数变化范围分别是0.47~0.73和0.37~0.75,平均值为0.56和0.53,表明研究地块土壤呈现中等质量水平。研究发现,在黄河南岸灌区典型地块基于少量样本进行土壤质量评价的过程中,使用聚类分析、相关分析、主成分分析3种分析方法可以构建最小数据集,且检验精度较高。

关键词 土壤质量; 少量样本; 最小数据集; 主成分分析; 聚类分析; 黄河南岸灌区

中图分类号 S114

文章编号 1007-4333(2022)06-0225-11

文献标志码 A

Construction of minimum data set with small number of samples for soil quality assessment: A case study of a typical land in the south bank of Yellow River irrigation area of Hangjin Banner, Inner Mongolia

ZHOU Wentao¹, LAN Tian^{1,2}, PAN Yue¹, GONG Yanli¹, GAO Yunyue³, LI Pinfang^{1,4*}

(1. College of Land Science and Technology, China Agricultural University, Beijing 100193, China;

2. College of Environmental Sciences, Tsinghua University, Beijing 100084, China;

3. College of Resources and Environmental Sciences, China Agricultural University, Beijing 100193, China;

4. Key Laboratory of Arable Land Conservation (North China) of Ministry of Agriculture and Rural Affairs, Beijing 100193, China)

Abstract In order to study the construction of minimum data set by using the small number of samples for soil quality evaluation, this research collected and measured twelve soil physical and chemical indicators of a typical field in the south bank irrigation area of the Yellow River in Hangjin Banner, Inner Mongolia. The methods for constructing the minimum data set with small number of samples by cluster analysis, correlation analysis and principal component analysis were investigated. The results showed that: 1) The indexes in the minimum data set established by cluster analysis, correlation analysis and principal component analysis are soil bulk density, saturated water content, soil conductivity, cation exchange capacity, nitrate nitrogen, available potassium and available phosphorus; 2) The soil

收稿日期: 2021-10-16

基金项目: 国家重点研发计划项目(2018YFF0213406)

第一作者: 周文涛, 硕士研究生, E-mail: zhouwentao971004@163.com

通讯作者: 李品芳, 教授, 主要从事盐碱地改良与利用及土壤健康评价等研究, E-mail: pfl@cau.edu.cn

quality index of both the total data set and minimum data set is significantly positively correlated, where R^2 is 0.735, the Nash effective coefficient is 0.917, and the deviation coefficient is 0.057; 3) Based on the total data set and minimum data set, the variation range of soil quality index are respectively 0.47–0.73 and 0.37–0.75 and with an average of 0.56 and 0.53, indicating the soil quality level of the study field is moderate. In conclusion, for soil quality evaluation in the typical irrigated area on the south bank of the Yellow River, the minimum data set can be constructed based on a small number of samples by using three methods including cluster analysis, correlation analysis and principal component analysis, and the test accuracy is high.

Keywords soil quality; small number of samples; minimum data set; principal component analysis; cluster analysis; south bank of Yellow River irrigation area

除水资源质量和空气质量之外,土壤质量是又一重要的环境质量参数^[1]。土壤质量体现了土壤在生态系统和土地利用中维持生物生产力、环境质量和促进动植物健康的能力^[2]。开展土壤质量评价工作,对种植模式优化和产业结构调整具有重要意义。自美国农业部土壤保护局于1961年发布土地能力分类系统以来,已开发出多种土壤质量评估方法,如:土壤质量检测卡^[3]、土壤质量指数法^[4]、空间插值法^[5]和土壤质量动态变化模型^[6]。其中,土壤质量指数法在进行土壤质量评价的工作中应用较为广泛,可以反应自然条件下土壤能力的复杂性。利用数理统计的方法通过定量模型对指标进行综合评估,能够更直接的代表土壤的质量情况。已经建立的定量模型有综合质量指数(IQI)^[7]和Nemoro质量指数(NQI)^[8]等,这些质量指数是通过数学方法来评估某些土壤因素与土地生产力之间的关系。

在土壤质量评价的过程中,常运用最小数据集理论,通过聚类分析、相关分析、主成分分析等统计方法对冗余指标进行筛选后,构建土壤质量评价的最小数据集来评价土壤质量^[10-11]。土壤质量的评价工作由于研究尺度的不同,土壤样品的采集数量有所差异,选择测定的指标也有所不同。如:在对南方地区冷浸田土壤质量进行评价的研究中,采集了7个省份的36个土壤样品,使用主成分分析法对测定的18项指标进行了筛选^[9];北方地区土默川平原不同利用类型土壤质量评价采集了20个土壤剖面的土壤样品,测定了32项分析指标,使用主成分分析法构建最小数据集^[10];在探讨伊朗干旱区将原生牧场转化为农田后土壤质量变化的研究中,同样采用主成分分析法建立土壤质量评价的最小数据集^[11]。主成分分析法是构建最小数据集广泛应用的统计方法,但评价指标是否适合采用主成分分析法需要经过KMO(Kaiser-Meyer-Olkin)与Bartlett's(Bartlett test of sphericity)检验^[12]。采用聚类分析或是其他分

析方法也可以进行最小数据集的构建,但需要足够多的研究样本及一定量的测定指标。如:在使用聚类分析和主成分分析对红壤坡耕地耕层土壤质量进行评价的研究中,采集了4个省份不同耕作方式的41个样点,测定了11项理化指标^[13];在对黄河三角洲地区土壤质量评价的工作中,利用模糊逻辑的数学方法对98个样点的10项土壤理化指标进行分析,构建了最小数据集对土壤质量进行评估^[14]。应用聚类分析构建最小数据集一般是在聚类分析的结果上结合相关分析对指标进行筛选,但是,仅依靠聚类分析方法无法完成对评价指标的精简,而大量的样点结合多个不同的评价指标意味着工作量成倍增加,耗费更长时间和增加更多成本。因此,如何在少量样本的情况下,通过分析方法的结合成功的构建土壤质量评价的少量样本最小数据集是有待深入探讨的研究课题。

内蒙古杭锦旗黄河南岸灌区位于河套平原黄河“几”字湾南岸,灌区内耕地面积广阔,土壤物理性质均一,但灌区内土壤质量状况尚未探明。因此,本研究拟以黄河南岸灌区单一地块少量土壤样本的12个理化指标的测定结果为基础,采用聚类分析对评价指标进行初筛,并且进行主成分分析的前提检验,在满足检验的条件下对冗余指标通过主成分分析进一步筛选,探讨土壤质量评价中基于少量样本最小数据集的构建方法,并与全体数据集进行了自相关的精度检验,以期土壤质量评价工作中少量样本最小数据集的构建方法提供参考。

1 材料与方法

1.1 研究区域概况

试验区域位于内蒙古鄂尔多斯市杭锦旗吉日嘎朗图镇,40°76′~40°81′N,107°65′~108°11′E,地处黄河“几”字湾南岸,杭锦旗北部,东接独贵塔拉镇,西邻呼和木独镇,南连库布其沙漠,北与巴彦淖

尔市隔黄河相望。全镇东西长 90 km、总面积 2 747 km²。研究区域北靠黄河南临沙漠，引用黄河水进行农业灌溉，故称为黄河南岸灌区。南岸灌区属于中温带半干旱大陆性气候，平均海拔 1 006 m，年降水量 150 mm 左右，60% 的降水集中在 7—9 月。平均气温 8.1 ℃，昼夜温差大，日照时间长，年日照时间约为 3 200 h，无霜期 130 d 左右，适宜农作物和牧草生长。粮食作物以玉米、小麦为主，经济作物以葵花为主。灌区内土壤质地较为黏重，透水性差，降雨后极易在地表形成积水，在作物生长初期频繁造成幼苗死亡。由于研究区域紧邻沙漠边缘，当地对质地黏重的土壤普遍采用掺沙的改良措施。

1.2 样品采集及理化指标测定

1.2.1 土样采集

通过对黄河南岸灌区的野外调查，确定本次试验选择的研究地块为灌区内典型的黏土质地，研究地块位于 40°77′ N, 107°50′ E，面积为 4 236.31 m²。于 2019 年在地表进行 30 cm 的掺沙后与原有耕层土壤进行翻耕，改良后耕层土壤质地经测定为壤质砂土。在 2020 年 7 月对所选地块耕层土壤样品进行采集后测定基本理化性质。如图 1 所示，使用基于网格的小规模土壤采样方案确定研究地块采样点的位置^[15]，共选取 24 个采样点，采集 0~20 cm 深土壤样品，每个样点 3 次重复采样，混合均匀后四分

法留取约 1 kg 土壤样品，实验室内风干后过 2 mm 筛用于土壤基本理化指标的测定与分析。共测定以下 12 项土壤理化指标：田间持水量(FMC)、饱和含水量(SMC)、土壤容重(BD)、有机质(OM)、速效钾(AK)、速效磷(AP)、铵态氮(NH₄⁺-N)、硝态氮(NO₃⁻-N)、土壤 pH、电导率值(EC)以及阳离子交换量(CEC)和钠吸附比(SAR)。

1.2.2 土壤理化指标的测定方法

具体测定指标为：田间持水量、饱和含水量和土壤容重采用环刀法；土壤有机质采用重铬酸钾容量法；速效钾采用乙酸铵浸提-火焰光度计法；速效磷采用碳酸氢钠浸提-钼锑抗比色法；铵态氮用 2 mol/L KCl 浸提-靛酚蓝比色法；硝态氮用双波长紫外分光光度法^[16]；阳离子交换量采用乙酸铵-火焰光度法；pH 采用 $m_{\text{水}} : m_{\text{土}} = 2.5 : 1$ 浸提电位法；电导率采用 $m_{\text{水}} : m_{\text{土}} = 5 : 1$ 浸提电导法；土壤盐基离子 Na⁺、Ca²⁺、Mg²⁺ 采用 $m_{\text{水}} : m_{\text{土}} = 5 : 1$ 浸提液相离子色谱仪测定；

钠吸附比(SAR)表征土壤的碱化程度，计算公式如下：

$$SAR = \frac{C_{\text{Na}^+}}{\sqrt{\frac{1}{2}(C_{\text{Ca}^{2+}} + C_{\text{Mg}^{2+}})}} \quad (1)$$

式中： C_{Na^+} 、 $C_{\text{Ca}^{2+}}$ 和 $C_{\text{Mg}^{2+}}$ 分别为土壤溶液中 Na⁺、Ca²⁺、Mg²⁺ 的相对浓度，mmol/L。

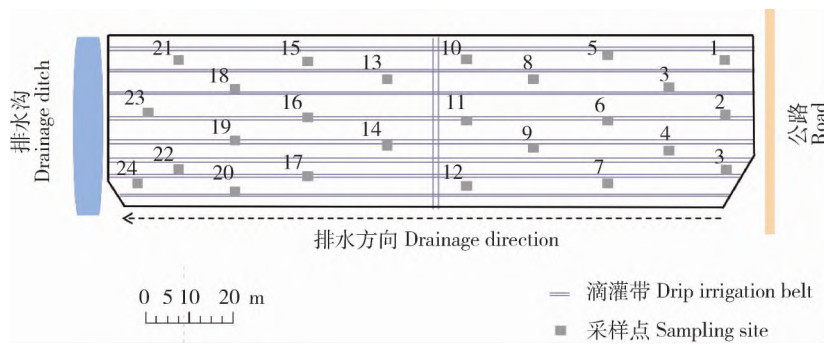


图 1 研究地块采样点位示意图

Fig. 1 Schematic diagram of sampling sites in the research field

1.3 最小数据集的构建方法

1.3.1 聚类分析

聚类分析是指将数据对象的集合分组为由类似对象所组成多个类的分析过程，其本质是寻找数据之间的内在关联结构。在本研究土壤质量评价的过程中，选择目前在实际应用中使用的最多的分层聚类

法对指标数据进行 R 型聚类，通过定义指标和不同类别之间的距离，合并距离最近的 2 类，再重新计算新类别与其他类之间的距离，采用最短距离法进行归类，使用欧式距离衡量评价指标之间的差异性^[17]。通过聚类将指标集合分组进行各组指标之间的相关性分析，建立土壤质量评价的最小数据

集^[13]。对评价指标进行聚类分析,以探讨聚类后树状图中各指标之间的聚合水平,将指标划分为反应土壤质量不同特征的若干组分,对相同组内的指标进行相关分析,相关性显著的指标可以根据研究区实际调查情况、专家意见和以往的研究结果互相代替,最终筛选出具有代表性的指标进入最小数据集。

1.3.2 主成分分析

主成分分析的核心是通过降维的统计原理精简评价指标的数量,消除多重指标之间的相关性,使最终筛选出来的指标具有较强的独立性^[14]。对土壤理化指标进行主成分分析,提取特征值 ≥ 1 的主成分,将在相同主成分中载荷 ≥ 0.5 的指标作为进入最小数据集中的备选指标。如果某项指标同时在2个主成分上的载荷都 ≥ 0.5 则该项指标应在与其他指标相关性较低的主成分中进行分析。为了避免仅使用指标载荷作为唯一的选择标准而导致部分指标信息被忽略,引入矢量常模值 Norm 值的计算,作为指标最终选入最小数据集的参考依据^[18]。某项指标的 Norm 值越大,则表明该指标在所有主成分的综合载荷越大,其解释综合信息的能力就越强。Norm 值计算公式如下:

$$N_{ik} = \sqrt{\sum_{i=1}^k (u_{ik}^2 \cdot \lambda_k)} \quad (2)$$

式中: N_{ik} 为第*i*个指标在特征值 ≥ 1 的前*k*个主成分上的综合载荷; u_{ik} 为第*i*个指标在第*k*个主成分上的载荷; λ_k 为第*k*个主成分的特征值。

计算各组指标的 Norm 值,选取每组中 Norm 值超过最高分值 90% 的指标。当每组主成分中符合要求的指标超过 1 时,需要对所选指标进行相关分析,若高度相关,则 Norm 值最大的指标进入最小数据集,相关度低则同组指标均进入最小数据集。在本研究中,由于部分土壤指标的变异性较大,在使用主成分分析法确定最小数据集时相关系数 > 0.3 即认为具有相关性^[9,13-14]。

1.4 土壤质量评价

1.4.1 土壤质量指数的计算

土壤质量指数是一个基于指标使用的多维概念,指标的综合评估比单个指标的选择更具有代表性,土壤质量指数能够以数值的形式体现土壤质量的优劣状况。根据评价指标对耕层土壤质量的正负效应,建立指标与土壤质量之间的隶属函数^[19],结合各数据集中的指标权重值最终计算得到土壤质量指数。

S 型隶属函数:

$$u(x) = \begin{cases} 1, & x \geq b \\ \frac{x-a}{b-a}, & a < x < b \\ 0, & x \leq a \end{cases} \quad (3)$$

反 S 型隶属函数:

$$u(x) = \begin{cases} 1, & x \leq a \\ \frac{x-b}{a-b}, & a < x < b \\ 0, & x \geq b \end{cases} \quad (4)$$

抛物线型隶属函数:

$$u(x) = \begin{cases} 1, & b_2 \geq x \geq b_1 \\ \frac{x-a_1}{b_1-a_1}, & a_1 < x < b_1 \\ \frac{x-a_2}{b_2-a_2}, & a_2 > x > b_2 \\ 0, & x \leq a_1 \text{ 或 } x \geq a_2 \end{cases} \quad (5)$$

式中: x 为评价指标实测值; a 、 b 分别为指标临界值的下限和上限,为实测的最小值和最大值; a_1 、 a_2 分别为指标临界值的下限和上限,为实测的最小值和最大值; b_1 、 b_2 为最适值的下限和上限。

分别对全体数据集 TDS 和最小数据集 MDS 中的指标做主成分分析,提取评价指标的公因子方差,各项指标公因子方差占公因子方差之和的比例为各数据集中评价指标的权重值,结合隶属度对不同数据集的土壤质量指数进行计算,公式^[20]如下:

$$SQI = \sum_{i=1}^n \omega_i \cdot N_i \quad (6)$$

式中: ω_i 为第*i*个评价指标权重; N_i 为第*i*个评价指标隶属度值; n 为评价指标个数。

1.4.2 土壤质量评价精度验证

利用 Nash 有效系数 E_f 和相对偏差系数 E_r 对最小数据集的精确度进行验证^[21]。公式如下:

$$E_f = 1 - \frac{\sum (R_T - R_M)^2}{\sum (R_T - \bar{R}_T)^2} \quad (7)$$

$$E_r = \left| \sum_{i=1}^n R_T - \sum_{i=1}^n R_M \right| / \sum_{i=1}^n R_T \quad (8)$$

式中: R_T 和 \bar{R}_T 为全体数据集计算出各样点的土壤质量指数值以及所有样点土壤质量指数的平均值; R_M 为基于最小数据集计算出各样点的土壤质量指数值, n 为样点个数。Nash 有效系数 E_f 越接近于 1,表示所构建的最小数据集计算出的土壤质量指数与全体数据集计算出的土壤质量指数越接近,模型

精确度越高。相对偏差系数 E_r 越接近于 0, 则表示模型偏差值越小。

1.5 数据处理

试验数据在 Microsoft Excel 2016 中计算处理后通过 SPSS 24.0 软件对土壤理化指标进行聚类分析、主成分分析和相关分析以及聚类分析树状图的制作, 其余图表均采用 Microsoft Excel 2016 制作完成。

2 结果与分析

2.1 最小数据集的构建

2.1.1 聚类分析构建指标分类

基于聚类分析对全体数据集中 12 个理化指标进行分类, 结果见图 2。可知在聚合水平为 23~24 将全体数据集分为 4 类: 土壤 pH、电导率值、钠吸附比、阳离子交换量、铵态氮、硝态氮为第 1 类, 代表土壤的盐渍化和氮素供应; 速效钾、速效磷、有机质为第 2 类, 代表土壤的速效养分和肥力水平; 土壤容重为第 3 类, 代表土壤的物理结构特性; 田间持水量、饱和含水量为第 4 类, 代表土壤对水分的保蓄能力。

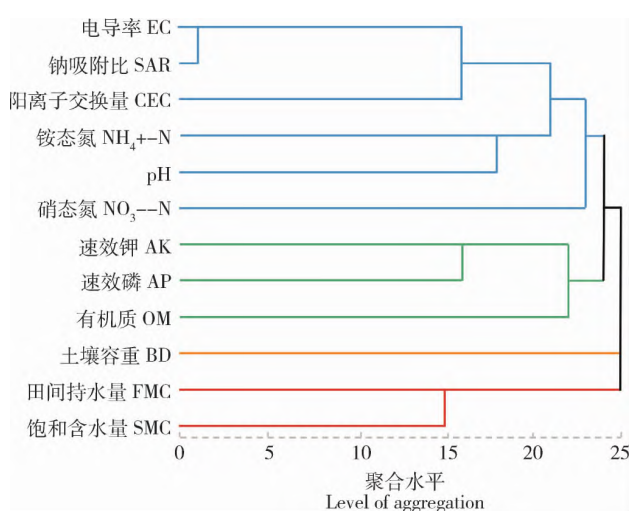


图 2 构建少量样本最小数据集的聚类分析树状图

Fig. 2 Cluster analysis tree for minimum data set construction with a small number of samples

2.1.2 聚类分析结果的初步筛选

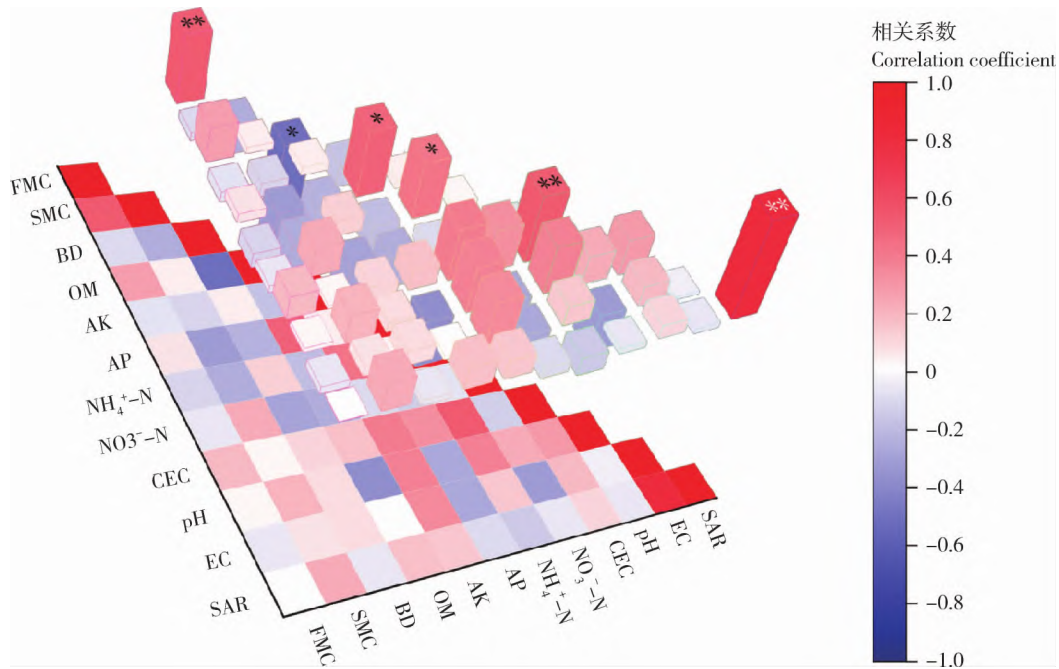
通过相关分析进行聚类结果的初步筛选, 对 12 项土壤理化指标进行正态分布检验, 结果表明: 田间持水量、有机质、阳离子交换量符合正态分布; 容重、有机质、速效钾、铵态氮、钠吸附比符合对数正态分布; 而饱和含水量、硝态氮、pH、电导率不满足正态

分布条件。所以本研究选择 Spearman 相关系数进行指标间的相关分析。图 3 为研究地块各评价指标间的 Spearman 相关系数矩阵, 在第 1 类指标中土壤电导率值与钠吸附比相关系数为 0.806**, 钠吸附比用于反应土壤中钙离子和镁离子的存在对交换性钠离子的中和作用, 而电导率值能更直观的说明土壤的盐渍化情况, 选择电导率值作为第 1 类指标中进入最小数据集的备选; 铵态氮与阳离子交换量相关系数为 0.527**, 阳离子交换量能够衡量土壤的缓冲性, 反应的是土壤胶体所能吸附各种阳离子的总量, 铵态氮代表 1 种氮素养分, 表示土壤的肥力水平, 2 类指标的实际关联性并不是十分明确, 为避免仅依据指标间相关性对指标进行筛选而造成选择上的误差, 暂时选择以上 2 种指标共同进入最小数据集备选, 而后通过主成分分析进一步筛选。第 1 类其余指标均无显著相关性, 一并选择进入最小数据集备选。第 2 类中速效磷与有机质相关系数为 0.511*, 速效磷是作物从土壤中获取的主要养分, 其含量变化可以反映土壤质量的变化^[22], 尽管有机质被普遍应用在土壤质量评价的过程中, 由于该指标结构和功能的非均一性, 有研究认为有机质在土壤质量评价过程中的相关性并不明确^[23], 故选择速效磷替代有机质进入最小数据集, 其余指标无显著相关性, 共同选入最小数据集。第 3 类中仅有土壤容重, 选入最小数据集。第 4 类中田间持水量与饱和含水量相关系数为 0.529**, 饱和含水量反映土壤的孔隙状况和最大的持水容量, 田间持水量代表了土壤所能稳定保持的最高水分含量, 该指标虽然可以在田间测定, 但不易再现, 并且本研究区域气候干旱, 降雨集中在固定月份, 田间持水量的测定极易受季节性降水的影响, 故选择饱和含水量列入最小数据集。

2.1.3 主成分分析进一步筛选冗余指标

通过对聚类结果进行相关分析后, 已将速效磷、速效钾、土壤容重、饱和含水量选入最小数据集, 但对于第 1 类选入最小数据集备选的指标还存在数据冗余, 故进一步采用主成分分析的方法确定第 1 类最终入选最小数据集的指标。

首先对主成分分析的指标进行 KMO 与 Bartlett's 检验, 确定所选指标是否适宜使用主成分分析的方法。检验后的 KMO 值为 0.562^[12], 满足进行主成分分析的条件; Bartlett's 球型检验中 $P=0.039 < 0.05$, 达到显著性水平, 因此所选指标适合



* 为在 $P \leq 0.05$ 水平上显著; ** 为 $P \leq 0.01$ 水平上的显著。

* means significance at $P \leq 0.05$ level; ** means significance at $P \leq 0.01$.

图3 研究地块评价指标的 Spearman 相关系数矩阵

Fig. 3 Spearman descriptive statistic of soil evaluating indicators of the study plot

采用主成分分析的方法。

进一步对备选指标土壤 pH、电导率值、阳离子交换量、铵态氮、硝态氮进行主成分分析,其结果见表 1。为保证主成分能够达到足够的累积贡献率,本次研究选择特征值大于 0.8 的主成分,并将结果分为 3 个主成分,累积贡献率达 81.35%,采用这 3 个主成分对指标进行筛选。主成分 1 特征根的贡献率为 40.08%,其中具有高载荷的指标为土壤 pH、电导率值、阳离子交换量、铵态氮。通过对高载荷指标 Norm 值的计算,选择 Norm 值最高的指标阳离子交换量(Norm=1.20)Norm 值 90% 内的指标进行筛选,结果显示主成分 1 中所有选择的指标均满足条件,进而对各指标进行相关分析。阳离子交换量与铵态氮和土壤 pH 的相关系数分别为 0.527** 和 0.286 \approx 0.3,故选择 Norm 值最高的阳离子交换量和与其相关性最低的电导率值进入最小数据集;主成分 2 特征根的贡献率为 24.29%,具有高载荷的指标为土壤 pH、电导率值、硝态氮;主成分 3 特征根的贡献率为 16.98%,具有高载荷的指标仅有硝态氮。土壤 pH 在主成分 1 的分析中被替代,电导率值也已进入最小数据集,仅讨论硝态氮是否进入最小数据集即可。由于主成分 3 中仅有硝态氮 1

个指标具有较高载荷,故选择该指标进入最小数据集。综合以上分析结果,最终确定 3 个主成分中进入最小数据集的指标为阳离子交换量、电导率值、硝态氮。

通过以上聚类分析、主成分分析、相关分析,综合计算与分析结果,最终从 12 个土壤理化指标中筛选出土壤容重、电导率值、饱和含水量、阳离子交换量、硝态氮、速效钾、速效磷 7 个指标,可作为评价研究地块土壤质量的最小数据集。

2.2 土壤质量指数的计算

建立土壤质量评价的最小数据集后,确定全体数据集与最小数据集中各指标的权重值与隶属度函数,进而计算分析两类数据集的土壤质量指数,结果见表 2,提取评价指标的公因子方差,通过各项指标公因子方差占公因子方差之和的比例分别计算出两类数据集中各指标的权重值,并使用隶属度函数确定各指标的隶属度。在本研究区域中,土壤有机质、硝态氮、铵态氮、阳离子交换量、速效钾、速效磷的含量与土壤质量呈正相关,界定为 S 型函数;而钠吸附比、电导率值的数值越高,越不利于作物生长,与土壤质量呈负相关,故界定为反 S 型函数。并且,由于本研究区域土壤为盐渍化土壤,pH 均在 8.5 以

表1 构建少量样本最小数据集的指标载荷矩阵和 Norm 值

Table 1 Index load matrix and Norm values to build the MDS with a small number of samples

评价指标 Evaluation index	主成分 Principal component			Norm 值
	第一因子 Principal component 1	第二因子 Principal component 2	第三因子 Principal component 3	
	pH	0.597	0.589	
EC	0.588	-0.585	0.442	1.129
CEC	0.841	-0.099	0.088	1.198
NO ₃ ⁻ -N	0.204	0.712	0.635	1.021
NH ₄ ⁺ -N	0.743	-0.093	-0.33	1.099
特征值 Eigen value	2.004	1.215	0.849	—
贡献率/% Variance	40.075	24.294	16.978	—
累积贡献率/% Cumulative variance	40.075	64.369	81.347	—

表2 全体数据集与最小数据集的指标权重和隶属度函数类型

Table 2 Index weight and membership function type of the TDS and the MDS

指标 Index	隶属函数类型 Membership function type	全体数据集 Total data set		最小数据集 Minimum data set	
		公因子方差 Common factor variance	因子权重 Factor weight	公因子方差 Common factor variance	因子权重 Factor weight
FMC	抛物线型	0.680	0.076		
SMC	抛物线型	0.725	0.812	0.327	0.069
BD	抛物线型	0.715	0.080	0.701	0.149
OM	S型	0.755	0.085		
AK	S型	0.679	0.076	0.613	0.130
AP	S型	0.768	0.086	0.802	0.171
NH ₄ ⁺ -N	S型	0.550	0.062		
NO ₃ ⁻ -N	S型	0.748	0.838	0.741	0.158
CEC	S型	0.713	0.079	0.772	0.164
EC	反S型	0.933	0.104	0.745	0.158
pH	反S型	0.714	0.079		
SAR	反S型	0.950	0.106		

上,所以 pH 越低则认为土壤质量相对越高,故在本研究中将 pH 指标界定为反 S 型函数;土壤容重、田间持水量、饱和含水量在土壤质量评价的过程中存在适宜的临界范围,则界定以上指标属于抛物线型函数。本研究中土壤容重均大于 1.3 g/cm^3 ,这与研究地块壤质砂土的土壤质地有关,结合实验数据与相关研究结果,判断土壤容重抛物线函数的适宜范围是 $1.4 \sim 1.6 \text{ g/cm}^3$ [24];田间持水量的测定不易再现,本次试验数据与目前关于河套平原土壤田间持水量的参考文献中也存在差异[25],结合本研究试验数据与以往的研究结果,认为田间持水量抛物线函数的适宜范围是 $20\% \sim 24\%$ [26];饱和含水量在不同土壤质地中的含量有较大的差异,本研究选取砂质土壤上饱和含水量的适宜范围 $25\% \sim 60\%$ 作为饱和含水量抛物线函数的临界范围[27]。结合各指标的权重值与隶属度,最终计算得出两类数据集各样点的土壤质量指数。全体数据集和最小数据集的土壤质量指数值分别在 $0.47 \sim 0.73$ 和 $0.37 \sim 0.75$,平均值分别为 0.56 和 0.53 。

2.3 最小数据集精度检验与土壤质量评价

以全体数据集计算得到的土壤质量指数值(SQI-TDS)为基准对最小数据集土壤质量指数值(SQI-MDS)进行精度检验。将两类数据集土壤质量指数值进行回归分析,其结果见图4。基于全体数据集和最小数据集计算得到的土壤质量指数值的范围分别为 $0.47 \sim 0.73$ 和 $0.37 \sim 0.75$,平均值为 0.56 和 0.53 ,差值较小,且从图4的拟合效果来看,两类数据集的土壤质量指数呈显著正相关, R^2 达到了 0.7354 。最小数据集土壤质量指数的 Nash 有效

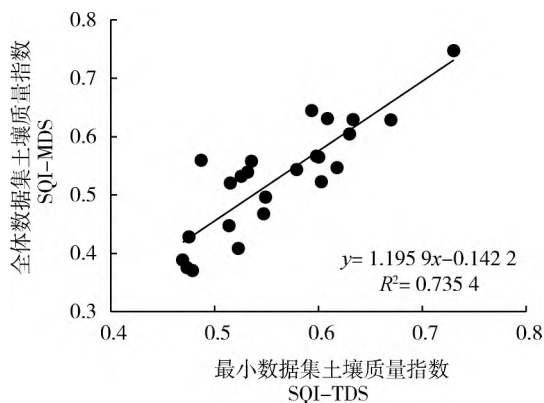


图4 基于少量样本最小数据集和全体数据集的研究区域土壤质量指数相关性

Fig. 4 Correlation of the study area soil quality index based on TDS and MDS

系数为 0.917 ,偏差系数为 0.057 ,说明本研究构建的最小数据集的检验精度较高。可以判断,结合聚类分析、主成分分析和相关分析构建少量样本的最小数据集能够代替全体数据集对研究地块土壤质量进行评价。

土壤质量指数越高,表明研究范围内的土壤理化性质越适宜作物的生长。采用等间距法将土壤质量划分为5个等级标准[28]进行比对分析,判断研究地块土壤属于3级,为中等质量水平。两类数据集土壤质量指数的变异系数分别为 12.10% 和 17.89% ,属于低度变异,这是由于采样范围较小,在单个地块上的土壤质量指数并未呈现出较大变异。

3 讨论

指标选择是土壤质量评价的关键。代表土壤物理性质的容重和反映土壤肥力水平的有机质、全氮、速效磷、速效钾等指标被广泛应用在土壤质量评价的工作中[11,29]。最小数据集中速效磷的选择能够在土壤质量评价的过程中直观的反应养分的供应情况,该指标的选择与已有研究一致[13,22,30]。对于反应土壤水分性质的指标,有研究选择田间持水量进入最小数据集[31],原因是由于其研究区域雨量充沛,田间持水量不易受季节性降水影响[32],能较为客观的表征土壤的水分性质,然而,由于本研究区域气候干旱,田间持水量的测定极易受季节性降水的影响,故选择土壤饱和含水量而不是田间持水量进入最小数据集。

土壤质量评价的尺度也有多种选择,一般以田块尺度、县域尺度、市域尺度、区域尺度[33]为主。评价单元的大小直接影响土壤质量评价的精度,已有研究中多基于大、中尺度对土壤质量进行评价[34],需要采集大量的土壤样品。而对于单一地块小范围内、以最小数据集理论为基础的土壤质量评价,往往因为样本点的数量较少受到主成分分析法前提检验的限制。本研究结合聚类分析、相关分析和主成分分析3种分析方法对内蒙古黄河南岸灌区单一地块土壤质量评价的最小数据集进行构建,最终构建的最小数据集指标包括:电导率值(EC)、容重(BD)、饱和含水量(SMC)、阳离子交换量(CEC)、速效钾(AK)、速效磷(AP)及硝态氮(NO_3^- -N)。上述指标可表征土壤的盐渍化程度、物理特性和养分供应状况。本研究通过与全体数据集的精度检验对比发现,最小数据集能够较为全面的代替全体数据集对

研究地块土壤质量进行评价。

在对红壤丘陵区坡耕地土壤质量评价的过程中,使用聚类分析比主成分分析建立的最小数据集评价精度更高^[13]。由于本研究样本数量较少,无法直接通过主成分分析进行最小数据集的构建,并且,仅使用聚类分析不能充分的对冗余指标进行筛选,所以本研究运用聚类分析对评价指标分类,从而达到了降低应用主成分分析所需指标数量的目的。分类后的指标满足主成分分析的前提条件,再采用主成分分析和相关分析对冗余指标进行筛选,最终得到最小数据集。对比过多的依靠相关分析以及经验判断进行指标选择,聚类分析与主成分分析的结合能够更有效的通过统计原理构建最小数据集,最大程度的利用数据信息,使评价结果更具代表性^[35],可为单一地块土壤质量评价过程中少量样本最小数据集的构建提供参考,也为小样本容量下无法直接通过主成分分析前提检验构建最小数据集的情况提供了新的研究方法。

本研究虽然成功构建了土壤质量评价的少量样本最小数据集,但是对于最佳样点数和测定指标数的设置并没有进行深入探讨,并且,合理应用主成分分析构建最小数据集所需的样本数量与指标数量一直以来没有一个明确的限定。目前的研究多集中在大尺度大样本的基础上,而对少量样本的定义还比较模糊,本研究中的少量样本是相对以往研究中的采样量而言,在今后的研究中还应深入探讨。

4 结 论

本研究主要结论如下:

1)使用聚类分析、主成分分析和相关分析,最终筛选出黄河南岸灌区典型地块土壤的理化指标中饱和含水量、土壤容重、阳离子交换量、电导率值、硝态氮、速效钾、速效磷可作为土壤质量评价的最小数据集指标。

2)全体数据集土壤质量指数变化幅度为 0.47~0.73,均值为 0.56,最小数据集土壤质量指数变化幅度为 0.37~0.75,均值为 0.53。根据土壤质量分级标准,判断黄河南岸灌区该地块属于中等质量水平。

3)对全体数据集与最小数据集分别得到的土壤质量指数进行相关分析和精度检验,二者呈显著正相关, R^2 达到 0.735,Nash 有效系数为 0.917,偏差系数为 0.057,表明结合聚类分析、主成分分析和相

关分析所构建的少量样本最小数据集可以代替全体数据集对研究地块进行土壤质量评价。

本研究为小样本容量下无法直接通过主成分分析前提检验构建最小数据集的情况提供了新的思路和方法。在今后关于土壤质量评价的此类研究中,还应进一步深入探究样本数量与评价指标数量之间的联动,以探明二者能否直接通过主成分分析构建最小数据集的比例界限等问题。

致谢

感谢内蒙古鄂尔多斯市杭锦旗中国农业大学教授工作站(20201701)的合作方杭锦旗水利局;感谢刘海全、刘瑞春、杨智等相关人员的支持以及冯志忠对本研究试验样品采集及测定工作的辛勤付出。

参考文献 References

- [1] Andrews S S, Karlen D L, Mitchell J P. A comparison of soil quality indexing methods for vegetable production systems in Northern California[J]. *Agriculture Ecosystems & Environment*, 2002, 90(1): 25-45
- [2] 杜振宇, 葛忠强, 王清华, 梁燕, 李宗泰, 刘桂民. 鲁中山地侧柏人工林土壤质量评价及林分密度的影响作用[J]. *水土保持研究*, 2020, 27(2): 128-135
Du Z Y, Ge Z Q, Wang Q H, Liang Y, Li Z T, Liu G M. Assessment on soil quality of plantations in central mountainous areas of Shandong Province and the influence of stand density [J]. *Research of Soil and Water Conservation*, 2020, 27(2): 128-135 (in Chinese)
- [3] Purakayastha T J, Pathak H, Kumari S, Biswas S, Chakrabarty B, Padaria R N, Kamble K, Pandey M, Sasmal S, Singh A. Soil health card development for efficient soil management in Haryana, India [J]. *Soil and Tillage Research*, 2019, 191: 294-305
- [4] Doran J W, Parkin T B. Quantitative indicators of soil quality: a minimum dataset[M]. In: Doran, J W, Jones, A J, eds. *Methods for Assessing Soil Quality*. Madison: Soil Science Society of America, 1996: 25-37
- [5] Diodato N, Ceccarelli M. Multivariate indicator Kriging approach using a GIS to classify soil degradation for mediterranean agricultural lands [J]. *Ecological Indicators*, 2004, 4(3): 177-187
- [6] Larson W E, Pierce F J. The dynamics of soil quality as a measure of sustainable management [M]. In: Doran J W, Coleman D C, Bezdicek D F, Stewart B A, eds. *Defining Soil Quality for a Sustainable Environment*. Madison: Soil Science Society of America, 1994: 37-51
- [7] Doran J W, Parkin T B. Defining and assessing soil quality

- [M]. In: Doran J W, Coleman D C, Bezdicek D F, Stewart B A, eds. *Defining Soil Quality for a Sustainable Environment*. Madison: Soil Science Society of America, 1994: 3-21
- [8] Qin M Z, Zhao J. Strategies for sustainable use and characteristics of soil quality changes in urban-rural marginal area: A case study of Kaifeng[J]. *Acta Geographica Sinica*, 2000, 67(5): 545-554
- [9] 邓绍欢, 曾令涛, 关强, 李鹏, 刘满强, 李辉信, 焦加国. 基于最小数据集的南方地区冷浸田土壤质量评价[J]. *土壤学报*, 2016, 53(5): 1326-1333
- Deng S H, Zeng L T, Guan Q, Li P, Liu M Q, Li H X, Jiao J G. Minimum dataset-based soil quality assessment of waterlogged paddy field in South China[J]. *Acta Pedologica Sinica*, 2016, 53(5): 1326-1333 (in Chinese)
- [10] 高娃, 景宇鹏, 樊明寿, 郜翻身, 郑海春, 陈玉海. 土默川平原盐渍化土地不同利用方式的土壤质量评价[J]. *中国土壤与肥料*, 2019(6): 22-31
- Gao W, Jing Y P, Fan M S, Gao F S, Zheng H C, Chen Y H. Soil quality assessment of different land use on the saline soils in Tumochuan plain[J]. *Soil and Fertilizer Sciences in China*, 2019(6): 22-31 (in Chinese)
- [11] Raiesi F. A minimum data set and soil quality index to quantify the effect of land use conversion on soil quality and degradation in native rangelands of upland arid and semiarid regions[J]. *Ecological Indicators*, 2017, 75: 307-320
- [12] 刘孟宜, 田博宇, 王滢颖, 王超, 李慧静. 基于主成分分析的不同小麦品种制作的韧性饼干品质评价[J]. *食品研究与开发*, 2021, 42(10): 44-53
- Liu M Y, Tian B Y, Wang Y Y, Wang C, Li H J. Comprehensive quality evaluation of semi-hard biscuits of different wheat varieties based on principal components analysis[J]. *Food Research and Development*, 2021, 42(10): 44-53 (in Chinese)
- [13] 金慧芳, 史东梅, 陈正发, 刘益军, 娄义宝, 杨旭. 基于聚类及PCA分析的红壤坡耕地耕层土壤质量评价指标[J]. *农业工程学报*, 2018, 34(7): 155-164
- Jin H F, Shi D M, Chen Z F, Liu Y J, Lou Y B, Yang X. Evaluation indicators of cultivated layer soil quality for red soil slope farmland based on cluster and PCA analysis [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2018, 34(7): 155-164 (in Chinese)
- [14] 吴春生, 刘高焕, 黄翀, 刘庆生, 管续栋. 基于MDS和模糊逻辑的黄河三角洲土壤质量评估[J]. *资源科学*, 2016, 38(7): 1275-1286
- Wu C S, Liu G H, Huang C, Liu Q S, Guan X D. Soil quality assessment of the Yellow River Delta based on MDS and fuzzy logic model[J]. *Resources Science*, 2016, 38(7): 1275-1286 (in Chinese)
- [15] Yao R J, Yang J S, Gao P, Zhang J B, Jin W H. Determining minimum data set for soil quality assessment of typical salt-affected farmland in the coastal reclamation area[J]. *Soil and Tillage Research*, 2013, 128: 137-148
- [16] 杜森, 高祥照, 全国农业技术推广服务中心. 土壤分析技术规范[M]. 第2版. 北京: 中国农业出版社, 2006
- Du S, Gao X Z, National Agricultural Technology Extension and Service Center. *Technical Specification for Soil Analysis* [M]. 2nd ed. Beijing: China Agriculture Press, 2006 (in Chinese)
- [17] 韩微, 翟盘茂. 三种聚类分析方法在中国温度区划分中的应用研究[J]. *气候与环境研究*, 2015, 20(1): 111-118
- Han W, Zhai P M. Three cluster methods in regionalization of temperature zones in China[J]. *Climatic and Environmental Research*, 2015, 20(1): 111-118 (in Chinese)
- [18] 蒋端生. 红壤丘陵区耕地肥力质量演变规律及其影响因素研究[D]. 长沙: 湖南农业大学, 2008
- Jiang D S. Studies on the evolution law of the fertility quality of arable land and its influencing factors in red soil hilly areas [D]. Changsha: Hunan Agricultural University, 2008 (in Chinese)
- [19] 陈正发, 史东梅, 金慧芳, 娄义宝, 何伟, 夏建荣. 基于土壤管理评估框架的云南坡耕地耕层土壤质量评价[J]. *农业工程学报*, 2019, 35(3): 256-267
- Chen Z F, Shi D M, Jin H F, Lou Y B, He W, Xia J R. Evaluation on cultivated-layer soil quality of sloping farmland in Yunnan based on soil management assessment framework (SMAF) [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2019, 35(3): 256-267 (in Chinese)
- [20] Askari M S, Holden N M. Quantitative soil quality indexing of temperate arable management systems[J]. *Soil and Tillage Research*, 2015, 150: 57-67
- [21] Nash J E, Sutcliffe J V. River flow forecasting through conceptual models part I: A discussion of principles [J]. *Journal of Hydrology*, 1970, 10(3): 282-290
- [22] 王华, 黄宇, 阳柏苏, 黄璜, 张芬. 中亚热带红壤地区稻-稻-草轮作系统稻田土壤质量评价[J]. *生态学报*, 2005(12): 3271-3281
- Wang H, Huang Y, Yang B S, Huang H, Zhang F. Paddy soil quality assessment under rice-ryegrass rotation system in red soil region of mid-subtropics[J]. *Acta Ecologica Sinica*, 2005, 25(12): 3271-3281 (in Chinese)
- [23] Bünemann E K, Bongiorno G, Bai Z G, Creamer R E, Deyn G D, Goede R, Flesskens L, Geissen V, Kuyper T W, Mäder P, Pulleman M, Sukkel W, van Groenigen J W, Brussaard L. Soil quality: A critical review [J]. *Soil Biology and Biochemistry*, 2018, 120: 105-125
- [24] Zheng K, Cheng J, Xia J F, Liu G Y, Xu L. Effects of soil bulk density and moisture content on the physico-mechanical properties of paddy soil in plough layer[J]. *Water*, 2021, 13(16): 2290
- [25] 游松财, 邸苏闯, 袁晔. 黄土高原地区土壤田间持水量的计算

- [J]. 自然资源学报, 2009, 24(3): 545-552
 You S C, Di S C, Yuan Y. Study on soil field capacity estimation in the loess plateau region[J]. *Journal of Natural Resources*, 2009, 24(3): 545-552 (in Chinese)
- [26] 许静. 不同试验方法测定田间持水量的对比研究[D]. 长春: 吉林大学, 2018
 Xu J. Comparative study on the field capacity based on different test methods[D]. Changchun: Jilin University, 2018 (in Chinese)
- [27] 劳家桢. 土壤农化分析手册[M]. 北京: 农业出版社, 1988
 Lao J C. *Handbook of Soil Agricultural Chemistry Analysis* [M]. Beijing: Agriculture Press, 1988 (in Chinese)
- [28] 张凤荣, 安萍莉, 王军艳, 张军连, 刘黎明, 陈焕伟. 耕地分等中的土壤质量指标体系与分等方法[J]. 资源科学, 2002, 24(2): 71-75
 Zhang F R, An P L, Wang J Y, Zhang J L, Liu L M, Chen H W. Soil quality criteria and methodologies of farmland grading [J]. *Resources Science*, 2002, 24(2): 71-75 (in Chinese)
- [29] Liu Z J, Zhou W, Shen J B, Li S T, He P, Liang G Q. Soil quality assessment of albic soils with different productivities for Eastern China[J]. *Soil and Tillage Research*, 2014, 140: 74-81
- [30] 许明祥. 黄土丘陵区生态恢复过程中土壤质量演变及调控 [D]. 杨凌: 西北农林科技大学, 2003
 Xu M X. Soil quality evolvement mechanism in the process of ecosystem restoration and its management in loess hilly-gully region[D]. Yangling: Northwest A & F University, 2003 (in Chinese)
- [31] 冯瑞琦, 潘萍, 欧阳勋志, 臧颖, 具琳静, 李坚锋. 基于最小数据集的闽楠天然次生林土壤质量评价[J]. 江西农业大学学报, 2021, 43(3): 585-597
 Feng R Q, Pan P, Ouyang X Z, Zang H, Ju L J, Li J F. Soil quality evaluation of natural secondary forest of Phoebe bournei based on minimum data set [J]. *Acta Agriculture Universitatis Jiangxiensis*, 2021, 43(3): 585-597 (in Chinese)
- [32] Rastgou M, Bayat H, Mansoorizadeh M, Gregory A S. Estimating the soil water retention curve: Comparison of multiple nonlinear regression approach and random forest data mining technique[J]. *Computers and Electronics in Agriculture*, 2020, 174: 105502
- [33] 崔潇潇, 高原, 吕贻忠. 北京市大兴区土壤肥力的空间变异 [J]. 农业工程学报, 2010, 26(9): 327-333
 Cui X X, Gao Y, Lv Y Z. Spatial variability of soil fertility in Daxing District of Beijing [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2010, 26(9): 327-333 (in Chinese)
- [34] 吴玉红, 田霄鸿, 侯永辉, 周密, 同延安. 基于田块尺度的土壤肥力模糊评价研究[J]. 自然资源学报, 2009, 24(8): 1422-1431
 Wu Y H, Tian X H, Hou Y H, Zhou M, Tong Y A. Preliminary study on the subject concerning fuzzy assessment method of soil fertility based on field scale [J]. *Journal of Natural Resources*, 2009, 24(8): 1422-1431 (in Chinese)
- [35] 陈梦迪, 陈奇伯, 黎建强, 李佳璇, 王瑞璋. 枯落物不同输入方式对云南松林地土壤质量的影响评价[J]. 云南农业大学学报: 自然科学, 2020, 35(1): 149-155
 Chen M D, Chen Q B, Li J Q, Li J X, Wang R Z. Evaluation of the effects of litter input managements on the soil quality in *Pinus yunnanensis* forest [J]. *Journal of Yunnan Agricultural University: Natural Science*, 2020, 35(1): 149-155 (in Chinese)

责任编辑: 杨爱东