

基于高斯-分类混合聚类方法的水稻区域化肥减施潜力研究

张耿苗¹, 麻万诸², 任周桥², 刘桃霞³, 项彬彬⁴, 黄银飞⁵, 吕晓男², 邓勋飞^{2*}

(1 浙江省诸暨市农业技术推广中心, 浙江诸暨 311800; 2 浙江省农业科学院数字农业研究所/农业农村部农产品信息溯源重点实验室, 杭州 310021; 3 浙江省诸暨市东和乡农技站, 浙江诸暨 311833; 4 浙江省诸暨市陈宅镇农技站, 浙江诸暨 311823; 5 浙江省诸暨市璜山镇农技站, 浙江诸暨 311809)

摘要:【目的】综合土壤养分、自然资源禀赋和种植管理方式等多源要素, 构建基于高斯-分类混合聚类模型的农田基础肥力分区, 为定性描述、定量估计化肥的“稳-控-增”和提高推荐施肥的准确性提供必要的补充。

【方法】2008—2017年, 共采集6382个农田土壤样品, 现场调查了距村距离、土壤类型、耕层质地、成土母质、地貌类型与种植模式等数据, 测定了土壤pH及有机质、全氮、有效磷和速效钾含量。同时, 在研究区内典型位置以当地主栽的‘甬优’、‘秀水’系列品种为试材, 布设6个“3414”试验点, 共开展了28个试验, 获得氮、磷、钾(N、P、K)不同施肥水平下的稻谷产量、最佳推荐施肥量等信息。基于高斯-分类混合聚类方法, 联合土壤、气候、地形等自然资源禀赋要素与区域性农业种植模式, 对农田基础肥力进行分区, 并将各区测土配方试验数据与农户典型施肥数据比较, 完成各不同分区的化肥减施潜力估计。【结果】整个研究区域稻田可划分为7个大区, 各区与已有农田地力评价等级具有较好的吻合度, 对应于较高、中等和较低三级别地力水平分别有4、1和2个分区。影响各农田基础肥力分区的重要因子基本一致, 主要为地面高程、土壤速效钾含量、降雨、温度和地貌分区。研究区内水稻的化肥减施潜力分别为减纯氮(N)26.2~71.3(平均48.7)kg/hm², 减磷肥(P₂O₅)9.9~23.5(平均16.7)kg/hm², 减钾肥(K₂O)18.5~39.9(平均29.2)kg/hm²。减肥潜力与分区基础肥力高低总体一致, 较高肥力区减肥潜力较大, 低肥力区减肥潜力相对较小, 但也有较高肥力分区减肥潜力低(如分区3)或较低肥力分区减肥潜力高(如分区6)的情形。从空间分布上看, 分区4、5、6、7减氮潜力较大, 分区2、7减磷钾肥潜力较大, 而分区1、5减磷钾肥潜力较小。【结论】综合自然资源禀赋、土壤及种植管理模式的分区划分, 比传统依靠土壤养分的肥力分区能更加全面客观地反映区域稻田的基础肥力差异。地形地貌、土壤速效钾含量、降雨和温度是区域基础肥力分区的关键影响因子。从空间分布上定量估计, 水稻的化肥减施潜力与基础肥力水平总体一致, 总体可减施氮肥(N)26.2~71.3 kg/hm²、磷肥(P₂O₅)9.9~23.5 kg/hm²、钾肥(K₂O)18.5~39.9 kg/hm²。

关键词:混合聚类; 减肥潜力; 基础肥力分区; 推荐施肥量

Estimation of chemical fertilizer reduction potential for paddy rice using Gaussian-categorical mixture clustering methods

ZHANG Geng-miao¹, MA Wan-zhu², REN Zhou-qiao², LIU Tao-xia³, XIANG Bin-bin⁴,
HUANG Yin-fei⁵, LÜ Xiao-nan², DENG Xun-fei^{2*}

(1 Agricultural Technology Extension Station of Zhuji City, Zhejiang 311800, China; 2 Institute of Digital Agriculture, Zhejiang Academy of Agricultural Sciences/Key Laboratory of Information Traceability for Agricultural Products, Ministry of Agriculture and Rural Affairs, Hangzhou 310021, China; 3 Agricultural Technology Extension Station of Donghe Town of Zhuji City, Zhejiang 311833, China; 4 Agricultural Technology Extension Station of Chenzhai Town of Zhuji City, Zhejiang 311823, China;
5 Agricultural Technology Extension Station of Huangshan Town of Zhuji City, Zhejiang 311809, China)

Abstract:【Objectives】Comprehensive considering soil nutrient, landscape, and cropping managements, the paddy fields were divided into fertility zones using Gaussian and categorical mixture model, which would help

收稿日期: 2019-06-26 接受日期: 2019-08-03

基金项目: 国家重点研发计划项目(2016YFD0200802, 2018YFD0200502)。

联系方式: 张耿苗 E-mail: 75022979@qq.com; *通信作者 邓勋飞 E-mail: dengxf@zaas.ac.cn

figurate and quantitative estimation of chemical fertilizer input, and increase the precision of fertilizer recommendation. **[Methods]** During 2008–2017, 6382 paddy field samples were collected in Zhuji City, Zhejiang Province. The distance to village, soil types, surface soil texture, parent materials, landscape and cropping system of the samples were investigated *in situ*. Then the soil pH, organic matter, total N, available P and available K contents of the samples were measured in laboratory. During the same period, six sites with "3414" experiment design were selected using main rice cultivars of 'Yongyou' and 'Xiushui' as tested materials, and rice yield and optimal NPK fertilizer application rates covering different fertilization levels were obtained from a total of 28 field experiments. A hybrid cluster model combined Gaussian and categorical mixture models were developed for dividing fertility zones with different fertility levels. The fertilizer reduction potential of different zones was estimated based on the optimal fertilization strategy and regular fertilization data of local farmers.

[Results] The paddy fields in the whole study area could be divided into seven fertility zones, each of which had a good coincidence with the existed level of farmland fertility. There were four, one and two zones corresponding to the fertility levels with high, moderate and low, respectively. The dominant factors influencing the basic fertility zoning of each farmland were the elevation, soil readily available potassium content, annual precipitation, annual mean temperature and geomorphological zoning. The fertilizer reduction potential in paddy fields was 26.2–71.3 (average 48.7) kg/hm² for N, 9.9–23.5 (average 16.7) kg/hm² for P₂O₅, and 18.5–39.9 (average 29.2) kg/hm² for K₂O, respectively. The levels of fertilizer reduction potential were basically consistent with the basic fertility levels of the zones, and the reduction potential in high fertility zones was greater than that in the low fertility zones, there were also zones of higher fertility with low fertilizer reduction potential (e.g., zone 3) or zones of lower fertility with high fertilizer reduction potential (e.g., zone 6). From the spatial perspective, there was great reduction potential on nitrogen fertilizer in zones 4, 5, 6 and 7, and on phosphorus and potassium fertilizer in zones 2 and 7, while low reduction potential on phosphorus and potassium in zones 1 and 5.

[Conclusions] The paddy fields in Zhuji City have been divided into seven fertility zones using the Gaussian-categorical mixture clustering model. The zone division estimated by the model is much closer to the practical situation than that by traditional methods depend on soil nutrient contents. Overall, the fertilizer reduction potential is well agree with the estimated fertility zones, and the fertilizer reduction potential is 26.2–71.3 kg/hm² for N, 9.9–23.5 kg/hm² for P₂O₅, and 18.5–39.9 kg/hm² for K₂O, respectively.

Key words: mixture clustering; fertilizer reduction potential; basic fertility zone;
recommended fertilizer use rate

我国作为水稻主产国，水稻上的化肥施用量普遍偏高^[1-2]，东南沿海部分地区稻田氮肥平均施用量达300 kg/hm²以上^[3]，水稻化肥过量与不合理施用造成面源污染等环境风险问题比较严峻^[2, 4]。要实现化肥的合理使用和精准减量，需要以农田的基础肥力状况为基准，科学合理地划定施肥(管理)分区^[5]，并在此基础上进行区域肥料减量潜力分析，这对开展区域精准施肥推荐具有重要的基础性意义。

目前已有的水稻化肥减量潜力研究主要基于肥效-产量相关关系的田间试验与农户调查^[6]，通常以典型试验田(点)为载体，区域性的水稻肥料减量潜力估算鲜有报道^[3]。对于区域化肥减量施用推荐研究，通常是从研究土壤养分(氮、磷、钾、有机质)

含量差异为依据的施肥(管理)分区开始^[5, 7-8]。传统的农田管理分区方法过度倚重土壤养分含量的高低，忽视了广义的农田基础肥力还受其所处的气候环境、地形、土壤类型等“自然禀赋”，以及区域性的农业种植模式、灌溉条件等因素的共同影响^[9-10]。有研究表明，肥料使用效率受土壤肥力和气候条件综合影响^[11]。这些非养分要素的差异所造成的农田基础肥力差异将直接影响肥料减量潜力估算的科学性和准确性。高斯-分类混合聚类模型是基于不同类型数据间的条件独立性假设，其条件分布函数同时调用高斯和分类模型实现聚类分析^[12]。它与高斯混合聚类、模糊(Fuzzy)聚类等方法类似，是一种数据驱动的聚类方法，其分区数是由数据本身的特性决定

的, 受人为主观因素影响小, 能够较好地解释自然界的连续变异现象^[8]。近年来, 相关聚类分析开始在土壤学中得到广泛应用, 如土壤肥力质量评价^[13-14]、农田作业管理单元分区^[15]和农田土壤碳景观分区^[16-17], 但目前主要集中在管理分区划定环节, 鲜有与多年施肥试验相结合开展肥料减量应用。随着近年来大量的测土配方施肥调查与肥效试验积累, 为开展农田分区与区域减量施肥提供了重要的基础数据支撑^[7, 18-19]。鉴于此, 将承载农田基础肥力的各因子进行融合, 通过聚类分析获得农田肥力及禀赋特征不同的综合性施肥分区, 然后结合各聚类区域当地农户的常规施肥习惯, 以及精准测土配方施肥推荐试验来综合确定化肥减量潜力, 用于指导区域化肥减量。

基于2008—2017年积累的测土配方施肥试验数据、农田调查数据, 开展浙江省诸暨市县域范围水稻的化肥减量潜力量化估算研究。主要完成以下两个目标: 1) 根据土壤条件、地形地貌、气候环境和种植管理等自然和人为影响因子, 联合使用高斯-分类混合聚类模型, 获得区域农田基础肥力分区; 2) 将各分区的测土配方试验数据与农户典型施肥数据比较, 量化估计各分区的水稻化肥减量潜力。该农田基础肥力分区与化肥减量潜力可服务于区域宏观化肥减施指导。

1 材料与方法

1.1 研究区域

本研究试验区域为浙江省绍兴市诸暨市全域范围 ($119^{\circ}52'58''\sim120^{\circ}32'13''E$, $29^{\circ}21'29''\sim$

$29^{\circ}58'59''N$)。该区域属于典型的亚热带季风气候, 常年积温5210 h, 年降水量1270 mm, 无霜期大于236天。区域内兼有河网平原、河谷平原、低丘大畈、丘陵山地, 其中水田面积约3.65万hm², 近年来主栽单季稻, 其主导品种为甬优系列、秀水与嘉优系列。水稻土亚类涵盖潜育型水稻土、渗育型水稻土和潴育型水稻土等, 土壤质地以粘土、粘壤土为主。研究区域位置、稻田空间分布及试验调查情况如图1所示。

1.2 数据获取与处理

1.2.1 农田调查与试验数据 2008—2017年, 以20~33 hm²为一个点, 共采集了6382个农田土壤样品, 并现场调查了距村距离、土壤类型、耕层土壤质地、成土母质、地貌类型与种植模式等数据, 以及部分典型农户(23户)的水稻习惯施肥数据。并对农田土壤pH及有机质、全氮、有效磷和速效钾含量等指标进行分析, 指标项按《土壤分析技术规范》进行测定^[20]。同时, 在研究区内典型位置布设6个点的“3414”试验, 共28个试验。获得当地主栽甬优、秀水系列品种氮(N)、磷(P)、钾(K)不同施肥水平下的稻谷产量、最佳推荐施肥量等信息。

1.2.2 其他数据 研究区内历年月平均温度和降雨量数据采用全球气象数据产品WorldClim2 (<http://worldclim.org/version2>)^[21], 该数据是基于1970—2000年全球9000~60000个气象站点, 并综合高程及土地利用等变量插值得到, 空间精度约1 km, 通过归一化处理成年降水量和年平均温度。此外, 研究区内地面数字高程数据集为美国航空航天

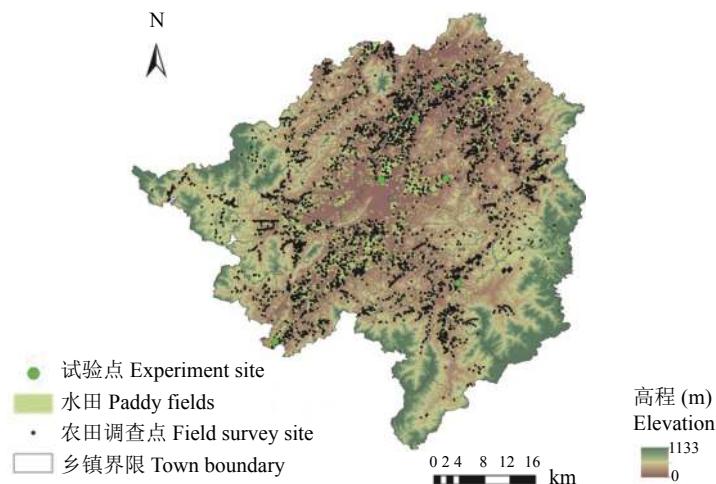


图1 研究区域调查点分布概况

Fig. 1 Sampling location in paddy fields of Zhuji City, Zhejiang Province

局 (NASA) 的 Shuttle Radar Topography Mission version 3 (SRTMGL1)^[22], 该数据最新的时间节点为 2000 年, 空间精度为 30 m (表 1)。此外, 获得了 2009 年的土地利用现状数据 (1 : 10000), 以及诸暨市耕地地力评价指数分级空间分布数据 (图 1), 并相应对水田类型进行了提取。

1.2.3 数据预处理 将研究区域范围的水田分布数据生成 90 m 空间格网, 并转换成栅格数据集。以该栅格数据为基准, 农田土壤养分、土壤类型、耕层土壤质地、年平均温度、地貌类型、种植模式、年降雨量和高程等数据通过栅格重采样 (resample)、邻近空间分析与克里格 (kriging) 插值等方法将其统一赋值/关联到 90 m 精度的基准格网中。

1.3 农田基础肥力分区方法

农田 pH、土壤养分 (全氮、有效磷、速效钾和有机质)、土壤类型、耕层土壤质地、地貌类型分区、种植模式、年平均温度、降雨量和地面高程, 是影响农田土壤基础肥力的重要因子, 这些因子往往通过交互或综合作用于土壤基础肥力空间变异^[9]。因此, 通过建立混合聚类模型, 将这些要素综合纳入考量, 对研究区域进行农田基础肥力分区。

1.3.1 聚类分区方法与步骤 本研究采用高斯-分类混合聚类模型 (Gaussian-categorical mixture clustering models)^[12], 该模型基于不同类型数据间的条件独立性假设, 同时调用不同模型高斯和分类函数模型实现聚类分析。该模型可以同时输入数值型变量 (如高程) 和分类变量 (如土壤类型), 以及自动

计算部分缺失数据。其核心算法 SemiSEM 是基于数学期望值最大化 (expectation maximum, EM) 实现^[12]。算法简要说明如下:

首先 (Step I), 初始化分布参数 (θ^0), 基于当前数据的条件分布密度函数 (t_{ik}^{m-1}) 和参数 (θ^{m-1}), 计算缺失数据 (X_{ik}^{m-1}) ($i=1, \dots, n$);

然后 (Step E), 计算在分布参数 (θ^{m-1}) 下的条件分布密度函数 (t_{ik}^m) ($1 < i < n; 1 < k < K$);

最后 (Step M), 基于条件混合权重 (p_k , $k=1, \dots, K$; $p_1+ \dots + p_K = 1$) 使条件分布密度函数最大化, 获得新的最大似然估计量 θ 。

重复迭代 Step E 和 Step M, 直至收敛。

1.3.2 聚类分区划定标准 聚类分区的判定主要依据贝叶斯信息准则 (bayesian information criterion, BIC)^[16], 设置分区数 1~25, BIC 越小, 分区特征聚合效果越好。在实际运用中, 需要考虑现实研究范围特点, 分区指导工作的实际可操作性, 以及现有“3414”试验点位分布、数量和农户肥料使用调查范围, 来综合确定合理的聚类分区数目。

1.3.3 基础肥力分区合理性验证 分区是否合理主要从以下两个方面进行验证: 1) 各分区内的主要变量的变异系数 (CV) 趋于减小; 2) 与已有的区域农田基础养分调查和耕地地力评价等级 (基于立地条件、土壤理化性状和灌溉条件等 14 项指标, 旨在反映农田的生产能力) 相互对照。具体地, 一方面通过比较不同分区下的稻田有机质、全氮、速效钾和有效磷含量 4 项养分指标的变异系数和综合养分含量水平

表 1 土壤属性、气象条件、地形地貌和种植模式

Table 1 Soil properties, meteorological conditions, topography and cropping patterns in the tested area

数据分类 Category	数据内容 Content	空间精度 Spatial resolution	时间跨度 Temporal coverage	数据来源 Data source
土壤属性 Soil properties	pH, 土壤养分 (有机质、全氮、速效钾、有效磷)	点 Point	2008—2017	农田调查 Field survey
	Soil nutrients (organic matter, total N, available K, and available P)			
	耕层土壤质地 (壤土、粘壤土、粘土)	点 Point	2008—2017	农田调查 Field survey
	Soil texture (loam, clay loam, and clay)			
气象 Meteorology	水稻土亚类 (潴育型、潜育型、渗育型)	点 Point	2008—2017	农田调查 Field survey
	Paddy soil subgroup (hydrogric, gleyed, and perco-genic)			
	降雨量、平均温度	1 km	1970—2000	WorldClim2 ^[21]
	Precipitation and average temperature			
地形地貌 Topography	地貌分区 (丘陵、低丘大畈、水网平原、河谷平原)	点 Point	2008—2017	农田调查 Field survey
	Topo-zone (hilly, hilly fields, alluvial plain, and valley plain)			
	地面高程 Elevation	30 m	2008—2017	农田调查 Field survey
种植模式 Cropping pattern	单季稻、双季稻、稻-麦 (油)	点 Point	2008—2017	农田调查 Field survey
	Single rice, double rice, and rice-wheat (oil rape)			

变化。其中综合养分含量水平的计算方法如下:

$$F_i^c = P_i^{tN} \times 0.25 + P_i^{vK} \times 0.25 + P_i^{vP} \times 0.25 + P_i^{oM} \times 0.25 \quad (1)$$

式中: P_i^{tN} 为分区 i 中全氮平均含量在所有全氮含量样本中所处的百分位 ($0 < P_i^{tN} < 100$); P_i^{vK} 、 P_i^{vP} 和 P_i^{oM} 分别为分区 i 中速效钾、有效磷和有机质平均含量在所有样本中所处百分位; 0.25 为平均权重 (均分)。 F_i^c 表示稻田综合养分指数, F_i^c 越大表示养分水平越高, 反之养分水平越低。

另一方面, 将得到的基础肥力分区与研究区域现有的耕地地力评价等级分布图进行面积加权空间叠加, 分别提取各分区内的耕地地力等级 [1~5, 其中: 1~2 表示高地力, 3~4 表示中等地力, ≥ 5 表示低地力], 并将各分区内的地力等级进行平均, 得到各分区内的平均地力等级, 平均地力等级越低表示耕地地力越好, 反之表示耕地地力越差。

1.3.4 输入变量重要性估计 随机森林 (random forest) 方法是构建多棵树 (森林) 对样本开展训练并进行预测分类 (或回归)^[23], 它可以用来估计各输入变量对预测 (分类) 结果的重要性。本研究以聚类输出的农田分区为因变量, 聚类中的各输入因子为自变量, 如温度、土壤类型、地貌类型分区等, 构建相应随机森林模型进行训练, 得到影响各农田基础肥力分区聚类的因子重要性排序。

1.4 稻田肥力减量潜力估算

根据确定的稻田基础肥力聚类分区数量, 通过空间匹配关联“3414”试验点(共 6 个, 图 1)。如某一聚类分区存在“3414”试验点, 则采用最优施肥方案 (N2P2K2) 条件下的纯 N、 P_2O_5 和 K_2O 施用量为水稻推荐施肥量 (一个分区多个试验点取平均值); 若某一分区内没有“3414”试验点, 则搜寻相邻分区中具有“3414”试验点且相邻面积较大者, 取该邻近分区的最优施肥方案 (N2P2K2) 下的纯 N、 P_2O_5 和 K_2O 施用量作为目标分区的水稻推荐施肥量。同时, 考虑到“3414”试验点只是某一分区内的典型代表, 本研究将该推荐施肥量上下浮动 5%, 作为整个分区的推荐施肥量的变化范围。此外, 将区域内的各典型农户习惯 (常规) 施肥数据与各分区关联, 得到各分区中当地农户常规施肥数据 (纯 N、 P_2O_5 和 K_2O)。同理, 将该典型的常规施肥数据上下浮动 5%, 来弥补常规施肥数据调查的不确定性。则整个研究区域水稻的减氮肥潜力 (Pr_N)、减磷肥潜力 ($Pr_{P_2O_5}$) 和减钾肥潜力 (Pr_{K_2O}) 估算如下:

$$Pr_N = \sum_{j=1}^n (F_{Nj} - R_{Nj}) \quad (2)$$

$$Pr_{P_2O_5} = \sum_{j=1}^n (F_{P_2O_5j} - R_{P_2O_5j}) \quad (3)$$

$$Pr_{K_2O} = \sum_{j=1}^n (F_{K_2Oj} - R_{K_2Oj}) \quad (4)$$

R_{Nj} 、 $R_{P_2O_5j}$ 、 R_{K_2Oj} 和 F_{Nj} 、 $F_{P_2O_5j}$ 、 F_{K_2Oj} 分别表示第 j ($j=1, \dots, n$) 个肥力分区下推荐纯氮 (N) 施用量和农户常规纯氮 (N) 施用量, 其中最大潜力 ($MaxPr$)=农户调查施肥量 $\times (1+5\%)$ -最优推荐量 $\times (1-5\%)$ 。平均潜力 (Pr)=农户调查施肥量-最优推荐量; 最小潜力为 ($MinPr$)=农户调查施肥量 $\times (1-5\%)$ -最优推荐量 $\times (1+5\%)$ 。整个研究区内的水稻化肥减量潜力是采用分区面积加权平均方法将各分区化肥减量潜力相加, 分别得到分区纯 N、 P_2O_5 和 K_2O 减量潜力变化范围 ($MinPr \sim MaxPr$) 和平均潜力值 (Pr); 类似地, 研究区内平均化肥减量幅度是分别将各分区纯 N、 P_2O_5 和 K_2O 减量潜力平均值除以该分区常规化肥施用量平均值计算得到相应减量百分比, 然后采用分区面积加权平均得到整个研究区的平均减量幅度。

1.5 数据统计分析

数据的处理、统计分析主要是基于 R 软件^[24], 应用到的主要程序包为 raster、stat、randomForst 和 MixAll, 分别进行栅格数据处理、统计分析、聚类模型分析与变量重要性分析等; 此外, 应用美国 Esri 公司 ArcGIS 10.3 桌面软件开展调查数据的空间插值、拓扑分析和部分数据的可视化作图。

2 结果与讨论

2.1 稻田基础肥力分区确立

2.1.1 分区数量确定 当聚类数量小于 5 时 BIC 值快速下降, 当聚类数量等于 7 后 BIC 值下降速度趋于平缓 (图 2)。该结果说明较多的聚类数量用于聚类分析模拟的效果较好, 但分区数量增加的边际效应在快速递减。本研究中, 考虑到分区的数量越多, 可用的代表性 6 个“3414”试验点落在所有分区所占比例越小, 同时还要兼顾过多的分区并不利于实际应用中可操作性技术指导方案的推行, 统筹考虑以生成 7 个聚类分区为宜。6 个“3414”试验点分别落在了其中 5 个聚类分区 (1、2、5、6、7 区) 中, 无“3414”试验点的两个分区是 3、4。通过比较相邻分区得到相关施肥试验数据。已有的研究表明, 在土壤景观分区或有机碳储量分区聚类分析中, 基于高斯混合或多项式混合聚类方法, 在分区数为 10 个左右, 能达到较好的应用效果^[16-17]。

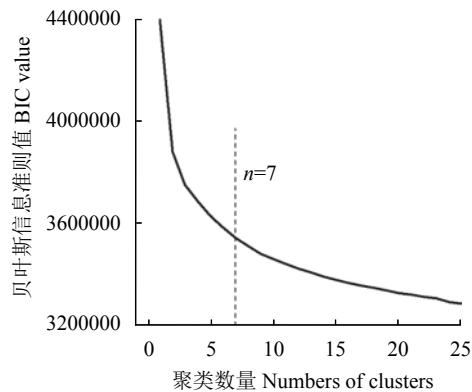


图 2 混合聚类分析中 BIC 与聚类数量设置的变化趋势
Fig. 2 The relationship between the number of clusters and BIC in mixture cluster analysis

2.1.2 基础肥力分区确立与验证 整个诸暨市县域范围内的水田被划分为 7 个分区，各分区之间总体相对独立连片，部分分区之间有穿插（图 3a）。分区 1、2、5 相对比较集中，主要分布于研究区域中部地区（东北—西南方向）的水网平原，分区面积较大；分区 6、7 分散在平原和丘陵山地交接地带，主要集中于该区域内的低丘大畈农田中，分区面积稍小；而分区 3、4 则基本散落在丘陵山区的农田中（图 1、图 3a），分区面积较小且细碎。各分区农田速效钾和有效磷含量差异较大，有机质和全氮含量差异小（表 2）。

为了验证农田基础肥力分区结果是否可以作为减量施肥潜力估算的依据，将建立的 7 个分区与已有诸暨市县域耕地地力评价图比较^[25]，分区趋势与耕地地力分级总体一致（图 3b）。通过两者的空间叠加计算各分区的平均耕地地力等级，发现分区（1、2、

3、5）、分区 7 和分区（4、6）分别大致对应于较高（1、2、3 级）、中等（3、4 级）和较低（3、4、5 级）3 个耕地地力等级（表 2）；从空间分布上看，除分区 3 外其余分区的基础肥力水平与耕地地力等级具有较好的一致性。各分区的综合养分指数（F）表明，分区 1、4、5 的稻田综合养分水平较高，分区 2、3、6 次之，分区 7 养分水平最低，这与耕地地力水平存在较大差异。主要原因是本研究中基础肥力分区是一个综合因素作用下的分区结果，稻田养分水平只是代表了其中的一部分作用。此外，土壤基础肥力分区内的土壤养分（速效钾、有机质、有效磷和全氮）含量变异情况表明，与分区前相比，多数分区土壤养分含量的变异系数有不同程度减小（表 2）。其中，分区 4、6 中部分养分含量的变异出现不同程度的增加，直接的原因可能是由于这两个分区空间跨域较大（图 3a）造成的。深层的原因主要是源于本研究中 7 个分区是权衡模型表现与可用数据综合确定的，并非是模型最优的聚类数量，并且，农田基础肥力分区中还将气候与地形地貌等因素纳入考量，某些主导的分区因子的变异系数降低（一致性）很可能掩盖了部分养分因子的变异增加所带来的效用。如分区 4、6 的地面高程变异系数分别从 104% 降低到了 41.8% 和 23.8%，降雨的变异系数分别从 1.9% 降低到 0.9% 和 1.4%。

一般来讲，农田的基础肥力受土壤基础养分外的气候、灌溉和地形等综合影响，尤其是在更大的区域尺度上（如省、全国、甚至全球）这些因素的作用更为突出^[9-10]。本研究中，综合自然资源禀赋、土

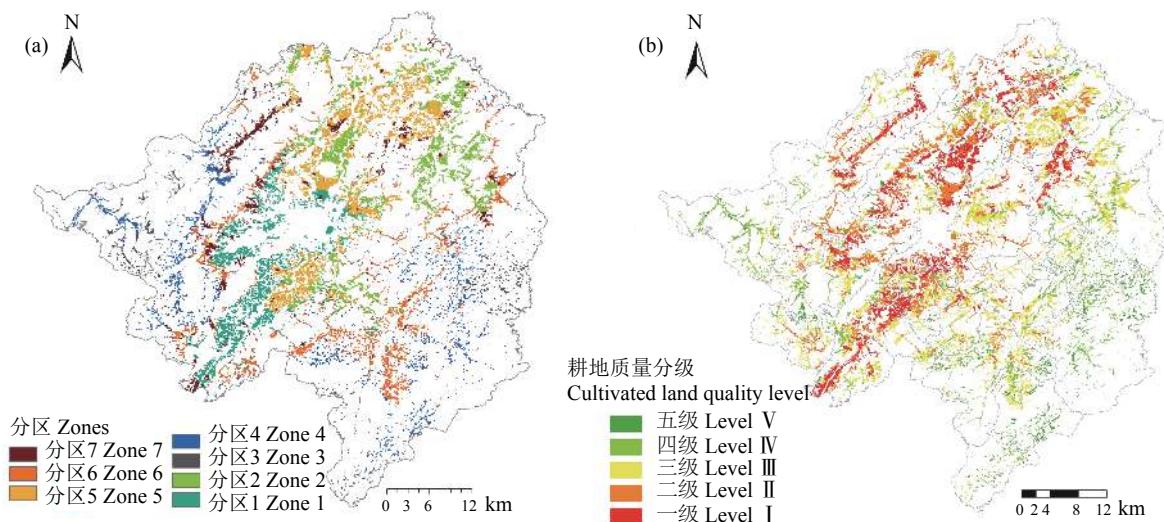


图 3 研究区域农田基础肥力分区 (a) 和农田耕地地力评价 (b) 空间数据分布
Fig. 3 Spatial distribution of basic fertility zone for paddy fields (a) and cultivated land quality in Zhuji City (b)

表 2 土壤分区中各养分数据统计
Table 2 Statistics of soil properties in zones

分区 Zone	样点数 Sample number	有机质 Organic matter		全氮 Total N		有效磷 Available P		速效钾 Available K		F	SFG
		平均 Mean (g/kg)	CV (%)	平均 Mean (g/kg)	CV (%)	平均 Mean (g/kg)	CV (%)	平均 Mean (g/kg)	CV (%)		
分区 1 Zone 1	592	35.7	21.2	2.1	20.3	17.7	89.9	121.4	18.2	68	2.59
分区 2 Zone 2	1452	36.0	16.5	2.1	16.7	19.1	91.7	64.0	16.2	53	2.79
分区 3 Zone 3	710	33.8	17.5	2.0	16.3	12.6	83.5	71.6	17.8	47	2.58
分区 4 Zone 4	148	37.3	22.1	1.9	30.7	15.6	145.1	107.6	60.4	63	3.86
分区 5 Zone 5	1226	38.7	13.3	2.2	13.8	22.6	92.7	92.3	12.2	73	2.62
分区 6 Zone 6	694	34.7	19.7	1.7	33.7	12.8	93.0	82.4	31.1	46	3.51
分区 7 Zone 7	1560	31.6	20.1	1.8	19.2	14.3	86.6	63.1	22.6	37	3.19
全部 All	6382	35.1	20.8	2.0	21.4	16.9	96.3	78.4	33.1	55	2.93

注 (Note) : F—综合养分指数 Combined nutrition index; SFG—平均地力等级 Average fertility grade.

壤及种植管理模式的分区划分, 比传统依靠土壤养分的农田管理/肥力分区^[5, 7-8, 26], 能更加全面客观的反映农田真正的基础肥力。在实际农田管理分区应用过程中, 由于影响实际分区聚类的输入数据的类型往往是多源的, 很可能同时受分类变量和连续数值型变量的叠加影响。目前常用的聚类模型, 如高斯混合聚类, 只能处理数值型数据, 不能直接输入分类变量, 或需将分类变量转换为虚拟变量 (dummy variables) 输入, 但是这个策略会因其数值型变量与虚拟变量可比性差而产生较大的模型应用偏差^[17]。高斯-分类混合聚类模型是基于不同类型数据间的条件独立性假设, 通过同时调用其不同条件分布函数 (即高斯与分类) 实现混合聚类分析^[12]。本研究中的农田土壤基础肥力分区经数据差异性验证及与现有耕地地力评价结果的对比验证, 说明了分区结果的有效性和合理性, 这与陈彦等^[8]、陈桂芬等^[26]基于土壤养分数据的农田管理分区及土壤肥力分区结果趋势类似。该分区结果可以为后续区域内化肥减量、稻田生产能力评价等提供辅助决策支持。

2.1.3 分区重要性因子 地面高程和速效钾含量是决定农田肥力分区聚类的 2 大重要因素, 其贡献率分别达到 30.6% 和 21.8%, 此外, 气象因素 (温度和降雨) 也共同解释了 29.1% 的分区变异影响。土壤全氮、有机质、有效磷含量和 pH 等数据对分区的影响较小 (合计 11.8%), 而土壤类型、耕层土壤质地和种植模式对分区的影响很小, 平均不足 1% (图 4)。

本研究地面高程因子对分区的重要性最高, 可

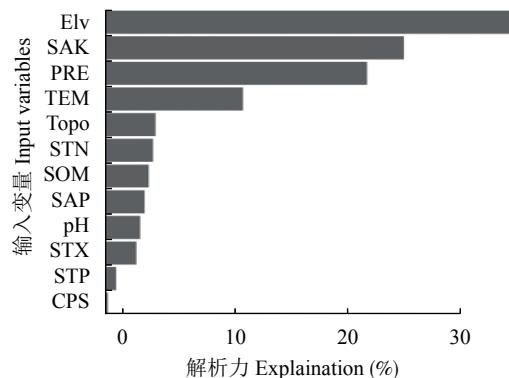


图 4 输入变量对农田基础肥力分区的重要性

Fig. 4 Importance of input predictors for dividing basic fertility zone of paddy fields

[注 (Note) : Elv—高程 Elevation; SAK—土壤速效钾 Soil available potassium; PRE—年降雨量 Annual precipitation; TEM—年平均温度 Annual mean temperature; Topo—地形分区 Topography; STN—土壤全氮 Soil total nitrogen; SOM—土壤有机质 Soil organic matter; SAP—土壤有效磷 Soil available phosphorus; STX—耕层土壤质地 Soil texture; STP—土壤类型 Soil type; CPS—种植模式 Cropping pattern.]

能与研究区内平原、丘陵在空间上都有不同程度的分布及空间上差异明显相关 (图 1), 其重要性也与聚类分区后地面高程变异系数的显著下降相一致 (从总体 104% 降到 25%~61%)。同时, 考虑到温度和降雨数据产品中将地面高程作为一个重要的因子进行空间插值预测^[21], 因此较高的温度和降雨重要性也受到了地面高程叠加影响。从表 2 可见, 速效钾作为土壤养分因子, 其总体变异程度较大, 并在聚类分区后绝大多数情况下速效钾变异系数都明显下降;

虽然有效磷总体变异更大，但分区后其变异程度差异变化不大，因此其重要性相对较低。此外，研究区内水稻种植模式以单季稻为主，土壤类型(水稻土亚类)和土壤耕层质地差异性较小，相应重要性很低。虽然特征变量之间的相关性可能对各因子的重要性判别产生潜在影响，然而随机森林基于决策树分类器和 Bagging 的集成学习机制，可以较好地避免变量之间的相关性、异常噪声等对各因子重要性判别的干扰，通常能取得让人信服的结果^[27]。本研究中明确农田基础肥力分区下各因子的重要性差异，可辅助决策用于筛选农田质量提升措施进行农田肥力改良。

2.2 稻田化肥减施潜力

研究区内水稻的化肥减施潜力分别为 N 26.2~71.3 kg/hm²，平均 48.7 kg/hm²；P₂O₅ 9.9~23.5 kg/hm²，平均 16.7 kg/hm²；K₂O 18.5~39.9 kg/hm²，平均 29.2 kg/hm²(表 3)。化肥平均减量幅度 N 为 16.8%、P₂O₅ 为 20.1%、K₂O 为 21.6%。分区基础肥力水平与单位面积化肥减量潜力总体一致，即较高基础肥力水平化肥减量潜力较大，较低肥力水平化肥减量潜力较小。但仍有不少分区并不一致性地符合上述情况，如基础肥力较高的分区 3 单位面积减氮、减钾潜力相对较低，而较低基础肥力分区 6 减氮、减钾潜力与高肥力水平分区 1、5 接近，中等基础肥力分区 7 的减磷、减钾的潜力又与高肥力水平分区相当。从空间分布上看(图 3a、图 5)，分区 4、5、6、7 减氮潜力较大(N > 65 kg/hm²)，分区 2、

7 磷钾肥减量潜力较大(P₂O₅ > 20 kg/hm²、K₂O > 40 kg/hm²)，而分区 1、5 磷钾肥减量潜力较小(P₂O₅ < 16 kg/hm²、K₂O < 25 kg/hm²)。

本研究揭示了该研究区内不同基础肥力分区下减施氮肥、磷肥和钾肥的潜力，可为当地农技部门开展区域肥料减量指导提供借鉴。高斯-分类混合聚类方法综合了农田基础肥力分区、农户调查和多年施肥推荐试验，对区域稻田化肥减量潜力进行了空间量化分布估算，是对传统依靠农户调查或依靠土壤养分划分农业管理分区进行“稳-控-增”定性描述、定量估计和推荐施肥的重要补充^[7-8, 19]。该方法将农田养分以外的信息纳入考量，得到的稻田基础肥力分区更趋合理，表 2 的结果也验证了本研究中基础肥力分区与传统常规养分分区的不一致性，优化了减肥潜力估算的分区前提；同时，该方法可以将分类变量和数值变量同时进行聚类计算，避免了分类变量虚拟化转换的信息损失或不匹配造成数据结果的偏差^[17]。并且，引入了当地多年肥料调查与测土配方试验数据，使得肥料减量参考依据的真实性和区域代表性较好；此外，该方法通过数据驱动将特性相近的区块聚合到一起，通过采集较少的定点施肥试验和调查数据即能得到区域性的肥料减量空间估计，比一般依赖大量稻田减肥试验点和复杂模型参数优化才能完成区域空间估算的稻田减肥模型的适用性和可操作性更强。

另一方面，由于受到现有可用数据的限制，本研究中的肥料潜力估算方法需要在最优聚类分区数

表 3 稻田化肥减量潜力估算
Table 3 Estimated potential of chemical fertilizer reduction in paddy fields

分区 Zone	基础肥力水平 Basic fertility level	面积 Area (× 10 ⁴ hm ²)	减量潜力 Reduction potential (kg/hm ²)			重要影响因子 Important predictors
			N	P ₂ O ₅	K ₂ O	
分区 1 Zone 1	较高 High	0.30	22.8~75.2 (52.4)	11.7~21.8 (15.0)	24.5~39.9 (29.1)	SAK/Elv/PRE/TEM
分区 2 Zone 2	较高 High	0.68	31.7~75.2 (48.5)	8.6~29.7 (21.5)	14.2~57.1 (43.4)	PRE/Elv/SAK/TEM
分区 3 Zone 3	较高 High	0.59	29.5~66.3 (45.9)	8.3~22.8 (16.7)	18.3~30.5 (21.6)	PRE/Elv/TEM/SAK
分区 4 Zone 4	较低 Low	0.21	21.7~65.0 (44.9)	13.2~19.1 (13.2)	29.6~28.7 (19.8)	Elv/PRE/TEM/SAK
分区 5 Zone 5	较高 High	0.75	25.5~69.4 (49.4)	10.5~18.9 (13.1)	12.7~30.5 (21.5)	SAK/Elv/PRE/TEM
分区 6 Zone 6	较低 Low	0.46	24.7~73.6 (52.7)	7.3~20.8 (14.7)	10.9~33.5 (23.9)	Elv/PRE/TEM/Topo
分区 7 Zone 7	中等 Moderate	0.66	29.3~72.6 (47.7)	7.2~27.0 (19.4)	12.4~49.3 (36.9)	Elv/SAK/PRE/TEM
全部 All		3.65	26.2~71.3 (48.7)	9.9~23.5 (16.7)	18.5~39.9 (29.2)	Elv/SAK/PRE/TEM

注 (Note) : Elv—高程 Elevation; SAK—土壤速效钾 Soil available potassium; PRE—年降雨量 Annual precipitation; TEM—年平均温度 Annual mean temperature; Topo—地形分区 Topography. 基础肥力水平是通过与已有诸暨市县域耕地地力评价图对应比较得到^[25]。The basic fertility level of different zones was estimated according to the spatial map of cultivated land quality for Zhuji County.

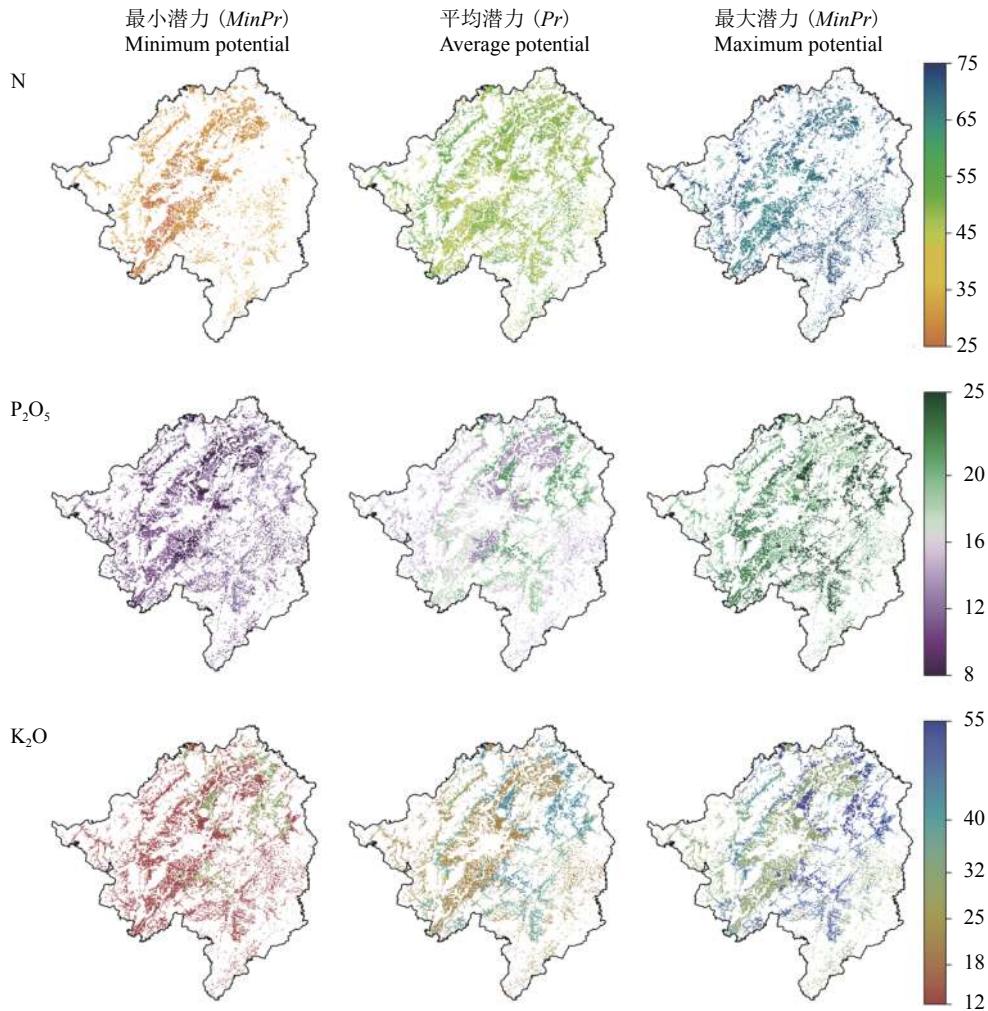
图 5 稻田化肥减量 (kg/hm^2) 潜力空间分布图

Fig. 5 Spatial distribution of reduction potential for chemical fertilizer in paddy fields

与实际可用数据之间进行综合权衡。通常分区划分结果随着聚类数量的增加而优化，因此这种非最优状态的分区数量的确定具有一定的主观性，尤其是在可用数量较少的时候。本研究确定的 7 个分区就是聚类数量和可用数据相互权衡的结果(图 2)，相应地造成了稻田化肥减量潜力估算的不确定性。如分区 3、4 施肥推荐方案采用了邻近分区的试验方案，减量潜力估计可能存在较大偏差；同时，每个分区内的代表性“3414”施肥试验点仍然偏少(一般只有一个)，虽然研究中对施肥试验推荐方案作了上下浮动 5% 的考量，仍有可能没有包括实际可能存在的大部分最优施肥推荐方案。此外，作为一个数据驱动的聚类方法，它的准确性和有效性不可避免的受输入数据可用性、数据量大小和数据本身准确性的影响^[16]。对于数据类型、精度比较一致，能涵盖分区主要影响因子的情况，聚类通常能发挥其最佳效用^[8]。

虽然本研究中采用了较高质量的气象数据、高精度地形数据以及大量的农田调查点所得到的养分含量数据等基础数据，使得稻田基础地力的划分具有较高的可靠性，然而仍有必要进一步挖掘可能对稻田基础肥力以及最优水稻施肥推荐量产生有效影响的因素，如目前的肥料减量潜力估计并未完全考虑所有水稻品种(只采用了施肥试验中当地的主栽品种)，其他品种的水稻由于作物养分吸收效率等差异^[28]，也可能会造成最优肥料推荐方案偏差，从而影响其减肥潜力估算的准确性。

本研究中区域的减肥潜力估算仅仅是提供优化宏观推荐施肥的一个基准性参考，实际应用中仍需要结合主栽作物的需肥特性和土壤养分状况进行精准减量试验。今后将进一步拓展的“3414”试验范围和水稻品种选择，每年轮回采集土壤样品，定期开展理化性状监测。同时在实际肥料减量试验推荐

中,充分考虑有机肥和化肥联合使用,进一步试验明确不同生育期的化肥减量潜力。随着大数据和人工智能在作物养分亏缺无损监测等方面的应用,为实现基准分区+精准修正的精细化减量施肥推荐打下基础^[29]。

3 结论

1)该分区拓展了原有依赖于土壤基础养分状况的分区模式,更全面地体现了区域稻田的基础肥力差异。并且明确了地形地貌、土壤速效钾含量、降雨和温度是区域基础肥力分区的关键影响因子,但不同分区下指标因子的重要性有所不同。2)水稻的化肥减施潜力高低与基础肥力水平高低总体一致。从空间分布上定量估计,总体可减氮肥(N)26.2~71.3 kg/hm²,减磷肥(P₂O₅)9.9~23.5 kg/hm²,减钾肥(K₂O)18.5~39.9 kg/hm²。

参 考 文 献:

- [1] Chen J, Huang Y, Tang Y H. Quantifying economically and ecologically optimum nitrogen rates for rice production in south-eastern China[J]. *Agriculture, Ecosystems & Environment*, 2011, 142: 195–204.
- [2] Peng S B, Tang Q Y, Zou Y B. Current status and challenges of rice production in China[J]. *Plant Production Science*, 2009, 12: 3–8.
- [3] 郭俊杰,柴以潇,李玲,等.江苏省水稻减肥增产的潜力与机制分析[J].*中国农业科学*,2019,52(5): 849–859.
Guo J J, Chai Y X, Li L, et al. The potential and related mechanisms of increasing rice yield by reducing chemical nitrogen application in Jiangsu Province[J]. *Scientia Agricultura Sinica*, 2019, 52(5): 849–859.
- [4] 孙铖,周华真,陈磊,等.农田化肥氮磷地表径流污染风险评估[J].*农业环境科学学报*,2017,36(7): 1266–1273.
Sun C, Zhou H Z, Chen L, et al. The pollution risk assessment of nitrogen and phosphorus loss in surface runoff from farmland fertilizer[J]. *Journal of Agro-Environment Science*, 2017, 36(7): 1266–1273.
- [5] 郭熙,赖锦春,赵小敏,等.基于GIS丘陵土壤分区的水稻施肥配方研究[J].*中国农业科学*,2011,44(2): 307–315.
Guo X, Lai J C, Zhao X M, et al. Fertilizer recommendations on rice of soil distribution at hilly area by GIS[J]. *Scientia Agricultura Sinica*, 2011, 44(2): 307–315.
- [6] 曹寒冰,王朝辉,赵护兵,等.基于产量的渭北旱地小麦施肥评价及减肥潜力分析[J].*中国农业科学*,2017,50(14): 2758–2768.
Cao H B, Wang Z H, Zhao H B, et al. Yield based evaluation on fertilizer application and analysis of its reduction potential in Weihei dryland wheat production[J]. *Scientia Agricultura Sinica*, 2017, 50(14): 2758–2768.
- [7] 陶云彬,章日亮,方守地,等.兰溪市农区土壤养分时空变化与施肥分区的研究[J].*浙江农业科学*,2010,(6): 1384–1388.
- Tao Y B, Zhang R L, Fang S D, et al. Study on spatial-temporal variation of soil nutrients and fertilization zoning in agricultural areas of Lanxi City[J]. *Journal of Zhejiang Agricultural Sciences*, 2010, (6): 1384–1388.
- [8] 陈彦,吕新.基于FCM的绿洲农田养分管理分区研究[J].*中国农业科学*,2008,41(7): 2016–2024.
Chen Y, LÜ X. Definition of management zones of soil nutrients based on FCM algorithm in Oasis field[J]. *Scientia Agricultura Sinica*, 2008, 41(7): 2016–2024.
- [9] 贺彩平.基于GIS和模糊聚类的巫山县烟草施肥区划研究[D].重庆:西南大学硕士学位论文,2013. 5–17.
He C P. Fertilizer regionalization of tobacco based on GIS and fuzzy clustering method in Wushan County[D]. Chongqing: MS Thesis of Southwest University, 2013, 5–17.
- [10] 陈丽,郝晋珉,艾东,等.基于改进农业生态区划法的耕地综合量能指数模型构建[J].*农业工程学报*,2014,30(20): 268–276.
Chen L, Hao J M, Ai D, et al. Comprehensive index model building of cultivated land yield and productivity based on improved agro-ecological zoning method[J]. *Transactions of the CSAE*, 2014, 30(20): 268–276.
- [11] 邹应斌,夏冰,蒋鹏,等.水稻生产目标产量确定的理论与方法探讨[J].*中国农业科学*,2015,48(20): 4021–4032.
Zou Y B, Xia B, Jiang P, et al. Discussion on the theory and methods for determining the target yield in rice production[J]. *Scientia Agricultura Sinica*, 2015, 48(20): 4021–4032.
- [12] Iovleff S. MixAll: Clustering mixed data with missing values [EB/OL]. (2016-11-24)[2019-03-30] <https://cran.r-project.org/web/packages/MixAll/vignettes/Introduction-Mixtures.pdf>.
- [13] 陈欢,曹承富,张存岭,等.基于主成分-聚类分析评价长期施肥对砂姜黑土肥力的影响[J].*土壤学报*,2014,51(3): 609–617.
Chen H, Cao C F, Zhang C L, et al. Principal component-cluster analysis of effects of long-term fertilization on fertility of lime concretion black soil[J]. *Acta Pedologica Sinica*, 2014, 51(3): 609–617.
- [14] 吴海燕,金荣德,范作伟,等.基于主成分和聚类分析的黑土肥力质量评价[J].*植物营养与肥料学报*,2018,24(2): 325–334.
Wu H Y, Jin R D, Fan Z W, et al. Assessment of fertility quality of black soil based on principal component and cluster analysis[J]. *Journal of Plant Nutrition and Fertilizers*, 2018, 24(2): 325–334.
- [15] 张泽.基于GIS的土壤氮素分区管理与施肥模型建立研究[D].新疆石河子:石河子大学博士学位论文,2015. 47–49.
Zhang Z. Study on partitioned management of soil nitrogen and fertilization model establishment based on GIS[D]. Shihezi, Xinjiang: PhD Dissertation of Shihezi University, 2015. 47–49.
- [16] Chen S C, Arrouays D, Angers D A, et al. National estimation of soil organic carbon storage potential for arable soils: A data-driven approach coupled with carbon-landscape zones[J]. *Science of the Total Environment*, 2019, 666: 355–367.
- [17] Mulder V L, Lacoste M, Martin M P, et al. Understanding large-extent controls of soil organic carbon storage in relation to soil depth and soil-landscape systems[J]. *Global Biogeochemical Cycles*, 2015, 29: 1210–1229.
- [18] 陆宏.聚类分析在测土配方施肥中的应用[J].*中国土壤与肥料*,

- 2012, (1): 99–102.
- Lu H. Application of cluster analysis in soil testing and formulated fertilization[J]. *Soil and Fertilizer Sciences in China*, 2012, (1): 99–102.
- [19] 李娟, 章明清, 孔庆波, 等. 构建县域早稻氮磷钾施肥的系统聚类方法研究[J]. *植物营养与肥料学报*, 2017, 23(2): 531–538.
- Li J, Zhang M Q, Kong Q B, et al. Building fertilization categories of N, P and K fertilization for early rice using systematic clustering method in county territory[J]. *Journal of Plant Nutrition and Fertilizers*, 2017, 23(2): 531–538.
- [20] 全国农业技术推广服务中心. 土壤分析技术规范(第二版)[M]. 北京: 中国农业出版社, 2006. 36–73.
- National Agro-Tech Extension & Service Center. Technical specification for soil analysis (2nd edition)[M]. Beijing: China Agriculture Press, 2006. 36–73.
- [21] Fick S E, Hijmans R J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas[J]. *International Journal of Climatology*, 2017, 37: 4302–4315.
- [22] NASA LP DAAC. SRTMGL1: NASA Shuttle Radar Topography Mission Global 1 arc second V003[DB/OL]. [2015-10-27]. <https://10.5067/MEaSUREs/SRTM/SRTMGL1.003>.
- [23] Breiman L. Random forests[J]. *Machine Learning*, 2001, 45: 5–32.
- [24] R Core Team. R: A language and environment for statistical computing[Z]. Vienna: R Foundation for Statistical Computing, 2018.
- [25] 吕晓男, 倪治华. 浙江省土壤资源与耕地地力等级地图集[M]. 哈尔滨: 哈尔滨地图出版社, 2013. 95–97.
- Lü X N, Ni Z H. Soil resources and cultivated land quality atlas of Zhejiang[M]. Harbin: Harbin Cartographical Publishing House, 2013. 95–97.
- [26] 陈桂芬, 曹丽英, 王国伟. 加权空间模糊动态聚类算法在土壤肥力评价中的应用[J]. *中国农业科学*, 2009, 42(10): 3559–3563.
- Chen G F, Cao L Y, Wang G W. Application of weighted spatially fuzzy dynamic clustering algorithm in evaluation of soil fertility[J]. *Scientia Agricultura Sinica*, 2009, 42(10): 3559–3563.
- [27] 何云, 黄翀, 李贺, 等. 基于 Sentinel-2A 影像特征优选的随机森林土地覆盖分类[J]. *资源科学*, 2019, 41(5): 992–1001.
- He Y, Huang C, Li H, et al. Land-cover classification of random forest based on Sentinel-2A image feature optimization[J]. *Resources Science*, 2019, 41(5): 992–1001.
- [28] 朱兆良, 张绍林, 尹斌, 等. 太湖地区单季晚稻产量-氮肥施用量反应曲线的历史比较[J]. *植物营养与肥料学报*, 2010, 16(1): 1–5.
- Zhu Z L, Zhang S L, Yin B, et al. Historical comparison on the response curves of rice yield-nitrogen application rate in Tai Lake Region[J]. *Journal of Plant Nutrition and Fertilizers*, 2010, 16(1): 1–5.
- [29] 何山, 孙媛媛, 沈掌泉, 等. 大数据时代精准施肥模式实现路径及其技术和方法研究展望[J]. *植物营养与肥料学报*, 2017, 23(6): 1514–1524.
- He S, Sun Y Y, Shen Z Q, et al. Advances in coupling big data technique with nutrient site-specific management: scheme, methods and outlook[J]. *Journal of Plant Nutrition and Fertilizers*, 2017, 23(6): 1514–1524.