

基于 FCM 的绿洲农田养分管理分区研究

陈彦, 吕新

(新疆生产建设兵团绿洲生态农业重点实验室, 新疆石河子 832003)

摘要: 【目的】以天山北麓绿洲农耕区——新疆农八师 148 团农田为研究对象, 以 193 个耕层土壤 (0~30 cm) 有机质、碱解氮、速效磷和速效钾含量的分析数据为变量进行农田土壤养分精确管理分区研究。【方法】模糊 c -均值聚类法被用来进行分区, 以棉田产量为外部变量, 采用 FPI、 c - ϕ 多次组合法及基于外部变量的多元回归法来确定适宜的模糊控制参数。【结果】研究区最佳分区数为 4, 模糊指数为 1.6。各管理分区土壤养分的变异系数都较分区前全研究区有所减小, 而分区间土壤养分差异显著。研究区的平均混乱度指数为 0.19, 不同模糊类别交叠程度小, 地理空间上土壤的隶属关系相对明确。【结论】通过选取适宜的外部变量, 模糊 c -均值聚类法可以较好地进行管理分区划分, 分区结果可以作为变量施肥的单独作业单元进行耕作管理。

关键词: 管理分区; 模糊聚类; 土壤养分; 外部变量; 绿洲农田

Definition of Management Zones of Soil Nutrients Based on FCM Algorithm in Oasis Field

CHEN Yan, LÜ Xin

(Key Laboratory of Oasis Ecology and Agriculture of Xinjiang Production and Construction Group, Shihezi 832003, Xinjiang)

Abstract: 【Objective】The objective of this research was to define management zones of oasis cotton field. The variables of organic matter, available N, available P and available K data determined in 193 top soil (0-30 cm) samples were selected as data sources. 【Method】Fuzzy c -means clustering algorithm was used to delineate management zones. In order to determine the optimum fuzzy control parameters, the fuzziness performance index (FPI), c - ϕ combinations and the multiple regression based on external variable were used. Meanwhile, the cotton yield was chosen as the external variable. 【Result】The whole field was divided into four management zones. And the fuzziness exponent was 1.6. The zoning statistics showed that variation coefficient of soil nutrients decreased, while the means of the soil nutrients differed markedly between management zones. The average confusion index was 0.19 in all management zones. The overlapping of fuzzy classes at points was low and the spatial distribution of membership grades was unambiguous. 【Conclusion】The results indicated that fuzzy c -means clustering algorithm could be used to delineate management zones by selecting the appropriate external variables. The defined management zones can be used for fertilizer recommendation to manage soil nutrients more efficiently.

Key words: Management zones; Fuzzy c -means clustering; Soil nutrients; External variable; Oasis field

0 引言

【研究意义】绿洲是干旱区三大地理景观 (山地、绿洲、荒漠) 之一。由于光、热、水、土的特殊组合, 在绿洲农耕区形成了独具特色的绿洲农业模式^[1,2]。在

内陆型绿洲进行基于多种土壤属性的农田精确管理分区研究, 不仅可以认识绿洲农业生态环境特征及其特殊的生产功能, 对绿洲农业的合理开发及可持续发展也有着重要作用。【前人研究进展】依据土壤属性差异, 将同一地块分成不同的均质性区域 (管理分区)

收稿日期: 2007-11-14; 接受日期: 2008-01-21

基金项目: “十一五” 国家科技支撑计划 (2006BAD21B02-2); 新疆兵团重大科技专项-精准农业与信息技术应用与示范 (2007ZX03)

作者简介: 陈彦 (1979-), 男, 河北安新人, 博士, 研究方向为绿洲生态和地理信息系统。Tel: 13999530263; E-mail: yanc1208@163.com。
通讯作者吕新 (1964-), 河北安国人, 博士, 研究方向为绿洲生态和地理信息系统。Tel: 13909931721; E-mail: lxshz@126.com

进行管理正成为目前精准农业的一个研究热点^[3], 而聚类法在管理分区的研究中应用最为广泛。但对于 K-均值聚类等硬聚类法, 其聚类对象被严格地区分出来, 一个对象只能划分到一个类中, 经过处理后原来数据中所蕴含的信息损失较多。国内李翔等^[4]采用基于改进 K-均值聚类法的空间连续性聚类算法 (SCKM) 进行管理分区研究, 兼顾了空间单元的空间位置关联性与属性数据的相似性, 为空间单元区域划分问题的解决提供了有效途径。与硬聚类方法相比, 模糊聚类方法则是具有一定柔性特征的分类方法, 它将样本对各个类的隶属度扩展到 $[0,1]$ 区间, 并且认为每一个样本与各个聚类中心都有一个隶属度关系, 能够有效地对类与类之间有交叉的数据集进行聚类, 其中又以模糊 c -均值聚类法 (FCM) 应用最广^[5]。但 FCM 法存在模糊控制参数 (模糊类别数、模糊加权指数) 不易确定、对处于不同类边界处的元素分辨能力不高等问题^[6], 这些成为众多研究者不断探索的课题。国内李艳等^[7,8]在海涂围垦区的管理分区研究中, 默认模糊加权指数为 2, 引用模糊性能指数 (FPI) 和归一化分类熵 (NCE) 对农田管理分区数 (模糊类别数) 进行定量化表达。檀满枝等^[9,10]在土壤重金属的分类研究中采用 c - φ 多次组合法同时确定了适宜模糊类别数和模糊加权指数。【本研究切入点】不过 Bragato^[11]和 de Bruin^[12]的研究指出以聚类过程自身获所获信息为基础的计算方法并不能总是有效地鉴别数据集的最适分类, 一些研究中出现了 FPI 和 c - φ 多次组合法对于同一研究区域计算出的模糊类别数不相同的情况。对此, de Bruin 等^[12]提出了基于外部变量的多元回归法, 该法并不借助从聚类过程所获的信息, 而是通过分析外部变量空间变异性与模糊数据集对应隶属度等级的空间分布之间的关系来确定适宜的模糊控制参数, 其计算结果可以作为 FPI 和 c - φ 多次组合法等方法的有效补充。此类方法在国内精确管理分区的研究应用中尚无报道。【拟解决的关键问题】本研究以内陆绿洲农业模式下大型农田为研究对象, 以多种土壤养分数据源, 应用模糊 c -均值聚类算法定义田间管理分区。并结合前人的研究基础, 以棉田产量为外部变量, 综合运用 FPI、 c - φ 多次组合法和基于外部变量的多元回归方法对模糊控制参数的选择进行对比研究, 同时利用混乱度指数 (CI) 表征样点模糊隶属关系的不确定性复杂程度, 评价对北疆绿洲农业模式下的大型农田进行精确管理分区的可行性。

1 材料与方法

1.1 研究区概况

试验区新疆生产建设兵团农八师 148 团垦区位于天山北麓中段准葛尔盆地南缘, 呼图壁河、塔西河、玛纳斯河冲积湖积平原中部, 地处 $44^{\circ}46'50'' \sim 44^{\circ}59'25''N$, $86^{\circ}9'10'' \sim 86^{\circ}22'51''E$ 。气候属典型的大陆性气候, 降水稀少, 光热资源丰富。全区地势平坦, 主要土壤类型有灰漠土、风沙土和盐土 3 个土类, 共计 5 个亚类, 9 个土属, 由于新疆自身特殊的气候、成土母质及地理原因, 土壤养分含量呈现缺氮、少磷、富钾的特点。区内种植作物以棉花为主, 现有耕地面积 1.28×10^4 ha, 其中滴灌棉田面积 1.1×10^4 ha, 农业技术采取集中管理, 不同地块施肥量 (包括追肥量) 一致。

1.2 土壤数据获取与样品分析

2006 年 10 月在 148 团滴灌棉区进行不规则取样, 以每个采样点的采样中心为圆心、4.5 m 为半径取 0~30 cm 耕作层土壤样品 5 个, 制成混合土样, 共采集土壤混合样 193 份。实测了以每个采样点为中心, 长 5.86 m、宽 2.25 m 的两条地膜内的棉花产量, 折合成公顷单产作为该样点的棉花产量数据, 采样点分布见图 1。获取的土壤样品经自然风干、过筛后进行室内养分测定。每个样品测定项目包括有机质、碱解氮、有效磷、速效钾, 按《土壤农化分析》^[13]中方法进行测定。

1.3 模糊聚类与插值

模糊 c -均值聚类法 (FCM) 是常用的一种非监督聚类方法, 被大量用于土壤、地形、地貌、产量和遥感数据等的分类中。对于一个空间样点数量为 n , 受测土壤属性数量为 p 的多源数据集, 模糊 c -均值聚类法能够确定其适宜模糊类别数 c , 定义聚类中心并计算出每一分类等级的模糊隶属度。它的基本思想是寻找目标函数的迭代最小化, 最常用的目标函数为:

$$J(MC) = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^{\varphi} d_{ij}^2 \quad i=1, \dots, n; \quad j=1, \dots, c \quad (1)$$

式中, C 是聚类类别质心矩阵 $c \times p$, M 是样本模糊隶属度矩阵 $n \times c$, μ_{ij} ($\mu_{ij} \in [0,1]$) 为第 i 个样点的土壤属性对于第 j 个聚类类别的隶属度值, φ ($\varphi \geq 1$) 是一个可控制聚类结果和类别交叠程度的常数^[14], 即模糊加权指数 (fuzziness exponent), d_{ij}^2 是第 i 个样点土壤属性与第 j 个类别中心之间距离的平方。

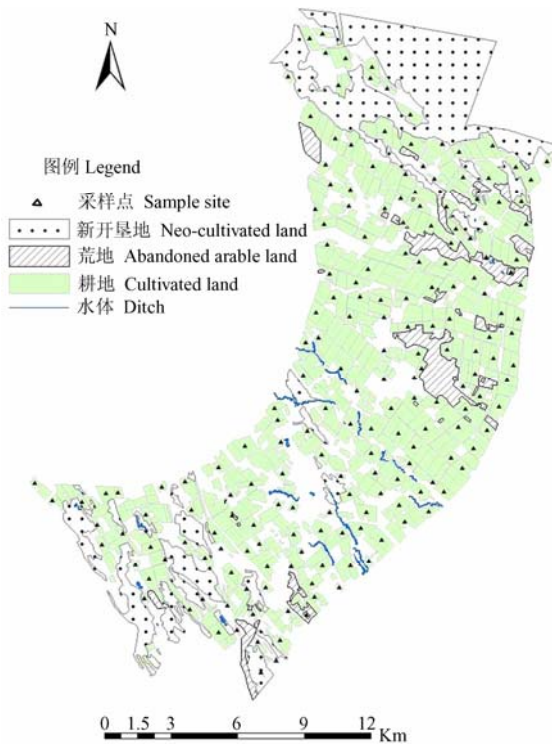


图 1 采样点空间分布图

Fig. 1 Distribution of sampling points

模糊类别数 c 、模糊加权指数 φ 两个参数的选择和确定直接影响连续分类的输出结果。McBratney 等^[15]提出用模糊效果指数 (fuzziness performance index, FPI) 来确定适宜的模糊类别数。FPI 是表示不同类别间共享的成员量的一个指数, 用来度量 c 个类别之间的分离程度, FPI 值的范围介于 0~1 之间, 值接近于 0 表示不同分区间共享的成员越少, 也即不同管理分区间差异越大; 当接近于 1 时表示分区间共享的成员越多, 管理分区间的差异越小。McBratney 和 Moore 还建议用 c - φ 多次组合最优取值法来同时确定 c 、 φ 。即对于一个相对于 φ 的派生函数 $[(\delta J / \delta \varphi) c^{0.5}]$, 使其峰值最小的 c 值为最优选择, 在选定 c 值的前提下, 峰值最大处的 φ 值为最优选择^[11]。

$$FPI = 1 - \frac{cF - 1}{c - 1} \quad (2)$$

式中, c 为模糊类别数, $F = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^2 / n$ 。

以上两种方法都是以从聚类过程自身获得的信息为基础, 被称为内部标准分析^[11]。内部标准分析并不能总是有效地鉴别数据集的最适分类, 有时甚至会出现计算最适模糊类别数时两种方法的计算结果不尽

一致的情况^[12]。相反, 外部标准分析所涉及的外部变量没有参与聚类分析, 但同样受影响聚类过程的外部条件所影响。对外部标准分析而言, 若所选外部变量的空间变异性与土壤养分对应隶属度等级的空间分布显示较强的相关关系, 则模糊 c -均值聚类能更好地表达数据的聚类效果, 所确定的模糊控制参数也适宜, 内部标准分析不能测试这种关系。因此, 当内部标准分析的计算结果出现矛盾时, 在选择合适外部变量的前提下采用外部标准分析可以确定适宜的模糊控制参数。de Bruin 和 Stein 提出一种多元回归方法来计算模糊隶属度与外部变量之间的关系。公式 (3) 即是模糊隶属度与外部变量的回归模型, 式中, β_0 、 β_1 、... β_c 是回归系数, ε 是随机误差, 回归模型的校正决定系数 r_a^2 被用来表达不同取值的 c 和 φ 之间的联系, 最大 r_a^2 处的 c 和 φ 为最优选择。在公式 (4) 中, r^2 是回归公式的决定系数, n_s 是外部变量的样点数目。

$$\text{外部变量} = \beta_0 + \beta_1 \mu_{A1}(x) + \dots + \beta_c \mu_{Ac}(x) + \varepsilon \quad (3)$$

$$r_a^2 = r^2 - \frac{c(1 - r^2)}{n_s - c - 1} \quad (4)$$

本研究选取的外部变量为研究区 2006 年的棉花产量数据, 其依据如下:

①农田土壤资源的空间位置分布除地带性分布规律外, 还受区域性地形、母质、水文、地质和人类耕作活动等条件的影 响。其中人类的耕作、施肥、灌溉等活动对农田土壤的演化起着越来越重要的作用, 并成为决定土壤肥力发展方向的基本动力之一。

②作为土壤生产力外在表现的作物产量是由土壤本身的肥力属性和发挥肥力作用的外界条件 (气候、地形、人为耕作、栽培等土壤管理措施及相关联的排水、供水条件) 所决定的。如前所述, 这些外界条件同样也影响土壤资源的空间分布和演化。

③产量不包含在用于聚类分析的数据集内。

根据模糊聚类结果, 将样点土壤对于不同土壤养分含量类别的部分隶属关系进行普通克里格插值, 获得研究区土壤养分单一类别模糊图。并依据最大隶属度原则确定样点所归属的模糊类别, 经插值和栅格转换, 生成研究区土壤属性管理分区图。同时, 为进一步明确不同土壤属性类别在各样点处的交叠程度, 利用混乱指数法 (confusion index, CI) 对样点土壤隶属关系进行模糊不确定性评价^[16]:

$$CI = 1 - [\mu_{\max i} - \mu_{(\max i - 1) i}] \quad (5)$$

式中, $\mu_{\max i}$ 为第 i 个样点土壤对于某土壤属性类别最大的隶属度值, $\mu_{(\max i - 1) i}$ 为第 i 个样点土壤第二大

隶属度值。

本研究中模糊 c -均值聚类是通过 MATLAB7.0 实现的；模糊隶属度的空间变异特征分析、半方差函数的计算、理论模型的拟合及 Kriging 插值和图形绘制是由 GS+3.5 和 ArcGIS9.2 共同完成的。

2 结果与分析

2.1 描述性统计结果

表 1 为研究区土壤养分及棉花产量数据的描述性统计结果，数据经域值识别法检验没有特异值存在。从均值来看，依据新疆土壤普查资料^[17]，研究区有机质和速效磷含量属于中等水平，碱解氮含量中等偏下，速效钾含量丰富。根据变异系数 (CV) 的大小可粗略估计变量的变异程度：弱变异性， $CV < 10\%$ ；中等变异性， $CV = 10\% \sim 100\%$ ；强变异性， $CV > 100\%$ ^[18]。

表 1 土壤属性及棉花产量的描述性统计分析

Table 1 Descriptive statistics of soil properties and cotton yield

土壤属性 Soil properties	最小值 Min	最大值 Max	平均值 Mean	标准差 SD	变异系数 CV(%)
有机质 OM ($\text{g}\cdot\text{kg}^{-1}$)	3.80	22.50	10.37	2.92	28.18
碱解氮 Alkali-hydrolyzable N ($\text{mg}\cdot\text{kg}^{-1}$)	20.00	102.00	47.63	12.85	26.98
速效磷 Available P ($\text{mg}\cdot\text{kg}^{-1}$)	3.60	40.60	12.29	5.66	46.03
速效钾 Available K ($\text{mg}\cdot\text{kg}^{-1}$)	161.00	519.33	318.02	58.26	18.32
棉花产量 Cotton yield ($\text{kg}\cdot\text{ha}^{-1}$)	1297.42	2925.15	1970.18	374.49	19.00

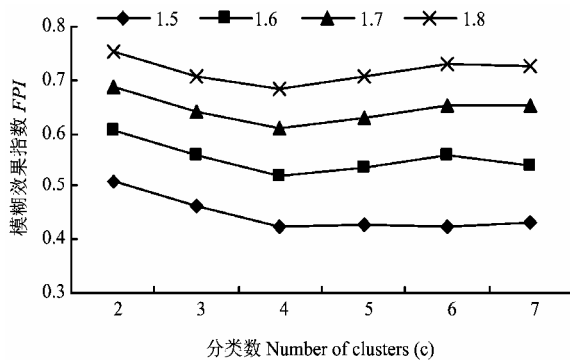


图 2 FPI 随管理分区数增加的变化趋势

Fig. 2 FPI plotted against the number fuzzy classes

当内部标准分析不能有效鉴别数据集的最适分类时，可以采用基于外部变量的多元回归法确定适宜的模糊控制参数。与 c - ϕ 多次组合法相同，此方法的结果允许同时选择 c 和 ϕ 。以研究区 2006 年的棉花产量

研究区土壤养分虽经长年大规模的农业机械化作业及膜下滴灌等人为因素的影响，在空间上仍表现为中等强度的变异，满足精准农业管理分区划分和实施变量施肥管理的前提条件。

2.2 土壤养分模糊分类结果输出

把 $n \times p$ ($n=193$, $p=4$) 的土壤养分模糊集带入模糊 c -均值算法中，依据前文讨论的相关方法进行模糊聚类。为了确定适宜的模糊类别数 c 及模糊加权指数 ϕ ，对比了 c 取 2~10, ϕ 取 1.1~2.0 时, FPI 及 $-\left[\frac{\delta J}{\delta \phi} c^{0.5}\right]$ 的结果。图 2 为 ϕ 取不同值时, FPI 随管理分区数增加的变化趋势图。可以看出，当 $c=4$ 时, FPI 降到了最低点。最适的模糊指数通过 ϕ 的派生函数 $-\left[\frac{\delta J}{\delta \phi} c^{0.5}\right]$ 来确定。图 3 显示，当 $\phi=1.5$ 时，曲线达到了波峰，但最缓的曲线为 $c=3$ ，与 FPI 获得的结果不同。

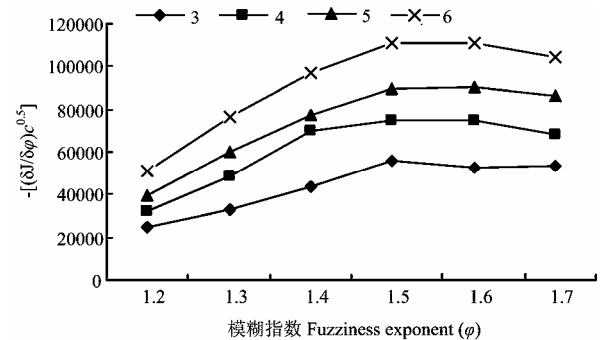


图 3 模糊类别 c 不同取值时的模糊指数

Fig. 3 Fuzziness exponent (ϕ) against $-\left[\frac{\delta J}{\delta \phi} c^{0.5}\right]$ for different class numbers of c

数据为外部变量，分别计算产量与不同 c 、 ϕ 的模糊隶属度的回归模型，模型的校正决定系数与 ϕ 的变化关系见图 4。可以看出， r_a^2 随着 ϕ 的增加而增加，并在

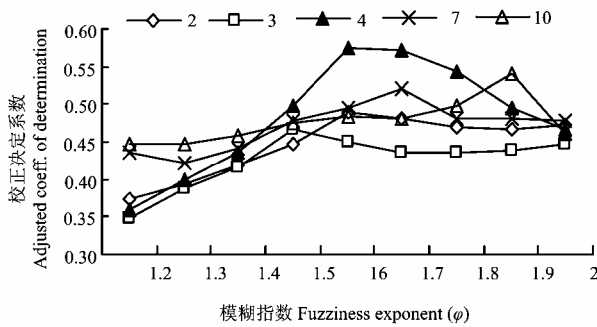


图 4 模糊类别与模糊指数不同取值时模糊隶属度回归方程的校正决定系数

Fig.4 Adjusted coefficient of determination (r_a^2) as a function of class partial memberships for different combination of c and ϕ

$c=4$ 、 $\phi=1.6$ 时达到最大值 0.57。

最终的模糊控制参数为： $c=4$ 、 $\phi=1.6$ ，即研究区可以划分为 4 个土壤养分管理分区，模糊 c -均值聚类算法生成的聚类中心矩阵 C 及样点对土壤各聚类类别的部分隶属度值分别见表 2、表 3。为了说明各样点在隶属关系上的交叠程度，将每个样点的混乱指数 CI 也列于表 3。聚类中心是多维空间上属性变量的平均值，能准确揭示不同类别中样品土壤属性数据以及数据组合之间的差异^[9]。根据表 2 中土壤属性类别中心值数据，对样品土壤分属的类别（管理分区）进行评价：管理分区间土壤属性含量变化平稳，分区 1 的有机质、碱解氮平均含量最高，具有最高土壤养分水平；分区 2 的总体养分水平则最低，分区 3、4 居中。模糊分类的输出结果包含了多种土壤养分在属性空间和地

表 2 研究区土壤养分模糊类别中心

Table 2 Cluster centers of soil nutrients

类别 Fuzzy class	有机质 OM($g \cdot kg^{-1}$)	碱解氮 Alkali-hydrolyzable N($mg \cdot kg^{-1}$)	速效磷 Available P($mg \cdot kg^{-1}$)	速效钾 Available K($mg \cdot kg^{-1}$)
C1	13.19	50.62	15.71	411.15
C2	8.19	42.52	10.04	222.89
C3	10.89	47.31	12.48	345.27
C4	9.95	48.18	11.96	293.50

表 3 研究区土壤类别隶属度和混乱度指数

Table 3 Partial memberships and confusion index of the sampled soils to the classes

土壤样点 Soil sample	分区 1 隶属度 Zone 1 memberships	分区 2 隶属度 Zone 2 memberships	分区 3 隶属度 Zone 3 memberships	分区 4 隶属度 Zone 4 memberships	混乱度指数 Confusion index
1	0.001	0.028	0.014	0.957	0.071
2	0.086	0.022	0.625	0.267	0.643
...					
193	0.001	0.944	0.004	0.052	0.108

理空间上的分布信息，对实施变量施肥及制定各管理分区田管措施具有重要的指导意义。

2.3 分类结果的可视化表达

模糊分类的输出结果所蕴含的空间信息只能作为空间预测的基础，并不能直接体现连续分类在空间预测中的重要作用^[9]。因此，必须利用地统计学手段实现样点土壤定量隶属关系在地理空间上的可视化表达，直观揭示土壤养分在整个研究区的空间分布与变异特征，进而对研究区土壤肥力状况进行评价。基于表 3 中的数据，获取隶属度半方差函数，然后对每个样点土壤各分区的部分隶属度值进行普通 Kriging 插

值，获得未采样土壤对于不同模糊类别的隶属度值，从而实现研究区土壤肥力分布状况的空间预测。

表 4 是 4 个分区模糊隶属度值的变异函数理论模型及相应参数。分区 1 符合高斯模型，分区 2 符合指数模型，分区 3、4 可用球状模型拟合。4 个分区的主轴变程在 4.35~6.81 km 之间，亚轴变程在 2.23~3.47 km 之间。块金值与基台值的比值均小于 25%，表现为强烈的空间自相关性^[19]。图 5 为研究区土壤肥力单一模糊隶属度空间分布图。

从图 5-b 看出，研究区北部土壤对于分区 2 的隶属度明显高于其它区域，此区域与古尔班通古特沙漠

表 4 模糊分区的半方差模型参数

Table 4 Semivariogram parameters of fuzzy classes

管理分区 Zones	模型 Model	块金值 C_0 Nugget	基台值 $C+C_0$ Sill	块金值/基台值 $C_0/(C+C_0)$ (%)	主轴变程 Major range (km)	亚轴变程 Minor range (km)
Zones 1	高斯 Gaussian model	0.37	2.75	13.46	5.57	2.88
Zones 2	指数 Exponential model	0.56	2.44	23.14	4.35	2.23
Zones 3	球状 Spherical model	0.34	2.93	11.60	6.81	3.47
Zones 4	球状 Spherical model	0.36	2.56	13.87	5.10	2.92

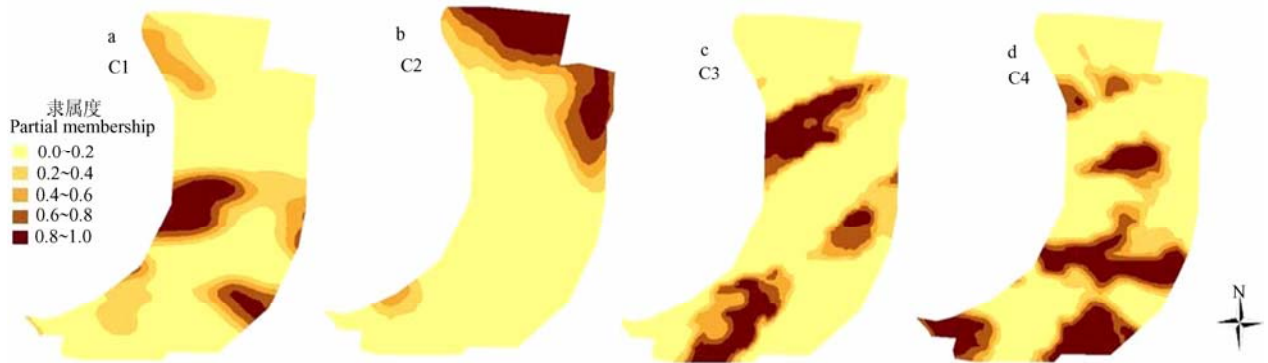


图 5 研究区土壤模糊隶属度值空间预测分布图

Fig. 5 Predictive maps showing spatial variation of partial memberships of soil to individual fuzzy classes derived respectively

接壤，主要为灌溉固定风沙土土属和灌溉苏打盐化风沙土土属，分布着大量的新开垦耕地，耕作年限短，耕作层不明显，保水保肥性差，土壤肥力相对其它区域贫瘠。研究区中部零星区域土壤对于分区 4 的隶属度最高（图 5-d），此区域是研究区的撂荒地分布区，主要为撂荒盐土土属，地下水位高，水质矿化度大，土层含盐量高，自然植被稀疏，土壤潜在养分含量低。西北及西南部土壤对于分区 3 的隶属度高（图 5-c），此区域分布有灌溉苏打盐化灰漠土土属、灌溉 $SO_4^{2-}-Cl$ 盐化灰漠土土属和灌溉 $Cl-SO_4^{2-}$ 盐化灰漠土土属，具有一定程度的次生盐化，土壤养分含量也因此呈现由西向东、由南向北递增趋势，整体肥力水平高于分区 4 低于分区 1。研究区中部偏南区域土壤对于分区 1 的隶属度高（图 5-a），达到 0.8 以上，该区域分布有灌耕灰漠土亚类的黄土状、红土质和砂砾质灌耕灰漠土土属，保水保肥性状较其它分区强，垦殖年限长，熟化程度高，耕层土壤养分含量较高。图 6 为依据最大隶属度原则绘制的绿洲管理分区图，对于土壤养分单一模糊隶属度空间分布图具有很好的补充说明作用。

混乱度指数 (CI) 表征样点模糊隶属关系的不确定性复杂程度。 CI 值越大，不同模糊类别在样点土壤上的交叠程度越高，样点土壤隶属关系不确定性越大；反之，则表明交叠程度低，土壤隶属关系趋向明确，当 $CI=0$ 时，表明样点土壤在属性空间和地理空间上

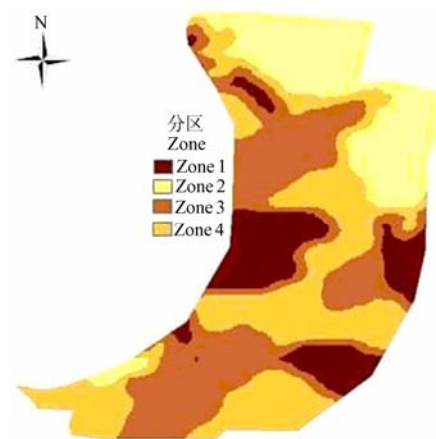


图 6 绿洲管理分区图

Fig. 6 Management zones maps of oasis field

都具有清晰边界。图 7 为研究区土壤混乱度指数图，可以看出研究区西北部、中部及西南部等新开垦地、荒地及水体等与主要耕地交界区域土壤隶属关系混乱指数明显较高，其中尤以中部和南部各一处最高， $CI > 0.8$ ，表明这些区域内土壤养分含量的隶属关系不确定性较大。研究区土壤隶属关系的 CI 平均值为 0.19。总体而言，土壤属性数据的模糊分类输出结果不确定性较小，地理空间上土壤的隶属关系相对明确。

2.4 管理分区的校验

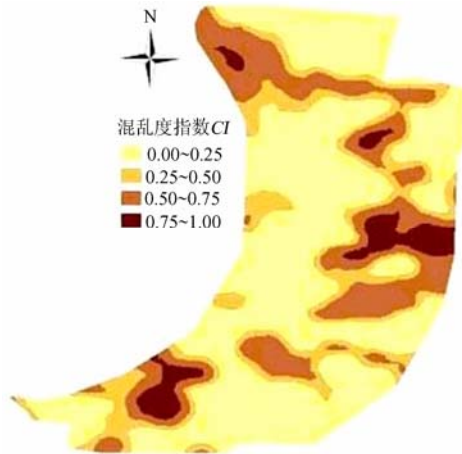


图 7 土壤隶属关系混乱度指数图

Fig. 7 Predictive maps showing spatial variation of confusion index

为了确定分区结果是否可以作为变量施肥的依据，对各管理分区的土壤养分数据进行了常规统计并用最小显著极差法 (LSR) 进行了各分区间差异显著性检验 (表 5)。与分区前相比，各管理分区土壤养分的变异系数都有所减小，有机质的变异系数由全研究区的 28.18% 减小到 18.94%~25.57% 之间，碱解氮的变异系数由 26.98% 减小到 18.81%~25.15% 之间，速效磷的变异系数由 46.03% 减小到 31.46%~42.28% 之间，总盐的变异系数由 18.32% 减小到 1.53%~13.55% 之间，表明各分区内的土壤属性变异程度均减小，含量分布趋于均一。分区间差异显著性检验表明，在 0.05 显著水平上，不同分区之间有机质和速效钾的均值均达到了显著差异；碱解氮在分区 1、3 之间，速效磷在分区 2、4 之间和分区 3、4 之间未达到显著，两种养分的其它分区间均达到显著差异。在 0.01 显著水平上，有机质在分区 3、4 之间，碱解氮在分区 1、3 之间未达到极显著，两种养分在其它分区间均达到极显著差异；速效磷在分区 2、3、4 之间均未达到极显著，但它们与分区 1 之间均达到了极显著差异；速效钾不同分区间的均值均达到了极显著差异。总体看来，分区后各管理分区内土壤属性含量空间分布差异较小，而管理分区间差异显著，因此可在同一管理分区内采用相同的施肥量，而在不同分区间实施变量施肥的管理措施。

表 5 分区土壤属性统计和 LSR 检验结果

Table 5 Zoning statistics for soil properties and LSR test

分区 Zone	样点数 No. of samples	有机质 OM		碱解氮 Alkali-hydrolyzable N		速效磷 Available P		速效钾 Available K	
		平均值 Mean ($\text{g}\cdot\text{kg}^{-1}$)	变异系数 CV (%)	平均值 Mean ($\text{mg}\cdot\text{kg}^{-1}$)	变异系数 CV (%)	平均值 Mean ($\text{mg}\cdot\text{kg}^{-1}$)	变异系数 CV (%)	平均值 Mean ($\text{mg}\cdot\text{kg}^{-1}$)	变异系数 CV (%)
Zone 1	27	13.17 a A	25.57	50.83 ac AC	21.28	15.27 a A	35.25	410.51 a A	7.68
Zone 2	27	7.99 d C	18.94	42.85 d D	20.74	9.92 c B	34.75	223.88 d D	13.55
Zone 3	68	10.82 b B	24.17	47.22 c C	25.15	12.51 b AB	42.28	344.80 b B	1.53
Zone 4	71	9.77 c B	25.18	48.62 b B	18.81	11.85 bc B	31.46	292.99 c C	6.12

均值后的不同小写字母和 大写字母分别表示 0.05、0.01 显著水平

Different small and capital letters after the means indicate significant difference at 0.05, and 0.01 level, respectively

3 讨论

基于模糊聚类法的农田精确管理分区研究，其分区数是由数据本身的特性决定的，不受人为因素影响，能够较好地解释自然界中发生的连续变异现象。本研究对各分区土壤养分数据进行常规统计并用 LSR 法进行各分区间差异显著性检验，表明管理分区可以作

为作业单元进行多元养分变量施肥管理，验证了分区结果的有效性与合理性，这与李艳等^[7,8]、李翔等^[20]的管理分区研究结果一致。

对于模糊 c -均值聚类法须事先确定的模糊类别数 c 、模糊加权指数 φ ，McBratney 和 Moore 提出 FPI 和相对于 φ 的派生函数 $-(\delta J/\delta \varphi)c^{0.5}$ 两个指标来计算。但这些以从聚类过程自身获得的信息为基础的指标并

不能总是有效地鉴别数据集的最适分类。本研究在以上两种指标所确定的模糊类别数不同的情况下, 采用基于外部变量的多元回归法来确定适宜的分区数及模糊指数, 此法能够较好地揭示外部变量的空间变异性与模糊数据集对应隶属度等级的空间分布之间的关系。该方法的关键是选择适宜的外部变量, 尽管 Milligan 明确地提出了选择标准, 但不同的聚类对象其选取标准应因地制宜地, 不能一概而论, 并且只有当外部变量与聚类数据对应隶属度等级显示较强的相关关系时, 所确定的模糊控制参数才能作为模糊聚类的前提。这些无疑会这给外部变量的实际选择带来一定难度, 增加了该法应用的复杂性, 给推广带来一定制约。

由于所选数据源的限制, 本研究只选取了一种外部变量——棉田产量数据进行基于外部变量的多元回归法的尝试, 在今后的研究中将拓宽外部变量的选择范围, 并针对基于不同外部变量的分区结果展开对比研究, 以期更全面地表达土壤肥力空间分布所具有的信息, 使分区结果更趋合理。对于各分区田间管措施的制定、实际经济效益等问题在研究中尚未涉及, 这些将在下一步的试验中进行论证说明。

4 结论

管理分区是实施精准农业变量投入的一个经济有效的手段。模糊 c -均值聚类法作为探测分析多源空间数据固有结构模式的一种有效方法, 目前正逐步受到国内管理分区研究者的重视。本研究采用模糊 c -均值聚类算法定义土壤养分精确管理分区, 并通过 FPI 、 $c-\phi$ 多次组合法和基于外部变量的多元回归法的比较论证确定了最佳模糊控制参数: 将研究区划分为 4 个养分管理分区, 模糊加权指数为 1.6。各管理分区内土壤属性的变异系数均较分区前全研究区有所减小, 且分区间土壤养分差异显著, 同时全研究区的平均混乱度指数较低, 不同模糊类别交叠程度小, 地理空间上土壤的隶属关系相对明确, 说明在研究区内, 模糊分类的输出结果符合管理分区的定义, 各管理分区均可以作为变量施肥的单独作业单元进行耕作管理, 证实了对北疆绿洲农业模式下的大型农田进行精确管理分区的可行性。分区结果可以有效地量化表达土壤属性的空间分布信息, 为研究区精准农业变量投入工作的实施提供理论基础, 为精确、高效的田间管理提供决策依据。

References

- [1] 张风华, 赖先齐. 西北干旱区内陆绿洲农业特征及发展认识. 干旱区资源与环境, 2003, 17(4): 19-23.
Zhang F H, Lai X Q. Oasis agriculture characters and development in arid area of northwestern China. *Journal of Arid Land Resources and Environment*, 2003, 17(4): 19-23. (in Chinese)
- [2] 赖先齐. 中国绿洲农业学. 北京: 中国农业出版社, 2005.
Lai X Q. *China Oasis Agronomy*. Beijing: China Agricultural Press, 2005. (in Chinese)
- [3] Schepers A R, Shanahan J F, Liebig M A, Schepers J S, Johnson S H, Luchiari A, Jr. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. *Agronomy Journal*, 2004, 96: 195-203.
- [4] 李翔, 潘瑜春, 赵春江, 王纪华, 鲍艳松, 刘良云, 王锦地. 基于空间连续性聚类算法的精准农业管理分区研究. 农业工程学报, 2005, 21(8): 78-82.
Li X, Pan Y C, Zhao C J, Wang J H, Bao Y S, Liu L Y, Wang J D. Delineating precision agriculture management zones based on spatial contiguous clustering algorithm. *Transactions of the CSAE*, 2005, 21(8): 78-82. (in Chinese)
- [5] Gorsevski P V, Gessler P E, Jankowski P. Integrating a fuzzy k -means classification and a Bayesian approach for spatial prediction of landslide hazard. *Journal of Geograph System*, 2003, 5: 223-251.
- [6] 石洪波, 于剑, 黄厚宽. 一种有效的 FCM 算法的实现方式. 铁道学报, 2003, 25(1): 63-67.
Shi H B, Yu J, Huang H K. An effective implementation method of Fuzzy c -Means algorithm. *Journal of the China Railway Society*, 2003, 25(1): 63-67. (in Chinese)
- [7] 李艳, 史舟, 吴次芳, 李锋, 程街亮. 基于模糊聚类分析的田间精确管理分区研究. 中国农业科学, 2007, 40(1): 114-122.
Li Y, Shi Z, Wu C F, Li F, Cheng J L. Definition of management zones based on fuzzy clustering analysis in coastal saline land. *Scientia Agricultura Sinica*, 2007, 40(1): 114-122. (in Chinese)
- [8] Li Y, Shi Z, Li F, Li H Y. Delineation of site-specific management zones using fuzzy clustering analysis in a coastal saline land. *Computers and Electronics in Agriculture*, 2007, 56: 174-186.
- [9] 檀满枝, 陈杰, 郑海龙, 张学雷. 模糊 c -均值聚类法在土壤重金属污染空间预测中的应用. 环境科学学报, 2006, 26(12): 2086-2092.
Tan M Z, Chen J, Zheng H L, Zhang X L. Spatial prediction of soil heavy metal pollution using fuzzy c -mean algorithm. *Acta Scientiae Circumstantiae*, 2006, 26(12): 2086-2092. (in Chinese)
- [10] 檀满枝, 陈杰, 徐方明, 郑海龙, 张学雷. 基于模糊集理论的土壤重金属污染空间预测. 土壤学报, 2006, 43(3): 389-396.

- Tan M Z, Chen J, Xu F M, Zheng H L, Zhang X L. Spatial prediction of soil heavy metal pollution based on fuzzy set theory. *Acta Pedologica Sinica*, 2006, 43(3): 389-396. (in Chinese)
- [11] Bragato G. Fuzzy continuous classification and spatial interpolation in conventional soil survey for soil mapping of the lower Piave plain. *Geoderma*, 2004, 118: 1-16.
- [12] de Bruin S, Stein A. Soil-landscape modeling using fuzzy *c*-means clustering of attribute data derived from a digital elevation model (DEM). *Geoderma*, 1998, 83: 17-33.
- [13] 鲍士旦. 土壤农化分析. 北京: 中国农业出版社, 2000.
- Bao S D. *Soil and Agricultural Chemistry Analysis*. Beijing: China Agricultural Press, 2000. (in Chinese)
- [14] 高新波, 裴继红, 谢维信. 模糊 *c*-均值聚类算法中加权指数 *m* 的研究. 电子学报, 2000, 28(4): 80-83.
- Gao X B, Pei J H, Xie W X. A study of weighting exponent *m* in a fuzzy *c*-means algorithm. *Acta Electronica Sinica*, 2000, 28(4): 80-83. (in Chinese)
- [15] McBratney A B, Moore A W. Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, 1985, 35: 165-185.
- [16] Burrough P A, van Gaans P F M, Hootsmans R. Continuous classification in soil survey: Spatial correlation, confusion and boundaries. *Geoderma*, 1977, 77: 115-135.
- [17] 新疆生产建设兵团土壤普查办公室. 新疆生产建设兵团垦区土壤. 乌鲁木齐: 新疆科技卫生出版社, 1993.
- Soil Survey Office of Xinjiang Production and Construction Corps. *Reclamation Soil of Xinjiang Production and Construction Group*. Urumqi: Xinjiang Science, Technology and Health Press, 1993. (in Chinese)
- [18] 王绍强, 朱松丽, 周成虎. 中国土壤土层厚度的空间变异特征. 地理研究, 2001, 20(2): 161-169.
- Wang S Q, Zhu S L, Zhou C H. Characteristics of spatial variability of soil thickness in China. *Geographical Research*, 2001, 20(2): 161-169. (in Chinese)
- [19] Chang Y H, Scrimshaw M D, Emmerson R H C, Lester J N. Geostatistical analysis of sampling uncertainty at the tollesbury managed retreat site in Blackwater Estuary, Essex, UK: Kriging and cokriging approach to minimize sampling density. *The Science of the Total Environment*, 1998, 221: 43-57.
- [20] 李 翔, 潘瑜春, 马景宇, 赵春江, 王纪华. 基于多种土壤养分的精准管理分区方法研究. 土壤学报, 2007, 44(1): 14-20.
- Li X, Pan Y C, Ma J Y, Zhao C J, Wang J H. Soil nutrients-based zoning for management of precision agriculture. *Acta Pedologica Sinica*, 2007, 44(1): 14-20. (in Chinese)

(责任编辑 李云霞)