

基于模糊 c -均值聚类法的绿洲农田精确管理分区研究

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 832003)

摘要:以新疆绿洲棉区为研究对象, 将 176 个土壤耕层属性数据作为变量进行精确管理分区研究。模糊 c -均值聚类法被用来进行分区, 采用 $c-\varphi$ 多次组合法来确定适宜的模糊控制参数。研究区最佳分区数为 4, 模糊指数为 1.5, 平均混乱度指数为 0.17, 不同模糊类别重叠程度小, 地理空间上土壤的隶属关系相对明确。为了评价分区结果的合理性, 对各管理分区土壤数据进行常规性统计, 并用 LSR 法进行各分区间差异显著性检验, 结果表明: 各管理分区土壤属性的变异系数都较分区前全研究区有所减小, 分区间土壤属性差异显著。通过选取适宜的模糊控制参数, 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; 土壤属性; 绿洲棉田

文章编号: 1000-0933(2008)07-3067-08 中图分类号: Q143 文献标识码: A

Definition of management zones based on fuzzy c -mean algorithm in oasis farmland

CHEN Yan, LÜ Xin*

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

Acta Ecologica Sinica, 2008, 28(7): 3067 ~ 3074.

Abstract: Percent organic matter, total nitrogen, available phosphorus, and salinity were estimated in 176 topsoil samples. These samples were collected from a 9600 hm² cotton field in northern Xinjiang Province. A fuzzy c -means clustering algorithm was used to assign these samples to management zones. The derivative of the objective function with respect to the fuzziness exponent, $(\delta J / \delta \varphi) c^{0.5}$, was used to determine the optimum fuzzy control parameters. The optimum number of the classes and the fuzziness exponent was 4 and 1.5, respectively. The average confusion index was 0.17 in all management zones. Thus, the overlapping of fuzzy classes at points was low and the spatial distribution of membership grades was unambiguous. To estimate the validity of zoning result, the general statistics analysis on the data was carried out. The zoning statistics showed that variation coefficients of soil properties decreased, while the means of the soil properties differed sharply between management zones. These results indicated that fuzzy c -means clustering algorithm can be used to delineate management zones by the optimum fuzzy control parameters. The management zones can then be used to help guide the rate of fertilizer application in an effort to manage soil nutrient levels more efficiently.

Key Words: management zones; fuzzy clustering; soil property; oasis cotton field

基金项目: “十一五”国家科技支撑计划资助项目(2006BAD21B02-2)

收稿日期: 2007-09-19; 修订日期: 2008-03-13

作者简介: 陈彦(1979~), 男, 河北安新人, 博士生, 主要从事绿洲生态和地理信息系统研究. E-mail: yanc1208@163.com

* 通讯作者 Corresponding author. E-mail: lxshz@126.com

Foundation item: The project was financially supported by the National “11 Five-Year-Plan” Science and Technology Program of China (No. 2006BAD21B02-2)

Received date: 2007-09-19; **Accepted date:** 2008-03-13

Biography: CHEN Yan, Ph. D. candidate, mainly engaged in application research of oasis ecosystem and GIS. E-mail: yanc1208@163.com

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。同我国东部河套平原绿洲区及柴达木高原绿洲区相比,西北干旱区内陆绿洲在地貌特征、气候特点及农业发展等方面均存在显著的地域差异性,同时内陆绿洲土壤具有非地带性分布的特征,土壤类型随地表水流向呈现自上而下的规律性变化^[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与其特殊的生产功能,对绿洲农业的合理开发及可持续发展也有着重要作用。

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)进行管理正成为目前精准农业的一个研究热点^[3],而模糊 c -均值聚类法(FCM)在管理分区的研究中应用最为广泛^[4]。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,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^[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。李翔等^[9]则引用模糊性能指数(FPI)和标准化分类熵(MPE)确定了适宜的管理分区数。檀满枝等^[10,11]在土壤重金属的分类研究中则采用 c - φ 多次组合最优取值法同时确定了适宜模糊类别数和模糊加权指数。将此法用于基于模糊聚类的农田管理分区研究中,能弥补上述模糊聚类法的固有不足,使分类结果更趋于合理。本研究以内陆绿洲农业模式下大型农田为研究对象,以多种土壤属性为数据源,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,采用 c - φ 多次组合最优取值法确定适宜的模糊控制参数,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七师 125 团垦区是典型的新疆绿洲农耕区,位于 $84^{\circ}20' \sim 84^{\circ}40'E$, $44^{\circ}36' \sim 44^{\circ}54'N$ 之间,地处天山北麓,准噶尔盆地西南缘,横跨奎屯河流域和古尔图河、四棵树河流域交汇的冲积扇、冲积-洪积平原。垦区自然植被稀少,土体干燥,发育为地带性灰棕漠土,农场光、热、水资源丰富,全年平均气温 $6.5^{\circ}C$, $\geq 10^{\circ}C$ 活动积温 $3622^{\circ}C$; 年均日照时数 2620h,日照百分率 63%,海拔 260 ~ 300m,无霜期 164d。现有耕地面积 1.17 万 hm^2 ,其中滴灌棉田面积 9600 hm^2 ,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

2005 年 10 月在 125 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、4.5m 为半径在 0 ~ 30cm 的主要耕作层取 5 个土壤样品的混合土样,共采集土壤样品 176 份。并实测了以每个采样点为中心,长 5.86m、宽 2.25m 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,采样点空间分布见图 1。获取的土壤样品经自然风干、过筛后进行室内含量测定。每个样品测定项目包括有机质、全氮、碱解氮、速效磷、速效钾和总盐,测定方法按《土壤农业分析》^[12]。

1.3 数据处理方法

为了得到影响棉花产量的土壤属性主要限制因子,对土壤属性数据与棉花产量做了相关性分析;为避免计算结果受变量量纲和数量级不同的影响,在进行运算和分析前,采用标准差标准变换法对原始数据进行标准化处理,均应用 SPSS12.0 完成。为分区的需要,分析了主要限制因子的空间变异特征,半方差函数的计算、理论模型的拟合及克里格插值和图形绘制是由 ArcGIS8.3 完成的,并采用克里格插值法对采样点插值生成土壤属性空间分布图。本研究中模糊 c -均值聚类及适宜分区数和模糊加权指数的计算都是通过 MATLAB6.0 编程实现的。

1.4 管理分区方法

1.4.1 模糊 c -均值聚类法

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:

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

$$i = 1, \dots, n; \quad j = 1, \dots, c \quad (1)$$

公式(1)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, 即模糊加权指数 (Fuzziness Exponent)^[13], 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。

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。国内李翔等和李艳等引用模糊效果指数 (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, 但 FPI 无法确定适宜的模糊加权指数^[14]。本研究采用 McBratney 和 Moore 提出的 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^[10, 15, 16]。

根据模糊分类结果, 并遵循最大隶属度原则确定样点所归属的模糊类别, 绘制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^[17]:

$$CI = 1 - [\mu_{\max i} - \mu_{(\max - 1)i}] \quad (2)$$

公式(2)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, $\mu_{(\min - 1)i}$ 为第 i 个样点土壤第 2 大隶属度值。

2 结果与分析

2.1 土壤数据与产量数据的相关性分析

表 1 为研究区土壤属性及棉花产量数据的描述性统计结果。从均值来看, 按照新疆土壤普查资料^[18], 研究区有机质、全氮、碱解氮含量属于中等水平, 速效磷、速效钾含量较丰富, 总盐含量较低。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, $CV > 100\%$ ^[19], 研究区土壤属性在空间上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。

为了得到影响棉花产量的土壤属性主要限制因子, 将所选取的土壤属性数据与棉花单产数据进行相关性分析。表 2 是样区棉花产量与土壤属性的相关关系矩阵, 有机质、全氮、碱解氮、速效磷与棉花产量呈显著的

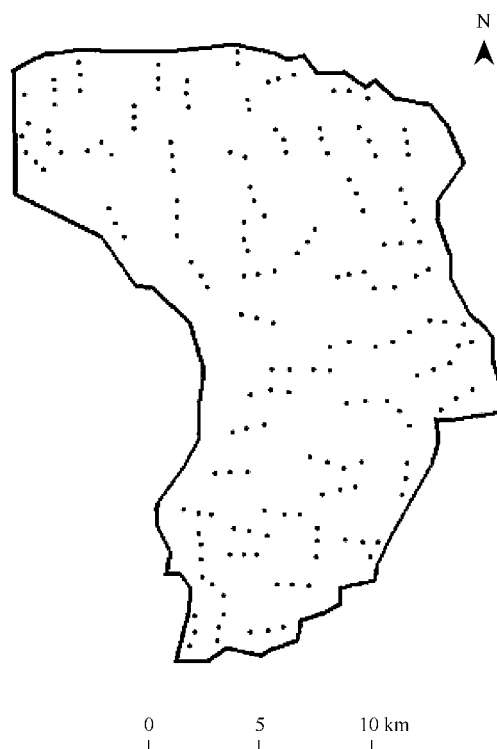


图 1 采样点分布图

Fig. 1 Distribution map of sample points

正相关关系;总盐与棉花产量呈显著负相关关系,相关系数为 -0.323 ;而速效钾与棉花产量关系不明显。虽然全氮与碱解氮相关系数均达到了极显著,但根据耕地分级的相关性原则^[20],弃去碱解氮保留全氮的数据作为评价指标,同时弃去对棉花产量影响较不显著的速效钾。

表 1 土壤属性和棉花产量传统统计分析

Table 1 Descriptive statistics of soil properties and crop yield

土壤属性 Soil properties	最小值 Min	最大值 Max	平均值 Mean	中值 Median	标准差 SD	偏度 Skewness	峰度 Kurtosis	变异系数 CV(%)
有机质 OM(g/kg)	4.33	38.30	14.99	14.42	6.05	0.84	3.88	40.37
全氮 TN(g/kg)	0.29	1.51	0.79	0.77	0.26	0.56	3.08	32.91
碱解氮 AN(mg/kg)	20.50	102.17	45.05	40.25	17.20	1.51	3.99	38.18
速效磷 AP(mg/kg)	9.33	60.67	28.32	24.67	14.59	0.46	3.13	51.52
速效钾 AK(mg/kg)	140.00	431.00	325.91	325.50	56.84	-0.70	3.58	17.44
总盐 Total salt(%)	0.02	1.36	0.25	0.20	0.15	1.31	3.48	60.00
产量 Yield(kg/hm ²)	1533.70	2808.10	2050.90	2068.10	224.42	-0.14	2.40	10.94

表 2 研究区土壤属性和作物产量的相关系数矩阵

Table 2 Correlation matrix for selected soil properties and yield in the study area

土壤属性 Soil properties	产量 Yield (kg/hm ²)	OM (g/kg)	TN (g/kg)	AN (mg/kg)	AP (mg/kg)	AK (mg/kg)	Total salt (%)
产量 Yield(kg/hm ²)	1						
有机质 OM(g/kg)	0.697 **	1					
全氮 TN(g/kg)	0.657 **	0.660 **	1				
碱解氮 AN(mg/kg)	0.523 **	0.356 **	0.301 *	1			
速效磷 AP(mg/kg)	0.317 *	0.500 **	0.319 *	0.182	1		
速效钾 AK(mg/kg)	0.120	0.275 *	0.057	0.150	0.185	1	
总盐 Total salt (%)	-0.323 *	-0.242	-0.473 **	-0.239	-0.155	-0.089	1

* 表示在 5% 的概率水平上有显著相关性; ** 表示在 1% 的概率水平上有极显著相关性 * Indicates correlation is significant at 5% probability level; ** indicates correlation is extremely significant at 1% probability level

2.2 土壤属性空间变异性分析

科学的土壤资源利用和管理是建立在对区域土壤属性空间变异性正确认识的基础之上,因此有必要了解 4 种影响棉花产量的土壤属性主要限制因子的空间变异状况。斜度和峰度反映样本的正态分布特点,标准正态分布斜度为 0,峰度为 3。由表 1 可知研究区土壤有机质、全氮、速效磷原始数据基本符合正态分布;总盐数据不服从正态分布,在进行对数转换后数据符合正态分布。

在考虑各向异性的情况下,拟合 4 种土壤属性的半方差模型并进行了交叉验证,模型及相关参数如表 3 所示。有机质和总盐符合球状模型,全氮、速效磷可用指数模型拟合。4 种土壤属性的块金值与基台值的比值在 25% ~ 75% 之间变化,表现为中等的空间自相关性^[13],说明其变异是由结构性因素和随机性因素共同作用的结果。在建立了半方差模型的基础上,采用克里格插值法对土壤属性数据进行插值并绘制土壤属性的空间分布图(图 2)。

2.3 基于模糊聚类分析的管理分区生成及校验

将 4 种土壤属性的空间分布图转换为 $0.1\text{km} \times 0.1\text{km}$ 分辨率的栅格图形,经研究区域边界图形切割后,得到 1175 个有效栅格值的图形。然后对栅格图进行转点运算,提取每个栅格重心点及其坐标,将栅格属性数据赋予重心点作为模糊聚类的数据源。把 $n \times p$ ($n = 1175, p = 4$) 的土壤属性模糊集带入模糊 c -均值聚类法中,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 ~ 6, φ 取 1.1 ~ 1.8 时, $-\left[(\delta J/\delta \varphi)c^{0.5}\right]$ 的结果,不同 c - φ 组合下 $-\left[(\delta J/\delta \varphi)c^{0.5}\right]$ 的取值结果见图 3。可以看出,当 $c = 4$ 时, $-\left[(\delta J/\delta \varphi)c^{0.5}\right]$ 曲线的峰值最小,即研究区最适宜分区数为 4;当 $c = 4, \varphi = 1.5$ 时, $-\left[(\delta J/\delta \varphi)c^{0.5}\right]$

$\delta\varphi)c^{0.5}]$ 出现峰值。因此研究区模糊聚类的最终控制参数应为: $c=4$ 、 $\varphi=1.5$ 。利用 MATLAB 6.0 的 FCM 函数,取分区数 $c=4$,模糊加权指数 $\varphi=1.5$,最大迭代次数为 100,允许误差 $E=10^{-5}$,对模糊集进行模糊聚类。经过 53 次迭代, $J(F,C)$ 达到最小值 1550.70,得到此时各样本的隶属度函数矩阵,依据最大隶属度原则,取每个栅格数据最大隶属度所在分区作为各样本分类依据,得到管理分区图(图 4a)。

表 3 土壤属性的半方差模型参数
Table 3 Semivariogram parameters of soil properties

土壤属性 Soil properties	模型 Model	块金值 C_0 Nugget	基台值 $C + C_0$ Sill	$C_0 / (C + C_0)$ (%)	主轴变程 (km) Major range	亚轴变程 (km) Minor range	主方向 Direction
有机质 OM	球状 Spherical model	4.97	17.73	27.02	5.21	1.96	NW45°
全氮 TN	指数 Exponential model	0.014	0.038	36.84	3.27	1.55	NW45°
速效磷 AP	指数 Exponential model	132.43	247.53	53.50	13.07	4.68	NE9°
总盐 Total salt	球状 Spherical model	0.51	0.93	54.84	3.18	0.50	NW18°

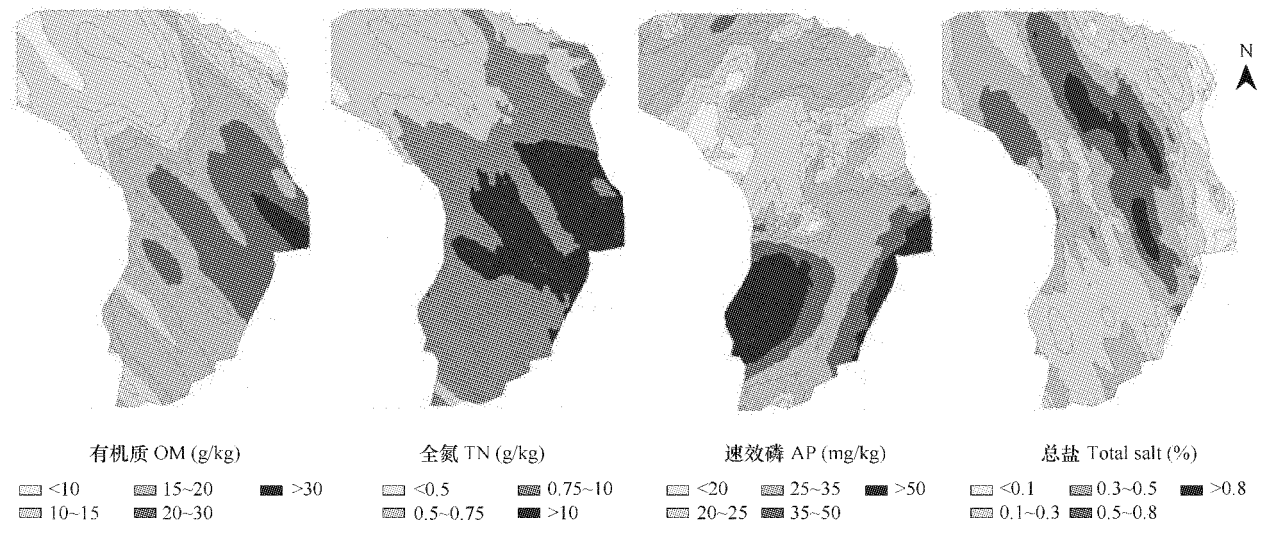


图 2 土壤属性插值分布图
Fig. 2 Interpolated distribution map of soil properties

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。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。图 4b 为混乱度指数图,可以看出在各管理分区交界处,即研究区中部和南部的零星小区域内土壤隶属关系混乱指数相对较高,尤以南部一隅最高, $CI>0.8$,表明这些区域内土壤属性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 0.17,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。

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属性数据进行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 4)。与分区前相比,各管理分区土壤属性的变异系数均有所减小,有机质的变异系数由全研究区的 40.37% 减小到 14.88%~24.28% 之间,全氮的变异系数由全研究区的 32.19% 减小到 10.21%~14.39% 之间,速效磷的变异系数由全研究区的 51.52% 减小到 13.88%~18.65% 之间,总盐的变异系数由全研究区的 60% 减小到 36.93%~50.32% 之间,表明各分区内的土壤属性变异程度均有减小,含量分布趋于均一。

分区间差异显著性检验表明,不同分区间土壤有机质、全氮的均值均达到极显著差异;速效磷除分区 2 与

分区 4 间未达到极显著,其他分区之间均达到既显著差异;对总盐而言,分区 1、2 之间和分区 2、4 之间未达到极显著,但各分区间均达到显著差异。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,适宜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管理。

2.4 分区的评价与管理

表 5 是土壤属性类别聚类中心矩阵,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^[14]。根据表 5 中土壤属性类别中心值数据及图 4a,对样品土壤分属的类别即管理分区进行定义:管理分区间土壤属性含量变化平稳,有机质、全氮等级总体呈逐区递增趋势。

分区 1 位于研究区西北部,该区分布有大量新开垦地,耕作年限短,耕作层不明显,保水保肥性差,是本地

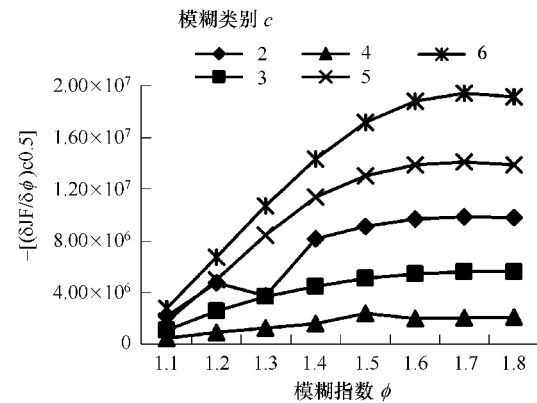


图 3 模糊类别 c 不同取值时模糊指数 ϕ 与 $-[(\delta J_F / \delta \phi) c^{0.5}]$ 的关系

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

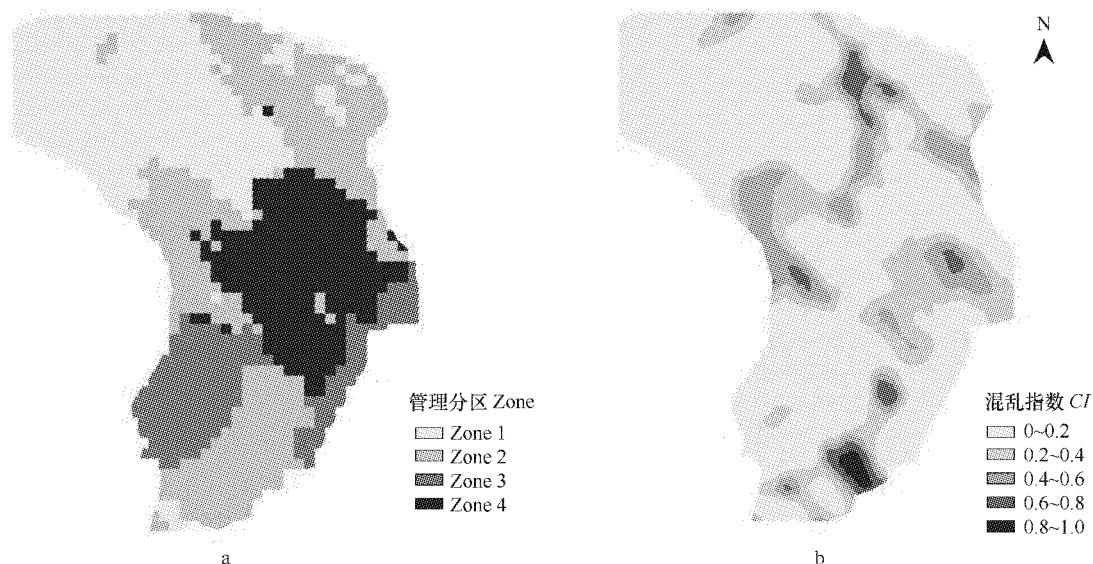


图 4 管理分区图(a)及混乱指数空间预测分布图(b)

Fig. 4 Management zones map and predictive map of confusion index

土壤肥力最低的土壤。多施氮肥、磷肥和有机肥应是改良土壤的主要措施。分区 2 分布于研究区中部及东北部零星区域,此区域为撂荒地分布区,主要为撂荒盐土土属,土壤质地粗、保水能力低,潜在养分含量低,是本地低产土壤之一。多施有机肥,客土掺粘,勤浇水、保墒情,是利用改良这类土壤的重点。分区 3 位于研究区南部,土壤的水、肥、气、热诸因素较协调,保水供肥性能也较好,今后应增施有机肥,协调土壤中各种营养元素的比例,以进一步提高产量。分区 4 位于研究区中部区域,该区土壤质地适中,土体水、肥、气、热诸因素协调,保水供肥性能好,适耕期长,产量高。由于长期耕种,熟化程度高,耕层土壤养分含量较高,具有最高的土壤肥力水平,多属高产稳产农田。

全研究区气候、地貌、水文、地形等自然地理条件基本一致,但土壤属性存在中等强度的变异,空间分布上表现为中等的空间自相关性,应以统一的管理分区作为变量施肥单元制定来施肥方案。由于研究区以滴灌为主要灌溉方式,采用基肥与随水滴施相结合的施肥措施,同时团场滴灌区的划分是以行政边界为依据,若完全

按管理分区的分布改造,成本及难度均较大,因此在制定分区变量施肥方案时也应考虑到滴灌区布局的限制因素,在对滴灌设施小幅度改造的前提下,尽量在同一管理分区采用同一标准的滴灌措施。对于在一种管理分区中出现的零星异质区域,出于成本及大田统一管理的考虑,可以考虑采用相同措施进行管理。

表 4 分区土壤属性统计和 LSR 检验结果

Table 4 Zoning statistics for soil properties and LSR test

分区 Zone	样点数 No. of samples	有机质 OM		全氮 TN		速效磷 AP		总盐 Total salt	
		平均值 Mean (g/kg)	变异系数 CV (%)	平均值 Mean (g/kg)	变异系数 CV (%)	平均值 Mean (g/kg)	变异系数 CV (%)	平均值 Mean (g/kg)	变异系数 CV (%)
Zone 1	417	10.89dD	18.58	0.65dD	11.59	25.83dC	13.88	0.47aA	36.93
Zone 2	355	15.51cC	19.42	0.82cC	10.21	27.58cB	18.65	0.45bA	38.49
Zone 3	235	19.01bB	24.28	0.95bB	14.39	51.49aA	15.6	0.23dB	41.58
Zone 4	168	22.51aA	14.88	1.04aA	10.44	28.71bB	16.14	0.25cB	50.32

* 均值后的小写字母和大写字母分别表示 5%、1% 显著水平 * Different small and capital letters after the means indicate significant difference at 5%, and 1% level, respectively

表 5 土壤属性聚类中心矩阵

Table 5 Class centroids for the OM, TN, AP and the total salt data set

分区 Zone	有机质 OM	全氮 TN	速效磷 AP	总盐 Total salt
Zone 1	-1.00	-1.06	-0.49	0.48
Zone 2	-0.02	0.01	-0.34	0.41
Zone 3	0.47	0.57	1.96	-0.60
Zone 4	1.25	1.19	-0.18	-0.45

3 结论

管理分区划分的依据是限制作物产量的多种因子的空间变异性。模糊聚类法结合地统计学手段可以有效降低土壤属性空间预测的不确定性、丰富样点实测数据所蕴涵的空间信息,实现样点土壤定量隶属关系在地理空间上的可视化表达,直观揭示土壤属性在整个样区的空间分布与变异特征,进而对研究区管理分区进行评价。本研究依据影响棉花产量的主要土壤属性限制因子的空间变异性,采用模糊 c -均值聚类算法定义农田精确管理分区,并以 $c-\varphi$ 多次组合最优取值法确定了适宜的模糊控制参数,试对北疆绿洲棉田管理分区的划分进行初步研究。全研究区可划分为 4 个管理分区,模糊加权指数为 1.5。结果表明:各管理分区土壤属性的变异系数均较分区前全研究区有所减小,分区间各土壤属性差异显著,同时全研究区的混乱度指数较低,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。基于多元土壤数据的精确管理分区在新疆绿洲农耕区的推广具有实际的意义,可避免大量重复性操作,提高农业管理效率,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。

本研究从确定适宜模糊控制参数入手,目的在于探讨定义北疆绿洲农田精确管理分区的可行性,对于管理分区变量施肥模型的建立、分区间具体田管措施的制定等问题尚未涉及,这些将在今后的实验中进行论证说明。另外,土壤属性的变化是一个长期渐变的过程,是时间和空间变异的综合反映,土地利用方式及其变化对土壤属性含量的影响也是一个十分复杂的过程,综合考虑多种变量和作物产量的时空变异信息,把空间单元的空间关联性和属性数据的演变特征相结合,对于管理分区的研究发展是十分必要的,这也是下一步研究的重点。

References:

- [1] 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.
- [2] Lai X Q. China oasis agronomy. Beijing: China Agricultural Press, 2005.
- [3] Schepers A R, Shanahan J F, Liebig M K, Schepers J S, Johnson S H, Jr A L. Appropriateness of management zones for characterizing spatial variability of soil properties and irrigated corn yields across years. *Agronomy Journal*, 2004, 96: 195—203.
- [4] Gorsevski P M, Gessler P E, Jankowski P. 2003. Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. *Journal of Geographical Systems*, 5: 223—251.
- [5] Diker K, Buchleiter G W, Farahani H J, Heermann D F, Brodahl M K. Frequency analysis of yield for delineating management zones. *Proceedings of the Sixth International Conference on Precision Agriculture and other resources management*, 2002. 817—827.
- [6] Shi H B, Yu J, Huang H K. An effective implementation method of Fuzzy *c*-Means algorithm. *Journal of China Railway Society*, 2003, 25(1): 63—67.
- [7] 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.
- [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] 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.
- [10] 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.
- [11] 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.
- [12] Bao S D. Soil and Agricultural Chemistry Analysis. Beijing: China Agricultural Press, 2000.
- [13] 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.
- [14] de Bruin S, Stein A. Soil-landscape modelling using fuzzy *c*-means clustering of attribute data derived from a Digital Elevation Model (DEM). *Geoderma*, 1998, 83: 17—33.
- [15] McBratney A B, Moore A W. Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, 1985, 35, 165—185.
- [16] 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.
- [17] Burrough P A, van Ganss P F M, Hootsmans R. Continuous classification in soil survey. Spatial correlation, confusion and boundaries. *Geoderma*, 1997, 77: 115—135.
- [18] Xinjiang Production and Construction Corps Soil Survey Office. Reclamation soil of Xinjiang Production and Construction Corps. Urumqi: Xinjiang Science and Technology Health Press, 1993.
- [19] 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.
- [20] Zhu D J. Land Evaluation. Beijing: China Agriculture Sciencetech Press, 2002.

参考文献:

- [1] 张凤华, 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.
- [2] 赖先齐. 中国绿洲农业学. 北京: 中国农业出版社, 2005.
- [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.
- [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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 鲍士旦. 土壤农化分析. 北京: 中国农业出版社, 2000.
- [13] 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.
- [18] 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.
- [19] 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.
- [20] 朱德举. 土地评价. 北京: 中国大地出版社, 2002.

$\delta\varphi)c^{0.5}]$ 出现峰值。因此研究区模糊聚类的最终控制参数应为: $c=4$ 、 $\varphi=1.5$ 。利用 MATLAB 6.0 的 FCM 函数,取分区数 $c=4$,模糊加权指数 $\varphi=1.5$,最大迭代次数为 100,允许误差 $E=10^{-5}$,对模糊集进行模糊聚类。经过 53 次迭代, $J(F,C)$ 达到最小值 1550.70,得到此时各样本的隶属度函数矩阵,依据最大隶属度原则,取每个栅格数据最大隶属度所在分区作为各样本分类依据,得到管理分区图(图 4a)。

表 3 土壤属性的半方差模型参数

Table 3 Semivariogram parameters of soil properties

土壤属性 Soil properties	模型 Model	块金值 C_0 Nugget	基台值 $C+C_0$ Sill	$C_0/(C+C_0)$ (%)	主轴变程(km) Major range	亚轴变程(km) Minor range	主方向 Direction
有机质 OM	球状 Spherical model	4.97	17.73	27.02	5.21	1.96	NW45°
全氮 TN	指数 Exponential model	0.014	0.038	36.84	3.27	1.55	NW45°
速效磷 AP	指数 Exponential model	132.43	247.53	53.50	13.07	4.68	NE9°
总盐 Total salt	球状 Spherical model	0.51	0.93	54.84	3.18	0.50	NW18°

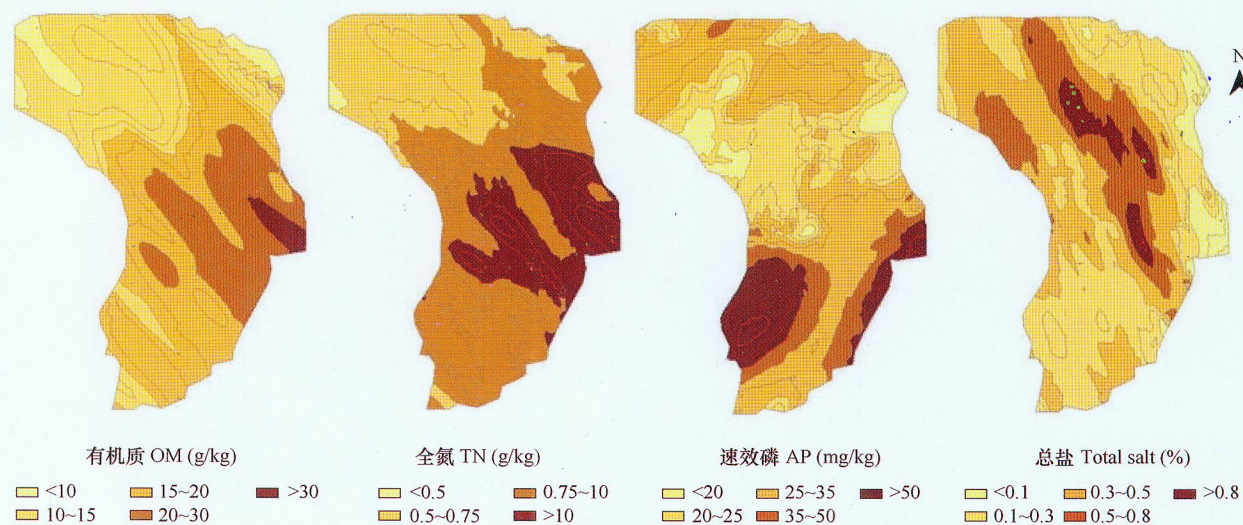


图 2 土壤属性插值分布图

Fig. 2 Interpolated distribution map of soil properties

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。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。图 4b 为混乱度指数图,可以看出在各管理分区交界处,即研究区中部和南部的零星小区域内土壤隶属关系混乱指数相对较高,尤以南部一隅最高, $CI>0.8$,表明这些区域内土壤属性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 0.17,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。

为了确定分区结果是否可以作为变量施肥的依据,对 4 个管理分区的土壤属性数据进行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 4)。与分区前相比,各管理分区土壤属性的变异系数均有所减小,有机质的变异系数由全研究区的 40.37% 减小到 14.88% ~ 24.28% 之间,全氮的变异系数由全研究区的 32.19% 减小到 10.21% ~ 14.39% 之间,速效磷的变异系数由全研究区的 51.52% 减小到 13.88% ~ 18.65% 之间,总盐的变异系数由全研究区的 60% 减小到 36.93% ~ 50.32% 之间,表明各分区内的土壤属性变异程度均有减小,含量分布趋于均一。

分区间差异显著性检验表明,不同分区间土壤有机质、全氮的均值均达到极显著差异;速效磷除分区 2 与

分区4间未达到极显著,其他分区之间均达到显著差异;对总盐而言,分区1、2之间和分区2、4之间未达到极显著,但各分区间均达到显著差异。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,适宜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管理。

2.4 分区的评价与管理

表5是土壤属性类别聚类中心矩阵,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^[14]。根据表5中土壤属性类别中心值数据及图4a,对样品土壤分属的类别即管理分区进行定义:管理分区间土壤属性含量变化平稳,有机质、全氮等级总体呈逐区递增趋势。

分区1位于研究区西北部,该区分布有大量新开垦地,耕作年限短,耕作层不明显,保水保肥性差,是本地

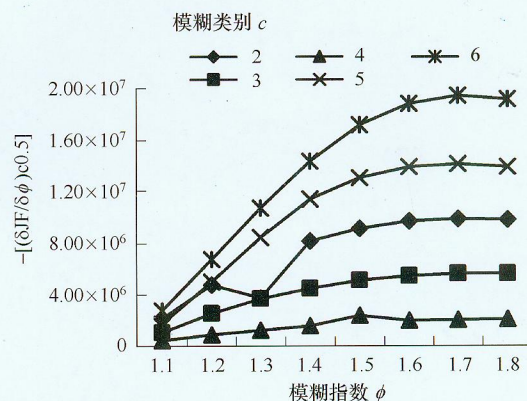


图3 模糊类别 c 不同取值时模糊指数 ϕ 与 $-[(\delta J_F / \delta \phi) c^{0.5}]$ 的关系

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

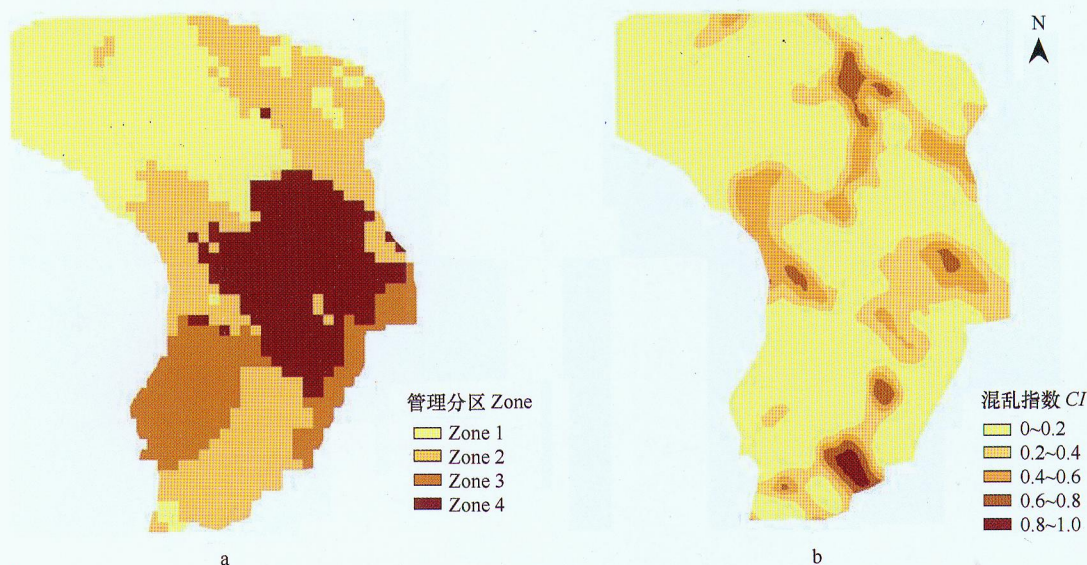


图4 管理分区图(a)及混乱指数空间预测分布图(b)

Fig. 4 Management zones map and predictive map of confusion index

土壤肥力最低的土壤。多施氮肥、磷肥和有机肥应是改良土壤的主要措施。分区2分布于研究区中部及东北部零星区域,此区域为撂荒地分布区,主要为撂荒盐土土属,土壤质地粗、保水能力低,潜在养分含量低,是本地低产土壤之一。多施有机肥,客土换粘,勤浇水、保墒情,是利用改良这类土壤的重点。分区3位于研究区南部,土壤的水、肥、气、热诸因素较协调,保水供肥性能也较好,今后应增施有机肥,协调土壤中各种营养元素的比例,以进一步提高产量。分区4位于研究区中部区域,该区土壤质地适中,土体水、肥、气、热诸因素协调,保水供肥性能好,适耕期长,产量高。由于长期耕种,熟化程度高,耕层土壤养分含量较高,具有最高的土壤肥力水平,多属高产稳产农田。

全研究区气候、地貌、水文、地形等自然地理条件基本一致,但土壤属性存在中等强度的变异,空间分布上表现为中等的空间自相关性,应以统一的管理分区作为变量施肥单元制定来施肥方案。由于研究区以滴灌为主要灌溉方式,采用基肥与随水滴施相结合的施肥措施,同时团场滴灌区的划分是以行政边界为依据,若完全