

# 基于空间连续性聚类算法的精准农业管理分区研究

李翔<sup>1,2</sup>, 潘瑜春<sup>2</sup>, 赵春江<sup>2</sup>, 王纪华<sup>2</sup>, 鲍艳松<sup>1</sup>, 刘良云<sup>2</sup>, 王锦地<sup>1,3</sup>

(1. 北京师范大学地理学与遥感科学学院遥感与地理信息系统研究中心, 北京 100875;

2. 国家农业信息化工程技术研究中心, 北京 100089; 3. 遥感科学国家重点实验室, 北京 100875)

**摘要:** 该研究在 K 均值算法 KM 的基础上, 根据空间单元位置的相互依赖关系, 提出了一种新的空间连续性聚类算法 SCKM。以北京精准农业示范基地获取的 OMIS 图像为数据源, 选用 K 均值算法、等间隔法、分位数法、自然断点法等传统分区方法和 SCKM 算法, 对肥水需求关键时期的小麦的长势差异进行了管理分区提取研究, 并引入了权重方差和聚集度两种分区效果评价指标, 对分区结果进行了比较和评价。结果表明: SCKM 算法与传统分区方法分区结果相比, 区内方差差异不显著, 而空间聚集度远好于后者, 利用 SCKM 法分区能够有效地去除大量的孤立单元或碎片。

**关键词:** 精准农业; 管理分区; SCKM 算法; 评价指标

**中图分类号:** S126; S11

**文献标识码:** A

**文章编号:** 1002-6819(2005)08-0078-05

李翔, 潘瑜春, 赵春江, 等. 基于空间连续性聚类算法的精准农业管理分区研究[J]. 农业工程学报, 2005, 21(8): 78-82

Li Xiang, Pan Yuchun, Zhao Chunjiang, et al. Delineating precision agriculture management zones based on spatial contiguous clustering algorithm [J]. Transactions of the CSAE, 2005, 21(8): 78-82 (in Chinese with English abstract)

## 0 引言

变量投入技术是实现农业精准管理的关键技术和手段, 初期的精准农业变量投入技术主要通过栅格采样方法实现的, 然而采样密度、采样方式以及插值方法的不同直接影响应用处方的精度和空间分布<sup>[1,2]</sup>。而且为了产生精确的处方图, 需要投入大量的人力、时间和成本获取高度密集的采样数据<sup>[3]</sup>。另一方面, 在实际田间变量实施中, 规则的格网处方并不实用。因此, 许多学者开始研究按照土壤或作物长势状况的差异性将同一地块分成不同的均质性区域进行管理, 即管理分区。管理分区技术是目前实施精准农业变量管理的一个经济有效的手段。科学合理的管理分区可以指导用户以管理分区为单元, 进行土壤和作物农学参数采样, 并根据不同单元间的空间变异性, 实施变量投入、精准管理决策。许多研究者采用不同的数据源(如高程、坡度、坡向等地形数据、土壤电导率、土壤耕层深度、产量数据等)进行了管理分区的划分研究<sup>[4-6]</sup>, 但目前的研究方法主要采用非监督分类或监督分类以及 GIS 软件提供的分类方法(如等间隔法、分位数法、自然断点法等)进行管理分区的勾画, 但由于在分区的过程中仅考虑了空间单元的属性数据, 并没有考虑单元的空间分布及空间相互依赖关系而使分区结果出现许多孤立的单元或碎片, 不利于精准农业机械设备的田间变量管理作业。为了解决此问题, 本研究在传统的 K 均值算法的基础上, 引入了空间

单元位置的相互依赖关系, 提出了一种新的空间连续性聚类算法 SCKM (Spatial Contiguous K Means cluster)。并以北京精准农业示范基地获取的高分辨率成像光谱图像 OMIS (Operative Modular Imaging Spectrometer) 为数据源, 选用 4 种传统的分区方法(K 均值、等间隔、分位数、自然断点法)和 SCKM 算法, 对肥水需求关键时期的小麦的长势差异进行了调优栽培管理分区提取研究, 并对传统方法与 SCKM 算法的分区效果进行了比较和评价。

## 1 传统的分区方法

### 1.1 GIS 软件提供的几种常用方法

ArcGIS 软件中提供的几种常见分类方法:

等间隔法(Equal Interval, EI): 根据空间单元的属性数据, 按等间隔距离将空间单元划分不同的类别。

分位数法(Quantile, QT): 按照每个类别具有相同的空间单元来划分的方法。

自然断点法(Natural Breaks, NB): 用 Jenk 优化公式确定不同类别间的断点, 该方法相当复杂, 但其核心思想是使类别内方差和最小。自然断点法能反映空间单元分布的固有模式或类别。

### 1.2 K 均值聚类算法

K 均值聚类算法(K-Means, KM)是一种经典的空间聚类算法。首先, 要给定聚类数目 K 创建一个初始划分, 然后根据聚类准则函数将空间对象与这些聚类中心和初始类逐一作比较, 判断对象的归属。K 均值算法是用每个聚类中所有对象的平均值作为该聚类(簇)的中心, 采用误差平方和最小准则判断对象的归属。

上述几种传统的分区方法, 都是根据空间单元属性数据的相似性程度划分为不同类型或区域, 但由于没有考虑单元的空间位置相互依赖关系而使分区结果出现许多孤立的单元或碎片, 不利于精准农业田间变量管理

收稿日期: 2004-11-04 修订日期: 2005-02-23

基金项目: 863 数字农业资助项目(2003AA209040)

作者简介: 李翔(1975-), 男, 河南鹿邑人, 博士生, 主要从事 GIS 空间分析与遥感应用研究。北京 北京师范大学地理学与遥感科学学院遥感与地理信息系统研究中心, 100875。Email: gm\_lxiang188@sohu.com

通讯作者: 赵春江(1964-), 研究员, 博士生导师, 北京 国家农业信息化工程技术研究中心, 100089。Email: zhaocj@nercita.org.cn

作业。然而一些空间单元区域划分问题的解决需要将空间单元的空间位置的关联性和属性数据的相似性结合起来进行聚类,这也是笔者提出空间连续性分区算法的目的所在。

## 2 基于空间连续性聚类算法 SCKM

Tobler (1978) 提出了地理学第一定律——空间自相关性,即“距离相近的物体比距离远的相关性大”,它是表征两个位置相近物体的相似程度,是自然界普遍存在的一种地理现象<sup>[7]</sup>。利用空间自相关性进行精准农业管理分区的提取可以使每个分区具有较高的空间自相关性。而实际上大部分空间自相关度量指标不能用于分类,如全局 Moran's I、Geary's C 等。因为这些指标是从全局角度来度量样本之间(在连续的栅格图像中,每个像元可看作一个样本)在整个区域的空间依赖程度,而不能度量每个样本对其周围样本的依赖性在空间上的差异。

Anselin (1995) 提出空间自相关的局部度量指标 LISA 算法<sup>[8]</sup>。

$$T_i = \frac{1}{N_j} \sum_{j=1}^N (W_{ij} Y_{ij}) \quad (1)$$

$$W_{ij} = C_{ij} / \sum_{j=1}^N C_{ij} \quad (2)$$

式中  $T_i$ ——局部空间关联指标;  $W_{ij}$ ——行标准化权重矩阵,可由公式(2)计算得到;  $C_{ij}$ ——二值空间权重矩阵,代表像元  $i$  与  $j$  的空间位置的关联性,  $N$  为像元总数,根据八邻域(eight-neighbor rule)判断像元是否相邻,像元  $i$  与  $j$  相邻时,  $C_{ij}$  为 1,否则为零;  $Y_{ij}$ ——像元  $i$  与  $j$  的属性值的关联性。

Getis 和 Ord 根据不同统计方法对  $Y_{ij}$  的假设和观念的差异,总结了几种不同的表达方式<sup>[9]</sup>,如: Getis 统计法,  $Y_{ij} = X_j$  或  $X_i + X_j$ ; Moran's I 统计法,  $Y_{ij} = (X_i - \bar{X})(X_j - \bar{X})$ ; Geary's C 统计法,  $Y_{ij} = (X_i - X_j)^2$ 。我们选择 Getis 统计表达方式,令  $Y_{ij} = X_j$ ,  $X_j$  表示像元  $j$  的属性值,因此像元  $i$  的空间关联指标可表示为:  $T_i =$

$(W_{ij} X_j), i = j$ 。每个像元都有相应的  $T_i$  值来表征与相邻像元的空间关联程度,  $T_i$  的变化可以反映变量在空间上分布的差异性。因此,以  $T_i$  为权重来修正每个像元的属性值(如公式(3)),然后再进行 K 均值聚类,这样在分类时不仅考虑了像元的属性值的相似性而且还兼顾了像元之间的空间关联性。

$$X_i = T_i X_i \quad (3)$$

基于此思想和 K 均值聚类算法,笔者提出了空间连续性聚类 SC-KM 算法:

根据反射率转换后的遥感影像提取每个像元的 NDVI 值  $X_i$ ;

根据八邻域(eight-neighbor rule)判断像元是否相邻,像元  $i$  与  $j$  相邻时,  $C_{ij}$  为 1,否则为零,计算归一化权重矩阵  $W_{ij} = C_{ij} / \sum_{j=1}^N C_{ij}$ ;

计算图像每个像元的局部空间关联指标  $T_i =$

$$(W_{ij} X_j), i = j;$$

修正图像每个像元的属性值  $X_i = T_i X_i$ ;

用 K 均值聚类算法对修正后图像进行分类。

SCKM 算法在空间聚类分析中,既考虑了空间单元的空间位置的关联性又考虑其属性数据的相似性,为空间单元区域划分问题的解决提供了有效途径。

## 3 分区效果评价指标

### 3.1 权重方差

方差是度量数据变异程度的一个重要指标。根据精准农业管理的思想,管理分区内的变异越小越便于在分区内实行统一管理,即越适宜作为变量管理的决策单元。为了表征各管理分区的区内变异对整个地块变异的贡献大小,进而评价整个地块区内方差随分区数变化的规律,笔者引入了权重方差的算法,如公式(4)。

$$WS_j = \left[ \frac{1}{N_j} \sum_{i=1}^{N_j} (Y_i - \bar{Y})^2 \right] \frac{N_j}{N_t} \quad (4)$$

式中  $WS_j$ ——管理分区  $j$  的权重方差;  $N_j$ ——管理分区  $j$  的像元数;  $Y_i$ ——像元  $i$  的属性值;  $\bar{Y}$ ——管理

分区  $j$  所有像元属性数据的平均值;  $\frac{1}{N_j} \sum_{i=1}^{N_j} (Y_i - \bar{Y})^2$ ——管理分区  $j$  的方差;  $N_t$ ——整个地块的像元总数。

整个地块的区内方差可以由公式  $TS = \sum_{j=1}^K WS_j$  计算得到,其中  $TS$  表示分区数为  $K$  时总的区内方差;  $WS_j$  为管理分区  $j$  的权重方差。

### 3.2 聚集度(CI)与破碎度(FI)

在景观生态格局分析中常用聚集度和破碎度来定量评价景观生态中缀块的聚集程度<sup>[10]</sup>,但其算法适于在同一分类图中不同景观要素之间的比较,而不适用于不同分类图中同一管理分区的分析和评价。因此,从管理分区像元之间的空间相邻性出发提出了两个评价指标的新算法。

$$CI_j = \frac{1}{N_j} \sum_{i=1}^{N_j} W_i$$

式中  $CI_j$  (Contagion Index)——管理分区  $j$  的聚集度;  $W_i$ —— $j$  分区中像元  $i$  与同一管理分区内的像元相邻的边数;  $N_j$ ——管理分区  $j$  的像元总数。  $CI$  越大,管理分区的聚集程度越高;相反,管理分区的碎片越多,聚集程度越低。

$$FI_j = \frac{1}{N_j} \sum_{i=1}^{N_j} C_i$$

式中  $FI_j$  (Fragment Index) 为管理分区  $j$  的破碎度;  $C_i$ —— $j$  分区中像元  $i$  与其它管理分区的像元相邻的边数;  $N_j$ ——管理分区  $j$  的像元总数。  $FI$  越大,管理分区的碎片或孤立的像元越多;相反,管理分区的聚集程度越高。

## 4 精准农业管理分区实例分析

### 4.1 实验区概况及数据源选择

本试验以北京精准农业示范基地为试验区,该基地位于北京昌平小汤山镇东北部,地处  $40^{\circ}10'31'' \sim 40^{\circ}11'18''N$ ,  $116^{\circ}26'10'' \sim 116^{\circ}27'05''E$ 。试验基地中的冬小麦种植面积约为  $39.2 \text{ hm}^2$ 。

实验采用了 2001 年 4 月 26 日(拔节期)在基地飞行获取的成像光谱图像 OMIS 数据。OMIS 系统采用线列探测器-光机扫描方式成像,该成像光谱仪瞬时视场为  $3 \text{ m rad}$ ,总视场为  $70^{\circ}$ ;覆盖了可见/近红外、中短波红外和热红外波谱区( $0.4 \sim 12.5 \mu\text{m}$ )划分为 5 个光谱段,共 128 个波段。其中可见/近红外波区( $0.46 \sim 1.1 \mu\text{m}$ )有 64 个波段,光谱分辨率  $10 \text{ nm}$ 。飞行高度  $1000 \text{ m}$  时,机下点分辨率约  $3 \text{ m}$ 。本研究在实验基地获取的 OMIS 数据中截取了一块地势平坦、大田生产力差异明显的小麦自然地块的影像(如图 1 所示)作为研究对象。该地块大小为  $62 \times 84$  像素,面积大约  $5 \text{ hm}^2$ 。

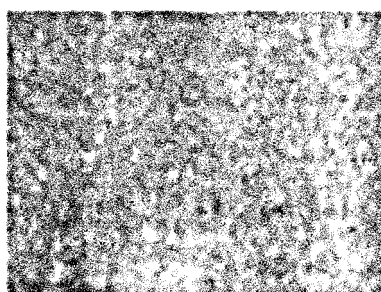


图 1 试验地 OMIS 原图

Fig. 1 Original OMIS image of studied wheat field

### 4.2 分区参数选择

拔节期是小麦肥水需求的关键时期,根据小麦的长势差异进行分区,及时实施变量管理来补充水肥是整个地块优质高产的有效途径。

归一化植被指数 NDVI 是作物长势遥感监测中应用最为广泛的指标,是作物生长状态及覆盖度的最佳指示因子,许多研究表明 NDVI 与 LAI 绿色生物量、作物覆盖度等参数有关,而且 NDVI 更适用于农作物发育中期的监测<sup>[11]</sup>。因此,在拔节期,根据 NDVI 进行管理分区的划分能较好的反映小麦长势的空间变异性,为农田变量投入、精准管理提供决策依据。本研究利用经过反射率转换后的 OMIS 图像,选择中心波长分别为

789.2 nm 和 675.8 nm 的两个波段,计算 NDVI。

$$NDVI = \frac{\rho_{NIR} - \rho_R}{\rho_{NIR} + \rho_R} \quad (5)$$

式中  $\rho_{NIR}, \rho_R$ ——分别代表近红外波段和红波段的反射率。

### 4.3 结果分析

由于地块在空间上存在肥力和管理因素等差异,致使地块不同区域间的小麦长势存在变异。整个地块不进行分区时的变异可被看作管理分区数为 1 的区内变异。由图 2 可以看出:随着管理分区数的增加,两种方法具有相同的趋势,区内小麦长势变异(用 NDVI 相对方差表示)逐渐减小。换句话说,管理分区数越多,管理分区作为变量管理决策单元的精度就越高。然而,这在实际生产中是不可行的,因为随着管理分区数的增多,必然带来管理成本的增加。因此,有必要选择精度和经济效益综合效果最佳的适宜管理分区数。以整个地块不进行分区时的变异即管理分区数为 1 的区内变异(方差)做参照(为 100%),计算不同分区数的相对方差。为了选择一个适宜的分区数,必须确定一个方差阈值。从图 2 可看出,当整个地块分为 5 个管理分区时,区内总的方差大致减少到原来的 10%,而且随分区数的进一步增加,方差相对变化不再明显,因此适宜的管理分区数为 5。

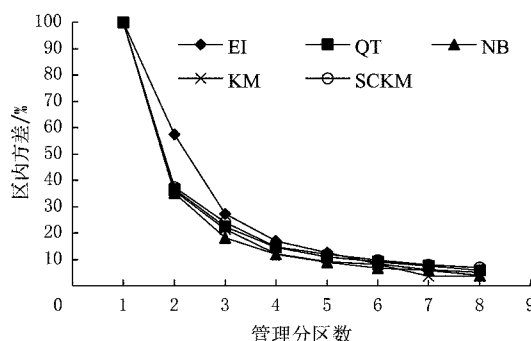


图 2 区内方差随管理分区数的变化

Fig. 2 Change trend of within-zone variance with number of management zone

管理分区的区内方差相比,分区数为 5 时, SCKM 法与传统的分区方法差异不显著(如图 2),这说明用 SCKM 法可以使各管理分区内差异较小而分区间差异较大,适宜在同一管理分区内实施同一管理,不同分区间实施变量管理,符合精准农业管理的思想。管理分区

表 1 SCKM 算法与 4 种传统方法的分区图的聚集度和破碎度的比较

Table 1 Comparison of the clustering index and fragment index of management zone maps based on five methods

分区	EI		QT		EI		KM		SCKM	
	CI	FI	CI	FI	NB	FI	CI	FI	CI	FI
1	2.16	1.52	3.32	0.62	3.18	0.70	3.34	0.59	3.61	0.32
2	2.87	1.04	2.79	1.16	3.05	0.90	2.80	1.15	3.59	0.37
3	3.12	0.83	2.66	1.26	3.15	0.79	2.38	1.53	3.33	0.59
4	3.45	0.49	2.96	0.99	2.93	1.00	2.89	1.05	3.53	0.41
5	3.54	0.42	3.51	0.46	3.33	0.64	3.41	0.56	3.76	0.21

的空间聚集度相比,两种方法的差异较大。从图 3 可以直观地看出,利用 EI、QT、NB、KM 等传统方法分区的结果,各管理分区都存在许多孤立的像元或碎片,而 SCKM 算法各管理分区则去除了大量的孤立单元或碎片,具有较好的空间连续性;从表 1 可以看出, SCKM

法各管理分区的聚集度均大于 4 种传统分区方法,而破碎度均小于后者。这表明 SCKM 算法由于分区时考虑像元间空间位置的关联性而大大提高了各管理分区的聚集度和连续性,适宜田间变量管理作业。

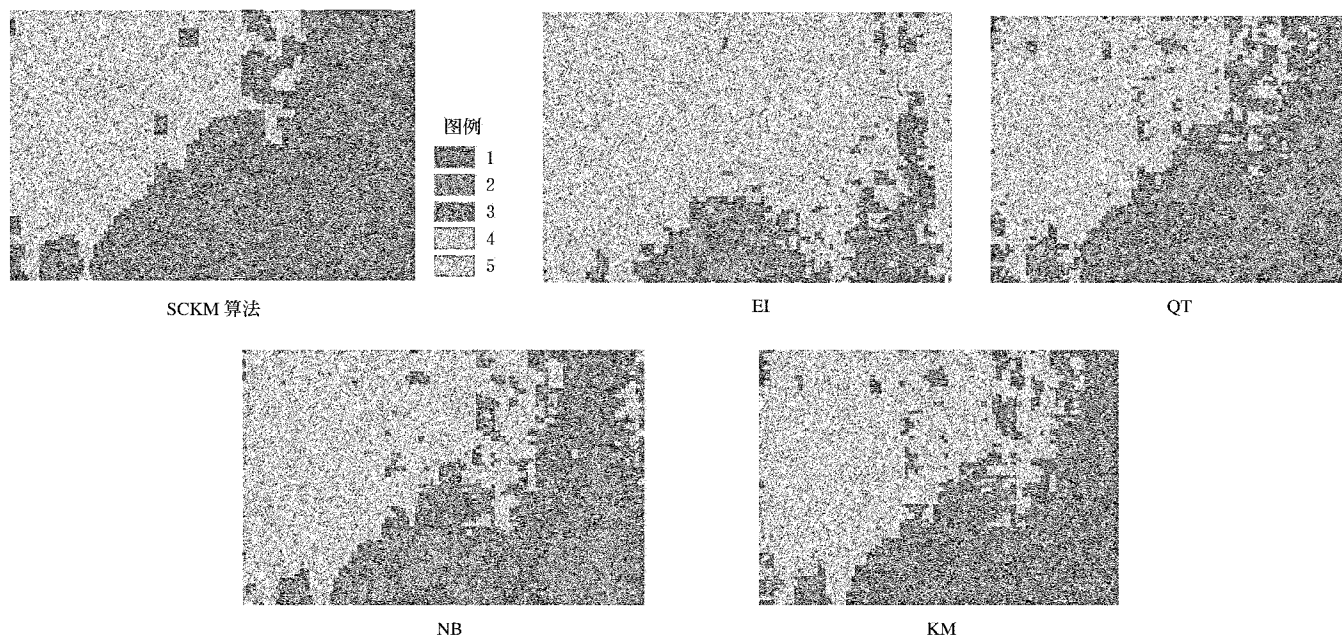


图 3 5 种方法分区结果

Fig 3 Zoned results based on five kinds of methods

## 5 结 语

本研究在 K 均值算法的基础上,提出了一种精准农业管理分区划分的新算法 SCKM,不仅考虑了空间单元的属性值之间的差异性,还考虑了空间单元的空间分布及其空间相互依赖性。验证结果表明 SCKM 算法可以使区内变异较小,而且去除了大量的碎片和孤立的像元,兼顾了管理分区的连续性,适宜精准农业田间变量管理作业。分区结果可以直接作为变量管理的决策单元,在同一管理分区内实施统一管理,不同分区间实施变量管理模式,如根据肥水需求关键时期的小麦长势差异,在不同管理分区间设计不同的目标产量进行产中变量追肥管理<sup>[12-14]</sup>。分区结果也可指导生产者和科技工作者进行土壤和作物农学参数采样,提高采样精度和效率<sup>[15]</sup>。因此,利用 SCKM 算法进行管理分区的划分为精准农业变量投入、精准管理提供了有效途径。

精准农业管理分区的划分往往需要考虑多种因素或变量,如多年的产量数据、地形数据、土壤数据等,本研究仅用单变量数据验证了 SC-KM 算法对精准农业管理分区的有效性,对多变量数据的适用性还需要进一步的验证。另一方面,管理分区的连续性或聚集程度不仅与土壤或作物长势状况的空间自相关程度有关,而且与选择的数据源空间尺度(即空间单元大小)有关。下一步将利用多源数据,考虑空间尺度对管理分区的影响因

素,研究精准农业管理分区的方法。

## [参 考 文 献]

- [1] Whelan B M, A B M & Ratney, R A Viscarra Rossel. Spatial prediction for precision farming [A]. Madison, WI Precision Agriculture [C]. ASA - CSSA - SSSA, 1996: 331- 342.
- [2] Kravchenko A N, Bullock D G. Comparison of interpolation methods for mapping soil P and K contents [A]. St Paul, MN, The 4th Inter Conf on Prec Agric [C]. 1998, 19.
- [3] Gotway C A, Ferguson R B, Hergert G W. The effects of mapping scale on variable rate fertilizer recommendations for corn [A]. Madison, WI Proc 3rd Inter Conf On Site-Specific Management for Agricultural Systems [C]. Am Soc Agron, 1996: 321- 330.
- [4] Fraisse C W, Sudduth K A, Kitchen N R. Use of unsupervised clustering algorithms for delineating within-field management zones [A]. ASA E Canada, 1999: 18- 21.
- [5] Kitchen N R, Sudduth K A, Drummond S T. An evaluation of methods for determining site-specific management zones [A]. St Louis, Proceedings of the North Central Extension-Industry Soil Fertility Conference, Potash and Phosphate Institute, Missouri, 1998: 133- 139.
- [6] Fridgen J J, Fraisse C W, Kitchen N R, et al

- Delineation and analysis of site-specific management zones [A]. Lake Buena Vista. The Second International Conference on Geospatial Information in Agriculture and Forestry[C]. Florida, 2000
- [7] Michael F, Goodchild. Spatial Autocorrelation[M]. U.K., Geo Books, Norwich, 1986: 3- 5
- [8] Anselin L. Local indicators of spatial association- LISA [J]. Geographical Analysis, 1995, 27(2): 93- 115
- [9] Getis A, Ord J K. Local spatial statistics: an overview [A]. Longley and M. Batty. Spatial analysis: modelling in a GIS environment[C]. 1996: 261- 277.
- [10] 邬建国. 景观生态学——格局、过程、尺度与等级[M]. 北京: 高等教育出版社, 2002: 100- 110
- [11] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003: 374- 375
- [12] Jiyul Chang, David E. Clay, Charles G. Carlson, et al. Different techniques to identify management zones impact nitrogen and phosphorus sampling variability [J]. Agronomy Journal, 2003, 95: 1550- 1559
- [13] Jiyul Chang, David E. Clay, Charles G. Carlson, et al. Defining yield goals and management zones to minimize yield and nitrogen and phosphorus fertilizer recommendation errors[J]. Agronomy Journal, 2004, 96: 825- 831.
- [14] Lukina E V, Freeman K W, Wynn K J, et al. Nitrogen fertilization optimization algorithm based on in-season estimates of yield and plant nitrogen uptake [J]. Plant Nutrition, 2001, 24: 885- 898
- [15] Brett Whelan. Moving to PA management in the grains industry [A]. Proceedings of the 1st Australian Conference on Geospatial Information in Agriculture[C]. NSW Agriculture, 2001: 406- 416

## Delineating precision agriculture management zones based on spatial contiguous clustering algorithm

Li Xiang<sup>1,2</sup>, Pan Yuchun<sup>2</sup>, Zhao Chunjiang<sup>2</sup>, Wang Jihua<sup>2</sup>, Bao Yansong<sup>1</sup>, Liu Liangyun<sup>2</sup>, Wang Jindi<sup>1,3</sup>

(1. Research Center for Remote Sensing and GIS, School of Geography, Beijing Normal University, Beijing 100875, China;

2. National Engineering Research Center for Information Technology in Agriculture, Beijing 100089, China;

3. State Key Laboratory of Remote Sensing Science, Beijing 100875, China)

**Abstract:** Based on the traditional K-Means cluster (KM) and the spatial autocorrelation, a new method, Spatial Contiguous K-Means clustering algorithm (SCKM), was developed in this study. According to the spatial variability of wheat growth under within-field level extracted from OMIS image of the key growth stage, precision agriculture management zones were delineated by using the SCKM method and the traditional methods such as KM, Equal Interval, Quantile and Natural Breaks method. Two evaluation indices were employed to evaluate the zoned results of the above mentioned methods. The results showed that the sum of the weighted variance of the corresponding within-zones based on these methods appeared no significant difference, and that the SCKM method could remove a lot of isolated cell or patch and improved the continuity of the corresponding management zone map, compared with the traditional methods. The zoned result based on the SCKM method can be used as the variable management unit for precision agriculture and can be used to advise the sampling of subsequent soil or crop.

**Key words:** precision agriculture; management zoning; SCKM Algorithm; evaluation index