

# 基于模糊聚类分析的田间精确管理分区研究

李艳<sup>1,2</sup>, 史舟<sup>1</sup>, 吴次芳<sup>2</sup>, 李锋<sup>3</sup>, 程街亮<sup>1</sup>

(<sup>1</sup> 浙江大学环境与资源学院农业遥感和信息技术应用研究所, 杭州 310029; <sup>2</sup> 浙江大学东南土地管理学院土地科学与不动产研究所, 杭州 310029;

<sup>3</sup> 浙江大学环境工程系, 杭州 310029)

**摘要:** 【目的】以海涂围垦区盐碱土为研究对象, 将从 SPOT 遥感影像提取出的 NDVI 数据和盐碱地土壤生产力的主要限制因子盐分数据及部分养分数据作为变量进行精确农业管理分区研究。【方法】模糊 c 均值聚类方法被用来进行分类分区, 并引入了模糊聚类指数和归一化分类熵两种分区效果评价指标, 对分区结果进行比较和评价。

【结果】本研究区, 最佳的分区数目为 3 个。对处于每一子区内土样的化学特性和实测棉花产量数据进行方差分析, 发现其均值在所定义的每个管理分区内都存在着统计意义上的显著差异性, 其中子区 3 具有最高的肥力水平和作物产量, 而子区 1 最低。【结论】利用所选取的变量, 模糊 c 均值聚类算法可以较好地进行管理分区划分。分区结果不但可以指导采样, 而且可用于实施变量投入和精确施肥推荐, 为样区土壤管理提供科学的决策依据。

关键词: 模糊聚类; 管理分区; 盐碱土; 精确农业

## Definition of Management Zones Based on Fuzzy Clustering Analysis in Coastal Saline Land

LI Yan<sup>1,2</sup>, SHI Zhou<sup>1</sup>, WU Ci-fang<sup>2</sup>, LI Feng<sup>3</sup>, CHENG Jie-liang<sup>1</sup>

(<sup>1</sup> Institute of Agricultural Remote Sensing and Information Technology Application, Zhejiang University, Hangzhou 310029;

<sup>2</sup> Institute of Land Science and Property Management, College of Southeast Land Management, Zhejiang University, Hangzhou

310029; <sup>3</sup> Department of Environmental Engineering, Zhejiang University, Hangzhou 310029)

**Abstract:** 【Objective】In this paper, the variables of NDVI data, soil salinity and some nutrient data were selected as data sources to study coastal saline field management zones. 【Method】Fuzzy c-means a clustering algorithm was used to delineate management zones. A fuzzy performance index and normalized classification entropy were used to select the optimal cluster numbers.

【Result】The optimal number of management zones for the study area was 3. To assess whether the defined management zones can be used to characterize spatial variability of soil chemical properties and crop productivity, 139 geo-referenced soil and yield sampling points across each management zone were examined by using one-way ANOVA. There were significantly statistical differences between the chemical properties of soil samples and yield data in each defined management. Management zone 3 presented the highest nutrient level and potential crop productivity, whereas management zone 1 was the lowest. 【Conclusion】The results revealed that fuzzy c-means clustering algorithm could be used to delineate management zones by using the given variables. The defined management zones not only direct soil-sampling design, but can be used for fertilizer recommendation to manage soil nutrient and salinity more efficiently in coastal saline field.

Key words: Fuzzy clustering; Management zones; Saline land; Precision agriculture

收稿日期: 2006-03-22; 接受日期: 2006-06-06

基金项目: 中德合作项目 (BMBF-AZ39742), 国家自然科学基金 (40001008, 40571066)

作者简介: 李艳 (1978-), 女, 河南南阳人, 讲师, 研究方向为遥感和地理信息系统、土地利用信息技术和精确农业。E-mail: liyan522@zju.edu.cn。  
通讯作者史舟 (1970-), 男, 浙江舟山人, 副教授, 研究方向为农业遥感与信息技术。Tel: 0571-86971831, 13067741224; E-mail: shizhou@zju.edu.cn

## 0 引言

【研究意义】定义管理分区来对土壤和农作物实施变量投入管理是精确农业近年来的研究热点<sup>[1]</sup>。管理分区就是由相似的地貌或土壤状况所导致的相似的作物生产潜力、相似的养分利用效率及相似的环境效应的子区域。按照田间每一管理分区的具体条件,采用不同的土壤管理办法,相应调整投入物质的施入量,这样既能提高土壤养分利用效率和管理精度,提高农产品产量和品质,又能节省资源,获得较好的经济效益,达到保护农业资源和环境质量的目的<sup>[2]</sup>。研究表明管理分区可以作为网格采样的一种替换手段在变量施肥中得到应用<sup>[3]</sup>。Fraisie 等认为管理分区在联系作物与土壤及地形参数,进行作物模拟和评价方面也是很有用的<sup>[4]</sup>。【前人研究进展】许多数据源可以被用来定义管理分区。土壤属性(肥力、电导率、有机质、质地等)、遥感影像、地形因子(坡度、坡向和高程)、产量数据等都被认为是定义田间管理分区的合乎逻辑的基础变量。很多研究者运用上述一种或几种因素结合来进行管理分区划分。如 Long 等人发现由作物生长期的遥感影像划分田块管理分区来预测作物产量是最精确的<sup>[5]</sup>。Johnson 等表明由土壤电导率所确定的管理区可被用于量化麦子和谷物产量的空间变异<sup>[6]</sup>。Kravchenko 等利用地形特征和土壤属性来定义管理分区<sup>[7]</sup>。Franzen 等利用地形数据、遥感数据、土壤电导率以及作物产量数据来建立土壤氮素管理分区<sup>[8]</sup>。国内近几年也开始了这方面的研究。如白由路等建立了区域土壤养分分区管理模型并将之应用于施肥推荐<sup>[9]</sup>。黄绍文等制订作物高产优质分区平衡施肥技术来指导养分管理和肥料的合理使用<sup>[10]</sup>。李翔等利用多年作物产量数据进行精准农业管理分区的提取研究<sup>[11]</sup>。【本研究切入点】聚类分析由于能有效地将具有相似作物产量、地形特征以及土壤肥力特性的区域进行归类而被较多地用于精确农业管理分区中。如 Stafford<sup>[12]</sup>, Boydell<sup>[13]</sup>和 Jaynes<sup>[14]</sup>等都是通过聚类分析利用作物产量数据来划分管理分区。但这些报道很少对最佳分区数目的确定进行定量化表达<sup>[15]</sup>。国内李翔等人利用聚类算法对小麦长势差异进行了管理分区研究,通过分析方差减少率来确定适宜的分区数<sup>[16]</sup>。

【拟解决的关键问题】本研究根据土壤质量和作物产量主要限制因子的空间变异性信息,利用模糊 c 均值聚类算法定义田间管理分区,引入了模糊性能指数和归一化分类熵两种分区效果评价指标确定最佳的分区

数目,利用方差分析对分区结果进行比较,评价对海涂围垦区进行精确管理分区的可能性。

## 1 材料与方法

### 1.1 研究区概况

研究区位于浙江省上虞市西北地区、杭州湾南岸的海涂实验农场。该地属于亚热带区域,自然植被为常绿阔叶林,年平均温度 16.5℃,年均降水量为 1 300 mm。土壤主要来自于海洋和河流沉积物,土壤质地以轻壤土或砂壤土为主。土壤类型是粗松咸砂土,剖面发育不明显,同一剖面上下层质地较为均匀。近年来该区已被连续垦种。研究田块围垦于 1996 年,2000 年开垦为棉田,面积为 10.5 ha。

### 1.2 数据采集

运用网格采样法(间距为 20~40 m),通过一个便携式电导率野外快速测定计,在表土(0~20 cm)共测得 139 个土壤体积电导率(bulk electrical conductivity,简称 EC<sub>b</sub>)数据。采样时间为 2003 年 11 月。在每个网格点 1 m 的圆周范围内共采集样品 5 个,取其平均值作为该网格点的 EC<sub>b</sub> 值。差分 GPS 被用来进行样点定位。同时,共 139 个土壤样品被采集,并利用常规方法进行了室内 pH、EC1:5、有机质、速效磷、速效钾、碱解氮、全氮及阳离子交换量的测定。

遥感影像采用获取于 2003 年 9 月 27 日的 SPOT5 卫星影像,提取归一化植被指数(NDVI)来反映当季棉花长势。并实测了共 139 个棉花产量数据(籽棉产量)来进行验证。在每个采样点 1 m 的圆周范围内共选择棉株 5 株,棉花采集晒干后取其平均值作为该样点的棉花籽棉产量数据。

### 1.3 研究方法

1.3.1 主成分变换 主成分分析是一种基于数据统计特征的多维正交线性变换。它可以将具有相关性的多个要素的信息量压缩到完全独立的较少的几个主成分要素中,这些主成分包含原始资料中尽可能多的信息,且彼此间互不相关,在特征空间中相互垂直,也即各个主成分包含的信息内容是不重叠的,从而大大减少总的的数据量并使特征信息得到增强<sup>[17]</sup>。

1.3.2 模糊 c 均值聚类 模糊 c 均值聚类是常用的一种非监督聚类方法,并被大量用于进行土壤、地形地貌、产量和遥感数据等的分类中<sup>[15]</sup>。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:

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m (d_{ik})^2 \quad (1)$$

其中,  $n$  是数据个数,  $c$  是类别数,  $m$  为模糊加权指数 ( $1 \leq m \leq \infty$ ), 它控制了不同类别间共用数据的数目, 当  $m$  趋于无穷大时, 共用数据数目增加, 最终的分类变得较不明显, 当  $m=1$  时, 就会发生硬聚类, 即没有数据共用现象。  $u_{ik}$  ( $1 \leq i \leq c, 1 \leq k \leq n$ ) 表示数据矩阵  $X$  中第  $k$  个样本  $x_k$  属于聚类中心矩阵  $V$  中第  $i$  个聚类中心  $v_i$  的隶属度值,  $U$  为隶属度矩阵。  $d_{ik}^2$  等于  $x_k$  与  $v_i$  在特征向量上距离的平方。

模糊  $c$  均值聚类算法是通过一个迭代优化的过程来实现分类。首先是对隶属度矩阵  $U$  的初始化, 然后根据公式 (2) 计算出  $c$  个聚类中心, 由这些聚类中心计算在类中各个数据点的隶属度值 (见公式 3), 调整隶属度矩阵, 依据调整完的隶属度矩阵计算出新的  $c$  个聚类中心, 比较聚类中心的变化, 如果两次循环之间的变化量小于预设的阈值, 则停止迭代计算, 否则重复前述步骤直至聚类中心变化小于设定的阈值。

$$v_i^{(b)} = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m} \quad (2)$$

$$u_{ik}^{(b+1)} = \left[ \frac{\sum_{j=1}^c \left( \frac{d_{ik}^{(b)}}{d_{jk}^{(b)}} \right)^{2/(m-1)}}{d_{jk}^{(b)}} \right]^{-1} \quad (3)$$

**1.3.3 聚类有效性检验** 聚类时, 经常产生的一个问题时是究竟划分多少个类别才合适。由聚类有效性可知, 好的聚类应提供尽可能明晰的划分。可采用模糊性能指数和归一化分类熵来确定合适的聚类数并进行

聚类效果优劣的检验。

模糊性能指数 (FPI) 是数据矩阵  $X$  的模糊  $c$ -分区分离程度的度量, 可定义为:

$$FPI = 1 - \frac{c}{(c-1)} \left[ 1 - \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^2 / n \right] \quad (4)$$

FPI 的值在 0 到 1 之间变动。若该值接近 0 表示聚类时共用数据较少, 类的划分明显。若该值接近 1 则表示具有较多的共用数据, 类的划分不明显。FPI 越小聚类效果越好。

另一种度量指标是归一化分类熵 (NCE), 它用来模拟数据矩阵  $X$  的模糊  $c$ -分区的分解量。其中分类熵 ( $H$ ) 可用下列函数定义:

$$H(U; c) = - \sum_{k=1}^n \sum_{i=1}^c u_{ik} \log_a(u_{ik}) / n \quad (5)$$

这里对数的底数  $a$  可为任意正整数。  $H$  的值在 0 到  $\log_a(c)$  之间变动。由于当  $c$  为 1 和  $n$  时,  $H$  都等于 0, 因此 Bezdek 在分类熵的基础上提出了归一化分类熵的概念<sup>[18]</sup>, NCE 越小则模糊  $c$ -分区的分解量越大, 分类效果越好。

$$NCE = H(U; c) / [1 - (c/n)] \quad (6)$$

## 2 结果与分析

### 2.1 土壤数据及产量数据的逐步回归分析

表 1 是对野外实测和实验室分析得到的各个成分进行传统分析的统计特征值。

研究区土壤呈中性至碱性, pH 值在 7.0~8.5 之间变动, 变异系数较小。电导率 ( $EC_b$  和  $EC_{1:5}$ ) 的

表 1 土壤化学性质和作物产量传统统计分析的特征值

Table 1 Descriptive statistics of soil chemical properties and crop yield

| 变量<br>Variables                | 均值<br>Mean | 中值<br>Median | 标准差<br>SD | 变异系数<br>CV(%) | 最小值<br>Min. | 最大值<br>Max. | 偏度<br>Skewness | 峰度<br>Kurtosis |
|--------------------------------|------------|--------------|-----------|---------------|-------------|-------------|----------------|----------------|
| pH                             | 7.82       | 7.81         | 0.229     | 2.9           | 7.19        | 8.50        | 0.472          | 0.724          |
| $EC_b$ (mS/m)                  | 135.7      | 119          | 87.6      | 64.6          | 15          | 345         | 0.436          | -0.842         |
| $EC_{1:5}$ ( mS/m)             | 134.0      | 111          | 98.02     | 73.1          | 7.40        | 513.5       | 1.239          | 1.802          |
| OM( $g \cdot kg^{-1}$ )        | 8.22       | 8.20         | 2.24      | 27.3          | 2.50        | 14.30       | 0.031          | -0.386         |
| TN ( $g \cdot kg^{-1}$ )       | 1.19       | 1.12         | 0.36      | 30.3          | 0.52        | 2.67        | 1.307          | 2.338          |
| AN ( $mg \cdot kg^{-1}$ )      | 46.80      | 42.81        | 23.77     | 50.8          | 13.66       | 95.8        | 3.021          | 15.188         |
| AP ( $mg \cdot kg^{-1}$ )      | 33.47      | 27.7         | 22.44     | 67.1          | 0.40        | 104.50      | 0.880          | 0.529          |
| AK ( $mg \cdot kg^{-1}$ )      | 105.53     | 92           | 54.44     | 51.6          | 28.50       | 249.81      | 0.687          | -0.520         |
| CEC( $cmol \cdot kg^{-1}$ )    | 7.75       | 7.72         | 2.10      | 27.1          | 3.93        | 19.14       | 0.206          | 4.998          |
| 棉花产量<br>Cotton yield (g/plant) | 213.7      | 182          | 148.9     | 69.7          | 14          | 714         | 1.176          | 0.931          |

$n=139$ ;  $EC_b$ : 体积电导率;  $EC_{1:5}$ : 土水比为 1:5 的电导率; OM: 有机质; TN: 全氮; AN: 碱解氮; AP: 速效磷; AK: 速效钾; CEC: 阳离子交换量  
The number of sample is 139;  $EC_b$ : Bulk electrical conductivity;  $EC_{1:5}$ : Electrical conductivity of solution comprising 1 litre of soil and 5 liters of water; OM: Organic matter; TN: Total nitrogen; AN: Alkaline nitrogen; AP: Available phosphorus; AK: Available kalium; CEC: Cation exchange capacity

变动范围极大, 变异系数达 70%左右, 平均含量高且分布不均匀。有机质平均含量较低, 属于低肥力水平, 与浙江省高产水稻土的有机质含量相比, 不及后者的 1/4<sup>[19]</sup>。全氮含量水平较低, 基础肥力差。海涂区由于砂性重且呈碱性, 氮素易损失, 样区碱解氮的含量在 14~95.8 mg·kg<sup>-1</sup> 之间, 仍处于很低的水平, 且变异系数较大, 各个区域含量不均。速效磷含量较高, 这与农户连年施用磷肥有关。速效钾含量处于中等水平, 基本能满足作物生长需要。整体来讲, 海涂区土壤盐

碱重, 砂性高, 有机质含量低, 氮磷等有效养分含量不高, 保肥力不强, 加之淡水资源缺乏, 而且常遇台风袭击, 导致土地产出率低而不稳, 农业经济效益不高。

表 2 是样区棉花产量与土壤化学特性的相关关系矩阵。盐分指标如 EC<sub>b</sub>, EC1:5 与棉花产量呈显著负相关关系。相反, 养分指标如有机质、全氮、碱解氮、速效磷和阳离子交换量与棉花产量呈显著的正相关关系, 而速效钾和 pH 与棉花产量关系不明显。

为了得到描述土壤属性对作物产量影响的最优模

表 2 样区土壤化学特性和作物产量的相关系数矩阵

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

| 观测样本<br>Observation<br>items | pH       | EC <sub>b</sub><br>(mS/m) | EC1:5<br>(mS/m) | OM<br>(g·kg <sup>-1</sup> ) | TN<br>(g·kg <sup>-1</sup> ) | AN<br>(mg·kg <sup>-1</sup> ) | AP<br>(mg·kg <sup>-1</sup> ) | AK<br>(mg·kg <sup>-1</sup> ) | CEC<br>(cmol·kg <sup>-1</sup> ) | 产量<br>Yield<br>(g/plant) |
|------------------------------|----------|---------------------------|-----------------|-----------------------------|-----------------------------|------------------------------|------------------------------|------------------------------|---------------------------------|--------------------------|
| pH                           | 1        |                           |                 |                             |                             |                              |                              |                              |                                 |                          |
| EC <sub>b</sub> (mS/m)       | -0.038   | 1                         |                 |                             |                             |                              |                              |                              |                                 |                          |
| EC1:5 (mS/m)                 | -0.272** | 0.852**                   | 1               |                             |                             |                              |                              |                              |                                 |                          |
| OM(g·kg <sup>-1</sup> )      | -0.242** | -0.182*                   | -0.232**        | 1                           |                             |                              |                              |                              |                                 |                          |
| TN(g·kg <sup>-1</sup> )      | -0.035   | -0.194*                   | -0.202*         | 0.606**                     | 1                           |                              |                              |                              |                                 |                          |
| AN(mg·kg <sup>-1</sup> )     | -0.495** | -0.053                    | -0.191*         | 0.335**                     | 0.508**                     | 1                            |                              |                              |                                 |                          |
| AP(mg·kg <sup>-1</sup> )     | -0.348** | -0.187*                   | -0.181*         | 0.098                       | 0.005                       | 0.077                        | 1                            |                              |                                 |                          |
| AK(mg·kg <sup>-1</sup> )     | 0.065    | 0.425**                   | 0.420**         | 0.151                       | 0.121                       | 0.107                        | 0.255**                      | 1                            |                                 |                          |
| CEC(cmol·kg <sup>-1</sup> )  | -0.156   | -0.204*                   | -0.291**        | 0.408**                     | 0.302**                     | 0.248**                      | 0.110                        | 0.066                        | 1                               |                          |
| 产量 Yield<br>(g/plant)        | -0.011   | -0.524**                  | -0.549**        | 0.335**                     | 0.431**                     | 0.255**                      | 0.271**                      | 0.112                        | 0.284**                         | 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

型, 将 EC<sub>b</sub>、有机质、全氮、碱解氮、速效磷、阳离子交换量作为自变量, 棉花产量作为因变量, 进行多元逐步回归分析。弃去对棉花产量影响较不显著的碱解氮和速效磷, 在 0.05 的显著水平下, 有如下回归方程:

$$Y=760.83-2.52 (EC_b)+1.47 (OM)+1.14 (TN)+0.893 (CEC) (R^2=42\%)$$

其中决定系数  $R^2=42\%$ , 表示利用式中的几个变量, 模型只能解译实测棉花产量总方差的 42%。公式中的回归系数表示每个变量对回归分析贡献的大小。EC<sub>b</sub> 具有最大的负的回归系数, 意味着 EC<sub>b</sub> 的增加会大大减少作物产量。傅庆林等对滨海盐土盐分对作物产量的影响研究也发现, 盐分与棉花、大豆和芥菜等作物的相对产量呈极显著负相关关系, 负相关系数高达 0.9<sup>[20]</sup>。这些研究表明, 在类似海涂围垦区, 土壤中过多的可溶性盐分是影响作物生长发育和产量的最重要的限制因素。有机质, 全氮和阳离子交换量与棉花产量是正相关关系, 土壤中这些成分的增加会提高

作物产量。

## 2.2 回归因子的空间分布格局图制作

图 1 列出了逐步回归分析得到的与棉花产量较为相关的几个变量: EC<sub>b</sub>、有机质、全氮、阳离子交换量的半方差图及其拟合模型参数。这 4 种变量都具有很好的空间结构。其中对 EC<sub>b</sub> 和有机质, 球状模型对其经验半方差值的拟合效果最好, 而全氮和阳离子交换量则是指数模型拟合的效果最好。4 种土壤属性的变程在 45~310 m 之间变化, 块金值与基台值的比值 22%~35% 之间变化, 表示其具有强烈或中等程度的空间自相关性<sup>[21]</sup>。

普通克里格插值法随后被用来对未采样点的 EC<sub>b</sub>、有机质、全氮和阳离子交换量进行空间预测, 形成这 4 个因子的空间分布图, 并将其转换为栅格大小为 10 m 的栅格图像 (图 2)。可以发现, 田块的东部区域盐分含量较高, 而北部和西南部含量低。全氮则刚好相反, 是东部区域含量较低, 而北部和西南区含量较高。有机质和阳离子交换量的区域分布较为一致,

西部区域含量高而东部区含量低。

归一化植被指数 (NDVI) 与叶面积系数、植被覆

盖度、生物量和作物长势等植被参数密切相关, 常用来进行地表植被动态调查、农业作物长势监测和产量

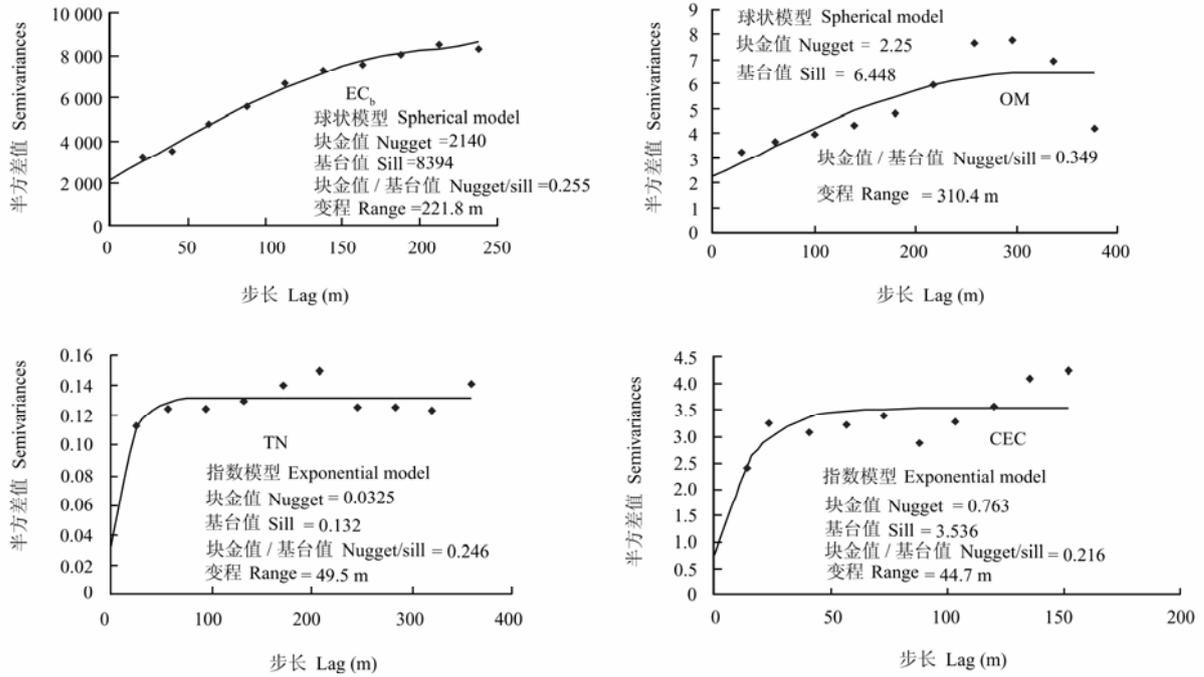


图 1 土壤特性的半方差值及其拟合曲线和参数

Fig. 1 Semivariogram of some soil properties and its fitted curve and parameters

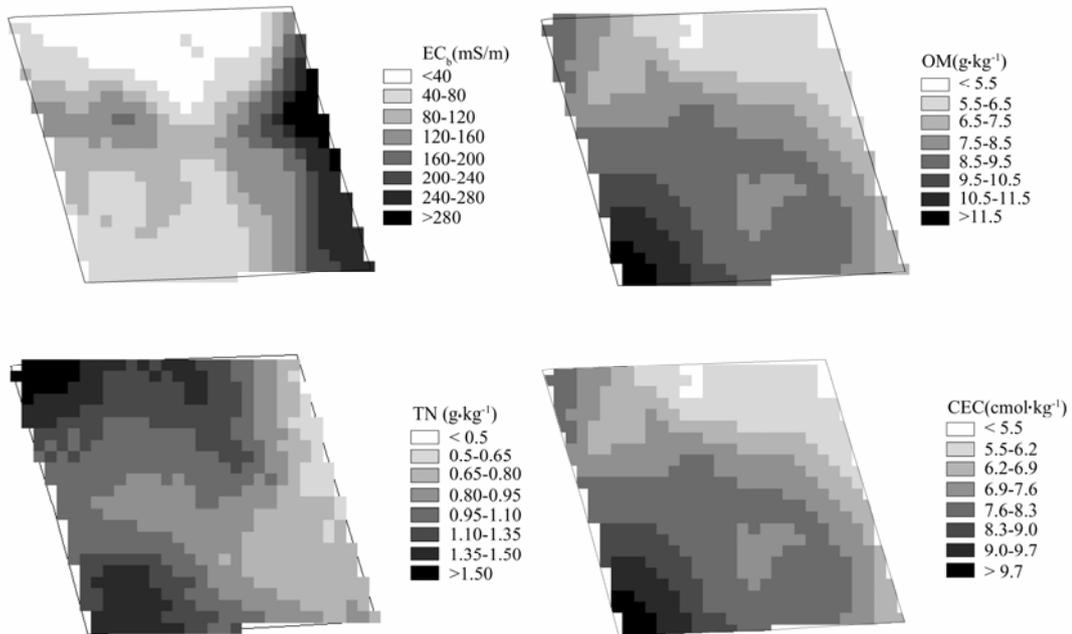


图 2 样区土壤 EC<sub>b</sub>、有机质、全氮和 CEC 的插值图像

Fig. 2 Smoothed contour maps produced by ordinary kriging for EC<sub>b</sub>, OM, TN and CEC

预测预报。本研究利用获取的 SPOT5 遥感图像提取 NDVI 值(图 3)来反映地面棉花盖度和长势。可以看出, NDVI 图像整体与盐分的空间分布图较为吻合。较高的 NDVI 值出现在盐分含量较低的北部和西南区域, 而较低的 NDVI 值则出现在盐分含量较高的东部区域, 这主要是由于盐分与作物生长的反比关系所引起的。棉花长势好的区域 NDVI 值较高, 盐分含量低, 相反棉花长势差的区域 NDVI 值较低, 盐分含量高。NDVI 图像作为一个反映地面作物长势和盖度的指标, 与图 2 中的  $EC_b$ 、有机质、全氮和阳离子交换量的分布图像一起用于主成分分析并进而用于聚类分析来形成精确农业管理分区。

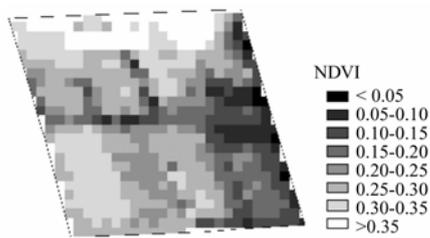


图 3 SPOT5 遥感影像提取的样区当季作物生长 NDVI 图像  
Fig. 3 NDVI images produced from SPOT5 image

### 2.3 空间分布格局图和 NDVI 影像图的主成分分析

由于用于聚类分析的  $EC_b$ 、有机质、全氮和阳离子交换量的栅格数据之间存在一定的相关性, 各指标反映的信息在一定程度上有重叠。本文采用主成分分析方法将原始栅格图像中的有用信息集中到尽可能少的新的主成分图像中来减少总的的数据量以使栅格图像信息得到增强。

表 3 研究区土壤养分含量的主成分描述及提取

Table 3 Principal component analysis of the five attributes in the study area

| 主成分<br>Principal components    | 特征根<br>Attribute value | 贡献率<br>Contribution rate (%) | 累积贡献率<br>Accumulative contribution rate (%) |
|--------------------------------|------------------------|------------------------------|---|
| 主成分 1<br>Principal component 1 | 2.861                  | 57.23                        | 57.227                                      |
| 主成分 2<br>Principal component 2 | 1.521                  | 30.42                        | 87.65                                       |
| 主成分 3<br>Principal component 3 | 0.263                  | 5.25                         | 92.90                                       |
| 主成分 4<br>Principal component 4 | 0.212                  | 4.23                         | 97.14                                       |
| 主成分 5<br>Principal component 5 | 0.143                  | 2.87                         | 100   |

表 4 主成分矩阵

Table 4 Principal component matrix

| 变量<br>Variables                | $EC_b$ | NDVI   | OM    | TN     | CEC   |
|--------------------------------|--------|--------|-------|--------|-------|
| 主成分 1<br>Principal component 1 | -0.849 | 0.824  | 0.481 | 0.883  | 0.370 |
| 主成分 2<br>Principal component 2 | 0.375  | -0.421 | 0.830 | -0.217 | 0.683 |

将图 2 中的土壤  $EC_b$ 、有机质、全氮、阳离子交换量的空间插值图像, 结合图 3 的 NDVI 图像共 5 个指标进行主成分分析, 得到各主成分得分值的栅格图。提取主成分的个数一般要求特征根大于 1, 累积贡献率超过 85%。利用这一标准, 本文提取了前 2 个主成分, 累积贡献率为 87.65% (表 4), 它们代表了反映植被长势和土壤特性变异的 87.65% 的原始信息, 可信度较高。

由表 4 可知, 第一主成分中土壤  $EC_b$ 、NDVI 和全氮所占的比重大于其它指标的系数, 故第 1 主成分是土壤盐分水平、全氮水平和植被覆盖程度的综合反映。第 2 主成分以有机质和阳离子交换量所占比重最大, 说明第 2 主成分是有有机质和阳离子交换量的综合反映。图 4 是得到的第 1 主成分影像 PC1 和第 2 主成分影像 PC2。

### 2.4 基于模糊聚类分析的管理分区生成及校验

第 1 主成分和第 2 主成分影像数据一起作为输入变量用于模糊 c 均值聚类分析中。取最大迭代次数为 300, 收敛阈值为 0.001, 模糊加权指数为 2。执行分类后, 即获得初步的分类结果。为了找出最佳分区数目, 分别产生 2、3、4、5 和 6 个类别, 计算并比较每一类的模糊性能指数 FPI 和归一化分类熵 NCE 的值。两者都是越小越好。FPI 越小表示聚类中的共用像元数越小, NCE 越小则表示分解量越大。若 FPI 和 NCE 的值在某个类别时都最小, 表示该类是研究区的最佳分类数。图 5 是将研究区划分为 2、3、4、5 和 6 个类别时所计算出的 FPI 和 NCE 值。可以看出, 当分类数目为 3 时, 产生的模糊性能指数 FPI 和归一化分类熵 NCE 都最小, 因此对本研究区, 最佳的分区数目为 3。图 6 是将研究区划分为 3 个区的管理分区图。

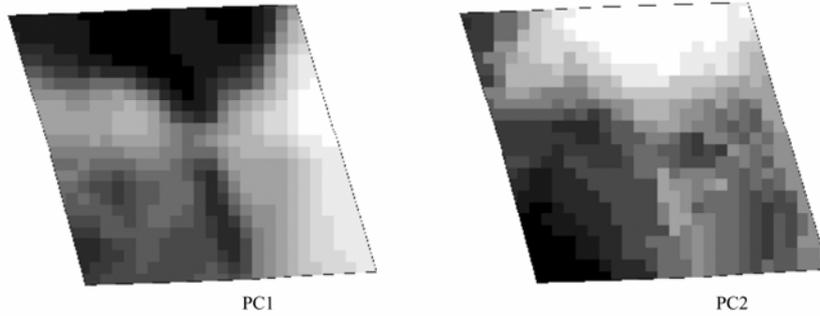


图 4 主成分分析得到的第 1 和第 2 主成分影像图  
Fig. 4 PC images produced by principal component analysis

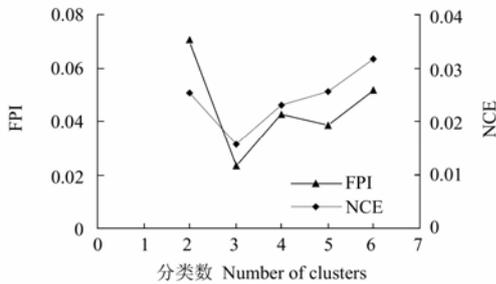


图 5 把研究区划分不同类别时的 FPI 和 NCE 的值  
Fig. 5 Calculated FPI and NCE for study area

为了确定分区结果是否可以有效量化那些影响作物生长的土壤化学性质的空间变异性，以最佳分区数目 3 个区为例，对分布于这 3 个管理区上的 139 个土壤数据和棉花产量数据进行均值统计和单向方差分析（表 5）。结果显示除了 pH 外，不同管理分区内平均的土壤 EC<sub>b</sub>、EC1 : 5、速效磷、速效钾、碱解氮、及实测的棉花产量数据在 99% 的置信水平上都具有极显著差异，而有机质、全氮、阳离子交换量在 95% 的置信水平上有显著差异。总的来说，分区 3 具有最高的土壤肥力水平，该区具有最低的盐分含量和最高的棉花产量，反映土壤肥力水平的指标如速效磷、有机质、碱解氮、全氮、阳离子交换量在该区也都是最高的，而速效钾含量水平则较低；相反，速效钾含量在盐分含量最高，总体肥力水平最低的分区 1 却很高。傅庆林等的研究也发现，速效钾含量在盐分含量较高的新围垦区比在盐分含量较低的老围垦区要高，这与不同区在土壤熟化过程中钾肥的施用情况有关<sup>[20]</sup>。分区 1 由于盐分过高，迟苗、缺苗和死苗等现象非常严重，

每年的播种面积、出苗面积和收获面积有很大的差距，产量不容乐观。农民对该区基本上处于放任状态，尽量少施肥或不施肥，再加上阳离子交换量低，保肥能力差，因此使分区 1 成为 3 个区中作物产量最低的。分区 2 的作物产量和肥力水平居中。

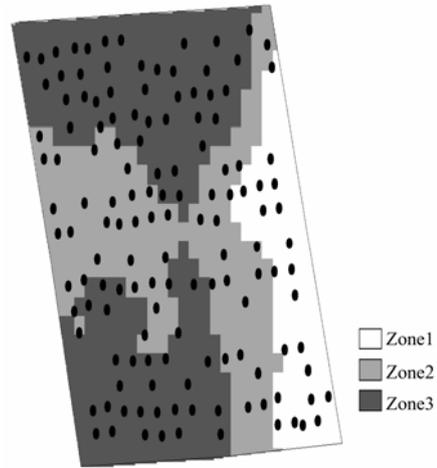


图 6 管理分区图及采样点在各区上的空间分布  
Fig. 6 Management zones map and spatial distribution of sampling point across each zone

### 3 讨论

精确农业管理分区的定义通常依赖于那些空间上可预测且具有时间稳定性的土壤和作物产量的空间变异信息。因此，管理分区的划分往往要考虑多年的产量数据、土壤属性和地形特征等多种因素和变量。本研究仅用单一年份的土壤、作物长势状况及产量数据的空间变异性验证了模糊 c 均值算法对精确农业管理

表 5 每个分区内土壤特性、棉花产量数据的均值统计及单因素方差分析

Table 5 One-way variance analysis of soil properties and cotton yield in the topsoil for the three management zones

| 管理分区<br>Zones             | 样点数目<br>No. of samples | 土壤特性<br>Soil properties |                           |                 |                              |                              |                             |                              |                             |                                 | 产量<br>(g/plant) |
|---------------------------|------------------------|-------------------------|---------------------------|-----------------|------------------------------|------------------------------|-----------------------------|------------------------------|-----------------------------|---------------------------------|-----------------|
|                           |                        | pH                      | EC <sub>b</sub><br>(mS/m) | EC1:5<br>(mS/m) | AP<br>(mg·kg <sup>-1</sup> ) | AK<br>(mg·kg <sup>-1</sup> ) | OM<br>(g·kg <sup>-1</sup> ) | AN<br>(mg·kg <sup>-1</sup> ) | TN<br>(g·kg <sup>-1</sup> ) | CEC<br>(cmol·kg <sup>-1</sup> ) |                 |
| 1                         | 20                     | 7.91                    | 251.29                    | 186.92          | 30.39                        | 145.64                       | 7.58                        | 36.99                        | 1.003                       | 6.79                            | 67.2            |
| 2                         | 47                     | 7.61                    | 158.12                    | 173.91          | 33.99                        | 114.82                       | 8.27                        | 46.68                        | 1.139                       | 7.88                            | 145.87          |
| 3                         | 72                     | 7.83                    | 83.27                     | 89.83           | 37.97                        | 88.32                        | 8.40                        | 51.13                        | 1.263                       | 7.96                            | 239.28          |
| 方差分析<br>Variance analysis | F<br>Prob>F            | 1.909<br>0.152          | 55.06<br>0.000            | 17.03<br>0.000  | 1.642<br>0.000               | 11.14<br>0.009               | 0.980<br>0.038              | 2.541<br>0.004               | 3.999<br>0.021              | 2.54<br>0.053                   | 13.46<br>0.000  |

分区的有效性,而未考虑这些土壤和产量数据空间变异的时间稳定性或时间变异性信息。另外,管理分区的空间连续性或聚集程度不仅与土壤或作物产量的时间和空间变异程度有关,而且与选择数据源的空间尺度(空间单元大小)有关<sup>[17]</sup>。下一步将根据土壤和作物产量的空间变异性及其时间稳定性来定义管理分区,并尝试在数据的不同空间尺度下,通过聚集度和破碎度等景观生态学指标来对分区结果的空间连续性和聚集程度进行评价,为精确农业管理分区的有效划分和精确管理决策提供手段和途径。

## 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 均值聚类作为模糊分类和模式识别方法,具有聚类信息丰富等特点从而在数据挖掘和图像处理领域被广泛使用。本文利用土壤质量和作物产量的限制因子的空间变异性,采用模糊 c 均值聚类算法来定义田间精确管理分区,并给出模糊性能指数和归一化分类熵有效地确定了最佳聚类效果和分区数目,证实了对海涂区盐碱地进行田间管理分区的可行性。分区结果不但可以代替网格取样成为一种新的采样方式——管理分区采样,而且可以有效地量化土壤特性和作物产量的空间变异性,从而为样区精确农业变量投入的实施提供了基础,为定点和精确的田间管理提供了决策依据。

## References

- [1] 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.
- [2] 赵其国. 我国现代农业发展中的若干问题. *土壤学报*, 1997, 34(1): 1-9.
- Zhao Q G. Problems challenging the development of modern

agriculture of China. *Acta Pedologica Sinica*, 1997, 34(1):1-9. (in Chinese)

- [3] Khosla R, Fleming K, Delgado J A, Shaver T M, Westfall D G. Use of site specific management zones to improve nitrogen management for precision agriculture. *Journal of Soil and Water Conservation*, 2002, 57: 513-518.
- [4] Fraisse C W, Sudduth K A, Kitchen N R. Calibration of the ceres-maize model for simulating site-specific crop development and yield on claypan soils. *Applied Engineering in Agriculture*, 2001, 17: 547-556.
- [5] Long D S, Carlson G R, DeGloria S D. Quality of field management maps. In: Robert P C, Rust R H, Larson W E, ed. *Proceedings of the 2nd International Conference on Site-specific Management for Agricultural Systems*. USA: ASA-CSSA-SSSA, Madison, 1994: 251-271.
- [6] Johnson C K, Mortensen D A, Wienhold B J, Shanahan J F, Doran J W. Site-specific management zones based on soil electrical conductivity in a semiarid cropping system. *Agronomy Journal*, 2003, 95: 303-315.
- [7] Kravchenko A N, Bullock D G. Correlation of corn and soybean grain yield with topography and soil properties. *Agronomy Journal*, 2000, 92: 75-83.
- [8] Franzen D W, Kitchen N R. Developing management zones to target nitrogen applications. SSMG-5. In: *Site-Specific Management Guidelines Series*. Potash & Phosphate Institute, 1999. <http://www.ppi-far.org/ssmg>.
- [9] 自由路, 金继运, 杨俐苹, 梁鸣早. 基于GIS的土壤养分分区管理模型研究. *中国农业科学*, 2001, 34: 46-50.
- Bai Y L, Jing J Y, Yang L P, Liang M Z. Research on the subarea management model of soil nutrients by GIS. *Scientia Agricultura Sinica*, 2001, 34: 46-50. (in Chinese)
- [10] 黄绍文, 金继运, 杨俐苹, 程明芳. 县级区域粮田土壤养分空间变异与分区管理技术研究. *土壤学报*, 2003, 40: 79-88.
- Huang S W, Jing J Y, Yang L P, Cheng M F. Spatial variability and

- regionalized management of soil nutrients in the grain crop region in Yutian County. *Acta Pedologica Sinica*, 2003, 40: 79-88. (in Chinese)
- [11] 李翔, 潘瑜春, 赵春江, 王纪华, 鲍艳松, 王锦地. 基于多年产量数据的精准农业管理分区提取与尺度效应评价. *中国农业科学*, 2005, 38: 1825-1833.
- Li X, Pan Y C, Zhao C J, Wang J H, Bao Y S, Wang J D. Delineation and scale effect of precision agriculture management zones using yield monitor data over four years. *Scientia Agricultura Sinica*, 2005, 38:1825-1833. (in Chinese)
- [12] Stafford J V, Lark R M, Bolam H C. Using yield maps to regionalize fields into potential management units. In: *Proceedings of the Fourth International Conference on Precision Agriculture*, USA: ASA-CSSA-SSSA, Madison, 1998: 225-237.
- [13] Boydell B, McBratney A B. Identifying potential within-field management zones from cotton yield estimates. In: *Proceedings of the Second European Conference on Precision Agriculture on Precision Agriculture*, Denmark: Odense Congress Cent, 1999: 331-341.
- [14] Jaynes D B, Kaspar T C, Colvin T S, James D E. Cluster analysis of spatiotemporal corn yield patterns in an Iowa field. *Agronomy Journal*, 2003, 95: 574-586.
- [15] Fridgen J J, Kitchen N R, Sudduth K A, Drummond S T, Wiebold W J, Fraisse C W. Management zone analyst (MZA): software for subfield management zone delineation. *Agronomy Journal*, 2004, 96: 100-108.
- [16] 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)
- [17] 朱小鸽. 多重主成分分析及在地质构造信息提取中的应用. *遥感学报*, 2000, 4(4): 299-303.
- Zhu X G. Extracting geological structure information by multi-principal component analysis. *Journal of Remote Sensing*, 2000, 4(4): 299-303. (in Chinese)
- [18] Bezdek J C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press, 1981.
- [19] 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)
- [20] 傅庆林, 厉仁安, 葛正豹. 浙江省海涂农业科技示范园区建设研究与实践. 杭州: 浙江大学出版社, 2000: 102-104, 122-123.
- Fu Q L, Li R A, Ge Z B. *Study and Practice on Agricultural Technology Demonstration in Coastal Saline Land in Zhejiang Province*. Hangzhou: Zhejiang University Press, 2000: 102-104, 122-123. (in Chinese)
- [21] 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 minimise sampling density. *The Science of the Total Environment*, 1998, 221: 43-57.

(责任编辑 李云霞)