

ESDA 支持下的城市地价分布信息提取

陈思源¹, 曲福田¹, 曹大贵², 李典友³

(1. 南京农业大学中国土地问题研究中心,南京 210095 ; 2. 南京大学城市与资源学系,南京 210093 ; 3. 南京农业大学资源与环境科学学院,南京 210095)

摘要 : 在城市地价分布信息提取中,因地价信息空间位置的相关性和样本数量的有限性,传统统计学的两个基本假设(样本独立与大样本)不易实现。为此,采用探索性空间数据分析(ESDA)技术,对地价数据作空间统计分析。在ESDA分析中,数据结构必须满足正态分布,同时,要对全局趋势分析和方向分析等予以考虑,才能得到误差较小的分析结果。镇江市实例研究表明,该地区地价样点的空间数据结构呈对数分布,存在二阶趋势和北北东方向,合理步长为375 m。

关键词 : 地理信息系统 ; 空间数据 ; 克力格插值 ; 空间统计学 ; 镇江市

中图分类号 : F 301.3 文献标识码 : A 文章编号 : 1001 - 070X(2006)03 - 0047 - 04

0 引言

土地已经作为社会主义市场经济重要的宏观调控手段,价格杠杆调控是城市土地资源优化配置的重要组成。城市地价分布信息的提取,可以量化反映城市地价信息,及时掌握城市地价的动态变化。

对于地价分布信息提取而言,传统统计学成立的两个基本假设(样本独立与大样本)很难满足^[1]。针对上述问题,新的研究思路是采用探索性空间数据分析(Exploratory Spatial Data Analysis,简称ESDA)技术研究地价。ESDA是指利用空间统计学原理和图形图表相结合对空间信息的性质进行分析、鉴别,用以引导确定性模型的结构和解法。ESDA分析对数据数量要求不高,但对数据结构要求严格,对数据相关性和模型适用范围也有具体规定。随着GIS空间统计分析技术的发展,ESDA技术也日益受到关注^[2,3]。然而,迄今为止的ESDA相关研究多侧重于理论探讨,实际运用也限于气候、环境污染等方面,将此技术用在地价研究的并不多见。据此,本研究从ESDA的数学建模出发,讨论其分析方法和步骤,并结合镇江市研究实例,探讨运用于地价信息提取,为今后此类研究提供借鉴。

1 ESDA 建模

ESDA是一种严格基于“数据驱动”的分析方法,实质上是一种“让数据说明其本身(Let the data speak for themselves)”的分析技术。它包括一整套分析技术和分析工具,如各种克力格(Kriging)内插、变率函数、多变量空间统计、模糊空间统计和优化采样设计等,并有统一的理论基础^[1]。

普通克力格内插(Ordinary Kriging)是ESDA较为常用的一种方法,它将空间数据分解为一般趋势和局部变化两个部分,分别称之为数据的平稳特性(Smooth)和非平稳特性(Rough)^[4]。建立模型 $Z(s) = \mu + \varepsilon(s)$,式中 $s = (x, y)$,即坐标位置; μ 为常均值(常期望); $\varepsilon(s)$ 为随机误差。考虑观测点的随机误差 $\varepsilon(s)$ 的内蕴平稳性,上式可简写为 $Z(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$,式中 $Z(s_i)$ 为第*i*点的观测值; λ_i 为第*i*点观测值的未知权重; s_0 是观测点的位置。

为了满足对空间数据相关性和求解,给出2个基本假设:①任意距离*h*两点间差值的数学期望为零;②任意距离*h*两点间差值的方差最小。对第1个假设进行数学转换,进行无偏内插估计,最小化统

计期望值为 $[Z(s_0) - \sum_{i=1}^n \lambda_i Z(s_i)]^2$, 获得求解。为满足第 2 个假设, 建立模型

$$\Gamma \times \lambda = g \quad (1)$$

式中 Γ 为所有样点对之间的半变异值矩阵; λ 为样点观测值的权重系数向量。通过构建理论半变异/协方差函数, 采用最小二乘法进行拟合, 获得半变异/协方差函数与样点对间距离的拟合曲线模型。根据模型, 对上述矩阵 Γ 和权重系数向量 λ 进行求解, 进而通过系数加权获得预测值向量 g , 实现了第 2 个假设的求解。

上述建模过程, 充分考虑了样点的空间相关性, 同时也获得预测点的数学预测值。

2 ESDA 城市地价分布信息提取方法

如前所述, ESDA 建立在一定的假设求解上, 对分析的存在条件有着特殊的要求, 并具有独特的研究方法^[5]。要采用 ESDA 提取地价分布信息, 必须运用这些方法, 才使得预测结果的误差较小。探索性的分析方法包括数据结构分析、趋势分析及方向分析等, 以下分别讨论。

2.1 数据结构分析

运用 ESDA 提取地价信息时, 首先, 要分析地价的数据结构。如果数据结构(或经过转换)为正态分布, 可以进行空间插值分析; 反之, 则不适合。

要分析地价空间数据是否符合正态分布, 可以从 2 方面来判断: ①通过直方图(Histogram)判断, 如果直方图的均值和中值近于相等, 则为正态分布; 反之, 为偏态分布; ②通过正态 QQ-Plot 分布图判断, 如果正态 QQ-Plot 分布图曲线趋于一条直线, 则符合正态分布。

2.2 全局趋势分析

一般而言, 因城市商业中心(CBD)的存在, 城市地价从城市中心向四周递减^[6]。将地价数值投影到与地图平面正交的平面上, 用一个多项式来拟合, 则会出现地价的空间分布在某一方向确定的形(二次函数曲线、指数函数曲线等), 这种函数的形被称为地价分布的全局趋势。在 ArcMap 视图中, 可将地价信息转化为以属性值为高度的三维曲面(图 1), 获得全局趋势。

在局部变异分析中, 对全局趋势进行剔除, 该过程称为趋势分离(Detrending), 其目的是为了在局部趋势分析中少受全局趋势影响。通过全局趋势分析

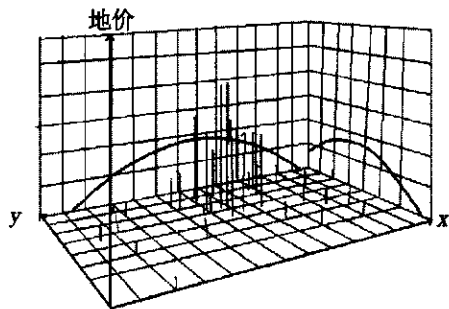


图 1 样点地价趋势分析

和趋势分离, 地价样点的空间相关性被很好考虑。

2.3 方向分析

与空间趋势类似, 地价分布在平面上也有一个趋势, 存在各向异性(Anisotropy)。如, 某一城市内由于沿南北向贯穿的交通干线存在, 使得该城市的地价分布明显呈南北方向延伸。地价分布各向异性的产生, 可能是商服繁华、交通便捷度及绿化等原因。

在提取地价信息时, 必须考虑各向异性。在 ESDA 中, 通过方向搜索(Search Direction)设置角度和搜索带宽, 进行方向分析, 量化和计算各向异性。

2.4 离群值查找

离群值是统计异常值在 ArcMap 视图中的图形表达。例如, 在地价统计中, 50 个样点数据有 49 个样点的值介于 200 ~ 300 元/m² 之间, 而另外一个点的值为 800 元/m², 则这个点就为全局离群值。离群值可能是因瞒价交易、年期不同、修正不当、记录错误等引起的异常, 也可能是客观存在的真实异常。

提取地价分布信息时, 可以通过图形数据库和属性数据库的关联, 查找离群值, 进而确认异常原因。如果确认是真实异常, 则不需剔除, 反之, 需要对样点进行剔除或重新调查。

2.5 步长确定

计算地价样点的半变异/协方差函数时, 需要对每一样点与其它样点间的距离和地价差进行曲线拟合, 如果数据集很大, 样点对将迅速增加, 因此, 将样点对按一定间距(步长)进行分组, 确定组别, 计算每一组的平均半变异值。

一般而言, 合理的步长分组规则如下: 如果是规则取样, 则格网的间距为步长大小; 如果是不规则取样, 则满足步长 × 步长数 = 样点间最大间距的 0.5 倍。

3 实例分析

以镇江市市区商业地价分布信息为对象, 对上述 ESDA 分析方法进行实例研究。在 Arc/Info 8.0

下,采用普通克力格插值,通过逐步比较,利用不同方法提取地价空间分布信息,通过误差分析,确定最佳分析方法。

3.1 分析步骤

结合“镇江市城市地价动态监测体系建设”项目,获得市场交易样点,共计 44 个,对其进行价格修正^[8,9],得到统一设定地价内涵条件下的地价,样点设立如表 1 所示,样点分布如图 2 所示。

表 1 镇江市商业用地样点设立

样点编号	所在区段	所在土地级别	交易价格或评估价格(元/m ²)	设定地价内涵条件下的地价(元/m ²)
S01	Sq03	1	1 945.84	1 869
S02	Sq02	1	2 754.91	2 864
⋮	⋮	⋮	⋮	⋮
S42	Sq13	5	507.89	505
S43	Sq15	6	349.89	357
S44	Sq15	6	339.62	332

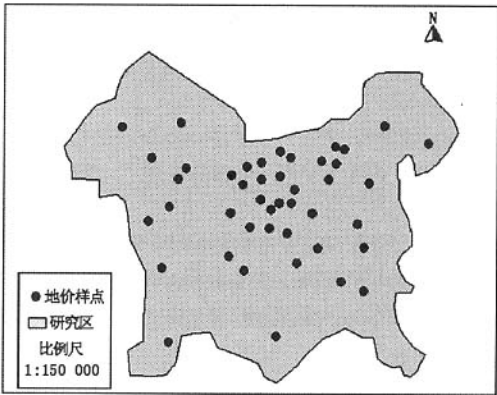


图 2 镇江市商业地价样点分布

首先,分析数据结构,由于不经转换的 QQ - Plot 图为曲线(图 3),要对其进行数据转换才能符合克力

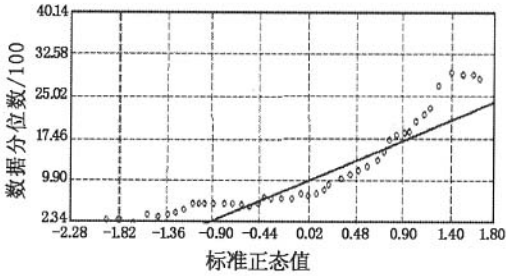


图 3 未经转换的 QQ - Plot 分布

格插值的要求。转换后的 QQ - Plot 分布见图 4。显然,对数转换后,样点概率分布趋于直线,基本呈正态分布。

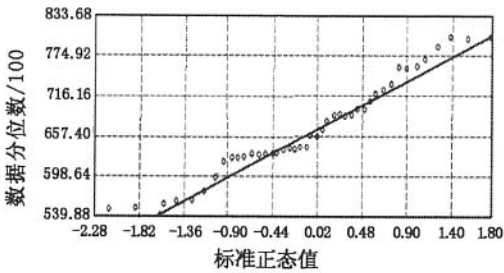


图 4 对数转换后的 QQ - Plot 分布

其次,进行趋势分析。分析结果显示,研究区样点地价存在东西和南北方向的二阶函数趋势,进行二阶函数趋势剔除,并比较趋势剔除前后的插值结果。采用类似步骤,对方法一逐步进行改进,即在方法一的基础上,依次进行数据转换得方法二,进行趋势分析得方法三,进行方向分析(方向角为北偏东 30°、椭圆四分搜索)得方法四(半变异函数如图 5 所示)。

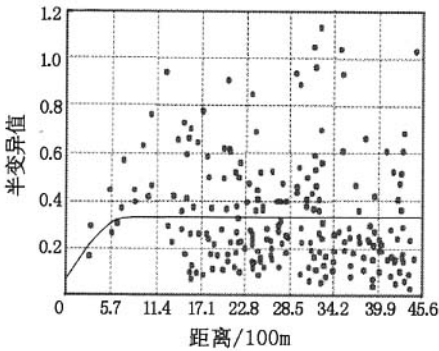


图 5 方法四的半变异函数云图

为了验证步长和分组的合理性,进行分组数相同而步长不同的比较。方法一到方法四采用的分组数是 10,步长为 375 m。在方法四的基础上,对样点数据进行分组数为 10,步长为 750 m 的插值,得方法五(表 2)。需要指出的是,由于本次探索性空间数据较少,地价为实际市场交易地价,且每一数据都进行了人工检验,所以,没有进行离群值查找。如果数据较多,要想获得较为准确的分析结果,必须进行离群值查找,以避免数据统计错误。

表 2 ESDA 的不同分析方法比较

方法	数据转换	趋势分析	方向分析	步长设定及分组		误差分析	
				步长/m	分组数	平均标准差	均方根标准误差
方法一	无	无	无	375	10	0.030	0.624
方法二	对数转换	无	无	375	10	0.078	1.108
方法三	对数转换	二阶趋势剔除	无	375	10	-0.103	1.150
方法四	对数转换	二阶趋势剔除	东北北方向	375	10	-0.036	1.003
方法五	对数转换	二阶趋势剔除	东北北方向	750	10	0.339	0.942

3.2 误差分析

对上述 5 种方法的误差进行比较,选用两个误差分析指标——平均预测误差和均方根标准误差,对比其插值精度(表 2)。根据地统计学原理,平均预测误差趋于 0、均方根标准误差趋于 1 时,插值结果满足两个基本假设,即插值结果可信度较高。

没经转换的插值分析(方法一),其平均预测误差为 0.030,均方根标准误差为 0.624,插值效果不理想。对数转换后(方法二),其平均预测的误差为 0.078,均方根标准误差为 1.108。可见,对数转换后,插值效果较好。方法一的两项误差指标都是最大,也就是说,方法一的插值结果最不理想,其原因显然与其数据结构不符合正态分布也没进行数据转换有关。

比较上述 5 种方法,可以发现,方法四的平均预测误差为 -0.036,最接近 0;均方根标准误差为 1.003,最接近 1,显然其插值误差最小,插值结果可信度高。同时可见,虽然方法五和方法四有着相同的数据转换、趋势分析、方向分析,但由于步长不同,方法五插值误差较方法四的大,插值结果不理想。综上所述,本研究的方法四为最佳方法。

4 结论

(1)ESDA 方法是地价分布信息提取的有效手段,该方法将样本数量的要求和地价样本的空间关联考虑,普通数学统计方法难以实现。

(2)ESDA 建立于一定的数学分析基础之上,而这些数学分析又由一定的假设组成,对这些假设的数学求解,是 ESDA 的理论基础,也是结果检验的重要手段。

(3)ESDA 分析方法包括空间数据结构分析、全局趋势分析、方向分析及离群值查找等。而数据结构分析最为重要,只有数据满足正态分布才能进行空间插值分析,否则,必须经过转换才能进行。在插值分析过程中,对于全局趋势分析和方向分析等方面要予以适当考虑,才能较好地提取地价的空间分布信息。

(4)在地价空间分布信息的提取中,通过 ESDA 分析,合理确定探索性分析的分析参数,插值结果的误差较小,可以得到较为满意的分析结果;反之,误差较大,可信度低。

参考文献

- [1] 柏延成,李新,冯学智. 空间数据分析与空间模型[J]. 地理研究,1999,18(2):185-187.
- [2] 李德仁,王树良,史文中,等. 论空间数据挖掘和知识发现[J]. 武汉大学学报(信息科学版),2000,12(6):492-495.
- [3] 李德仁,王树良,李德毅,等. 论空间数据挖掘和知识发现的理论与方法[J]. 武汉大学学报(信息科学版),2002,27(3):221-222.
- [4] 郭达志,胡召玲,陈云浩. GIS 中空间对象的不确定性研究[J]. 中国矿业大学学报(自然科学版),2000,29(1):20-24.
- [5] 马洪超,李德仁. 基于空间统计学的空间数据窗口大小的确定[J]. 武汉大学学报(信息科学版),2001,26(1):18-19.
- [6] 杨山,沈宁泽. 基于遥感技术的无锡市城镇形态分形研究[J]. 国土资源遥感,2003(3):41-44.
- [7] 张景雄, Roger P Kirby. 模糊类别制图的空间统计学方法[J]. 测绘学报,1998,27(2):167-168.
- [8] 国土资源部. GB/T 18508-2001 城镇土地估价规程[S]. 北京:中国标准出版社,2002.
- [9] 杜小娅,陆跃进. 南京市地价空间分布及其影响因素分析[J]. 国土资源遥感,2004(2):51-55.

A TENTATIVE MEASUREMENT STUDY OF GIS EXPLORATORY SPATIAL DATA ANALYSIS IN URBAN LAND DISTRIBUTION INFORMATION EXTRACTION

CHEN Si-yuan¹, QU Fu-tian¹, CAO Da-gui², LI Dian-you³

(1. China Center for Land Policy Research, Nanjing Agriculture University, Nanjing 210095, China; 2. Department of Urban and Resource Sciences, Nanjing University, Nanjing 210093, China; 3. College of Resources and Environmental Sciences, Nanjing Agricultural University, Nanjing 210095, China)

Abstract: The measurement of GIS Exploratory Spatial Data Analysis in the land price distribution information extraction was study in this paper. As it is very difficult to satisfy the two premises in the traditional statistics for land price samples, other methods should be sought for. The Exploratory Spatial Data Analysis (ESDA) has been set up on the mathematic basis, which has some spatial methods for land price distribution. Besides, a few samples should

for Land Cover Classification[J]. IEEE Transactions on Geoscience and Remote Sensing ,1995 ,33(1) :59 - 66.

[10] 李 爽 ,张二勋. 基于决策树的遥感影像分类方法研究[J]. 地域研究与开发 ,2003 ,22(1) :17 - 21.

[11] Yang Chun - chieh , Shiv O Prasher , Peter Enright , et al. Application of decision tree technology for image classification using remote sensing data[J]. Agriculture Systems ,2003 (76) :1101 - 1117.

[12] 陈松林. 基于 GIS 的土壤侵蚀与土地利用关系研究[J]. 福建师范大学学报(自然科学版) 2002 ,16(1) :106 - 109.

[13] 聂 芹 ,李连运. 多步骤分类法在上海市闵行区土地利用/覆盖信息提取中的应用[J]. 国土资源遥感 ,2006 (1) :79 - 83.

[14] 曹文洪. 土壤侵蚀的坡度界限研究[J]. 水土保持通报 ,1993 ,13 (40) :1 - 5.

[15] 靳长兴. 论坡面侵蚀的临界坡度[J]. 地理学报 ,1995 ,50(3) :234 - 239.

[16] 胡世雄 ,靳长兴. 坡面土壤侵蚀临界坡度问题的理论与实验研究[J]. 地理学报 ,1999 ,54(4) :347 - 356.

[17] 王秀英 ,曹文洪 ,陈 东. 土壤侵蚀与地表坡度关系研究[J]. 泥沙研究 ,1998 ,21(2) :36 - 41.

RESEARCHES ON METHODS FOR LAND USE
INFORMATION EXTRACTION IN COMPLICATED
TERRAIN AREAS ON LOESS PLATRAU

LIANG Wei¹ , YANG Qin - ke¹ , LIU Yong - mei^{1 2}

(1. Northwestern Science and Technology University of Agriculture and Forestry ,Research Center of Soil and Water Conservation of CAS and Ministry of Water Resources ,Yangling 712100 , China ; 2. Department of Urban and Resource Sciences , Northwest University ,Xi 'an 710069 , China)

Abstract :Soil erosion is one of the important environment problems ,and land use classification is an important process in this aspect. The normal automatic classification (supervised classification and unsupervised classification) based on spectral characteristics cannot meet the accuracy needed. Therefore ,the slope (produced by DEM) which affects the land use type location and the NDVI (produced by TM images) which reflects the vegetation coverage sensitivity should be taken into account. With Yanhe basin as the study area ,the authors carried out the research on methods for extracting the land use information in complicated areas on the loess plateau. The result indicates that the division of the image into several parts according to the TM image characteristics and the extraction of the land use type one by one assisted by the slope are suitable for the complicated terrain area. The result obtained serves as an important reference to the remote sensing classification technical system.

Key words :Remote sensing ; Slope ; Land use ; Yanhe basin ; Area division ; Multi - stage

第一作者简介 :梁 伟 (1976 -) ,女 ,硕士 ,主要从事遥感与地理信息系统在水土保持中的应用研究。

(责任编辑 :肖继春)

=====

(上接第 50 页)

be needed and the spatial relationship has been considered. It can therefore be used to study the land price distribution. The actuality research of ESDA was expounded in a case of urban commercial land price by GIS Arc / Info 8.0 in this paper. First ,its mathematic basis was described. Then the measurements of GIS Exploratory spatial Data Analysis were studied ,which included the data structure research ,full trend analysis ,searching direction and so on. In the end ,the GIS Exploratory Spatial Data Analysis measurement was investigated with the Kriging Analysis as an example. The error analysis indicates that ,if the frequency distribution of the samples assumes Normal Q - Q Plots and the measurements are used properly ,the result is acceptable.

Key words :GIS ; Spatial data ; Kriging analysis ; Spatial statistics ; Zhenjiang City

第一作者简介 :陈思源 (1972 -) ,男 ,博士生 ,从事土地资源、地理信息系统等方面研究。

(责任编辑 :肖继春)