

基于 GIS 与 Surfer 的甘肃省人口趋势面分析^①

Trend Surface Analysis of Population Density of Gansu Based on GIS and Surfer

朱瑜馨¹ 张锦宗^{1,2} (1 聊城大学环境与规划学院 山东 聊城 252059;
2 北京师范大学地理与遥感学院 北京 100875)

摘要: 借助 MapInfo 提供的空间分析功能和 Surfer7.0 格网文件的数据内插方法,根据空间的采样数据可以拟合一个人口密度的数学曲面,可以模拟人口在空间上的分布及变化趋势。文章选择 Quadratic surface 回归类型确定趋势面方程,利用 2006 年甘肃省人口与计划生育网站提供的人口及几何数据,建立了甘肃省地市级人口密度二阶空间趋势面模型,并制作了趋势等值线图与剩余值散点图。分析表明:甘肃省的人口高密度区在平凉、天水、临夏、兰州围成的西北-东南走向的四边形区域。以此区域为核心远端向西北、毗邻向东北和西南两侧密度递减。南北两侧的密度变化明显;西北部地区密度梯度较小,人口密度最低。模型的拟合程度较高,可以应用于其它地区实践之中。

关键词: 甘肃省 人口趋势面 人口密度 Surfer GIS

人口在一定的地域环境下,呈现一定的空间分布形态。人口空间分布是指一定时间内人口在一定地区范围的空间分布状况,它是人口过程在空间上的表现形式^[1]。受自然环境、生产力水平、生产布局特点、历史、社会经济和政治等因素的综合制约。

趋势面分析,是通过回归分析原理(在多要素的地理系统中,因变量 Z 受 K 个自变量 x_1, x_2, \dots, x_k 的影响,根据其 n 组观测值 $(Z_i, x_{i1}, x_{i2}, \dots, x_{ik})$ 建立回归方程),运用最小二乘法,拟合一个二维非线性函数,展示地理要素在地域空间上的变化趋势^[2,3]。文章以甘肃省各地级市为例,选择区域几何中心、人口密度等指标,借助 Surfer 及 GIS 软件建立 2 阶趋势面模型进行人口分布的空间拟合。

1 趋势面分析方法的基本原理

1.1 趋势面分析方法过程

空间趋势面是一个模拟地理要素空间分布的近似曲面。通常把实际的地理曲面分解为趋势面和剩余面两部分,趋势面反映地理要素的宏观分布规律,剩余面对应于微观局部,是随机因素影响的结果。

1.1.1 趋势面模型的建立

设 $z_i(x_i, y_i)$ ($i=1, 2, \dots, n$) 为某地理要素的实际观测数据,趋势值拟合值为 $\hat{z}_i(x_i, y_i)$, 则有 $z_i(x_i, y_i) = \hat{z}_i(x_i, y_i) + \bar{z}_i$, 式中: \bar{z}_i 为剩余值(残差值)。

用来建立趋势面的数学方程有多项式函数和傅立叶级数,最常用的是多项式函数形式,通过调整多项式的次数,可使所求的回归方程适合实际问题的需要。文章选择多项式函数作为回归方程。

在实际应用中,应先考虑用一次趋势面模型拟合,再用二次趋势面模型拟合,如果还不能满足研究需求,则需选用三次趋势面、四次趋势面甚至更高次趋势面进行拟合。

1.1.2 趋势面模型的参数估计

趋势面模型的参数估计就是根据实际观测值 $z_i(x_i, y_i)$ ($i=1, 2, \dots, n$) 确定多项式的系数 a_0, a_1, \dots, a_p , 使剩余平方和最小。根据实际观测值 $z_i(x_i, y_i)$ ($i=1, 2, \dots, n$) 用回归分析方法求得的回归曲面 $\hat{z} = f(x, y)$, 对应于回归曲面的值 \hat{z}_i 作为趋势值,以残

① 基金项目:聊城大学青年基金项目(X071015)

差 $\bar{z}_i = z_i - \hat{z}_i$ 作为剩余值。

1.1.3 趋势面模型的显著性检验

不同次数的趋势面方程对原始数据的拟合程度是不一样的,但趋势面分析的核心是从实际观测值出发推算趋势面,一般采用回归分析方法,使得剩余平方和 $s_s = \sum_{i=1}^n (z_i - \hat{z}_i)^2 \rightarrow \min$ 估计趋势面参数。

(1) F - 分布检验

检验统计量为:

$$F = \frac{U/P}{Q/(n-p-1)}$$

式中 U 为回归平方和, Q 为剩余平方和, P 为多项式的项数(不含常数项), n 为观测点数。在给定置信水平 α 的条件下,若 $F > F_\alpha$, 则趋势面拟合效果显著, 否则不显著。

(2) 拟合度指数公式检验

拟合度可以用公式(1-6)进行检验^[4]:

$$c = \frac{\sum_{i=1}^n (\hat{z}_i - \bar{z}_i)^2}{\sum_{i=1}^n (z_i - \bar{z}_i)^2} \times 100\%$$

一般而言, $C \geq 80\% \sim 85\%$ 认为可以满足要求。本文采用拟合度指数进行显著性检验。

2 基于 GIS 与 Surfer 的趋势面分析方法实现

2.1 GIS 数据准备

以甘肃省 1:5000000 政区图为基础资料,在 MapInfo 中配准,建立各个地级市多边形矢量数据图层,借助 MapInfo 提供的空间分析功能提取多边形的几何中心坐标(X,Y)作为自变量,选择 2006 年年底甘肃省行政区划人口密度数据为因变量,建立趋势面方程。

2.2 Surfer 软件实现

2.2.1 趋势值提取

在 Surfer7.0 环境下,Grid 菜单下的 Data 命令,提供了根据离散数据建立格网文件的数据内插方法,"Polynomial Regression" 方法可根据空间的采样数据拟合一个数学曲面,该内插方法实际上是一个趋势面分析过程。该方法又提供了多种回归类型: Simple planar surf、Bi - linear saddle、Quadratic surface、Cubic surface、User defined polynomial。文章通过对各种回

归类型拟合程度的比较计算,最终选择 Quadratic surface 回归类型确定趋势面方程,同时建立格网数据文件,以便在此基础上生成趋势等值线图和空间分布趋势图。

2.2.2 剩余值计算

剩余值是实际观测值与趋势值之差。选择 Grid 菜单下的 Residuals 命令,首先选择人口分布的格网文件(*.grd),然后打开相应的 dat 文件,分别指定文件中的 X,Y,Z 列,计算后的剩余值保存在相应的 dat 文件最后一列。

2.2.3 制作趋势面与剩余面等值线图

首先对生成的网格化文件(*.grd)进行圆滑, Suffer 中提供了矩阵圆滑和样条圆滑方法。一般等值线图用样条圆滑取得的圆滑效果较好。然后通过 Map 菜单下的 Contour Map 的 New Contour Map 命令或工具栏中的 Contour Map 打开圆滑后网格文件(out.grd),设置等值线间隔,并根据需要设置等值线标签开关、颜色填充等参数,即可绘制等值线图^[5]。

3 甘肃省人口趋势面分析

甘肃省位于我国的地理中心,地处黄河上游, $32^{\circ}11'N \sim 42^{\circ}57'N$ 、 $92^{\circ}13'E \sim 108^{\circ}46'E$,地域辽阔。全省地貌复杂多样,地势自西南向东北倾斜,地形呈狭长状。全省总面积 454236.37KM²,共包括 14 个地级市^[6]。由于嘉峪关市为省直辖市,腹地极小,作为飞地被包围在酒泉市域之内,故本文将二者合并为一研究。

3.1 数据预处理

文章的空间数据在 MapInfo 环境下矢量完成,建立 Region.tab,并通过表/转出菜单将 Region.tab 转换为 Region.mif 交换格式文件,以便在 Sufer 软件中打开。人口数据,最小统计单位为县(市),数据存储为 Micorsoft Visual FoxPox 6.0 的.dbf 格式。

3.1.1 人口统计数据处理。由统计资料的人口普查区的人口数据及地市的行政区面积得到人口普查区的人口密度^[5];

3.1.2 空间数据的处理。以甘肃省 1:5000000 政区划图为底图,通过 MapInfo 提供的 centroidX 与 centroidY

函数得到各地市的几何中心坐标(表1)。

3.1.3 将 Region. tab 对应的属性数据通过表/转出菜单转换为 Region. dbf, 并在 Excel 中打开保存为 Region. xls, 在 Sufer 中作为离散数据内插生成. grd 文件, 再生成空间分布趋势图。

表1 甘肃省地市几何中心与人口密度

地市	X(经度)	Y(纬度)	人口密度 (人/KM2)
兰州	103.658	36.103	232.019
金昌	101.942	38.273	47.110
白银	104.554	36.373	81.423
天水	105.822	34.533	223.909
武威	102.873	37.780	55.266
定西	104.615	34.920	138.690
陇南	105.502	33.416	92.630
平凉	106.781	35.274	181.791
庆阳	107.629	36.203	89.109
临夏	103.269	35.409	231.680
甘南	102.809	34.116	16.636
张掖	99.736	38.442	30.964
酒泉	96.222	40.172	5.87

3.2 人口趋势面模型的建立

趋势面方程对观测值的拟合程度越好,趋势面拟合效果也就越好。建立人口趋势面模型,首先应根据趋势面方程显著性检验结果、拟合优度等指标选择合适的趋势面模型的阶次。

3.2.1 趋势面函数的拟合及显著性检验

结合甘肃省人口地理分布特征以及文章采用的因变量和自变量,先后以县市和地市为单位分阶次进行比较分析,结果表明以地市为单位建立的二阶趋势面模型拟合程度高。趋势面方程为:

$$z = -2110.91.91 + 5801.86y - 39.74y^2 + 2068.06x - 28.58xy - 5.01x^2$$

将实际观测值代入模型中得到甘肃省13地市人口空间分布的趋势值和剩余值(表2)。

显著性检验 C 值为 83.2995%。满足 ≥ 80% ~ 85% 的要求范围,表明建立的方程能够拟合甘肃省13地市人口密度的空间分布格局。

表2 13 地市人口趋势面分布的趋势值和剩余值(人/KM2)

地市	观测值	趋势值	剩余值
兰州	232.019	156.2917	75.72725
金昌	47.110	-0.45136	47.56136
白银	81.423	132.1519	-50.7289
天水	223.909	174.0839	49.82511
武威	55.266	30.39873	24.86727
定西	138.690	161.8009	-23.1109
陇南	92.630	70.50536	22.12464
平凉	181.791	171.5584	10.23261
庆阳	89.109	52.93062	36.17838
临夏	231.680	150.5435	81.1365
甘南	16.636	9.531027	7.104973
张掖	30.964	62.94676	-31.9828
酒泉	5.87	-18.6666	24.53664

3.2.2 人口趋势面分析

① 趋势分析

根据趋势面模型计算的趋势值与剩余值,通过 Map 菜单下的 Wireframe 或工具栏中的 Wirefram Map 制作趋势图,再利用 Map 菜单下的 Base Map 叠加 Region. mif, 作出甘肃省人口空间分布趋势图(图1)。

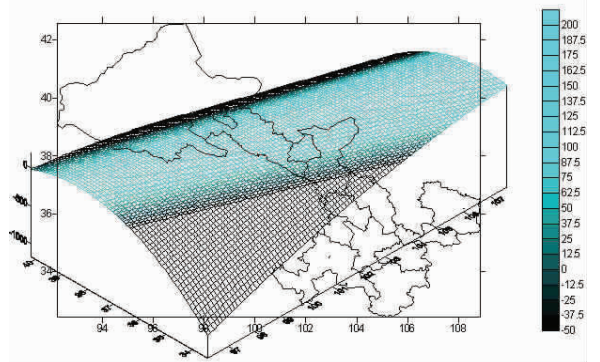


图1 人口空间分布趋势面立体图

由图1分析可知:甘肃省的人口密度高地在平凉、天水、临夏、兰州围成的西北-东南走向的四边形区域。以此区域为核心远端向西北、毗邻向东北和西南两侧密度递减。南北两侧的密度变化明显;西北部地区密度梯度较小,人口密度最低。

② 剩余值分析

剩余值是指人口密度的实际观测值与趋势值之差,反映了扣除地理因素作用后局部综合因素作用下

人口密度的变化情况。以实际观测值为横坐标, 剩余值为纵坐标做散点图(图 2), 图中各点随机分布, 说明该模型拟合程度好。

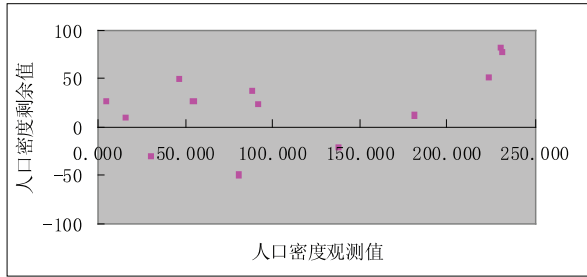


图 2 人口密度趋势面剩余值散点图

根据已得到的剩余值就可以分析自相关剩余区域, 首先确定剩余值的上限(1)和下限(2)^[7]。

$$L(+) = \frac{\sum_{i=1}^n \bar{z}_i(+)}{k} \quad (1) \quad L(-) = \frac{\sum_{i=1}^n \bar{z}_i(-)}{k} \quad (2)$$

式中 $k = \frac{n}{x} (x=1, 2, 3, \dots, m)$ 。

正自相关剩余区域说明人口密度偏高, 负自相关剩余区域则说明人口密度偏低。将表 2 中的人口密度剩余值分为正负两组, 其中正剩余值为 10 个, 负剩余值为 3 个, 由公式 1 和 2 计算得到 $L(+) = 58.353$, $L(-) = -16.2804$ (确定 k 时, $x=2$)。在 10 个正剩余区域中, 有 2 个区域的人口密度高于 $L(+)$, 分别是兰州和临夏。在 3 个负剩余区域中, 3 个区域的人口密度均低于 $L(-)$, 分别是白银、定西和张掖。

笔者分析后发现, 影响现状人口分布的因素主要有两个: 地区耕地面积占地区国土面积的比例, 地区地均 GDP 的数值。其相关分析和回归分析的结果如表 3:

表 3 耕地比例、地均 GDP 与人口密度的回归分析

	回归方程	拟合度 (R2)	相关系数 (R)
密度与耕地比例	$D = 22.383 + 6.291L$	0.632	0.795
密度与地均 GDP	$D = 72.537 + 0.391E$	0.313	0.559
密度与经济、地均 GDP	$D = 4.447 + 5.599L + 0.285E$	0.79	0.889

注: D 表示人口密度; L 表示耕地占国土面积比例, 取值范围为 0 - 30; E 表示地均 GDP, 取值范围为 0 - 500(万元)。

结果显示, 影响地区人口密度的因素地区耕地面积占地区国土面积的比例和地区地均 GDP, 两者均不同程度的与人口密度成正相关。在不同地区耕地面积比例和地均 GDP 起作用大小不同, 耕地面积比例相对于地均 GDP 的重要性依次是兰州(0.67)、金昌(0.6)、白银(3.4)、天水(4.6)、酒泉(1.8)、张掖(1.78)、武威(2.65)、定西(12.8)、陇南(6.2)、平凉(6.17)、庆阳(5.15)、临夏(4.59)、甘南(4.5)。总体来看, 确定人口密度的格局决定性因素是地区耕地面积占地区国土面积的比例, 而地均 GDP 可以对人口密度起二次调整作用。这种状况与甘肃省经济发展水平低下的现实吻合, 随着经济的进一步发展, GDP 对人口密度的二次调整作用会逐渐加强。

兰州和临夏的人口密度高于 $L(+)$, 兰州是全省唯一特大城市, 经济发达, 工作生活条件优越, 吸引了大量人口, 因此人口密度偏高; 临夏则是历史原因造成的少数民族集聚的结果。白银、定西、张掖的人口密度低于 $L(-)$, 白银耕地比例和地均 GDP 都偏低, 又临近密度高地, 故显得人口密度偏低。张掖的耕地比例和地均 GDP 都偏低, 又相对临近密度高地, 也显得人口密度偏低。定西地均 GDP 相对于临近地区过于偏低, 因此使得人口密度偏低。

4 结论

趋势面分析从整体出发, 以数学模型来模拟(或拟合)地理数据的空间分布及其区域变化趋势, 将与空间地理坐标相关的数据分解为趋势部分和剩余部分, 可集中反映地理数据在大范围内的空间变化趋势。文中建立的趋势面方程的 C 值大于满足要求范围, 表明建立的趋势面方程拟合度高; 剩余值散点图中各点随机分布, 也表明该模型拟合程度较好, 较好的模拟和预测了甘肃省人口密度空间分布的总的变化。

一般来说, 人口最稠密地区都是自然条件优越、资源丰富、经济发达、历史悠久的地区。资源是基础支撑, 环境是限制条件, 会造成人口的分布不均^[8]。影响甘肃人口密度的因素是耕地面积比例和地区地均 GDP, 两者均不同程度的与人 (下转第 53 页)

(上接第 68 页)

口密度成正相关。总体来看,确定甘肃人口密度的格局决定性因素是地区耕地面积占地区国土面积的比例,而地均 GDP 可以对人口密度起二次调整作用。这种状况与甘肃省经济发展水平低下的现实吻合,随着经济的进一步发展,GDP 对人口密度的二次调整作用会逐渐加强。

参考文献

- 1 胡焕庸,张善余. 中国人口地理. 上海:华东师范大学出版社,1984:196.
- 2 徐建华. 现代地理学中的数学方法. 北京:高等教育出版社,2002: 47 - 60.
- 3 庄宇,胡晓蕊,马贤娣. 水环境承载力与经济效率的多元回归模型及应用. 干旱区资源与环境,2007,21

(9):41 - 45.

- 4 陈江,韩文喜,高雪梅. 趋势面分析在滑坡防治工程设计中的应用. 水土保持研究,2007,14(6):115 - 119.
- 5 赵荣军,和向丽. SURFER 在地球化学图制图中的应用. 物探与化探,2004,24(2):76 - 78.
- 6 国家统计局农村社会经济调查司. 2006 中国县(市)社会经济统计年鉴. 北京:国家统计局出版社,2006: 163 - 181.
- 7 王晓燕,沈毅,陈坤,等. 趋势面分析法在肺癌死亡率地理分布研究中的应用. 中华流行病学杂志,2007,28(6):608 - 612.
- 8 张志强. 区域 PRED 的系统分析与决策制定方法. 地理研究,1995,14(4):62 - 68.