

引用格式:黄安,许月卿,孙丕苓,等.基于多源数据人口分布空间化研究——以河北省张家口市为例[J].资源科学,2017,39(11):2186-2196. [Huang A, Xu Y Q, Sun P L, et al. Spatial distribution of population specialization based on multi-source data: a case study of Zhangjiakou City[J]. *Resources Science*, 2017, 39(11): 2186-2196.] DOI: 10.18402/resci.2017.11.16

基于多源数据人口分布空间化研究 ——以河北省张家口市为例

黄安^{1,2},许月卿^{1,2},孙丕苓^{1,2},刘超^{1,2},郑伟然^{1,2}

(1. 中国农业大学资源与环境学院, 北京 100193;

2. 国土资源部农用地质量与监控重点实验室, 北京 100193)

摘要:人口数据空间化研究能够发掘和展现人口空间分布信息,有助于为区域可持续发展研究提供数据支持。研究基于GIS与RS技术,运用主成分分析模型整合多源影响因素获取人口空间分布影响贡献系数,遵循以乡镇级行政区域为控制单元,由控制单元内完整土地空间共同分配乡镇统计人口的原则,实现张家口市统计人口数1km×1km格网空间化。结果表明:空间化人口密度分布特征符合张家口人口空间分布的实际情况,并随各影响因素空间分布变化而变化的趋势,与HSI指数、建设用地板显著相关,与其他指标之间存在显著相关;高值区主要集中在桥西区、桥东区和宣化区,以及各乡镇的镇中心,大致分布在与HSI高值区、RDLS低值区和建设用地集中区。以乡镇级尺度为控制单元进行人口空间化,所得空间化结果的精度优于县级控制单元空间化结果,可为多尺度的综合研究提供数据支持。

关键词:GIS;RS;多源数据;人口;空间化;张家口市

DOI: 10.18402/resci.2017.11.16

1 引言

人口数据是表征人类活动最直接的指标之一,是自然科学、社会科学等研究领域的重要数据源。长期以来,以各级行政区划为单元进行统计的人口数据很难与空间格网形式存储的自然、生态数据进行融合、叠加分析,人口统计数据空间化能够有效解决这一问题^[1,2]。人口数据空间化,是基于影响人口空间分布的要素和关系模型,对人口统计数据进行空间离散化,挖掘并展现人口在空间上的分布信息,以便模拟客观世界人口空间地理分布^[3];离散化后的人口数据能真实地刻画人口空间分布,信息容量显著增加。格网化的人口数据能更好地与土地利用、生态、环境、地理要素等空间数据进行融合,为综合定量分析区域自然、社会经济问题提供支撑。

目前,国内外学者对人口空间化进行了大量研究,完成了全球人口数据库(Gridded Population of the World, GPW)^[4]、全球1km分辨率人口数据集LandScan^[5]、中国1995年、2000年、2005年等全国每平方公里格网人口分布数据集^[6]等,这些成果为其他研究领域提供了重要的基础数据源。随着研究不断深入^[7],人口空间化从单因素影响研究^[8-11]逐渐过渡到多因素集成研究,主要影响因素涉及土地利用^[11]、DMPS/OLS夜间灯光数据^[8-10]、交通、地形^[12,13]等;研究方法方面从传统的多元回归^[4]、土地利用类型法^[11]等逐渐发展为多源数据融合模型^[14]、核密度估算模型^[15]、作息空间法^[16]等。近年来,随着GIS (Geography Information System) 与RS (Remote Sensing) 技术的快速发展,为人口空间化提供了有

收稿日期:2016-10-27;修订日期:2017-08-25

基金项目:国家自然科学基金项目(41571087)。

作者简介:黄安,男,四川雅安人,博士生,主要研究方向为土地资源可持续利用。E-mail: hanner@163.com

通讯作者:许月卿, E-mail: xmoonq@sina.com

2017年11月

效工具^[17]。基于GIS与RS技术,选择科学合理的方法并结合多源因素进行人口统计数据空间化,成为当今国际学术前沿^[11,14,18,19]。在空间化方案设计研究方面主要形成以下两种思路:①以县域行政区域为单元,利用多元回归、地理加权回归等模型,综合计算各种因素对人口空间分布的影响,构建总体模型,并根据此模型及参数,以各影响因素空间分布特征为因变量,反向推导,计算各栅格单元人口数,实现人口空间化^[9,20,21];②以栅格为单元,将各影响因素通过空间叠加,组合为影响人口空间分布的贡献系数,再将各区域人口数分配到各格网中^[7,22],其中,贡献系数可通过专家打分法、层次分析法等模型计算获得。

前人研究为人口数据空间化奠定了较好的基础,但仍存在诸多不足之处,例如:空间化模型以线性回归模型为主,然而县级行政区域范围内的影响因素与人口统计数之间线性相关性较弱,人口空间化结果误差较大^[23,24];以贡献系数分配行政单元人口数的空间化方案中,众多学者采用专家打分法^[7]、层次分析法等方法对多源数据进行简单叠加^[22],忽略了不同数据源之间内部关系,未实现多源因素的有机整合,各因素对人口空间分布的控制信息尚未完全挖掘。可见,对多源数据进行科学处理,并采用科学的综合模型充分挖掘多源数据对人口空间分布的控制作用,是人口数据空间化的重要环节和内容,更是人口空间化结果科学性和精确性的重要保障。

河北省张家口市位于北京市西北,是京津地区重要供水源地,承担着天然生态屏障作用^[25];随着京津冀协同发展的推进,张家口社会经济得到迅速发展,土地利用矛盾日益突出,急需探明该区人口空间分布特征,为该区域土地资源管理和城市发展战略的制定提供决策依据^[26]。以乡镇级尺度为控制单元进行人口空间化,有利于获得更高精度的人口空间化结果,可为多尺度的综合研究提供数据支持,本文以河北省张家口市为例,基于2010年张家口市乡镇级人口统计数据与GIS/RS技术,采用主成分分析法整合影响人口空间分布的多源数据,确定各因素对人口空间分布的综合影响,并设计贡献系数分配人口统计数据的空间化方案,实现张家口市1km²

格网人口空间化。

2 研究区概况

河北省张家口市位于113°50'E-116°30'E,39°30'N-42°10'N之间,与京、冀、晋、内蒙古四省、直辖市(自治区)相接壤,全市3.68万km²,辖4区、13县209个乡镇。张家口市地势呈现西北高、东南低的阶梯状分布态势,海拔320~2841m(图1),分坝上、坝下两大区域,坝上高原为内蒙古高原向南延伸的区域,海拔在1400m左右,地势北低南高,崇礼、赤城山地面积较大;坝下山间盆地南部太行山向北延伸以及中部燕山山脉向西延伸的余脉所在,山地与小面积盆地相间分布,海拔在1000~2800m之间,地形复杂,主要有洋河、壶流河、桑干河等河谷谷地^[27]。气候以半干旱大陆性季风气候为主,四季分明,冬季寒冷而漫长^[25]。至研究基期2010年,张家口市GDP为954.097亿元,全社会固定资产投资完成903.67亿元,总人口达465.966万人,其中,农业人口为313.8万人。张家口市土地利用类型以耕地、林地、草地为主,建设用地主要集中在坝下洋河、壶流河、桑干河等谷地,以及坝上水源较好的区域。

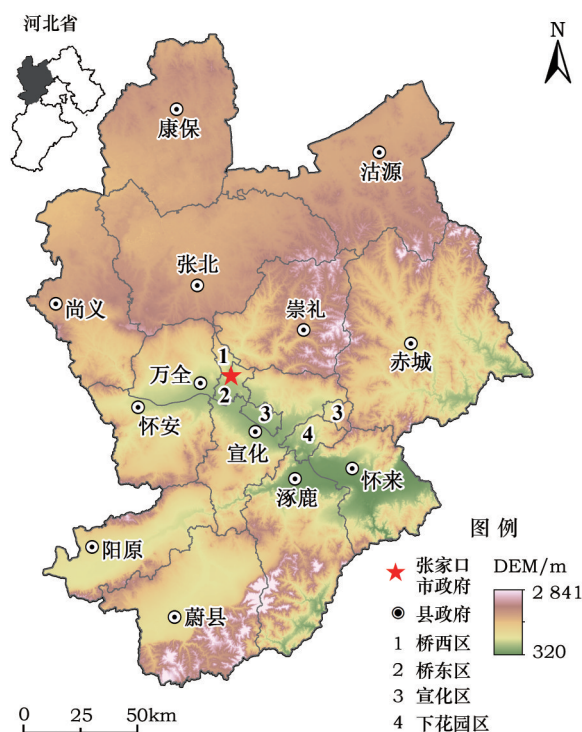


图1 研究区区位

Figure 1 Location of study area

3 数据来源与研究方法

3.1 研究思路与数据来源

人口空间分布受自然和社会、经济等众多因素的影响。综合考虑张家口市自然、社会经济实际情况,借鉴国内外已有研究,确定影响人口空间分布的主要因素,通过GIS与RS技术,得到基于栅格单元的各影响因素空间分布特征;利用主成分分析模型,确定各种影响因素对栅格单元的人口空间分布贡献系数;遵循以行政区域为控制单元,由完整行政单元的土地空间共同承担统计人口的原则,将各行政单元的统计人口离散化到栅格单元上,实现统计人口空间化。2010年是新近全国人口数据普查的关键年份,以2010年为研究基期,可搜集到相对科学和全面的乡镇统计数据。考虑到研究区实际情况及数据可得性,确定人口空间化栅格单元为1km×1km。

本文所涉及数据主要有:由中国科学院数据共享平台提供的2010年Landsat5遥感影像(30m分辨率)、DEM数据(30m分辨率)^[28];由张家口市统计局提供2010年乡镇级统计人口数(内部资料);由美国国家地球物理数据中心(NGDC)提供的2010年DMSP/OLS稳定夜间平均灯光数据(约900m分辨率)^[29]。遥感影像通过辐射校正、大气校正、影像裁剪等预处理后,用于土地利用图、植被指数NDVI提取;土地利用类型图通过ENVI5.1软件中神经网络分类算法对2010年8月Landsat5影像分类后,结合SPOT5影像目视解译获得,共分为林地、草地、耕地、园地、建设用地、水域、其他土地7种土地利用类型;NDVI数据通过NDVI计算公式对Landsat5影像近红和红波段进行波段计算获得。

3.2 影响因素确定与处理

根据人类居住分布特征以及活动的分布规律,并参考已有研究成果^[10,11,13,22],选择人居适宜性指数(Human suitable index, HSI, 下文简称人居指数)、地形起伏度(Relief Degree of Land Surface, RDLS)、格网内建设用地比例、格网内耕地比例、格网内园地比例、格网内水域比例6个因素作为张家口市人口空间化研究的影响因素。人口密度是表征单位面积上所承载的人口数量,因此,本文遵循行政单元内所有面积共同承担行政区域内所有人口数,以影

响因素控制人口空间分布密度的原则进行因素处理方案设计。

(1)HSI指数是表征当前环境下,现有人口居住的适宜空间^[13],该指数可直观反映出人口居住的集中区域,由DMPS/OLS稳定夜间灯光数据和NDVI数据通过GIS空间分析方法获得。DMPS/OLS数据表征区域夜间灯光强度,高亮区域表示人口集中分布的区域,可最大限度的反映现有人口居住的区域;已有研究中,众多学者将DMPS/OLS作为区域统计人口分配的贡献系数对区域人口统计数据进行了空间化^[11,10];将归一化植被指数NDVI与DMPS/OLS数据相结合制作人居指数,能避免由于夜间灯光指数数据过饱和而引起的结果偏差问题^[10]。HSI计算公式为:

$$HSI = \frac{(1 - NDVI) + OLS_{nor}}{(1 - OLS_{nor}) + NDVI + OLS_{nor} \times NDVI} \quad (1)$$

式中NDVI为标准化到0~1之间的植被指数;OLS_{nor}为标准化到0~1之间夜间灯光数据。人居指数反映了人口集中的区域,但由于人口空间活动不仅在有灯光的区域,且通过遥感影像所获得夜间灯光数据和NDVI整合数据反映人口空间分布具有一定的不确定性,例如,在农村地区,由于灯光数据较弱,很难被遥感卫星识别。因此,需要进一步借助地形因素、土地利用类型等因素对人口分布进行控制,以提高贡献系数的客观性和科学性。

(2)RDLS是表征区域地形地貌起伏特征的重要指标^[3]。地形起伏度与人口空间分布存在极显著负相关关系,是影响人口空间分布的重要因素,即RDLS越大,地形起伏度越大,人口密度越小,反之,地形起伏度越小,地势越平坦,人口密度越大^[22,30]。常用的RDLS采用单位面积内高程极值之差进行计算,国内学者结合高程、坡度以及计算的单位面积将计算公式进行改进,使得RDLS成为具备地理学意义的独立指标^[22],其公式为:

$$RDLS = \frac{H_{mean}}{1\ 000} + \frac{(H_{max} - H_{min}) \times (1 - p_A/A)}{500} \quad (2)$$

式中RDLS为地形起伏度;H_{mean}为以某一栅格单元为中心一定区域内的平均海拔(m);H_{max}、H_{min}分别为区域内的最高与最低海拔(m);p_A为区域内的平地面积(km²),研究将坡度小于等于5°的区域确

2017年11月

定为平地; A 为区域总面积, 研究结果格网分辨率为 1km, 因此, A 值为 1km²。

(3) 土地利用类型是人类活动对地表下垫面影响最直接和最强烈的表现, 也是区域人口分布的重要影响因素^[11]。建设用地、耕地、水域、园地是人类活动频繁区域^[11,19], 以此 4 种地类作为人口空间化重点控制区域, 可反映土地利用类型对人口空间分布的影响作用, 有助于提高人口空间化的客观性。本文基于张家口市 2010 年土地利用数据, 运用 ArcGIS 10.4 空间分析功能, 分别统计各格网内的建设用地、耕地、园地、水域面积比例, 以此作为各地类对人口空间分布的贡献值。由于水域面积过大的区域, 人口较难分布, 且存在一定的风险, 因此, 在对水域因素进行处理时, 将水域比例大于 70% 的栅格予以剔除^[23], 以确保水域因素的控制作用具有现实意义。

3.3 影响因素整合

选择科学的方法对影响因素进行综合整理, 是利用贡献系数分配人口统计数据空间化方法的关键, 决定了所选择的影响因素控制作用是否被充分挖掘^[7]。主成分分析法是基于各因子之间的特征进行权重赋值, 并获取各因素综合值的方法; 与层次分析法、专家打分法相比, 该法根据各因素之间特征关系进行权重打分, 可充分整合 HSI 、 $RDLS$ 、以及格网内建设用地、耕地、水域、园地面积比例的控制信息。因此, 本文运用主成分分析模型, 整合多源影响因素, 获得人口空间分布的综合贡献系数 IFI 值。具体计算公式为:

$$IFI = \frac{x_1}{\sum_{i=1}^m x_i} F_1 + \frac{x_2}{\sum_{i=1}^m x_i} F_2 + \cdots + \frac{x_m}{\sum_{i=1}^m x_i} F_m \quad (3)$$

公式(3)可简化为

$$IFI = \lambda_1 F_1 + \lambda_2 F_2 + \cdots + \lambda_m F_m \quad (4)$$

式中 IFI 为影响人口空间分布的综合贡献值; x_1, x_2, \cdots, x_m 为确定的影响因素; F 为单个主成分得分; λ 为对应主成分的贡献率。

3.4 人口空间化模型

本文采用人口空间分配模型进行人口空间化。即基于 IFI 综合贡献系数, 遵循以行政单元控制为原则, 将各行政区域人口统计数据分配到栅格

单元中, 具体分配模型为:

$$P_y = \frac{IFI_n}{IFI_s} \times P_{ss} \quad (5)$$

式中 P_y 为栅格单元人口数; IFI_n 为各栅格单元贡献系数; IFI_s 为栅格所在行政单元的总贡献系数; P_{ss} 为对应行政单元的统计总人口。

4 结果与分析

4.1 影响因素特征分析

根据已有数据和上述公式计算得到各影响因素空间分布(图2)。其中, 人居指数 HSI 值域介于 0.093~3.081 之间, 均值为 0.323, 标准差为 0.190。 HSI 高值区域主要分布于张家口市市辖区、县城中心及其周围人口集中区域, 在林地、草地、未利用地等集中分布区域 HSI 值均较小(图2a)。

$RDLS$ 是地形地貌特征最直接的反应, 直接影响着人口空间分部特征。由图2b可知, 研究区 $RDLS$ 值域在 0.473~3.874 之间, 均值为 1.46, 标准差为 0.44。 $RDLS$ 低值区与 HSI 高值区具有较高的空间一致性, 主要分布在张家口市市区及其周围区域、洋河、壶流河、桑干河等河谷谷地、山间盆地等区域; 高值区主要集中在谷地两侧的山地地区。北部坝上高原 $RDLS$ 差异较小, 南部山地丘陵地区 $RDLS$ 差异较大。

4 种土地利用类型的格网面积比例如图2c-图2f。其中格网建设用地面积比例空间分布图表明建设用地集中分布在坝下洋河、桑干河、壶流河谷地, 以及坝上水源较好的地区(图2c), 该数据可弥补由于农村区域灯光效果不明显所引起 HSI 指数对人口空间分布控制的不足作用^[10]。耕地主要集中在坝下山间谷地以及坝上高原地区(图2d); 坝上地区水域面积比例高值区呈点状分布集中于淖湖区域, 坝下地区高值区呈带状分布集中于河流谷地(图2e); 园地主要集中在南部坝下山间盆地, 尤以涿鹿-怀来盆地、蔚县-阳原盆地园地比例较高(图2f)。4 种土地利用类型分布相对集中, 坝下高值区主要集中在洋河谷地、壶流河谷地以及桑干河谷地等, 坝上地区地形较为平坦, 4 种地类分布较为均一。

4.2 综合贡献系数分析

根据主成分分析模型, 计算 6 种因素对人口空

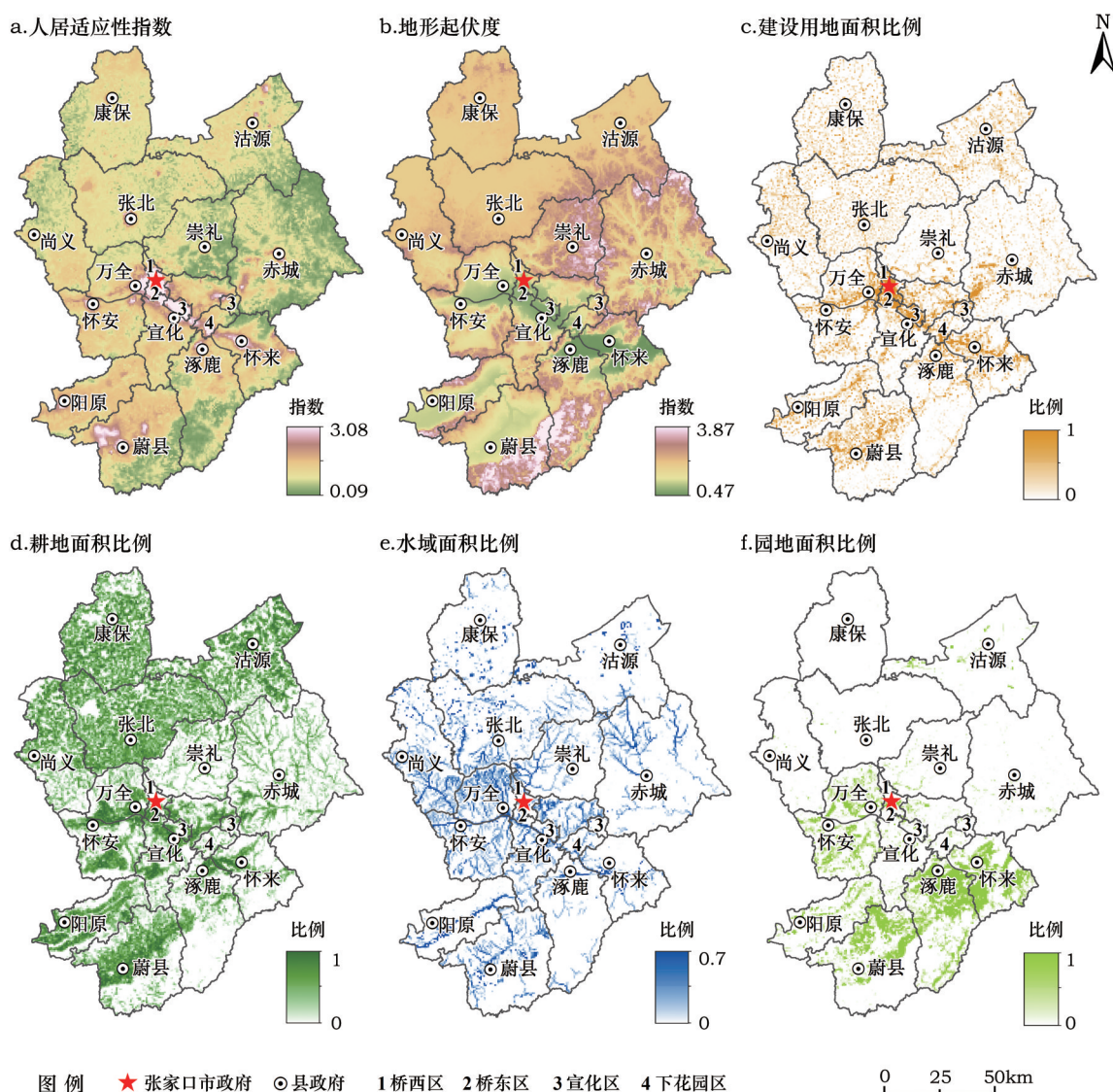


图2 影响因素空间分布特征

Figure 2 The spatial distribution characteristics of influencing factors

间分布影响的综合贡献系数 IFI 值及各主成分特征向量与成分参数(表1)。其中,第一成分包含了6种因素的37.283%的信息量,该成分特征根最大,为2.237,反映了各因素对综合贡献系数的贡献大小及趋势,其贡献大小为 $HSI > RDLS > \text{建设用地比例面积} > \text{耕地面积比例} > \text{园地面面积比例} > \text{水域面积比例}$,其中 $RDLS$ 为负向贡献,即 $RDLS$ 越小,其对人口空间分布的影响越大,人口密度越高,其余因素均为正向贡献。到第四主成分时,主成分累计贡献率达到88.923%,说明前四个成分包含了所有因素88.923%的信息量。因此,选择前4个主成分计算各因素人口空间分布的贡献系数 IFI 值,主成分得分

系数矩阵见表2。

根据主成分模型拟合参数,可获得各因素综合贡献系数 IFI 值,其值域介于-3.235~3.982之间。综合贡献系数出现0或小于0的情况,具有一定的物理意义,即在部分土地空间人口难以活动或生存,则在贡献系数中呈现负向贡献。但在实际中,人口空间分布在区域中不可能出现负人口分布,因此,为了避免人口分布空间化结果出现负值,将 IFI 值小于0的栅格单元,均调整为0,即对人口分配为零贡献,使 IFI 值均大于等于0,从而获得最终综合贡献系数(图3)。

调整后的 IFI 值介于0~3.982之间,均值为

2017年11月

表1 主成分特征向量及累计贡献率

影响因素	主成份					
	1	2	3	4	5	6
<i>HSI</i>	0.848	0.003	-0.199	0.264	-0.186	0.369
<i>RDLS</i>	-0.818	-0.046	-0.266	0.264	0.336	0.273
建设用地面积比例	0.721	-0.091	-0.443	0.354	0.309	-0.233
耕地面积比例	0.411	-0.678	0.311	-0.452	0.242	0.112
水域面积比例	0.217	0.279	0.787	0.476	0.169	0.019
园地面积比例	0.334	0.729	-0.084	-0.542	0.221	0.084
特征根	2.237	1.079	1.030	0.989	0.379	0.285
贡献率/%	37.283	17.988	17.168	16.484	6.323	4.754
累计贡献率/%	37.283	55.271	72.439	88.923	95.246	100.000

表2 主成分得分系数矩阵

影响因素	主成份			
	1	2	3	4
<i>HSI</i>	0.379	0.003	-0.193	0.267
<i>RDLS</i>	-0.366	-0.043	-0.258	0.267
建设用地面积比例	0.322	-0.085	-0.430	0.358
耕地面积比例	0.184	-0.628	0.302	-0.457
水域面积比例	0.097	0.258	0.764	0.481
园地面积比例	0.149	0.676	-0.082	-0.548

0.561,标准差为0.494。由图3可知,张家口市建设用地、耕地、河流集中分布区域综合贡献系数 IFI 值较大,林地、草地集中分布区域 IFI 值较小。 IFI 系数高值区与 HSI 高值区、 $RDLS$ 低值区以及建设用地、耕地、水域、园地集中分布区基本一致,而 HSI 低值区、 $RDLS$ 高值区及林地、草地、未利用地集中分布区 IFI 值均较小。可见,贡献系数 IFI 值随着各因素变化而变化的空间分布趋势,其空间分布特征融合了6种影响因素的空间分布特征,表明6种因素对人口空间分布的控制作用被有效挖掘。

4.3 人口空间化结果分析

根据人口分配模型,基于2010年各乡镇人口统计数据,结合 IFI 值空间分布结果,获得基于栅格单元的人口空间分布(图4)。乡镇级离散化人口空间化分布特征值介于(0~4601.65)人/ km^2 之间,平均值为123.41人/ km^2 ,标准差为271.641。从空间化结果可以看出,高值区主要集中在桥西区、桥东区、宣化区以及各区、县的县城中心;其他区县人口空间分布随各影响因素空间分布变化而变化,人口高值区

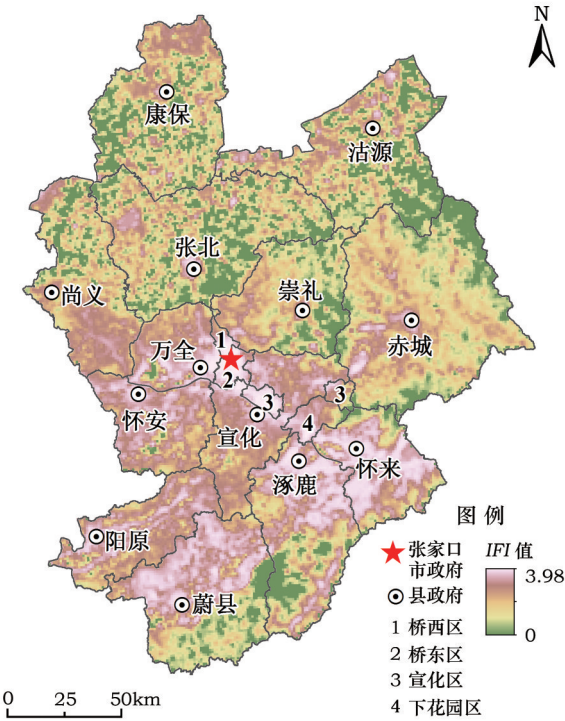


图3 影响因素综合贡献系数

Figure 3 The IFI value of the spatial distribution of population

集中在乡镇中心所在地,这些区域 HSI 值和建设用地比例较高、 $RDLS$ 值较低。由于人口离散化以行政区为控制单元,在离散化人口空间分布栅格图中,人口空间分布在各行政区范围内呈现出高低分布的面板效应。从人口空间化分布特征可以看出,其结果一定程度上符合张家口市实际人口空间分布特征,坝上高原区人口空间分布低于坝下山间盆地,坝下区域土地空间承载的人口数高于坝上

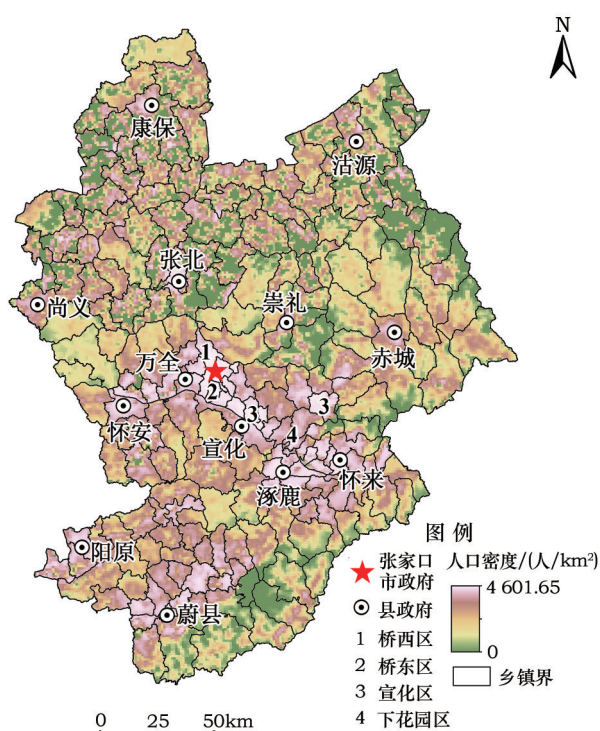


图4 张家口市人口空间化结果

Figure 4 The result of population specialization in Zhangjiakou City

高原区。坝上五县中,张北县人口空间分布高于其他区域,主要是由于张北县为张家口市重点发展县之一,近年来,尚义、康保、沽源、赤城四县的人口逐渐向张北县迁移定居,从而导致张北县人口空间分布高于其他县。坝下各县人口集中分布在较为平坦的洋河、桑干河、壶流河谷地,以及官厅水库周边区域,即怀来、涿鹿、蔚县、阳原等人口密度高值区域,这些区域地形平坦、水源充足、耕地质量好、人居环境适宜性高;坝下地区是张家口市经济发展的核心区,交通便捷、基础设施较为完善,得天独厚的自然、社会、经济条件,是坝下各县人口空间分布密集的主要原因。受负向贡献系数的影响,南部太行山脉、东北部的燕山山脉以及部分海拔较高的山地地区人口分布稀少,甚至为0人口分布;这些区域主要受地形的影响,实际人口分布亦较少。从以上分析可知,人口空间化分布结果客观性较强。

为了进一步验证研究结果的客观性,将离散化人口空间分布分别与影响因素进行pearson相关分析,与 HSI 、 $RDLS$ 、建设用地面积比例、耕地面积比例、水域面积比例和园地面积比例的相关系数分别

为 0.689^{**} 、 -0.371^{**} 、 0.455^{**} 、 0.085^{*} 、 0.276^{**} 和 0.230^{**} (** 为0.01水平极显著相关),表明:所选取的指标中,人口离散化结果与 HSI 指数、建设用地极显著相关,与其他指标之间呈现显著相关关系。换句话说:本文人口空间化结果具有较强的客观性,符合城镇区域人口密度较高、其他区域较低的空间分布态势,且其人口空间分布特征随着各因素变化而变化。

5 讨论

人口空间化研究中,多源空间数据控制信息的充分挖掘是科学、准确地实现人口统计数据空间离散化的重要环节,空间化模型的选择是确保人口空间化结果具有足够精度的关键。目前,已有研究大多采用回归分析模型构建人口统计数与影响因素之间的多源回归模型,通过人口数回归分析整合各因素对人口进行空间化,但由于该方法为全局空间模型,将全域范围内的影响因素与人口统计数据视为线性相关关系进行建模,得到的结果往往存在较大误差,尤其是在大中尺度区域人口空间化时,各因素的空间异质性和复杂性显著,得到的人口空间化结果误差更大,陈晴等利用回归分析法预测区县级人口空间化,所得到的区县级尺度预测误差为15%,乡镇级预测误差在20%左右^[23,31];吴建生等利用贡献系数分配京津冀地区统计总人口数^[32],得到京津冀地区空间化总人口比统计总人口多出153.29万人,各市人口误差范围在-24.65%~25.67%之间。与前人研究相比,本文采用乡镇级控制单元内贡献系数分配人口统计数的方法,可将人口误差控制在乡镇级尺度范围内,即在乡镇级尺度内其人口误差基本为0,以此推演县级尺度的人口空间化误差均较小。

本文涉及的人口空间化方案是影响因素分配控制单元统计人口的方法,主要涉及三个基本变量,即控制单元内的栅格单元数、栅格单元的贡献系数以及控制单元的人口总数。模型应用的关键在于控制单元的尺度以及影响因素整合的准确性与科学性,其中影响因素是贡献系数的基础,直接影响单位栅格内人口分配量,因此,影响因素选择的科学性,是决定栅格单元精度的关键。在贡献系数一定的情况下,控制单元尺度越小,则预测精度

2017年11月

越好,精度误差控制在控制单元内。人口空间化研究带有较强的不确定性,本文人口空间化建模思路是以研究区内所有土地空间共同承担行政单元内的统计人口为原则,采用主成分分析法,有效挖掘了各因素的影响特征信息,获得的 IFI 贡献系数客观的反应了人口空间分布的现实情况,所获得的人口空间化分布特征符合城镇区人口密度较高、其他区域较低的空间分布特征,采用乡镇级尺度的统计人口数据进行了人口空间化,所得空间化结果的精度优于县级控制单元空间化结果。由于乡镇级尺度离散化结果具有绝对的精度优势,在向上一级行政单元推演时,依然具有较强的精度优势,因此,在具体工作中,可克服行政尺度效应,适应于县级尺度和市级尺度的应用研究。此外,这一方法可避免由影响因素所引起的非线性误差问题,将误差控制在行政单元尺度内,在控制单元越小的情况下,可获得较高精度的人口空间化结果。但由于地理学研究对象格局与过程存在空间上尺度依存问题^[33],本研究中人口空间化控制指标如夜间灯光数据等尺度的限制^[34,35],最大栅格尺度只能采用1km分辨率,在相关研究中,可根据具体工作需要、以及数据可获取性,利用本文研究方案,设计与研究尺度相匹配的控制指标和栅格单元,进行人口离散化工作。此外,由于村级数据较难获取,即下一级行政单元统计人口检验控制单元,难以采用传统方法进行精度验证。本研究采用了乡镇级行政单元作为精度验证控制单元,较为科学的精度验证方法是通过栅格抽样进行验证,此方法为理想的空间化验证方法,但实施难度较大;进一步研究,应根据研究需求,优化人口分布的影响因素,从而保证人口离散化结果的客观性和准确性,以确保人口空间化具有一定的精度。

6 结论

本文基于GIS/RS技术,运用主成分分析计算人口空间分布影响贡献系数 IFI 值,通过贡献系数分配人口总数的空间化方法对张家口市2010年各区、乡镇统计人口进行了1km×1km 网格空间化研究,主要结论如下:

(1)6种影响因素空间分布中,人居指数 HSI 高值区分布于张家口市市辖区、县城中心及其周围人

口集中区域,偏远区域 HSI 值均较小。地形起伏度 $RDLS$ 低值区主要分布在洋河、壶流河、桑干河等河谷谷地及山间盆地等区域,高值区主要集中在谷地两侧的山地地区。建设用地等4种土地利用类型格网面积比例介于0~1之间,主要分布在于洋河、壶流河、桑干河等河流谷地及坝上高原区域。

(2)6种因素对人口空间分布贡献大小为: $HSI>RDLS>$ 建设用地比例面积>耕地面积比例>园地面积比例>水域面积比例,其中 $RDLS$ 为负向贡献,其余因素均为正向贡献。 HSI 高值区、 $RDLS$ 低值区及建设用地、耕地、水域和园地集中区域 IFI 值较高,而林地、草地、未利用地等人类活动较少的区域 IFI 值较小。6因素对人口空间分布的影响作用较好,其对人口空间分布的影响信息被有效挖掘。

(3)空间化人口密度高低分布特征符合张家口实际人口空间分布,有随各影响因素空间分布变化而变化的趋势,人口离散化结果与 HSI 指数、建设用地高度相关,与其他指标之间联系较为密切;高值区主要集中于包括桥西区、桥东区和宣化区在内的市辖区,以及各乡镇的城中心,主要集中于 HSI 高值区、 $RDLS$ 低值区和建设用地集中区域。

(4)通过乡镇级控制单元内贡献系数分配人口统计数的方法,可将人口误差控制在乡镇级尺度范围内,与前人研究结果相比,具有绝对的精度优势。

致谢:真诚感谢《资源科学》编辑部王立新老师和两位匿名专家在论文完善过程中提出的宝贵建议,以及张家市政府、国土局、统计局等相关部门给予的数据支持。

参考文献(References):

- [1] Cao X, Wang J M, Chen J, et al. Spatialization of electricity consumption of China using saturation- corrected DMSP- OLS data[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2014, 28(1): 193-200.
- [2] Spielman S E, Logan J R. Using high-resolution population data to identify neighborhoods and establish their boundaries[J]. *Annals of the Association of American Geographers*, 2013, 103(1): 67-84.
- [3] 柏中强,王卷乐,杨飞. 人口数据空间化研究综述[J]. *地理科学*

- 进展, 2013, 32(11): 1692-1701. [Bai Z Q, Wang J L, Yang F. Research progress in spatialization of population data[J]. *Progress in Geography*, 2013, 32(11): 1692-1701.]
- [4] 廖顺宝, 孙九林. 基于 GIS 的青藏高原人口统计数据空间化[J]. 地理学报, 2003, 58(1): 25-33. [Liao S B, Sun J L. GIS based spatialization of population census data in Qinghai-Tibet Plateau[J]. *Acta Geographica Sinica*, 2003, 58(1): 25-33.]
- [5] Dobson J E, Bright E A, Coleman P R, *et al.* LandScan: a global population database for estimating populations at risk[J]. *Photogrammetric Engineering and Remote Sensing*, 2000, 66(7): 849-857.
- [6] 中国科学院资源环境科学数据中心. 社会经济数据库[EB/OL]. (2012-02-20)[2016-06-06]. <http://www.resdc.cn/DataList1.aspx?FieldTypeId=8>, 7. [Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (RESDC). Social and Economy Database [EB/OL]. (2012-02-20)[2016-06-06]. <http://www.resdc.cn/DataList1.aspx?FieldTypeId=8>, 7.]
- [7] 葛美玲, 封志明. 基于 GIS 的中国 2000 年人口之分布格局研究: 兼与胡焕庸 1935 年之研究对比[J]. 人口研究, 2008, 32(1): 51-57. [Ge M L, Feng Z M. Study of China's 2000 population distribution pattern based on GIS: and contrast with the population of Hu Huanyong line in 1935[J]. *Population Research*, 2008, 32(1): 51-57.]
- [8] Zeng C Q, Zhou Y, Wang S X, *et al.* Population spatialization in China based on night-time imagery and land use data[J]. *International Journal of Remote Sensing*, 2011, 32(24): 9599-9620.
- [9] Saha A K, Gupta R P, Sarkar I, *et al.* An approach for GIS-based statistical landslide susceptibility zonation: with a case study in the Himalayas[J]. *Landslides*, 2005, 2(1): 61-69.
- [10] Letu H, Hara M, Yagi H, *et al.* Estimating energy consumption from night time DMPS/OLS imagery after correcting for saturation effects[J]. *International Journal of Remote Sensing*, 2010, 31(16): 4443-4458.
- [11] Xiong J N, Wei F Q, Jiang Y H, *et al.* Classification simulating population distribution and disaster emergency management considering the disparity between city and rural area taking Sichuan as an example [J]. *Disaster Advances*, 2012, 5(4): 1718-1723.
- [12] Harvey J T. Estimating census district populations from satellite imagery: some approaches and limitations[J]. *International Journal of Remote Sensing*, 2010, 23(10): 2071-2095.
- [13] 赵利利, 孟芬, 马才学. 基于多源遥感数据的武汉市人口空间分布格局演化[J]. 地域研究与开发, 2016, 35(3): 165-160. [Zhao L L, Meng F, Ma C X. The analysis on spatial distribution and evolution of the population in Wuhan city based on multi-source Remote Sensing data[J]. *Areal Research and Development*, 2016, 35(3): 165-169.]
- [14] Lu D, Tian H, Zhou G, *et al.* Regional mapping of human settlements in southeastern China with multisensor remotely sensed data[J]. *Remote Sensing of Environment*, 2008, 112(9): 3668-3679.
- [15] 闫庆武, 卞正富, 张萍, 等. 基于居民点密度的人口密度空间化[J]. 地理与地理信息科学, 2011, 27(5): 95-98. [Yan Q W, Bian Z F, Zhang P, *et al.* Census spatialization based on settlements density[J]. *Geography and Geo-information Science*, 2011, 27(5): 95-98.]
- [16] 康敏捷, 曹可, 王辉, 等. 基于作息空间思想的人口统计数据空间化方法[J]. 地理与地理信息科学, 2012, 28(4): 60-62. [Kang M J, Cao K, Wang H, *et al.* Spatialization method of demographic data based on work and residence space conception[J]. *Geography and Geo-information Science*, 2012, 28(4): 60-62.]
- [17] Zhang Y, Dong C, Liu J P, *et al.* Gridded population distribution map for the Hebei province of China[J]. *Environmental Engineering and Management Journal*, 2015, 14(3): 673-680.
- [18] 高义, 王辉, 王培涛, 等. 基于人口普查与多源夜间灯管数据的海岸带人口空间化分析[J]. 资源科学, 2013, 35(12): 2517-2523. [Gao Y, Wang H, Wang P T, *et al.* Population spatial processing for Chinese Coastal Zones based on census and multiple night light data [J]. *Resources Science*, 2013, 35(12): 2517-2523.]
- [19] 田永中, 陈述彭, 岳天祥, 等. 基于土地利用的中国人口密度模拟[J]. 地理学报, 2004, 59(2): 283-292. [Tian Y Z, Chen S P, Yue T X, *et al.* Simulation of Chinese population density based on land use[J]. *Acta Geographica Sinica*, 2004, 59(2): 283-292.]
- [20] Townsend A C, Bruce D A. The use of night-time lights satellite imagery as a measure of Australia's regional electricity consumption and population distribution[J]. *International Journal of Remote Sensing*, 2010, 31(16): 4459-4480.
- [21] Yue T X, Wang Y A, Chen S P, *et al.* Numerical simulation of population distribution in China[J]. *Population and Environment*, 2003, 25(2): 141-163.
- [22] 周自翔, 李晶, 任志远. 基于 GIS 的关中-天水经济区地形起伏度与人口分布研究[J]. 地理科学, 2012, 32(8): 951-957. [Zhou Z X, Li J, Ren Z Y. The relief degree of land surface and population distribution in Guanzhong-Tianshui economic region using GIS[J]. *Scientia Geographica Sinica*, 2012, 32(8): 951-957.]
- [23] 陈晴, 侯西勇. 集成土地利用数据和夜间灯光数据优化人口空间化模型[J]. 地球信息科学学报, 2015, 17(11): 1370-1377. [Chen Q, Hou X Y. An improved population spatialization model by combining land use data and DMSP/OLS data[J]. *Journal of Geo-information Science*, 2015, 17(11): 1370-1377.]
- [24] 杨续超, 高大伟, 丁明军, 等. 基于多源遥感数据及 DEM 的人口统计数据空间化-以浙江省为例[J]. 长江流域资源与环境, 2013, 22(6): 729-735. [Yang X C, Gao D W, Ding M J, *et al.* Modeling population density using multi-sensor remote sensing

2017年11月

- data and DEM: a case study of Zhejiang Province[J]. *Resources and Environment in the Yangtze Basin*, 2013, 22(6): 729-735.]
- [25] 许月卿, 田媛, 孙丕苓. 基于 Logistic 回归模型的张家口市土地利用变化驱动力及建设用地增加空间模拟研究[J]. 北京大学学报(自然科学版), 2015, 51(5): 955-964. [Xu Y Q, Tian Y, Sun P L. Study on driving forces and spatial simulation of land use change in Zhangjiakou city based on logistic regression model[J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2015, 51(5): 955-964.]
- [26] Sun P L, Xu Y Q, Yu Z L, et al. Scenario simulation and landscape pattern dynamic changes of land use in the Poverty Belt around Beijing and Tianjin: a case study of Zhangjiakou city, Hebei Province [J]. *Journal of Geographical Sciences*, 2016, 26(3): 272-296.
- [27] 孙丕苓, 许月卿, 王数. 环京津贫困带土地利用变化的地形梯度效应分析[J]. 农业工程学报, 2014, 30(14): 277-288. [Sun P L, Xu Y Q, Wang S. Terrain gradient effect analysis of land use change in poverty area around Beijing and Tianjin[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2014, 30(14): 277-288.]
- [28] 中国科学院计算机网络信息中心. 地理空间数据云[EB/OL]. (2011-04-03)[2017-03-30]. <http://www.gscloud.cn/>. [Computer Network Information Center, Chinese Academy of Sciences. Geospatial Data Cloud [EB/OL]. (2011-04-03)[2017-03-30]. <http://www.gscloud.cn/>.]
- [29] National Centers for Environmental Information. Version 4 DMSP-OLS Nighttime Lights Time Series[EB/OL]. (2011-05-24)[2014-09-30]. [https://ngdc.noaa.gov/eog/dmsp/download V4/composites.html](https://ngdc.noaa.gov/eog/dmsp/download/V4/composites.html).
- [30] 刘颖, 邓伟, 宋雪茜, 等. 基于地形起伏度的山区人口密度修正-以岷江上游为例[J]. 地理科学, 2015, 35(4): 404-470. [Liu Y, Deng W, Song X Q, et al. Population density correction method in mountain areas based on relief degree of land surface: a case study in the upper Minjiang River basin[J]. *Scientia Geographica Sinica*, 2015, 35(4): 464-470.]
- [31] 陈晴, 侯西勇, 吴莉. 基于土地利用数据和夜间灯光数据的人口空间化模型对比分析-以黄河三角洲高效生态经济区为例[J]. 人文地理, 2014, 29(5): 94-90. [Chen Q, Hou X Y, Wu L. Comparing of population spatialization models based on land use data and DMSP/OLS data respectively: a case study in the efficient ecological economic zone of the Yellow River Delta[J]. *Humen Geography*, 2014, 29(5): 94-90.]
- [32] 吴健生, 许多, 谢舞丹, 等. 基于遥感影像的中尺度人口统计数据空间化-以京津冀地区为例[J]. 北京大学学报(自然科学版) 2015, 51(4): 707-717. [Wu J S, Xu D, Xie W D, et al. Spatialization of demographic data at medium scale based on Remote Sensing images: Regarding Beijing-Tianjin-Hebei as an example[J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2015, 51(4): 707-717.]
- [33] 李双成, 蔡运龙. 地理尺度转换若干问题的初步探讨[J]. 地理研究, 2005, 24(1): 11-18. [Li S C, Cai Y L. Some scaling issues of geography[J]. *Geographical Research*, 2005, 24(1): 11-18.]
- [34] 刘明亮, 唐先明, 刘纪远, 等. 基于 1 km 格网的空间数据尺度效应研究[J]. 遥感学报, 2001, 5(3): 183-190. [Liu M L, Tang X M, Liu J Y, et al. Research on scaling effect based on 1km grid cell data[J]. *Journal of Remote Sensing*, 2001, 5(3): 183-190.]
- [35] 白燕, 廖顺宝, 孙九林. 栅格化属性精度损失的评估方法及其尺度效应分析-以四川省 1:25 万土地覆被数据为例[J]. 地理学报, 2011, 66(5): 709-717. [Bai Y, Liao S B, Sun J L. Evaluating methods and scale effects of attribute information loss in rasterization: a case study of 1:250 000 land cover data of Sichuan[J]. *Acta Geographica Sinica*, 2011, 66(5): 709-717.]

Spatial distribution of population specialization based on multi-source data : a case study of Zhangjiakou City

HUANG An^{1,2}, XU Yueqing^{1,2}, SUN Piling^{1,2}, LIU Chao^{1,2}, ZHENG Weiran^{1,2}

(1. College of Resource and Environmental Science, China Agricultural University, Beijing 100193, China;

2. Key Laboratory of Agricultural Land Quality, Monitoring and Control, Ministry of Land and Resources, Beijing 100193, China)

Abstract: The study of population can reveal the spatial information behind the demographic data, and helps to provide data support for the research on regional sustainable development. With the aid of technology of GIS and RS, we selected six factors that affect the distribution of population: Human Settlements Index (HSI), Relief Degree of Land Surface (RDLS), and the area ratio of construction land, cultivated land, water area, and garden in a regular grid. We also obtained the contribution coefficient IFI value of population spatial distribution through principal component analysis model, and allocated the population of Zhangjiakou City to a 1km×1km grid, which obeys the principles that the statistical population of the township is assumed by the land space of the whole area, and the administrative area is taken as a model control unit to achieve the goal of specialization of population. The results showed that the population density obtained in this study conformed to the actual situation of the city. There are changes over various factors influencing the spatial distribution tendency, and there is a significant relationship between result and HSI and construction land, which was closely related to the other factors. High value population area mainly concentrated in the area includes Qiaoxi, Qiaodong and Xuanhua districts, as well as downtown center. The specific area is distributed in the area of HSI high value, low RDLS and high value of construction land ratio area. Based on the township level scale as the control unit to forecast the spatial population, the accuracy of the result is better than that of the county level as control unit, which can provide data support for multi-scale comprehensive research.

Key words: GIS; RS; multi-source data; population; specialization; Zhangjiakou City