

DOI: 10.5846/stxb201812212776

王铮,夏海斌,田园,王魁,花卉,耿文均,田丽,郑保利,赵金彩.胡焕庸线存在性的大数据分析——中国人口分布特征的生态学及新经济地理学认识.生态学报,2019,39(14): - .

Wang Z, Xia H B, Tian Y, Wang K, Hua H, Geng W J, Tian L, Zheng B L, Zhao J C. A big-data analysis of Hu-line's existence in the ecology view and new economic geographical understanding based on population distribution. Acta Ecologica Sinica, 2019, 39(14): - .

胡焕庸线存在性的大数据分析 ——中国人口分布特征的生态学及新经济地理学认识

王 铮^{1,2,*}, 夏海斌¹, 田 园², 王 魁¹, 花 卉¹, 耿文均³, 田 丽¹, 郑保利¹, 赵金彩¹

1 华东师范大学 地理信息科学教育部重点实验室, 上海 200241

2 中国科学院科技战略咨询研究院, 北京 100190

3 河南大学黄河文明与可持续发展研究中心, 开封 475001

摘要: 本文基于大数据分析思路和数据挖掘工具, 在县级尺度上, 利用 2010 年第六次人口普查数据, 计算各县的平均人口密度, 以及合成海拔(地带性因素)、环境脆弱性、人生气候指数、农业生产潜力、适宜水资源偏离度、交通便捷性、区位指数等, 克鲁格曼 Krugman 所谓的区域地理本性特征, 对胡焕庸线存在的地理基础和生态学基础进行了分析。研究结果表明: (1) 中国人口密度与合成海拔、环境脆弱性、人生气候指数、适宜水资源偏离度、农业生产潜力等生态学因素密切相关, 与交通便捷性、GDP 和区位指数等显著相关。(2) 对大多数省的人口密度的影响的贡献率排在强烈的是合成海拔、农业生产潜力和和水资源适应度, 其次是交通便捷性和区位指数; 它们是克鲁格曼 Krugman 新经济地理学认识的第一次区域本性和第二次地理本性因素。(3) 新疆、山东、宁夏、内蒙古、黑龙江、江苏、北京和天津的人口分布比较独特, 还需引入新的因子来解释其人口分布的影响因素; (4) 胡焕庸线沿线是一个生态脆弱地带。本文的研究基本展现了中国人口的空间分布的胡焕庸线的地理学与生态基础。

关键词: 人口密度; 空间分布; 胡焕庸线; 大数据分析

A big-data analysis of Hu-line's existence in the ecology view and new economic geographical understanding based on population distribution

WANG Zheng^{1,2,*}, XIA Haibin¹, TIAN Yuan², WANG Kui¹, HUA Hui¹, GENG Wenjun³, TIAN Li¹, ZHENG Biao¹, ZHAO Jincai¹

1. Key Laboratory of Geographic Information Science, East China Normal University, Shanghai, 200241, China

2. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China

3. Institute of Yellow River Civilization and Sustainable Development, Henan University, Kaifeng 475001, China

Abstract: Based on big-data analysis, this paper used the data of the sixth census in 2010 to calculate the average population density of each county. Based on Krugman's geographic nature features, including synthesis of altitude, environmental vulnerability, and productive potentiality of agriculture, human climate index, and appropriate deviation degree of water resources, traffic convenience, the locational index and GDP were selected, the geographical and ecological basis of Hu's line existence are analyzed. The results showed that: (1) China's population density was closely related to ecological factors such as synthetic altitude, environmental vulnerability, human climate index, appropriate deviation degree of water resources, and productive potentiality of agriculture, and was significantly related to traffic convenience, GDP, and location index. (2) In most provinces, the synthetic altitude, productive potentiality of agriculture and appropriate deviation

基金项目: 国家社会科学重大项目 (18ZDA132)

收稿日期: 2018-12-21; 修订日期: 2019-05-30

* 通讯作者 Corresponding author. E-mail: wangzheng@casipm.ac.cn

degree of water resources contributed most to the population density, followed by transportation convenience and location index. They were the first nature and the second nature factors recognized by Krugman's new geography. (3) The population distribution is special in Xinjiang, Shandong, Ningxia, Inner Mongolia, Heilongjiang, Jiangsu, Beijing and Tianjin, and new factors should be introduced to explain the population distribution. (4) The Hu's line is an ecologically fragile zone. This paper presents the geographical and ecological basis of Hu's line, which is the dividing line of spatial distribution of Chinese population.

Key Words: population density; spatial distribution; Hu's Line; Big-data analysis

早在 1935 年,地理学家胡焕庸先生发现了黑河—腾冲一线,刻画了中国人口分布差异的空间格局^[1],后来被人们称为胡焕庸线。王铮、张丕远、刘燕华^[2]提出胡焕庸线对应于中国的生态环境脆弱带,将胡焕庸线与生态条件的变化明确地联系了起来。吴静、王铮等^[3]研究发现,在气候变化引起生态条件和社会经济发展等多要素的共同作用下,从汉晋时期,全国人口与经济重心持续向东南迁移,逐渐形成了东南地狭人稠、西北地广人稀的人口分布格局,最终在 1230—60 年代形成了胡焕庸线。胡焕庸线代表着中国人地生态系统的基本特征,因此有必要结合生态学和地理学观点认识胡焕庸线。

目前国际上也有一些学者对人口空间分布的生态背景做了研究。例如 Lovemore 和 Richard^[4]研究了经济发展政策和土地利用模式等对津巴布韦人口分布的影响。在国内,封志明等^[5]研究了中国人口的空间分布与地形起伏度的关系,研究结果表明,他们定义的地形起伏度与人口密度有较好的对数拟合关系,拟合度高达 0.91。方瑜,欧阳志云等^[6]从生态学观点将在这些研究的基础上,更详细地认识各种地理与生态要素人口分布的影响。更进一步,刘清春,王铮^[7],夏海斌,王铮^[8]借助 Krugman 的新经济地理学的地理本性(nature)的观点,阐述了中国人口分布与地理要素或本性(nature)有关。

立足于上述认识,本研究将以县级行政区为基本研究单元,基于目前最新的 2010 年第 6 次全国人口普查数据,在全国省级尺度对人口分布及空间背后的空间经济格局进行了分析。研究将参考夏海斌、王铮等^[8]提出等三次地理本性所包含的指标。基于数据的可获得性,选取了 8 个综合反映生态环境和社会经济发展状况的指标作为地理要素,探讨人口与胡焕庸线的生态学、地理学成因,以图从根本上说明胡焕庸线的存在性。

1 数据及其分析方法

1.1 数据来源

人口数据主要来自于全国第六次人口普查数据,行政区划数据是基于 1:400 万矢量图,数据以县级行政单位为界线,利用 ArcGIS10.2 对人口统计数据 and 行政区划数据进行空间关联和数据修订,基础数据首先以中国地理信息数据库 1:25 万 DEM 数据库为基础,气候数据资料来源于中国气象局网站,其中人生气候指数中的相关数据来源于中国 824 个气象站 1961—2012 年间 50 年的观测数据;霜期数据是利用 1951—2012 年中国 824 个气象站观测数据计算得到;降水数据来自气象站点观测的 1981—2010 年降水资料;地表岩性数据(喀斯特数据)来源于《中国岩溶环境地质图》;土壤类型数据来源于中国科学院寒区旱区研究所的相关网站,其数据粗度是 1km×1km;交通数据来源于国家 2010 年交通数字地图与 12306 网站。人均 GDP 数据和人均财政收入数据资料来源于《中国城市统计年鉴》和《中国县(市)社会经济统计年鉴》以及各省的统计年鉴,其中某些经济数据有一定保密性,统计合成后可以公开。特别是地形地貌数据来自于 NASA 遥感观察数据,计算出全国 1km×1km 的 DEM 数据,用于计算全国各县旗区的平均海拔、平均坡度和地表崎岖度。由此可见,这个数据体系由于其数据量和数据源的多样性,构成了一个大数据集合,构成我们使用大数据挖掘方法分析的科学基础。不过由于众所周知的原因,某些数据内容缺少台湾省的数据,故在图中采用虚线表示我国台湾地区的疆域。

1.2 胡焕庸线的计算

根据胡焕庸线的表述,胡焕庸线是从爱辉(今黑龙江黑河)到云南腾冲的一条直线,在图 1 中我们绘制了地图上连接着这两个地方的大圆曲线的线段,显示了其经过的县(旗、区),及其毗邻县的情况,胡焕庸线经过的地区在全国艰苦边远地区辨识中全部属于艰苦地区^[9],包含全国 66 个县(旗、区)。除黑河、腾冲外,比较著名的还有黑龙江嫩江、内蒙古准格尔、陕西榆林、志丹、甘肃陇南、天水、四川康定、北川、云南香格里拉、贡山等。

胡焕庸当年发现胡焕庸线的一个基础是人口密度。人口密度反映的不仅是人口分布数量,也是一个区域生态综合区域经济发展水平的测度与生态环境的质量的综合特征;一个可以供应密集人口的区域其经济发展水平是好的,生态条件也不可能恶劣;一般来说,人口稀少的地方,其经济水平也必然低。人口密度的存在特点代表了经济发展水平,代表了环境的生态质量水平。通常一个区域的平均人口密度的计算公式如下:

$$P = N/M \quad (1)$$

式中: P 为人口密度(人/ km^2); N 为区域人口数量(人), M 为区域的面积(km^2),图 1 是 2010 年中国人口密度的分布情况,图中白色实线是胡焕庸线。

1.3 生态与地理因子的选择和计算方法

根据数据挖掘原理,好的大数据分析,不仅需要数据量大,而且需要数据特征的物理、生态、经济或社会变量意义明确。本文根据依据 Krugman^[10] 的新经济地理学原理,以区域的地理本性来透视区域人口分布密度与分布特征。对于地理本性,选取了 8 个综合地理-生态环境和社会经济发展状况的指标表征生态与地理特征。在这些指标中,先天海拔和纬度综合反映一个地区所处的地理位置的地带性生态环境条件;环境脆弱性则直接刻画地区的生态环境的艰苦性以及自然灾害对人类生活的影响;人生气候指数反映地区的环境适合人类生存发展的程度;适宜水资源偏离度综合反映地区水资源状况对居民生产、生活的影响;农业生产潜力反映当地的农业生态条件对人类生活的支撑力;这些是克鲁格曼的理论所强调的区域第一次地理本性的特征。第一次地理本性除矿产条件外可以称为区域的生态本性。由人类适应行为产生的区域第二本性,是人类适应环境及其对环境改造的结果,包括交通便捷性指数、区位指数和人均 GDP,它们分别表达当地的通达水平、区位状况和经济发展水平、产业发展水平。对这些因素的选取和计算方法如下^[8]:

1.3.1 合成海拔

合成海拔指标结合了海拔和纬度二个方面的生态因素,高纬度地区与高海拔地区具有一定的自然地理或生态环境相似性(包括热量、土壤、植被等)。在本文的研究中,以全国 1:25 万 DEM 数据库为基础,通过 ARCGIS10.2 进行地理计算,得出全国各县市的平均海拔高度。考虑到随着地区纬度的升高,其接受太阳辐射量随之降低,与海拔高度类似。通过地区获得的有效太阳辐射量,可以计算其纬度位置对应的海拔高度,进一步考虑城市是人口主要居住地,在计算一个城市(或县城)合成海拔时,县城海拔高度的计算权重较大。合成海拔的公式如下:

$$\text{合成海拔} = 0.3 \text{ 平均海拔} + 0.7 \text{ 县城海拔} + 2/3 \text{ 纬度热量带折算海拔} \quad (2.2)$$

1.3.2 环境脆弱性

环境脆弱性表征一个地区各种自然环境生态特征,本文运用由地表崎岖度、无霜期、旱涝指数和喀斯特特征复合而成的生态环境脆弱性,它表征一个区域地表的生态环境艰苦性一面。各特征计算方法如下:

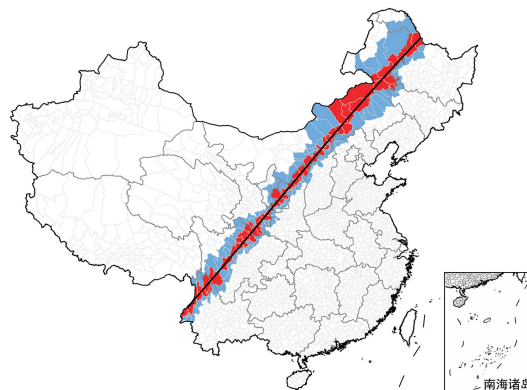


图 1 在大圆曲线上的胡焕庸线(色带是其毗邻地区)

Fig.1 Hu's line on the big round curve

地表崎岖度指标,本文的地表崎岖度 pSA 是运用夏海斌,王铮等^[8]的计算方法获取的。为了能够更加真实的展示现状,对坡度较大的地区提高其地表崎岖度指标,采用 Burrough^[11]提出的坡度算法用公式(2.3),得到的地表崎岖度:

$$pSA_{adj} = \begin{cases} pSA & S \leq 10^\circ \\ (S/20 + 1) \times pSA & S > 10^\circ \end{cases} \quad (2.3)$$

式中, pSA 是地表崎岖度, S 表示平均坡度, adj 表示调整得到的地表崎岖度。

霜期指标:霜期是自然地理环境不良的特征,是利用气象站观测数据,本文得到每个站点多年的有霜期,取其平均值得到每个地方的霜期的平均天数。由于某些站点的缺失,本文采用空间插值算法得到全国空间分布,最后对全国各县进行区域统计汇总,得到各县多年有霜期平均值,最后对其标准化,即为有霜期指标。

旱涝指数:本文采用标准化降水指数(Standardized Precipitation Index, SPI)表征区域旱涝特征,通常这是一种只需利用连续时间段(30年以上)的降水量数据就可以很好的表达的指数,可以反映不同地域和不同时间尺度的干旱降水情况,能够较好的反映干湿等级标准,可进行多时空尺度的对比分析。本文采用张岳军和郝智文等^[12]的计算公式计算 SPI 。从农业生态角度,考虑到冬季降水量对农作物生长的影响不大,因此在分季节计算 SPI 指数时,并未计算冬季的 SPI 。将春、夏和秋三个季节的 SPI 指数,按 1:1:1 的权重合成,可得到全国各县的复合 SPI 指数,其只大于 0 表示干旱,小于 0 表示雨水充足。然后对复合的 SPI 指数取绝对值,值越大表明灾害程度越大,最后对其标准化,得到旱涝指数。其值越大,越易发生旱涝灾害,值越小则表明水分条件越适宜。

岩溶指标:本文基于矢量化《中国岩溶环境地质图》,并根据岩溶类别进行分级,然后基于岩溶对当地的危害程度赋值。

最后对计算得到的地表崎岖度、有霜期、旱涝指数以 3:1:1 的权重复合,然后利用岩溶指标进行校正,即为最终的环境脆弱性指标,这个指标显然是一个环境的生态指标。

1.3.3 人生气候指数

人生气候指数是衡量地区人们生产、生活环境的舒适度的一个重要指标,它往往是人们日常评价区域生态条件的一个标准。王铮等^[13]论证,知识型产业和高技术人员追求舒适的生活环境,气候舒适有利于创造性劳动。舒适度可以体现在多个方面,气温、风速、湿度等都是重要的指标。本文采用刘清春,王铮等^[14]的气候舒适指数偏离度来表示气候舒适程度,即人生气候指数。该指数越大表示偏离舒适气候的程度越大,越不适宜人类生活,反之,越小越舒适,越适宜人类的生产和生活。

1.3.4 农业生产潜力

农业生产潜力是一个个有经济意义的生态指标,用它可推算出一个地区的期望农业产量,实际上是对陆地生态系统初级生态潜力的一种估计,是区域对人口活动的基本支撑力的测度。本文采用钟章奇和王铮等^[15]的模型,具体计算方法参照钟章奇的研究。

1.3.5 适宜水资源偏离度

我国水资源分布不均,南北差异较大。过度缺乏和充沛都会对生产和生活产生一定的影响。适宜水资源偏离度是反映一个地区水资源状况的综合指标,主要通过一个地区的湿润程度和由降水量引起的地表径流这两个方面来反映。本文采用夏海斌,王铮等^[8]的方法计算水资源适宜度。该指数越小水资源状况越适宜,指数越大表示水资源过度缺乏或者过度充沛,它们都不利于人们的生产和生活。

1.3.6 交通便捷性

交通便捷性指标不是自然生态系统特征,它的合成是以各种交通工具的质与量的乘积也即是以各种交通工具的速度与载客量的乘积作为各自的权重。交通便捷性的计算公式主要采用夏海斌,戴霄烨等^[16]的交通便利性的计算方法。

1.3.7 区位指数

区位指数是一个地区其与中心城市、省会城市以至首都的联系。因此,计算各县市的区位指数以衡量每

个县(市)在全国空间分布的相对位置。研究采用夏海斌等^[17] 威尔逊空间相互作用模型计算每个县(市)的区位指数,它表征城市或区域的贸易条件和经济辐射力,为计算这个数据,需要先行获得城市 GDP 数值、财政收入,城市间球面空间距离和道路交通距离等背景数据。

总之,这里选取合成海拔、环境脆弱性、人生气候指数、农业生产潜力、适宜水资源偏离度、交通便捷性、区位指数和 GDP 等区域指标,具有明确的地理学和生态学意义,这种具体指标的学科意义,我们认为是最终理解大数据分析意义必要的基础。而 GDP 等社会经济指标,有助于理解众多数据因子与人口密度关系。例如为了由数据挖掘信息,我们可以与人口密度影响因素为自变量,以人口密度为因变量,利用逐步回归分析方法,得到分省区的拟合方程,进而对各省的人口空间分布的地理或生态因素进行分析。

2 人口分布的空间特征

2.1 人口分布的空间格局

现在我们可以进一步来挖掘作为表象特征的图 2。从图 2 可以看出,中国人口主要聚集在中国的大城市,包括东部、中部和东南部的地级市,以及西部的某些省会城市,四川盆地、黄淮海平原、长江中下游平原和东南沿海地区等发展较好的地方。这些区域的附近人口分布也较多。人口相对稀疏的地区大多分布在胡焕庸线附近及其以西,东北和云贵高原的一些自然条件恶劣的地区。实际上当年胡焕庸线的分析就是胡焕庸学生利用当年的数据采集条件,完成的一项大数据分析的结果。只要有数据统计存在,大数据分析就会出现。

从统计结果看(表 1),中国人口密度 ≤ 25 人/ km^2 的地区的人口约占全国的 1.81%,土地面积却占全国的 51.59%,仍然占据了中国一半以上的国土面积;人口密度 ≥ 200 人/ km^2 的地区的人口占全国的 78%,面积却只占 19.8%。更精确地说,大数据的统计可以发现人口密度为 1—25 人/ km^2 低密度区域占了中国的 33.85% 的区域,其次是人口密度为 100—200 人的中密度区,占全国约 12% 的面积。密度达到 500 人/ km^2 以上的高密度区,占中国总面积的 6.7%。人口是聚集的,按经济早期发展的观点^[18],人口聚集是区域经济发展的最初条件。从可持续发展看,区域的生态适应,是调控人口过于聚集的因子。

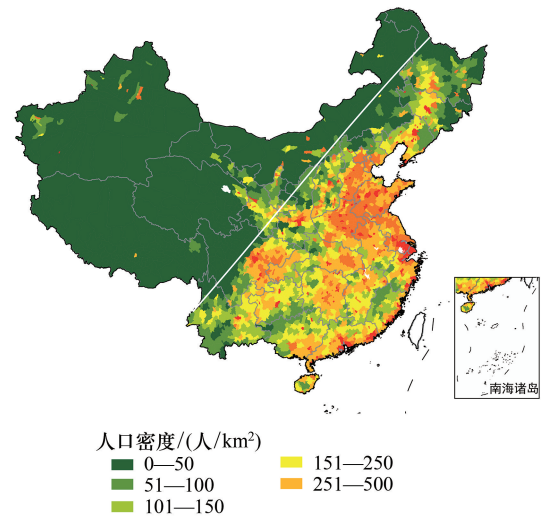


图 2 2010 年中国县级人口密度分布图(缺港澳台数据),白线为胡焕庸线

Fig.2 The population density at county level in China in 2010, The white line is Hu's line

表 1 人口密度各值域范围总人口和面积统计

Table 1 The total population and area of each population density range at county

人口密度值域 Population density range/(人/ km^2)	单元个数 Number of units	总人口数/万 Total population	人口占比/% Population proportion	总面积/ m^2 Total area	面积占比 Area proportion/%
0—1	30	66	0.05	1671447.0	17.74
1—25	228	2249	1.76	3189806.3	33.85
25—50	104	2235	1.75	606990.9	6.44
50—100	278	6872	5.38	922109.1	9.79
100—200	462	16590	12.99	1166725.7	12.38
200—400	461	26538	20.78	939082.7	9.97
400—500	188	13126	10.28	294639.6	3.13
500—1000	383	33574	26.29	492151.3	5.22
1000—	169	26479	20.73	140687.6	1.49

进一步的我们采用了地理学传统的地图对比分析方法,计算给出了中国合成海拔(地带性因子)和水资源适宜度、农业生产潜力等合成的中国第一地理本性图(这里没有使用费生态的矿产条件),它基本上表征的是区域的生态条件(图3),对比图2,图3,可以发现具有生态学意义的合成海拔、水资源适宜度和农业生产潜力确定的中国第一地理本性,决定了中国的人口密度分布。由于气候变化导致地表水资源、农业生产潜力和地带性热量分布都发生变化,因此气候变化驱动着中国的人口分布,这一点得到了吴静,王铮^[3]的证实。在中国的早期发展历史中这一点得到了证实。距今500年前黄河文明出现在一个疏林地带,而当时的温暖湿润气候在安阳、洛阳这些带“阳”字的地方由于太阳辐射的生态优势产生了农业剩余,导致了城市发育。相反的在农业生态条件过于优越的中国南方,由于“江南湿热,丈夫早夭”,在当年的技术水平下,人口适应自然生态的努力不足,人口得不到发展。

花卉^[19]计算得到了第六次人口普查中国人口累计百分比和面积累积百分比的 Lorenz 曲线, Lorenz 曲线它的弯曲程度很大,明显偏离对角线,人口占比占全国的40%时,面积占比却不到6%,与表1所得结论相一致,中国人口的空间分布表现出空间的不均衡性,从第一地理本性表征的生态条件看,这种不平衡是其先天性自然地理基础的。

2.2 人口分布的聚集程度和空间自相关

中国人口分布数据还能通过数据方法挖掘。众所周知,空间全局 Moran's I 指数反映的是人口分布的总体特征。经过花卉^[19]的计算发现中国的全局 Moran's I 为 0.200495。Z 检验值为正,且 P 为 0.05,即在 99.5% 的置信度下,中国的人口密度存在自相关。借助区域的地理本性认识,容易理解,由于第一地理本性的逐渐过渡性,这种人口密度自相关是必然的,连续过渡的第一地理本性,导致人口分布的倾向合适区域聚集,从而表现了人口分布自相关。

LISA 图(图4)是衡量空间单元与周围单元属性的相似和相异程度及其显著性的指标,它反映了四种不同的空间自相关的关系类别:分别为高-高(H-H)、低-低(L-L)、低-高(L-G)、高-低(G-L)。高-高表示研究区及其周围地区的人口密度较高,低-低表示研究区及其周围地区的人口密度较低;高-低表示研究区的人口密度较高而周围人口密度较低,低-高表示研究区人口密度较低而周围人口密度较高。花卉^[19]通过计算中国县级人口密度的局部 Moran's I 值,在通过 Z 值检验并且 $P = 0.05$ 的基础上绘制了 2010 年中国县级人口密度的 LISA 图,如图4。在图4中,高-高聚集主要集中在长三角、珠三角、河南、四川、安徽、河北等地;低-低聚集主要分布在西北、西南、东北等地区,注意到这些地区地理本性的生态因素渐变性,这种空间自相关地理起因是具有生态学起因的。进一步的, LISA 值大的区域大部分位

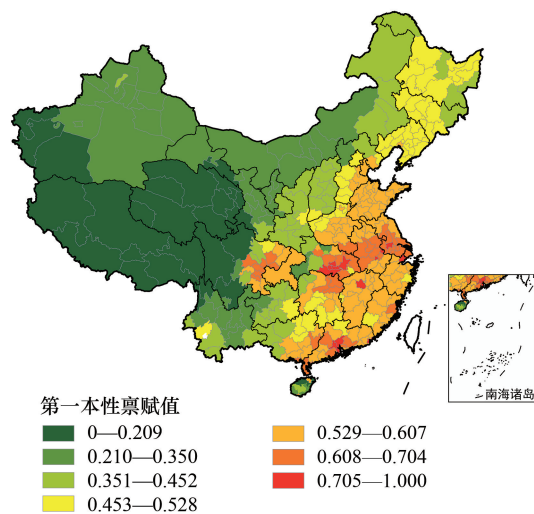


图3 2009年中国的第一地理本性特征分布图

Fig.3 The distribution of the first geographical nature of China in 2009

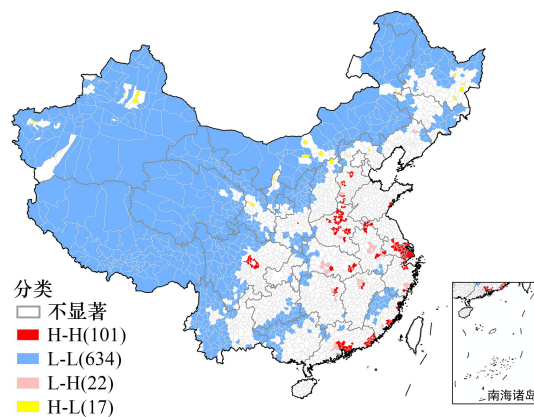


图4 2010年人口密度 LISA 图^[19]

Fig.4 The LISA maps for population density in China in 2010

于胡焕庸线以西,LISA 小的部分位于胡焕庸线东侧,这表明了胡焕庸线意味着地理本性中的生态因素突变性。张丕远,王铮,刘啸雷等^[20]指出大约发生在 1230—1260 年代的气候变化,是全球中世纪温暖期的结束,中国历史时期最大的气候突变,它以降水减少和干旱沙漠化为特征,人口发生大规模迁移,蒙古地区,牧草自燃,中国气温下降约 1°C。迁移人口分布表现的这种空间自相关性,说明了中国区域人口是因经济原因聚集形式存在,人口稀疏的贫困区则因生态原因连片出现,在图 4 中,可以分析胡焕庸线西部人口是低低聚集的,这是中国存在贫困区连片区域地理本性基础^[21]。

3 人口分布的地理-生态因素分析

3.1 地理本性的作用

根据上述的现象学特征,我们进一步通过数据分析,分别探讨区域地理本性的指标和人口密度的空间分布之间的关系。图 5 是在县级行政区基础上得到的 8 个地理本性要素在中国大陆的分布特征(由于缺乏详细的数据资料,图中对我国台湾省的图斑做了非彩色处理)。

将这 8 个特征表征的地理本性的与胡焕庸线展布形式这些特征做直观分析,我们发现,胡焕庸线与农业生产潜力的分布特征最为一致,其次相对海拔和水资源适宜度也是影响影响胡焕庸线的根本性原因。一些人想象的胡焕庸线是中国绝对海拔变化产生的地貌学三个台阶的产物,是错误的。作为自然因素的地形起伏度、环境脆弱性、纬度地带性都影响着胡焕庸线的基本分布。除了海拔因素,这些因素在气候变化下会发生变化,事实吴静,王铮^[21]就发现,胡焕庸线很可能出现于 13 世纪中叶的气候突变.当时,中世纪温暖期结束,中国境内气候大部分地区变干^[23],中国境内人类为了适应自己生态环境的自然变化而呈现出胡焕庸线特征。因此可以说,胡焕庸线是气候变化的产物。当然这个问题值得进一步探讨,在当前气候变化的条件下,这或许是我们讨论胡焕庸线的稳定性起点^[24]。

3.2 影响因子的统计分析

为了进一步胡焕庸线的成因,我们将各个区域的人口分布密度与地理要素做了统计分析,以省区为单元的计算结果见表 2。表 2 数据显示,中国大陆各省的人口密度与这 8 个指标的某些特征存在明显相关关系。这里选入的 GDP 是人均 GDP,它代表了区域发展能力。

从各区域人口密度与 8 个因子的相关性特征看,华东地区的上海、安徽和浙江等东部地区的人口密度与合成海拔和环境脆弱性成较强的负相关关系,华南地区的 4 个省份的人口密度都与合成海拔和环境脆弱性具有较强的负相关关系,在中国北方的华北地区的河北和山西的人口密度与合成海拔和环境脆弱性都存在较强的负相关关系,东北地区的黑龙江、吉林和辽宁的人口密度都与合成海拔具有较强的负相关关系。所有这些说明中国大部分地区人口分布受自然地理的作用大。第一次本性的先天决定作用仍然是重要的,因为相对海拔(含地带性因素)和环境脆弱性,决定的发展的先天差异。进一步研究发现,华中地区的湖北、湖南和河南等中部地区的人口密度与合成海拔和环境脆弱性有较较强的负相关关系,江西的人口密度也与合成海拔具有负相关关系。说明中国人口分布在中国中部地区仍然受到生态意义第一次本性的控制。但是发展内生的第二本性没有从根本上决定区域的人口分布

继续的分析发现,浙江、上海、安徽和江苏人口密度与交通便捷性成较强的正相关关系在这些地区,第二本性的交通便利性根本地强化了区域人口聚集水平和经济密度。华南多区域或地区与 3 个社会经济指标都有较强的正相关关系;这是华南地区长期以来贸易发达、开放活跃的结果。东北地区各省区与 3 个社会经济指标具有正相关关系,说明人类为适应自然而发展的第二地理本性是,在发展意义上是积极的,对东部地区的发展有一定决定作用。

值得注意的是,西南地区较为发达的重庆和四川的人口密度主要与合成海拔和环境脆弱性呈较强的负相关关系,与东部地区一致,但是与农业生产潜力和区位指数具有正相关关系,说明西南地区的发展受到环境的制约本质上与它受到区位的限制和山地环境中农业生产潜力的限制作用具有相似性。其中相对落后的处于

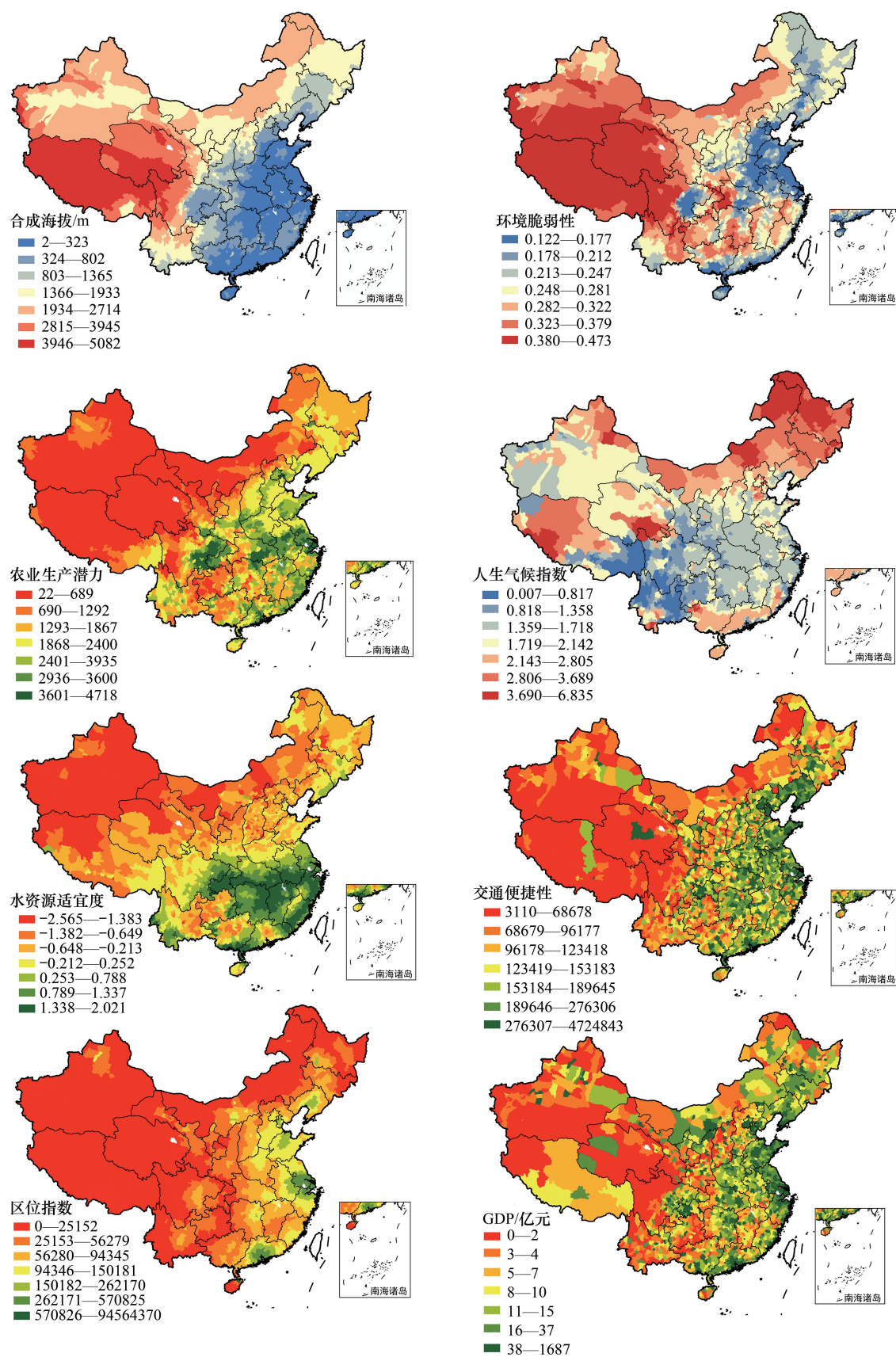


图5 中国大陆县级单位的合成海拔、环境脆弱性、农业生产潜力、人生气候指数、适宜水资源偏离度、交通便捷性、区位指数和 GDP 分布图

Fig.5 The synthesis of altitude, environmental vulnerability, productive potentiality of agriculture, human climate index, appropriate deviation degree of water resources, traffic convenience, locational index and GDP distribute at county level in China

表 2 各省人口密度与各影响因素的相关系数

Table 2 Correlation coefficients of population density and factors in provinces

省份 Provinces	合成海拔 Synthesis altitude	环境脆弱性 Environmental vulnerability	人生气候指数 Human climate index	水资源适宜度 Water resource suitability	农业生产潜力 Agricultural potential production	交通便捷 性指数 Traffic convenience	区位指数 Location	国内生 产总值 GDP
重庆	-0.881 **	-0.853 **	0.546 *	-0.503 *	0.877 **	0.112	0.526 *	0.097
浙江	-0.821 **	-0.771 **	-0.460 **	0.112	0.346 **	0.404 **	0.475 **	0.306 *
云南	-0.223 *	-0.264 **	-0.296 **	0.330 *	0.062	0.478 **	0.479 **	0.466 **
新疆	-0.446 **	-0.374 **	-0.056	-0.312 **	0.404 **	0.301 **	0.194	0.337 **
西藏	0.108	0.421 **	-0.48 **	-0.31 **	0.148	0.281 *	0.562 **	0.491 **
天津	0.170	0.180	0.048	-0.006	-0.063	0.666	0.369	0.667
四川	-0.942 **	-0.660 **	0.172 *	-0.579 **	0.737 **	0.379 **	0.785 **	0.342 **
上海	-1.00 **	-1.00 **	-1.00 **	-1.00 **	1.00 **	1.00 **	1.00 **	1.00 **
陕西	-0.785 **	-0.306 **	-0.051	-0.150	0.484 **	0.380 **	0.687 **	0.273 **
山西	-0.522 **	-0.406 **	-0.282 **	-0.204 *	0.428 **	0.573 **	0.368 **	0.483 **
山东	-0.232 *	-0.004	-0.034	0.088	0.085	0.520 **	0.094	0.449 **
青海	-0.775 **	0.255	-0.461 **	-0.237	0.580 **	0.472 **	0.868 **	0.449 **
宁夏	-0.236	0.395	0.390	-0.586 *	0.184	0.003	0.626 **	0.597 **
内蒙古	-0.366 **	-0.055	-0.396 **	-0.317 **	0.310 **	0.437 **	0.543 **	0.404 **
辽宁	-0.503 **	-0.339 **	-0.087	0.189	-0.108	0.641 **	0.596 **	0.475 **
江西	-0.529 **	-0.115	0.055	0.069	0.142	0.557 **	0.700 **	0.648 **
吉林	-0.552 **	-0.366 *	-0.116	0.195	0.298 *	0.554 **	0.575 **	0.459 **
湖南	-0.617 **	-0.462 **	0.171	-0.397 **	0.132	0.731 **	0.691 **	0.622 **
湖北	-0.746 **	-0.621 **	0.267 *	-0.314 **	0.599 **	0.417 **	0.610 **	0.256 **
黑龙江	-0.619 **	-0.249 *	-0.365 **	0.060	0.491 **	0.479 **	0.413 **	0.370 **
河南	-0.419 **	-0.352 **	0.346 **	0.285 **	-0.250 **	0.509 **	0.571 **	0.475 **
河北	-0.765 **	-0.700 **	-0.057	-0.049	0.520 **	0.495 **	0.649 **	0.369 **
海南	-0.732 **	-0.612 **	-0.503 **	-0.242	0.200	0.809 **	0.550 *	0.691 **
贵州	0.497 **	-0.002	-0.472 **	0.421 **	-0.423 **	0.456 **	0.744 **	0.661 **
广西	-0.545 **	-0.619 **	-0.049	-0.14	0.181	0.622 **	0.632 **	0.569 **
广东	-0.718 **	-0.562 **	-0.344 **	-0.200	0.397 **	0.487 **	0.517 **	0.458 **
甘肃	-0.496 **	-0.372 **	-0.402 **	-0.491 **	0.303 **	0.290 **	0.333 **	0.249 *
福建	-0.743 **	-0.579 **	-0.553 **	0.112	0.367 **	0.571 **	0.606 **	0.674 **
安徽	-0.603 **	-0.578 **	0.149	0.499 **	0.037	0.456 **	0.326 **	0.677 **
北京	-0.675	-0.886	-0.836	0.496	-0.695	0.983	0.974	0.986
江苏	-0.082	-0.249 *	-0.205	-0.255 *	0.229	0.364 **	0.607 **	0.610 **

* * 表示在 0.01 的水平上相关, * 表示在 0.05 的水平上相关

山地环境的西藏、云南和贵州的人口密度与区位指数和 GDP 具有正相关关系,人口密度主要受区位和经济的影响较大,应该说是山地环境的海拔、地表崎岖度,严重地影响了发展水平,只有靠第二本性的改变来发展。类似的西北地区的陕西和青海的人口密度主要与合成海拔具有较强的负相关关系,与区位指数具有较强的正相关关系。甘肃的人口密度与除了农业生产潜力以外的第一次本性指标都具有一般的负相关关系,整个西北地区的人口密度还与第二本性的 3 个社会经济指标和农业生产潜力具有正相关关系,说明农业生产潜力在我国西部的重要性。新疆的人口密度与合成海拔、水资源适应度具有突出的负相关关系,与农业生产潜力、交通便捷性和 GDP 具有正相关关系;宁夏的人口密度与适宜水资源偏离度具有较强的负相关关系,与区位指数和 GDP 具有正相关关系,说明西北地区人口密度主要受水资源和区位因数的影响。

事实上,各省份的人口密度与第二本性的三个社会经济指标都有较强的正相关关系,说明人口聚集的基础是三个社会经济指标反映的第二地理特征。第一次本性决定了区域发展的生态条件,其中海拔、水资源和

农业潜力具有突出的意义。第二本性的因素推动着人口与经济发展,其中交通和区位因素在某些地区具有突出意义。

值得注意的是,北京、上海和天津与 8 个因子都不具有显著的相关关系,这可能由于其他因素的影响,例如国家政策、地缘政治地位和海运条件等,当然也可以说人口聚集到一定规模后,人口密度与两个本性的联系不再显著,是否存在完全无生态学意义的第三本性的作用,如信息化水平,这还需进一步对其进行研究。

从整体上看,通过计算表 2 中显著性水平为 0.01 的各省人口密度与合成海拔、环境脆弱性、人生气候指数、适宜水资源偏离度、农业生产潜力、交通便捷性、GDP 和区位指数的相关系数的平均值,分别为-0.53、-0.50、-0.40、-0.23、0.42、0.51、0.51、0.60。可以认为区位指数、合成海拔、交通便捷性、GDP 和环境脆弱性与人口密度具有较强的关系,对人口空间分布具有较强的影响。整体上看,正相关关系最大的是区位指数,负相关关系最大的是合成海拔,其次是水资源偏离度。合成海拔代表的自然地理条件是难以改造的,但是水资源偏离度,特别是区位指数是具有明显可调节性的。这是中国未来国土建设必须注意的地理学特征。

基于以上表征区域生态与地理特征的 8 个指标,计算得到地级市尺度的生态环境特征,它代表环境的宜居-艰苦性。宜居和艰苦性是在人口学争执不休的概念,这里作为我们用它反映相应区域人口生态意义及区域适应人类生存的可能性。计算并采用自然断点法将艰苦性指数分为 8 大类,它反映如图 6 所示。从 I 类到 VIII 类,颜色越来越深,表示从生态条件看该区域越来越不适宜居住。从图 6 中可以看出,胡焕庸线以西的区域多为 VII 类和 VIII 类,胡焕庸线经过的区域以 V/VI 类为主,由此可见胡焕庸线分割了中国人口生态宜居区和环境艰苦区,胡焕庸线以西的云南、贵州全省和东北与内蒙古东部地区几乎都落在生态环境较为艰苦的 V 区,或者更高,是生态环境的过渡地带^[20]。值得引起重视的是,沿胡焕庸线,基本上是人口生态环境艰苦的地区。

为了大量数据计算,我们将省区内各地级市人口分布密度与各地级市的按面积加权的各地理要素值,利用各以逐步回归分析法,分析影响人口分布省区的显著相关的因子。在用 SPSS 进行逐步回归分析时,各因子进入回归方程的显著性 p 值设为 0.05,剔除的显著性 p 值设为 0.1。

通过这部回归在地级市为数据基础的分析分析,影响各个区域人口分布的因素,重庆、浙江、云南、四川、陕西、山西、吉林、湖北、河南、河北、海南、广东和福建等 14 个省份的贡献率第一、第二的影响因子都包含合成海拔,这 14 个省区再加上典型山地省区西藏、青海、贵州,占了中国省区数的一半多,表明在其他条件相同的情况下,海拔、地形起伏度和纬度等地带性因子,显著影响了我国的人口密度分布乃至经济,这就揭示了我国作为山地国家的存在基本地理特征。上海、辽宁、湖南、江西、江苏、浙江和广西等省份的海拔不是主要的影响因素,这些省区的交通便捷性对人口密度差异贡献率最大,其省份内部交通便利程度决定了各个省内部的人口分布状况。值得注意的是青海、江西和贵州等 3 个省份的区位指数贡献率是最大的,说明这些相对落后的地区的人口分布受经济区位的影响较大,需要发展城市经济。各县市与中心城市或者省会城市的联系以及它们在全国空间分布位置影响了这些省份人口的空间密度。加强这些区域的各县市与周边发展较好的城市的经济交流而不仅仅是大陆交通联系,是促进它们城市经济的发展是关键。数据还揭示,广东、福建、河南、湖北的水资源丰富度的影响都是负向因子,反映这些地区,水资源过于丰富意味着洪涝发生,影响了人口分布。但是水资源供应不足,是影响区域农业生产潜力的重要因素。典型的是甘肃,甘肃省贡献率第一的因

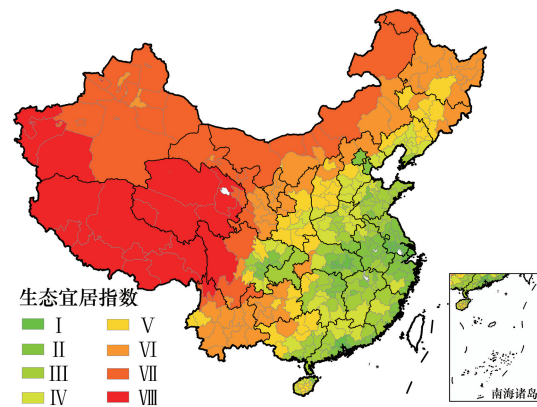


图 6

Fig. 6 The distribute map of ecological livability index at prefecture level in China

子是适宜水资源偏离度,表明甘肃省内部水资源分布差异较大,各地区的水资源分布的差异影响了甘肃各地区的人口密度,甘肃省属于我国水资源较少的省份,又属于农业经济区,因此其人口问题,较我国水资源不足的新疆和宁夏相比就更加突出。这就是说在胡焕庸线以西的北部,更应该对水资源设施进行优化配置,并进行合理引水,这或许是水利学家们提出“红旗河”的意义所在之处。

值得注意的是,影响西藏人口密度前二位的因子而不是海拔,这是很正常的,因为西藏、青海、贵州地区能够影响人口分布的海拔内部差异已经变小,在人口总体分布格局形成的条件下,西藏、青海内部的气温、湿度和水资源条件等自然环境条件差异影响了青藏高原地区各县市的人口密度。青海、云南、贵州由于城市的地貌的封闭性使得区位影响特别突出。可持续发展的关键应该促进人口在人生气候指数较好的地区的聚集。

作为中部地区代表的安徽省贡献率第一位的因子是合成海拔,GDP 排在第二位。可见安徽地区在已经突破了海拔影响的限制基础上,需要突破的是人均 GDP。这种情况在逐步的河南省、湖南省等同样存在,这说明通过聚集来提高人均 GDP 是逐步地区发展的关键。

在以每个县委数据采样单元基础上分析发现,北京、上海和天津跟每一个影响因子的关系都不显著(重庆的显著影响因子是合成海拔及人生气候),究其原因,主要是北京、上海和天津属于平原地区直辖市,它们已经是中国人口最密集的地区,人口密度的影响因子更为多样化,比如信息化程度可能已经影响了这个分布,其中天津武清的兴起就是一例。

4 结论

本文借助大数据分析的思路和数据挖掘方法,以中国人口分布密度为典型对象,借助 Krugman 的新经济地理学的地理本性观点,对中国区域地理特征和胡焕庸线的存在性做了再分析,发现了下列规律:

(1)中国人口的空间分布非常不均衡;中国人口密度存在一定的自相关性,高-高聚集主要集中在长三角,珠三角,河南,四川、安徽、河北等地,低-低聚集主要分布在西北、西南、东北等地区,大部分位于胡焕庸线以西,少部分位于胡焕庸线东侧。沿胡焕庸线是一个生态环境过渡带

(2)胡焕庸线形成的地理学基础最直接的是农业生产潜力,即陆地生态系统原始生产力,驱动农业生产潜力的主要是气候变化,换言之,胡焕庸线可以认为气候变化的产物。突破胡焕庸线的关键在于适合由于气候变化带来的农业生产潜力(即陆地生态系统原始生产力)变化,关于这个变化,王铮等^[24]给出了一个初步研究,在气候变化新的动态下,这个问题需要更多研究。

(3)在第一本性指标中,合成海拔对各省人口密度的影响贡献率排在第一位的省份最多,有 13 个,考虑到西藏、青海、云贵高原,总体海拔高,我国的山地国家特征突出。其次,可以认为区位指数、交通便捷性、GDP 和环境脆弱性与人口密度具有较强的相关关系,对人口空间分布具有较强的影响。其中,合成海拔和环境脆弱性具有较强的负相关关系,而交通便捷性、区位指数和 GDP 与人口密度具有较强的正相关关系。

(4)逐步回归分析的结果显示,在四川、青海、陕西、宁夏、辽宁、江西、江西、云南、贵州、内蒙古、广西、甘肃、安徽等省区,第二本性因素的区位和交通便利性是影响人口发展前两位因素,因此需要城市化来推动聚集和交通发展。而新疆、山东、宁夏、内蒙古、黑龙江、江苏、北京、天津和重庆的人口分布比较独特,这些指标不能很好的刻画这些省份的人口分布规律,所以针对这几个省份还需引入新的因子来刻画人口分布的影响因素。

(5)胡焕庸线沿线是中国的生态环境过渡带。

致谢:本文第一作者感谢过去 12 年,先后参加本项目研究的研究生等刘清春、戴晓晔、吴静、田园、刘涛、刘筱、薛俊波、顾高翔、汪晶、翟石艳、余芳琳、蒋欣、杨念、利果、张焕波、李山、赵金彩等,以及国家公务员局、新疆维吾尔自治区人事厅、宁夏回族自治区人事厅、云南省人事厅、四川省人事厅、宁夏回族自治区人事厅、黑龙江省人事厅、青海省人事厅、甘肃省人事厅、广西壮族自治区人事厅、海南人事厅及云南省民委、中国农业大学、云南农业大学、西北师

范大学、中科院地理所、中科院寒旱所、中科院情报中心、中国铁道总公司等单位人员刘丽、王远飞、李秀山等,他们为大数据的获取做出了贡献,在此一并致谢。

参考文献(References):

- [1] 胡焕庸. 中国人口之分布. 地理学报, 1935, 2(1): 33-74.
- [2] 王铮, 张丕远, 刘燕华. 中国生态环境过渡的一个重要地带. 生态学报, 1995, 15(3): 319-326.
- [3] 吴静, 王铮. 2000 年来中国人口地理演变的 Agent 模拟分析. 地理学报, 2008, 63(2): 185-194.
- [4] Lovemore Z, Richard W. Changing patterns of population distribution in Zimbabwe. *GeoJournal*, 1986, 13(4): 365-384.
- [5] 封志明, 唐焰, 杨艳昭. 中国地形起伏度及其与人口分布的相关性. 地理学报, 2007, 62(10): 1073-1082.
- [6] 方瑜, 欧阳志云, 郑华. 中国人口分布的自然成因. 应用生态学报, 2012, 23(12): 3488-3495.
- [7] 刘青春, 王铮. 中国区域经济差异形成的三次地理要素. 地理研究, 2009, 28(02): 430-440.
- [8] 夏海斌, 王铮. 中国大陆空间结构分异的进化. 地理研究, 2012, 31(12): 2123-2138.
- [9] 王铮, 夏海斌, 吴静, 刘青春, 戴小晔, 刘丽, 田园. 中国艰苦边远地区辨识(2006-2016), 国家人力资源于社会保障部(内部报告), 2006-2016
- [10] Paul Krugman. Increasing Returns and Economic Geography. *The Journal of Political Economy*, 1991, 99(3): 483-499.
- [11] Burrough P. Principles of Geographical Information Systems for Land Resources Assessment. *SOIL SCI*, 1987, 144(4): 306.
- [12] 张岳军, 郝智文, 王雁. 基于 SPEI 和 SPI 指数的太原多尺度干旱特征与气候指数的关系. 生态环境学报, 2014, 23(09): 1418-1424.
- [13] 王铮, 孙枫, 王瑛, 吴慧娟, 刘丽. 知识型产业区位的实证分析. 科研管理, 1999, 20(03): 101-108.
- [14] 刘青春, 王铮, 许世远. 中国城市旅游气候舒适性分析. 资源科学, 2007, 29(01): 133-141.
- [15] 钟章奇, 王铮, 夏海斌. 全球气候变化下中国农业生产潜力的空间演变. 自然资源学报, 2015, 30(12): 2018-2032.
- [16] 夏海斌, 戴霄晔, 王莹, 王铮. 基于 GIS 的中国县级尺度交通便利性分析. 地域研究与开发, 2006, 25(03): 120-124.
- [17] 夏海斌. 基于自主体建模的中国区域结构进化分析. 华东师范大学博士论文, 2013.
- [18] Goodfriend, M., McDermott, J. Early development, *The American economic review*. 1995, 85: 116-133.
- [19] 花卉. 基于地理本性的国人口空间分布的影响因素及其合理性, 华东师范大学硕士论文, 2018.
- [20] 王铮, 张丕远, 周清波. 历史气候变化对中国社会发展的影响——兼论人地关系. 地理学报, 1996, 63(4): 329-339.
- [21] Tian Y, Wang Z, Zhao J, Jiang X, Guo R. A Geographical Analysis of the Poverty Causes in China's Contiguous Destitute Areas, *Sustainability* 2018, 10: 1895-1897.
- [22] Wu J, Mahamod, R, Wang, Z. Agent-based simulation of the spatial evolution of the historical population in China, *J. History Geography*, 2011, 37(1): 1-10
- [23] 张丕远, 王铮, 刘啸雷, 张时煌. 中国近 2000 年来气候演变的阶段性. 中国科学(B 辑 化学 生命科学 地学), 1994(09): 998-1008.
- [24] 王铮, 乐群, 夏海斌, 孙翊, 钟章奇, 徐程瑾, 顾高翔, 王远飞. 中国 2050: 气候情景与胡焕庸线的稳定性. 中国科学: 地球科学, 2016, 46(11): 1505-1521.