

冯喆, 蒋洪强, 卢亚灵. 基于大数据方法和SOFM聚类的中国经济-环境综合分区研究[J]. 地理科学, 2019, 39(2): 242-251. [Feng Zhe, Jiang Hongqiang, Lu Yaling. China's Economic-environment Comprehensive Zoning Based on Big Data Method and SOFM Clustering. Scientia Geographica Sinica, 2019, 39(2): 242-251.] doi: 10.13249/j.cnki.sgs.2019.02.008

# 基于大数据方法和SOFM聚类的中国 经济-环境综合分区研究

冯喆<sup>1,2</sup>, 蒋洪强<sup>1</sup>, 卢亚灵<sup>1</sup>

(1. 环境保护部环境规划院国家环境保护环境规划与政策模拟重点实验室, 北京 100012; 2. 中国地质大学(北京) 土地科学技术学院, 北京 100083)

**摘要:** 研究使用经济和环境多源大数据, 建立包含人口、GDP等经济指标和空气质量等环境指标的中国经济-环境关联体系, 识别各指标的热点、冷点时空变化特征, 采用人工神经网络聚类方法对中国现阶段经济-环境进行综合分区。研究结果如下: ① 灯光平均强度较高的省份主要集中在沿海地区, 经济以长三角、珠三角和环渤海区域为主要拉动引擎, 呈东南高西北低的发展态势, 东部沿海地区经济发展优于东北、中部和西南地区。② PM<sub>2.5</sub>浓度呈现先增后减趋势, 高污染区主要集中在华北、华中等区域; 东北方向逐步扩散, 污染热点地区从辽东半岛、山海关一线向东北扩张; 南方地区基本保持稳定态势。③ 采用自组织特征映射模型对2015年全国各城市OLS灯光指数、人口、城市自然边界和年均PM<sub>2.5</sub>浓度4类指标进行聚类, 第I类为经济极发达-环境恶化地区, 主要位于华北平原和长江三角洲; 第II类为经济发达-环境趋恶化地区, 主要分布在第I类区域周边, 特别是京津冀周边地区; 第III类为经济发达-环境良好地区, 广东、海南、江西、福建以及重庆等省市多属此类型; 第IV类为经济不发达-环境优质地区, 主要分布于东北地区北部、内蒙古、甘肃、贵州、新疆、青海、西藏等地。

**关键词:** 经济-环境关联体系; 夜间灯光指数; PM<sub>2.5</sub>; 自组织特征映射模型(SOFM)

**中图分类号:** S601.9      **文献标识码:** A      **文章编号:** 1000-0690(2019)02-0242-10

改革开放以来, 中国经济取得了快速发展, 但也因此付出了沉重的环境代价。部分地区经济发展与生态环境保护脱钩, 表现为东部沿海发达地区如华北、华东等大城市发展迅速, 造成了城市热岛效应显著, 污染严重, 环境恶劣<sup>[1-5]</sup>。而中西部欠发达地区, 由于资源过度开发, 造成污染围城、土壤肥力下降、耕地砂砾化、水土流失等<sup>[6-8]</sup>, 已经严重影响了区域发展的完整和均衡。日趋紧迫的资源供应压力、环境保护压力迫使国家由传统的粗放型经济增长方式向环境友好的经济发展方向转型<sup>[9,10]</sup>。同时, 中国地域辽阔, 地区间经济社会发展水平不均衡, 环境问题的种类和紧迫性也有差异, 这就要求政府部门实施有针对性的管理措施。因此, 如何充分利用现有信息, 准确描述地区经济发

展水平和环境质量, 成为学者关注的热点问题。

过去10 a来, 大数据已成为全球关注的焦点, 日益引起学界、行业、政府等组织的关注。“大数据”一词首次出现在20世纪90年代中期, 自2008年逐渐开始流行, 并于2010年开始得到认可。如今, 大数据以庞大的规模为基本特征, 数据类型多样, 传输及时迅速, 挖掘潜力大, 被认为是创新、竞争和生产力的主要来源, 并引起了向数据驱动的范式研究转变<sup>[11-13]</sup>。大数据方法已经成为反映区域环境质量和经济发展的重要手段, 在资源管理、环境保护、政策制定等研究领域得到广泛应用<sup>[14,15]</sup>。大数据方法的发展, 为科学客观的度量区域环境与经济发展间关联, 更为有效地实现经济-环境综合分区提供了方法论基础。

**收稿日期:** 2018-01-29; **修订日期:** 2018-04-21

**基金项目:** 国家环境保护环境规划与政策模拟重点实验室开放基金(ZDSYS201701)、国家自然科学基金项目(71603097, 71433007, 41771204)资助。[Foundation: Open Foundation of Key Laboratory of National Environmental Protection Environmental Planning and Policy Simulation (ZDSYS201701), National Natural Science Foundation of China (71603097, 71433007, 41771204).]

**作者简介:** 冯喆(1984-), 男, 北京人, 讲师, 博士, 主要研究方向为地理大数据。E-mail: zhefeng@cugb.edu.cn

**通讯作者:** 蒋洪强, 研究员。E-mail: jianghq@caep.org.cn

尽管大数据方法在环境形势分析中具有重要意义,但其在经济与环境综合分区研究中的应用尚待挖掘。本研究以人类社会经济活动与生态保护之间的平衡作为目标,使用经济和环境多源大数据,构建包含人口、GDP等经济指标和空气质量等环境指标的中国经济-环境关联体,采用人工神经网络聚类的方法对2015年经济-环境综合体进行分区,针对不同区域提出相应的政策建议,有助于区域的环境与经济协调发展,为制定有针对性的环境政策提供参考。

## 1 数据与方法

### 1.1 数据来源与处理

本研究采用中国灯光数据作为区域经济发展的表征(不含港澳台数据)。夜间灯光数据可以更多地反映人类活动情况,已被验证与一国的GDP、人口等社会因子间存在较强的正相关关系<sup>[6]</sup>,在能源消耗、碳排放等指标的估算以及城市化监测等领域获得广泛应用<sup>[17-19]</sup>。

本研究的数据来源为美国国防气象卫星计划(Defense Meteorological Satellite Program, DMSP)系列卫星搭载的业务型线扫描传感器(Operation- al Linescan System, OLS)及苏奥米国家极轨合作伙伴(Suomi NPP)可见光红外成像辐射仪(The Visible Infrared Imaging Radiometer Suite, VIIRS)的夜间灯光数据。其中,OLS数据目前发布了1992~2013年的全球数据,VIIRS数据(<http://earth-data.nasa.gov/earth-observation-data/near-real-time/download-nrt-data/viirs-nrt>)发布了2012年4月至今的全球数据。在本研究中,2000~2010年3期使用夜间灯光数据源自OLS数据集(<http://ngdc.noaa.gov/eog/dmsp>),是经NOAA-NGDC科学家的除尘、去噪、全年平均等处理后获得的;2015年数据源自VIIRS提供的2015年年度数据,但此数据未经过滤机噪声去除,参考Dai等<sup>[20]</sup>的处理方法,以2015年北京、上海、广州和深圳市区灯光最大值为阈值,对数据进行处理矫正。在数据的分析统计中,下文采用的Getis-Ord( $G_i^*$ )指数计算及SOFM分类过程均不受数据绝对值影响,因此两套不同来源数据不会影响空间分析及聚类结果。以上两套数据集均由NOAA-NGDC网站提供影像下载,空间分辨率为1 km。

人口的空间分布是表征区域发展水平的重要

社会因素。本研究人口的空间数据由美国国家橡树岭实验室全球人口项目制作提供的Landscan数据采集,空间分辨率为1 km。Landscan绝对人口数值不可用于年际比较,但可用于空间结构对比。

城市自然边界从形态学角度衡量区域城市化发展,数据集由OpenStreetMap(OSM)数据库(<http://www.openstreetmap.org>)提供,该数据库采用GPS踪迹和未经版权保护的卫星图像产生地理空间数据,从OSM数据中可提取街道节点和街区,经与实际验证精度较高。

本研究采用PM<sub>2.5</sub>数据表征区域环境质量情况,数据源自van Donkelaar等通过大数据方法模拟的中国PM<sub>2.5</sub>年均浓度<sup>[21]</sup>。所有数据经重采样处理后空间分辨率为1 km。

### 1.2 研究方法

本研究的技术路线如图1所示,首先,通过整合夜间灯光数据、城市边界拟数据、人口统计数据 and 空气质量数据,构建基础数据集。其次进行时空格局分析,分别进行空间制图后,利用 $G_i^*$ 指数分析各指标的冷热点及空间转移情况,最后使用SOFM人工神经网络方法,选择环境和经济指标进行聚类,形成以省级行政区为评价单元的经济-环境综合分区。

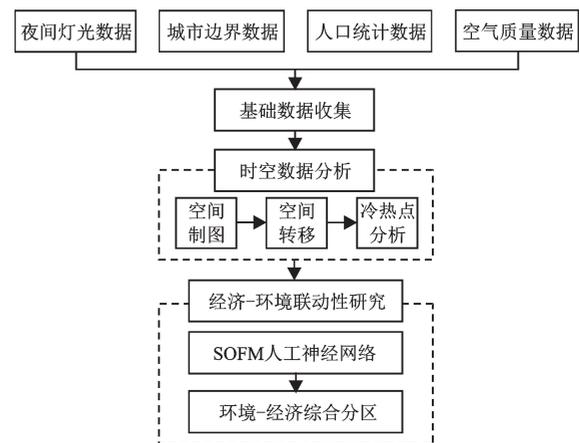


图1 技术路线

Fig.1 Technical framework

#### 1.2.1 空间集聚分析

本研究采用Getis-Ord( $G_i^*$ )识别具有统计显著性的高值(热点)和低值(冷点)的空间聚类。其中,具有显著统计学意义的热点,要素本身应具有高值,且被其他同样具有高值的要素所包围;冷点亦然。Getis-Ord( $G_i^*$ )的局部统计可以表示为:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - \bar{X} \sum_{j=1}^n w_{ij}}{S \sqrt{\frac{[n \sum_{j=1}^n w_{ij}^2 - (\sum_{j=1}^n w_{ij})^2]}{n-1}}} \quad (1)$$

式中,  $x_j$  是要素  $j$  的属性值,  $w_{ij}$  是要素  $i$  和  $j$  之间的空间权重,  $n$  为要素总量, 且:

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (3)$$

$G_i^*$  值就是  $z$  得分。对于具有显著统计学意义的正的  $z$  得分,  $z$  得分越高, 高值(热点)的聚类就越紧密。对于统计学上的显著性负  $z$  得分,  $z$  得分越低, 低值(冷点)的聚类就越紧密。

### 1.2.2 自组织特征映射模型

自组织特征映射模型(Self-Organizing Feature Map, SOFM)是一种非监督型的人工神经网络。SOFM学习规则是一种与大脑皮层竞争(抑制)相似的算法。竞争学习中每一时刻只有一个输出单元激活, 或每组输出单元中只有一个激活, 输出单元为了激活而竞争, 聚类中心映射到一个曲面或平面上, 并且保持拓扑结构不变, 因此, 能根据学习规则对输入的模式自动进行分类, 降低了指标和权重中的主观性, 提高了分类的客观性和准确性, 在地理学和生态学领域, 尤其是数据结构复杂、数据量大和非线性特征明显的问题上得到广泛应用<sup>[25,26]</sup>。

SOFM网络学习过程由以下几步构成: ① 权值初始化, 用小的随机数对各权向量赋予初值, 各节点权值应取为不一样的。② 在样本集中随机选择一个样本  $x$  作为输入。③ 在时刻  $t$ , 选择最佳匹配单元  $i$  (竞争过程)。这里是选定输入向量  $x$  与所有权向量之间的最不相似者作为获胜单元, 用欧氏距离表示, 其中  $c$  表示获胜单元, 则有:

$$\|x - w_c\| = \min\|x_i - w_i\| \quad (4)$$

## 2 研究结果与分析

### 2.1 中国经济要素时空变化分析

#### 2.1.1 灯光强度时空分布变化分析

通过计算  $G_i^*$  值, 得到 2000~2015 年经济要素

空间聚集和转移情况(图2)。2000年, 经济要素的主要热点区域在长三角地区、珠三角地区, 此外在北京及山东半岛也有分布; 广大中部内陆地区, 以及黑龙江等地则表现为冷点。至2005年, 经济热点地区在原有基础上逐步扩张, 除原有的长三角、珠三角地区外, 环渤海区域也表现出了明显的经济热点; 相对来说, 经济冷点区域变化不大。2010年, 随着经济的进一步发展, 东部沿海地区经济热点区域呈现出集中连片态势, 环渤海、长三角、珠三角3个经济增长极拉动作用明显。至2015年, 灯光强度已呈现集中连片态势, 内陆地区的强度提升明显。

#### 2.1.2 人口时空分布变化分析

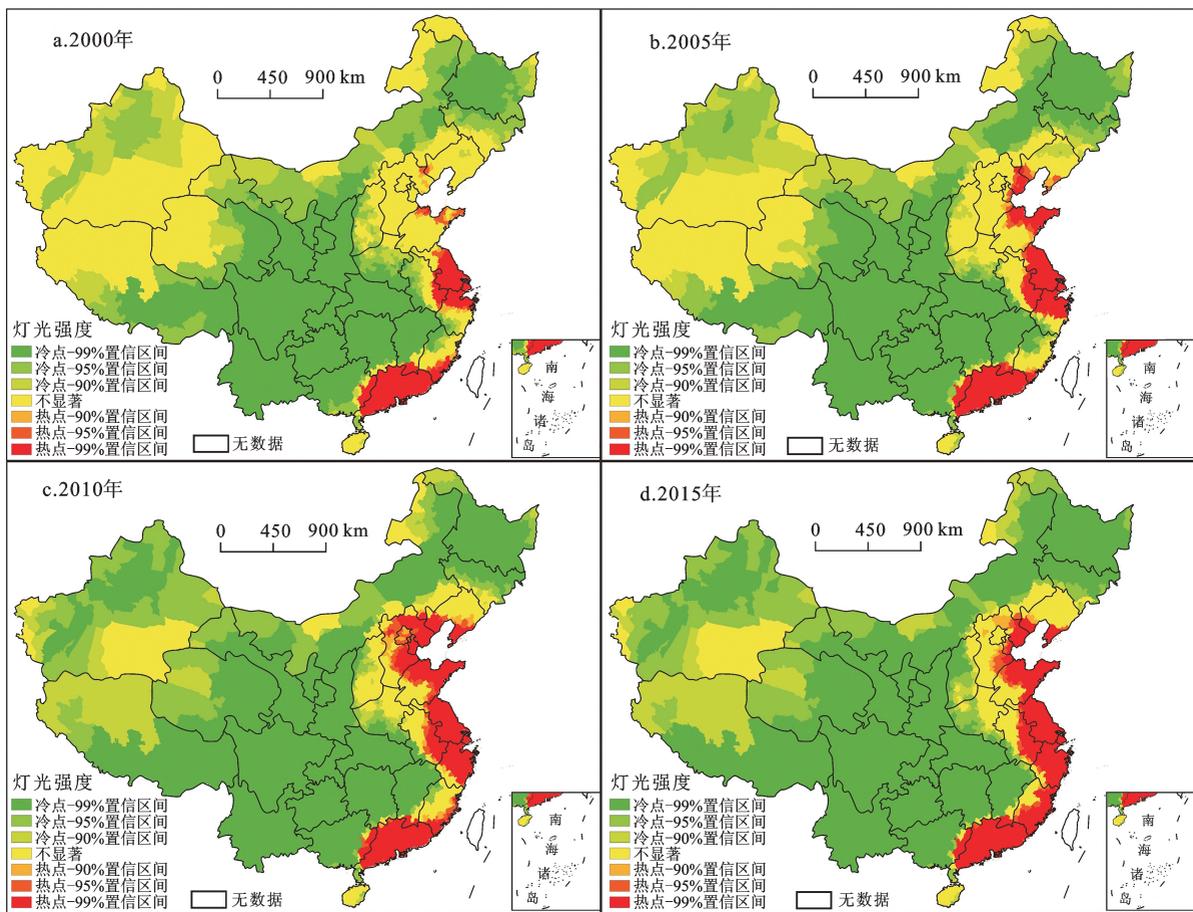
2000~2015年, 中国人口分布依然遵循胡焕庸线以东人口密集, 胡焕庸线以西则人口较为稀疏的规律。图3显示了用  $G_i^*$  值代表的2000~2015年人口聚集和空间转移情况, 2000年, 人口聚集的热点区域主要集中在华北、华中地区以及四川盆地。2005年起, 四川盆地人口集聚效应减弱, 人口进一步向华北和华中区域转移。至2015年, 人口分布热点地区在2005年的格局上向东南沿海方向扩展, 但总体态势并未发生剧烈变化。

#### 2.1.3 经济要素时空变化综合分析

综合上述2个指标, 可以反映出中国经济格局的时空变化特征。自2000年起, 随着加入世界贸易组织和西部大开发、东北老工业基地振兴、中部崛起等一系列区域协调发展战略的实施, 经济空间格局东、中、西部均衡程度有所提高, 但受历史基础、区位特征及发展政策的影响, 经济重心始终在东部沿海地区。随着开放的不断深化, 各类经济要素流动愈发顺畅, 在一定程度上强化了东部沿海地区经济吸纳能力。灯光强度和人口指标都反映了这一趋势。2000~2015年, 东部沿海的经济热点区和人口聚集区面积都呈现显著扩张态势; 相对而言, 中西部地区的变化幅度较小, 部分地区(如四川盆地)人口聚集程度出现下降趋势, 经济发展均衡程度还有待进一步加强。

## 2.2 中国各地市环境要素时空变化分析

中国幅员辽阔, 环境类型复杂多样。考虑数据可得性和代表性, 本研究选取近年来的研究热点大气污染中的PM2.5年均浓度作为环境要素指标代表。PM2.5污染空间聚集程度较高的区域位于华北和华中地区, 而广大西部地区污染程度较

图2 2000~2015年中国灯光指数 $G^*$ 集聚Fig.2 Light index  $G^*$  gathering in China from 2000 to 2015

低。2000~2015年,PM<sub>2.5</sub>浓度呈现先增加后平稳的趋势(图4),2000~2005年,PM<sub>2.5</sub>浓度呈现增长的趋势,污染区域逐渐向华南一带扩散,污染高值区最南端从湖南北部、江西北部、浙江北部一线扩散至广西北部、湖南全境、江西西北部;同时,山西南部地区逐渐减弱,而四川东部、重庆一带则逐渐变化高值区域;与2005年相比,2010年PM<sub>2.5</sub>污染保持基本平稳态势,高污染区仍主要集中在华北、华中区域。东北地区高值区域不断扩散,从2000年的辽东半岛、山海关一线逐步向东北方向扩展,至2015年,已经扩散到辽宁全境、吉林中部乃至黑龙江南部部分地区。从总体趋势看,华北、华中区域始终为污染热点区域;东北方向逐步扩散,污染热点地区从辽东半岛、山海关一线向东北扩张;热点区西线变化较大,从2000年的山西南部向南移动至2010年的四川东部,至2015年逐步移出四川;南线在2000年位于湖南北部、江西北部、

浙江北部一线,2005~2015年已经逐渐南移进入广西、广东北部地区。PM<sub>2.5</sub>污染受到经济发展、自然条件等因素共同作用。随着对环境重视程度的提高,污染集中的区域有收缩的趋势,中部地区污染改善的程度较高,而污染集中区域向东北方向移动,一方面与当地的产业结构可能存在一定关系,另一方面也有其他地区污染治理步伐更大的因素。

### 2.3 基于SOFM人工神经网络的各地市经济-环境分类

本研究使用Matlab 2017a作为SOFM网络聚类计算工具。使用ArcGIS 10.5在研究区内随机选取50 000个样点进行分析,剔除边缘空白样点后得到49 613个样点。将每个样点2015年的OLS灯光指数、人口、城市自然边界、年均PM<sub>2.5</sub>浓度5类指标作为图层输入网络,训练次数为1 000次。为保持类别间具有较大差异性,经反复测试,确定类

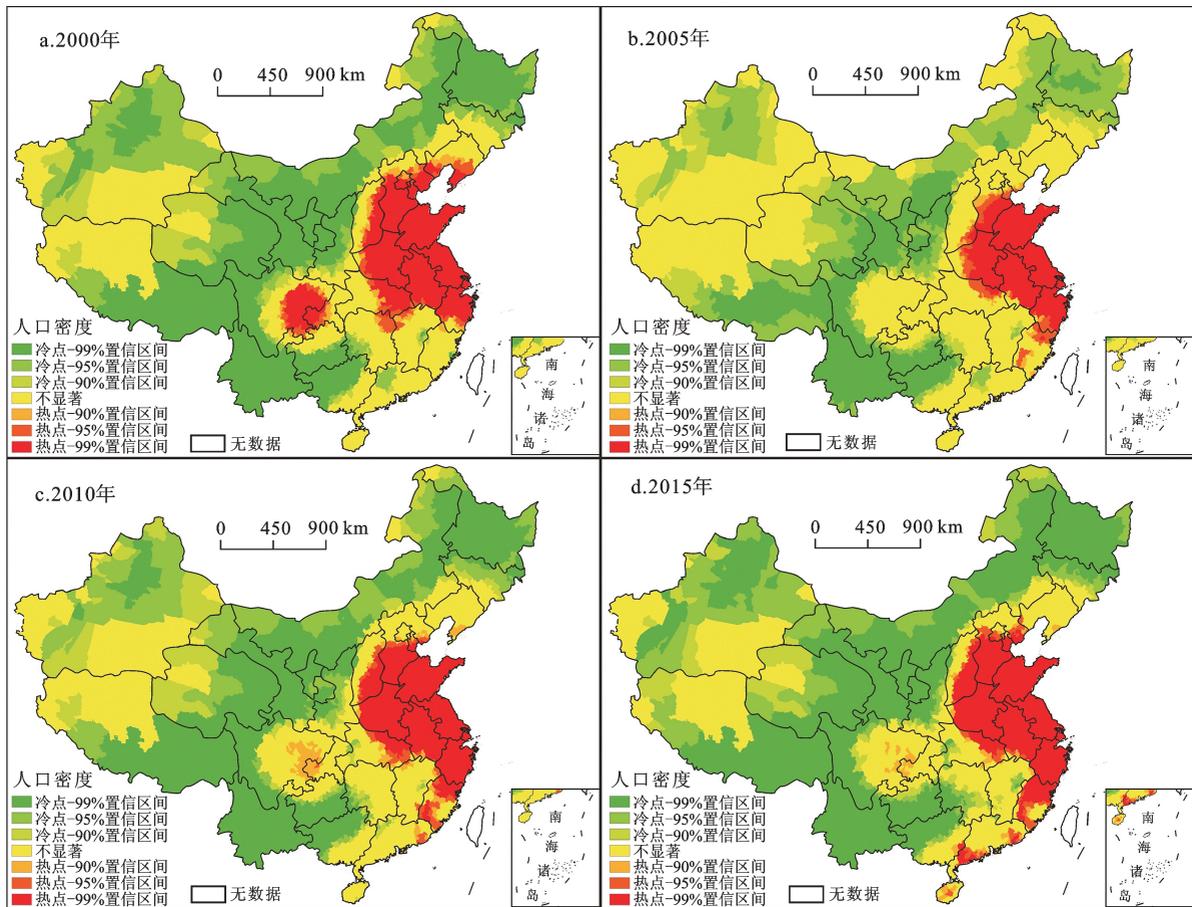


图3 2000~2015年中国人口分布G\*集聚

Fig.3 The G\* gathering of population distribution in China from 2000 to 2015

别数量为3类。

SOFM 聚类将全国抽样点分为3类,并按照欧氏距离将其他栅格赋予最近邻样本点的类别。按类别统计3类地区经济-环境特征(表1),第I类为城市地域比例高、灯光强度高、PM2.5浓度高、人口稠密的经济极发达、环境恶化地区,占国土比例最小,从人口总数上看,生活在第I类地区的人口占全国人口比例26.37%;第II类为城市地域比例较低、灯光强度较低、PM2.5浓度高、人口密度中等的经济发达、环境污染地区,占国土面积比例的9.08%,约1/3的人口生活在此类地区;第III类为城市地域比例极低、灯光强度低、PM2.5浓度低、人口密度低的经济欠发达、环境优质地区,占国土面积比例大,全国约有43.65%的居民生活在此类地区。

从分类结果可见(图5),中国经济发达和极发达的地区占国土面积的12.56%,但集聚了超过50%的人口,极发达地区和发达地区在经济水平上

有一定差距,但在污染程度上差距不大,即中等发达程度地区面临“未富先脏”问题,环境污染状况尤为突出。

依据以上分类结果,按照经济-环境维度将全国各地级市划分为4类(图6),其中第I类为经济极发达-环境恶化地区,主要位于华北平原地区、长江三角洲地区和四川盆地东部,包括北京、上海、广州、石家庄、唐山等38个直辖市或地级行政单位;第II类为经济发达-环境趋恶化地区,主要分布在第I类区域周边,特别是京津冀周边地区,包括西安、郑州、安阳、广安、马鞍山等37个地级行政单位;第III类为经济发达-环境良好地区,广东、海南、江西、福建以及重庆等省域内多属此类型,包括三亚、茂名、宜春、福州等122个地级行政单位;第IV类为经济不发达-环境优质地区,主要分布于东北地区北部、内蒙古、甘肃、贵州、新疆、青海、西藏等地,包括牡丹江、武威、铜仁、乌鲁木齐、

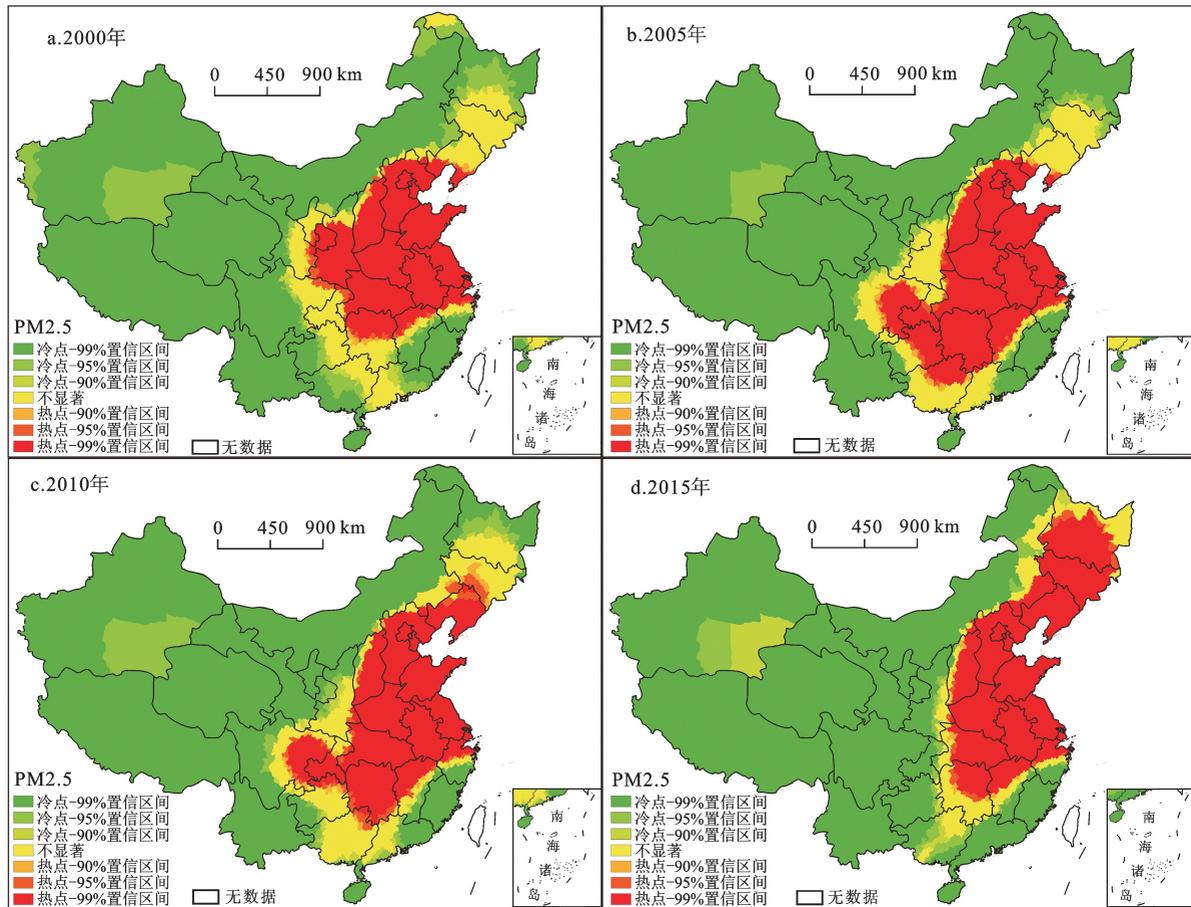


图4 2000~2015年中国PM2.5冷热点集聚

Fig.4 PM2.5 hot and cold spots gathered in China from 2000 to 2015

表1 基于神经网络分类的各类别特征统计

Table 1 Statistics of different types of features based on artificial neural networks

类别	面积 (10 <sup>4</sup> km <sup>2</sup> )	占国土面积 比例(%)	城市地域 比例(%)	灯光 指数	PM2.5浓度 (μg/m <sup>3</sup> )	人口 (人/km <sup>2</sup> )	占比 (%)
I	32.92	3.48	30.22	23.00	44.13	1049.11	26.37
II	85.86	9.08	3.16	4.94	41.89	454.38	29.99
III	826.50	87.44	0.49	0.97	14.36	68.66	43.65

注:不含港澳台数据。

西宁、拉萨等148个地级行政单位。

改革开放40 a来,中国经济取得快速发展,但也付出了巨大的环境成本。从分类结果中可以看出,经济最为发达的京津冀、长三角、珠三角地区都面临着一定的环境问题。一方面,这些区域是经济热点,吸引大量人口聚集,城市化速度较快,增加了环境的压力。同时,一些经济欠发达地区,具有较好的环境条件,但在产业转移过程中也面临相当大的风险。因此,针对不同的经济-环境特

征采取不同的发展策略,采用生态补偿等转移支付措施平衡区域间差异,是推动生态文明建设的重要手段。

### 3 结论与讨论

本研究以大数据作为切入点,选取灯光遥感数据、人口空间分布数据、PM2.5空间反演数据等指标,描述了2000~2015年中国经济、人口、环境质量的格局和变化趋势;在地统计方法的支持

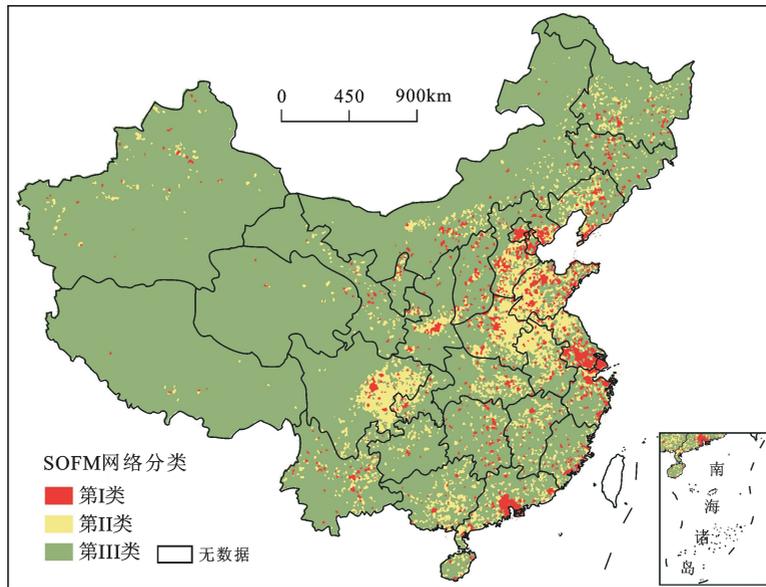


图5 基于SOFM人工神经网络的中国各地市经济-环境分类

Fig.5 Economic-environmental classification of regions based on SOFM artificial neural network in China

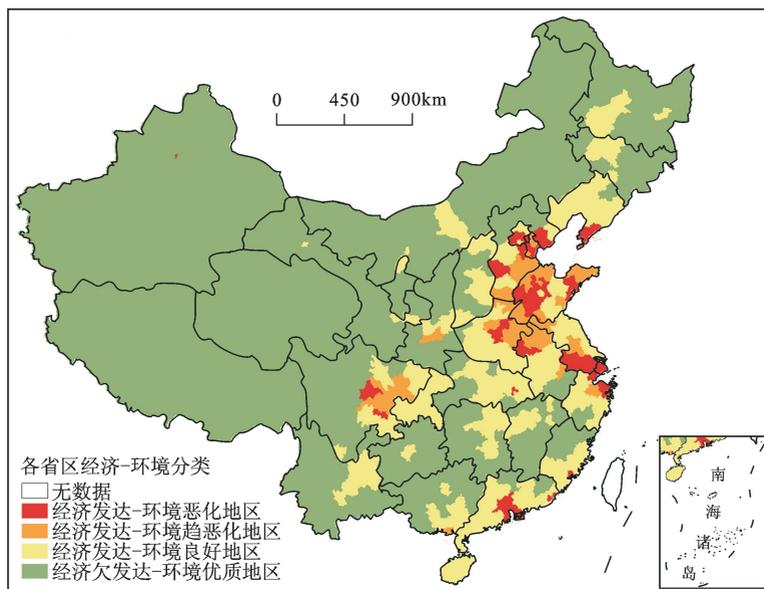


图6 中国2015年各地市经济-环境体系分类

Fig.6 Classification of economic-environmental system in China in 2015

下,揭示了经济、人口、环境质量的冷热点地区及变化趋势,根据经济-环境要素的空间分布特征将各省级行政区分成四类,研究结果表明:

1) 灯光平均强度较高的省份主要集中在沿海地区,广大中西部地区灯光强度相对较弱。2000~2015年,灯光高强度区域由沿海地区向内陆扩散,京津冀、长三角、珠三角等经济增长极集

聚效应明显,四川、重庆、湖北、湖南等中西部地区发展迅速。2000~2015年人口分布呈现出向特定区域集中的态势。在经济较为发达的京津冀、长三角、珠三角地区,人口愈加稠密,而东北地区、湖北湖南等地,人口则出现下降态势。主要经济热点区域在环渤海地区、长三角地区、珠三角地区等地,呈现出东南高西北低的发展态势,

东北、中部和西南地区经济发展相对来说不如东部沿海地区,环渤海在2000~2015年期间取得了较大发展。

2) PM<sub>2.5</sub>浓度呈现先增加后稳定的趋势,高污染区主要集中在华北、华中等区域;自2000~2015年,四川东部地区污染浓度有较大幅度下降,东北地区污染浓度有显著增长。从以PM<sub>2.5</sub>浓度为表征的环境污染情况看,华北、华中区域始终为污染热点区域;东北方向逐步扩散,污染热点地区从辽东半岛、山海关一线向东北扩张;热点区西线变化较大,从2000年的山西南部向南移动至2015年的四川东部;南线从湖南北部、江西北部、浙江北部一线南移,至2005~2015年已经进入广西、广东北部地区。

3) 根据2015年各地市OLS灯光指数、人口、城市自然边界、年均PM<sub>2.5</sub>浓度4类指标SOFM网络分类结果,全国可划分为4类区域:第I类为经济极发达-环境恶化地区,主要位于华北平原地区、长江三角洲地区和四川盆地东部;第II类为经济发达-环境趋恶化地区,主要分布在第I类区域周边,特别是京津冀周边地区;第III类为经济发达-环境良好地区,广东、海南、江西、福建以及重庆等省域内多属此类型;第IV类为经济不发达-环境优质地区,主要分布于东北地区北部、内蒙古、甘肃、贵州、新疆、青海、西藏等地。

地理空间数据具有大容量、快速度和多样性的特点,大数据方法在分析中国经济-环境关联方面有着较强的优势和可行性。在大数据方法支持下,环境和经济监测数据分析结果可以有效的提高科学决策水平。各类数据的相互配合与分析更加完整的体现了大数据的完整性与可靠性,将各种环境和经济监测数据成功的转换为有价值的分析结果,为政府决策者、企业管理者、环保组织和社会公众提供了有效的数据,促进其决策、管理、行为的有效性和科学性。而经济与环境大数据的结合应用,则进一步为不同区域间政策制定与实施提供了依据。由以上结果提出如下建议:①在经济高度发达地区,应进一步依靠地方财政优势,采用高新技术开展环境治理工作。②在中等发达程度地区面临“未富先脏”问题,环境污染状况尤为突出,该地区主要位于经济发达地区周边,应防止污染转移等问题,高度重视地区间污染治理协同防治的作用。③在经济欠发达地区,

应注意根据不同区域的资源禀赋协调具有比较优势的产业,发展清洁行业,借助发达地区生态补偿等政策,在提高经济发展水平的同时继续既有清洁环境。

## 参考文献(References):

- [1] 周雅清, 任国玉. 城市化对华北地区最高、最低气温和日较差变化趋势的影响[J]. 高原气象, 2009, 28(5):1158-1166. [Zhou Yaqing, Ren Guoyu. The effect of urbanization on maximum, minimum temperatures and daily temperature range in north China. Plateau Meteorology, 2009, 28(5):1158-1166.]
- [2] 王芳, 葛全胜. 根据卫星观测的城市用地变化估算中国1980~2009年城市热岛效应[J]. 科学通报, 2012, 57(11):951-958. [Wang Fang, Ge Quansheng. Estimation of urbanization bias in observed surface temperature change in China from 1980 to 2009 using satellite land-use data. Science China Press, 2012, 57(11):951-958.]
- [3] 石忆邵. 中国“城市病”的测度指标体系及其实证分析[J]. 经济地理, 2014, 34(10):1-6. [Shi Yishao. Measurement index system and empirical analysis of China's urban. Economic Geography, 2014, 34(10):1-6.]
- [4] 周元, 孙新章. 中国城镇化道路的反思与对策[J]. 中国人口·资源与环境, 2012, 22(4):56-59. [Zhou Yuan, Sun Xinzhang. Rethinking and countermeasures on China's urban road. China Population, Resources and Environment, 2012, 22(4):56-59.]
- [5] 李芬, 李岱青, 张林波, 等. 中日韩城市化过程中的资源环境变化比较研究[J]. 中国人口·资源与环境, 2015, 25(4):125-131. [Li Fen, Li Daiqing, Zhang Linbo et al. Comparative study on urbanization process and changes of resource and environment in China, Japan and Korea. China Population, Resources and Environment, 2015, 25(4):125-131.]
- [6] 张欢, 成金华, 陈军, 等. 中国省域生态文明建设差异分析[J]. 中国人口·资源与环境, 2014, 24(6):22-29. [Zhang Huan, Cheng Jinhua, Chen Jun et al. Difference analysis on the ecological civilization construction of China's provinces. China Population, Resources and Environment, 2014, 24(6):22-29.]
- [7] 张荣天, 焦华富. 中国省际城镇化与生态环境的耦合协调与优化探讨[J]. 干旱区资源与环境, 2015, 29(7):12-17. [Zhang Rongtian, Jiao Huafu. Coupling and coordination between urbanization and ecological environment in China. Journal of Arid Land Resources and Environment, 2015, 29(7):12-17.]
- [8] 成金华, 李悦, 陈军. 中国生态文明发展水平的空间差异与趋同性[J]. 中国人口·资源与环境, 2015, 25(5):1-9. [Cheng Jinhua, Li Yue, Chen Jun. Regional difference and convergence of ecological civilization level in China. China Population, Resources and Environment, 2015, 25(5):1-9.]
- [9] 杜祥琬, 杨波, 刘晓龙, 等. 中国经济发展与能源消费及碳排放解耦分析[J]. 中国人口·资源与环境, 2015, 25(12):1-7. [Du

- Xiangwan, Yang Bo, Liu Xiaolong et al. Decoupling analysis of China's economic development, energy consumption and carbon dioxide emission. *China Population, Resources and Environment*, 2015, 25(12):1-7.]
- [10] 李福柱, 赵长林. 中国经济发展方式的转变动力及其作用途径[J]. *中国人口·资源与环境*, 2016, 26(2):152-162. [Li Fuzhu, Zhao Changlin. Changing power of China's economic development mode and its pathway. *China Population, Resources and Environment*, 2016, 26(2):152-162.]
- [11] Kitchin R. Big data, new epistemologies and paradigm shifts[J]. *Big Data & Society*, 2014, 1(1):1-12.
- [12] Song M, Cen L, Zheng Z et al. Improving natural resource management and human health to ensure sustainable societal development based upon insights gained from working within 'Big Data Environments' [J]. *Journal of Cleaner Production*, 2015, 94:1-4.
- [13] Song M, Fisher R, Wang J et al. Environmental performance evaluation with big data: theories and methods[J]. *Annals of Operations Research*, 2016:1-14.
- [14] Li L, Hao T, Chi T. Evaluation on China's forestry resources efficiency based on big data[J]. *Journal of Cleaner Production*, 2016, 142: 513-523.
- [15] Xie H, He Y, Xie X. Exploring the factors influencing ecological land change for China's Beijing-Tianjin-Hebei Region using big data[J]. *Journal of Cleaner Production*, 2017,142: 677-687.
- [16] Wu J, Wang Z, Li W et al. Exploring factors affecting the relationship between light consumption and GDP based on DMSP/OLS nighttime satellite imagery[J]. *Remote Sensing of Environment*, 2013, 134: 111-119.
- [17] 李德仁, 李熙. 论夜光遥感数据挖掘[J]. *测绘学报*, 2015, 44(6): 591-601. [Li Deren, Li Xi. An overview on data mining of Nighttime light remote sensing. *Acta Geodaetica et Cartographica Sinica*, 2015, 44(6): 591-601.]
- [18] Clark H, Pinkovskiy M, Sala-i-Martin X, China's GDP growth may be understated[EB/OL]. 2018-4-23. <http://www.nber.org/papers/w23323.pdf>. [2018-4-23.]
- [19] 李德仁, 余涵若, 李熙. 基于夜光遥感影像的“一带一路”沿线国家城市发展时空格局分析[J]. *武汉大学学报·信息科学版*, 2017, 42(6): 711-720. [Li Deren, Yu Hanruo, Li Xi. The spatial-temporal pattern analysis of city development in countries along the Belt and Road Initiative based on nighttime light data. *Geomatics and Information Science of Wuhan University*, 2017, 42(6): 711-720.]
- [20] Dai Z, Hu Y, Zhao G. The suitability of different nighttime light data for GDP estimation at different spatial scales and regional levels[J]. *Sustainability*, 2017, 9(2):305.
- [21] van Donkelaar A, Martin R V, Brauer M et al. Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors[J]. *Environmental Science & Technology*, 2016, 50(7): 3762.
- [22] 柯文前, 陆玉麒. 基于县域的福建省经济空间格局演化[J]. *经济地理*, 2011, 31(7): 1081-1087.[Ke Wenqian, Lu Yulin. Evolution of spatial pattern of economy at county level in Fujian Province. *Economic Geography*, 2011, 31(7): 1081-1087.]
- [23] 李莎, 刘卫东. 俄罗斯人口分布及其空间格局演化[J]. *经济地理*, 2014, 34(2): 42-49. [Li Sha, Liu Weidong. Spatial distribution of population in Russia and its evolution. *Economic Geography*, 2014, 34(2): 42-49.]
- [24] 管卫华, 彭鑫, 张惠, 等. 不同空间尺度的江苏省区域经济差异研究[J]. *长江流域资源与环境*, 2015, 24(12): 2003-2011. [Guan Weihua, Peng Xin, Zhang Hui et al. Regional economic inequalities in Jiangsu at different spatial scales. *Resources and Environment in the Yangtze Basin*, 2015, 24(12): 2003-2011.]
- [25] 黄姣, 高阳, 赵志强, 等. 基于GIS和SOFM网络的中国综合自然区划[J]. *地理研究*, 2011, 30(9): 1648-1658. [Huang Jiao, Gao Yang, Zhao Zhiqiang et al. Comprehensive physiographic regionalization of China using GIS and SOFM neural network. *Geographical Research*, 2011, 30(9): 1648-1658.]
- [26] 冯喆, 吴健生, 高阳, 等. 基于SOFM网络的景观功能分类——以北京及周边地区为例[J]. *地球信息科学学报*, 2012, 14(6): 800-806. [Feng Zhe, Wu Jiansheng, Gao Yang et al. Classification of landscape functions using SOFM neural network: A case study from Beijing and its peripheral area. *Journal of Geo-information Science*, 2012, 14(6): 800-806.]

## China's Economic-environment Comprehensive Zoning Based on Big Data Method and SOFM Clustering

Feng Zhe<sup>1,2</sup>, Jiang Hongqiang<sup>1</sup>, Lu Yaling<sup>1</sup>

(1. *State Environmental Protection Key Laboratory of Environmental Planning and Policy Simulation, Chinese Academy for Environmental Planning, Beijing 100012, China*; 2. *School of Land Science and Technology, China University of Geosciences, Beijing 100083, China*)

**Abstract:** In the past 40 years, China's economy has developed rapidly but has paid a heavy environmental cost. China's environment and economy have undergone tremendous changes. The imbalance in regional economic development has intensified, while the various kinds of environmental problems have taken place in different area. Therefore, it is necessary to formulate environmental policies in a targeted manner. To this end, big data methods should be applied to reveal the relationship between the environment and economy, which forms the basis of comprehensive zoning. In this study, multi-source big data such as nighttime light remote sensing, spatialized population data, urban natural boundary and simulated PM<sub>2.5</sub> concentration was used to establish an economic-environmental linkage system. The time and spatial variation rules of economic and environment hotspots were identified by Getis-Ord index. The provincial administrative areas of China were divided into four groups by using an artificial neural network (Self-Organizing Feature Map, SOFM) clustering of both economic and environmental indicators. The results show that: 1) The provinces with higher average light intensity in China are mainly concentrated in coastal areas. The economy is mainly driven by the Yangtze River Delta, the Pearl River Delta, and the Bohai Rim. It is a developing trend that is low in the southeast, high in the northwest and the eastern coastal areas. The economic development of coastal area is superior to the northeast, central and southwest regions. 2) The concentration of PM<sub>2.5</sub> in China shows a trend of increasing first and then decreasing. Highly polluted areas are mainly concentrated in the North China and Central China regions. In northeast China, the pollution hotspots expand from the Liaodong Peninsula and Shanhaiguan to the northeast, while the south China maintains stability. 3) Based on OLS light index, population, urban natural boundary, and annual average PM<sub>2.5</sub> concentration in 2015, the prefecture-level cities can be grouped into four types by using the self-organizing feature mapping model. Type I region is highly economic developed and environment deteriorating areas, mainly located in the North China Plain and the Yangtze River Delta. Type II region is the economic developed and environment deteriorating areas, which is mainly distributed in the periphery of the Type I region, especially around Jing-jin-ji area. Type III region is economically developed and environmentally friendly, such as Guangdong, Hainan, Jiangxi, Fujian, and Chongqing. Type IV is economically underdeveloped and high environment quality areas, which are mainly distributed in the north of the Northeast, Inner Mongolia, Gansu, Guizhou, Xinjiang, Qinghai, and Tibet. The research results are hoped to provide reference for balancing economic development and environment conservation in different regions of China.

**Key words:** economic-environment system; big data; nighttime lighting index; PM<sub>2.5</sub>; Self-Organizing Feature Map (SOFM)