www.csdata.org

2010 年中国工业产值公里格网数据集

薛倩[1,2]，宋伟[1*]，朱会义[1]

1. 中国科学院地理科学与资源研究所，陆地表层格局与模拟院重点实验室，北京 100101

2. 重庆交通大学建筑与城市规划学院，重庆 400074

摘要：空间化工业产值数据的缺乏，使得全球变化背景下工业经济系统风险与灾害评估等工作受到很大限制。我们发展了一套融合美国国防气象卫星计划/线性扫描业务系统（DMSP/OLS，Defense meteorological satellite program/operational linescan system）夜间灯光数据、中分辨率成像光谱仪（MODIS，Moderate resolution imaging spectroradiometer）全年植被数据集、工业用地空间分布数据、城市化率数据的工业产值数据空间化方法，形成了 2010 年中国工业产值公里格网数据集。本数据集的主要构建步骤为：（1）数据预处理与稳定灯光数据选取；（2）构建基于增强型植被指数的夜间灯光调整指数（EANTLI，Enhanced vegetation index adjusted nighttime light index）；（3）利用工业用地空间分布数据获取工业用地最佳灯光指数；（4）构建工业产值空间分配模型；（5）数据质量精度验证。全国 105 个城市的随机样本验证结果表明，空间化工业产值数据的相对误差在 0%～39.6%之间，且大部分样本的相对误差小于 15%，数据集平均精度为 81.40%。本数据集解决了二三产业产值空间化数据难以区分的问题，空间化数据打破了行政边界的限制，可以直观地反映出工业产值在时空数量差异以及分布特征，对于划分中国工业重点区域划分、厘清工业产业变化趋势等有重要价值。

关键词：工业产值；中国；DMSP/OLS；空间化

数据库（集）基本信息简介

数据库（集）名称	2010 年中国工业产值公里格网数据集
数据作者	薛倩、宋伟、朱会义
数据通信作者	宋伟（songw@igsnrr.ac.cn）
数据时间范围	2010年
地理区域	中国大陆地区
空间分辨率	1000 m
数据量	15.1 MB
数据格式	*.tif ,*.shp, *.mxd, *.xlsx
数据服务系统网址	http://www.sciencedb.cn/dataSet/handle/482
基金项目	国家重点研发计划，全球变化人口与经济系统风险形成机制及评估研究（2016YFA0602402）

文献 DOI:
10.11922/csdata.2017.14.zh
数据 DOI：
10.11922/sciencedb.551
文献分类：地球科学

收稿日期：2017-09-20
开放同评：2017-10-17
发表日期：2018-06-05

* 论文通信作者
宋伟：songw@igsnrr.ac.cn

数据库（集）组成	数据集由3部分数据组成，其中：1 2010年中国工业产值公里格网数据集.gdb.zip是包含行政边界信息以及栅格格式的中国2010年每公里工业产值数据（13.8 MB）。2. 精度验证.xls是全国90个城市工业产值统计数据和验证精度说明（23 KB）。3. 2010年中国工业产值公里格网数据集.mxd是可编写的ArcGIS数据文件（730 KB）。

引　言

辨析气候变化下的社会经济系统风险是目前气候变化研究的重点内容之一[1]。就工业系统而言，气候变化一方面通过均值波动间接对工业行业的原料存储、加工以及运输过程产生影响[2]，另一方面通过极端气候事件对工业生产的各个环节及工作人员产生破坏性的影响（例如工厂基础设施破坏、人员伤亡等）[3-4]。准确评估气候变化带来的工业经济系统风险与损失，依赖于较高精度的工业经济系统暴露分析[5]，工业产值的空间分布制图则是工业经济系统暴露分析的重要基础数据之一。

近年来，国内外在国内生产总值（GDP，Gross domestic product）空间分布制图方面已经取得了比较多的进展[6]。但是，由于常规的遥感手段很难在空间上实现第二、三产业产值的准确识别，全球以及中国工业产值格网数据仍然相对缺乏。目前，已有的一些工业经济系统产值数据空间化制图，多为省、市、县级尺度[7]，且空间分辨率多以行政区为最小单元，无法表征省或者城市内部工业产值的差异及空间分布，在风险评估中很难与气候格网数据等开展叠加分析；基于此，也有一些针对工业某一具体行业产值的空间化研究数据[8-9]，但是总体上缺少大尺度（中国）、高分辨率、综合性的工业产值空间化数据。所以，本数据集结合 DMSP/OLS 夜间灯光数据、中国各省工业产值统计数据以及 MODIS 中的植被指数产品，对中国 2010 年工业产值进行遥感反演，并利用工业用地空间分布数据对其进行修正，形成中国 1 km 分辨率的工业产值格网数据集。本数据集可以为全球变化中工业经济系统的风险与灾害评估等提供基础数据支撑。

1　数据采集和处理方法

本数据集构建过程包括数据收集、夜间灯光数据和植被指数数据预处理、构建基于增强型植被指数的夜间灯光调整指数（EANTLI）及利用工业用地空间分布数据进行灯光修正、构建分配模型、数据质量精度验证几个步骤（图 1）。

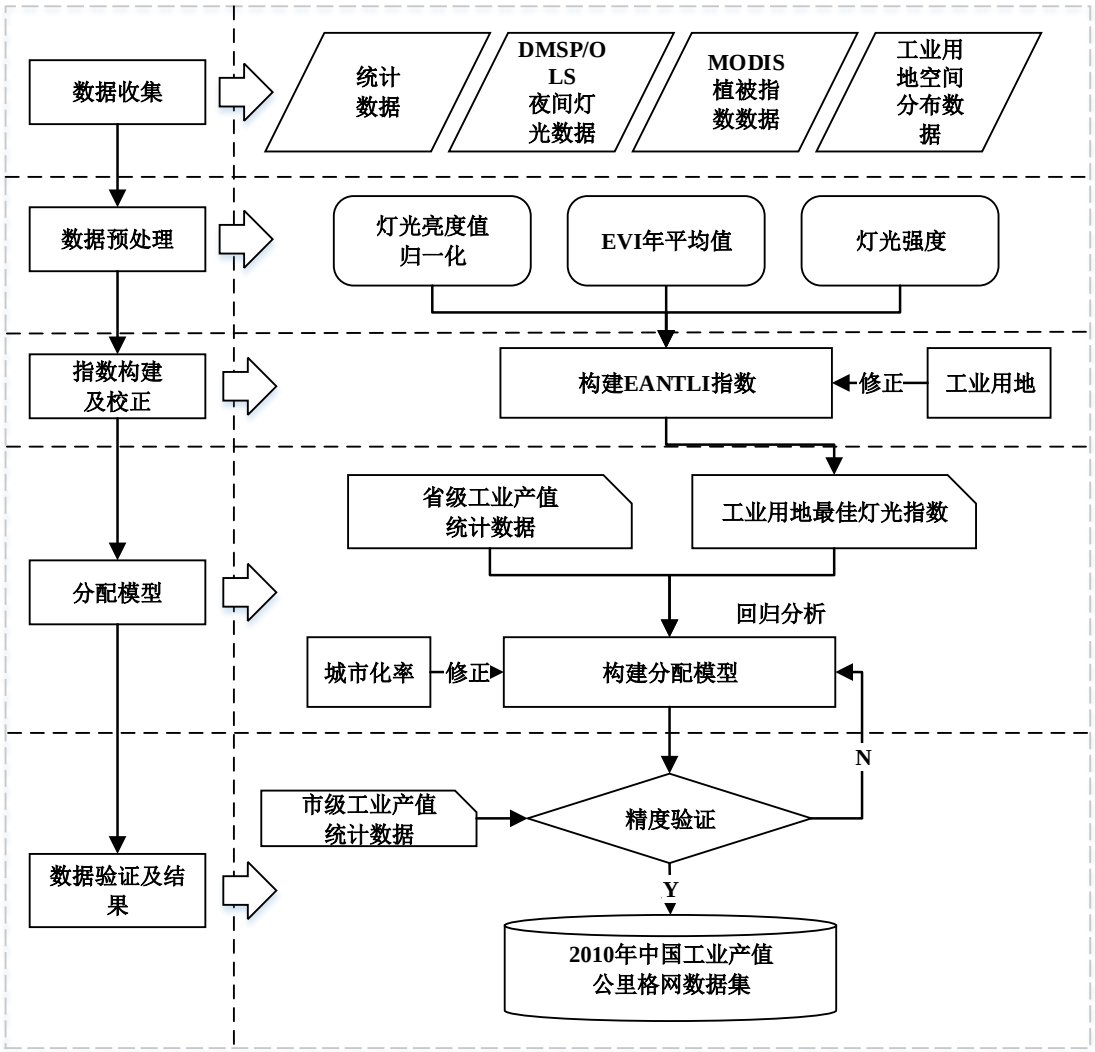


图 1　数据集构建及处理流程

1.1　数据来源

本数据集对工业的定义以世界银行数据统计标准为主，包含采矿、制造、建筑、电力、水和天然气生产和供应业，与中国统计局差别在于将建筑业纳入了工业。所以，本数据集中所采用的 2010 年全国各省份工业产值和市级工业产值数据为工业产值和建筑业产值之和。数据均来自于国家统计局发布的《中国统计年鉴》和各省统计年鉴（http://www.stats.gov.cn/tjsj/ndsj/）。

夜间灯光数据选取美国气象卫星 DMSP/OLS 成像得到的非辐射定标夜间平均灯光产品。其优势在于能通过灯光强度变化反应地理实体信息，特别地，该产品中的稳定灯光数据通常被用于提取城市区域[10-11]、反演社会经济数据[12-13]等研究中。本数据集所选取的夜间灯光数据为稳定灯光数据（https://ngdc.noaa.gov），其空间分辨率为 1 km，像元灰度值（DN，Digital number）范围为 0～63，数据时间为 2010 年。由于 DMSP/OLS 数据存在的饱和与溢出现象会削弱灯光数据和社会经济数据的相关性[14]，并影响反演精度。而且有研究表明，植被指数和灯光强度值变化存在较好相关关系[15]，因此可以通过植被指数来消除夜间灯光数据的饱和与溢出现象对社会经济数据反演的影响。所以本数据集选取空间分辨率为 1 km 的 MODIS

月合成植被指数产品（MOD13A3）中的增强型植被指数（EVI，Enhanced vegetation index）（http://ladsweb.nascom.nasa.gov/data）对灯光饱和与溢出现象进行预处理。

此外，选取中国科学院资源环境科学数据中心 2010 年工业用地空间分布数据（http://www.resdc.cn）对 EANTLI 指数进行精度修正。

1.2 数据处理

利用夜间灯光数据与植被指数之间关系消除灯光数据饱和与溢出现象，并构建最佳灯光指数，同时利用工业用地空间分布数据对其进行修正，得到工业用地上的最佳灯光数据，并与工业产值数据构建分配模型。此外，针对西部地区工业产值较低且工业用地分布较为离散造成的较大低估现象问题，利用城市化率对新疆、西藏等四个灯光值较弱的省份进行修正，最终形成 2010 年中国工业产值公里格网数据集。具体流程如下：

（1）对夜间灯光数据进行预处理，包括按行政边界剪裁、二值化、归一化等，提取灯光强度值，并计算归一化的灯光强度值；

（2）对 MODIS 植被指数数据进行处理，从 MOD13A3 产品中提取出中国 2010 年 1–12 月的 EVI，然后进行初步的剪裁、镶嵌、投影等工作；

（3）EVI 数据处理，计算 EVI 的均值，为去除月份不同对夜光数据造成的敏感性，所以本数据集选取 EVI 全年 12 个月的均值数据，并剔除 EVI 小于 0.01 的区域，此区域包括水体、裸岩等无人区域；

（4）构建最佳灯光指数，其计算公式如式（1）[14]：

$$EANTLI = \left(\frac{2}{1 - NTL_{\mathrm{n}} + EVI_{\mathrm{i}}} - 1 \right) \times NTL \tag{1}$$

其中，*EANTLI* 为最佳灯光指数，NTL_{n} 为归一化灯光强度，EVI_{i} 指处理过后的 EVI 数据，*NTL* 为原始灯光强度。

EANTLI 能够较大程度上削减灯光饱和与溢出现象，以凸显城市内部灯光强度差异，更好地对经济数据进行反演。

（5）通过工业用地空间分布数据对第 4 步得到的 EANTLI 指数结果进行修正，本数据集认为仅工业用地上存在工业产值，通过工业用地数据对 EANTLI 指数进行工业工地上的灯光值提取，得到工业用地最佳灯光指数 *EI*。

（6）利用国家统计局《中国统计年鉴》[16]中计算得到的全国各省 2010 年工业产值数据与工业用地最佳灯光指数进行回归分析，构建分配模型：

$$I = \frac{I_i \times EI}{EI_i} (i = 1, 2, 3 \ldots n) \tag{2}$$

其中，*I* 代表每个栅格上的工业产值，I_i 代表每个各省的工业产值，EI_i 代表每个省的工业用地最佳灯光指数。

经过该步骤，初步获得到工业产值公里格网数据。

（7）通过对第 6 步获得的数据验证发现，由于新疆、青海、西藏、云南四个省份的工业产值较低、工业用地较少且分布较为离散，导致存在较大的产值低估现象。针对这个问题，本数据集利用与工业产值在市级尺度上存在显著相关（0.01 置信度水平下）的辅助数据—

一四个省份的土地城市化率进行构建多元线性回归模型进行数据修正，模型如下：

$$I_g = aI_n + bU_n + c \quad (3)$$

其中 I_g 代表修正后的工业产值，I_n 代表第（6）步中得到第 n 个栅格的工业产值，U_n 代表第 n 个栅格的城市化率，a、b 分别为 I_n、U_n 的参数，c 为常量。

（8）利用市级统计年鉴验证各城市工业产值数据精度，满足则得到最终 2010 年中国工业产值公里格网数据集，否则重新构建分配模型。

2 数据样本描述

经过数据处理，得到 2010 年中国工业产值公里格网数据集。本数据集可以比较清晰地反映每公里工业产值的空间分布特征。由于地理位置及政策优势，我国工业产值的高值区域主要分布在我国沿海区域，特别是长三角、珠三角、环渤海等几个典型地区（图 2）；而西北地区由于地处内陆，交通运输条件较差，工业产值普遍较低，且呈点状零星分布，且相较于沿海区域数值明显较低。

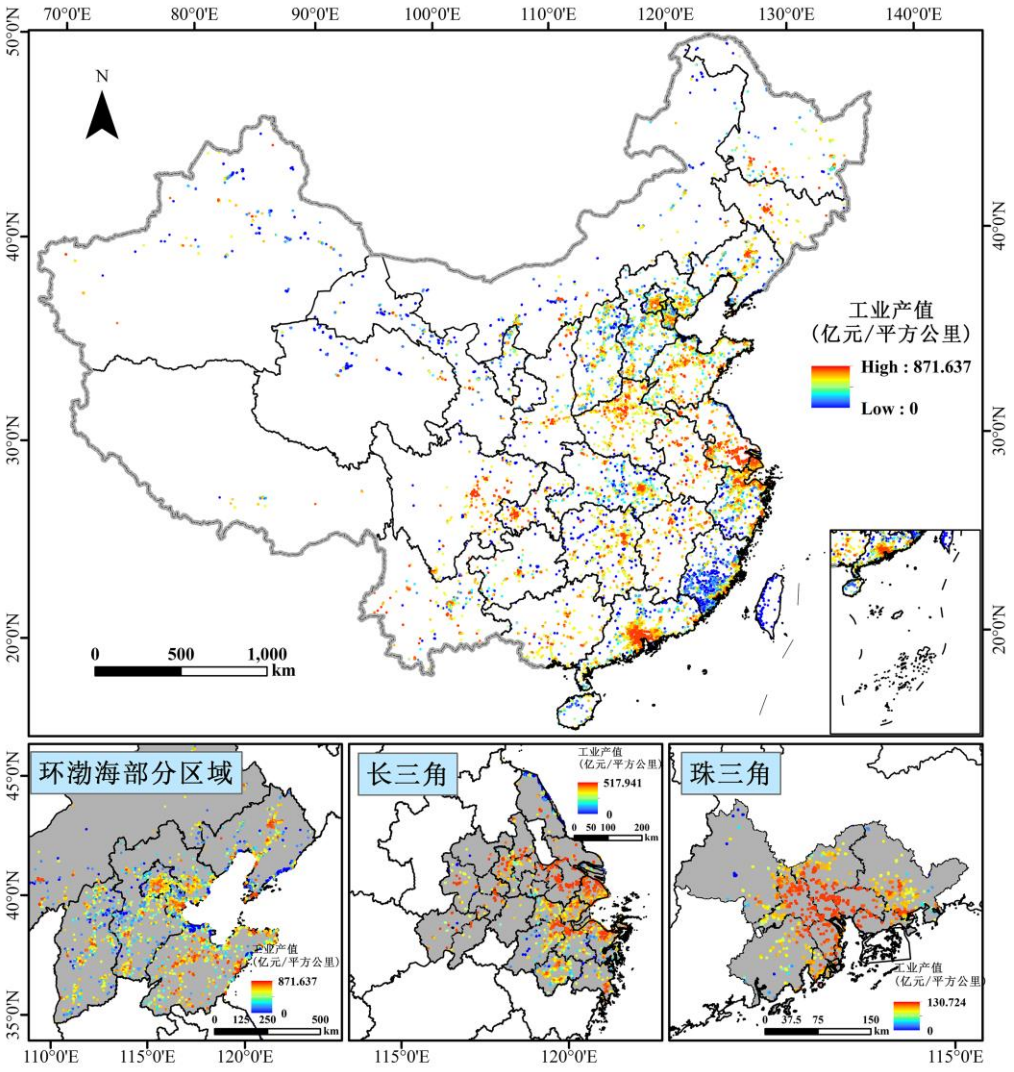


图 2 2010 年中国工业产值公里格网数据（审图号：GS（2018）2341 号）

3 数据质量控制和评估

数据验证时，开展了省级工业用地最佳灯光数据与工业产值的相关性分析，结果显示两者显著相关（在 0.01 置信度水平下），且相关系数达到 0.72。同时，随机选取了全国 105 个不同市级行政单位（图 3），利用 ArcGIS 的统计工具，将工业产值公里格网数据进行统计，并与市级统计年鉴中计算得到的工业产值数据进行对比。整体而言，平均精度达到 81.40%（图 4）。为了使数据能够给其他行业深度分析带来参考，我们又按照城市主导功能的不同，从 105 个验证样本中挑选了不同产业类型城市进行了精度验证比较。按照全国资源型城市可持续发展规划（2013–2020 年）的分类标准和相关文献[17-18]，选取了 11 个资源型城市、10 个综合型城市和 17 个工业型城市（表 1）进行验证比较，结果表明资源型城市整体平均精度为 82.27%，其中 7 个煤炭型城市平均精度为 82.75%，1 个有色金属城市精度为 94.08%，3 个石油城市平均精度 77.22%；总体而言，综合型城市整体平均精度为 82.34%；工业型城市整体平均精度为 77.25%。不过，部分西部欠发达区域精度略低，环渤海区域、长三角、珠三角等工业较发达区域工业产值反演精度较高。西部地区精度偏低的主要原因在于经济发展水平偏低导致区域灯光强度较弱，且内部差异较小，使得反演精度偏低。为了解决这个问题，我们利用与工业产值有较好相关性的城市化率数据对新疆、西藏、青海、云南四个省份进行了数据校正，校正后精度有较大的提升，提高了 12.1 个百分点（由 62.4%提高到 74.5%）。

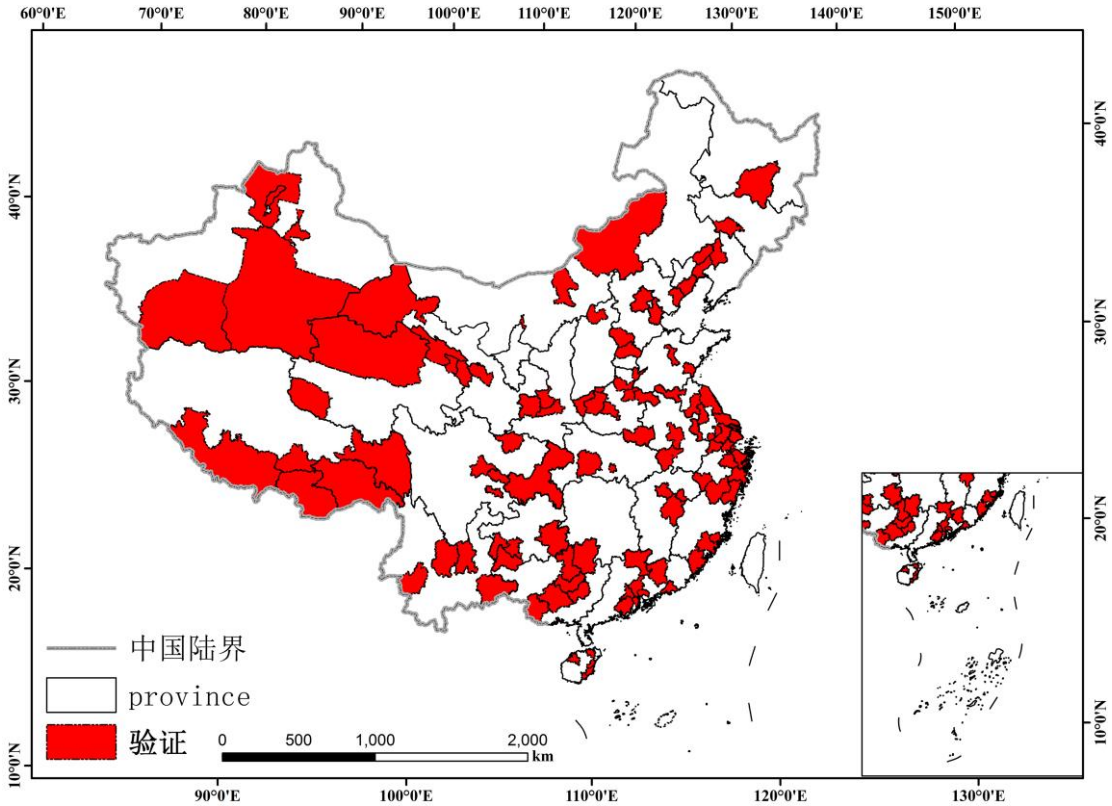
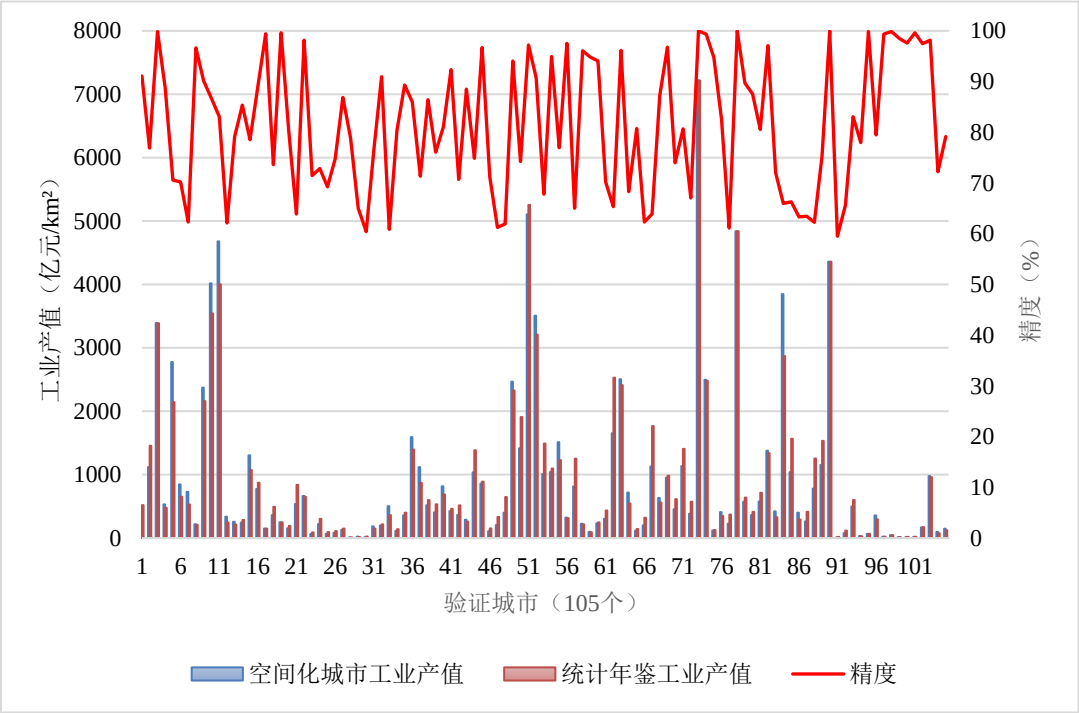


图 3 精度验证选择的 105 个样本城市的位置（审图号：GS（2018）2341 号）



注：横坐标轴为所选验证城市的序号，左侧纵坐标轴为对应该城市的空间化与统计年鉴工业产值，右侧纵坐标轴为该城市的空间化产值数据的验证精度。

图 4　数据集中数据与统计年鉴数据对比及精度

表 1　分类型城市精度验证

	资源型城市		综合型城市		工业型城市			
	城市	精度（%）	城市	精度（%）	城市	精度（%）	城市	精度（%）
煤炭型城市	邢台	80.34	北京	99.87	扬州	76.99	宁波	65.96
	朔州	96.80	上海	100.00	自贡	61.10	柳州	63.88
	乌海	68.32	天津	99.99	江门	88.85	淄博	63.88
	阜新	94.85	哈尔滨	74.87	新乡	81.07	台州	62.26
	平顶山	71.35	西安	80.67	开封	89.36	包头	96.14
	广元	94.75	沈阳	65.33	盐城	94.96	湖州	80.55
	六盘水	72.84	东莞	90.18	南通	74.22		
有色金属城市	葫芦岛	94.08	佛山	86.60	温州	75.29		
石油城市	濮阳	70.71	南昌	65.01	中山	78.52		
	潜江	71.24	石家庄	60.87	无锡	90.70		
	克拉玛依	89.73			洛阳	86.00		
平均精度（%）	82.27		82.34		77..5			

本数据处理过程在地理信息系统软件 ArcGIS 平台下进行，数据处理过程科学合理规范，保证了数据空间分辨率的准确性和可靠性。

4 数据价值

工业产值公里格网数据打破了行政边界的限制，通过夜间灯光数据对工业产值数据进行大尺度反演，得到高精度、高分辨率空间化的栅格数据，从而可以直观地分析工业产值区域间数量差异以及时空分布特征，为划分中国工业重点区域、厘清工业产业变化趋势、评估工业用地效率等提供了数据支撑，特别是为气候变化下工业产业暴露及脆弱性评估提供了基础数据，从而为工业布局规划及极端灾害条件下工业产业受灾预警及灾害评估提供参考。

此外，本数据集中工业定义以世界银行中工业定义为准，可用于全球尺度的工业产值变化评估研究。特别地，虽然全球或者区域空间化产品较多，但大都是 GDP、人口的空间化产品，特别是现有 GDP 空间化数据集大多是对二三产业合并处理，鲜有单独的二产或三产空间化数据集产品。本数据集提出的遥感反演、工业用地修正、城市化率修正等方法，能够有效在空间上区分第二与第三产业产值，丰富了社会经济数据的空间化方法，为以后的类似研究提供了借鉴。在精度上，本数据集的平均精度达到了 81.40%，相较于以往类似研究有了明显提升。

5 数据使用方法和建议

本数据集中包含工业产值公里格网文件（.tif）和检验数据文件（.xls）以及处理过程数据库（.gdb），可用 ArcGIS 软件对数据进行读写，也可用 Python 等主流编程语言调用相关函数库读写，实现数据批处理。

对于全球气候变化的相关研究工作，1km 的全国工业产值数据集可以满足目前研究的精度。同时，1km 也是夜间灯光影像数据、MODIS 植被指数数据、土地利用栅格数据（全国尺度）常用的空间分辨率，这个分辨率能够最好地匹配制作工业产值所需的空间基础数据。鉴于制作工业产值空间化反演的基础数据在空间分辨率均为 1km，数据集的分辨率已达到最高。如果后期需要匹配大尺度低分辨的气象数据（如 0.1°，0.25°等），可以借助于百分比栅格的方法对数据集进行升尺度，这个方法不会影响降低分辨率之后的数据集精度。

此外，在数据集更新方面，我们在此数据集的基础上会根据基础数据——土地利用（更新周期：5 年）、夜间灯光影像数据（更新周期：1 年）、MODIS 植被指数数据（更新周期：1 年）、统计年鉴数据（更新周期：1 年）的更新周期进行数据更新，拟依据更新周期最长的土地利用数据的更新周期（5 年）进行数据集更新。同时，对未来工业产值格网格化数据，我们会通过动力降尺度的方法进行多源数据融合模拟，制作未来不同气候变化情景下的工业产值空间数据。

数据作者分工职责

薛倩（1993—），女，新疆乌鲁木齐市人，硕士，研究方向为国土资源与 3S 技术。主要承担工作：数据采集及数据处理。

宋伟（1981—），男，山东沂源人，博士，副研究员，研究方向为土地利用变化及其生态效应。主要承担工作：数据集设计及技术指导。

朱会义（1966—），男，江苏省响水县人，博士，副研究员，研究方向为土地变化。主要承担工作：数据集设计。

参考文献

[1] IPCC, Working Group I. Climate Change 2013: The Physical Science Basis[J]. Contribution of Working, 2013, 43(22): 866-871.

[2] 王守荣. 气候变化对中国经济社会可持续发展的影响与应对[M]. 北京: 科学出版社, 2011.

[3] CRUZ A M, KRAUSMANN E. Vulnerability of the oil and gas sector to climate change and extreme weather events[J]. Climatic Change, 2013, 121(1): 41-53.

[4] KEPPENNE C L, GHIL M. Extreme weather events[J]. Nature, 1992, 358(6387): 547.

[5] LEREBOULLET A, BELTRANDO G, BARDSLEY D. Socio-ecological adaptation to climate change: A comparative case study from the Mediterranean wine industry in France and Australia[J]. Agriculture Ecosystems & Environment, 2013, 164(164): 273-285.

[6] SUTTON P C, COSTANZA R. Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation[J]. Ecological Economics, 2002, 41(3): 509-527.

[7] 刘永伟, 闫庆武. 基于SLM模型的中国碳排放格网化空间分布模拟[J]. 地理与地理信息科学, 2015, 31(3): 76-80.

[8] LIU X L, FANG C L. Wind energy resources distribution and spatial differences of wind power industry in China[C]. Marietta: American Scholars Press, 2007.

[9] DONG L, LIANG H, GAO Z, et al. Spatial distribution of China's renewable energy industry: Regional features and implications for a harmonious development future[J]. Renewable & Sustainable Energy Reviews, 2016, 58(4): 1521-1531.

[10] HUANG X, SCHNEIDER A, FRIEDL M A. Mapping sub-pixel urban expansion in China using MODIS and DMSP/OLS nighttime lights[J]. Remote Sensing of Environment, 2016, 175: 92-108.

[11] CHEN Z, YU B, SONG W, et al. A New Approach for Detecting Urban Centers and Their Spatial Structure With Nighttime Light Remote Sensing[J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, PP(99): 1-15.

[12] WANG W, CHENG H, ZHANG L. Poverty assessment using DMSP/OLS night-time light satellite imagery at a provincial scale in China[J]. Advances in Space Research, 2012, 49(8): 1253-1264.

[13] 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(7): 111-119.

[14] ZHUO L, ZHENG J, ZHANG X, et al. An improved method of night-time light saturation reduction based on EVI[C]. AAG, Chicago, USA, 2015.

[15] WANG Z, YAO F, LI W, et al. Saturation Correction for Nighttime Lights Data Based on the Relative NDVI[J]. Remote Sensing, 2017, 9(7): 759-772.

[16] 国家统计局. 中国统计年鉴-2011[M]. 2011.

[17] 张文举, 刘嗣明, 郑永丹. 国内资源型城市经济转型文献综述[J]. 资源与产业, 2015, 17(3):22-26.

[18] 许锋, 周一星. 科学划分我国城市的职能类型,建立分类指导的扩大内需政策[J]. 城市发展研究, 2010, 17(2): 88-97.

论文引用格式

薛倩, 宋伟, 朱会义. 2010 年中国工业产值公里格网数据集[J/OL]. 中国科学数据, 2018, 3(2). (2018-06-05). DOI: 10.11922/csdata.2017.14.zh.

数据引用格式

薛倩, 宋伟, 朱会义. 2010年中国工业产值公里格网数据集[DB/OL]. Science Data Bank, 2017. (2017-12-29). DOI: 10.11922/sciencedb.551.

1 km grid dataset of industrial output value in China (2010)

Xue Qian[1,2], Song Wei[1*], Zhu Huiyi[1]

1. Key Laboratory of Land Surface Pattern and Simulation, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;

2. Institute of Architecture and Urban Planning of Chongqing Jiaotong University, Chongqing 400074, China

*Email: songw@igsnrr.ac.cn

Abstract: The lack of spatial industrial output value data limited the risk and disaster assessment of industrial economy responding to global change. Therefore, we developed a new method to spatialize industrial output value coupling DMSP/OLS (Defense meteorological satellite program/operational linescan system) nighttime light data, MODIS (Moderate-resolution imaging spectroradiometer) annual vegetation data, industrial land distribution data and urbanization rate. A grid data set of 1 km industrial output value of China was created using this method. The main steps creating the data set were as follows: (1) data preprocessing and selecting stable lighting data; (2) constructing an Enhanced Vegetation Index (EVI) adjusted nighttime light index (EANTLI); (3)

obtaining optimum light index by industrial land distribution data; (4) constructing spatial distribution model of industrial output value; (5) verifying data accuracy. We randomly selected 105 cities nationwide to assess the accuracy of the data set. The results show that the relative errors of whole samples ranged from 0% to 39.6%,the relative errors of most samples were less than 15%, and the average accuracy of the data set was as high as 81.40%. The dataset solved the problem that the industrial output value and service output value are difficult to be distinguished in value spatialization. The dataset broke the limits of administrative boundaries so as to directly reflect the spatial and temporal disparities and distribution features of industrial output value. The advances of the dataset could contribute to the identification of China's key industrial distribution areas and discern the change trend of industry.

Keywords: industrial output value; China; DMSP/OLS; spatialization

Dataset Profile

Title	1 km industrial output value grid data set of China（2010）
Data author(s)	Xue Qian, Song Wei, Zhu Huiyi
Data corresponding author	Song Wei (songw@igsnrr.ac.cn)
Time range	2010
Geographical scope	Mainland China
Spatial resolution	1000 m
Data volume	15.1 MB
Data format	*.tif ,*.shp, *.mxd, *.xlsx
Data service system	http://www.sciencedb.cn/dataSet/handle/482
Source(s) of funding	National Key Research and Development Program--Global Change and Mitigation Project: Global change risk of population and economic system mechanism and assessment （2016YFA0602402）
Dataset/Database composition	The dataset is composed of three subsets. (1) 1 km industrial output value grid data set of China(2010).gab.zip includes administrative border data and spatial distribution data of industrial output value of mainland China in 2010. The format of the spatial distribution data is 1 km grid with a data volume of 13.8 MB. (2) Accuracy verification.xls is an accuracy file including the randomly selected 90 sample cities for accuracy assessment in China and their statistical industrial output value, spatialized value and assessed accuracy. The data volume is 23 kB. (3) 1 km industrial output value grid data set of China(2010). mxd is a programmable ArcGIS data file, with a data volume of 730 KB.