

# 不同样本尺度和分区方案的粮食产量空间化及误差修正

姬广兴<sup>1</sup>, 廖顺宝<sup>1,2\*</sup>, 岳艳琳<sup>1</sup>, 侯鹏敏<sup>1</sup>, 杨旭<sup>1</sup>

(1. 河南大学环境与规划学院, 开封 475004; 2. 中国科学院地理科学与资源研究所, 北京 100101)

**摘要:** 粮食产量数据空间化有助于粮食产量数据与其他自然、人文数据进行综合分析, 但空间化过程中必然会产生误差。该文按照 3 种分区方案 (全国不分区、全国分为 7 个区以及按省分区), 选择 3 种尺度上 (县级、地市级和公里网格) 的总产及平均产量数据 (即 4 种样本: 县级粮食总产、县级平均粮食产量、地市级粮食总产、地市级平均粮食产量) 分别为因变量, 以对应的 3 种农田类型 (水田、水浇地、旱地) 面积数据为自变量, 利用多元线性回归分析方法, 得到 15 种空间化模型。采用两阶段误差分析方法, 选取 2 个模型误差评价因子和 5 个空间化结果误差评价因子, 对模型和空间化结果进行误差分析。结果表明: 1) 空间化过程中, 模型精度与空间化结果的精度存在不一致性; 2) 对于采用同一样本的模型 (常数项为 0) 而言, 空间化结果精度随着分区方案的细化先提高再降低, 而对于采用同一样本的模型 (常数项非 0) 而言, 空间化结果精度随着分区方案的细化而降低; 3) 在全国不分区和分为 7 个区 2 种情况下, 空间化结果精度随着分析样本尺度的细化 (从地市级到县级再到公里网格) 先提高后降低。根据上述分析结果, 最终以县级粮食总产为样本、常数项为 0、全国分 7 个区建模的方案实现全国粮食产量数据空间化, 并通过修正, 得到 2005 年中国粮食产量公里网格分布图。该研究弥补了粮食产量空间化误差分析的不足, 探寻了不同样本尺度和分区方案与空间化误差的关系, 提高了空间化精度, 同时对其他类型的社会经济统计数据空间化研究具有一定的参考价值。

**关键词:** 粮食; 误差修正; 模型; 样本尺度; 分区方案; 多元回归; 产量; 空间化

doi: 10.11975/j.issn.1002-6819.2015.15.037

中图分类号: F762.1

文献标志码: A

文章编号: 1002-6819(2015)-15-0272-07

姬广兴, 廖顺宝, 岳艳琳, 侯鹏敏, 杨旭. 不同样本尺度和分区方案的粮食产量空间化及误差修正[J]. 农业工程学报, 2015, 31(15): 272-278. doi: 10.11975/j.issn.1002-6819.2015.15.037 <http://www.tcsae.org>

Ji Guangxing, Liao Shunbao, Yue Yanlin, Hou Pengmin, Yang Xu. Spatial distribution of grain yield based on different sample scales and partitioning schemes and its error correction[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2015, 31(15): 272-278. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2015.15.037 <http://www.tcsae.org>

## 0 引言

农业是一个国家国民经济的基础产业, 而粮食则是人们日常生活的必需品, 关系着一个国家的安全、稳定和发展。中国作为世界上人口最多的国家, 同时也是世界上粮食生产和消费最大的国家, 其粮食产量统计数据具有重要的研究价值和意义。粮食产量统计数据是农业经济统计数据的重要组成部分, 同时也是社会经济统计数据不可或缺的一部分。截至目前, 中国基本上所有的社会经济统计数据都是基于行政区划 (县、市) 进行统计和发布的, 空间精度低, 缺乏对数据空间分布特征的描述, 同时还具有数据不完整等缺点。在实际应用中, 社会经济统计数据难以和其他数据 (如自然资源数据) 进行综合分析, 这就限制了社会经济统计数据在学科交叉等方面的应用。对社会经济统计数据空间化, 在

一定程度上恢复或者重构其空间分布特征<sup>[1]</sup>, 不仅可以弥补数据的不完整, 而且可以提高数据的空间精度和扩展数据的应用范围。

1990s 以来, 国内外学者对社会经济统计数据空间化问题进行了许多研究和探讨<sup>[2-6]</sup>, 但是关注的重点主要在于 GIS、RS 的人口数据空间化<sup>[7-9]</sup>、基于人口分布和土地利用的人口数据空间化<sup>[10-11]</sup>、国内生产总值的空间化<sup>[12-14]</sup>和森林蓄积量调查数据、草地载畜量调查数据的空间化<sup>[15-16]</sup>等方面。农业生产活动受多种自然条件因素 (海陆位置、气候特征、水文环境、土壤机理、自然灾害、地理变化等) 和社会经济条件因素 (人口、文化、经济、政治等) 的影响, 因此对农业经济统计数据空间化的难度很大, 导致在很长时间内, 国内外对于农业经济统计数据空间化的研究还比较少。

遥感和 GIS 技术的蓬勃发展为农业经济统计数据空间化提供了新的方法。国内外学者开始在农作物种植面积的空间化<sup>[17-27]</sup>、农业生产投入方面的空间化<sup>[28-29]</sup>和农业产值的空间化<sup>[30]</sup>等方面进行了一些研究, 但对粮食产量统计数据空间化的研究较少。2012 年, 刘忠等<sup>[1]</sup>对中国 2000 年粮食产量统计数据进行了空间化, 并在省级、地市级和县级 3 个尺度上, 对其空间化精度进行了一些分析; 2015 年, 廖顺宝等对粮食产量数据空间化建模过程

收稿日期: 2015-05-17 修订日期: 2015-06-25

基金项目: 中国科学院战略性先导科技专项 (XDA05050000)

作者简介: 姬广兴, 男, 河南商丘人, 研究方向为属性数据空间化。开封 河南大学环境与规划学院, 475004。Email: 284174322@qq.com

\*通信作者: 廖顺宝, 男, 四川德阳人, 博士, 教授, 博士生导师, 主要从事遥感与 GIS 应用以及属性数据空间化方面的教学与研究工作。开封 河南大学环境与规划学院, 475004。Email: liaosb@igsnnr.ac.cn



中样本尺度及线性回归方程常数项设置问题进行了分析和探讨。但这些研究未深入分析不同的样本尺度和分区方案与空间化误差的关系。

因此，本研究基于不同分区方案和粮食产量数据，研究探寻不同样本尺度和分区方案与空间化误差的关系，对提高空间化精度具有重要意义。

1 数据与分析

1.1 数据来源

本研究使用的数据主要包括 3 部分：

1) 2005 年全国各县粮食产量统计数据 该数据来源于全国 31 个省、直辖市和自治区（除台湾、香港和澳门）的 2006 年统计年鉴。

2) 2005 年全国 1 km×1 km 土地覆被类型数据集 该数据集由国家科技基础条件平台下属的地球系统科学数据共享平台提供，包括 25 种土地覆被类型在全国每个栅格（1 km×1 km）中所占面积百分比。

3) 2005 年全国县级行政区划图 该数据包括全国各县行政区划界线矢量数据以及行政区名称、行政区划代码等属性数据。

1.2 数据预处理

数据预处理主要包括以下内容：

1) 统计各县市不同农田类型的面积。在 ArcGIS 软件中，利用县级行政区划图，对不同农田类型的栅格图层进行区域统计运算，得到每个行政区划（县、市、区）的不同农田类型（水田、水浇地和旱地，其分类信息及依据见表 1）的面积数据。

表 1 全国 1:25 万农田覆被分类系统  
Table 1 Farmland classification system of land cover of China at scale of 1:250000

一级类型 First classification	二级类型 Second classification	编码 Code	含义 Definition
农田 Farmland	水田	31	有水源保证和灌溉设施，在一般年景能正常灌溉，用以种植水稻、莲藕等水生农作物的耕地，包括实行水稻和旱地作物轮种的耕地
	水浇地	32	有水源和灌溉设施，在一般年景下能正常灌溉的耕地；以种菜为主的耕地；正常轮作的休闲地和轮歇地
	旱地	33	无灌溉水源及设施，靠天然降水生长作物的耕地

2) 将全国各县级粮食产量统计数据与其不同农田类型面积数据进行匹配。以名称和行政代码一致为标准进行匹配，然后对匹配结果进行检验。首先，以名称和行政代码一致为标准，删除数据的重复项，防止建模时在同一数据多次计算，影响模型精度；其次，删除一些不合理的数据，包括以下 2 种情况：①该县市有农田面积却无粮食产量统计数据；②该县市无农田面积却有粮食产量统计数据。

通过上述方法检验和甄别，本研究最终获得 2321 条县级数据记录。地市级的不同农田类型面积数据和粮食产量数据可由该地级市下属县区的数据加和所得，以此方法得到 349 条地市级数据记录。

1.3 建模及模型验证方案

1.3.1 建模方案

在自然条件和农业生产技术条件基本一致的情况下，粮食产量与农田面积成正比例关系，而不同的农田类型对粮食产量的影响不同，同时统计分析建模是实现粮食产量数据空间化的一种常用方法，因此，本研究选用多元线性回归分析方法建模。在建模时，既可以选取地市级尺度的数据建模，也可以选取县级尺度的数据建模，还可以选取公里网格尺度上（区域平均粮食产量）的数据建模，但利用不同尺度的样本数据建模必然会导致不同的空间化误差。同时，在实际情况中不同区域的自然条件和农业生产技术条件难以一致，不同的分区方案必然导致不同的空间化误差。另外，全国 1 km×1 km 土地覆被类型数据集来源于全国 1:25 万土地覆被类型数据集，该数据集在实际制图中，依照制图的基本规则，会把面积较小的农田类型划分为其他土地覆被类型，存在一定的误差。因此，建模时，需考虑样本尺度、分区方案和回归方程常数项设置（0 与非 0）对模型的影响。

基于获得的 2 321 条县级数据和 349 条地市级数据建立粮食产量预测模型。建模时，按照全国不分区、全国分为 7 个区以及按省分区 3 种分区方案，选择 3 种尺度（地市级尺度、县级尺度、公里网格尺度）上的县级粮食总产、县平均粮食产量、地市粮食总产和地市平均粮食产量 4 种样本分别作为因变量，以其对应的 3 种农田类型（水田、水浇地和旱地）面积数据为自变量，在 Excel 中，运用多元线性回归分析方法，建立 20 种模型。根据廖顺宝等<sup>[31]</sup>关于样本尺度和常数项取值问题的研究结果，去掉 5 种模型，最终，得到 15 模型（表 2），141 个回归方程。其中，平均粮食产量等于区域粮食总产除以总面积。在按省分区方案下，以地市粮食总产和地市平均粮食产量为分析样本，样本数量较少，构建的模型不可信，因此未对此情况建模。7 个区包括：东北地区、华东地区（除台湾省）、华南地区（除香港和澳门）、华北地区、华中地区、西南地区和西北地区。

表 2 15 种全国粮食估算模型描述  
Table 2 Description of 15 models for national crop yield estimation

模型 Models	常数项 Constant term	样本 Samples	分区方案 Partitioning schemes
1	无 No	县级粮食总产	不分区
2			7 个区
3			按省分区
4		地市粮食总产	不分区
5			7 个区
6	有 Have	县平均粮食产量	不分区
7			7 个区
8			按省分区
9		地市平均粮食产量	不分区
10			7 个区
11	无 No	县平均粮食产量	不分区
12			7 个区
13			按省分区
14		地市平均粮食产量	不分区
15			7 个区



在 ArcGIS 软件中,使用 Raster Calculator 工具,以 3 个农田类型(水田、水浇地、旱地)的栅格图层作为输入参数,利用构建的 15 种模型进行栅格计算,得到 15 种粮食产量公里网格分布图。其中,台湾、香港和澳门在运算时都设为 NoData。

1.3.2 模型验证

在自然要素数据(如气温、降水等)空间化误差分析中,建模数据不能用于验证,需单独选取一些样本(不参加建模)验证空间化结果。但社会经济统计数据空间化与自然要素空间化有一定差异,社会经济数据空间化的最终结果要保证在建模单元尺度或更小行政尺度(地市、县等)上使空间化结果与统计数据一致。因此,再单独选取一些样本(不参加建模)进行验证的意义不是很大,因为空间化最终结果与统计数据基本是完全一致的。基于上述考虑,本文在建模时选择全部样本而不是部分样本,验证也是基于全部样本建模时所得到的误差评判参数。

1.3.3 模型评价

作为一种数据处理方法,属性数据空间化必然会存在一定的误差<sup>[31]</sup>。因此,需选取误差评价因子,对空间化误差进行分析,探寻减少误差的方法或方案,提高其精度。本文采用两阶段误差分析方法对空间化误差进行分析,包括模型误差分析和空间化结果误差分析。回归模型的评价指标有多种,常用的有复相关系数、决定系数、校正决定系数、显著性检验、剩余标准差和相对误差等指标<sup>[31-35]</sup>。复相关系数和决定系数只能评价自身模型的拟合程度,不能作为评价不同模型优劣的指标;对 15 模型进行  $F$  值检验,发现 15 种模型都在  $\alpha=0.005$  水平上显著相关;由于本文 15 模型的变量单位不完全一致,因此剩余标准差不适合被用来评价 15 模型的优劣。因此,

本文选取模型的校正决定系数和平均相对误差 2 个因子作为模型的评价因子,对 15 种模型进行误差分析。

在已有研究中,学者们主要选取空间化结果的决定系数、相对误差和绝对误差等指标的一种或几种<sup>[31,36-39]</sup>评价空间化结果。本文除考虑上述评价指标以外,还根据实际粮食生产情况选取 2 个指标,空间化结果的零值栅格所占百分比(实际粮食生产中,粮食产量不会为负值)和栅格单元最大值与全国水稻单位面积产量的相对误差(水稻是中国单位面积产量最高的农作物,理论上,水稻单位面积产量可被认为是实际生产中公里网格上最高的粮食产量值)。综上,本文选取空间化结果的零值栅格所占百分比、粮食总产量相对误差、栅格单元最大值与全国水稻单位面积产量的相对误差、平均绝对误差和决定系数 5 个误差评价因子对空间化结果进行误差分析。其中,2005 年全国粮食总产量统计数据为  $4.84 \times 10^8$  t、2005 年水稻单位面积产量约为 626 t/km<sup>2</sup>。

2 结果与分析

2.1 模型模拟精度比较

15 种粮食产量估算模型模拟精度用模型校正决定系数和平均相对误差来评价,如表 3 所示。结果表明:1)常数项为 0 的模型,模型精度随着分区方案的细化而提高,而且比常数项非 0 的模型精度高;2)在全国不分区和分为 7 个区 2 种情况下,不论以粮食总产还是平均粮食产量为分析样本,以地市级数据构建的模型比以县级数据构建的模型精度高;3)无论采用哪种分区方案,以县级总产为分析样本的模型精度比以县平均粮食产量为分析样本的模型精度高;4)以地市粮食总产为分析样本、常数项为 0、全国分为 7 个区建模获得的模拟精度最高。

表 3 15 种中国公里网格粮食产量模型精度评价及其空间分布的误差评价  
Table 3 Simulation assessment of 15 models for national grain yield estimation and error evaluation of 15 distribution of grain yield in China at 1 km by 1 km

模型 Models	模型拟合评价 Assessment on model simulation		空间化分布评价 Assessment on spatial distribution				
	校正决定系数 Adjusted determination coefficient	平均相对误差 Average of relative error/%	负值栅格所占百分比 Percentage of negative grid/%	粮食总产量相对误差 Relative error of total grain yield/%	栅格单元粮食产量最大值与水稻单位面积产量的相对误差 Relative error between maximum grain yield of grid and rice yield per unit area/%	平均绝对误差 Mean absolute difference/t	决定系数 Determination coefficient $R^2$
1	0.556	185.16	0	5.67	35.46	93716	0.550
2	0.657	140.42	0	3.31	20.77	71852	0.655
3	0.762	117.62	0.52	9.72	678.59	75077	0.518
4	0.698	102.32	0	10.43	29.87	94522	0.549
5	0.806	99.51	0	7.28	13.58	79925	0.647
6	0.316	515.82	0	37.73	21.88	111753	0.526
7	0.357	2031.44	1.14	34.23	16.13	125638	0.474
8	0.405	390.84	2.59	2.66	1008.15	127867	0.350
9	0.735	1054.41	66.92	7.94	29.07	118352	0.451
10	0.818	1158.44	65.77	1.18	187.22	120700	0.427
11	0.318	251.14	0	31.58	23.64	106631	0.523
12	0.346	228.14	0	30.24	7.67	99410	0.589
13	0.380	168.37	0.74	23.98	1011.98	82579	0.676
14	0.734	134.67	0	26.49	29.71	103846	0.524
15	0.800	211.04	0.46	21.85	134.82	94002	0.596



## 2.2 粮食空间化结果误差分析

根据建立的 15 种粮食预测模型, 获得 15 种粮食产量空间分布图, 应用零值栅格所占百分比、粮食总产量相对误差、栅格单元最大值与全国水稻单位面积产量的相对误差、平均绝对误差和决定系数 5 个误差评价因子空间化结果进行评价, 如表 3 所示。结果表明:

1) 不论模型是基于县级粮食总产、县平均粮食产量, 还是地市粮食总产、地市平均粮食产量, 对常数项为 0 的模型而言, 全国分为 7 个区的空间化结果精度最优, 全国不分区的空间化结果的精度次之, 分省模型计算出的空间化结果的精度最差。

2) 不论模型是基于县平均粮食产量, 还是地市平均粮食产量, 对常数项非 0 的模型而言, 全国不分区的空间化结果精度最优, 全国分为 7 个区的空间化结果的精度次之, 分省模型计算出的空间化结果的精度最差。

3) 在全国不分区和分为 7 个区 2 种情况下, 以粮食总产为样本的模型的空间化结果比以其平均粮食产量为样本的模型的空间化结果精度要高。在这 2 种情况下, 以粮食总产为分析样本, 基于县级数据的模型计算出的空间化结果的精度比基于地市级数据的模型计算出的空间化结果的精度高。

综上, 以县级粮食总产为样本、常数项为 0、全国分为 7 个区的建模方案计算出的粮食产量公里网格分布图精度最高。

## 2.3 粮食产量空间化模拟误差修正

表 3 表明, 以县级粮食总产为样本、常数项为 0、全国分为 7 个区的建模方案计算的粮食产量精度最高, 但与其与统计值的相关性仍然较低, 决定系数只有 0.655, 有待于进一步提高。因此, 笔者基于比例系数法对空间化结果进行了误差修正, 即把每个县级行政单元的误差按比例系数分配到该县的每个栅格上, 比例系数等于各县粮食产量的统计值除以预测值, 具体处理步骤如下: 1) 统计各县粮食产量预测值, 与粮食产量统计值匹配; 2) 用各县的粮食产量统计值除以预测值, 得到各县的比例系数, 其中一些区域无粮食产量统计数据, 比例系数均设为 1, 即不进行修正; 3) 以各县的比例系数为栅格化属性代码, 对县级行政区划进行栅格化, 得到以县级行政区划为单元的比例系数栅格图层; 4) 用空间化结果乘以比例系数栅格图层, 得到修正后的粮食产量空间化结果。理论上, 修正后的结果即是最终的粮食产量公里网格分布图, 但通过对其栅格单元数值的观察, 发现有些栅格单元的数值相当大, 最高达 10 000 t/km<sup>2</sup>, 不符合实际粮食生产情况, 因此需对这些栅格单元进行处理。分析空间化结果中公里网格粮食产量与其对应栅格单元数量的关系, 发现两者存在对数函数拟合关系 (如图 1)。由图 1 方程计算可知, 当栅格单元的数值为 767 t/km<sup>2</sup>, 其栅格单元的数量趋近于 0。其中, 粮食产量 > 767 t/km<sup>2</sup> 的栅格单元面积仅占全国总面积的 0.14%。因此, 把大于 767 t/km<sup>2</sup> 的栅格单元数值设定为 767 t/km<sup>2</sup>, 其余保持不变, 使空间化结果更符合实际。因此, 用已获得的栅格

图层乘以 1 个系数 0.924 (粮食产量统计总值除以预测总值), 得到最终的粮食产量公里网格分布图。

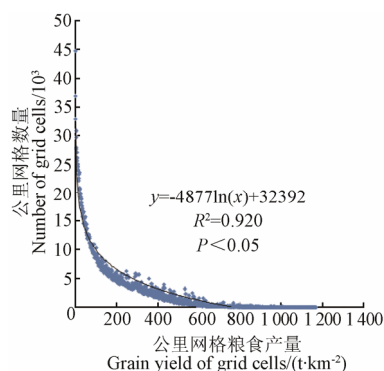


图 1 公里网格粮食产量与相应网格数量分布情况

Fig.1 Relationship between grain yield of grid cells and number of corresponding grid cells

最终粮食产量公里网格分布如图 2 所示, 中国粮食生产的地理分界线呈现东北-西南走势。

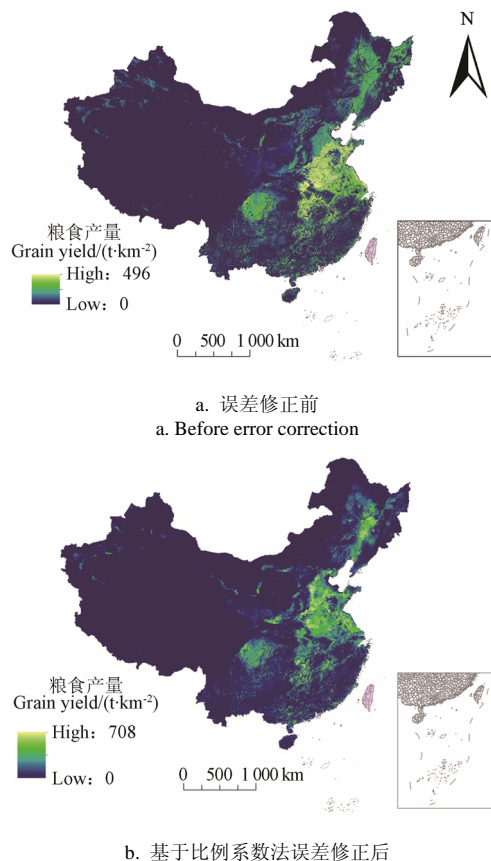


图 2 2005 年中国公里网格粮食产量分布图修正前后比较  
Fig.2 Comparison of grain yield distribution in China at 1 km by 1 km in 2005 before and after error correction

中国粮食主产区分布在华北平原、长江中下游平原、东北平原和成都平原, 而华北平原是粮食产量最高



的地区。与 2005 年全国十大粮食生产先进县中部分县市的平均粮食产量相比 (湘潭县为  $742.46 \text{ t/km}^2$ , 滑县为  $658.16 \text{ t/km}^2$ , 襄阳区为  $700.49 \text{ t/km}^2$ , 五常市为  $685.86 \text{ t/km}^2$ ), 图 2a 中公里网格最大值为  $496 \text{ t/km}^2$ , 比实际的粮食产量低, 而图 2b 中公里网格最大值为  $708 \text{ t/km}^2$ , 与实际的粮食产量相符。同时, 修正后, 各县预测值与统计值的决定系数由 0.655 提高到 0.968, 平均绝对误差从 71 852 t 下降到 21573t, 粮食总产量相对误差也由 3.31% 降到 0.04%, 表明利用比例系数法修正的粮食产量空间化结果的精度在县级尺度上有了很大的提高。刘忠等<sup>[1]</sup>基于土地利用和人口密度进行粮食空间化, 县级尺度上空间化的决定系数只有 0.640。比较而言, 本文基于样本尺度和分区方案的空间化方法精度更高。

### 3 结论与讨论

本文在综合考虑样本尺度 (地市级尺度、县级尺度、公里网格尺度)、分区方案 (全国不分区、全国分为 7 个区和按省分区) 和回归方程常数项 (0 与非 0) 等因素的情况下, 以县级粮食总产、县平均粮食产量、地市粮食总产和地市平均粮食产量 4 种样本分别作为因变量, 以对应的 3 种农田类型 (水田、水浇地和旱地) 面积数据为自变量, 运用多元线性回归分析方法, 建立 15 种空间化模型, 并获得 15 种粮食产量公里网格分布。基于误差评价, 采用比例系数法修正了空间化误差。结果表明: 1) 模型精度与空间化结果精度存在不一致性。精度最高的模型计算出的空间化结果精度不是最高; 2) 常数项为 0 的模型, 其空间化结果精度随着分区方案的细化先提高再降低。常数项非 0 的模型, 其空间化结果精度随着分区方案的细化而降低; 3) 在全国不分区和分为 7 个区 2 种情况下, 空间化结果精度随着样本尺度的细化 (从地市级到县级再到公里网格) 先提高后降低; 4) 与其他模型相比, 在以县级粮食总产为分析样本、常数项为 0、全国分为 7 个区建模的情况下, 空间化结果的精度最高 (决定系数为 0.655), 利用比例系数法对其误差进行修正, 进一步提高了空间化精度 (决定系数为 0.968)。

该研究弥补了粮食产量空间化误差分析的不足, 探寻了不同样本尺度和分区方案与空间化误差的关系, 提高了空间化精度, 同时对其他类型的社会经济统计数据 (如人口、GDP 等) 空间化研究具有一定的参考价值。在今后的研究中, 有必要考虑更多的因子和方法, 以进一步提高空间化结果的精度。可以从以下 2 方面进行改进:

1) 细分不同农作物的产量。不同的农田类型种植的农作物不同, 因此, 可以考虑细分不同农作物的产量, 利用不同农作物的产量数据和对应的农田类型面积数据建模, 进行空间化。

2) 利用不同的修正方法对空间化结果修正。本文所用的误差修正方法为比例系数法, 也可以利用其他的修正方法 (如残差修正) 对结果进行修正并对比分析, 提高空间化精度。

### 【参 考 文 献】

- [1] 刘忠, 李保国. 基于土地利用和人口密度的中国粮食产量空间化[J]. 农业工程学报, 2012, 28(9): 1—7.  
Liu Zhong, Li Baoguo. Spatial distribution of China grain output based on land use and population density[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(9): 1—7. (in Chinese with English abstract)
- [2] Elvidge C D, Baugh K E, Kihn E A, et al. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption[J]. International Journal of Remote Sensing, 1997, 18(6): 1373—1379.
- [3] Harvey J T. Population estimation models based on individual TM pixels[J]. Photogrammetric Engineering & Remote Sensing, 2002, 68(11): 1181—1192.
- [4] 刘纪远, 岳天祥, 王英安, 等. 中国人口密度数字模拟[J]. 地理学报, 2003, 58(1): 17—24.  
Liu Jiyuan, Yue Tianxiang, Wang Ying'an, et al. Digital simulation of population density of China[J]. Acta Geographica Sinica, 2003, 58(1): 17—24. (in Chinese with English abstract)
- [5] 江东, 杨小唤, 王乃斌, 等. 基于 RS、GIS 的人口空间分布研究[J]. 地球科学进展, 2002, 17(5): 734—738.  
Jiang Dong, Yang Xiaohuan, Wang Naibin, et al. Study on spatial distribution of population based on remote sensing and GIS[J]. Advance in Earth Sciences, 2002, 17(5): 734—738. (in Chinese with English abstract)
- [6] 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 Shunbao, Sun Jiulin. GIS based spatialization of population census data in Qinghai-Tibet plateau[J]. Acta Geographica Sinica, 2003, 58(1): 25—33. (in Chinese with English abstract)
- [7] Lo C P. Modeling the population of China using DMSP operational linescan system nighttime data[J]. Photogrammetric Engineering and Remote Sensing, 2001, 67(9): 1037—1047.
- [8] Sutton P, Roberts D, Elvidge C, et al. Census from Heaven: An estimate of the global human population using night-time satellite imagery[J]. International Journal of Remote Sensing, 2001, 22(16): 3061—3076.
- [9] 吕安民, 李成名, 林宗坚, 等. 基于 GIS 的人口信息提取[J]. 清华大学学报: 自然科学版, 2005, 45(9): 1189—1192.  
Lu Anmin, Li Chengming, Lin Zongjian, et al. Population data abstraction based on GIS[J]. Journal of Tsinghua University: Science and Technology, 2005, 45(9): 1189—1192. (in Chinese with English abstract)
- [10] 廖顺宝, 李泽辉. 基于人口分布与土地利用关系的人口数据空间化研究——以西藏自治区为例[J]. 自然资源学报, 2003, 18(6): 659—665.  
Liao Shunbao, Li Zehui. Study on spatialization of population census data based on relationship between population distribution and land use: Taking Tibet as an example[J]. Journal of Natural Resources, 2003, 18(6): 659—665. (in Chinese with English abstract)
- [11] 廖顺宝, 李泽辉. 四川省人口分布与土地利用的关系及人口数据空间化试验[J]. 长江流域资源与环境, 2004, 13(6): 557—561.  
Liao Shunbao, Li Zehui. Relationship between population



- distribution and land use and spatialization of population census data[J]. *Resources and Environment in the Yangtze Basin*, 2004, 13(6): 557—561. (in Chinese with English abstract)
- [12] 易玲, 熊利亚, 杨小唤. 基于 GIS 技术的 GDP 空间化处理方法[J]. *甘肃科学学报*, 2006, 18(2): 54—58.  
Yi Ling, Xiong Liya, Yang Xiaohuan. Method of pixelizing GDP data based on the GIS[J]. *Journal of Gansu Sciences*, 2006, 18(2): 54—58. (in Chinese with English abstract)
- [13] 刘红辉, 江东, 杨小唤, 等. 基于遥感的全国 GDP 1km 格网的空间化表达[J]. *地球信息科学*, 2005, 7(2): 120—123.  
Liu Honghui, Jiang Dong, Yang Xiaohuan, et al. Spatialization approach to 1 km grid GDP supported by remote sensing[J]. *Geo-information Science*, 2005, 7(2): 120—123. (in Chinese with English abstract)
- [14] 黄莹, 包安明, 陈曦, 等. 基于绿洲土地利用的区域 GDP 公里格网化研究[J]. *冰川冻土*, 2009, 31(1): 158—165.  
Huang Ying, Bao Anming, Chen Xi, et al. Studying the GDP in 1 km square grid-cells based on oasis land use[J]. *Journal of Glaciology and Geocryology*, 2009, 31(1): 158—165. (in Chinese with English abstract)
- [15] 廖顺宝, 许立民. 森林蓄积量调查数据空间化的方法研究[J]. *中南林业科技大学学报*, 2013, 33(11): 1—7.  
Liao Shunbao, Xu Limin. Methodology for spatialization of timber volume data from forest inventory[J]. *Journal of Central South University of Forestry & Technology*, 2013, 33(11): 1—7. (in Chinese with English abstract)
- [16] 廖顺宝, 秦耀辰. 草地理论载畜量调查数据空间化方法及应用[J]. *地理研究*, 2014, 33(1): 179—190.  
Liao Shunbao, Qin Yaochen. A spatialization method for survey data of theoretical stock-carrying capacity of grassland in China and its application[J]. *Geographical Research*, 2014, 33(1): 179—190. (in Chinese with English abstract)
- [17] Froking S, Qiu J, Boles S, et al. Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China[J]. *Global Biogeochemical Cycles*, 2002, 16(4): 38-1—38-10.
- [18] Qiu J, Tang H, Froking S, et al. Mapping single -, double -, and triple - crop agriculture in China at  $0.5^\circ \times 0.5^\circ$  by combining county-scale census data with a remote sensing - derived land cover map[J]. *Geocarto International*, 2003, 18(2): 3—13.
- [19] Leff B, Ramankutty N, Foley J A. Geographic distribution of major crops across the world[J]. *Global Biogeochemical Cycles*, 2004, 18(1): 1—27.
- [20] Liangzhi You, Stanley Wood. Assessing the spatial distribution of crop areas using a cross-entropy method[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2005, 7(4): 310—323.
- [21] Liangzhi You, Stanley Wood. An entropy approach to spatial disaggregation of agricultural production[J]. *Agricultural Systems*, 2006, 90(1/2/3): 329—347.
- [22] Liangzhi You, Stanley Wood, Wood-Sichra U. Generating plausible crop distribution maps for Sub-Saharan Africa using a spatially disaggregated data fusion and optimization approach[J]. *Agricultural Systems*, 2009, 99(2): 126—140.
- [23] Maxwell S K, Hoffer R M. Mapping agriculture crops with multi-date Landsat data[C]. *Proceedings of the National ASPRS/ACSM 1996 Annual Convention*. Baltimore, MD: American Society for Photogrammetry and Remote sensing, 1996: 433—443.
- [24] Maxwell S K, Nuckols J R, Ward M H, et al. An automated approach to mapping corn from Landsat imagery[J]. *Computers and Electronics in Agriculture*, 2004, 43(1): 43—54.
- [25] Maxwell S K, Nuckols J R, Ward M H. A method for mapping corn using the US Geological Survey 1992 National Land Cover Dataset[J]. *Computers and Electronics in Agriculture*, 2006, 51(1): 54—65.
- [26] Monfreda C, Ramankutty N, Foley J A. Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000[J]. *Global Biogeochemical Cycles*, 2008, 22(1): 1—19.
- [27] Khan M R, De Bie C, Van Keulen H, et al. Disaggregating and mapping crop statistics using hypertemporal remote sensing[J]. *International Journal of Applied Earth Observation and Geo-information*, 2010, 12(1): 36—46.
- [28] Potter P, Ramankutty N, Bennett E M, et al. Characterizing the spatial patterns of global fertilizer application and manure production[J]. *Earth Interactions*, 2010, 14(2): 1—22.
- [29] 孙丹峰, 王雅, 李红, 等. 基于 MODIS NDVI 年序列的区域化肥投入空间化方法[J]. *农业工程学报*, 2010, 26(6): 175—180.  
Sun Danfeng, Wang Ya, Li Hong, et al. Spatializing regional fertilizer input based on MODIS NDVI time series[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2010, 26(6): 175—180. (in Chinese with English abstract)
- [30] 张晶, 吴绍洪, 刘燕华, 等. 土地利用和地形因子影响下的西藏农业产值空间化模拟[J]. *农业工程学报*, 2007, 23(4): 59—65.  
Zhang Jing, Wu Shaohong, Liu Yanhua, et al. Simulation of distribution of agricultural output value influenced by landuse and topographical indices in Tibet[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2007, 23(4): 59—65. (in Chinese with English abstract)
- [31] 廖顺宝, 张赛. 属性数据空间化误差评价指标体系研究[J]. *地球信息科学学报*, 2009, 11(2): 176—182.  
Liao Shunbao, Zhang Sai. Study on error evaluating index for spatialisation of attribute data[J]. *Journal of Geo-information Science*, 2009, 11(2): 176—182. (in Chinese with English abstract)
- [32] Vicente Serrano S M, Sánchez S, Cuadrat J M. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature[J]. *Climate Research*, 2003, 24(2): 161—180.
- [33] Stahl K, Moore R D, Floyer J A, et al. Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density[J]. *Agricultural & Forest Meteorology*, 2006, 139: 224—236.
- [34] Zeng Chuiqing, Zhou Yi, Wang Shixin, 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.
- [35] 徐建华. 计量地理学[M]. 北京: 高等教育出版社, 2006: 47—71.
- [36] Liu X H, Kyriakidis P C, Goodchild M F. Population-density estimation using regression and area-to-point residual kriging[J]. *International Journal of Geographical Information Science*, 2008, 22(4): 431—447.



- [37] Yang Xiaohuan, Huang Yaohuan, Dong Pinliang, et al. An updating system for the gridded population database of China based on remote sensing, GIS and spatial database technologies[J]. *Sensors*, 2009, 9(2): 1128—1140.
- [38] 叶靖, 杨小唤, 江东. 乡镇级人口统计数据空间化的格网尺度效应分析——以义乌市为例[J]. *地球信息科学学报*, 2010, 12(1): 40—47.
- Ye Jing, Yang Xiaohuan, Jiang Dong. The grid scale effect analysis on town leveled population statistical data spatialization[J]. *Journal of Geo-information Science*, 2010, 12(1): 40—47. (in Chinese with English abstract)
- [39] 廖顺宝, 张赛. 多年平均气温数据空间化误差的尺度效应[J]. *地球信息科学学报*, 2014, 16(1): 8—14.
- Liao Shunbao, Zhang Sai. Scale effect of errors on spatialization of annual mean air temperature data[J]. *Journal of Geo-information Science*, 2014, 16(1): 8—14. (in Chinese with English abstract)

## Spatial distribution of grain yield based on different sample scales and partitioning schemes and its error correction

Ji Guangxing<sup>1</sup>, Liao Shunbao<sup>1,2\*</sup>, Yue Yanlin<sup>1</sup>, Hou Pengmin<sup>1</sup>, Yang Xu<sup>1</sup>

(1. College of Environment and Planning, Henan University, Kaifeng 475004, China;

2. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China)

**Abstract:** Spatialization of grain yield can contribute to comprehensive analysis of grain yield with other natural and cultural data. Grain production has a close relationship with the distribution of farmland. Therefore, information on spatial distribution of farmland is an important parameter for spatialization of grain yield, and the statistical analysis and modeling are the basic means to realize spatialization of grain yield. Spatialization of nationwide grain yield relates to sample scales and partitioning schemes. Different sample scales and partitioning schemes will inevitably lead to different errors of spatialization. In this paper, models considering farmland distribution and sample scales and partition schemes were proposed to estimate grain yield and its spatial distribution. The grain yield data were collected from 2005 Yellow Book of China. Data of paddy field, irrigated land, and dry land areas in each county or district were calculated. Four datasets of 3 scales were selected including total grain yields of counties, total grain yields of prefectures and their average grain yields. A total of 2321 county data and 349 prefecture-level data were obtained. 3 partitioning schemes (no partition of China, 7 regions of China, partitions of China by province) were considered. A total of 15 kinds of multiple variable linear models were constructed with area of different types of farmland as independent variables, grain yields as dependent variables. The results showed that: 1) Based on model fitness of grain yield and its spatialization results, optimal models could be selected since the model fitness suggested that the model without constant term based on prefecture-level data and 7 regions was best but the spatialization results indicated that the model without constant term based on county-level data and 7 regions was best; 2) For models without constant term, precision of spatialization results increased first and then decreased with scaling down of partitioning scheme; For models with constant term, precision of spatialization results decreased with scaling down of partitioning scheme; 3) In the 2 partitioning schemes (no partition of China and 7 regions of China), the precision of spatialization results increased first and then decreased with scaling down of samples from prefecture level to county level and 1 km by 1 km level; and 4) Compared with other models, in the case of county grain yields as samples, the model without constant term and 7 regions of China had the highest precision with coefficient of determination of 0.655. The spatialization results were modified with error by a proportional coefficient method, and the precision was improved to coefficient of determination of 0.968. This research made up for the deficiency of spatial error analysis of grain yield, explored the relationship between different sample scales and partitioning schemes and spatial error. Meanwhile, it also provided valuable information for other types of social and economic statistical data.

**Key words:** grain; error correction; models; sample scale; partitioning scheme; multiple variable regression; yield; spatialization