

中国传统农区历史耕地数据网格化方法

林珊珊^{1,2}, 郑景云¹, 何凡能¹

(1. 中国科学院地理科学与资源研究所, 北京 100101;

2. 中国科学院研究生院, 北京 100049)

摘要: 建立具有空间属性的历史时期土地覆被数据集有助于更好地模拟土地覆被变化的气候与生态效应。根据我国历史时期土地开发利用的特点, 深入分析了影响我国历史时期土地开发利用的主导因子, 量化了海拔高度、坡度、人口分布等自然及人文要素与耕地分布的关系, 并以此为依据设计了一套将我国传统农区历史耕地数据网格化的方法与算法模型, 计算各网格 (分辨率为 $60\text{ km} \times 60\text{ km}$) 耕地面积占整个农区耕地总面积的比重、以及各网格的耕地面积与垦殖率; 同时利用基于《嘉庆重修一统志》重建的中国传统农区历史耕地资料, 采用上述方法与算法建立了研究区内 1820 年 (清嘉庆二十五年) $60\text{ km} \times 60\text{ km}$ 空间尺度的耕地数据集, 并绘制了分布图。重建结果与册载府级田亩数据的比较分析表明: 该方法与算法可以有效地将以行政区域为统计单元的中国历史耕地数据量化为具有统一且更高空间分辨率的网格化数据集。

关键词: 中国传统农区; 历史时期; 耕地数据; 网格化方法

1 引言

人类活动已经在很大程度上改变了地表格局, 大范围的土地覆被变化对大气成分、气候变化、区域水循环与碳循环、环境质量、生物多样性及陆地生态系统的脆弱性与生产力都有深刻的影响^[1]。但土地系统对地球系统的影响是极为复杂的非线性过程, 目前的土地覆被变化环境效应 (特别是气候效应) 的模拟也很薄弱, 仍有许多关键问题亟待解决; 而如何获取高分辨率与高精度的土地覆被数据是解决这些问题的基础^[2]。尽管在此方面, 遥感技术的应用已为各种尺度土地覆被数据的获取提供了重要途径, 但遥感数据不过覆盖近几十年, 而在此之前, 特别是历史时期的土地覆被变化数据仍非常有限。

中国由于拥有悠久的农业开发史和丰富的历史文献被认为是最有条件开展历史时期土地利用 / 覆被变化重建的国家之一。近年来, 有许多学者对我国历史时期, 特别是清代以来不同地区的土地利用进行了定量重建研究^[3-6], 但要深入分析土地利用 / 覆被变化的环境效应, 则还需要土地覆被的空间属性, 即需要建立一种将历史土地利用属性数据在地图网格中进行分配的方法^[7]。Ramnkutty^[7]与 Goldewijk^[8]曾分别设计了两套耕地统计数据的分配方法, 并建立了过去 300 年的全球土地覆被数据集 (分辨率为 $0.5^\circ \times 0.5^\circ$), 但从区域尺度看, 这两套数据集无论是引用的耕地资料精度, 还是耕地数据分配方法都过于粗略。因此, 本文希望通过借鉴他们的思路, 发挥中国史料丰富的优势, 设计一套更适合于中国历史时期耕地数据网格化再现的方法。

本文选取我国传统农区为研究对象, 大致包括华北、江南与西南地区 (不包括台湾省); 由于我们采用清嘉庆二十五年 (1820 年) 作为研究实例, 故而主要采用 1820 年的行

收稿日期: 2007-03-26; 修订日期: 2007-09-27

基金项目: 国家自然科学基金项目 (40471007); 中国科学院知识创新工程重要方向项目 (KZCX2-YW-315)

[Foundation: National Natural Science Foundation of China, No.40471007; Knowledge Innovation Program of CAS, No.KZCX2-YW-315]

作者简介: 林珊珊 (1982-), 女, 福建闽侯人, 硕士。主要从事历史时期环境变化研究。E-mail: summer.lin33@gmail.com

政边界作为研究区范围的界定依据,大致包括当时内地 17 个行省,其中直隶、山西不包括长城以北地区,福建省不包括台湾府(图 1)。研究内容主要包括:遴选量化影响耕地分布的主导因子,并利用 GIS 手段设计、建立历史耕地数据网格化重建模型;在此基础上,根据这一模型对 1820 年省级耕地数据进行网格化处理,建立起该区域 1820 年网格化耕地数据集,并绘制分布图;为检验结果有效性,我们将耕地数据的网格化结果与 1820 年府级册载田亩数据进行比较验证。



图 1 研究区范围

(依据 1820 年行政区划图)

Fig. 1 Location of the study area

2 中国传统农区历史耕地数据网格化模型设计

中国历史时期的田亩数据多按行政区划(如省、府、县等)进行记载,因此选择适当的指标量化某一地区内各网格的耕地面积比例成为模型设计的关键。为与目前常用的区域气候模式分辨率对应^[9],本文将研究区的基本单元,即网格大小设定为 $60\text{ km}\times 60\text{ km}$ 。

2.1 影响因子的遴选

耕地分布受地形、热量、水分、土壤、植被等自然因素以及人口状况、经济发展水平、农业政策、战争等社会因素的共同影响。但在不同地区或同一地区不同空间尺度上,其主导影响因子是不同的。首先,从水热条件看,研究区位于东部季风区,虽然水热条件对农业生产有极其重要的影响,但其作用主要体现在作物种类、熟制及农业灾害的差异上,而对耕地的分布并无明显影响。其次,这一地区作为历史悠久的农垦区,经过长期的森林开采与农业耕种,地势相对平坦的地方已经几乎没有自然植被残留,土壤普遍熟化,甚至许多低产土壤也开垦为耕地。因此,在本文的研究尺度上,地形是影响耕地分布的最主要自然因素,其中包括海拔高度、地面坡度、地势起伏和坡向等 4 个因子;但由于本研究的分辨率为 $60\text{ km}\times 60\text{ km}$,而在这一尺度上,地势的起伏和坡向对耕地分布影响作用极小,因此,海拔高度与地面坡度是影响本研究区耕地分布的主要自然因素。

虽然地势平坦之地适宜耕种,但土地的开发却也并非一蹴而就,人口、生产力水平、战乱、政策等社会因素也直接影响着历史时期的土地开垦。其中又以人口状况与耕地面积及分布格局联系最为紧密:对于某一特定区域而言,人口总量的消长一般就意味着耕地面积的增减;对于整个研究区而言,人口的迁移往往就指示了农耕区的扩展。从内在机制上看,历史时期中国是一个农业社会,而耕地面积的大小也主要取决于劳动力的多少。虽然生产力水平的进步可以提高土地耕种的效率与产量,但就历史上的中国而言,自从大约 2500 年以前开始使用铁器以来,生产力水平就始终没有显著提高,要想获得更多的粮食就必须仰仗耕地面积与劳动力的增加。因此,对于同一个时期而言,地区性的人口数量变化能够很好地指示耕地的开垦过程。此外,政策与战乱对耕地数量与分布都有十分重要的影响,只是二者的影响都难以直接量化,但仔细分析它们对耕地变化影响的过程不难发现,这种影响首先是作用于进行农事生产的劳动者,进而才导致一个地区是抛荒或是拓垦。因此,政策与战乱对耕地的影响实际上已经间接地反映在人口状况的变化上。从资料 and 数据的获取可能性来看,中国历代典籍中拥有大量有关人口与田亩数据的记载,这是现今学者深入研究中国历史时期社会经济状况的重要资料来源;然而,在历史学和历史地理学研究中,人口史的研究却远较历史耕地的研究系统和详尽。因此,选择人口分布状况作为影响历史时期耕地分布的主导性社会因素,不仅具有代表性,而且也具有可行性。

2.2 影响因子的量化

2.2.1 地形 为量化地形因素对耕地分布的影响，我们以现代资料为依据来分析研究区内海拔高度与坡度这 2 个主要地形因子与耕地分布的关系。首先，从海拔高度看，研究区主要位于海拔 2000 m 以下的一级与二级阶梯上，其中太行上、巫山、雪峰山一线以东地区

的海拔多在 500 m 以下，集中了我国大部分耕地。海拔高度因其热量与水分条件的垂直变化，而造成土地开垦的垂直差异，包括耕地比重、农作物的组合、耕作制度和复种条件等。但对于不同的海拔变化范围，这些差异的显著性大不相同，只有当海拔达到一定高度时，水热条件限制了作物的生长，才会显著地影响耕地面积的分布。其次，从地表坡度看，耕地按地面坡度大小可分为平耕地（坡度 $\leq 2^\circ$ ）和坡耕地（坡度 $> 2^\circ$ ），其中坡耕地又可分缓坡耕地（ $2^\circ \sim 6^\circ$ ）、坡耕地（ $6^\circ \sim 15^\circ$ ）、陡坡耕地（ $15^\circ \sim 25^\circ$ ）与很陡坡耕地（ $> 25^\circ$ ）四级（根据中国科学院自然资源综合考察委员会调查资料分级方法^[10]）。由于地面坡度不同，耕地的水土流失状况也有明显差别，因而会明显影响农业土地利用的方式与土地的垦殖状况。一般讲，地面坡度在 3° 以下的土地，基本无水土侵蚀，可作为较好的耕地； $3 \sim 15^\circ$ 的土地，水土流失比较普遍，但采取水土保持措施后，仍宜于开垦；而 25° 则是坡地开垦的极限，超限开垦会引起严重的水土流失，因而极少被作为耕地垦殖^[10]。为计算 $60 \text{ km} \times 60 \text{ km}$ 网格内的平均海拔 $D(i)$ 与平均坡度 $S(i)$ ，我们借鉴《中国农业土地利用》^[10] 中的海拔高度和坡度分级标准（表 1），以全国 1 km 的 DEM 数据^①与 1 km 坡度数据^②为基础资料，采用面积加权平均方法，计算出 $60 \text{ km} \times 60 \text{ km}$ 网格内的平均海拔 $D(i)$ 与平均坡度 $S(i)$ ；同时根据全国 1 km 土地利用数据库^③，计算每个 $60 \text{ km} \times 60 \text{ km}$ 网格的实有耕地面积和垦殖率 $C(i)$ 。

计算结果显示：在 $D(i)$ 大于 3500 m 的区域内（主要在川西滇北的高原区），其耕地面积占研究区总耕地面积的比重不足 0.1%，且其垦殖率 $C(i)$ 皆低于 5%；因此为计算方便，我们在下文分析海拔高度 $S(i)$ 与坡度 $S(i)$ 这 2 个因子与垦殖率 $C(i)$ 的关系时，将 $D(i) = 3500 \text{ m}$ 作为区内耕地分布（农垦）的上限。

在 $D(i) < 3500 \text{ m}$ 的区域中，我们将网格平均海拔 $D(i)$ 和平均坡度 $S(i)$ 分别与网格垦殖率 $C(i)$ 进行相关分析（表 2），发现二者与垦殖率 $C(i)$ 均呈显著的负相关（相关系数分别为 -0.488 与 -0.694）。但仔细分析研究区的地貌特征不难发现，海拔与坡度的变化具有较好的一致性（二者相关系数达到 0.736）：华北平原、

表 1 中国传统农区海拔与坡度分级

Tab. 1 The classification of Elevation and Slope of the traditional cultivated region in China

海拔高程分级 (m)	计算时的海拔高程值 (m)	坡度分级	计算时的坡度值
≤ 100	100	$\leq 2^\circ$	2°
100~250	250	$2^\circ \sim 6^\circ$	6°
250~500	500	$6^\circ \sim 15^\circ$	15°
500~750	750	$15^\circ \sim 25^\circ$	25°
750~1000	1000	$> 25^\circ$	45°
1000~1500	1500		
1500~2000	2000		
2000~3000	3000		
> 3000	4000		

表 2 中国传统农区海拔 $D(i)$ 、坡度 $S(i)$ 与垦殖率 $C(i)$ 的相关关系

Tab. 2 The correlations between Elevation $D(i)$, Slope $S(i)$ or Reclamation Ratio $C(i)$ of the traditional cultivated region in China

	$D(i) - C(i)$	$S(i) - C(i)$	$D(i) - S(i)$
相关系数	-0.488	-0.694	0.736
Sig. (2-tailed)	0.000	0.000	0.000
固定 $S(i)/D(i)$ 后的偏相关系数	0.046	-0.567	
Sig. (2-tailed)	0.180	0.000	

① 数据来源：全球 1 km DEM，由国家科学数据共享工程 - 地球系统科学数据共享网提供，www.geodata.cn。

② 根据 1:25 万 DEM 数据计算而得。

③ 由 1980s 全国 1km 土地利用数据库计算而得，该数据是由 1980s1:10 万土地利用数据计算所得，由国家科学数据共享工程 - 地球系统科学数据共享网提供，www.geodata.cn。

长江中下游平原以及四川盆地海拔较低且坡度平缓；南方丘陵与华北山地海拔相对较高，地面坡度也相对较大；云贵高原是研究区内海拔最高的地区（不包括海拔大于 3500 m 的川西地区），同时地面坡度也最大。因此，为了更准确地反映海拔、坡度与垦殖率之间的关系，需要进一步进行偏相关分析，结果表明：在控制 $D(i)$ 情况下， $S(i)$ 与 $C(i)$ 仍然存在着显著的负相关（偏相关系数为 -0.567），但在控制 $S(i)$ 的情况下， $D(i)$ 与 $C(i)$ 的偏相关系数仅为 0.046（双尾检验概率 $P = 0.180$ ），可以认为海拔高度对耕地分布的影响并不显著。因此，在研究区 $D(i) < 3500$ m 的区域中，可以选择平均坡度 $S(i)$ 代表地形对耕地分布的影响。作为历史悠久的农业垦殖区，在现代巨大的人口压力下，研究区内宜耕土地已几乎开垦殆尽，因此现代耕地分布可以近似看作农业开发前的潜在垦殖区，通过拟合现代网格垦殖率与坡度之间的关系，可以很好地定量化揭示坡度对耕地分布的影响。

我们仍采用上述 $60\text{km} \times 60\text{km}$ 平均坡度 $S(i)$ 与 1980s 网格垦殖率 $C(i)$ 进行分析计算，为了方便比较，首先对 $C(i)$ 进行标准化处理：

$$\alpha(i) = C(i)/\text{Max}(C(i)) \quad (1)$$

式中： $\alpha(i)$ 为网格 i 的宜垦程度， $C(i)$ 与 $\text{Max}(C(i))$ 分别为网格垦殖率及其最大值（对整个研究区而言）。 $\alpha(i)$ 值越大表示该网格越适宜开垦，其分布区间为 $[0, 1]$ 。对网格宜垦程度 $\alpha(i)$ 与平均坡度 $S(i)$ 之间的关系进行曲线拟合，发现 $\alpha(i)$ 随 $S(i)$ 呈明显的逆函数变化趋势（图 2），通过曲线回归得到拟合函数（为了区分实际的 $\alpha(i)$ ，用 $\hat{\alpha}(i)$ 表示拟合值）：

$$\hat{\alpha}(i) = 0.047 + 1.504/S(i) \quad (2)$$

式中： $\hat{\alpha}(i)$ 为网格宜垦程度的理论值，其大小与地面平均坡度 $S(i)$ 值呈反比，即坡度越小的地方越适宜开垦，对农垦开展的吸引力也就越大，从历史农业开发的进程来看，这样的地区往往也最先被开发。由于 $\hat{\alpha}(i)$ 的大小反映其对农垦开展的吸引力大小，并且只受地形因素（坡度）控制，故而将其称为农垦地形引力。

2.2.2 人口分布 假设某一时段，人口数据集的空间分辨率为 h 级（如统计到府一级），而耕地数据集的分辨率仅为 k 级（如统计到省）， h 、 k 指行政区划级别；由于在农业社会中，人口分布状况是影响历史时期耕地分布的主导社会因素，且一般情况下是人口密度越高的地方，土地被开垦的程度越大，因此，当人口数据集的空间分辨率高于耕地数据的分辨率时，人口空间密度的分布就可以在在一定程度上反映出不同地区土地垦殖的空间差异。但由于当土地垦殖达到一定强度以后，人口密度的升高并不会带来耕地的显著增加，因此通过人口密度指示土地垦殖程度具有一定的局限性。例如清代华北平原的人口密度要低于地面平坦程度相当的长江中下游平原，但前者的开垦程度并不低于后者。那么，要消除这一局限性的影响，就必须将比较尺度从整个研究区降到 k 级行政区上。因为，只有在同一个 k 级行政区内，其下一级的人口密集程度与耕地的垦殖程度才具有更显著的正相关关系。设 t_v 年 h_m 地区的人口密度为 $P(h_m, t_v)$ ，为便于计算，我们先对人口密度 $P(h_m, t_v)$ 进行标准化处理：

$$\beta(h_m, k_n, t_v) = P(h_m, k_n, t_v)/\text{Max}(P(h_m, k_n, t_v)) \quad (3)$$

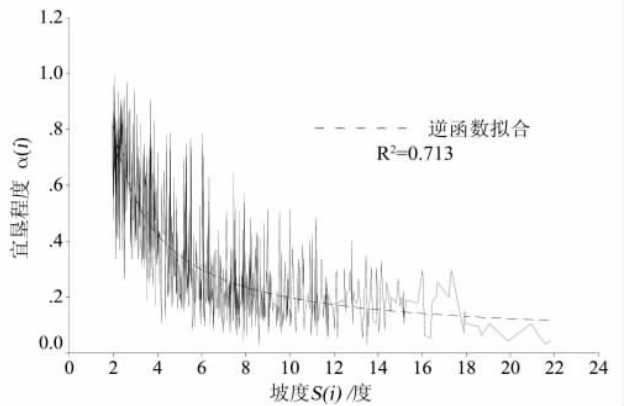


图 2 中国传统农区宜垦程度 $\alpha(i)$ 随坡度 $S(i)$ 变化趋势的逆函数拟合

Fig.1 The inverse regression of Cultivability ($\alpha(i)$) and Slope ($S(i)$) of the traditional cultivated region in China

式中: $\beta(h_m, k_n, t_v)$ 表示地区在其所属 k_n 行政区人口的相对聚集度, $P(h_m, k_n, t_v)$ 与 $Max(P(h_m, k_n, t_v))$ 分别代表 t_v 年 h_m 地区 (隶属行政区) 的人口密度及其最大值 (针对所属的同一个行政区而言)。标准化后得到人口相对聚集度 $\beta(h_m, k_n, t_v)$ 的变化区间为 $[0, 1]$ 。必须指出的是, 这里所指的人口聚集度是一个相对的概念, 只有对同一个 k 级行政区, β 值的大小才指示人口聚集程度的高低, 即人口聚集度 $\beta(h_m, k_n, t_v)$ 越大的 h_m 地区土地被开垦的程度越高。由于 β 值的大小可以用以指示特定区域内人口聚集程度对土地开垦的驱动力的大小, 因此将 $\beta(h_m, k_n, t_v)$ 称为农垦人口引力。

尽管中国历史文献中关于人口的记载多不胜数, 但古代毕竟缺乏人口普查数据与政府统计报告这类现代或接近现代的人口数据, 册载的人口数据往往范围大小不一且质量良莠不齐, 因此要想获得比较完整且精确的历史时期人口密度 (或人口数量) 对历史时期的很多年份而言是比较困难的。但从公式 (3) 中可以看出, 历史时期耕地数据网格化模型对人口资料的要求也并非绝对的人口密度, 而是特定区域内的相对人口聚集程度, 因此对于人口密度 $P(h_m, t_v)$ 缺失的年份, 可以通过以下方法进行估计:

(1) 可获得相近时期人口数据的年份: 假设 t_2 年为研究时间点 (人口密度 $P(h_m, t_2)$ 未知), t_1 年为与 t_2 年相近的年份 (人口密度 $P(h_m, t_1)$ 已知)。一般情况下, 除非受到战争、流行疾病或是大的自然灾害等影响, 短时间内行政区的人口分布格局是比较稳定的, 因此可以通过 t_1 年的人口数据计算 t_2 年的农垦人口引力:

$$\beta(h_m, k_n, t_2) \approx \beta(h_m, k_n, t_1) = P(h_m, k_n, t_1) / Max(P(h_m, k_n, t_1)) \quad (4)$$

(2) 无相近时期人口数据的年份: 对于人口资料相对缺乏的年份, 可以通过历史文献中记载的大致人口分布状况或社会经济状况直接估计农垦人口引力 $\beta(h_m, k_n, t_v)$ 。

2.3 建模

在量化影响耕地分布的自然因子 (农垦地形引力) 与社会因子 (农垦人口引力) 后, 用垦殖倾向 $\delta(i, h_m, k_n, t_v)$ 代表二者的综合影响, 其计算公式为:

$$\delta(i, h_m, k_n, t_v) = \alpha^{\wedge}(i) \cdot \beta(h_m, k_n, t_v) \quad (5)$$

式中: $\delta(i, h_m, k_n, t_v)$ 表示 t_v 网格 i 的垦殖倾向, $\alpha^{\wedge}(i)$ 与 $\beta(h_m, k_n, t_v)$ 分别代表相应的农垦地形引力与农垦人口引力。由公式 (2) ~ (3) 与公式 (5) 可以看出, $\delta(i, h_m, k_n, t_v)$ 的大小与坡度成反比而与人口聚集程度成正比, 也就是说, 地势平坦且人口密集的地方土地被垦殖的倾向相对较大。由于不具有时间变量, 对于同一个网格 i 而言, $\alpha^{\wedge}(i)$ 是一个常量, 因而 δ 值随年份的变化受网格所在地区的 $\beta(h_m, k_n, t_v)$ 控制。

要注意的是, 由公式 (5) 所得到的土地垦殖倾向并未考虑到水域的影响, 虽然对于 $60 \text{ km} \times 60 \text{ km}$ 的网格而言, 大部分河流与湖泊所占的比重较小, 其影响可以忽略不计, 但一些大型湖泊 (如洞庭湖、太湖等) 以及海洋 (指位于海岸线的网格) 在网格中所占得比重较大, 影响不容忽视。此外, 位于国界上的网格在进行计算时其实际陆地面积也并未达到 3600 km^2 。因此需要对公式 (5) 进行修正: 假设网格 i 的实际陆地面积比重为 $\varepsilon(i, t_v)$, 则修正后的垦殖倾向 $\delta'(i, h_m, k_n, t_v)$ 为:

$$\delta'(i, h_m, k_n, t_v) = \delta(i, h_m, k_n, t_v) \cdot \varepsilon(i, t_v) = \alpha^{\wedge}(i) \cdot \beta(h_m, k_n, t_v) \cdot \varepsilon(i, t_v) \quad (6)$$

由于 k_n 行政区的耕地面积总数 $A(k_n, t_v)$ 是已知的, 因此要得到网格 i 的耕地面积 $X(i, t_v)$, 关键在于计算网格 i 的耕地面积在其所属 k_n 行政区总耕地面积所占的比重 $\theta(i, k_n, t_v)$ 。 $\theta(i, k_n, t_v)$ 受土地的垦殖倾向 $\delta'(i, h_m, k_n, t_v)$ 影响, 对于同一个 k_n 行政区而言, $\delta'(i, h_m, k_n, t_v)$ 值越大的网格其耕地面积比重越大, 因此通过直接求取比例系数方法求得 $\theta(i, k_n, t_v)$:

$$\theta(i, k_n, t_v) = \delta'(i, h_m, k_n, t_v) / \sum_i \delta'(i, h_m, k_n, t_v) \quad (7)$$

根据 $\theta(i, k_n, t_v)$, 得到 k_n 区内网格 i 的耕地面积 $X(i, k_n, t_v)$ 与垦殖率 $D(i, t_v)$ 分别为:

$$X(i, k_n, t_v) = \theta(i, k_n, t_v) \cdot A(k_n, t_v) \quad (8)$$

$$D(i, t_v) = X(i, t_v) / \text{area}(i) \quad (9)$$

式中: $\text{area}(i)$ 为网格的面积, 大小为 3600 km^2 。

理论上网格的垦殖率 $D(i, t_v)$ 分布区间应为 $[0, 1]$, 但实际上土地垦殖并不会达到百分百, 从现代 (1980s) 耕地的分布状况来看, 垦殖率 $D(i)$ 的最大值为 0.95, 而 99.5% 以上的网格垦殖率均低于 0.90, 由于现代耕地分布可近似于垦殖饱和下的耕地分布状况, 因此可以认为历史时期的网格垦殖率 $D(i, t_v)$ 不会超过 0.90。

当土地的开垦到达一定强度以后, 人口聚集程度的升高并不会带来耕地的显著增加。而根据上述方法, 在人口极为密集的地方, 计算结果有时可能会出现 $D(i)$ 值过大、甚至大于 1 的情况, 因此需要对结果进一步处理, 具体方法为:

① 将 0.90 作为网格垦殖率 $D(i, t_v)$ 的上限。

② 根据公式 (6) ~ (9), 垦殖率 $D(i, t_v)$ 大小受垦殖倾向 $\delta'(i, h_m, k_n, t_v)$ 控制, 因此对于 $D(i, t_v)$ 值超出其上限值 (0.90) 的网格, 可以通过适当减小 $\delta'(i, h_m, k_n, t_v)$ 值以对结果进行修正。过程如下: 提取 k_n 区域内垦殖率 $D(i, t_v) < 0.9$ 的所有网格中最大垦殖倾向 $\delta'(i, h_m, k_n, t_v)$ 值, 将其取代 $D(i, t_v) \geq 0.9$ 的网格 $\delta'(i, h_m, k_n, t_v)$ 值, 根据公式 (6) ~ (9) 重新计算 $X(i, k_n, t_v)$ 与 $D(i, t_v)$, 依次循环, 直至所有 $D(i, t_v)$ 值均小于 0.9。

3 实例应用与检验

本文选择资料较为完整的 1820 年 (清嘉庆二十五年) 作为应用实例, 利用上述历史耕地数据网格化重建模型绘制研究区 1820 年分辨率为 $60 \text{ km} \times 60 \text{ km}$ 耕地分布图, 并将模型的计算结果与册载府级田亩数据进行比较, 通过统计分析, 检验评价模型的有效性。

3.1 1820 年耕地分布状况重建

本文引用的 1820 年耕地数据与人口数据分别来自葛全胜等所重建的过去 300 年中国部分省区耕地数据集^[3]与曹树基整理校正的 1776 年至 1953 年中国 18 个省的分府人口数据集 (表 1)^[11], 用以检验模型的册载府级田亩数据来自《嘉庆重修一统志》(转引自文献^[12]), 其中府级的人口数据分辨率要高于省级的人口分辨率, 满足历史耕地数据网格化模型的应用条件。由于耕地数据与人口数据均由《嘉庆重修一统志》的相应数据统计校正而得, 因此这三个数据集具有较强的统一性与可比性, 避免了由于不同数据源统计口径等因素造成过大误差。

根据前文中所设计的历史耕地数据网格化方法, 计算了研究区 1820 年 $60 \text{ km} \times 60 \text{ km}$ 耕地面积 $X(i, 1820)$ 与网格垦殖率 $D(i, 1820)$ (图 3)。在模型应用过程中, 我们对有关问题进行如下处理:

(1) 本研究中, h_m 为府级行政区划, k_n 为省级行政区划;

(2) 计算过程中, 将 1820 年耕地数据 $A(k_n, 1820)$ 与人口数据 $\beta(h_m, k_n, 1820)$ 分别加载至 1820 年行政区划图上, 并与 $60 \text{ km} \times 60 \text{ km}$ 网格图进行叠加处理。由于行政边界与网格边界并不重合, 因此对政区交界的处理遵循以下原则: ① 为多个省区共有的网格将其划入最大面积所属的省区; ② 对同一省内为多个府所共有的网格, $\beta(h_m, k_n, 1820)$ 值按面积加权法取其均值;

(3) 网格实际陆地面积比重 $\varepsilon(i, t_v)$ 。位于研究区边界 (包括国界与海岸线) 上的网格

表 3 中国传统农区 1820 年各省耕地面积 [3]

Tab. 3 Cropland area in the traditional cultivated regions of China in 1820

行省	1820 年耕地面积 (ha)	行省	1820 年耕地面积 (ha)	行省	1820 年耕地面积 (ha)
直隶省	5966908	江苏省	5380440	福建省	1354320
山西省	3632080	浙江省	2003120	江西省	2702640
陕西省	4536560	湖南省	3331010	四川省	5500820
河南省	9714830	湖北省	3825910	贵州省	1885960
山东省	6652990	广西省	3267910	云南省	2469200
安徽省	4669210	广东省	2569300		

注：受研究区范围限制，福建省不包括台湾府；直隶省不包括承德府与口北三厅；山西省不包括长城以北地区（大致包括归绥六厅、朔平府的宁远厅及大同府的丰镇厅）；因此将原文献中直隶省与山西省的耕地数据分别减去相应地区的耕地面积（直隶省与山西省北部 1820 年耕地数据根据田砚宇硕士论文《热察绥地区过去 300 年土地利用/覆被变化》进行估算而得（北京：中国科学院地理科学与资源研究所，2004）。由于缺乏台湾府较为准确的资料，故而不对福建省耕地数据另行计算。

$\varepsilon(i, t_i)$ 值在 GIS 软件直接进行结算；清中期至今的陆地水域面积的变化，河流可以忽略不计，直接采用现代河流面积进行计算^④；对一些面积有较大变化的大型湖泊（如洞庭湖等），文中采用清代的湖泊分布图^⑤与研究区网格图进行叠加计算。

3.2 模型检验

根据上文得到的垦殖率 $D(i)$ 计算 1820 年分府耕地面积的模拟值，由于文献^[3]中的清代的数据经过校正处理后，无法直接与《嘉庆一统志》中数据进行比较，但原文献的校正过程是为了连接清代的数据与民国及现代的耕地数据，而对各省的耕地总和乘以相应的校正系数，虽然各省的校正系数不同，但并不涉及府一级数据的变化。也就是说，同一个省的耕地总面积在两个数据集中虽然有差异，但各府的耕地面积占其所在省域的耕地总面积的比重应该是相同的，因此通过检验这一假设是否成立就可以评价模型的有效性。此外，虽然历史文献中田亩数据受到广泛的置疑，“明清的土地数据实为纳税单位”已经成为大部分学者的共识^[13, 14]，但田亩数据的失真程度在一个省内比较相似，因此能比较真实地反映彼时的耕地分布状况，从这个层面上来看，利用册载数据进行检验仍是可行的。

用 $F(h_m)$ 与 $F(h_m)$ 分别代表 府耕地占其所在省耕地总面积百分比的模型拟合值与册载值，要判断二者之间是否存在显著差异，可以通过连续型资料假设检验的方法判断其差别是否具有统计学意义。由于 $F(h_m)$ 与 $F(h_m)$ 之间是一一对应的关系，因此选用配对样本 T 检验方法进行检验。

配对样本 T 检验要求 $F(h_m)$ 与 $F(h_m)$ 之间的差值 $E(h_m)$ (即模型的绝对误差) 服从正态

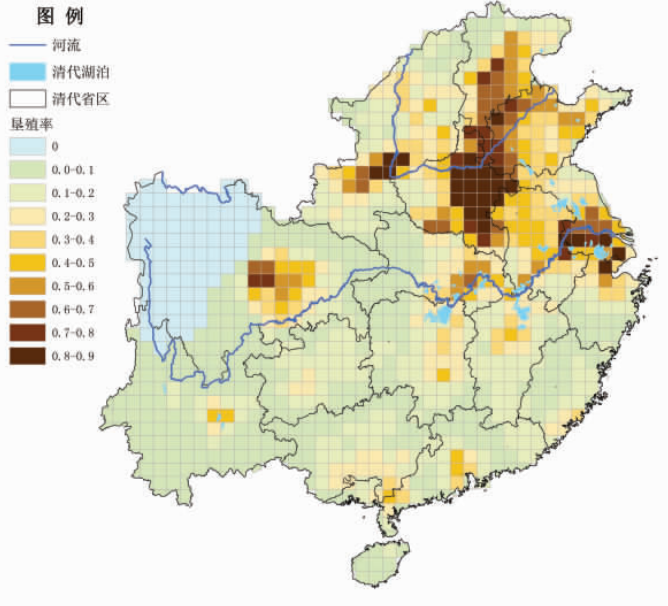


图 3 中国大陆传统农区 1820 年 60 km×60 km 网格垦殖率
Fig. 3 Reclamation ratio of the traditional cultivated region of China mainland in 1820 (at a resolution of 60 km×60 km)

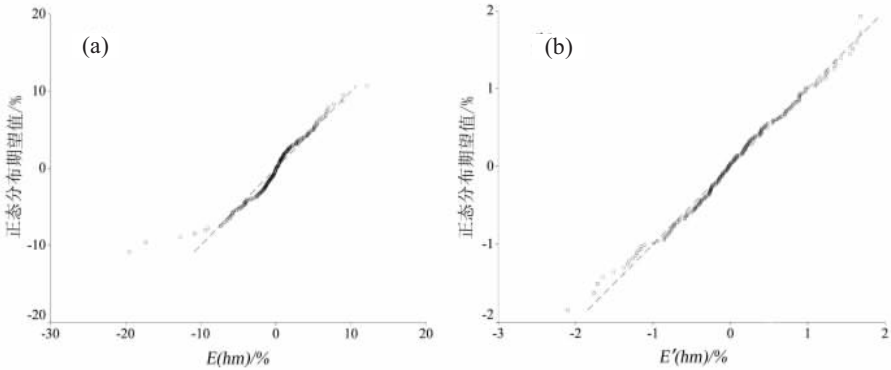


图 4 正态变换前后绝对误差 $E(h_m)$ (a) 与 $E'(h_m)$ (b) 的 Q-Q 概率图

Fig. 4 Q-Q probability plot of the absolute error $E(h_m)$ (a) and $E\phi(h_m)$ (b) before / after transformation to normality

分布, 从 $E(h_m)$ 的 Q-Q 概率图上看 (图 4a), $E(h_m)$ 的分布并不完全近似于一条斜线, 有个别离群值的存在; 进一步对其进行正态分布 K-S 检验, 结果表明 $E(h_m)$ 并不服从正态分布 (K-S 的 Z 统计量为 1.664, 其相伴概率为 0.008, 小于显著水平 0.05); 因此需要首先对数据进行正态变换。对 $\hat{F}(h_m)$ 与 $F(h_m)$ 分别进行平方根变换后得到 $\hat{F}'(h_m)$ 与 $F'(h_m)$, 重新计算二者差值 $E'(h_m)$; 从 $E'(h_m)$ 的 Q-Q 概率图 (图 4b) 以及正态分布 K-S 检验结果 (K-S 的 Z 统计量为 0.735, 其相伴概率为 0.652, 大于显著水平 0.05) 可以看出, 经由平方根变换后, $E'(h_m)$ 服从正态分布。

在确认 $E'(h_m)$ 的正态性后, 对 $\hat{F}(h_m)$ 与 $F(h_m)$ 进行配对样本 T 检验 (表 4), 检验的零假设为: $\hat{F}(h_m)$ 与 $F'(h_m)$ 的总体分布不存在显著差异; t 统计量为 -1.008, 其相伴概率 (双尾) 为 0.314, 大于显著水平 0.05, 因此不能拒绝零假设, 即可以认为 $\hat{F}(h_m)$ 与 $F(h_m)$ 的差异不具有统计意义。

通过统计检验, 可以得到如下结论: 在 0.05 的显著水平下, 1820 年研究区 h_m 府耕地占其所在省耕地总面积百分比的模型拟合值 $\hat{F}(h_m)$ 与册载值 $F(h_m)$ 并不存在显著差异。

上述配对 T 检验实际是对模型拟合的绝对误差均值为 0 的检验, 由于 $F(h_m)$ 的变异程度较大 (变异系数为 0.87), 为了更好地分析模型的误差来源, 还要进行相对误差 $E^*(h_m)$ 的分析:

$$E^*(h_m) = (\hat{F}(h_m) - F(h_m))/F(h_m)$$

从图 5 可以看出, 较大的绝对误差更多的是产生在 $F(h_m)$ 较小的地区, 也就是说对耕地面积相对较小的地区更易产生偏差。从地域上看主要分布在西部的陕西、山西、四川、云南及广西, 这些地区耕地分布的影响因素要更为复杂一些, 例如陕北气候干旱, 宜垦性较差, 广西、云南地区少数民族聚居, 无论是对土地利用方式还是对历史文献中田亩数据的失真程度都

表 4 正态变换后 $F'(h_m)$ 与 $\hat{F}(h_m)$ 的配对 T 检验

Tab. 4 Paired sample T test between $F'(h_m)$ and $\hat{F}(h_m)$ after transformation to normality

配对差				t	自由度	双尾检验概率
均值 (%)	标准差	标准误	95%置信区间			
			下限 上限			
-0.043	0.673	-0.043	-0.013 0.041	-1.008	246	0.314

④ 由 1980s 全国 1km 土地利用数据库计算而得, 该数据是由 1980s:10 万土地利用数据计算所得。由国家科学数据共享工程 - 地球系统科学数据共享网提供, www.geodata.cn.

⑤ 根据 《中国历史地图集·清时期》 中所绘制的主要湖泊分布。

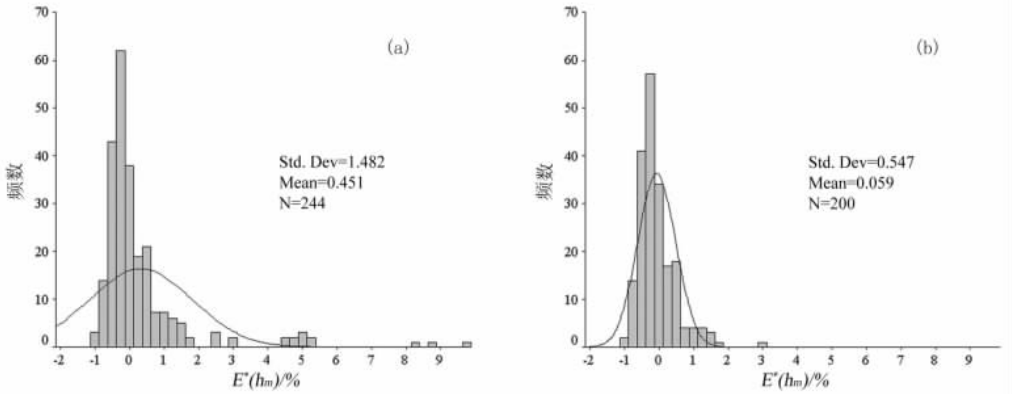


图 5 相对误差分布图 (a) 所有府级行政区 * (b) $F(h_m) > 2\%$ 的府级行政区
(*: 为显示方便, 不包括大于 10 的三个值)

Fig. 5 Histogram of relative error (a) All prefectures*; (b) The prefectures of $F(h_m) > 2\%$

有较大的影响, 以相对误差较大的云南广南府为例 ($E^*(h_m)$ 达到 61), 《嘉庆一统志》中记载该府的耕地仅 6142 亩, 占全省的 0.06%, 人口记载则缺失, 仅标注为“向因蛮民杂处, 未经编丁”, 由此可以田亩数据偏低的幅度也应较大。对 $F(h_m) > 2\%$ 的地区而言, $E^*(h_m)$ 的分布区间与均值分别缩小为 $[-0.83, 3.22]$ 及 0.059。对于历史时期耕地分布估计而言, 可接受这样的误差。

4 结论与讨论

(1) 历史耕地数据网格化模型的设计旨在将历史耕地面积的统计数据转化为具有空间属性的数据集, 对历史时期土地覆被变化研究以及气候变化等生态系统模型研究都有重要意义, 本文针对中国传统农区 (华北、江南与西南) 设计的历史耕地数据网格化方法, 在精度上要较已有的两个全球土地覆被数据集有很大的提高。

(2) 影响历史时期我国传统农区耕地分布因素虽然纷繁复杂, 但在不同的时空尺度上, 这些因素的影响效应差异十分显著, 因此, 在经过仔细遴选后, 模型选择地面坡度与人口分布状况分别代表影响该区历史时期耕地分布的自然因素与社会因素, 既能最大程度地简化模型, 又能保证在 $60 \text{ km} \times 60 \text{ km}$ 空间分辨率上较充分地模拟影响因子作用。

(3) 中国传统农区历史耕地数据网格化模型表明: 该区网格 i 的垦殖率与地面坡度呈反比而与人口的相对聚集度成正比, 即在一定区域内地势平缓且人口聚集度大的网格垦殖率相对较大。

(4) 选择数据匹配度高的 1820 年作为研究实例对模型进行检验, 统计检验与绝对误差分析的结果表明模型可以有效地将以行政区域为统计单元的中国历史耕地数据量化为具有统一且更高空间分辨率的网格化数据集。

参考文献 (References)

- [1] Foley J A, Ruth D, Gregory P A et al. Global consequences of land use. *Science*, 2005, 309: 570-574.
- [2] IGBP Report No.53/IHDP Report No.19. GLP (2005) Science Plan and Implementation Strategy. Stockholm: IGBP Secretariat, 2005. 1-64.
- [3] Ge Q S, Dai J H, He F N. Spatiotemporal dynamics of reclamation and cultivation and its driving factors in parts of China during the last three centuries. *Progress in Natural Science*, 2003, 13(8): 825 -832. [葛全胜, 戴君虎, 何凡能. 过去 300 年中国部分省区耕地资源数量变化及驱动因素分析. *自然科学进展*, 2003, 13(8): 825 -832.]
- [4] Ge Quansheng, Zhao Mingcha, Zheng Jingyun. Land use change of China during the 20th century. *Acta Geographica Sinica*, 2000, 55(6): 698-706. [葛全胜, 赵名茶, 郑景云. 20 世纪中国土地利用变化研究. *地理学报*, 2000, 55(6):

698-706.]

- [5] He Fanneng, Ge Quansheng, Dai Junhu. Quantitative analysis on forest dynamics of China in recent 300 years. *Acta Geographica Sinica*, 2007, 62(1): 30-40. [何凡能, 葛全胜, 戴君虎. 近三百年来中国森林的变迁. *地理学报*, 2007, 62(1): 30-40.]
- [6] He Fanneng, Ge Quansheng, Zheng Jingyun. Reckoning the areas of urban land use and their comparison in the Qing dynasty in China. *Acta Geographica Sinica*, 2002, 57(6): 709-716. [何凡能, 葛全胜, 郑景云. 中国清代城镇用地面积估算及其比较. *地理学报*, 2002, 57(6): 709-716.]
- [7] Ramankutty N, Foley J A. Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochemical Cycles*, 1999, 13(4): 997-1027.
- [8] Goldewijk K K. Estimating global land use change over the past 300 years: The HYDE database. *Global Biogeochemical Cycles*, 2001, 15(2): 417-433.
- [9] Wei W C, Gong W, Wei H L. A regional model simulation of the 1991 severe precipitation event over the Yangtze-Huai river valley. Part I: Precipitation and circulation statistics. *Journal of Climate*, 2000, 13(1): 74-92.
- [10] Sun Han, Shi Yulin. *Agricultural Land Use in China*. Nanjing: Phoenix Science Press, 2003. 3-84. [孙颌, 石玉林. 中国农业土地利用. 南京: 江苏科学技术出版社, 2003. 3-84.]
- [11] Cao Shuji. *Population History of China (Vol. 5, Qing Dynasty Period)*. Shanghai: Fudan University Press, 2001. 690-720. [曹树基. 中国人口史 (第五卷, 清时期). 上海: 复旦大学出版社, 2001. 690-720.]
- [12] Liang Fangzhong. *Statistics of Registered Permanent Residence, Plowlands and Land Taxes in Past Dynasties in China*. Shanghai: Shanghai People's Press, 1980. 401-413. [梁方仲. 中国历代户口、田地、田赋统计. 上海: 上海人民出版社, 1980. 401-413.]
- [13] He Bingdi. *Textual Research and Estimation of Land Data in China*. Beijing: China Social Science Press, 1988. 61-114. [何炳棣. 中国古今土地数字的考释和评价. 北京: 中国社会科学出版社, 1988. 61-114.]
- [14] Zhou Rong. A general inspection and re-appraise on area under cultivation in the early period of the Qing. *The Journal of Chinese Social and Economic History*, 2001, (3): 39-49. [周荣. 对清前期耕地面积的综合考察和重新估价. *中国社会经济史研究*, 2001, (3): 39-49.]

The Approach for Gridding Data Derived from Historical Cropland Records of the Traditional Cultivated Region in China

LIN Shanshan^{1,2}, ZHENG Jingyun¹, HE Fanneng¹

(1. *Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;*

2. *Graduate University of the Chinese Academy of Sciences, Beijing 100039, China)*

Abstract: Recent studies have demonstrated the importance of LUCC change with climate and ecosystem simulation, but the result could only be determined precisely if a high-resolution underlying land cover map is used. While the efforts based on satellites have provided a good baseline for present land cover, what the next advancement in the research about LUCC change required is the development of reconstruction of historical LUCC change, especially spatially-explicit historical dataset. Being different from other similar studies, this study is based on the analysis of historical land use patterns in the traditional cultivated region of China. Taking no account of the less important factors, slope and population patterns were selected as the major drivers of reclamation in ancient China, and were used to design the Chinese Historical Cropland Data Gridding Model (at a 60 km×60 km resolution), which is an empirical model for allocating the historical cropland inventory data spatially to grid cells in each political unit. Then we use this model to reconstruct the cropland distribution of the study area in 1820, and test the result by prefectural cropland data of 1820, which is from the historical documents. The statistical analyzing result shows that the model can simulate the patterns of the historical cropland's distribution in the traditional cultivated region efficiently.

Key words: approach; gridding data; Chinese historical cropland records