中国科学: 地球科学 2010 年 第 40 卷 第 8 期: 1048 ~ 1059
www.scichina.com earth.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS

论 文

全球土地利用数据集精度的区域评估——以中国东北地区为例

李蓓蓓[①], 方修琦[①②*], 叶瑜[①], 张学珍[②]

① 北京师范大学地理学与遥感科学学院, 北京 100875;
② 中国科学院地理科学与资源研究所, 北京 100101
* 联系人, E-mail: xfang@bnu.edu.cn

收稿日期: 2010-02-07; 接受日期: 2010-07-13
国家自然科学基金(批准号: 40571165, 40901099)、北京师范大学自主科研基金资助项目(编号: 2009SAP-2)和国家科技支撑计划(编号: 2007BAC03A11)资助

摘要 历史土地利用数据集是解释过去人类活动及其影响、定量评估土地利用变化对气候变化和碳循环影响、检验模型模拟结果的基础. SAGE("全球土地利用数据集")和 HYDE("全球历史环境数据集")是国际上已被广泛应用的代表性的全球土地利用数据集, 但其数据的准确性并未得到充分地评估. CNEC 数据(Cropland data of Northeast China)重建的是东北三省过去 300 年耕地覆盖变化数据, 该数据所利用的数据源为历史文献和调查、统计资料, 并经过订正和归一化处理, 能够较好地反映耕地变化的实际情况. 以 CNEC 数据为参照, 分别从东北地区、分省和网格 3 个空间尺度对 SAGE 和 HYDE 数据集中东北地区的耕地数据精度进行了评估. 结果表明, SAGE 和 HYDE 数据集中东北地区的耕地重建结果无论总量还是空间分布都存在显著的误差, 其反映的耕地面积增长与空间变化过程也与研究区历史开垦过程相左. 其中, SAGE 数据集对东北历史耕地数据的估计普遍偏高, 并且没有反映出东北地区耕地面积阶段性增长的特点; HYDE 数据集的耕地总量虽然具有一定可比性, 但是年代际变化和各时间断面空间分布与历史真实情形不符. 全球数据集在东北地区的重建误差可能来源于两方面: 一是依据现代耕地分布格局回推历史耕地分布的方法不能反映现实开垦进程, 二是遥感资料提取的现代耕地数据存在误差.

关键词
历史耕地
数据集
精度评估
中国东北

土地利用/覆盖变化(LUCC)是人类活动对地球系统的重要影响因素之一, 具有全球范围的环境效应[1,2]. 土地利用/覆盖变化能够直接或间接影响陆地生态系统的碳循环过程和地-气之间能量交换, 其对碳收支和气候变化的影响备受关注. 土地利用/覆盖类型决定了陆地生态系统的碳储量, 土地利用/覆盖由一种类型转变为另一种类型往往伴随着大量的碳交换[3]. 土地利用/覆盖变化亦会改变气候系统的下垫面属性, 从而影响区域乃至全球的气候, 其中人类土地利用活动(主要是农业活动)导致的地表反照率变化是全球气候变化的重要辐射强迫因子之一[4~6].

过去长时间序列的土地利用/覆盖变化过程及其环境效应研究, 是解答人类活动对地球系统影响历史过程的一把钥匙. 研究表明, 1850~2000 年土地利

引用格式: Li B B, Fang X Q, Ye Y, et al. Accuracy assessment of global historical cropland datasets based on regional reconstructed historical data—A case study in Northeast China. Sci China Earth Sci, 2010, doi: 10.1007/s11430-010-4053-5

用活动使大气净增加了 156 Pg碳[7], 是仅次于化石燃料燃烧的第二大碳排放源, 工业革命以来土地利用变化所引起的CO_2排放量约占人类活动总排放量的1/4[6]. HadAM3 AGCM模型模拟结果表明, 1700 年以来土地利用变化造成全球降温, 冬季和春季平均温度大约降低 1~2℃, 该模型考虑了土地覆盖类型变化所导致的地表反照率和蒸散量的变化, 其中地表反照率是全球降温的主要因素[8].

无论是研究历史土地利用变化与碳循环关系, 还是研究历史土地利用变化的气候效应, 共同的基础都是准确而完备的历史土地利用变化数据[9,10]. 在IGBP的BIOME300, LUCC, GCTE (Global Change and Terrestrial Ecosystem), GLP 和 iLEAPS (Integrated Land Ecosystem Atmosphere Processes Study)等研究计划共同推动下, 重建过去环境, 特别是重建过去300 年全球土地覆盖的工作取得重大进展[11], 已经建立了许多全球及区域尺度上的历史土地利用/覆盖数据集[12]. 其中, 2 个代表性的全球数据集分别是美国威斯康星大学全球环境和可持续发展中心(SAGE)的Ramankutty和Foley建立的“全球土地利用数据集”(以下称为SAGE数据集)[13]和荷兰环境评价局(Netherlands Environmental Assessment Agency)建立的“全球历史环境数据集” (Historical Database of the Global Environment, 以下称为HYDE数据集)[14]. 它们作为定量评估土地利用环境效应的重要数据, 已被广为采用[6,9,10]. 在土地利用变化与碳循环研究方面, Tian等[15]将SAGE数据集的历史耕地数据作为TEM模型(Terrestrial Ecosystem Model 4.2)的历史环境变量, 估算了1860~1990年开垦和弃耕对亚洲季风区碳收支的影响. Brovkin等[16]使用SAGE和HYDE2.0数据模拟了过去 150 年由于土地利用变化导致的大气CO_2增加, 以及CO_2浓度增加产生的气候效应. Oost等[17]从HYDE3.0数据集提取1990年垦殖率和牧场比例网格数据作为水蚀-耕作侵蚀模型(water and tillage erosion models)的变量, 估计了全球农业土壤侵蚀对碳循环的影响. 在土地利用变化的气候效应研究方面, Matthews等[18]根据SAGE数据集 1700~1900 年全球耕地数据计算了历史地表反照率变化, 利用Uvic Earth System Climate Model模拟了农田面积扩展对全球气候造成的影响. 石正国等[19]以SAGE数据集1700~1900 年耕地变化近似森林砍伐过程, 使用McGill Paleoclimate Model (MPM-2)模拟了全球毁林开荒的气候效应. 国内也有学者利用全球数据集研究中国土地利用变化的区域气候效应, 陈星等[20]使用HYDE2.0 中的潜在植被和耕地、草地数据研究了亚欧大陆植被变化引起的气候响应. 李巧萍等[21]也利用HYDE2.0 数据, 基于国家气候中心改进的区域气候模式(RegCM-NCC)研究了中国近代土地利用变化的气候效应.

土地利用数据的准确性和精度直接影响到土地利用环境效应模拟的结果. 如果数据本身的时间过程和空间分布存在显著误差, 必然会影响使用这些数据的模式输出结果, 增加土地利用环境效应定量评估的不确定性. SAGE 和 HYDE 数据集具有明确的地理属性, 可以作为模型输入的土地利用变化基础变量, 目前这 2 个数据集已经在全球或区域模型中有所应用, 但是为了模式研究而快速建立的全球数据集, 其数据精度还有待评估. 一般认为, SAGE 和HYDE 数据集能够反映全球大势, SAGE 数据集明确声明其数据不能作为局地尺度研究的依据. 但 2 个全球数据集在区域尺度上的数据精度并没有引起重视, 尚未有研究对其区域尺度的准确性进行论证, 这必然会影响使用这些数据所得到的模式输出结果.

1700 年来, 东北三省经历了大规模的移民开垦, 是中国土地利用/覆盖变化最显著的地区之一. 叶瑜等[22]从历史文献、政府档案、日俄调查资料、现代统计和调查数据中提取土地利用信息, 经过历史数据订正和多源耕地数据关系模型校核, 重建了归一化的东北地区过去 300 年耕地面积和垦殖率数据(Cropland data of Northeast China, 简称CNEC数据). 该套区域数据能够较好地代表历史真实土地开垦进程. 本文利用CNEC数据对 2 个代表性全球土地利用数据集中东北地区耕地数据进行评估. 从区域角度对全球数据集的数据精度进行检测. 并且通过对全球数据集与CNEC数据, 从资料来源、重建方法和土地覆盖变化时空过程等方面的比较, 尝试分析全球数据集在假设条件、基础数据及估算方法等方面的可能误差来源.

1 数据和方法

1.1 数据来源与重建方法

国际土地利用数据选取的 2 个代表性全球数据集分别是: SAGE数据集[13], 来自美国威斯康星大学全球

环境和可持续发展中心网站http://www.sage.wisc.edu/iamdata; 以及HYDE数据集[14](本文使用的是2008年8月发布的HYDE3.0版本), 来自荷兰环境评价局HYDE数据集网站http://www.pbl.nl/en/themasites/hyde/index.html. SAGE数据集的空间分辨率为0.5°经纬网格, 时间覆盖1700~1992年, 时间分辨率为10年; HYDE数据空间分辨率为5′经纬网格, 时间跨度为10000 BC~2000 AD, 1700年以来的时间分辨率为10年. 区域重建数据采用叶瑜等[22]建立的东北过去300年耕地面积和垦殖率数据(CNEC数据), 空间分辨率为县级单元, 约合0.5°~1°经纬网格, 时间范围为1683~2002年. 由于受历史资料限制, 同时考虑重建结果的前后对比, 该套数据20世纪前的时间分辨率约为50~100年, 20世纪以来的时间分辨率为5~10年(表1).

SAGE数据集重建历史耕地的基本思路是: 利用遥感资料建立现代耕地分布图, 收集按行政单元统计的历史耕地数据, 计算各统计单元的历史耕地覆盖比率图, 将现代高空间分辨率耕地分布图与历史耕地覆盖比率图相乘, 获得具有地理属性的历史耕地数据(图1). 现代耕地数据决定了历史数据的空间分辨率和耕地空间分布型. 为了保持数据集均一的时间间隔, 对缺少历史清查数据的时间点, 用有资料支撑的临近时间断面的耕地数据进行线性插补[13].

HYDE数据集自发布以来数据不断更新. 初期版本HYDE2.0的建立思路是: 利用联合国人口统计

表1 近300年来SAGE, HYDE数据集和CNEC数据的时空分辨率

数据集	时间分辨率	重建时间断面	空间分辨率
SAGE	10年	1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 1992	0.5°
HYDE	10年	1700, 1710, 1720, 1730, 1740, 1750, 1760, 1770, 1780, 1790, 1800, 1810, 1820, 1830, 1840, 1850, 1860, 1870, 1880, 1890, 1900, 1910, 1920, 1930, 1940, 1950, 1960, 1970, 1980, 1990, 2000, 2005	5′
CNEC	100~5年	1683, 1735, 1780, 1908, 1914, 1931, 1940, 1950, 1955, 1960, 1970, 1980, 1990, 2002	分县, 约0.5°~1°

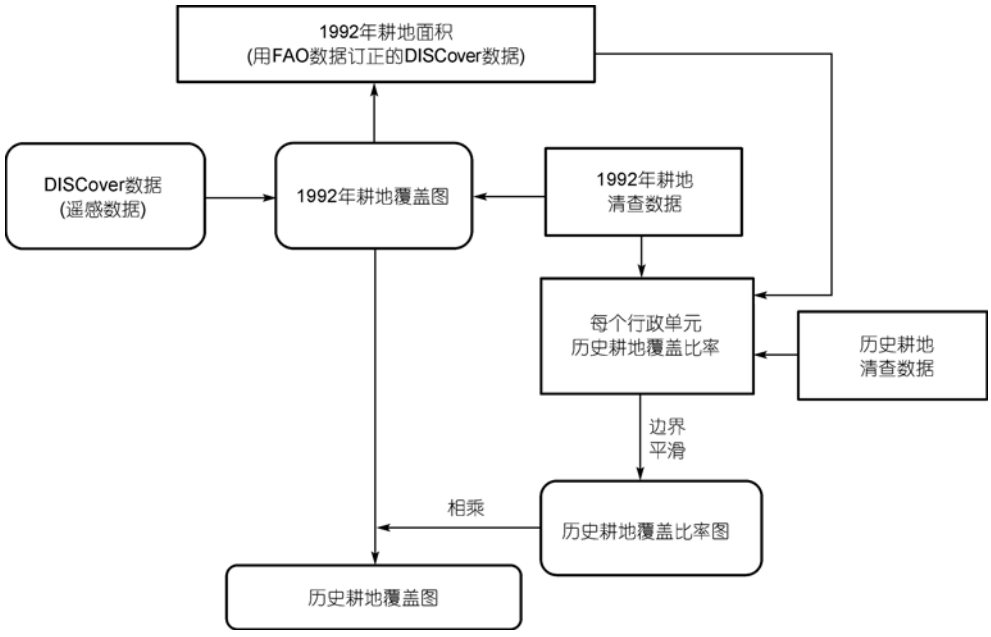

图1 SAGE历史耕地覆盖图重建算法

引自Ramankutty和Foley[13]. 方形框表示按行政单元统计的数据, 圆角框表示有地理属性的分布图. DISCover数据是利用1 km AVHRR遥感影像建立的全球土地利用数据. 提取1992年耕地覆盖时用同期统计数据对DISCover数据进行了订正; 1992年以前的耕地数据按行政单元统计, 通过行政单元间的边界平滑计算获得具有地理属性的耕地比率图

数据订正现代人口密度图(1994 年), 在此基础上结合联合国统计数据、国家普查数据及相关研究, 建立历史人口密度分布图, 以人口密度为权重分配历史统计单元内的耕地[23]. 本文使用的是HYDE3.0 版本, 它更新了人口数据和现代土地利用数据, 空间分辨率由HYDE2.0 的 0.5°提高为 5′经纬度网格, 并改进了耕地算法, 采用分时段权重图对耕地历史统计数据进行空间分配[12,14,24]. 其权重算法如下:

$$W_{\mathrm{crop},t}=\frac{G_{\mathrm{area}}-U_{\mathrm{area},t}}{G_{\mathrm{area\ max}}}\times W_{\mathrm{pop},t}\times W_{\mathrm{suit}}\times W_{\mathrm{river}}\times W_{\mathrm{slope}}\times W_{\mathrm{temp_crop}},$$

式中W表示权重, t表示历史的年份, crop为耕地, G_{area}为除去冰雪覆盖的所有土地面积, U_{area}为城市用地面积, pop为人口密度, suit为土壤宜耕性, river距水体(河流和海岸)距离, slope为坡度, temp_crop是农作物生长温度, 0℃为阈值.

上述两套HYDE数据的重建方法都以历史人口数据集为重要基础, 在重建历史人口数据时, 以地理属性明确的现代人口分布图为依据, 分配按行政单元统计的历史人口数据, 即认为现代人口密集的地区同样是历史上人口密集的地区[23,24]. 这使得以此为基础建立的耕地数据表现为, 历史耕地与现代耕地具有相似的空间分布格局.

CNEC数据是以现代调查数据为标准的东北三省过去 300 年分县耕地数据, 时间分辨率从 17~19 世纪的 50~100 年到 20 世纪的 5~10 年. 不同时段使用了不同类型的原始数据. 清代(17~19 世纪)主要是来自历史文献的原始田亩数据, 20 世纪前半叶为省政府档案和日本、俄罗斯调查数据, 1949 年以来是国家统计数据和 20 世纪 80 年代第一次土地概查数据. CNEC数据根据各段资料的特点进行了订正和归一化处理[22,25], 主要包括: 对于清代数据, 需要将原始田亩数据转化为统一的标准计量单位, 并进行亩制转换和订正, 同时采用瞒田率对耕地总量进行订正; 对于 20 世纪的数据, 通过建立多源耕地数据关系模型, 将统计数据订正到同期土地概查数据或调查数据(即实际耕地数据)水平; 利用清代官方统计数据与 20 世纪调查数据重叠的 1908 年, 将 300 年的数据衔接起来; 各时期耕地数据统计单元都被调整到现代县域行政单元, 得到以现代调查数据水平为标准的东北三省过去 300 年分县耕地数据.

1.2 数据预处理

为了评估 2 个全球数据集中东北地区历史耕地数据的准确性, 分别对其进行了总量评估、分省评估和网格评估, 分析了各时间断面上全球数据集与 CNEC 数据在不同空间尺度上的差异. 由于 SAGE 和 HYDE 数据集的耕地数据与 CNEC 数据格式存在差别(表 1), 在进行评估前对数据做了相应处理.

在选取评估时间断面时, 以 CNEC 数据中具有原始耕地记录的时间断面(表 1)为依据, 挑选了以下时间断面: 1700 年, 1740 年, 1780 年, 1910 年, 1920 年, 1930 年, 1940 年, 1950 年, 1980 年, 1990 年和 2005 年. 在数据评估中, 需要对 SAGE 和 HYDE 数据与 CNEC 数据的同期耕地面积作对比分析, 当 CNEC 数据的重建时间断面与挑选时间断面不一致时, 将 CNEC 数据中临近时间断面数据对应到所挑选的时间断面上, 其中 1683 年的 CNEC 数据对应 1700 年, 1908 年的 CNEC 数据对应 1910 年, 1914 年的 CNEC 数据对应 1920 年, 1931 年的 CNEC 数据对应 1930 年, 2002 年的 CNEC 数据对应 2005 年.

总量评估是以东北三省为单元进行统计对比. 分省评估是将SAGE和HYDE数据集中辽宁、吉林和黑龙江 3 省的耕地面积, 分别与CNEC分省数据进行比较. SAGE数据集提供了中国按省级行政单元输出的耕地数据, 因此可以直接得到分省耕地数据, 将分省数据加和得到东北三省耕地总量. HYDE数据集首先用东北三省行政边界进行切割, 再统计得到东北三省耕地总量和分省耕地面积. CNEC东北三省耕地总量和分省数据直接使用叶瑜等[22]发表文章中的数据.

为了揭示全球数据集与区域数据差异的空间分布, 对耕地数据进行了网格评估. 由于 SAGE 与 HYDE 数据集的耕地数据是网格数据, 而 CNEC 数据是以县为统计单元, 因此网格评估时需要将 3 组数据转化成统一的空间表达形式. 东北县域平均面积与 1°经纬网格面积较为接近, 为了控制 CNEC 数据转为网格数据的误差, 选取了 1°经纬网格作为 3 组数据共同的网格分析单位. SAGE 和 HYDE 数据直接升尺度, 生成空间分辨率为 1°的经纬网格数据. CNEC 数据处理是按照县域内使用统一垦殖率, 将 1°经纬网格与分县耕地垦殖率图叠加, 计算网格内各斑块耕地面积, 重新按网格统计转为 1°经纬网格数据形式.

1.3 评估方法

总量评估和分省评估以 CNEC 数据为基数, 分别计算各时间断面上 SAGE 与 HYDE 耕地数据相当于 CNEC 数据的差异倍数(即相对差异比值), 其计算方法是将同一时间断面上的 SAGE 与 HYDE 耕地数据分别除以 CNEC 数据.

基于分省数据评估结果,选择 HYDE 数据集的耕地数据做了进一步的网格评估. 对 HYDE 数据集和 CNEC 数据提取转化的 1°经纬网格耕地数据分别进行了时间变化图谱对比、网格绝对误差和相对误差、分时段耕地增长重心分析. 由于 CNEC 耕地数据转换为网格数据时可能会产生系统和计算误差, 使用 CNEC 网格耕地面积对 HYDE 数据评估时, 只对其计算得出的绝对误差和相对误差的空间分布进行分析, 不考虑具体数值. 绝对误差和相对误差计算公式如下:

$$\text{绝对误差: } Ea = X_{Hi} - X_{Ci},$$

$$\text{相对误差: } Er = (X_{Hi} - X_{Ci})/X_{Ci} \times 100\%,$$

式中X_{Hi}是i时间断面HYDE数据集的耕地面积, X_{Ci}是对应时间断面CENC数据的耕地面积.

2 结果与分析

2.1 总量评估

东北地区是过去 300 年我国土地利用/覆盖变化最显著的区域. 受清代移民开垦政策的影响, 东北三省 17~19 世纪农垦区扩张十分缓慢, 大规模的耕地增长发生在 20 世纪以来, CNEC数据耕地总量变化曲线显示东北三省 17~20 世纪耕地呈现指数型增长. 17~19 世纪CNEC数据东北三省耕地面积增长缓慢, 耕地总量与现代相比少得多, 东北三省在 1683, 1735, 1780 和 1908 年的总耕地面积分别为 5396, 15266, 21867 和 81016 km^2; 这段时间耕地面积的增长主要发生在 1840~1908 年, 耕地面积翻了近 3 倍, 年平均增长率约为 1.6%, 而其他时段 50~100 年间耕地面积增长量均小于 1×10^4 km^2, 年平均增长率均在 1%以下(除 1683~1735 年因基数小而增长率高于 2.0%外). 20 世纪以来, 东北三省耕地面积呈阶段性快速增长, 1914 年东北三省总的耕地面积为 102245 km^2, 至 1960 年已达到 196519 km^2, 接近现代耕地面积水平, 1914~1931 年和 1950~1960 年为两个增长最为迅速的阶段, 其年平均增长率分别达到 2.3%和 3.1%, 其间 1940~1950 年耕地面积持平后小幅减少; 1960~1980 年呈缓慢增长趋势, 20 年间耕地面积增加 0.68×10^4 km^2, 年平均增长率为 0.2%; 1980~2000 年呈下降趋势, 年平均增长率为–0.5%, 尤其是 1990~2000 年下降迅速, 1990 年和 2002 年耕地面积分别为 211516 和 182140 km^2, 年平均增长率达到–1.2%[22].

SAGE数据集东北三省耕地面积的时间变化序列显示, 1700 年耕地面积已达 113210 km^2, 1700~1950 年耕地总量以每年 0.4%的速率线性增加(拟合方程为y_i=112.36i–179852.62, r^2=0.99, i为年份, y_i为i年耕地面积), 1950 年后呈下降趋势(图 2). 1700~1950 年SAGE数据集耕地面积的持续线性增长与东北地区先慢后快、阶段性增长的耕地增长过程相去甚远. 20 世纪 50 年代以来SAGE数据呈下降趋势, 1950~ 1970 年、1970~1980 年和 1980~1990 年耕地面积分别减少 2.43×10^4, 4.81×10^{4}和 0.75×10^4 km^2, 而东北地区自建国来农垦区持续扩展, 耕地总量仍有增长, 直至 20 世纪 80 年代受到退耕还林政策影响耕地面积才呈现下降趋势, 1980~1990 年耕地面积减少仅 0.18×10^4 km^2. 显然 1950 年后SAGE数据耕地面积变化亦与东北地区近代开垦进程不符. 与CNEC数据对比, 1700 年以来SAGE数据集重建的耕地面积在绝对量上一直明显大于CNEC数据. 从SAGE耕地数据与CNEC数据的相对差异比值变化来看, 两组数据在 18~19 世纪差异巨大, 1700 年和 1780 年相对差异比值分别为 20.98 和 9.22, 虽然随着时间推移, SAGE数据与CNEC数据的相对差异比值逐渐小, 1910 年为 4.22, 但直到 1990 年SAGE数据仍是CNEC数据的 1.6 倍(表 2).

HYDE 数据集东北三省耕地面积变化曲线显示, 除了 1960~1980 年耕地面积有小幅减少外, 其余时段耕地面积都呈现增加趋势(图 3), 不同时段耕地增长速率各不相同, 1900 年以前增加较为缓慢, 1900~1960 年耕地总量增加速率有所提高, 1980 年以来耕地增长率比以往都要高. HYDE 数据集的东北耕地总量与 CNEC 数据比较表明, 18~20 世纪初 HYDE 数据的耕地面积大于 CNEC 数据, 两者相对差异比值在 18~19 世纪最大, 1700 年和 1780 年分别为 9.87 和 2.99, 1910 年 HYDE 数据的耕地总量与 CNEC 数据最为接近, 相对差异比值为 1.11; 20 世纪 30 年代以来 HYDE 数据则小于 CNEC 数据, 两者相对差异比值均小于 1, 平均约为 0.69, 即 HYDE 数据比 CNEC 数据

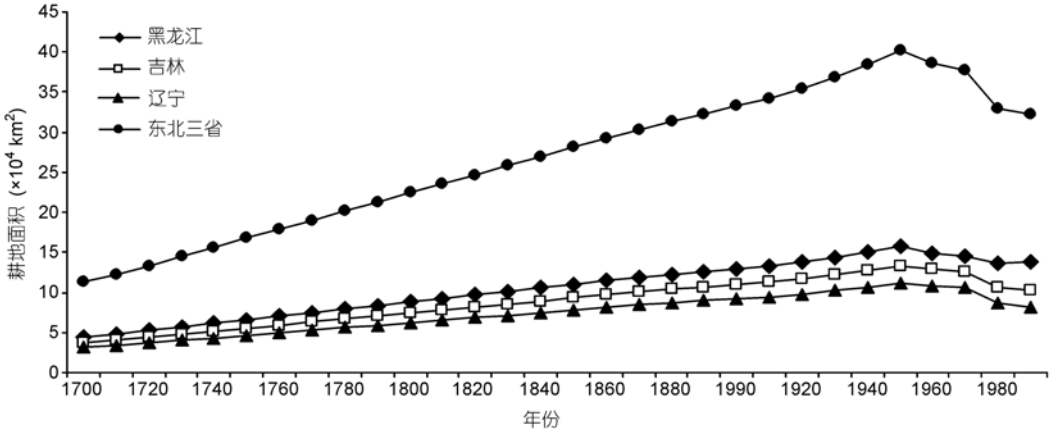

图 2　SAGE 数据集 1700~1900 年东北三省耕地面积变化序列

表 2　SAGE, HYDE数据集和CNEC数据东北分省耕地数据(km²)及相对差异比值(括号内, 单位: %)

区域	数据集	1700 年	1780 年	1910 年	1930 年	1950 年	1990 年
东北三省	SAGE	113210(20.98)	201650(9.22)	341800(4.22)	367800(2.45)	401600(2.77)	321810(1.60)
	HYDE	53244(9.87)	65417(2.99)	89672(1.11)	101219(0.68)	109405(0.75)	140252(0.70)
	CNEC	5396(1.00)	21867(1.00)	81016(1.00)	150128(1.00)	145058(1.00)	201516(1.00)
辽宁	SAGE	31370(5.81)	55900(3.15)	94780(4.18)	101990(2.54)	111340(2.40)	80630(1.86)
	HYDE	11938(2.21)	14792(0.83)	20445(0.90)	23206(0.58)	25173(0.54)	32673(0.76)
	CNEC	5390(1.00)	17749(1.00)	22683(1.00)	40212(1.00)	46312(1.00)	42992(1.00)
吉林	SAGE	37510(–)	66820(26.21)	113290(3.32)	121940(2.22)	133160(3.14)	102990(2.09)
	HYDE	17276(–)	21111(8.28)	28796(0.84)	32578(0.59)	35270(0.83)	44980(0.91)
	CNEC	0(–)	2549(1.00)	34172(1.00)	54813(1.00)	42342(1.00)	49168(1.00)
黑龙江	SAGE	44330(–)	78930(50.31)	133730(5.53)	143870(2.61)	157100(2.79)	138190(1.26)
	HYDE	24030(–)	29514(18.81)	40431(1.67)	45435(0.82)	48962(0.87)	62599(0.57)
	CNEC	0(–)	1569(1.00)	24161(1.00)	55103(1.00)	56404(1.00)	109356(1.00)

小约 30%(表 2). 1700~1910 年, HYDE数据集的东北三省耕地总量增加了 3.6×10^4 km², 年平均增长率为 0.3%, 远小于同一时期CNEC数据的增长率(1683~1908 年CNEC的耕地年平均增长率为 6.2%). 1910~1960 年, HYDE数据集的耕地面积呈持续线性增长, 年增长速率为 0.5%, 而CNEC数据显示这 50 年间耕地面积增长具有明显的年代际差异, 1914~1931 年耕地面积年平均增长率为 2.3%, 1940~1950 年基本持平后有小幅减少, 1950~1960 年又快速增加, 速率达 3.1%. 1960~1980 年HYDE耕地面积呈现小幅下降, 减少约 0.55×10^4 km², 而同期CNEC数据的耕地面积没有减少而是增加 0.68×10^4 km². 1980 年以来HYDE数据耕地面积以前所未有的速率快速增加, 年平均增长率约 2.0%, 而CNEC数据则呈下降趋势, 年平均增长率为–0.5%. 虽然HYDE数据集在总体增加趋势上反映出 20 世纪以来耕地开垦速率提高的特点, 但是HYDE数据集耕地百年尺度上的增长速率均低于CNEC数据, 并且其耕地增长的具体时段也与CNEC数据不同.

2.2　分省评估

CNEC数据东北分省耕地面积时间序列显示, 辽宁、吉林和黑龙江三省的耕地增长特点各不相同, 并且东北地区耕地增长重心随着时间由南到北推移. 清初东北地区的开垦仅局限在辽宁, 1683 年辽宁耕地面积为 5396 km², 而黑龙江和吉林的耕地面积为 0. 17~18 世纪辽宁的耕地面积明显大于吉林和黑龙江, 而且增长幅度最大, 1683~1780 年辽宁耕地面积增加了 1.2×10^4 km²; 至 1780 年, 辽宁、吉林和黑龙江的耕地面积分别为 17749, 2549 和 1569 km². 19 世纪吉林和黑龙江 2 省耕地大幅增加, 至 19 世纪末 20 世纪初 2 省的耕地面积均已超过辽宁. 1908 年东北

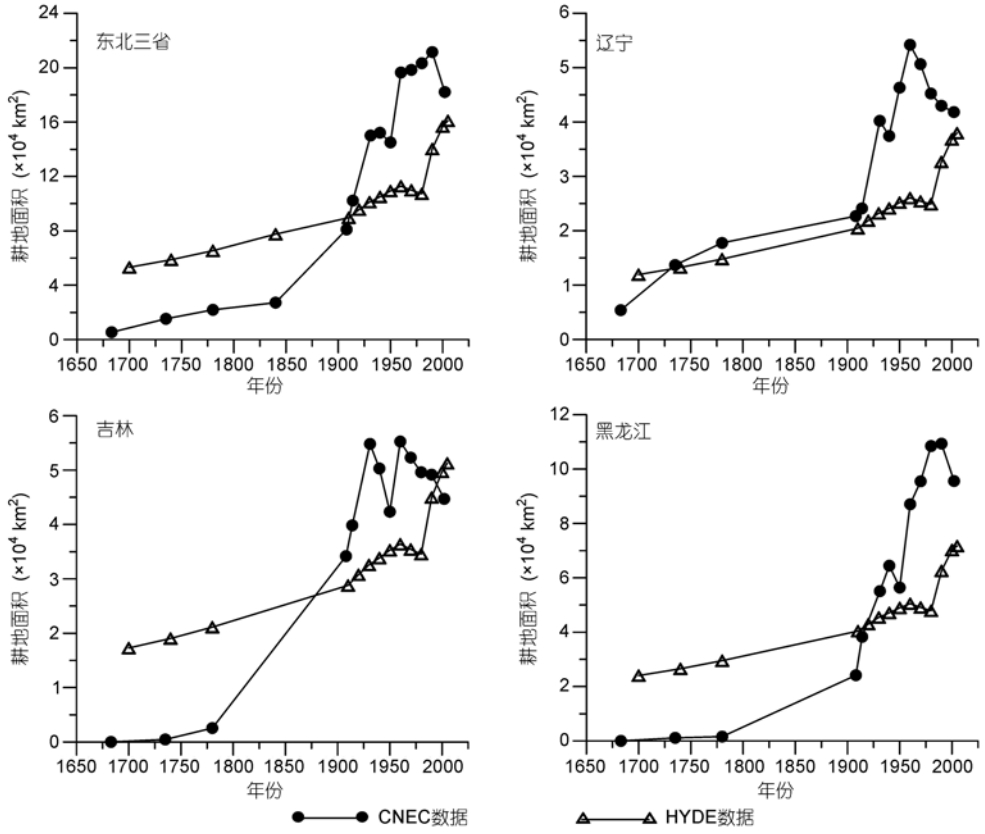

图 3　HYDE 数据集和 CNEC 数据东北三省耕地面积变化对比

三省中耕地面积最多的是吉林, 为 34172 km^2, 1780~1908 年增加约 3.16×10^4 km^2; 其次为黑龙江 24161 km^2, 增幅约 2.26×10^4 km^2; 辽宁耕地面积最少, 为 22683 km^2, 增长也有限, 约为 0.49×10^4 km^2. 20 世纪黑龙江耕地面积增长幅度最大, 至 20 世纪 30 年代黑龙江的耕地面积已经超过了吉林和辽宁, 20 世纪末更是超过了 2 省之和. 20 世纪东北三省耕地增长的年代际变化各有特点. 黑龙江耕地增长时段持续最长且增长率高, 包括 1914~1940, 1950~1960 和 1960~1980 年, 年平均增长率分别为 2.0%, 4.4%和 1.1%, 1940~1950 年和 1990~2000 年呈下降趋势, 年平均增长率分别为–1.3%和–1.1%, 1980~1990 年基本持平; 吉林耕地增长阶段有两个, 即 1914~1931 年和 1950~1960 年, 年平均增长率分别达到 1.9%和 2.7%, 1931~1950 年和 1960~2000 年呈持续下降趋势, 年平均增长率分别为–1.3%和–0.5%; 辽宁耕地增长的两个阶段是 1914~1931 年和 1940~1960 年, 年平均增长率分为 3.1%和 1.6%, 1931~1940 年和 1960~2000 年耕地面积有所减少, 年平均增长率分别为–0.8%和–0.6%[22].

SAGE数据集 1700~1950 年分省耕地面积变化曲线均可以用线性方程拟合(3 个拟合方程的r^2均达到 0.99), 方程斜率大小说明黑龙江耕地年平均增长率最大、吉林次之、辽宁最小; 1950 年以后 3 省耕地面积都呈现下降趋势, 只有黑龙江在 1980~1990 年间略有增加(图 2). SAGE分省数据没有反映出东北各省耕地面积阶段性增长的特点. SAGE数据集东北各省耕地绝对数量均比CNEC数据大. 从相对差异比值来看, 吉林和黑龙江 2 省累计相对差异比值远大于辽宁, 贡献最大的时间段是 18~19 世纪, 1700 年CENC数据中吉林和黑龙江耕地面积为 0 而无法计算相对差异比值, 同期辽宁的相对差异比值为 5.81; 1780 年吉林和黑龙江的相对差异比值分别为 26.21 和 50.31, 辽宁仅为 3.15. 说明SAGE数据集过高估计了东北各省耕地数量, 最大重建误差来自对吉林和黑龙江 18~19 世纪耕地面积的估算.

HYDE 数据集的分省耕地面积时间序列与东北耕地总量的变化型一致, 1700~1910 年缓慢线性增加, 1910~1960 年增长速度加快, 1960~1980 年略呈下降趋势, 1980 年后快速增长, 显然 HYDE 分省耕地变化曲线与 CNEC 数据东北三省各具特点的耕地增长方式不一致, 没有反映出各省年代际尺度上耕地面积的变化(图 3). 1700 年来, HYDE 数据同一时期辽宁、吉林和黑龙江 3 省耕地面积的比值基本保持不变, 约为 1∶1.40∶1.95, 因此 HYDE 数据没有记录下东北历史开垦重心由南向北迁移的特点. HYDE 数据集各省耕地面积与 CNEC 数据的绝对数量差值和相对差异比值的时间变化各不相同, 总的来说, 吉林和黑龙江 2 省在 1700 年和 1780 年相对误差最大(表 2). 辽宁相对差异比值仅在 1700 年大于 1, 约为 2.21; 自 1740 年之后均小于 1, 其中 1740~1920 年和 1990~2000 年两段时间较接近于 1, 平均差异比值分别为 0.90 和 0.82, 在这 2 段时间 HYDE 数据中辽宁耕地面积误差较小. 由于 CNEC 数据的耕地面积为 0, 吉林 1700 年相对差异比值无法计算, 1780 年的比值为 8.28, 说明 HYDE 数据集吉林的耕地面积重建结果在 18~19 世纪误差最大, 1910 年以来相对差异比值小于 1, 其中 1910 年、1950 年和 1990 年的比值分别为 0.84, 0.83 和 0.91, 在这 3 个时间断面上 HYDE 数据误差较小. 由于 CNEC 数据的耕地数量为 0, 黑龙江 1700 年相对差异比值无法计算, 1780 年的比值为 18.81, 表明黑龙江也是 20 世纪前的重建结果误差最大, 1930 年以后其比值开始小于 1, 其中 1930 年和 1950 年的比值分别为 0.82 和 0.87, 相对接近 1, 即在这 2 个时间断面上误差较小.

2.3 网格评估

总量评估和分省评估表明 SAGE 数据集明显高估了东北耕地面积, 存在显著误差, 没有进一步分析的必要; HYDE 数据集和 CNEC 数据在部分时间断面上重建结果较为接近, 因此对 HYDE 数据集进行了网格评估, 以分析其耕地数据误差的空间分布.

CNEC 网格数据时间变化图谱反映了东北地区近 300 年来耕地覆盖的空间变化. 17~19 世纪东北主要农垦区的空间变化不大, 1683 年耕地仅分布在辽宁省中部的辽河平原下游地区, 1780 年东北的主要农垦区仍在辽宁, 此外在吉林省中部相当于今吉林市所辖的区域有垦殖率超过 5%的垦区. 19 世纪末至 20 世纪受到政府“移民放垦”政策影响, 东北耕地的空间分布发生了显著变化, 垦殖北界延伸至黑龙江省中部. 20 世纪有两次明显的耕地空间扩张和垦殖率增加过程: 20 世纪初至 20 世纪 30 年代耕地由吉林、黑龙江中部向东西两侧扩展, 其中西侧的耕地增长更多, 原中部农垦区的垦殖率增加; 20 世纪 50~80 年代新增耕地主要分布在三江平原和小兴安岭南麓, 原有农垦区垦殖率进一步加大.

对比HYDE和CNEC耕地数据的时间变化图谱, 发现两组数据记录的耕地空间扩展和增长过程存在明显差异(图 4). HYDE耕地数据各时期耕地在区域的空间分布大势与现代东北耕地分布空间格局一致, 耕地集中分布在松嫩平原和下辽河平原; 近 300 年来耕地覆盖空间变化显示, 辽河-松嫩平原耕地面积持续增加, 垦殖范围沿此线向东西两侧扩展; 黑龙江中西部地区在 1700 年已有耕地分布, 三江平原在 20 世纪初开垦率已达到 5%~10%. HYDE数据并没有反映出开垦沿东北中部平原地区自南向北扩展的空间变化特点, 并且黑龙江大规模开垦的起始时间明显早于历史文献记录的农耕北界首次北上时期(1796~1901 年)[22].

HYDE 网格数据绝对误差的空间分布随着时间推移而变化, 总体可以分为两个阶段: 18~19 世纪, HYDE 数据绝对误差大部分为正, 误差值较大的地区集中在吉林西部和黑龙江中南-西南部. 20 世纪以来, 绝对误差逐渐变成以负值居多, 1910 年正的绝对误差主要分布在黑龙江西南部和三江平原西部, 负值分布在黑龙江中部和吉林中部; 1930 年以来绝对误差大部分为负值, 连片分布在松嫩平原. 说明 HYDE 数据在 1900 年以前对吉林和黑龙江的耕地面积估算偏大, 而 1900 年以来对松嫩平原的耕地面积普遍估计偏小, 并且总量和分省评估中误差较小的 1910 年数据, 网格绝对误差十分明显.

相对误差能够去除不同网格耕地面积差异的影响, 使区域内各网格误差可以相互比较. 通过统计各时间断面HYDE数据的相对误差, 表明18~19世纪大部分网格耕地面积的相对误差大于 100%, 20 世纪初相对误差主要集中在–10%~–100%和>100%两个等级, 20 世纪 30 年代以来多数网格的误差在–10%~–100%之间(表 3). 1700 年和 1780 年 HYDE 数据的相对误差大于 100%的网格数占总网格数比例分别是 63.55%和 53.27%(包括由于 CNEC 吉林和黑龙江耕

表 3　HYDE 数据相对误差统计表

年份	占总网格数的比例(%)							最小值(%)	最大值(%)	均值(%)	标准差
	−100~−50	−50~−10	−10~0	0~10[a)]	10~50	50~100	>100[b)]				
1700	0.00	2.80	0.00	27.10	3.74	2.80	63.55	41.55	1417.43	245.31	3.43
1780	13.08	6.54	0.93	20.56	0.93	4.67	53.27	−100	21449.86	2142.75	53.11
1910	14.95	13.08	4.67	20.56	9.35	5.61	31.78	−100	3405.62	288.63	6.97
1930	42.06	26.17	0.93	4.67	9.35	3.74	13.08	−100	1125.08	8.3	1.94
1950	25.23	27.10	2.80	18.69	12.15	4.67	9.35	−100	1671.72	9.37	1.96
1980	50.47	37.38	2.80	0.93	6.54	0.93	0.93	−100	283.74	−51.2	0.50
1990	36.45	33.64	11.21	4.67	7.48	4.67	1.87	−100	305.04	−34.37	0.59
2005	19.63	25.23	6.54	9.35	21.50	8.41	9.35	−97	899.73	16.39	1.26

a) 包括 HYDE 数据集和 CNEC 数据中耕地面积均为 0 的网格; b) 包括了因 CNEC 数据中耕地面积为 0 而无法计算相对误差的网格

地面积为 0 而无法计算相对误差的网格), 它们主要分布在吉林和黑龙江现代农垦区范围内; 2 个时间断面上相对误差为 0~10%的网格分别约占 27.10%和 20.56%, 主要是 HYDE 和 CNEC 数据中耕地面积为 0 的未开垦区, 分布在小兴安岭地区. 20 世纪以来, HYDE 网格数据的相对误差总体减小, 无论是误差范围、均值和方差, 都比 18~19 世纪有所下降. 1910 年相对误差大于 100%的网格约占 31.78%, 主要分布在黑龙江松嫩平原西部和三江平原西部; 相对误差为−10%~−50%和−50%~−100%的网格比例分别为 13.08%和 14.95%, 主要分布在辽宁东部和黑龙江中部; 相对误差为 0~10%的网格占到 20.56%, 大多是 HYDE 和 CNEC 数据中未开垦区域, 分布在小兴安岭北段. 鉴于以上结果, 虽然 1910 年 HYDE 数据与 CNEC 数据的东北地区耕地总量接近, 但耕地空间分布存在明显差异. 1930 年以来相对误差大于 100%的网格所占比例小于 15%, 零散分布在东北区域边缘, 如三江平原东部和辽河-松嫩平原西部; 多数网格的相对误差在−10%~−100%之间, 相对误差为−10%~−50%的网格主要分布在辽河-松嫩平原, 相对误差在−50%~−100%的网格主要分布在小兴安岭和长白山一带. 2005 年是 20 世纪中除了 1910 年外 HYDE 数据与 CNEC 数据在总量上较为接近的一年, 并且具有统计数据和遥感资料支持, 但这一年网格相对误差在−50%~−100%, −10%~−50%和 10%~50%分别为 19.63%, 25.23%和 21.50%, 说明 HYDE 现代数据耕地空间分布存在一定误差, 对黑龙江省的耕地面积估计过低. 总之, 相对误差时间变化和空间分布表明, 18~19 世纪 HYDE 数据集的重建结果精度远低于 20 世纪, 误差较大的网格主要分布在 CNEC 数据的新增农垦区; 20 世纪以来误差的空间分布表明, HYDE 数据集对新增农垦区估计偏大, 而对旧垦区垦殖率的增加估计偏小.

3　结论与讨论

本文采用空间分辨率较高的 CNEC 数据, 对 SAGE 和 HYDE 数据集东北地区近 300 年来耕地数据的区域重建精度进行了评估, 得到以下结论:

(1) 总量评估和分省评估表明, SAGE 和 HYDE 数据集对东北地区耕地重建均存在明显误差, 不能真实反映历史耕地面积及其变化特点, 在东北地区直接使用这 2 个全球数据集的耕地数据进行土地利用变化影响定量评估, 会影响模拟结果的准确性.

(2) SAGE 数据集对东北近 300 年历史耕地面积估算过高, 无论总量还是分省统计, SAGE 数据集重建值均大于 CNEC 数据, 尤其是 18~19 世纪数据误差最大. 1700~1950 年 SAGE 数据集的耕地面积呈线性增长, 与东北地区先慢后快、阶段性增长过程不同. 并且 SAGE 数据集同时段东北三省耕地面积大小排序不变, 没有体现东北历史开垦重心由南向北推移的空间变化.

(3) HYDE 数据集的总量和分省评估中均表明其 18 世纪耕地数据重建误差最大, 主要来自黑龙江和吉林 2 省; 其耕地面积变化在世纪尺度的增长率和年代际变化上都与 CNEC 数据不同. HYDE 数据集中同一时间断面上黑龙江、吉林、辽宁三省间耕地比例几乎不变, 没有反映出开垦重心在 3 个省份中自南向北更替的空间变化.

(4) 网格评估结果表明, HYDE 数据集在 18~19 世纪绝对误差和相对误差最大, 主要分布在吉林和黑龙江 2 省; 20 世纪以来对整个东北估算普遍偏小, 主要分布在东北中部平原和小兴安岭地区, 而辽河-松嫩

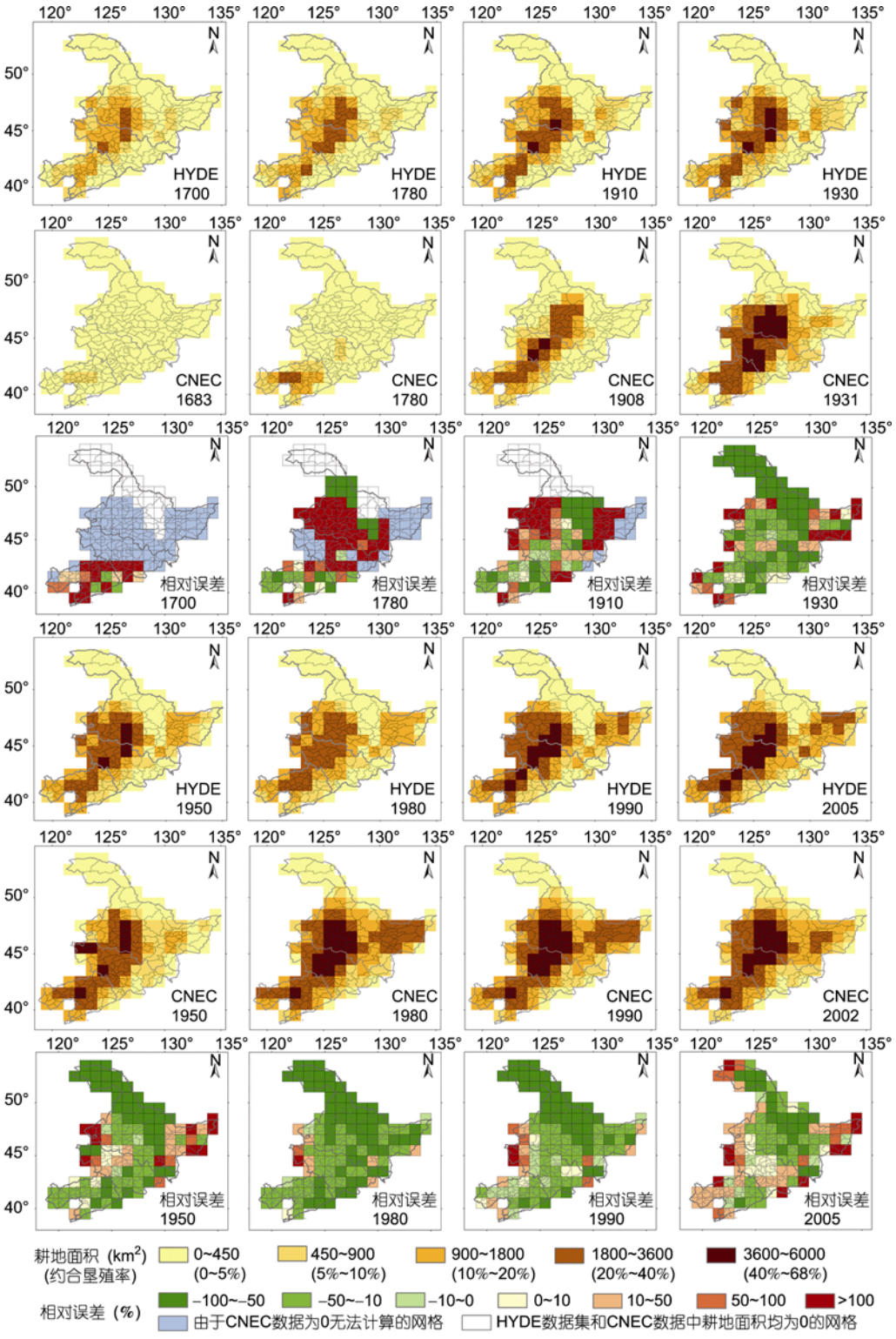

图 4　HYDE 数据集与 CNEC 数据网格耕地面积变化与相对误差时间图谱(统计单元为 1°经纬网格)

平原西部、三江平原-长白山地区的耕地面积估计偏大.

上述评估结果表明SAGE和HYDE数据集在东北地区耕地数据估算中存在相当程度的误差, 可能的误差来源主要有以下两方面: 一是全球数据集的历史耕地重建方法. 全球数据集建立的基本思路是利用现代耕地分布的空间型进行历史复原, 采用线性插补建立均一时间间隔的数据. SAGE和HYDE数据集分别以遥感资料提取的现代耕地覆盖图和人口密度图为基础, 获得耕地空间分布权重图, 分配以国家或区域为统计单元的历史数据, 即认为历史耕地数据空间分布格局与现代相似. 这种理想化的推断忽略了移民、战争和政策等多种自然社会耦合因素对开垦空间扩展和耕地增长速率的影响, 对东北地区这样由于大规模移民开垦而形成的农垦区而言, 不能反映真实的人类土地利用历史进程, 最终导致重建数据偏离历史真相. 另一个主要误差来源是全球数据集采用的现代基础底图. 从全球数据集历史数据的重建方案来看, 现代基础地图对重建结果的准确性具有重要影响, 而作为现代底图建立基础的遥感资料本身既存在误差. SAGE数据集使用了IGBP建立的DISCover数据建立了 5′×5′全球耕地数据[26], 对DISCover数据的评估表明其全球精度为66.9%[27], 尽管SAGE数据集利用联合国粮农组织(FAO, 1995)的统计数据对其耕地数据进行了订正, Ramankutty和Foley[26]承认其全球耕地重建结果仍然比FAO的数据大得多, 特别是前苏联、中国等地区数据要大1.5倍, 甚至更多. HYDE现代耕地数据建立在DISCover数据和GLC2000数据两个遥感数据基础上[28], GLC2000数据评估显示其全球精度(68.6%)与DISCover数据相近[29], HYDE也将遥感数据提取的耕地面积订正到了FAO统计数据上(FAO, 2005), 而本文网格评估显示1990年和2000年HYDE东北地区的耕地数据依然存在较大误差, 特别是黑龙江省1990年和2000年的耕地面积分别比现代调查数据小42.76%和26.51%.

全球土地利用数据库在区域尺度上的数据精度还有待进一步评估, 从历史数据来源和重建方法来看, 其无法保证低于行政统计单元(国家或者次一级行政单元)重建结果的准确性. 因此, 区域尺度上的历史土地利用重建研究是对全球数据的验证和必要补充. 完善全球历史时期的土地利用变化数据, 需要更多的土地统计数据、历史时期土地利用图和其他影响土地利用的重大历史事件信息, 而这些资料收集和整理需要巨大的投入. 区域尺度上的细致研究及集成, 将是建设有精度保证的全球土地利用数据更为便捷的途径. 因此, 进一步研究应关注如何利用已有区域数据对全球数据集进行订正, 可以从两方面切入: 一是使用区域历史真实数据的变化趋势替代全球数据集线性趋势, 纠正简单线性回推的误差; 二是分析全球数据集中历史资料的精度, 采用空间分辨率更高的历史区域数据改进全球数据集.

致谢 感谢审稿专家提出的宝贵意见.

参考文献

1 Vitousek P M, Mooney H A, Lubchenco J, et al. Human domination of the Earth's ecosystems. Science, 1997, 277: 494—499
2 Lambin E F, Turner B L, Geist H J, et al. The causes of land-use and land-cover change: Moving beyond the myths. Glob Environ Change, 2001, 11: 261—269
3 Bolin B, Sukumar R. Global perspective. In: Watson R T, Noble I R, Bolin B, et al., eds. Land Use, Land-use Change, and Forestry. UK Cambridge: Cambridge University Press, 2000. 25—45
4 Hansen J E, Sato M, Lacis A, et al. Climate forcings in the Industial era. Proc Natl Acad Sci, 1998, 95: 12753—12758
5 Feddema J, Oleson K, Bonan G, et al. A comparison of a GCM response to historical anthropogenic land cover change and model sensitivity to uncertainty in present-day land cover representations. Climate Dyn, 2005, 25: 581—609
6 IPCC. Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge and New York: Cambridge University Press, 2007. 180—185, 512
7 Houghton R A. Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850—2000. Tellus Ser B-Chem Phys Meteorol, 2003, 55: 378—390
8 Betts R A, Falloon P D, Goldewijk K K, et al. Biogeophysical effects of land use on climate: Model simulations of radiative forcing and

larg-scale temperature change. Agric Forest Meteorol, 2007, 142: 216—233
9 Zhen J Y, Lin S S, He F N. Recent progress in studies on land cover change and its regional climatic effects over China during historical times. Adv Atmos Sci, 2009, 26: 793—802
10 葛全胜, 戴君虎, 何凡能, 等. 过去 300 年中国土地利用、土地覆被变化与碳循环研究. 中国科学 D 辑: 地球科学, 2008, 38: 197—210
11 LUCC Scientific Steering Committee. Key findings of LUCC on its research questions. Glob Change Newsl, 2004, 63: 12—14
12 Goldewijk K K, Ramankutty N. Land cover change over the last three centuries due to human activities: The availability of new global data sets. Geo J, 2004, 61: 335—344
13 Ramankutty N, Foley J A. Estimating historical changes in global land cover: Croplands from 1700 to 1992. Glob Biogeochem Cycle, 1999, 13: 997—1027
14 Goldewijk K K, Drecht G V. HYDE 3: Current and historical population and land cover. In: Bouwman A F, Kram T, Goldewijk K K, ed. Integrated Modelling of Global Environmental Change: An Overview of IMAGE 2.4. Bilthoven: Netherlands Environmental Assessment Agency (MNP), 2006
15 Tian H , Melillo J M, Kicklighterc D W, et al. Regional carbon dynamics in monsoon Asia and its implications for the global carbon cycle. Glob Planet Change, 2003, 37: 201—217
16 Brovkin V, Sitch S, Bloh W V, et al. Role of land cover changes for atmospheric CO_2 increase and climate change during the last 150 years. Glob Change Biol, 2004, 10: 1253—1266
17 Oost K V, Quine T A, Govers G, et al. The impact of agricultural soil erosion on the global carbon cycle. Science, 2007, 318: 626—629
18 Matthews H D, Weaver A J, Eby M, et al. Radiative forcing of climate by historical land cover change. Geophys Res Lett, 2003, 30: 1055—1059
19 石正国, 延晓东, 尹崇华, 等. 人类土地利用的历史变化对气候的影响. 科学通报, 2007, 52: 1436—1444
20 陈星, 雷鸣, 汤剑平. 地表植被改变对气候变化影响的模拟研究. 地球科学进展, 2006, 21: 1075—1082
21 李巧萍, 丁一汇, 董文杰. 中国近代土地利用变化对区域气候影响的数值模拟. 气象学报, 2006, 64: 257—270
22 叶瑜, 方修琦, 任玉玉, 等. 东北地区过去 300 年耕地覆盖变化. 中国科学 D 辑: 地球科学, 2009, 39: 340—350
23 Goldewijk K K. Estimating global land use change over the past 300 years: The HYDE database. Glob Biogeochem Cycle, 2001, 15: 417—433
24 Goldewijk K K. Three centuries of global population growth: A spatial referenced population density database for 1700－2000. Popul Environ, 2005, 26: 343—367
25 叶瑜, 方修琦, 戴玉娟, 等. 东北三省民国时期耕地数据的同化与垦殖率重建. 自然科学进展, 2006, 16: 1419—1427
26 Ramankutty N, Foley J A. Characterizing patterns of global land use: An analysis of global croplands data. Glob Biogeochem Cycles, 1998, 12: 667—685
27 Scepan J. Thematic validation of high-resolution global land-cover data sets. Photogramm Eng Rem Sens, 1999, 65: 1051—1060
28 Goldewijk K K, Drecht G V, Bouwman A F. Mapping contemporary global cropland and grassland distributions on a 5×5 minute resolution. J Land Use Sci, 2007, 2: 167—190
29 Mayaux P, Eva H, Gallego J, et al. Validation of the global land cover 2000 map. IEEE Trans Geosci Remote Sensing, 2006, 44: 1728—1739