中国科学: 地球科学 2016 年 第 46 卷 第 11 期: 1459 ~ 1471
SCIENTIA SINICA Terrae earthcn.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS

论 文 **GlobeLand30 遥感制图创新与大数据分析专题**

不同耕地数据集在中国区域的对比研究

陆苗[①], 吴文斌[①*], 张莉[①], 廖安平[②], 彭舒[②], 唐华俊[①†]

① 中国农业科学院农业资源与农业区划研究所, 农业部农业信息技术重点实验室, 北京 100081;
② 国家基础地理信息中心, 北京 100830
* 通讯作者, E-mail: wuwenbin@caas.cn;
† 通讯作者, tanghuajun@caas.cn

收稿日期: 2015-09-18; 接受日期: 2016-04-25; 网络版发表日期: 2016-09-27
国家自然科学基金项目(批准号: 41501483)和测绘地理信息公益性行业科研专项项目(编号: 201512028)资助

摘要 耕地的数量和空间分布信息是国家粮食安全、全球环境变化、陆地生态系统地球物理和地球化学循环过程等研究的基础. 本文以中国区域耕地为研究对象, 从耕地的数量和空间位置两方面, 对 2010 基准年的 4 套全球地表覆盖产品 GlobeLand30、FROM-GLC、GlobCover 和 MODIS Collection 5, 以及耕地数据集 MODIS Cropland 进行了对比研究. 结果表明, 不管是从耕地数量还是空间位置, GlobeLand30 的准确度都要高于其他 4 套产品. 从耕地的数量来看, 5 套数据集计算的耕地面积比例在多数省份都有差异, 和统计数据相比, GlobeLand30 和统计数据的吻合度最好, MODIS Collection 5 和 FROM-GLC 次之, MODIS Cropland 和 GlobCover 较低. 在空间位置方面, GlobeLand30 的空间位置精度最高, 其次是 FROM-GLC 和 MODIS Collection 5, GlobCover 和 MODIS Cropland 的精度最低; 同时, 5 套数据的空间一致性从农业主产区到草原牧区逐渐降低, 且受高程和坡度的影响明显. 本文发现, 虽然 MODIS Collection 5 的空间分辨率最低, 但其耕地的数量精度和空间位置精度都高于 300m 空间分辨率的 GlobCover 和 250m 空间分辨率的 MODIS Cropland. 因此, 在进行耕地遥感制图时, 选择高的空间分辨率固然可以提高数据精度, 选择合适的分类方法也同样重要. 同时, 地形复杂的西北和西南区域光谱混合现象严重, 给制图带来更多的不确定性, 导致数据间的一致性较差, 这将是以后耕地制图需要重点关注的区域.

关键词 地表覆盖, 耕地, 耕地数量, 空间位置精度, 空间一致性

1 引言

地表覆盖是地球表面各种物质类型及其自然属性与特征的综合体(陈军等, 2011). 随着卫星传感器的不断发展和计算机制图技术的进步, 遥感技术已经成为全球或区域地表覆盖空间分布信息获取的重要手段(Giri等, 2013; Ban等, 2015; Chen等, 2013). 最早, 美国马里兰大学和美国地质调查局以AVHRR

中文引用格式: 陆苗, 吴文斌, 张莉, 廖安平, 彭舒, 唐华俊. 2016. 不同耕地数据集在中国区域的对比研究. 中国科学: 地球科学, 46: 1459–1471, doi: 10.1360/N072015-00274
英文引用格式: Lu M, Wu W B, Zhang L, Liao A P, Peng S, Tang H J. 2016. A comparative analysis of five global cropland datasets in China. Science China Earth Sciences, 59: 2307−2317, doi: 10.1007/s11430-016-5327-3

© 2016《中国科学》杂志社 www.scichina.com

(Advanced Very High Resolution Radiometer)影像数据为基础, 分别生产了20世纪90年代初1km空间分辨率的全球地表覆盖数据, 即UMD(University of Maryland)数据集和IGBP-DISCover(International Geosphere Biosphere Programme Data and Information Systems Cover)数据集(Hansen等, 2000; Loveland等, 2000); 之后, 美国波士顿大学研制了以MODIS (MODerate Resolution Imaging Spectroradiometer)影像为基础的全球土地覆盖数据(Friedl等, 2002), 欧盟联合研究中心基于SPOT4/VEGETATION数据研制了GLC2000(Global Land Cover 2000)全球地表覆盖数据(Bartholome和Belward, 2005), 这两套数据集的空间分辨率也是1km, 数据来源为2000年左右. 进入21世纪以来, 国际社会将研制更高空间分辨率的全球地表覆盖数据提上议事日程, 全球地表覆盖遥感数据产品的空间分辨率不断提高(陈军等, 2011). 美国波士顿大学利用2000~2012年的MODIS数据研制了500m空间分辨率的新一期全球地表覆盖数据MODIS Collection 5(Friedl等, 2010), 欧洲空间局基于MERIS(Medium Resolution Imaging Spectromete)反射率数据先后研制了2005年和2009年的全球300m空间分辨率的地表覆盖数据GlobCover(Bicheron等, 2008; Bontemps等, 2011). 在国家科技计划支持下, 中国科学家在世界上率先研制了空间分辨率为30m的2010年全球地表覆盖遥感数据产品, 即清华大学的FROM-GLC(Finer Resolution Observation and Monitoring-Global Land Cover)数据集和国家基础地理信息中心的GlobeLand30数据集(Gong等, 2013; 曹鑫等, 2014; 廖安平等, 2014; Chen等, 2015).

卫星传感器数据、空间分辨率、分类系统与制图技术的不同, 使得以上全球地表覆盖数据集不可避免的存在差异, 开展对比分析、弄清这些数据产品之间的异同对数据生成者和数据使用者具有重要意义(Congalton等, 2014; Wang和Liu, 2014). 已有部分国内外学者在全球、区域或国家尺度上对这些全球地表覆盖数据集的一致性进行了比较研究(Hansen和Reed, 2000; Giri等, 2005; Jung等, 2006; Herold等, 2008; Tchuenté等, 2011; 吴文斌等, 2009; 宋宏利和张晓楠, 2012; 李蓓蓓等, 2010; 杨永可等, 2014). 但分析发现, 这些研究多针对2000年前后1km空间分辨率的UMD、IGBP-DISCover、GLC2000和MODIS数据集, 以及2005年300m分辨率的Globcover数据集, 而缺乏对新近2010年左右的全球地表覆盖数据产品的对比研究, 尤其针对30m空间分辨率的GlobeLand30和FROM-GLC数据集的类似研究未见报道.

作为重要的地表覆盖类型, 耕地是人类社会赖以生存和发展的基础, 其数量和空间分布格局直接关乎人类粮食安全、全球环境变化、陆地生态系统地球物理和地球化学循环过程(Fritz等, 2013; 唐华俊等, 2015). 因此, 本研究选择耕地为研究对象, 在中国区域, 对2010基准年的4套全球地表覆盖数据集GlobeLand30、FROM-GLC、GlobCover和MODIS Collection 5的耕地数据, 以及1套全球耕地数据集MODIS Cropland进行对比分析研究. 在耕地数量上, 以耕地统计数据为基准, 分析上述5套耕地数据集计算的耕地面积比例和统计数据的差异; 在空间位置上, 基于检验样本对5套耕地数据集的精度进行验证, 同时结合高程和坡度数据, 分析5套数据的空间一致性. 该研究结果有助于评估和对比不同耕地数据集在中国区域的精度, 揭示不同耕地数据集的数量和空间差异特征, 一方面为数据使用者在中国区域选择适宜耕地数据提供依据, 另一方面也为数据生产者完善数据和未来遥感制图提供新的方向.

2 数据来源与预处理

2.1 数据来源

表1列出了2010基准年5套数据集的基本信息. GlobeLand30数据集利用美国Landsat卫星和中国环境卫星(HJ)的30m多光谱影像数据, 采用新型的POK(Pixel-Object-Knowledge)分层分类方法, 将全球地表覆盖划分10大类(Chen等, 2015). FROM-GLC数据集也使用了30m分辨率的Landsat数据, 在计算机自动分类基础上, 结合MODIS时间序列、DEM和土壤水分等多源数据对分类结果进行改进(Yu等, 2013). 2009年的GlobCover数据集按照联合国粮食与农业组织倡导的LCCS(Land Cover Classification System)分类体系, 在全球地理分区基础上, 基于影像的时间和光谱特征, 采用监督分类和非监督分类相结合的方法将地表覆盖分为22个类型(Bontemps等, 2011). MODIS Collection 5数据集遵循了IGBP土

地覆盖分类系统, 基于500m的MODIS时间序列数据, 利用决策树分类方法, 将全球地表覆盖分成17个类别(Friedl等, 2010). 此外, 美国南达科他州立大学利用2000~2008年250m的MODIS时间序列数据, 采用决策树分类方法并结合统计数据, 生成了全球耕地分布数据, 即MODIS Cropland数据集(Pittman等, 2010).

耕地统计数据来自2010年《中国区域经济统计年鉴》, 该数据来源于国土资源部, 统计了中国各省、市和自治区(不含港、澳、台地区)的耕地面积及其占全国的比重(盛来运和严建辉, 2010). 同时, 选用SRTM(Shuttle Radar Topography Mission)90m分辨率的DEM(Digital Elevation Model)数据, 并在此数据基础上计算坡度数据.

2.2　数据预处理

5套数据集的坐标系统、空间分辨率和分类体系不同, 为方便对比分析, 需要进行数据预处理和转换. 首先, 利用中国边界矢量数据裁切5套全球数据产品, 得到中国区域数据集. 5套数据集在生产过程中都有严格的几何精度控制策略, 如基于Landsat影像的GlobeLand30几何精度在25m以内(Rengarajan等, 2015), GlobCover的几何精度在150m以内(Bontemps等, 2011), 相比之下, GlobeLand30的几何精度最高. 为了降低数据间几何精度差异的影响, 以GlobeLand30数据的参考椭球为基准, 将其他4套数据集统一到WGS-84(World Geodetic System-1984)地理坐标系统. 其次, MODIS Collection 5的空间分辨率最低, 为500m, 为便于比较, 将其他4套数据集采用最邻近方法重采样到相同的500m空间分辨率. 最后, 本研究主要对耕地数据进行对比分析, 需要对5类数据集的分类体系进行合并处理, 重新划分为耕地和非耕地, 保留耕地类别信息, 而剔除其余土地覆盖类别的影响. 以联合国粮农组织耕地定义为参考, 选取各数据集中的耕地类型进行合并, 如表1所示. GlobeLand30和MODIS Cropland仅包括了1个耕地类型, FROM-GLC包括了耕地和裸耕地2个类型, GlobCover数据集包括有灌溉耕地、雨养耕地、耕地(50~70%)/植被(20~50%)混合和植被(50~70%)/耕地(20~50%)混合等4个类型, MODIS Collection 5数据集包括耕地、耕地/自然植被混合类型等2个类型. 图1是从5套数据集中提取的中国耕地空间分布数据. 同时, 将SRTM-DEM数据利用中国边界矢量数据裁切, 并重采样到500m.

3　对比方法

本研究从耕地数量和空间位置两个方面在中国区域对5套耕地数据集进行对比分析. 在耕地数量方面, 分析5套产品在省级和区域尺度耕地面积比例的不同, 以及和统计数据的差异. 在空间位置方面, 基于检验数据评估各套数据集的耕地空间分布精度, 并进行5套数据集的空间一致性分析.

3.1　耕地数量对比

以中国省级行政边界矢量图为基础, 分别从5套耕地数据集中计算不同省份的耕地面积, 汇总全国耕地总面积, 并计算各省耕地面积占全国耕地总面积的比例. 同时, 按照地理位置, 将中国分为东北、华北、东南、华中、西南和西北6个区域, 在5套数据各省耕地面积比例计算的基础上, 计算各个区域每套产品的耕地面积比例. 然后, 以2010年《中国区域经济统计年鉴》的耕地面积比例为基准, 计算5套产品各省和区域耕地面积比例与统计数据比例的偏差, 并从总体上计算5套产品和统计数据的均方根误差(root mean square error, RMSE), 以反映5套产品的耕地数据和统计数据的离散度, 偏差Δx_i和均方根误差RMSE的计算方法如下所示:

$$\Delta x_i = x_i - y_i, \tag{1}$$

表 1　5 套 2010 基准年全球地表覆盖(或耕地)数据集的基本信息

数据集名称	卫星和传感器	空间分辨率(m)	耕地相关类型
GlobeLand30	Landsat/TM和HJ/CCD	30	耕地
FROM-GLC	Landsat/TM	30	耕地, 裸耕地
GlobCover	ENVISAT/MERIS	300	灌溉耕地, 雨养耕地, 耕地(50~70%)/植被(20~50%)混合, 植被(50~70%)/耕地(20~50%)混合
MODIS Collection 5	TERRA/MODIS	500	耕地, 耕地/自然植被混合
MODIS Cropland	TERRA/MODIS	250	耕地

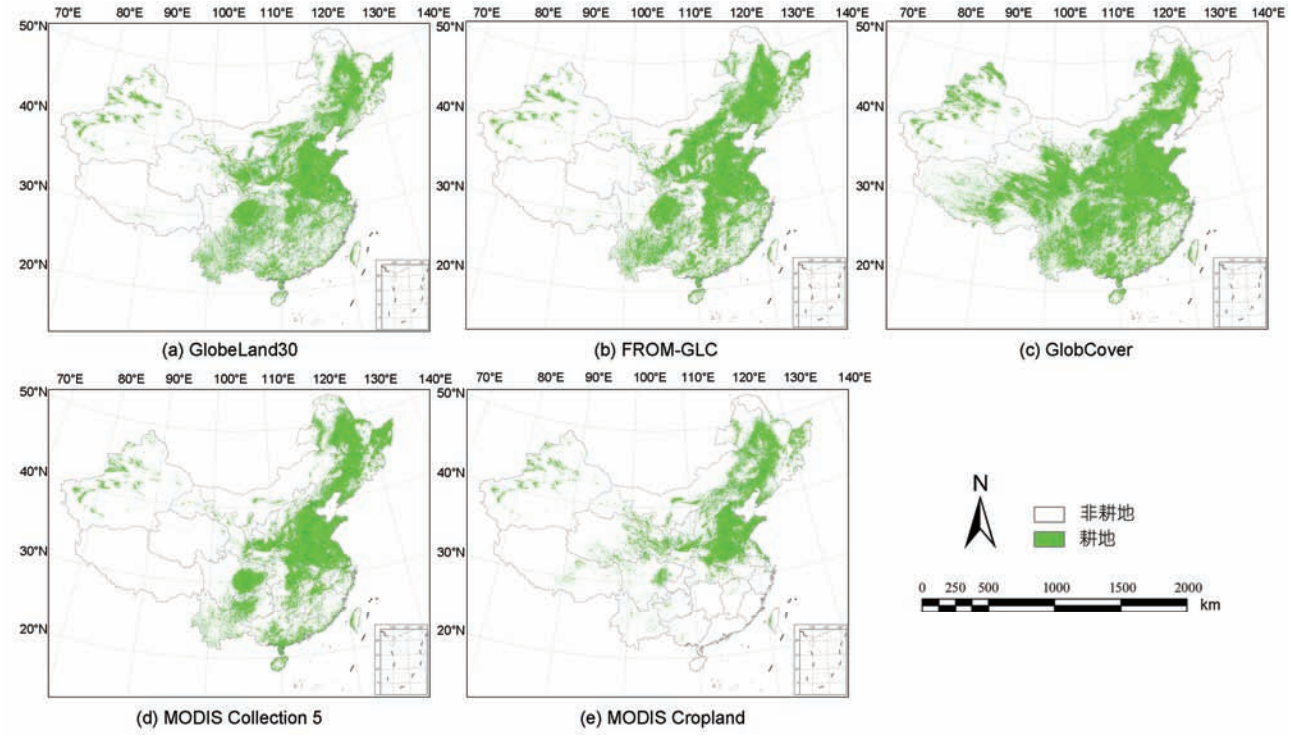


图 1　从 5 套全球耕地数据集中提取的中国耕地空间分布

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-y_i)^2}{n}}, \tag{2}$$

式中, x_i是耕地数据集计算的i省耕地面积比例, y_i是i省的耕地统计面积比例, n是省、市和自治区的总数.

最后, 为反映5套数据集的耕地数量和统计数据的吻合度, 对5套数据的耕地面积比例和统计数据的耕地面积比例进行相关性分析, 相关系数计算公式为

$$R=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}}, \tag{3}$$

式中, $\bar{x}$ 是耕地数据集计算的各省耕地面积比例的平均值, $\bar{y}$ 是耕地统计面积比例的平均值. 相关系数越大, 表明和统计数据的吻合度越高.

3.2　空间位置对比

仅仅比较遥感数据产品的数量特征具有相当的局限性, 因为遥感制图可以相对容易地描述区域地表覆盖类型的数量特征, 但其空间位置制图可能并不准确. 因此, 数量特征比较之外的空间位置比较十分必要(Wu等, 2008). 传统的空间位置比较通常以某种数据集为参照, 通过分析目标数据产品和参照数据产品间的一致性, 进行空间位置比较(李蓓蓓等, 2010; Pflugmacher等, 2011). 该方法依赖于参考数据的准确性, 然而通常参考数据本身也可能存在错误和不足, 尤其是在大范围制图中(Herold等, 2008; 杨永可等, 2014). 本文以检验样本为基准比较5套数据的空间位置精度, 并分析数据间的空间分布一致性.

空间位置精度利用相同的耕地检验样本对5套耕地数据集精度进行估算. 样本的数量、质量和分布对空间位置精度评价影响重大(Foody, 2010). 本研究中的耕地样本数据一部分来自于FROM-GLC数据生产中的检验样本(Gong等, 2013; 俞乐等, 2014), Gong等(2013)将全球陆地地表分为约7000个等面积的六边形, 在每个六边形内随机抽取5个检验样本, 采用目视解译方法确定样本的类别属性. 该采样方法保证了检验样本分布的均匀性和客观性, 但是在中国区域的耕地样本数据有443个、非耕地样本1687个, 难

以满足耕地精度评价的数量要求. 因此, 为进一步补充耕地样本, 基于5套耕地数据空间叠置的一致性分布图, 在不同的一致性层次上随机抽样0.001%的像元作为补充样本, 利用Google Earth高分影像解译确定样本的类别属性, 采集了耕地样本2716个、非耕地样本858个. 将上述两种来源的样本进行组合得到了图2中的检验样本空间分布, 其中耕地样本数据为3159个、非耕地样本数据为2545个. 在检验样本的基础上, 采用混淆矩阵的方法, 计算5套数据的总体精度、Kappa系数、漏分率和错分率.

空间一致性是对不同土地覆盖数据集在相同空间位置上地物类别一致性的描述, 用于不同数据集之间的对比和相互验证(杨永可等, 2014). 本研究将5套耕地数据集进行空间叠加得到耕地空间一致性分布图, 图中的像元值描述了在相同地理位置上5套耕地数据集的一致性状况. 如像元值5表示在该像元上5套数据产品完全一致; 像元值4表示在该像元上有4套数据一致, 像元值越大, 数据产品之间的一致性越高. 根据像元值将其划分为完全一致性、高度一致性、中度一致性、低度一致性和不一致性等5个级别, 可以在像元尺度上描述5套全球耕地数据集的空间一致性. 在区域尺度, 计算5套耕地数据集的空间一致性比例, 分析各个区域的一致性差异. 此外, 耕地数量特征虽然受人类活动影响大, 但其空间分布往往受地形、降水等生态环境因子的制约, 这给耕地遥感制图带来一定的挑战. 为此, 本研究进一步分析了5套耕地数据集空间一致性随高程和坡度的变化情况. 根据高程将地貌形态分为平原(<20m)、丘陵(20~200m)、低山(200~500m)、中山(500~1500m)和高山(>1500m)五种类型(柴宗新, 1989), 分析在每种地貌类型中5套耕地数据的一致性分布. 同时, 按照《土地利用现状调查技术规程》, 将全国坡度数据分为五个等级, 即≤2°、2°~6°、6°~15°、15°~25°和>25°, 分析在每个坡度区间5套耕地数据的一致性分布特征.

4　结果与分析

4.1　省级和区域耕地数量对比

表2是从5套耕地数据集中计算得到的不同省份和区域耕地面积占全国耕地总面积的比例, 及其和

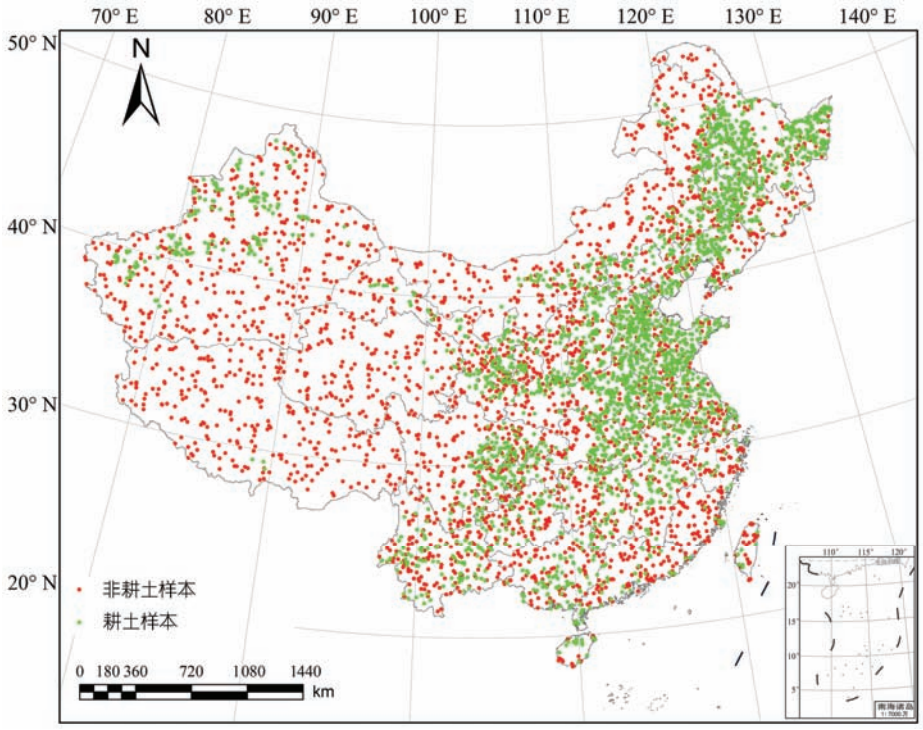


图 2　检验样本空间分布

表 2　5 套耕地数据集计算的不同省份和区域的耕地面积比例及其和统计数据的偏差

区域	省、市、自治区	Globeland30		FROM-GLC		GlobCover 2009		MODIS collection 5		MODIS cropland		耕地统计数据(%)
		耕地比例(%)	与统计数据的偏差(%)	耕地比例(%)	与统计数据的偏差(%)	耕地比例(%)	与统计数据的偏差(%)	耕地比例(%)	与统计数据的偏差(%)	耕地比例(%)	与统计数据的偏差(%)	
东北区域	内蒙古	6.71	0.84	12.31	6.44	6.34	0.47	5.72	−0.15	11.93	6.05	5.87
	黑龙江	9.15	−0.57	8.80	−0.92	3.30	−6.42	13.82	4.10	12.31	2.59	9.72
	吉林	4.35	−0.20	4.37	−0.18	1.47	−3.08	5.18	0.64	5.67	1.12	4.55
	辽宁	3.50	0.15	3.44	0.09	2.17	−1.19	5.19	1.84	6.48	3.12	3.36
	汇总	23.72	0.23	28.92	5.42	13.28	−10.21	29.92	6.42	36.38	12.89	23.49
华北区域	北京	0.27	0.08	0.26	0.07	0.29	0.10	0.28	0.09	0.33	0.14	0.19
	天津	0.38	0.01	0.34	−0.02	0.30	−0.06	0.44	0.08	0.51	0.15	0.36
	河北	4.92	−0.27	5.24	0.05	5.46	0.27	5.01	−0.18	8.68	3.49	5.19
	山西	3.42	0.09	3.79	0.46	4.51	1.18	2.89	−0.44	3.13	−0.20	3.33
	山东	6.10	−0.08	4.94	−1.23	5.38	−0.79	7.26	1.08	10.93	4.76	6.17
	河南	5.36	−1.15	5.46	−1.06	5.78	−0.73	7.14	0.62	9.76	3.25	6.51
	汇总	20.44	−1.32	20.04	−1.72	21.72	−0.04	23.02	1.26	33.34	11.58	21.76
东南区域	上海	0.17	−0.03	0.17	−0.03	0.20	0.00	0.22	0.02	0.05	−0.15	0.20
	江苏	3.60	−0.32	3.22	−0.70	3.31	−0.60	4.33	0.42	2.40	−1.52	3.91
	浙江	1.61	0.04	1.46	−0.12	1.39	−0.19	1.14	−0.44	0.07	−1.50	1.58
	福建	1.21	0.11	1.01	−0.08	0.52	−0.58	0.65	−0.44	0.02	−1.07	1.09
	广东	2.36	0.03	2.01	−0.32	1.48	−0.84	2.79	0.46	0.04	−2.29	2.33
	海南	0.39	−0.21	0.50	−0.10	0.33	−0.27	0.72	0.12	0.03	−0.57	0.60
	汇总	9.34	−0.37	8.36	−1.35	7.23	−2.48	9.84	0.13	2.60	−7.11	9.71
华中区域	安徽	4.12	−0.59	3.99	−0.72	3.94	−0.77	4.97	0.26	3.73	−0.98	4.71
	江西	2.32	−0.01	2.44	0.12	2.16	−0.17	2.22	−0.10	0.16	−2.16	2.32
	湖北	4.04	0.21	3.74	−0.09	4.39	0.55	4.16	0.32	1.81	−2.02	3.83
	湖南	3.24	0.13	4.26	1.15	3.87	0.76	2.65	−0.46	0.06	−3.05	3.11
	汇总	13.72	−0.26	14.43	0.45	14.35	0.38	14.00	0.02	5.76	−8.21	13.98
西南区域	广西	3.31	−0.16	2.73	−0.74	2.48	−0.98	2.37	−1.10	0.14	−3.33	3.47
	重庆	2.01	0.18	1.44	−0.39	1.64	−0.20	1.68	−0.16	0.05	−1.79	1.84
	四川	5.82	0.93	5.41	0.52	7.65	2.77	5.23	0.35	3.56	−1.33	4.89
	贵州	2.95	−0.74	2.72	−0.96	3.06	−0.62	2.41	−1.27	0.32	−3.37	3.69
	云南	5.23	0.24	4.57	−0.42	4.76	−0.23	3.13	−1.86	0.68	−4.31	4.99
	汇总	19.31	0.45	16.88	−1.99	19.60	0.73	14.81	−4.05	4.74	−14.12	18.86
西北区域	西藏	0.22	−0.08	0.22	−0.08	3.88	3.58	0.11	−0.19	1.26	0.96	0.30
	陕西	3.10	−0.22	3.97	0.64	4.54	1.21	2.79	−0.54	2.75	−0.58	3.33
	甘肃	4.16	0.34	2.74	−1.09	4.46	0.64	1.75	−2.08	4.49	0.66	3.83
	青海	0.57	0.13	0.19	−0.26	5.54	5.09	0.11	−0.34	3.27	2.82	0.45
	宁夏	1.09	0.18	0.84	−0.07	0.40	−0.51	0.26	−0.65	0.70	−0.22	0.91
	新疆	4.33	0.94	3.41	0.03	5.00	1.61	3.40	0.01	4.71	1.32	3.39
	汇总	13.47	1.28	11.38	−0.82	23.82	11.62	8.42	−3.78	17.18	4.98	12.20

耕地统计数据的偏差. 总体上看, 在多数省份5套产品计算的耕地面积比例都有所差异, 差异较大的省份有青海、甘肃和云南等省, 相比较而言, 安徽、山西和河北等省的耕地面积比例数据差异较小. 和统计数据相比, GlobeLand30计算的耕地面积比例和统计数据比较接近, 其他4套产品都有不同程度的高估和低估. FROM-GLC在东北区域和统计数据相差较大, 耕地面积高估5.42%, 华北、东南和西南区域存在不同程度的低估, 和统计数据的偏差分别是1.72%、1.35%和1.99%, 华中和西北区域和统计数据比较接近. GlobCover在东北区域明显低估了耕地面积, 和统计数据的偏差是10.21%, 而高估的西北区域耕地面积达到11.62%. MODIS Collection 5也高估了东北区域耕地面积约6.42%, 在华北、华南和华中3个区域与统计数据的一致性较好, 同时不同程度地低估了西南和西北地区耕地面积比例. MODIS Cropland和统计数据差异较大, 计算的东北、华北和西北耕地面积分别比统计数据高出12.89%、11.58%和4.98%, 而

在华南、华中和西南区域存在较大的耕地面积低估, 偏差是7.11%、8.21%和14.12%.

从均方根误差和相关性分析两方面, 评估5套数据集的耕地数量和统计数据的总体离散度和吻合度. 图3显示了各耕地数据集的面积比例和统计数据的均方根误差和相关系数, 图中虚线表示1∶1线, 实线是各数据集的耕地比例和统计数据拟合的趋势线. GlobeLand30、FROM-GLC、GlobCover、MODIS Collection 5、MODIS Cropland和统计数据的RMSE分别是0.43%、1.29%、1.89%、1.07%和2.46%, RMSE的值越大, 表示和统计数据的离散度越高, 其中GlobeLand30的RMSE最小, 和统计数据的离散度最低, 从图3a可看出, GlobeLand30的数据点大多数分布在1∶1线上及其附近; MODIS Collection 5和统计数据的离散度较低, 如图3d所示, 大部分点在1∶1线及其附近, 但也有少数点偏离度较大; MODIS Cropland的RMSE值最高, 和统计数据的偏离度最大, 如图3e所示. GlobeLand30、FROM-GLC、GlobCover、MODIS Collection 5、MODIS Cropland和统计数据的相关系数分别是0.96、0.70、0.29、0.82和0.25, 以相关系数为判定系数, 相关系数越大, 表示和统计数据的吻合度越好. GlobeLand30计算的各省耕地面积比例数据和统计数据的吻合度最好, 其次是MODIS Collection 5, MODIS Cropland的吻合度最差. 通过离散度和吻合度的分析可看出, GlobeLand30和统计数据的离散度最低, 吻合度最高, 因此耕地数量较为准确. 同时, 500m空间分辨率的MODIS Collection 5数据集和统计数据的相关系数高于30m空间分辨率的FROM-GLC数据集、300m分辨率的GlobCover和250m分辨率的MODIS Cropland, 这表明技术方法适宜时, 低空间分辨率数据也可以取得不错的分类效果.

4.2　空间位置精度比较

表3列出了5套耕地数据集的空间位置精度评价结果. 其中, GlobeLand30空间位置精度最高, 该数据的总体精度是79.61%, Kappa系数为0.58. FROM-GLC和MODIS Collection 5的总体精度和Kappa系数较高, 分别为76.23%、0.52和73.49%、0.47. 相对而言, GlobCover和MODIS Cropland的空间位置精度较低, 它们的总体精度和Kappa系数分别是70.13%、0.39和67.99%、0.38. 具体来看, GlobCover的耕地错分率最高, 即28.97%, 从表2可看出GlobCover在西北区域耕地面积明显高估, 因此可推断该数据集耕地错分率较高的原因可能是在西北区域将其他地表覆盖类型错分为耕地. MODIS Cropland的耕地漏分率最高, 达到49.86%, 从与耕地统计数据的对比中可看出(表2),

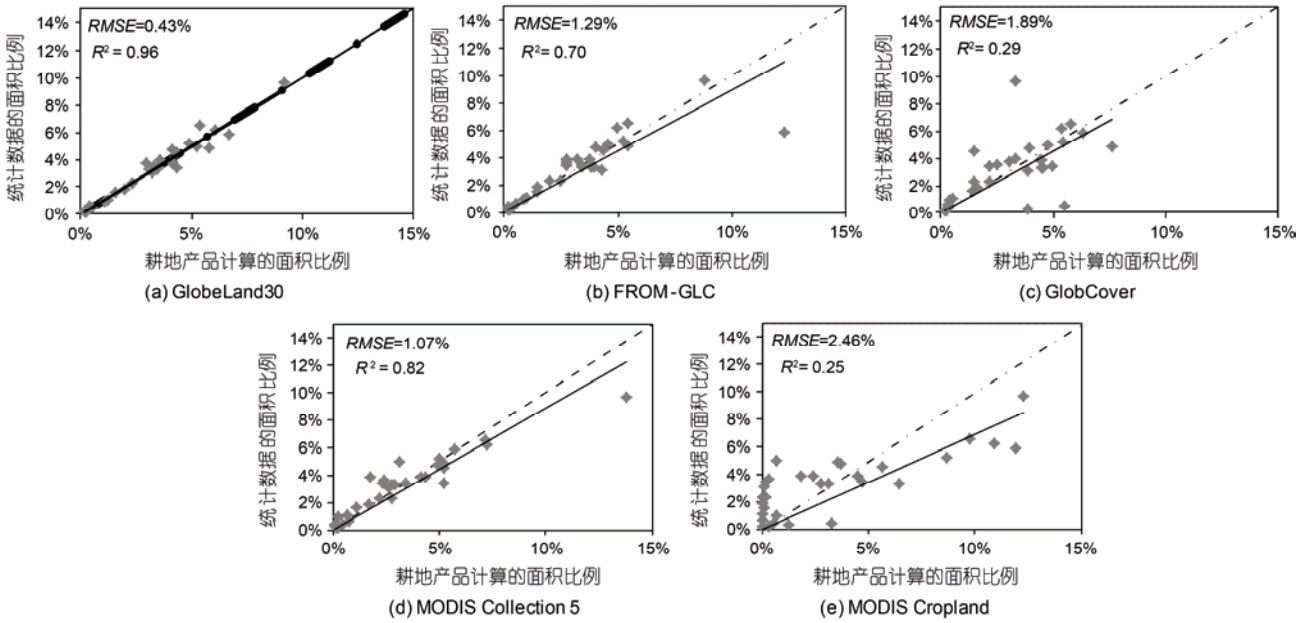


图 3　5 套耕地数据集计算的省级耕地面积比例和统计数据的离散度和吻合度分析

该数据集在东南、华中和西南三个区域耕地面积比例都被明显低估, 因此可判断MODIS Cropland在中国南方区域存在大面积的耕地漏分.

图4显示了5套耕地数据在中国东北、华北、东南、华中、西南和西北6个区域的总体精度. 在东北区域, FROM-GLC的总体精度最高, 即79.40%, GlobCover在该区域耕地漏分情况比较严重, 因此总体精度最低, 即64.25%; 在华北区域GlobCover的总体精度是78.90%, 高于其他4套产品; 在华南区域, MODIS Collection 5的总体精度最高, 即68.72%; 在华中、西南和西北3个区域GlobeLand30的总体精度最高, 分别是78.73%、66.41%和86.62%, 相比而言, MODIS Cropland在这些区域的精度较低, 分别是48.17%、51.51%和77.72%. 总体来看, 在6个区域中, 5套产品在西北和华中区域的总体精度普遍较高, 在西南区域的总体精度相对较低; 在5套产品中, GlobeLand30在各个区域都有较高的总体精度, MODIS Cropland在各个区域的总体精度偏低.

4.3 空间一致性比较

图5显示了5套耕地数据集的空间一致性分布情况. 从总体上看, 在东北平原、黄淮海平原、长江流域和渭河平原等粮食主产区5套耕地数据的一致性较好, 这些区域耕地分布集中, 连续成片, 地理景观比较单一, 因此各套产品都能在这些区域比较准确地提取出耕地. 从粮食主产区到农牧交错带, 空间一致性以中度一致和低度一致为主. 到草地牧区及其周边, 一致性最差. 以东北平原粮食主产区为例, 在黑龙江、吉林和辽宁的粮食主产区5套数据空间一致性较好; 到北方农牧交错带, 即内蒙古高原的南缘和长城沿线, 5套产品以中度一致性和低度一致性为主; 在内蒙古高原的草原牧区, 5套产品的一致性最差.

在图5的基础上, 从区域尺度定量分析5套耕地产品的空间一致性分布情况(图6). 在6个区域中, 只

表 3　5 套耕地数据集的空间位置精度评价结果

	GlobeLand30	FROM-GLC	GlobCover	MODIS Collection 5	MODIS Cropland
总体精度(%)	79.61	76.23	70.13	73.49	67.99
Kappa系数	0.58	0.52	0.39	0.47	0.38
耕地错分率(%)	19.70	21.66	28.97	20.79	13.68
耕地漏分率(%)	21.33	21.11	22.22	29.31	49.86

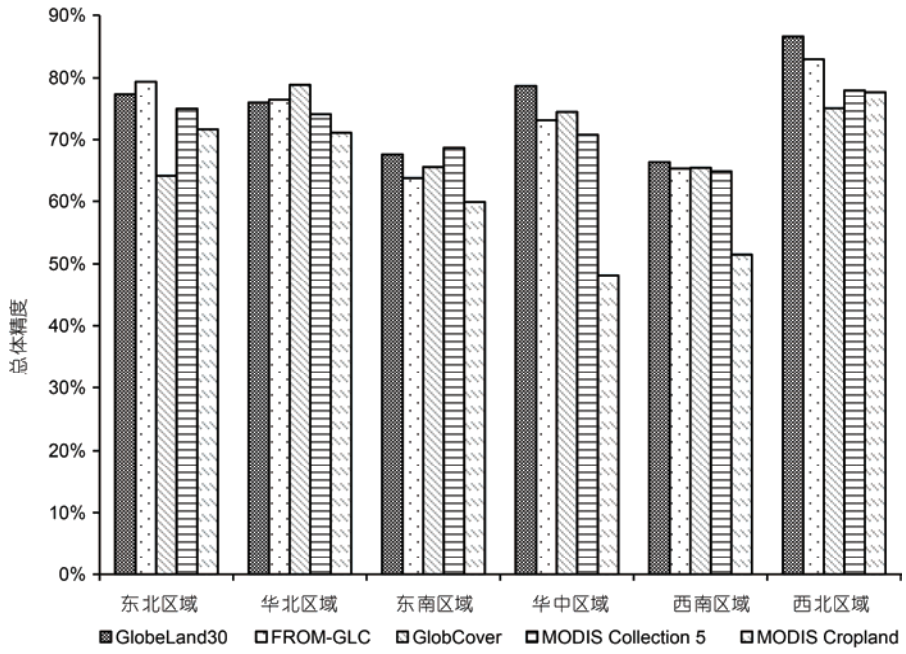


图 4　5 套耕地数据在 6 个区域的精度分析

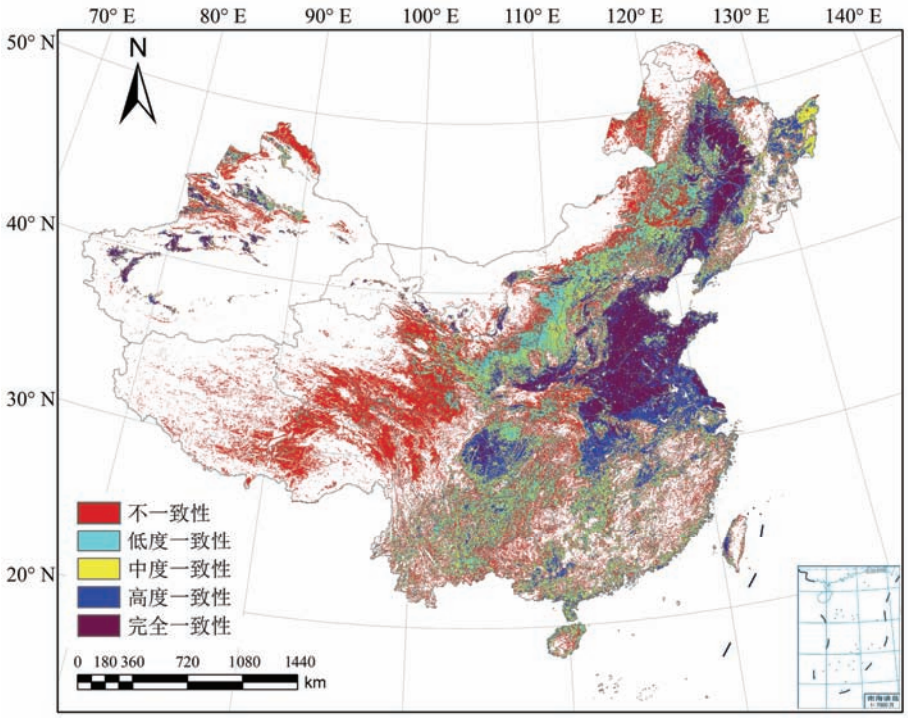


图 5　5 套耕地数据集的空间分布一致性

有华北区域的耕地面积比例随一致性的增加而增加, 其中完全一致的耕地面积比例高达39.63%, 因此可看在华北区域5套产品的一致性最好. 与华北区域相反, 在西北和西南区域耕地面积比例随着一致性的增加而减少, 一致性较差. 西北区域的不一致性比例最高, 达到61.79%, 而完全一致区域仅有5.75%, 西南区域的不一致耕地面积比例是41.27%, 完全一致性仅占2.22%. 西南和西北区域地形复杂, 地理景观破碎, 耕地和其他地物混合比较严重, 在遥感影像上耕地的光谱特征可分离度较差, 并且混合像元较多, 因此这些区域遥感分类难度较大, 使得5套耕地产品间的一致性较差.

图7显示了5套耕地数据集空间分布一致性随高程的变化情况. 高程低于20m的平原地区和高程在20~200m间的丘陵地区, 高度一致性和完全一致性所占比例较大, 说明5套数据集在这些区域的一致性较好. 平原和丘陵区域主要分布在华北和华中的大部分省份, 如东北的黑龙江、辽宁和吉林, 以及华南的江苏、上海和浙江等, 这些区域是我国主要的粮食生产区, 耕地面积广阔, 各种产品在这些区域都能比较准确地提取出耕地范围, 如表2所示. 从200~500m的低山区域开始, 数据一致性随着高程的增加逐渐降低. 500~1500m的中山区域主要分布在大兴安岭以西的内蒙古高原、新疆的塔里木盆地、黄土高原以及云贵高原, 除内蒙古高原和塔里木盆地外, 其他地区多山, 且地形比较破碎, 耕地提取难度较大, 因此该区域的不一致性较高, 即40.16%. 高程高于1500m的高山主要分布在西北区域的青藏高原, 不一致性高达70.76%.

坡度是耕地利用的主要影响因素之一, 图8统计了不同坡度区间的空间一致性分布情况. 总体上看, 随着坡度的增加, 5种数据集间的一致性逐渐降低. 坡度小于2°的区域主要分布在我国各大平原和盆地, 地势平缓, 适合农业耕作, 地理景观相对比较简单, 因此耕地提取的一致性较好, 高度一致和完全一致区域分别占24.09%和22.55%. 在第2坡度区间, 即2°~6°, 不一致性上升到41.11%, 完全一致性下降到3.66%. 15°~25°和坡度大于25°的区域主要分布在青藏高原的山区, 5套数据集不一致性在这些区域所占比例较高, 分别达到77.09%和81.92%.

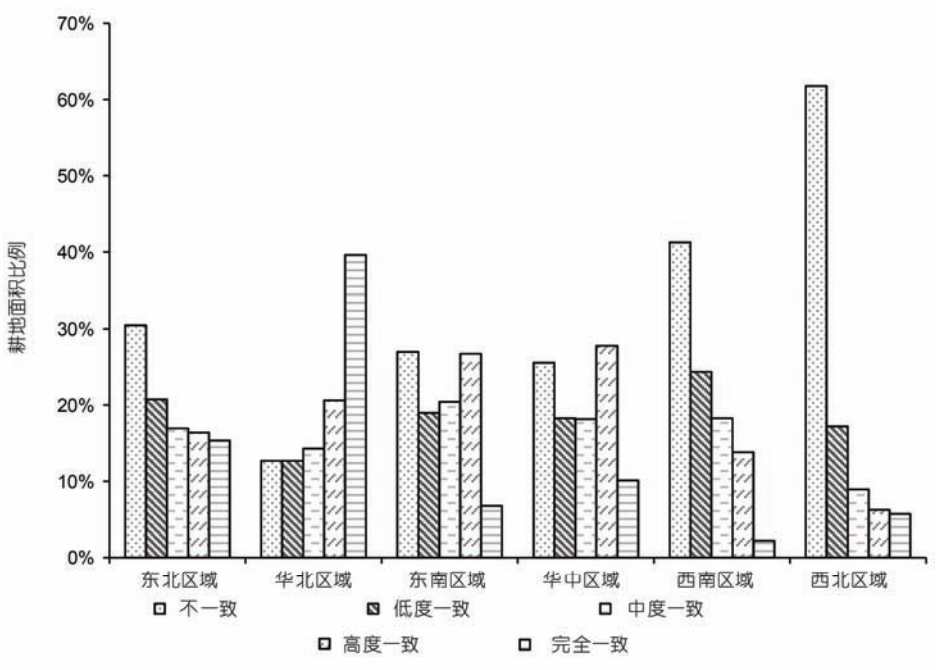


图 6　不同区域的耕地数据空间分布一致性统计

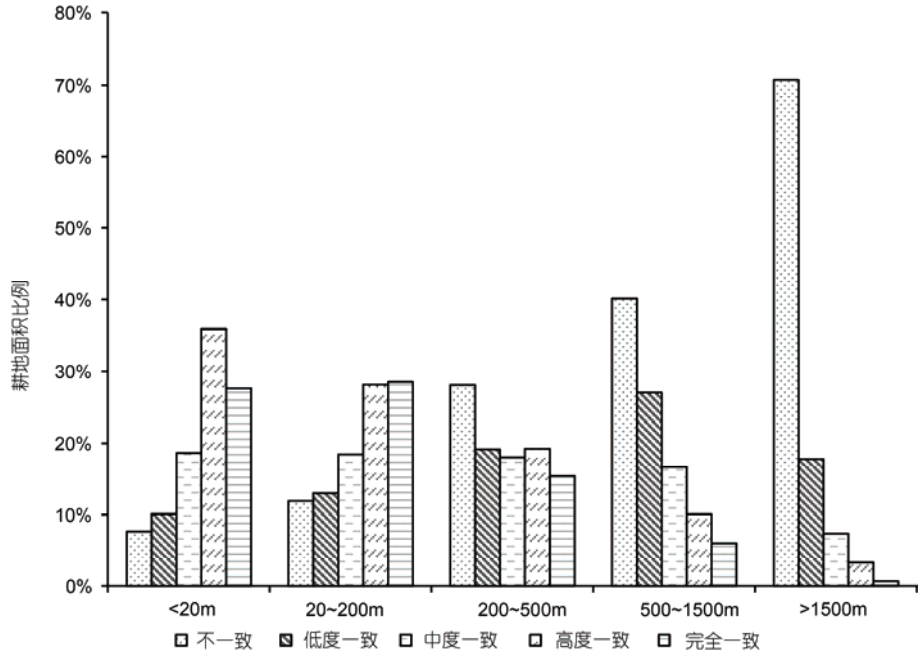


图 7　不同高程区间的一致性分布

5　结论

本文以中国的耕地类型为研究对象，对2010基准年GlobeLand30、FROM-GLC、GlobCover、MODIS Collection 5和MODIS Cropland这5套耕地制图产品进行了对比分析. 结果表明，无论是耕地数量计算，还是耕地空间位置描述，GlobeLand30耕地数据集总体上要优于其他4套耕地数据产品. 单从耕地数量对比

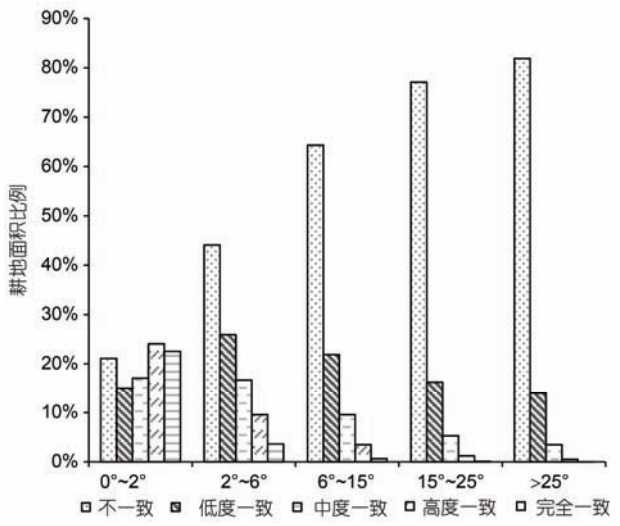


图 8　不同坡度区间的一致性分布

看, 在多数省份5套产品计算的耕地面积比例都有所差异; 和统计数据相比, GlobeLand30计算的各省和区域耕地面积比例与耕地统计数据吻合度最高, MODIS Collection 5和FROM-GLC次之, MODIS Cropland和GlobCover的吻合度较低. 在空间位置精度方面, GlobeLand30耕地数据集精度最高, 其次是FROM-GLC和MODIS Collection 5, GlobCover和MODIS Cropland的精度最低. 5套数据的空间分布一致性从农业主产区到草原牧区呈现逐渐降低的趋势, 各个区域的一致性有所不同, 且受高程和坡度影响明显.

一般认为, 影像空间分辨率对地表覆盖遥感产品的制图精度影响较大(Hansen和Loveland, 2012; Ban等, 2015). 本研究中的GlobeLand30和FROM-GLC都利用了相同空间分辨率的30m数据源, 分类体系大体一致, 但其分类方法存在较大差异, 使得两类制图产品差异较大(Chen等, 2015; Gong等, 2013). FROM-GLC以计算机自动分类为主, 可以快速获取全球地表覆盖制图产品; GlobeLand30采用POK分类方法, 在计算机自动分类的基础上加入了人工检核和知识修改, 制图效率有所降低, 但精度明显提升. 同时, 本研究也发现在耕地数量方面, 空间分辨率最低的数据集MODIS Collection 5和统计数据的吻合度要高于数据集FROM-GLC, GlobCover和MODIS Cropland; 在空间位置精度方面, MODIS Collection 5在中国区域的总体精度和Kappa系数高于GlobCover和MODIS Cropland. 这表明在重点关注全球地表覆盖制图品的空间分辨率的同时, 也要重视适宜的分类技术方法, 若采用合适的分类方法, 较低空间分辨率数据也可以取得不错的制图结果.

本研究发现, 5套耕地数据产品在我国华北和东北地区的制图一致性最好, 这些区域是主要粮食主产区, 海拔较低, 坡度较缓, 耕地集中连片分布, 地物类型较为单一, 遥感影像可分性较好, 即便利用空间分辨率较低的数据也可以取得较好制图效果. 但是, 西北和西南地区海拔较高、坡度大, 复杂的地形因素使得地理景观破碎, 耕地和其他地物混合较为严重, 遥感光谱的可分离度较差, 造成区域分类结果差异大, 5套耕地产品之间的空间分布一致性较差. 如FROM-GLC在内蒙古将大量牧草地错分为耕地, MODIS Cropland在东南和西南区域大量漏分耕地, GlobCover在西南和西北区域无法很好区分耕地和其他植被类型. 可见, 这些耕地破碎、零散分布和混合像元严重区域将是耕地遥感制图将来需要重点关注的区域.

参考文献

曹鑫, 陈军, 陈利军, 廖安平, 孙芳蒂, 李阳, 李磊, 林忠辉, 庞治国, 陈晋, 何超英, 彭舒. 2014. 全球陆表水体空间格局与波动初步分析. 中国科学: 地球科学, 44: 1661–1670

柴宗新. 1989. 按相对高度划分地貌基本形态的建议. 北京: 测绘出版社

陈军, 陈晋, 宫鹏, 廖安平, 何超英. 2011. 全球地表覆盖高分辨率遥感制图. 地理信息世界, 2: 12–14

李蓓蓓, 方修琦, 叶瑜, 张学珍. 2010. 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. 中国科学: 地球科学, 40: 1048–1059

廖安平, 陈利军, 陈军, 何超英, 曹鑫, 陈晋, 彭舒, 孙芳蒂, 宫鹏. 2014. 全球陆地水表高分辨率遥感制图. 中国科学: 地球科学, 44: 1634–1645

盛来运, 严建辉. 2010. 中国区域经济统计年鉴2010. 北京: 中国统计出版社. 86

宋宏利, 张晓楠. 2012. 中国区域多源土地覆被遥感产品精度分析与验证. 农业工程学报, 28: 207–214

唐华俊, 吴文斌, 余强毅, 夏天, 杨鹏, 李正国. 2015. 农业土地系统研究及其关键科学问题. 中国农业科学, 48: 900–910

吴文斌, 杨鹏, 张莉, 唐华俊, 周清波. 2009. 四类全球土地覆盖数据在中国区域的精度评价. 农业工程学报, 25: 167–173

杨永可, 肖鹏峰, 冯学智, 李海星, 常潇, 冯威丁. 2014. 大尺度土地覆盖数据集在中国及周边区域的精度评价. 遥感学报, 18: 453–475

俞乐, 王杰, 李雪草, 李丛丛, 赵圆圆, 宫鹏. 2014. 基于多源数据集成的多分辨率全球地表覆盖制图. 中国科学: 地球科学, 44: 1646–1660

Ban Y F, Gong P, Giri C. 2015. Global land cover mapping using Earth observation satellite data: Recent progresses and challenges. ISPRS-J Photogramm Remote Sens, 103: 1–6

Bartholome E, Belward A S. 2005. GLC2000: A new approach to global land cover mapping from Earth observation data. Int J Remote Sens, 26: 1959–1977

Bicheron P, Defourny P, Brockmann C, Schouten L, Vancutsem C, Huc M, Bontemps S, Leroy M, Achard F, Herold M, Ranera F, Arino O. 2008. GlobCover Products Description and Validation Report

Bontemps S, Defourny P, Bogaert E V, Arino O, Kalogirou V, Perez J R. 2011. GLOBCOVER 2009 Products Description and Validation Report

Chen J, Chen J, Liao A P, Cao X, Chen L J, Chen X H, He C Y, Han G, Peng S, Lu M, Zhang W, Tong X H, Mills J. 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. ISPRS-J Photogramm Remote Sens, 103: 7–27

Chen J, Lu M, Chen X H, Chen J, Chen L J. 2013. A spectral gradient difference based approach for land cover change detection. ISPRS-J Photogramm Remote Sens, 85: 1–12

Congalton R G, Gu J, Yadav K, Thenkabail P, Ozdogan M. 2014. Global land cover mapping: A review and uncertainty analysis. Remote Sens-Basel, 6: 12070–12093

Foody G M. 2010. Assessing the accuracy of land cover change with imperfect ground reference data. Remote Sens Environ, 114: 2271–2285

Friedl M A, Mciver D K, Hodges J C F, Zhang X Y, Muchoney D, Strahler A H, Woodcock C E, Gopal S, Schneider A, Cooper A, Baccini A, Gao F, Schaaf C. 2002. Global land cover mapping from MODIS: Algorithms and early results. Remote Sens Environ, 83: 287–302

Friedl M A, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X. 2010. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. Remote Sens Environ, 114: 168–182

Fritz S, See L, You L, Justice C, Becker-Reshef I, Bydekerke L, Cumani R, Defourny P, Erb K, Foley J, Gilliams S, Gong P, Hansen M, Hertel T, Herold M, Herrero M, Kayitakire F, Latham J, Leo O, McCallum I, Obersteiner M, RamanKutty N, Rocha J, Tang H, Thornton P, Vancutsem C, Velde M, Wood S, Woodcock C. 2013. The need for improved maps of global cropland. Eos Trans AGU, 94: 31–32

Giri C, Pengra B, Long J, Loveland T R. 2013. Next generation of global land cover characterization, mapping, and monitoring. Int J Appl Earth Obs, 25: 30–37

Giri C, Zhu Z, Reed B. 2005. A comparative analysis of the global land cover 2000 and MODIS land cover datasets. Remote Sens Environ, 94: 123–132

Gong P, Wang J, Yu, L, Zhao Y C, Zhao Y Y, Liang L, Niu Z G, Huang X M, Fu H H, Liu S, Li C C, Li X Y, Fu W, Liu C X, Yue X, Wang X Y, Cheng Q, Hu L Y, Yao W B, Zhang H, Zhu P, Zhao Z Y, Zhang H Y, Zheng Y M, Ji L Y, Zhang Y W, Chen H, Yan A, Guo J H, Liang Y, Wang L, Liu X J, Shi T T, Zhu M H, Chen Y L, Yang G W, Tang P, Xu B, Giri C, Clinton N, Zhu Z L, Chen J, Chen J. 2013. Finer resolution observation and monitoring of GLC: first mapping results with Landsat TM and ETM+ data. Int J Remote Sens, 34: 2607–2654

Hansen M C, Loveland T R. 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens Environ, 122: 66–74

Hansen M C, Reed B. 2000. A comparison of the IGBP DISCover and university of Maryland 1 km global land cover products. Int J Remote Sens, 21: 1365–1373

Hansen M C, Defries R S, Townshend J R G, Sohlberg R. 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. Int J Remote Sens, 21: 1331–1364

Herold M, Mayaux P, Woodcock C E, Baccini A, Schmullius C. 2008. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. Remote Sens Environ, 112: 2538–2556

Jung M, Henkel K, Herold M, Churkina G. 2006. Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sens Environ, 101: 534–553

Loveland T R, Reed B C, Brown J F, Ohlen D O, Zhu Z, Yang L, Merchant J W. 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. Int J Remote Sens, 21: 1303–1330

Pflugmacher D, Krankina O N, Cohen W B, Friedl M A, Sulla-Menashe D, Kennedy R E, Nelson P, Loboda T V, Kuemmerle T, Dyukarev E, Elsakov V, Kharuk V I. 2011. Comparison and assessment of coarse resolution land cover maps for northern Eurasia. Remote Sens Environ, 115: 3539–3553

Pittman K, Hansen M C, Becker-Reshef I, Potapov P V, Justice C O. 2010. Estimating global cropland extent with multi-year MODIS data. Remote Sens-Basel, 2: 1844–1863

Rengarajan R, Sampath A, Storey J, Choate M. 2015. Validation of geometric accuracy of global land survey (GLS) 2000 data. Photogramm Eng Rem S, 81: 131–141

Tchuenté A T K, Roujean J L, Jong S M D. 2011. Comparison and relative quality assessment of the GLC2000, GLOBCOVER,

MODIS and ECOCLIMAP land cover data sets at the African continental scale. Int J Appl Earth Obs, 13: 207–219

Wang Z, Liu L. 2014. Assessment of coarse-resolution land cover products using CASI hyperspectral data in an arid zone in northwestern China. Remote Sens-Basel, 6: 2864–2883

Wu W B, Shibasaki R, Yang P, Ongaro L, Zhou Q B, Tang H. 2008. Validation and comparison of 1 km global land cover products in China. Int J Remote Sens, 29: 3769–3785

Yu L, Wang J, Gong P. 2013. Improving 30 meter global land cover map FROM-GLC with time series MODIS and auxiliary datasets: A segmentation based approach. Int J Remote Sens, 34: 5851–5867