

全球尺度多源土地覆被数据融合与评价研究

白 燕^{1,2}, 冯 敏³

(1. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101; 2. 江苏省地理信息资源开发与利用协同创新中心, 南京 210023; 3. 马里兰大学地理科学系, 美国马里兰 20742)

摘要: 精确的全球及区域尺度土地覆被遥感分类数据是全球变化、陆地表层过程模拟、生态文明建设及区域可持续发展等研究的重要基础数据。本文以5套全球土地覆被数据集GLCC、UMD、GLC2000、MODIS LC、GlobCover为研究对象, 结合MODIS VCF、MODIS Cropland Probability以及AVHRR CFTC数据集, 设计一种基于模糊逻辑思想的证据融合方法实现上述多源土地覆被信息的决策融合, 生成一套依据植物功能型分类的全球1 km土地覆被融合数据SYNLcover。结果显示, 与5套源土地覆被数据集相比: ①在总体一致性精度上, SYNLcover的8个生物形态类型和12个目标类型的平均总体一致性精度最高, 分别约为65.6%和59.4%, 其次依次是MODIS LC、GLC2000、GLCC和GlobCover, UMD的最低, 分别约为48.9%和42.6%, 而且SYNLcover与5套源土地覆被数据集两两相比的总体一致性都是最好的; ②在类型一致性精度上, 除灌丛类型外, SYNLcover中包括森林、草地、耕地、湿地、水体、城镇建筑和其他7种生物形态类型, 以及森林类型的5种叶属性的平均一致性精度也是最高的, 如其他类型的平均一致性精度可达67.73%; ③除灌丛和湿地类型外, SYNLcover的其余6种生物形态类型的平均一致性精度均比其在5套源数据中相应的一致性精度的最大值提高了10%~15%左右; 森林类型的5种叶属性的一致性精度也提高了约10%。SYNLcover分类精度的提高反映了本研究设计的多源数据融合方法的可行性和有效性。

关键词: 土地覆被; 模糊逻辑; 相关性分值; 数据融合; 一致性精度评价; 多源信息

DOI: 10.11821/dlxb201811013

1 引言

土地覆被空间分布及其变化反映了人类社会经济活动过程, 被视为全球陆地生态系统变化最为重要的影响因素之一^[1]。大尺度土地覆被分类制图工作尤为重要, 它为全球陆地表层过程与生态系统碳水循环模拟^[2-5]、气候变化^[6]、土壤侵蚀^[7]、生物多样性^[8]、人地关系及区域可持续发展^[9]等研究提供了基础的输入数据源。国际社会十分重视全球地表覆被及其变化研究, 自20世纪90年代起, 世界各国许多组织(如美国地质调查局、欧空局、中国国家测绘地理信息局等)相继运用不同的遥感影像和分类技术, 开展了以土地覆被为主体的遥感制图研究, 形成了一系列不同时空分辨率的全球土地覆被数据集^[10-19],

收稿日期: 2018-01-24; 修订日期: 2018-08-22

基金项目: 科技基础资源调查专项课题(2017FY100900); 资源与环境信息系统国家重点实验室青年人才培养基金(Y6V60220YZ); 国家科技基础条件平台项目—国家地球系统科学数据共享服务平台(2005DKA32300)
[Foundation: Basic Resources Investigation of Science and Technology, No.2017FY100900; Young Talents Training Fund of State Key Laboratory of Resources and Environmental Information System of China, No. Y6V60220YZ; National Earth System Science Data Sharing Infrastructure, National Science & Technology Infrastructure of China, No.2005DKA32300]

作者简介: 白燕(1985-), 女, 山西偏关人, 博士, 工程师, 主要从事生态遥感和数据质量评价研究。E-mail: baiy@lreis.ac.cn

并在全球或区域尺度的科学研究和应用中发挥了重要作用。然而,由于数据生产者采用了不同的输入数据源、分类系统和制图技术,目前已经建立的全球土地覆被数据集彼此之间存在明显的差异,包括类型数量及其空间分布在相同区域尺度上的不一致性^[20-23],从而导致用户在具体应用中使用这些数据时存在很多不确定性,主要体现在:①全球土地覆被数据集的第三方精度评价并不高;②数据生产者各自使用了不同的验证框架和参考数据,使得这些独立完成的土地覆被数据集精度评价结果之间没有可比性。因此,在具体应用中,用户面临如何选择以及为什么选择某种土地覆被数据的不确定性挑战。

此外,单一的土地覆被数据集不能满足研究应用需求的问题日益凸显出来。在多源数据共存和集成研究的背景下,采用信息融合技术进行土地覆被分类制图是全球变化研究的主要发展趋势之一。数据融合是一个多层次、多方面的数据处理过程,可用来处理不同数据之间分类系统的不兼容,并在一定程度上提高数据的精度^[24]。许多国际组织,如全球森林和土地覆被动态变化监测站(Global Observation of Forest and Land Cover Dynamics, GOFC-GOLD)、国际粮农组织(Food and Agricultural Organizations, FAO)以及全球陆地观测系统(Global Terrestrial Observing Systems, GTOS)等共同合作,致力于融合两种或两种以上全球土地覆被数据集信息,以期生产精度更高的土地覆被数据产品的研究^[25],并于2014年共同发布了一套全球土地覆被数据库GLC-SHARE(Global Land Cover SHARE),主要是以欧洲空间局的GlobCover-2009为主要信息,采用数据融合技术将质量较高的全球不同国家级尺度土地覆被数据以图注的形式实现了全球范围内的可视化展示^[16]。

在已有的对5套全球宏观尺度土地覆被数据集GLCC、UMD、GLC2000、MODIS LC和GlobCover的类型、空间一致性及其精度进行系统的对比评价,在了解它们各自的优缺点的基础上^[22, 26],本文从相关性分值的定义和融合规则出发,结合利用土地覆被辅助数据集,提出并设计一种基于模糊逻辑思想的证据融合方法实现多源土地覆被信息的综合,产出一套精度较高的全球土地覆被融合数据产品,并对其与5套源土地覆被数据集的一致性精度进行对比评价与分析。本研究不仅能够有效解决现存的土地覆被数据集之间不兼容及其数据精度有限的问题,同时能为在多种数据并存的背景下,利用多源数据决策融合技术进行大尺度土地覆被制图研究提供技术支持与科学指导。

2 数据源及其预处理

2.1 多源土地覆被信息

本文用到的多源土地覆被相关数据集包括以下8个,其基本特征如表1所示:

(1) 美国地质调查局(United States Geological Survey, USGS)建立的国际地圈生物圈计划(International Geosphere-Biosphere Programme, IGBP)分类(17类)的全球土地覆被数据(Global Land Cover Characterization, GLCC)^[10];

(2) 美国马里兰大学建立的简化IGBP分类(14类)的全球土地覆被数据UMD^[12];

(3) 欧洲联合研究中心建立的FAO LCCS(Land Cover Classification System)分类(23类)的全球土地覆被数据GLC2000(Global Land Cover 2000 project data)^[13];

(4) 波士顿大学建立的IGBP分类(17类)的全球土地覆被数据MODIS LC(Moderate Resolution Imaging Spectro-radiometer annual land cover product, Collection 5)^[11];

(5) 欧洲空间局建立的联合国LCCS分类(22类)的全球土地覆被GlobCover

表1 多源土地覆被数据集的基本特征
Tab. 1 Characteristics of multi-source land cover datasets

数据集	传感器	数据时间	空间分辨率	分类方法
GLCC	AVHRR	1992.4-1993.3	1 km	非监督分类, 分类后处理
UMD	AVHRR	1992.4-1993.3	1 km	监督分类决策树
GLC2000	SPOT-4	1999.11-2000.12	1 km	非监督分类
MODIS LC	MODIS	2001.1-2002.12	500 m	决策树, 人工神经网络
GlobCover	MERIS	2004.12-2006.7	300 m	非监督分类
MODIS VCF	MODIS	2000	250 m	回归树
MODIS Cropland Probability	MODIS	2000	250 m	决策树
AVHRR CFTC	AVHRR	1992.4-1993.3	1 km	线性混合

(Global Land Cover Map)^[27];
(6) 美国马里兰大学建立的全球植被连续覆盖数据 MODIS VCF (Vegetation Continuous Field)^[28];
(7) 美国南达科他州立大学建立的全球农田分布概率数据 MODIS Cropland Probability^[29];
(8) 美国马里兰大学建立的全球森林连续覆盖数据 AVHRR CFTC (Advanced Very High Resolution Radiometer-Continuous Fields of Tree Cover)^[30], 包括森林覆盖类型的叶型 (阔叶/针叶) 和叶物候 (常绿/落叶) 的覆盖率信息。

2.2 数据预处理

为了便于后续多源信息融合工作的开展, 首先需要将上述8个具有不同空间坐标参考和空间分辨率的土地覆被相关数据源进行相应的预处理, 主要包括: ① 数据统一转换为正弦曲线 Sinusoidal 投影, 其空间范围为 (180°W~180°E、55°S~90°N), 即不包含主要分布着海洋和南极洲的 55°S 以南区域; ② 利用最邻近算法将数据统一转换为 1 km 空间分辨率。

同时, 为了便于从 AVHRR CFTC 数据中获取森林类型的叶型及叶物候属性信息, 本文将该数据中的叶型和叶物候覆盖率数据层分别进行合并处理, 即将阔叶与针叶覆盖率数据层、常绿与落叶覆盖率数据层分别合并成对应的叶型、叶物候数据, 合并规则如表2所示。

表2 AVHRR CFTC 叶型和叶物候数据合并规则
Tab. 2 Reclassifying rules for leaf type and leaf longevity of AVHRR CFTC

AVHRR CFTC	叶型	叶物候
覆盖率> 66%	针叶	常绿
覆盖率>66%	阔叶	落叶
覆盖率< 66%	混交	混交

3 多源土地覆被数据融合决策方法

多源土地覆被信息融合的目标就是采用一种能够灵活处理源数据不同分类体系与目标分类系统之间模糊性的方法, 将几种不同分类系统的土地覆被数据源进行融合处理, 生成一套在目标分类系统下最佳估计的土地覆被综合数据集。本文设计一种基于模糊逻辑思想的证据融合方法, 并结合 MODIS VCF、MODIS Cropland Probability、AVHRR CFTC 等辅助数据, 实现对5套全球土地覆被数据集的高效融合。

3.1 目标分类系统的定义

本文面向陆面过程模型应用，并依据植物功能型 PFTs (Plant Functional Types) 的划分标准，包括生物形态（如乔木、灌木和草本等）、乔木叶型（针叶、阔叶和混交）及其叶物候（常绿、落叶和混交）综合定义了一个新的目标土地覆被分类系统，共包括 12 类（表 3）。该分类系统满足以下两个基本要求：① 所有的分类信息都能从输入数据中获得；② 分类易于在遥感影像上识别。

3.2 相关性分值的定义

相关性分值是建立多源土地覆被数据集中各类型与目标类型之间的模糊隶属关系的一种量化表达方法，也是基于模糊逻辑思想的多源数据融合方法的核心内容。本文从以下 3 个层面分别定义了源数据集中不同的土地覆被类型与目标分类（生物形态和叶属性）之间的相关性分值。

3.2.1 源数据集中土地覆被类型与森林类型的相关性分值 参考 IGBP 和 FAO 森林资源评估对森林类型的定义描述^[31-32]，本文依据森林树冠覆盖度大小和类型之间的语义相关关系，用 0~100 的分值确定源土地覆被数据集中各类型与森林类型的相关性分值。规则如下（示例如表 4 所示）。

表 3 目标土地覆被分类系统

Tab. 3 Land cover legends defined by life forms and leaf attributes

目标类型	生物形态	叶属性	
		叶型	叶物候
常绿针叶林	森林	针叶	常绿
常绿阔叶林	森林	阔叶	常绿
落叶针叶林	森林	针叶	落叶
落叶阔叶林	森林	阔叶	落叶
针阔混交林	森林	针叶	混交
	森林	阔叶	混交
	森林	混交	常绿
	森林	混交	落叶
	森林	混交	混交
灌丛	灌丛	-	-
草地	草地	-	-
耕地	耕地	-	-
城镇建筑	城镇建筑	-	-
湿地	湿地	-	-
水体	水体	-	-
其他(如冰雪、裸地等)	其他	-	-

表 4 根据树冠覆盖度定义的与森林的相关性分值示例

Tab. 4 Definition example of affinity scores for trees according to percentage of canopy cover

源土地覆被类型	树冠覆盖度(%)	相关性分值	目标类型(生物形态)
裸地	0	0	森林
草地	< 10	0	森林
火灾烧毁林地	> 15	30	森林
稀疏林地	40~60	50	森林
常绿针叶林	> 60	80	森林

- (1) 若源数据集中的某种土地覆被类型在语义上与森林类型完全匹配，那么其与森林的相关性分值取该类型的定义中树冠覆盖度大小的中值。
- (2) 若源数据集中的土地覆被类型为森林和其他植被的混合类型，并且其定义描述中的树冠覆盖度> 15%，那么根据混合类型中森林所占的比例及其与森林类型的语义关系，该混合类型与森林类型的相关性分值取其树冠覆盖度的最小值和中值之间的分值；如果植被混合类型的定义中树冠覆盖度< 10%，那么其与森林类型的相关性分值为 0。
- (3) 若源数据集中的土地覆被类型与森林类型没有任何语义相关关系，那么其与森林类型的相关性分值为 0。

3.2.2 源数据集中土地覆被类型与耕地类型的相关性分值 本文根据耕地的分布概率大小和类型的语义相关关系, 定义了源土地覆被数据集中各类型与耕地的相关性分值, 分值范围为0~100。规则如下 (示例如表5所示):

表5 根据语义关系定义的与耕地的相关性分值示例
Tab. 5 Definition example of affinity scores for cropland according to semantic rules

源土地覆被类型	语义规则	相关性分值	目标类型(生物形态)
稀疏草原	不是	0	耕地
自然植被/耕地混合体: 草地/灌木/林地(50%~70%)/耕地(20%~50%)	占少数	40	耕地
耕地与自然植被混合体	占半数	50	耕地
耕地/灌木/草地混合体	占多数	60	耕地
耕地	是	100	耕地

- (1) 若源数据集中的土地覆被类型在语义上与耕地完全匹配, 那么其与耕地的相关性分值为100, 即是耕地的概率为100%;
- (2) 若源数据集中的土地覆被类型为耕地和其他植被的混合类型, 那么其与耕地的相关性分值根据混合类型中耕地所占的比例及其与耕地的语义关系来灵活确定;
- (3) 若源数据集中的土地覆被类型在语义上与耕地没有任何关系, 那么其与耕地的相关性分值为0, 即是耕地的概率为0。

3.2.3 源数据集中土地覆被类型与其他目标分类之间的相关性分值 本文分别从土地覆被类型的生物形态和叶属性(叶型/叶物候)两方面, 根据语义来定义源数据集中各土地覆被类型与目标分类之间的相关性分值, 即依据如表6所示的语义规则, 用0~100内的5个分值来表示全球土地覆被数据集中各个类型与一个或多个目标分类系统中定义的土地覆被类型之间的模糊隶属度。

表6 根据语义规则定义的与其他分类之间的相关性分值示例
Tab. 6 Definition example of affinity scores for other legends according to sematic rules

源土地覆被类型	语义规则	相关性分值	目标类型(生物形态)
裸地	不是	0	耕地
稀疏林地	占少数	25	草地
稀疏草原	占半数	50	森林
稀疏灌木林	占多数	75	灌丛
城镇	是	100	城镇建筑

3.3 数据融合方法

3.3.1 多源土地覆被数据融合的技术方法 本文采用基于模糊逻辑思想的证据融合方法, 结合MODIS VCF、MODIS Cropland Probability 和 AVHRR CFTC 数据, 将GLCC、UMD、GLC2000、MODIS LC 和 GlobCover 5套全球土地覆被数据集进行融合, 生成包含两级类的全球土地覆被融合数据, 包括土地覆被生物形态类型融合数据(8类), 以及目标类型融合数据(12类)。多源数据融合方法的总体技术路线如图1所示。

3.3.2 多源土地覆被数据融合的计算方法 (1) 森林/非森林的确定

- 土地覆被融合数据中森林和非森林类型的确定方法如下: ① 将MODIS VCF数据所反映的树冠覆盖度大小(0~100), 作为该信息源与森林类型的相关性分值大小(0~100)。
- ② 结合利用5套土地覆被数据集中各类型与森林的相关性分值以及MODIS VCF数

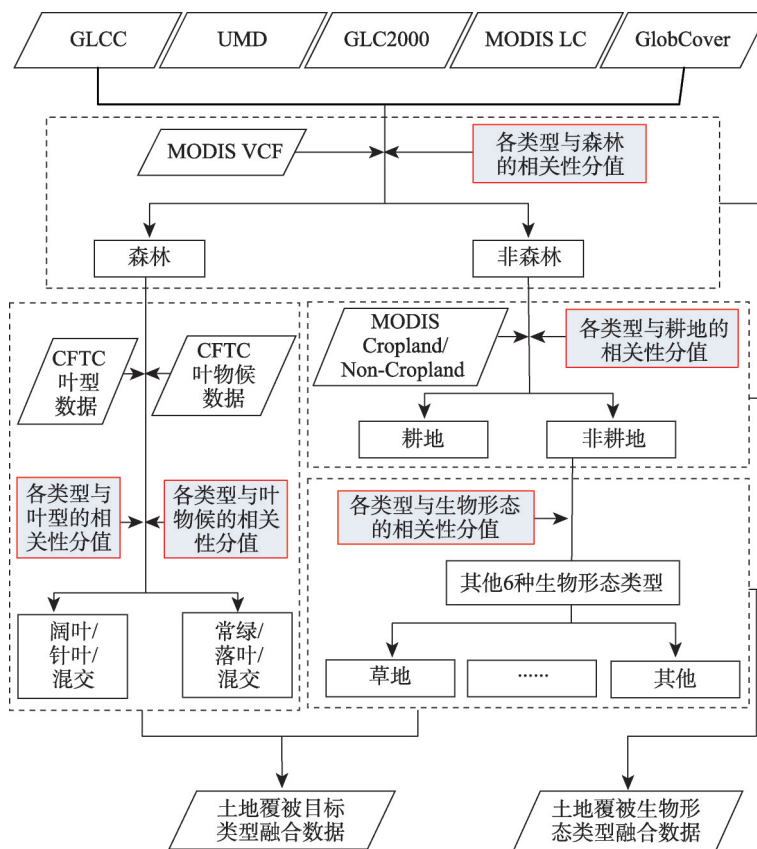


图1 基于模糊逻辑思想的多源证据融合技术方法

Fig. 1 Principle of the decision-fuse method with multi-source land cover information using fuzzy logic

据中相应的相关性分值,按照如下公式逐像元计算其与森林的相关性分值:

$$S_{Mean}^{Trees}(i, j) = \left(\sum_{M=1}^6 S_M^{Trees}(i, j) \right) / 6 \quad (1)$$

式中: $S_{Mean}^{Trees}(i, j)$ 是融合数据中像元 (i, j) 与森林类型相关性分值的平均值; $S_M^{Trees}(i, j)$ 是数据集 M 中像元 (i, j) 的类型与森林的相关性分值; M 代表 5 套全球土地覆被数据集和 MODIS VCF 数据; i 为像元的行号, j 为像元的列号。

③ 融合数据像元类型森林/非森林的确定规则是: 参考 IGBP 分类中将森林类型定义为“树冠覆盖度在 30% 以上的乔木植被”, 如果像元 (i, j) 按照公式 (1) 计算得到的与森林类型的分值 $S_{Mean}^{Trees}(i, j) \geq 30$, 那么该像元最终的土地覆被类型为森林; 反之, 若其分值 $S_{Mean}^{Trees}(i, j) < 30$, 那么该像元的土地覆被类型为非森林。

(2) 耕地/非耕地的确定

对于土地覆被融合数据中非森林类型的像元, 本文结合利用 MODIS Cropland Probability 数据的重分类数据 MODIS Cropland/Non-Cropland, 以及 5 套源土地覆被数据集中各类型与耕地的相关性分值, 将其进一步分为耕地和非耕地类型, 方法如下:

① 将 MODIS Cropland/Non-Cropland 数据中的值 0 和 1, 分别对应转换成各土地覆被类型与耕地类型的相关性分值大小 0 和 100;

② 利用5套土地覆被数据集中各类型与耕地的相关性分值及其在MODIS Cropland/Non-Cropland数据中相应的分值,按照如下公式逐像元计算其与耕地的相关性分值:

$$S_{Mean}^{Crop}(i,j) = \left(\sum_{M=1}^6 S_M^{Crop}(i,j) \right) / 6 \quad (2)$$

式中: $S_{Mean}^{Crop}(i,j)$ 是融合数据中像元 (i,j) 与耕地类型相关性分值的平均值; $S_M^{Crop}(i,j)$ 是数据集 M 中像元 (i,j) 的类型与耕地的相关性分值; M 代表5套全球土地覆被数据集和MODIS Cropland/Non-Cropland数据; i 为像元的行号; j 为像元的列号。

③ 参考美国南达科他州立大学将MODIS Cropland Probability重分类成Cropland/Non-Cropland数据时所使用的全球尺度的阈值,融合数据中像元耕地/非耕地类型的确定规则是:如果像元 (i,j) 按照公式(2)计算得到的分值 $S_{Mean}^{Crop}(i,j) \geq 43$,那么该像元最终的土地覆被类型为耕地;反之,若其分值 $S_{Mean}^{Crop}(i,j) < 43$,那么该像元的土地覆被类型为非耕地。

(3) 其他生物形态类型的确定

对于土地覆被融合数据中非耕地类型的像元,本文利用5套源土地覆被数据集中各类型与目标分类中生物形态类型的相关性分值,将其进一步划分为除森林和耕地之外的其他6种类型,计算方法如下:

$$S_{Total}^{Lf}(i,j) = \sum_{M=1}^5 S_M^{Lf}(i,j) \quad (3)$$

式中: $S_{Total}^{Lf}(i,j)$ 是融合数据中像元 (i,j) 的生物形态为Lf类型的相关性分值之和; $S_M^{Lf}(i,j)$ 是数据集 M 中像元 (i,j) 的类型与生物形态为Lf类型的相关性分值; Lf为各生物形态类型(表3); M 代表5套全球土地覆被数据集; i 为像元的行号; j 为像元的列号。

利用公式(3)计算得到的 $S_{Total}^{Lf}(i,j)$ 的最大值所对应的生物形态类型Lf,就是融合数据中像元 (i,j) 的生物形态类型。

(4) 森林类型叶属性的确定

对于土地覆被融合数据中类型为森林的像元,结合利用AVHRR CFTC叶型和叶物候数据,以及5套源土地覆被数据集中各类型与叶属性之间的相关性分值,确定了其叶型(阔叶/针叶/混交)和叶物候(常绿/落叶/混交)的分类信息,从而将森林类型进一步划分为表1中所示的5种类型。

融合数据中某像元的叶型和叶物候分值的计算方法分别如下:

$$S_{Total}^{LT}(i,j) = \sum_{M=1}^6 S_M^{LT}(i,j) \quad (4)$$

$$S_{Total}^{LL}(i,j) = \sum_{M=1}^6 S_M^{LL}(i,j) \quad (5)$$

式中: $S_{Total}^{LT}(i,j)$ 和 $S_{Total}^{LL}(i,j)$ 分别是融合数据中像元 (i,j) 森林类型的叶型和叶物候分别为LT和LL的相关性分值之和; $S_M^{LT}(i,j)$ 和 $S_M^{LL}(i,j)$ 分别是数据集 M 中像元 (i,j) 的类型与叶型为LT、叶物候为LL的相关性分值; LT和LL分别为叶型和叶物候属性(表3); M 代表5套全球土地覆被数据集和AVHRR CFTC叶型/叶物候数据; i 为像元的行号, j 为像元的列号。

从公式(4)和公式(5)可以得出,融合数据中像元 (i,j) 为森林类型的叶型和叶物候,分别就是 $S_{Total}^{LT}(i,j)$ 和 $S_{Total}^{LL}(i,j)$ 的最大值所对应的叶型LT和叶物候LL。

4 土地覆被融合数据的一致性精度评价

利用上述基于模糊逻辑思想的多源证据融合方法得到的全球 1 km 土地覆被融合数据集 SYNLCover (Synergy of Land Cover) 如图 2 所示。其中, 图 2a 和图 2b 显示的分别是土地覆被生物形态类型和包含叶属性信息的目标分类融合数据。

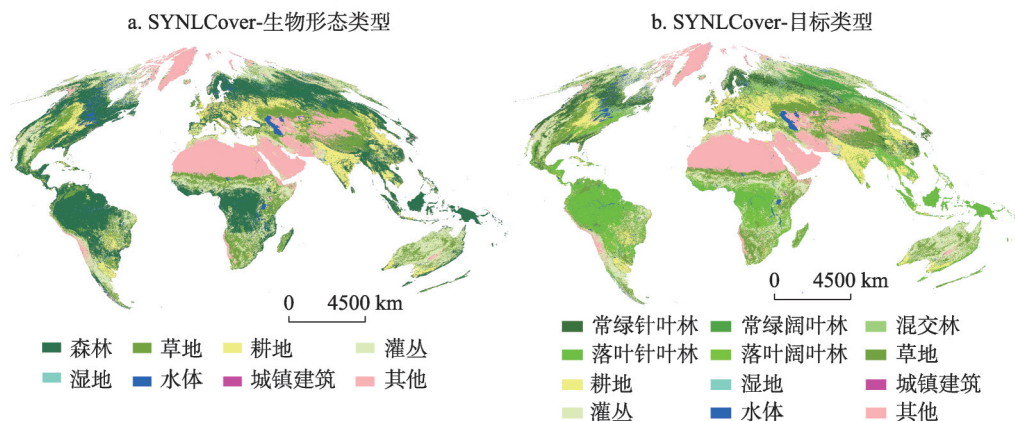


图2 全球土地覆被融合数据 SYNLCover

Fig. 2 The SYNLCover datasets with: (a) life forms, and (b) objective legends

本文分别从土地覆被生物形态类型和目标分类两个层面, 对比分析全球土地覆被融合数据 (SYNLCover) 与 5 套源土地覆被数据集 (GLCC、UMD、GLC2000、MODIS LC 和 GlobCover) 彼此之间的一致性精度, 包括: ① 数据集的平均总体一致性; ② 土地覆被类型的平均一致性。

4.1 土地覆被数据集的平均总体一致性

土地覆被数据集的平均总体一致性的计算分两个步骤: ① 分别计算 SYNLCover 和 GLCC、UMD、GLC2000、MODIS LC 以及 GlobCover 两两之间的一致性, 即两套土地覆被数据集中相同土地覆被类型的个数与土地覆被类型总数的百分比; ② 将其中一套土地覆被数据集与其他 5 套数据集两两之间的一致性求平均, 即可得到该土地覆被数据的平均总体一致性:

$$MeanC_a = (C_{ab} + C_{ac} + C_{ad} + C_{ae} + C_{af})/5 \quad (6)$$

式中: $MeanC_a$ 土地覆被数据集 a 的平均总体一致性; C_{a*} 为土地覆被数据集 a 分别与其他 5 套数据集之间的一致性; $a \sim f$ 为 6 套土地覆被数据集, 包括 GLCC、UMD、GLC2000、MODIS LC、GlobCover 以及 SYNLCover。

图 3a 和图 3b 分别显示了在中国区域尺度上, SYNLCover 和 5 套源土地覆被数据集在生物形态类型和目标类型两个层面上的总体一致性精度对比结果。其中, 两幅图中对角线上的值分别为每套土地覆被数据集的平均总体一致性精度 (%), 而非对角线的数值则是上述 6 套土地覆被数据集之间两两比较的总体一致性精度 (%)。

(1) 整体上来说, 6 套全球土地覆被生物形态类型数据集的平均总体一致性精度高于其目标类型的平均总体一致性精度;

(2) 在 6 套土地覆被数据集中, SYNLCover 的生物形态类型和目标类型的平均总体一致性精度都是最高的, 分别约为 65.6% 和 59.4%, 其次依次是 MODIS LC、GLC2000、

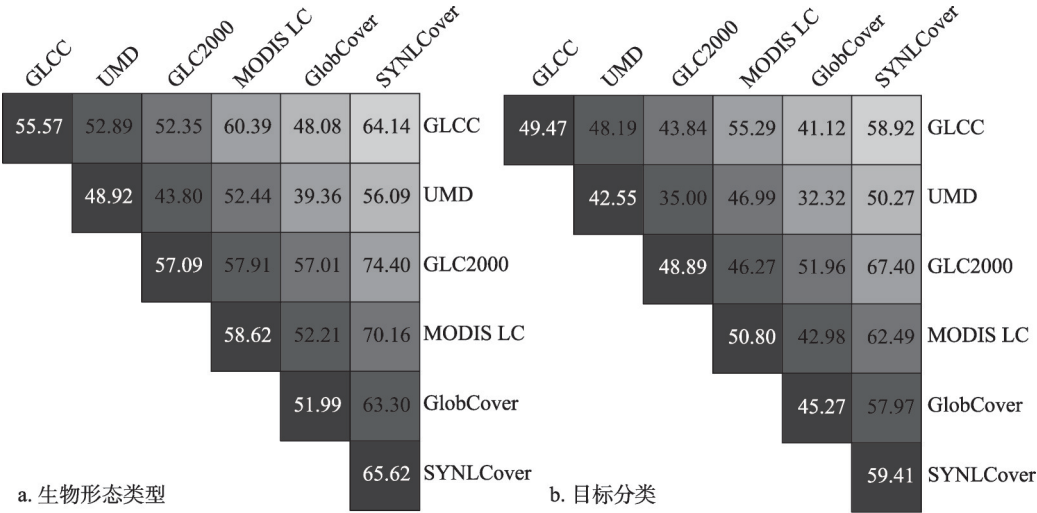


图3 6套土地覆被数据集的总体一致性精度对比

Fig. 3 Overall consistencies between SYNLCover, GLCC, UMD, GLC2000, MODIS LC, and GlobCover based on (a) life forms, and (b) objective legends

GLCC 和 GlobCover，UMD 的生物形态类型和目标类型的平均总体一致性精度最低，分别为 48.9% 和 42.6% 左右；

（3）就 6 套数据集两两比较的总体一致性精度而言，在土地覆被生物形态类型和目标类型两个层次上，SYNLCover 与 GLCC、UMD、GLC2000、MODIS LC 和 GlobCover 两两之间的总体一致性都是最好的，而且生物形态类型总体一致性均在 55% 以上，目标类型总体一致性也都超过 50%；特别是 SYNLCover 与 GLC2000 的生物形态类型和目标类型总体一致性最好，分别高达 74.4% 和 67.4%。

4.2 土地覆被类型的平均一致性

任意两套土地覆被数据集中各类型的一致性计算方法如下：

$$CT_{ab}^t = \frac{n_{ab}^t}{n_a^t + n_b^t - n_{ab}^t} \times 100 \tag{7}$$

式中： CT_{ab}^t 为两套数据集中土地覆被类型为 t 的一致性； n_{ab}^t 为两套数据集中土地覆被类型均为 t 的数量； n_a^t 和 n_b^t 分别为土地覆被数据 a 和 b 中类型为 t 的数量； a 、 b 为 GLCC、UMD、GLC2000、MODIS LC、GlobCover、以及 SYNLCover 其中任意两套数据集的组合。

类似地，每套土地覆被数据集中的各类型的平均一致性可以利用公式（6）计算得到，将该式中的 C_{a*} 替换成 CT_{a*}^t 即可。

结合利用公式（6）和公式（7）计算得到的 SYNLCover 和 5 套源土地覆被数据集中各生物形态类型和森林类型叶属性的平均一致性精度（%）对比分别如图 4 和图 5 所示。

从图 4 可以看出，在中国区域尺度上：① 与 5 套源土地覆被数据集相比，除灌丛类型外，SYNLCover 中其他的 7 种生物形态类型的平均一致性精度均是最高。其中，SYNLCover 中森林的平均一致性精度为 60.23%，草地的平均一致性精度为 43.89%，耕地的平均一致性精度为 54.13%，水体的平均一致性精度为 50.32%，其他的平均一致性精度为 67.73%，均高于相应类型在任意一套源土地覆被数据中的一致性精度；② 除灌丛和湿

地类型外,森林、草地、耕地、水体、城镇建筑和其他6种类型在SYNLCover中的平均一致性精度均比其在5套源数据集中的平均一致性精度的最大值提高了10%~15%左右。

就森林叶属性的平均一致性精度(图5)而言:①与5套源土地覆被数据集相比,SYNLCover中森林类型的5种叶属性的平均一致性精度最高。其中,SYNLCover中常绿针叶林的一致性精度最高,为29.21%;②5套源土地覆被数据集的叶属性的平均一致性精度都很低,其最大值不超过20%;但是,5种叶属性在融合数据SYNLCover中的平均一致性精度比其在源土地覆被数据集中的一致性精度提高了约10%。

5 结论与讨论

5.1 结论

在多源数据共存的背景下,信息融合技术能在一定程度上减少源数据集之间的不一致性,并生成精度较高的融合数据产品,从而有利于扩展这些数据集的应用范围。本文从相关性分值的概念出发,设计了一种基于模糊逻辑思想的证据融合方法,实现了GLCC、UMD、GLC2000、MODIS LC、GlobCover、MODIS VCF、MODIS Cropland Probability以及AVHRR CFTC等多源土地覆被信息的综合集成,生成了一套全球1 km土地覆被融合数据集SYNLCover。在此基础上,本文从土地覆被生物形态类型和目标分类两个层面对融合数据SYNLCover与源土地覆被数据集的总体一致性和类型一致性精度进行了评价。

结果显示:在中国区域尺度上,与5套源土地覆被数据集相比,SYNLCover的平均总体一致性精度最高;除灌丛和湿地类型之外的6种生物形态类型在SYNLCover中的平均一致性精度均比其在5套源数据集中的一致性精度的最大值提高了10%~15%左右,而且5种叶属性信息在SYNLCover中的平均一致性精度比其在源数据集中的一致性精度提

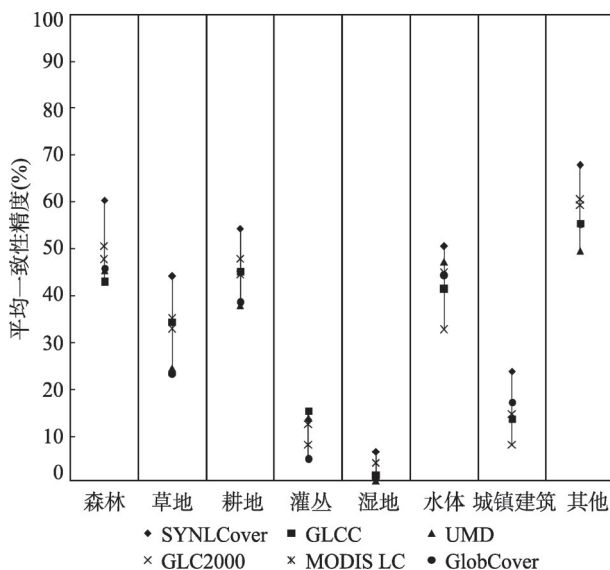


图4 SYNLCover与5套源土地覆被数据集中各生物形态类型的平均一致性精度对比

Fig. 4 Comparison of average consistencies for life forms among SYNLCover and five original land cover datasets

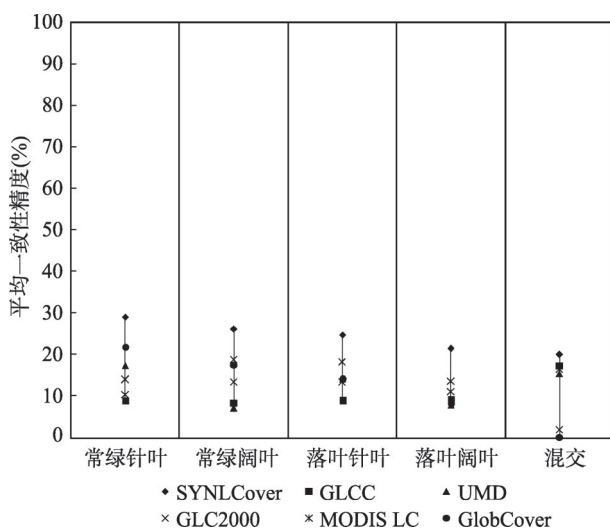


图5 SYNLCover与5套源土地覆被数据集中森林叶属性的平均一致性精度对比

Fig. 5 Comparison of average consistencies for leaf attributes of trees among SYNLCover and five original land cover datasets

高了10%左右。研究结果在一定程度上反映出本研究设计的基于模糊逻辑思想的证据融合方法的可行性和有效性,同时为该数据融合方法的使用和推广提供了实践基础和技术支持。

5.2 讨论

首先,全球土地覆被数据集的一致性精度评价对比是建立在统一的分类系统基础上的。因此,本文结合PFTs定义了一个面向陆面过程模型应用的土地覆被分类系统,作为数据集精度对比所需要的统一的目标分类。但与IGBP分类或FAO LCCS分类不同的是,该目标分类系统中没有关于各土地覆被类型的定义描述,不同数据的分类系统与目标分类之间的映射仅仅依据语义关系进行转换,从而导致会对数据集之间的一致性精度对比评价结果产生一定的误差影响。然而,从已有的相关研究和实际经验来看,不同分类系统中关于各个土地覆被类型指标的定量描述只是个理论值^[33],在目前关于宏观尺度的土地覆被遥感制图研究中,这些定量指标信息并不能被准确的反映出来。

其次,本文设计的多源土地覆被信息决策融合方法的关键和核心是相关性分值的定义,该分值是基于模糊逻辑思想并结合专家经验知识建立的量化表达了源土地覆被类型与目标类型之间的模糊隶属度关系,具有一定的主观性,可通过采用多位专家经验知识定义的分值综合考虑的方式最终确定相关性分值的大小,从而尽量将主观认识对融合结果的影响减少到最低。此外,基于模糊逻辑思想的证据融合方法从分层次、分类型的建立相关性分值,到定义约束规则从而实现多源土地覆被数据融合,可以为多源信息决策融合相关研究提供技术支撑和科学指导。

最后,利用高精度的土地覆被验证样点数据对全球土地覆被融合数据SYNLCover与5套源数据的精度定量评价与对比分析将会是下一步的研究内容,以此来补充完善这6套数据直接验证的精度评价体系,更加充分、突出反映本研究设计的基于模糊逻辑思想的证据融合方法的应用价值。

参考文献(References)

- [1] Chen Jun, Chen Jin, Gong Peng, et al. Higher resolution global land cover mapping. *Geomatics World*, 2011(2): 12-14.
[陈军, 陈晋, 宫鹏, 等. 全球地表覆盖高分辨率遥感制图. *地理信息世界*, 2011(2): 12-14.]
- [2] Cramer W, Kicklighter D W, Bondeau A, et al. Comparing global models of terrestrial net primary productivity (NPP): Overview and key results. *Global Change Biology*, 1999, 5(Suppl.1): 1-15.
- [3] Foley J A, DeFries R, Asner G P, et al. Global consequences of land use. *Science*, 2005, 309(5734): 570-574.
- [4] Alcamo J, Flörke M, Märker M. Future long-term changes in global water resources driven by socio-economic and climatic changes. *Hydrological Sciences Journal*, 2007, 52(2): 247-275.
- [5] Liu J, Vogelmann J E, Zhu Z, et al. Estimating California ecosystem carbon change using process model and land cover disturbance data: 1951-2000. *Ecological Modelling*, 2011, 222(14): 2333-2341.
- [6] Bounoua, L, DeFries, R, Collatz, G J, et al. Effects of land cover conversion on surface climate. *Climatic Change*, 2002, 52(1/2): 29-64.
- [7] Douglas I. Hydrological investigations of forest disturbance and land cover impacts in South-East Asia: A review. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 1999, 354(1391): 1725-1738.
- [8] Chapin F S, Zavaleta E S, Eviner V T, et al. Consequences of changing biodiversity. *Nature*, 2000, 405(6783): 234-242.
- [9] Lambin E F, Geist H J, Lepers E. Dynamics of land-use and land-cover change in tropical regions. *Annual Review of Environment and Resources*. 2003, 28: 205-241.
- [10] Loveland T, Reed B, Brown J, et al. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR Data. *International Journal of Remote Sensing*, 2000, 21(6/7): 1303-1330.
- [11] Friedl M A, Sulla-Menashe D, Tan B, et al. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 2010, 114(1): 168-182.
- [12] Hansen M C, Defries R S, Townshend J R G, et al. Global land cover classification at 1 km spatial resolution using a

- classification tree approach. *International Journal of Remote Sensing*, 2000, 21(6/7): 1331-1364.
- [13] Bartholome E, Belward A. GLC2000: A new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, 2005, 26(9): 1959-1977.
- [14] Bontemps S, Defourny P, Bogaert E, et al. Globcover 2009: Products Description and Validation Reports. 2011.
- [15] Tateishi R, Hoan N, Kobayashi T, et al. Production of global land cover data-GLCNMO2008. *Journal of Geography and Geology*, 2014, 6(3): 99-122.
- [16] Latham J, Cumani R, Rosati I, et al. Global Land Cover SHARE (GLC-SHARE) database Beta-Release Version 1.0. Rome: FAO. 2014.
- [17] ESA. CCI-LC Product User Guide v2.4. http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf. 2014.
- [18] Gong P, Wang J, Yu L, et al. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 2013, 34(7): 2607-2654.
- [19] Chen J, Chen J, Liao A, et al. Global land cover mapping at 30 m resolution: A pok-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 103: 7-27.
- [20] Congalton G R, Gu J, Yadav K, et al. Global land cover mapping: A review and uncertainty analysis. *Remote Sensing*, 2014, 6(12): 12070-12093.
- [21] Herold M, Mayaux P, Woodcock C, et al. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, 2008, 112: 2538-2556.
- [22] Bai Y, Feng M, Jiang H, et al. Assessing consistency of five global land cover data sets in China. *Remote Sensing*, 2014, 6(9): 8739-8759.
- [23] Tchunte A, Roujean J, Jong S. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. *International Journal of Applied Earth Observation and Geoinformation*, 2011, 13(2): 207-219.
- [24] Ran Youhua, Li Xin, Lu Ling. China land cover classification at 1 km spatial resolution based on a multi-source data fusion approach. *Advances in Earth Science*, 2009, 24(2): 192-203. [冉有华, 李新, 卢玲. 基于多源数据融合方法的中国 1 km 土地覆盖分类制图. *地球科学进展*, 2009, 24(2): 192-203.]
- [25] Herold M, Woodcock C E, Antonio D G, et al. A joint initiative for harmonization and validation of land cover datasets. *IEEE Transactions on Geoscience and Remote Sensing*, 2006, 44(7): 1719-1727.
- [26] Bai Y, Feng M, Jiang H, et al. Validation of land cover maps in China using a sampling-based labeling approach. *Remote Sensing*, 2015, 7: 10589-10606.
- [27] Arino O, Gross D, Ranera F, et al. GlobCover: ESA service for global land cover from MERIS//Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, 2017: 2412-2415.
- [28] DiMiceli C M, Carroll M L, Sohlberg R A, et al. Annual global automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m spatial resolution for data years beginning day 65, 1000-2010, Collection 5 Percent Tree Cover. Maryland: University of Maryland, 2011.
- [29] Pittman K, Hansen M C, Becker-Reshef I, et al. Estimating global cropland extent with multi-year MODIS data. *Remote Sensing*, 2010, 2(7): 1844-1863.
- [30] Defries R S, Hansen M C, Townshend J R G, et al. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Global Change Biology*, 2000, 6(2): 247-254.
- [31] Belward A. S. The IGBP-DIS global 1 km land cover data set "DISCover": Proposal and implementation plans: Report of the Land Cover Working Group of IGBP-DIS. IGBP-DIS Office, 1996.
- [32] FAO. Global Forest Resources Assessment 2010: Country Report. Forest Resources Assessment Programme, 2010.
- [33] Jung M, Henkel K, Herold M, et al. Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment*, 2006, 101(4): 534-553.

Data fusion and accuracy evaluation of multi-source global land cover datasets

BAI Yan^{1,2}, FENG Min³

(1. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 2. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; 3. Global Land Cover Facility, Department of Geographic Sciences, University of Maryland, College Park, MD 20742, USA)

Abstract: Accurate global and regional land cover classification datasets based on remote sensing are of fundamental importance in research on global changes, land surface process modeling, ecological progress, and regional sustainable development and so on. The overall objective of this study is to present a decision-fuse method that integrates existing multi-source land cover information into a 'best-estimate' dataset using fuzzy logic. Combined with another three global datasets, i.e., MODIS VCF (Vegetation Continuous Field), MODIS Cropland Probability, and AVHRR CFTC (Continuous Fields of Tree Cover), this method is applied to five global land cover datasets (GLCC, UMD, GLC2000, MODIS LC, and GlobCover) to generate a new 1-km global land cover product SYNLCover with desired legends, which are properly defined in terms of plant functional types. Pixel-based comparisons among these six global land cover datasets are performed, and results reveal that compared with five original global land cover datasets: (1) In terms of map-specific consistency, overall consistencies of both eight life forms and twelve objective legends of SYNLCover are the highest, accounting for about 65.6% and 59.4%, respectively; followed by the accuracy of MODIS LC, GLC2000, GLCC, and GlobCover in a descending order, and the lowest map-specific consistencies of life forms and objective legends are separately 48.9% and 42.6% in UMD. Besides, among all dataset pairs, SYNLCover agrees best with each original land cover dataset regarding the occurrences of life forms and leaf attributes. (2) In terms of class-specific consistency, it is suggested that SYNLCover gets the highest average class consistencies for all the five leaf attributes, as well as major life forms except Shrubland, among which the consistency for Others in SYNLCover is up to 67.73%. (3) For Trees, Grassland, Cropland, Water, Urban and built-up and Others, SYNLCover shows particular improved average class-consistencies by about 10% to 15% over the maximum consistency of original datasets, and the consistencies of five leaf attributes in SYNLCover also increases by about 10%. This study indicates a successful integration of multi-source land cover information into a new refined dataset with improved characteristics scientifically.

Keywords: land cover; fuzzy logic; affinity scores; data integration; consistency assessment; multi-source information