

基于 ArcPy 的林业与国土“一张图”数据批量自动化融合技术研究

代劲松, 王剑武, 徐达, 宋盛

(浙江省森林资源监测中心, 浙江 杭州 310020)

摘要: 森林资源专项调查是自然资源调查工作的重要组成内容。为解决浙江省森林资源管理“一张图”(以下简称林业“一张图”)与第三次全国国土调查数据(以下简称国土“三调”)数据融合过程中的技术问题, 基于 ArcPy 开发实现了数据的空间融合、属性融合及数据质量检查, 并利用 ArcToolbox 的定制功能, 对该一体化流程进行封装与可视化, 提高了其交互性和有效性。以浙江省淳安县为例, 进行淳安县林业“一张图”与国土“三调”数据融合, 经验证, 该技术方法可以有效提高数据融合的效率, 大大减轻技术人员的工作量, 对更好地开展森林、草地、湿地调查监测工作具有借鉴意义。

关键词: ArcPy; 森林资源管理“一张图”; 批量自动化; 第三次全国国土调查; 数据融合

中图分类号: S758.62 **文献标识码:** A **文章编号:** 1001-3776(2022)06-0069-08

Fusion of Data from Forest and National Land Resources Survey by ArcPy

DAI Jin-song, WANG Jian-wu, XU Da, SONG Sheng

(Zhejiang Forest Resources Monitoring Center, Hangzhou 310020, China)

Abstract: Fusion of data from forest resources survey in 2020 and the third national land resource survey in 2019 in Chun'an county, Zhejiang province was carried out by ArcPy, and a series of operation processes such as data spatial fusion, attribute fusion and data quality inspection were developed. It was verified that this technical method could effectively improve the efficiency of data fusion, greatly reduce the workload of technicians.

Key words: Arcpy; one map of forest resources management; batch automation; the third national land resource survey; data fusion

2021 年, 国家林业和草原局提出了开展林地、草地、湿地数据与第三次全国国土调查数据(以下简称国土“三调”)对接融合和国家级公益林优化的工作, 即充分发挥“三调”数据在国土空间管理中的“统一底版”作用, 全面厘清林地、草地、湿地与其他土地范围界线, 构建与国土“三调”数据无缝对接的全国林草资源图^[1-2]。长期以来, 我国的自然资源调查和监测工作实行各部门分头管理, 林业部门和国土部门在林地属性的理解上, 由于认定标准、调查方式等差异, 导致调查成果空间分布及面积数量有较大歧义^[3], 从而增加了整个数据融合的工作难度, 主要问题包括: (1) 各种界线的不一致, 包括各级行政界线、森林经营界线、国土图斑与林业小班界线等的差异, 由此会导致融合后的小班数量成倍增加, 大量无规则和小面积细碎图斑产生; (2) 小班面积平差规则及计算方式的不一致, 使得同一地块出现不同的面积; (3) 关键地类的重叠交叉, 主要表现为国土“三调”数据中的耕地、园地与森林资源管理“一张图”中林地的重叠交叉^[4-5]。针对这些问题及各地的实际情况, 需要一套统一的数据处理软件来处理融合过程中的大数据, 以提高工作效率和保障数据质量。

本研究以浙江省淳安县为例, 依据浙江省森林资源管理“一张图”(以下简称林业“一张图”)与国土“三

收稿日期: 2022-07-03; 修回日期: 2022-10-09

基金项目: 浙江省省院合作林业科技项目(2021SY05)

作者简介: 代劲松, 博士, 工程师, 从事森林资源监测工作; E-mail: Aarondjs@outlook.com。

调”数据对接融合的相关技术要求,基于 ArcPy 进行 GIS 二次开发,利用 ArcToolbox 的定制功能,对工具进行封装与可视化,以实现融合工作中数据自动化批量处理。

1 数据来源与研究方法

1.1 研究区概况及数据来源

研究区位于浙江省杭州市淳安县,地处浙江省西北部,地理坐标为 $29^{\circ}11' \sim 30^{\circ}02' \text{N}$, $118^{\circ}20' \sim 119^{\circ}20' \text{E}$, 是著名的国家 5A 级风景区千岛湖所在地,又是浙江省政府批准的革命老根据地县^[6]。全县森林资源丰富,是全国南方重点林区县(市)之一,截至 2019 年底,全县森林面积有 $339\,642.267 \text{ hm}^2$,森林覆盖率为 76.86%,林木绿化率为 78.93%。属中亚热带季风气候,温暖湿润,雨量充沛,四季分明,年平均气温为 17.2°C ,最冷月 1 月平均气温为 5.3°C ,最热月 7 月平均气温为 28.6°C ,年平均降水量为 $1\,515.0 \text{ mm}$ 。

研究采用的数据主要包括淳安县 2020 年森林资源管理“一张图”数据和 2019 年第三次全国国土调查统一时点的数据更新成果。

1.2 数据融合要求

林业与国土“一张图”数据融合主要包括三个方面:图形数据界线融合、属性数据融合及融合数据质量检查^[1]。

1.2.1 图形数据界线融合 主要包括两个方面:一是各级行政界线的统一,按照融合技术规程要求严格采用“三调”的行政界线及行政信息,可补充林业相关经营界线;二是林业小班界线与“三调”图斑界线的融合。

1.2.2 属性数据融合 以国土“三调”数据为基础底图,在保留“三调”图斑所有属性因子的基础上,进一步融入森林资源专项调查的特有因子,从而反映出森林资源的总体情况。对于两套数据属性因子有不一致的小班,运用遥感影像或现地核实的方法,根据实际情况确定小班因子。

1.2.3 融合数据质量检查 为了保证数据融合的质量,在完成数据的图形和属性数据融合后,需对数据的空间关系及属性逻辑关系进行质量检查。质量检查内容包括融合后图斑与“三调”图斑的拓扑空间关系检查、图斑属性因子完整性和逻辑性检查。

1.3 开发技术方法

ArcPy 是 ESRI 公司针对有批处理脚本及二次开发需要的 ArcGIS 用户提供的利用 Python 语言编写数据处理站点包,为使用 Python 执行地理数据分析、数据转换、数据管理和地图自动化创建等功能提出了实用高效的解决方案^[7]。Python 是一种解释型及通用的编程语言,具有语法简单、功能强大、跨平台等优点。尽管 ArcGIS 中也提供模型构造器来编辑和创建模型,并实现一体化集成运行地理处理工具,但是它只能调用 ArcGIS 本身所提供的内置工具,对于一些特殊的处理任务,仍需使用 ArcPy 进行定制化的二次开发。

1.4 系统数据融合总体设计

本研究共设计 3 大模块,共 11 个脚本工具,来完成融合过程中的自动化处理任务。其中,图形数据融合模块包含“消除细碎图斑”和“合并属性图斑”2 个工具,用于处理林业“一张图”和国土“三调”数据叠加分析以后所产生的大量细小图斑,并对空间相连且属性因子相同的图斑进行重新归并;属性数据融合模块包括“赋值邻近图斑属性”“面积平差”和“小班编号”3 个工具;数据质检模块包含“尖锐角检查”“狭长及扁担面检查”“拓扑关系检查”“多部件检查”“行政信息检查”和“因子逻辑检查”6 个工具。所有工具可在 ArcGIS 内置模型构造器中组合实现一体化地理处理,如图 1。

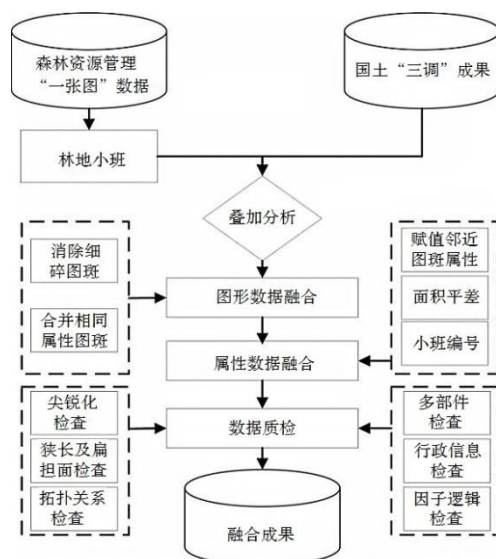


图 1 林业与国土“一张图”数据融合系统总体设计流程

Fig. 1 Overall design of data fusion in one map of forest and land resources

2 结果与分析

2.1 图形数据融合

由于林业“一张图”和国土“三调”成果数据在矢量界线上存在大量不一致情况,使得 GIS 空间叠加分析后产生了数倍无意义细碎斑块。本研究开发的“按属性归并细碎”工具自动解决了这一问题。工具的算法流程、功能实现界面及运行结果见图 2。核心实现过程主要包括:(1)设定细碎的最小面积,用 ArcPy 内置的 Update Cursor 函数搜索待处理的细碎图斑;(2)遍历搜索到的所有结果,对每一条记录使用 Search Cursor 函数查找与细碎图斑属性相同且面积大于最小面积的图斑作为归并的候选目标图斑,通过 Touches 函数判断细碎图斑与每一个候选目标图斑是否相邻,结合 Union 函数及 Part Count 属性判断细碎图斑与每一个候选目标图斑归并后是否会产生多部件;(3)记录目标图斑中与细碎图斑相邻且归并后不会产生多部件的面积最大的作为目标图斑,使用 Union 函数将细碎图斑归并到目标图斑。

主要应用情景包括:(1)消除国土中林地、草地、湿地中的细碎图斑,通过属性查询选择需要处理的图斑,按技术规程要求设定最小面积为 400 m² (推荐),属性值中选择国土的图斑标识码字段;(2)消除国土地类中除林地、草地、湿地外的其他细碎图斑,设定最小面积(推荐 667 m² 或以上,原林业调查的小班最小勾绘面积),属性值中选择国土的图斑标识码字段;(3)商品林图斑的处理,按照“行政/经营界线+国土标识码+林业地类+起源+树种+权属”进行归类,设定为新的属性值,根据需要适当调整细碎最小面积,减小商品林中图斑的破碎化。

在林业“一张图”中对于林地内的界线划分较为精细,但非林地部分比较粗糙,而在国土“三调”中,林地的界线相对较为粗放,一些非林地比如建筑、道路、河流等用地比较精细。因此,在两张图融合的过程中,要充分利用各自的数据特点,对部分图斑进行归并处理。

“按属性归并图斑”工具的算法流程、功能实现界面及运行结果见图 3。核心实现过程主要包括:(1)根据设定的属性值,调用 ArcPy 内置的 Dissolve_management 函数融合成新的多边形矢量,并统计新多边形矢量中每一个多边形由原矢量中图斑组成的数量;(2)遍历新的多边形矢量数据,使用 Search Cursor 和 Contains 函数在原数据中搜索出属性相同且空间落入新图形内的待归并的所有图斑;(3)使用 Union 函数将所有其他图斑归并到面积最大图斑中。

主要的应用情景包括:(1)对于部分国土地类中明显不会有植被覆盖的图斑进行归并,比如河流水面、坑



图 2 细碎图斑融合算法流程 (A)、功能实现界面 (B) 及运行结果 (C)

Fig. 2 Algorithm flowchat (A), function interface (B) and operation result for eliminating mini sub compartments

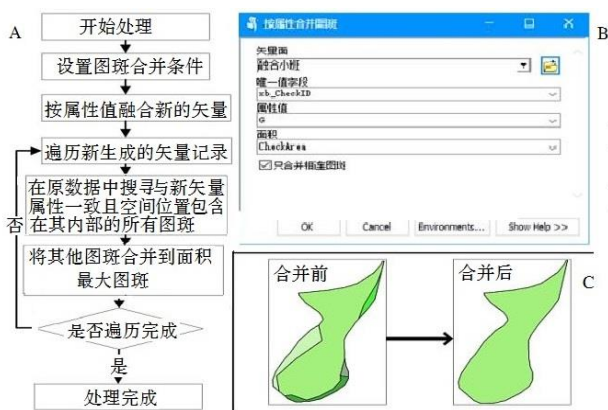


图 3 按属性归并图斑算法流程 (A)、功能实现界面 (B) 及运行结果 (C)

Fig. 3 Algorithm flowchat (A), function interface (B) and operation result for merging polygons by attributes

塘水面、沟渠、农村道路、农村宅基地等；(2)对于国土林地中林业地类为无立木林地的图斑,可以根据属性因子的相似性(林业一级地类相同、森林类别、事权等级、林地保护等级、权属等一致)进行归并。

2.2 属性数据融合

图形数据融合后,进一步完善属性信息是一项重要工作,其中“小班编号”“赋值邻近图斑属性”和“面积平差”3个工具可自动化完成这一工作,其算法流程、功能实现界面及运行结果见图4。

2.2.1 小班编号 “小班编号”工具的核心实现过程:根据不同行政单元(如建制村、林班等)及小班的空间位置信息,按照从北向南、从西向东的“S”形顺序确定小班的编号。通过设置分组的行政信息字段、用于升序和降序的小班空间位置信息的字段、起始小班号及结果写入字段,系统将调用 ArcPy 的内置 Sort_management 函数对数据进行重新编排,再使用 Setvalue 将编号写入结果字段中。

2.2.2 赋值邻近图斑属性 “赋值邻近图斑属性”工具的核心实现过程:(1)调用 Arcpy 内置的 Select LayerByLocation_management 函数搜索与需要赋值小班空间位置相近的作为候选数据源小班;(2)在所有候选数据源小班中查找国土土地类与需赋值小班相同且面积最大的作为数据源;(3)使用 SetValue 函数将数据源中相应字段值复制到需赋值的小班中。

2.2.3 面积平差 “面积平差”工具的核心实现过程:(1)按照国土“三调”面积计算要求,调用 CalculateField_management 函数,参数设置为 SHAPE.geodesicArea@SQUAREMETERS,重新计算每个小班的实际椭圆面积;(2)按国土“三调”标识码筛选出相关小班,计算每个小班在国土“三调”图斑内的面积占比,根据比例将原国土“三调”图斑面积分解到各小班;(3)重新汇总小班总面积,与原国土“三调”图斑面积比较计算面积差,将相差的面积增加到最大面积小班中。

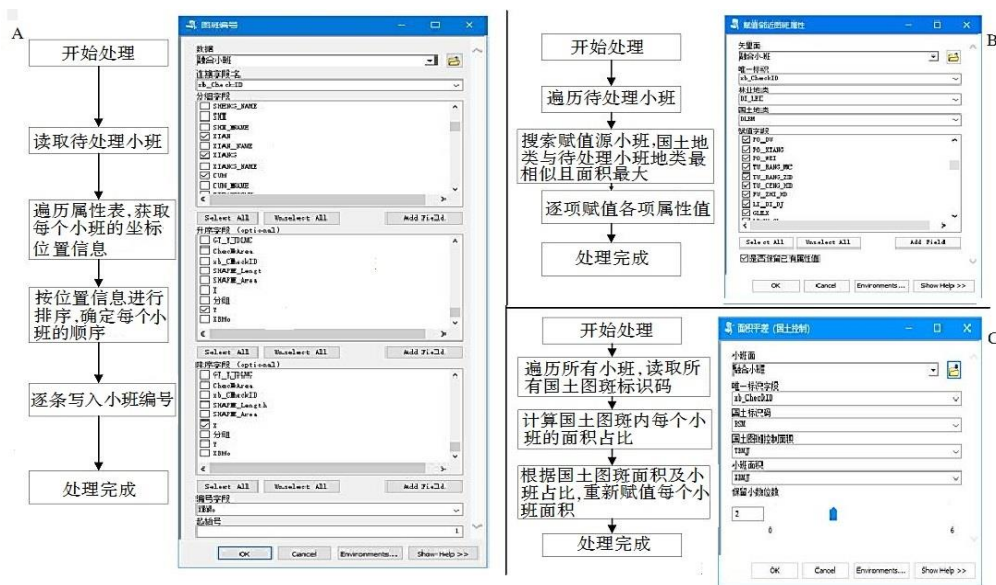


图4 小班编号(A)、赋值邻近图斑属性(B)及面积平差(C)算法流程和功能实行界面

Fig. 4 Algorithm flowchat and function interface for subcompartment numbering (A), assignment adjacent polygon attributes (B) and area adjustment (C)

2.3 数据质量检查

数据质量检查是保证融合成果的重要一环,包括尖锐角的检查和自动修复两部分。本研究共开发了6个主要的数据质检工具,其算法流程、功能实现界面及运行结果见图5至图7。

“尖锐角的检查及修复”工具包含对小班尖锐角的检查和自动修复两个部分。检查的核心实现过程主要包括:(1)调用 ArcPy 内置的 SearchCursor 函数对所有的小班进行遍历;(2)提取每个 Shape 对象的所有顶点,对所有顶点进行遍历,以当前点为中心,与其前后相连的两个顶点一起形成三角形;(3)计算三角形中当前顶点的内角值,根据阈值判定是否为尖锐角,调用 InsertCursor 插入尖锐角错误点矢量。自动修复需在检查的基础

上完成,其核心过程主要包括:(1)在尖锐角顶点及其前后两个顶点所形成的三角形上,以尖锐角顶点所在的短边向长边做垂线,调用 Polygon 函数生成新的直角三角形(如图 5C);(2)根据三角形与原图斑的位置关系,调用 Difference 将原图斑中直角三角形部分擦除,使用 Union 将直角三角形合并到相应图斑当中。

融合后的小班中会产生较多不规则多边形,其中以狭长面和扁担面(两头大中间小)最为普遍。狭长面的检查通过计算图斑的紧凑系数(CI, compaction index, 计算方法见公式 1)来判断。CI 是区域形状特征的一种测度量,取值在 0~1 之间,值越小表示区域形状越不紧凑,越分散;当区域为圆形时,CI 值=1,此时的区域形状最为紧凑;反之,当 CI 值趋近于 0 时,区域形状趋近于一条直线^[8-9]。

2.3.1 狭长面检查 狭长面检查的核心实现过程包括:

(1)调用 Shape.area 和 Shape.length 属性分别获得小班的面积和周长值;(2)通过公式计算小班的 CI 值;(3)将 CI 值与设定的阈值进行比较并判断小班是否狭长。

$$CI = \frac{2\sqrt{S}}{L} \quad (1)$$

式中, S 为图斑面积, L 为图斑周长。

2.3.2 扁担面 扁担面检查的核心实现过程包括:(1)根据设定的阈值调用 Shape.buffer 函数对小班进行负距离缓冲;(2)通过 PartCount 属性获得缓冲后小班的部件数量,如果数量大于 1 则判断其为扁担面。

2.3.3 小班拓扑检查 “小班拓扑检查”工具主要用于检查融合后的小班有无重叠,小班是否跨国土“三调”图斑界线,其核心实现过程包括:(1)调用 Add feature class To topology_management 将融合后的小班和原国土图斑矢量数据添加到拓扑关系集当中;(2)使用 Add rule to topology_management 添加检查规则,分别为 Must not overlap (area)、Must cover each other (area-area)及 Must be covered by (area-area);(3)调用 Validate topology_management 进行验证,统计错误数量。

2.3.4 多部件检查 “多部件检查”工具的核心实现过程包括:(1)调用 Multipart to single part_management 将原小班矢量数据转换为单部件矢量数据;(2)使用 Frequency_analysis 统计原小班面在新矢量数据中的多边形数量;(3)结合 Select layer by attribute_management 和 Delete rows_management 将统计结果中数量小于等于 1 的记录删除,剩余的数据即为有多部件错误的检查结果。

2.3.5 行政信息检查 “行政信息检查”工具用于检查小班填写的行政代码与名称是否一一对应,如果存在代码与名称一对多或多对一的情况,即表示存在行政信息填写错误。其核心实现过程包括:(1)调用



图 5 小班尖锐角检查及修复算法流程(A)、功能实现界面(B)及运行结果(C)

Fig. 5 Algorithm flowchart (A), function interface (B) and operation result for checking and repairing sharp angle of subcompartment

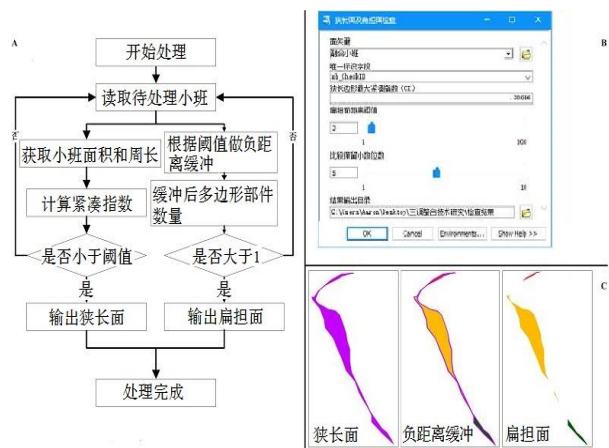


图 6 狭长面和扁担面的检查算法流程(A)、功能实现界面(B)及运行结果(C)

Fig. 6 Algorithm flowchart (A), function interface (B) and operation result for checking of long narrow and shoulder pole polygon

Dissolve_management 按设置的行政代码及名称字段进行融合,生成新的矢量数据;(2)使用 Frequency_analysis 分别按行政代码和名称组合统计记录数量;(3)统计记录数量大于 2 条的就表示存在行政代码一对多或行政名称存在一对多的错误。

2.3.6 因子逻辑关系检查 “因子逻辑关系检查”工具是保证融合数据因子逻辑正确性的关键。其核心实现过程包括:(1)通过 Open 函数读取预先写的检查逻辑条件文件,该文件为文本文件格式,奇数行是逻辑条件说明,偶数行是具体的检查 SQL 表达式,例如,逻辑条件说明为“林业地类为乔木林、竹林、疏林地、灌木林及未成林造林地,树种组成不能空”,其对应的 SQL 表达式为“(DI_LEI IN ('111','113','120','131','132','141')) and (SZ_ZC IS NULL or SZ_ZC = '')”; (2)使用 Select layer by attribute_management 和 Get count_management 根据检查逻辑条件进行筛选并统计记录数,如果记录数大于 0,表示有小班存在该条的逻辑错误。

2.4 研究区应用结果

2.4.1 淳安县数据融合 淳安县林业“一张图”与国土“三调”数据融合过程中的图斑数量统计结果见表 1。融合前国土“三调”共有 119 309 个图斑,其中林地图斑有 35 971 个,占总图斑数的 30.15%,林业“一张图”共有 137 704 个小班,其中林地小班有 129 885 个,占总小班数的 94.32%,可见两部门在调查时各自区划重点不相同。通过 ArcGIS 的叠加分析,总的图斑数量达到 898 209 个,分别是国土“三调”图斑和林业“一张图”小班总数的 7.5 和 6.5 倍,其中林地图斑有 466 709 个,分别是国土“三调”林地图斑和林业“一张图”林地小班总数的 12.8 和 3.6 倍。融合成果的各类图斑数量与叠加分析的结果相比有了大幅减少,总图斑数为 247 713 个,减少了 72.4%,其中林地、草地、湿地、种植园地及其他地类分别减少了 68.4%、71.5%、44.4%、76.7%和 77.1%。

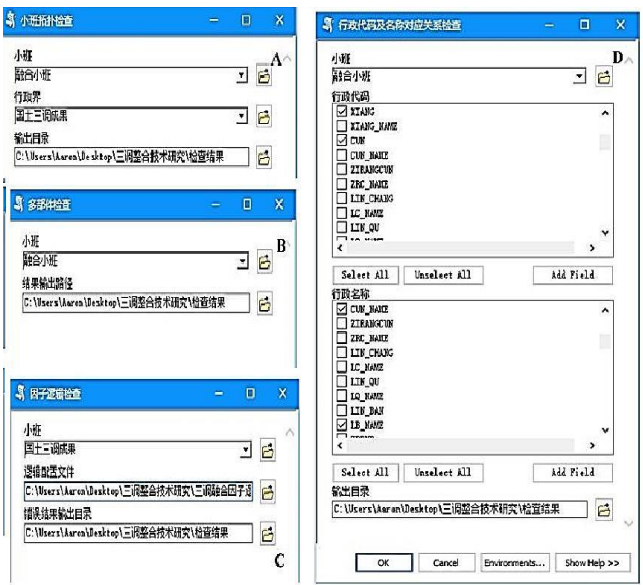


图 7 拓扑 (A)、多部件 (B)、因子逻辑 (C)、行政信息 (D)、检查功能实现界面

Fig. 7 Function interfaces of topology (A), multipart (B), logical checking (C), administrative information (D)

表 1 不同地类图斑数量统计

Tab. 1 Amount of polygon by different land types

地类	国土“三调”图斑数量/个	林业“一张图”小班数量/个	叠加分析图斑数量/个	融合成果图斑数量/个
合计	119 309	137 704	898 209	247 713
林地	35 971	129 885	466 709	147 342
草地	2 453		9 580	2 735
湿地	248		462	257
种植园地	27 732	7 819	185 608	43 298
其他	52 905		235 850	54 081

2.4.2 中洲镇数据融合 淳安县中洲镇数据融合成果见图 8。全镇原始国土“三调”图斑有 4 096 个,原始林业“一张图”小班有 6 644 个,融合后小班为 10 678 个(分别为国土“三调”和林业“一张图”小班数的 2.6 和 1.6 倍)。融合后的所有小班均落入国土“三调”图斑范围内,其中有 3 238 个小班与国土“三调”图斑完全一致,占国土“三调”总图斑数的 79.05%,749 个国土“三调”图斑被分割为 2~10 个小班,占比 18.29%,109 个国土“三调”图斑被分割为 10 个以上小班,占比 2.66%。

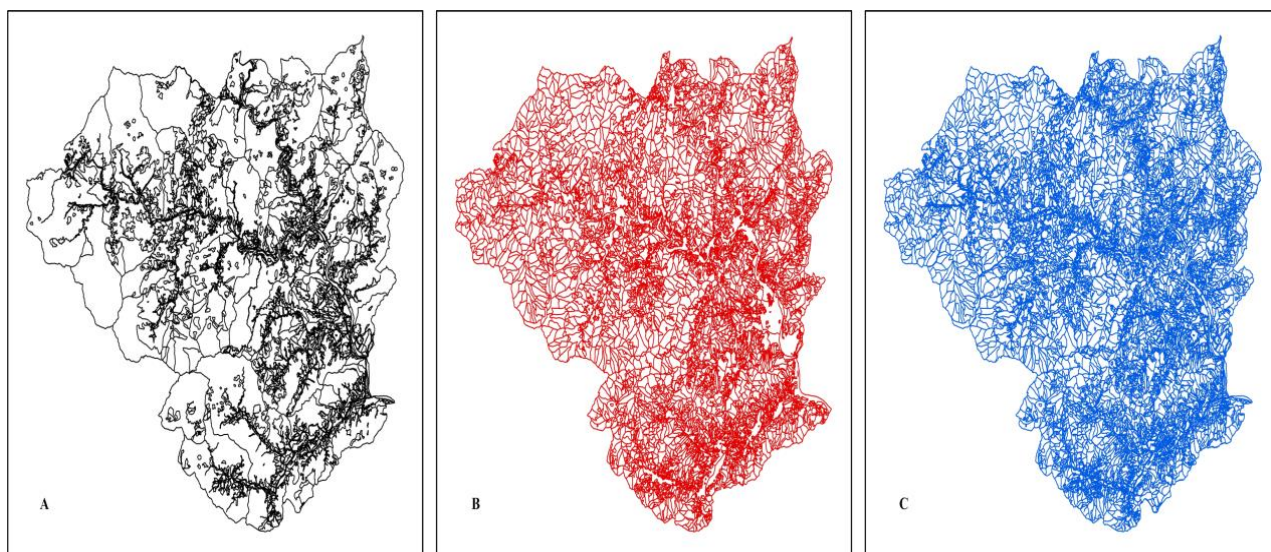


图 8 淳安县中洲镇国土“三调”(A)与林业“一张图”(B)数据融合成果(C)

Fig. 8 Data fusion results (C) of the third land resource investigation(A) and forest resource (B) of Zhongzhou Town of Chun'an County

3 结论与讨论

依据林业“一张图”与国土“三调”成果对接融合技术规程要求,利用 ArcPy 结合定制 ArcToolbox,提出了融合过程中数据的批量自动化处理方法,简化了整个操作流程,并以淳安县的实例数据开展应用研究,结果表明对接融合工作效率得到显著提高,主要完成了以下几个方面的内容:

(1) ArcPy 在测绘行业已经得到了较为广泛的使用,包括利用脚本自动处理错误多边形、图斑的自动编号、快速制图等^[7-11],但在林业行业中应用起步较晚,目前还主要用于编写批处理脚本,实现一些简单的小班编号和制图^[12-15],功能相对单一,由于操作时需要直接修改代码,使得应用难度较高。本文基于 ArcPy 编写代码,将以前在 ArcGIS 需要一系列手动操作的部分进行了集成,并且将代码整合进入 ArcToolbox,不仅可以在 ArcGIS 中直接使用,同时也可以与内置的工具相互组合并在 ModelBuilder 中构建各种新的批处理工具,整个程序可以分发和应用在 ArcGIS 9.0 以上及 ArcGIS Pro 的各种版本,使得程序更加便捷高效,减少了大量的重复劳动,工作效率大大提升。

(2) 针对 ArcGIS 的二次开发,林业行业以往主要采用 ArcEngine 和 ArcObjects 两种方式,开发的程序往往只能针对当时的技术要求,各个功能模块相互独立,难以应对多变的应用需要,同时程序的批处理能力较差,需要通过重新修改代码来增加批处理的功能模块^[16-19]。对于以后可能发生的技术要求变化,本研究所开发的各种工具只需要对其调用顺序进行重新组合即可实现程序的功能更新,重新编写代码的工作量 and 时间成本将会大大降低。

(3) 开发的各种工具不仅适用于本次融合的各项工作需要,也能够适用于林业其他工作的需要。例如:因子逻辑关系检查是森林资源管理各项工作的重要一环,每一项工作都有不同的数据库格式及因子逻辑关系需要控制,本文开发的工具只需要针对不同的应用情景,利用文本文件编写相应的逻辑检查条件即可满足要求;多源数据的整合管理也是森林资源管理中经常遇到的问题,包括国土年度变更、森林督查、林地征占用、历史森林资源档案等,这些数据与现有的森林资源数据往往都会因为空间坐标参考及区划尺度等一系列差异,在数据进行叠加分析时出现大量细碎图斑需要处理,利用本文开发的空間数据融合工具,只需要适当调整参数即可满足应用需求。

