

# Research on the Quality Analysis Method of Basic Geographical National Conditions Monitoring Results

Jingjing Zhao

The First Surveying and Mapping Institute of Xinjiang Uygur Autonomous Region, Changji, Xinjiang, 831100, China

## Abstract

In 2017 and 2018, China's Xinjiang Uygur Autonomous Region's basic geographic national conditions monitoring results acceptance work, the "monitoring results error data set" with errors in surface coverage and national conditions elements as the research object, the use of Python and ArcPy to extract and integrate the data elements, conducted an in-depth research on the quality status of the land cover classification data results and geographic national conditions element data results in the basic geographic national conditions monitoring results, evaluated the overall quality change trend of the results, and put forward a new idea for effective quality control of continuous geographic national conditions monitoring work.

## Keywords

geographic national conditions monitoring; quality control; error data set; result quality database

## 基础性地理国情监测成果质量分析方法的研究

赵晶晶

新疆维吾尔自治区第一测绘院, 中国·新疆 昌吉 831100

## 摘要

以2017年和2018年中国新疆维吾尔自治区基础性地理国情监测成果验收工作、地表覆盖与国情要素错误的“监测成果错误数据集”为研究对象,利用Python和ArcPy对数据进行要素提取、整合,对基础性地理国情监测成果中的地表覆盖分类数据成果和地理国情要素数据成果质量现状进行了深入研究,评价成果总体质量变化趋势,对持续性地理国情监测工作进行有效质量控制提出一个新思路。

## 关键词

地理国情监测; 质量控制; 错误数据集; 成果质量数据库

## 1 引言

中国新疆维吾尔自治区在2017年和2018年的基础性地理国情监测成果验收工作中自主创新,设计包含了地表覆盖与国情要素错误的“监测成果错误数据集”。两年的验收工作形成了数量可观的成果检验数据,成果每年更新,检验工作每年进行。利用成果检验数据、有针对性地为后续开展的监测工作进行技术指导和质量控制,是值得质检工作者深入分析的问题。

## 2 数据来源

### 2.1 验收工作数据成果

成果检查采用程序自动检查、人机交互检查和人工比对检查三种方式。检查中发现的数据质量问题形成“错误数据集.mdb”。中国新疆维吾尔自治区验收总面积为18.59

万 $\text{km}^2$ ,包括103个县级行政区划。根据规定要求,2017年、2018年分别抽取了11个县级任务区作为样本进行验收,其中地表覆盖分类数据抽样采用了多阶段抽样方法,对于成果采集精度和分类精度两项质量元素的检查,在县级测区范围内抽取不低于样本10%面积的图幅进行详查<sup>[1]</sup>。验收抽样面积达3.1万 $\text{km}^2$ ,2017年和2018年形成了包含县级任务区检查意见和分幅图检查意见在内的236个样本“错误数据集.mdb”。

### 2.2 错误数据集设计

论文采用以Access数据库为基础的个人地理数据库格式.mdb,设计样本错误数据集。该数据集中包括3个数据层,层名和字段定义如表1所示。

## 3 监测成果质量研究

质量研究的主要工作包括成果质量数据库建设、数据处理和数据分析。主要流程如图1所示。

### 3.1 成果质量数据库构建

第一,数据合并。利用Python脚本语言编写批量处理命令代码,调用ArcPy模块中的ListFiles函数、

【作者简介】赵晶晶(1979-),女,中国新疆乌鲁木齐人,本科,高级工程师,从事地理信息系统、航空摄影测量、遥感影像、地图制图研究。

表 1 错误数据集图层和字段定义

图层	属性项	描述	数据类型	数据长度
地表覆盖点状错误图层	ErrorDescription	错误描述	text	50
	ErrorCode	所属质量元素	text	50
	ErrorCount 计	计分错误个数	double	—
	ErrorField	错误字段	text	50
	RealValue	正确值	text	50
	ErrorCC	当前填写 CC 值	Test	50
地表覆盖分类精度错误图层	RealCC	正确 CC 值	Test	50
	ErrorDescription	错误描述	Test	50
地表覆盖分类精度错误图层	ErrorCode	所属质量元素	Test	50
	ErrorImportance	错误重要程度	Test	50
	ErrorDescription	错误描述	Test	50
	ErrorCode	所属质量元素	Test	50
	ErrorCount	计分错误个数	Double	—
	SelectableAttribute	必选属性	text	50
国情要素错误图层	CCcode	CC 值	text	50
	ErrorLayer	错误图层	Text	50
	ErrorField	错误字段	Text	50
	RealValue	正确值	Text	50

ListFeatureClasses 函数、Append\_management 函数。可对 236 个样本“错误数据集.mdb”进行批量同图层合并，并存储到新建的个人地理数据库中(.mdb),即成果质量数据库<sup>[2]</sup>。

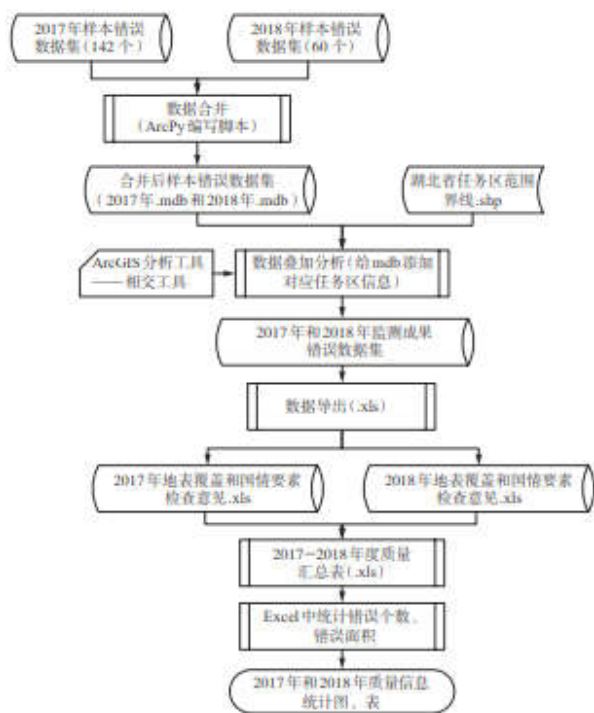


图 1 监测成果质量研究流程图

第二，数据叠加。利用 ArcMap 叠加分析功能将湖北省任务区范围界线数据链接到成果质量数据库，为成果质量数据库中所有图层添加样本名称、生产单位、任务区面积、样本量等字段。

### 3.2 数据处理

对成果质量数据库进行数据处理，利用 ArcGIS 转换数

据格式，输出地表覆盖点状错误图层 (.xls)、地表覆盖分类精度错误图层 (.xls)、国情要素错误图层 (.xls) 3 个文件。人工编辑提取“质量问题描述”“所属质量元素”“错误个数”“错误字段”“正确 CC 值”“错误 CC 值”“错误重要程度”等字段内容，按照一定规则统一规范质量问题描述，对各差错类别的主要问题描述进行关键字提取，合并同类问题，整理形成质量问题汇总表<sup>[3]</sup>。

### 3.3 数据分析

质量问题汇总表包括 2017 年、2018 年地表覆盖分类数据成果与地理国情要素数据成果的质量问题，分别对两类数据成果进行质量分析。地表覆盖分类数据成果在进行质量问题记录时，分类精度按照面积记录<sup>[4]</sup>，其他质量元素按照个数记录。因此，地表覆盖分类数据分为两类进行质量透视分析。

第一，分类精度错误统计。①对一级类分类错误与二、三级类分类错误的面积进行统计，得到 2017 年一级类分类错误的面积占比为 91.16%，二、三级类分类错误的面积占比为 8.84%；2018 年一级类分类错误的面积占比为 72.02%，二、三级类分类错误的面积占比为 27.98%。②对分类错误所属地类类别进行统计，按照一级类种植土地(01)、林草覆盖(03)、房屋建筑(05)、铁路与道路(06)、构筑物(07)、人工堆掘地(08)、荒漠与裸露地表(09)、水域(10)、地理单元(11)、地形(12)共 10 类统计错误面积，占比如图 2 所示。③对分类错误所属地类类别进行细分统计，统计每个一级类中被错分的地类类别，结果如图 2 所示<sup>[5]</sup>。

第二，其他质量元素错漏统计。①对错漏类型所属质量元素进行统计，分别统计表征质量、采集精度、拓扑一致性、属性精度的错漏个数占比，2017 年分别为 3.39%、63.29%、1.13% 和 32.19%；2018 年分别为 4.15%、61.83%、7.88% 和 26.14%。②对错漏类型所属检查项目

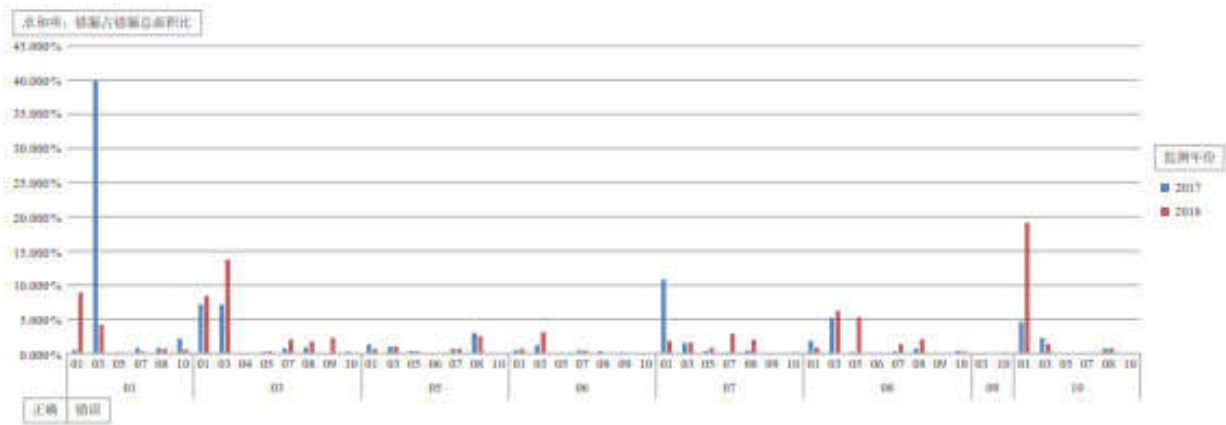


图2 地表覆盖分类数据分类错误所属地类类别细分统计

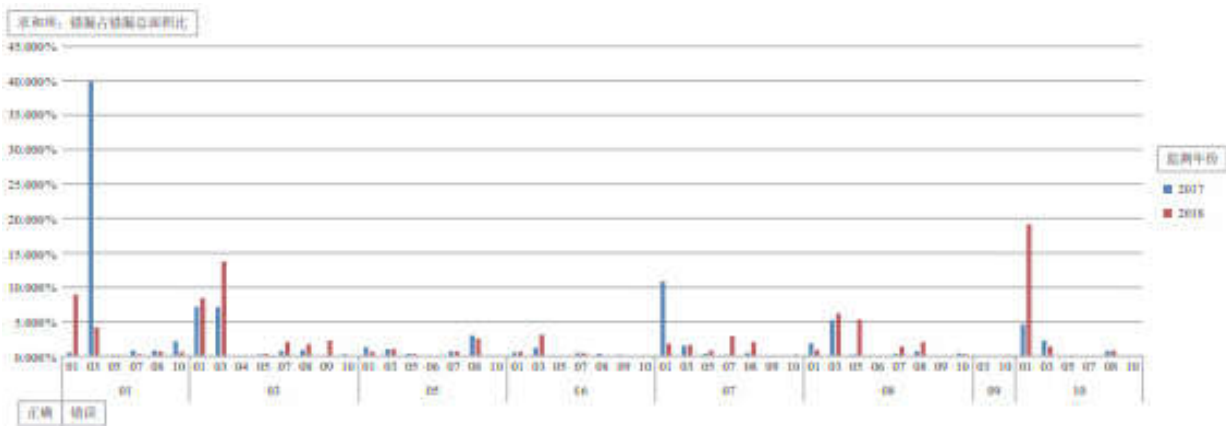


图3 地表覆盖分类数据错漏类型所属检查项目统计

进行统计,错漏类型包括属性不接边、Tag 赋值错误、ChangeType 赋值错误、面连续、图斑与影像套合超限、几何不接边、面折刺、极小面,错漏所属质量元素与个数占比如图3所示<sup>[6]</sup>。

#### 4 结语

论文建立统一标准的错误数据集和规范化的错误描述库,收集整理各级质检形成“错误数据集.mdb”,结合Python中的ArcPy模块建立年度成果质量数据库。论文通过对成果质量数据库构建方法的研究以及对建库数据的统计分析,确定了质量问题产生的原因并提出了科学有效的建议,对持续性地理国情监测工作进行有效质量控制提出了一个新思路,为建立科学合理的自然资源质量管控体系起到了一定的参考作用<sup>[7]</sup>。

#### 参考文献

[1] 国家测绘产品质量检验测试中心.GQJC11—2017基础性地理国

情监测检查验收与质量评定规定[S].北京:国家测绘产品质量检验测试中心,2017.

[2] 李强,白建荣,李振林,等.基于Python的数据批处理技术探讨及实现[J].地理空间信息,2015,13(2):54-56.

[3] 国家基础地理信息中心.GQJC01—2018基础性地理国情监测数据技术规定[S].北京:国家基础地理信息中心,2018.

[4] 陈海鹏,张莉,郭婧,等.地表覆盖分类数据成果质量检查技术与方法[J].测绘与空间地理信息,2018,41(7):72-74.

[5] 陈青,许庆领,李巍巍.基础性地理国情监测技术方法研究[J].地理空间信息,2016,14(2):28-30.

[6] 陈海鹏,程鹏飞,张莉,等.地理国情普查数据成果质量评价标准的探讨[J].测绘与空间地理信息,2017,40(1):43-45.

[7] 张继贤,翟亮.关于常态化地理国情监测的思考[J].地理空间信息,2016,14(4):1-3.