

Data Quality Evaluation for Database of the First Chinese Glacier Inventory

Lizong Wu, Xin Li

Cold and Arid Environmental and Engineering Research Institute, Chinese Academy of Sciences
Lanzhou, China 730000
wulizong@lzb.ac.cn

Abstract—The first glacier inventory in China which accomplished in 2002 made great contribution to World Glacier Inventory (WGI). The publication series of *Glacier Inventory of China* are the representative accomplishments of this project, which contain 21 books and about 200 attached maps. The database of Chinese glacier inventory has been developed on the basis of the publication of first glacier inventory of China in 2004. This paper introduced the development of the Chinese Glacier Information System in which a strict quality control was conducted to and most of the editing errors in the inventory were corrected. To overview the date quality of this database, evaluation was done for glacier area, spatial position and glacier state date, respectively. The results show that the errors meet the request to resource assessment for glacier and water resources. However, spatial error brought by traditional mapping method and digitized process.

Keywords—glacier; glacier inventory; database; data quality

I. 引言

冰川是气候变化信息指示器和存储器。冰川编目是冰川学研究中的一项非常基础和重要的工作，既可为冰川学研究和冰川水资源评估提供基础的数据，也可为冰川变化研究和冰川对气候变化响应研究提供长时间序列和高分辨的信息[1,2]。1965-1974年的“国际水文十年”提出对全球冰川和永久性积雪的分布和储量进行系统调查和评估，并于1973年在瑞士成立了国际冰川编目临时技术委员会，负责全球冰川编目的组织和协调。为响应全球冰川编目的号召，中国于1979年开始系统性的开展冰川编目工作，经过近50位冰川学家和工作者的努力，于2002年完成了这项艰巨的工程[3,4]。

为促进冰川编目成果的应用，2000年起开始对冰川编目成果进行数字化，并建立了数据库和信息查询系统。该成果通过《中国冰川信息系统》进行了正式发布[5]，并被世界数据中心兰州冰川（冰雪）冻土数据中心 (<http://wddcgg.westgis.ac.cn>)、美国雪冰数据中心 (<http://nsidc.org/>) 和世界冰川监测服务中心所收集，也被 GLIMS (Global Land Ice Measurements from Space, <http://www.glims.org>) 计划集成到全球冰川数据库中。中国冰川编目及其数据库的建立为全球冰川编目作出了重要贡献，据 GLIMS 计划建立的全球冰川数据库统计，约有 49% 的冰川记录来源于中国冰川编目。

中国冰川编目数据库发布后得到广泛的应用，但《中国冰川信息系统》一书仅对数据库的结构和数据处理过程进行了说明，并没有对数据质量进行系统的评价，研究者在使用数据过程中有诸多疑问，因此有必要对数据质量进行系统的评价和说明。本文将在简要介绍中国冰川编目的基础上，介绍数据处理过程和质量控制手段，着重对数据质量进行评价。2006年，我国新一轮的冰川资源及其变化调查研究正式启动，为了区别新一轮的冰川编目，本文将1979-2002年的冰川编目称之为中国第一次冰川编目。

II. 中国第一次冰川编目介绍

中国冰川编目是分区阶段完成的，以流域和山脉为编目单元，中国境内的冰川作用区被分为12个调查区。因此，中国冰川编目的最终成果-《中国冰川目录》被分为12卷21册，且出版年代不同。具体书目请参考文献[3,5]。

作为中国第一次冰川编目最终成果的《中国冰川目录》分为两部分：以表格形式表现的冰川基本属性和以附图形式表现的冰川分布图。冰川基本属性信息共33项，可分为3种类型：1) 编码信息，包括每条冰川的编码和流域编码；2) 23项对每条冰川属性进行描述的参数，包括冰川名称、冰川所在位置(经度和纬度)、冰川面积(总面积、裸露区面积和消融区面积)、冰川平均宽度、冰川平均长度、冰川最大长度(总长度、裸露区长度、消融区长度)、冰川朝向(积累区朝向和消融区朝向)、冰川高度(最大高度、中值高度冰舌、末端高度、冰川裸露区最低高度)、冰川类型、冰碛类型、雪线高度、雪线高度测量日期、冰川平均厚度和冰储量；3) 2项描述测量精度的参数，分别是冰川面积精度和雪线高度精度；4) 6项描述冰川编目所采用数据源的参数，包括地形图类型、地形图比例尺、地形图出版年代、影像类型(遥感图片或航片)、影像拍摄时间、影像比例尺。详细的数据参数说明可参考文献[3,5]。

冰川分布图是以《中国冰川目录》附图的形式出版的，每个冰川调查区的冰川分布图被分为若干图幅。冰川分布图中包含河流、湖泊、道路、基本居民点、流域分界线、山脊线、山峰和冰川信息(图1)，每幅图上都详细标注了五级流域编码和冰川编码。冰川分布图共122幅，比例尺介于1:25万-1:40万之间。详细的信息请参考文献[5]。

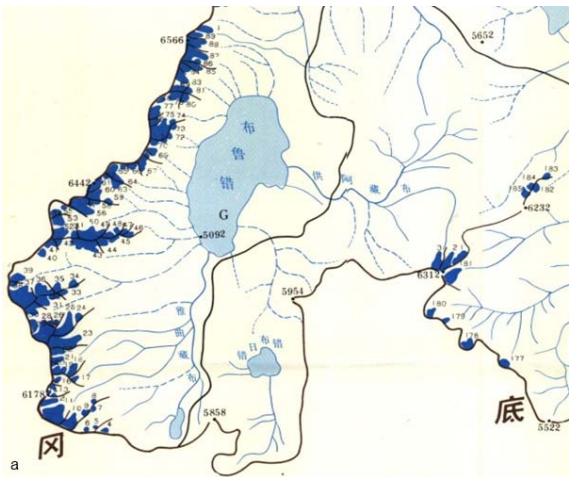


图1 《中国冰川目录》随书附图

III. 数据处理和质量控制

中国第一次冰川编目数据处理过程包括三个关键步骤：1) 录入并检查《中国冰川目录》上的冰川属性信息，获得每条冰川的参数和属性；2) 数字化冰川分布图，并恢复其地理坐标，对分区与的冰川分布图进行拼接，获得空间真实坐标的全中国范围内的冰川分布图；3) 利用冰川编码，将冰川属性信息与冰川空间分布图进行关联，获得以 GIS 图层形式表达的矢量化数据。

A. 冰川属性信息的录入和质量控制

《中国冰川目录》共 12 卷 21 册，其中，帕米尔高原地区利用不同的资料开展过两次编目，数据库实际参考的资料为 20 册。冰川属性信息的录入依靠手工方式录入，在初步录入的基础上又进行了 2 遍人工检查和 1 遍计算机逻辑检查（如通过比较录入数据的统计值和原书中的统计值来查找录入错误），尽量减少原书排版和印刷带来的错误。

《中国冰川目录》中的绝大多数冰川属性都对应于每条冰川，但有两类信息是与冰川所在的五级流域进行对应。其处理方法如下：

1) 《中国冰川目录》以五级流域为单元对冰川属性以表格的形式进行汇总，但冰川编目所用的基础资料信息是在表头部分进行注释，用于说明该流域冰川编目所采用的主要参考资料类型、比例尺和时间。如“AM-5-70, AP-4-56”，其中 AM 和 AP 分别代表地形图和影像的类型，5 和 4 分别代表地形图和影像的比例尺，70 和 56 分别代表地形图的出版时间和影像的摄影时间。在建立数据库时每条冰川记录都包含有参考资料信息，且都等于流域所参考的主要资料信息。

2) 由于部分地区缺乏大比例尺的地形图和航空相片，冰川编目只能参考比例尺较小的地形图或 Landsat MSS 卫星像片，冰川编目所统计的属性只包括冰川总条数和冰川总面积（如班公湖地区，流域编码 5Z42）。为了保证数据库与《中国冰川目录》统计结果的一致性，在数据录入时依然以每条冰川为单位，冰川的面积等于该流域内冰川的平均面积，其他信息缺省。

B. 冰川空间信息的处理与质量控制

冰川分布图是通过转绘的方法绘制而成的，在制图过程成带来了很大误差。为了提高数字化的精度，应尽量采用冰川编目所采用的工作底图（即经过航空相片修订过冰川边界的地形图）。在实际工作中只收集到恒河流域(501)、雅鲁藏布江流域(502)和印度河流域(503)的工作底图，共 500 余幅，比例尺为 1:5 万和 1:10 万。上述图件经过扫描后在 ArcView 3.2 环境下完成了矢量化，几何校正（或坐标恢复）和地图拼接过程是整个数字化过程的关键步骤。

由于《中国冰川目录》附图没有包含任何坐标信息，需要从其他地图采集地面控制点来进行校正。在实际工作中地面控制点均采集于冰川编目同时期同比例尺的地形图，共 1000 余幅。由于《中国冰川目录》附图中可用于几何校正的地物（如河流、湖泊和道路）都经过了地图概况，其形状与地形图上的同名地物不尽相同，因此几何校正误差较大。几何校正计算获得的 RMS 误差为 200-500m，实际误差可能大于该误差。地形图的校正较为简单，其精度主要取决于纸质地形图的变形程度，除部分折痕明显的地形图外，大部分地形图的几何校正误差在 100m 以内。

由于几何校正精度较差，不同图幅相邻冰川在几何校正后存在叠加、分离和错位等现象，这种现象在不同时期绘制的冰川分布图之间尤其明显。在实际的数据处理过程中，同一地区冰川分布图不同图幅间的相邻冰川进行了接边拼接处理。在当时条件下，对不同时期绘制的冰川分布图无法判断哪个更为准确，因此没有进行接边处理，留与数据用户参考其他资料进行处理。

C. 冰川属性数据与空间数据的关联

理论上通过冰川编码信息可以将冰川的空间信息和属性信息一一对应。但实际操作中存在一些问题，表现为：1) 约有 71 条冰川由 2-4 个多边形组成，冰川属性不能与冰川多边形一一对应；2) 参考资料不足时，只统计了流域冰川总条数和总面积的冰川属性不能与矢量化的冰川多边形一一对应；3) 由于排版或印刷错误，冰川分布图中的冰川编码顺序发生错误，冰川属性数据与冰川多边形不能一一对应。

IV. 数据质量评价

应用目标不同，研究者对冰川编目数据库的数据质量关注的角度也不同。从冰川资源评估或冰川水资源评估的角度，研究者主要关注的要素包括冰川面积、长度、海拔高度、雪线高度和冰储量的精度，从冰川变化分析的角度，研究者主要关心要素包括冰川面积、冰川长度、冰川空间几何形状和冰川编目所代表的冰川状态时间的精度。本文重点从冰川变化研究需要的角度对冰川面积精度、空间位置精度和冰川编目所代表的冰川状态时间精度进行评价。

A. 冰川面积精度

冰川面积的量算是通过求积仪和米格纸在地形图上测量获得的，冰川面积并不是地形图上绘制的冰川图形的原始面积，而是航空相片校正后的面积。因此在分析冰川变化时，不能简单的用从地形图上数字化的冰川面积代表冰川的实际面积。地形图比例尺越小，对冰川面积误差的影响越大。世界冰川编目规范中要求只对面积大于 0.02km^2 的冰川进行记录，冰川内部大于 0.02km^2 的裸露地区的面积也被排除在外。因此，在进行冰川面积变化分析时，要注意冰川面积测量方法的统一。

对比《中国冰川目录》中的冰川面积和根据数字化的冰川分布图计算获得的冰川面积，后者比前者大 10.29%。考虑误差的影响，建议采信前者作为冰川的真实面积。第一次冰川编目提出了 5 级冰川面积精度指标体系[6]，用数字 1-5 表示不同的面积精度。实际操作中绝大多数冰川面积误差都标注为 1，即面积误差小于 5%。冰川面积精度受参考资料类型、比例尺、编目者主观因素等多种因素的综合影响，面积精度应有很大的不同，需要重新进行评价。

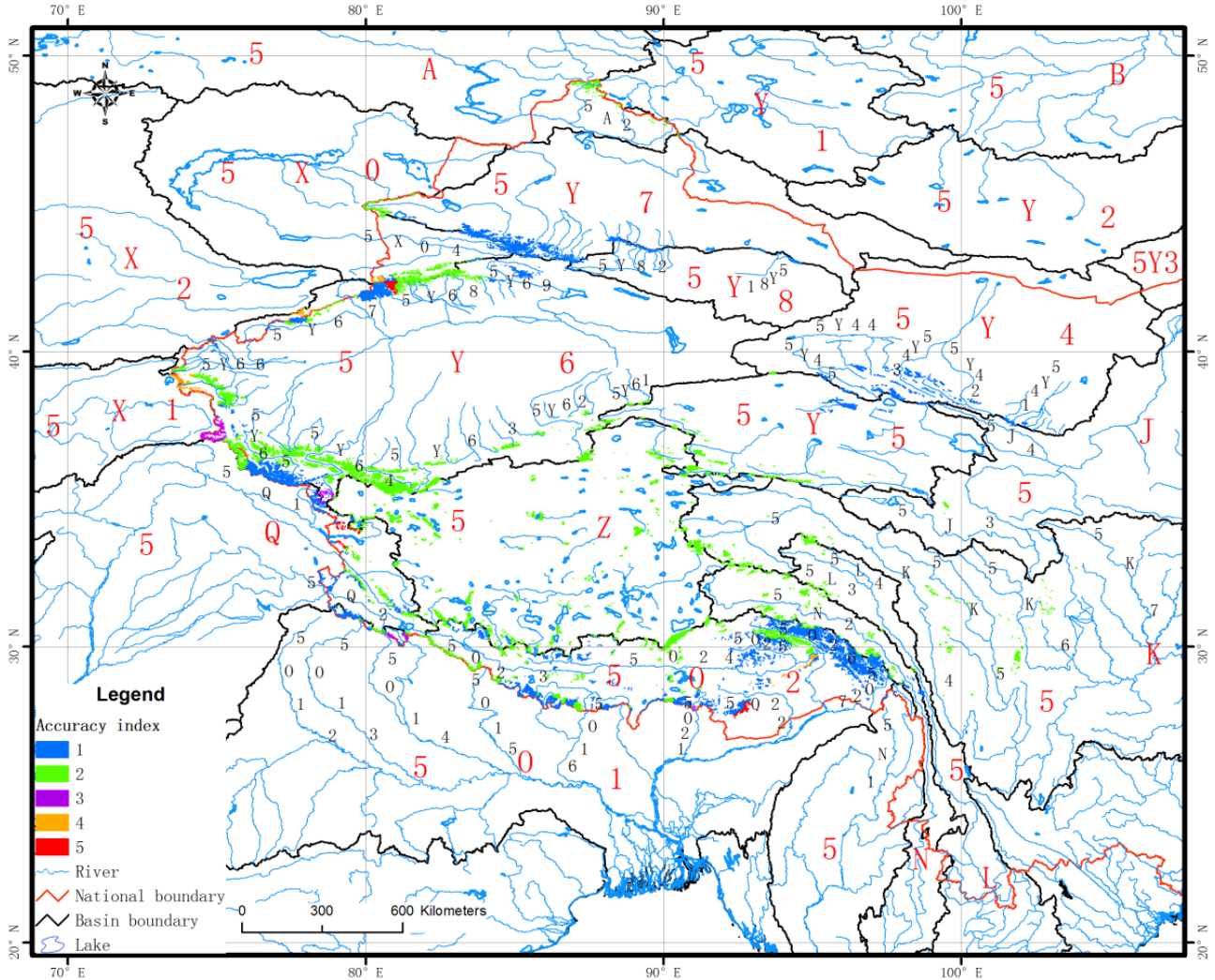


图 2 中国第一次冰川编目冰川面积精度分布图

除了主观因素外，参考资料类型和比例尺对冰川面积精度的影响最大。在重新收集的全国范围冰川编目工作底图的基础上，利用 GIS 方法选取 2000 个样本重新测量了冰川的面积，并与《中国冰川目录》中的冰川面积进行了对比，将冰川面积精度划分为 5 类，其面积精度分布图见图 2。

1: 根据 1:5 万地形图测量，并经过 1:5 万航片的校正，当冰川面积 $>0.2\text{ km}^2$ 时，平均面积误差 $<\pm 10\%$ ；

2: 根据 1:10 万地形图测量，并经过 1:5 万或 1:10 万航片校正，当冰川面积 $>0.5\text{ km}^2$ 时，平均面积误差 $<\pm 10\%$ ；

3: 根据 1:5 万或 1:10 万地形图测量，并经过 Landsat MSS 卫星像片校正，当面积 $>0.5\text{ km}^2$ 时，平均面积误差 $\leq \pm 10\%$ ，但存在较多的错误；

4: 根据 1:5 万或 1:10 万地形图测量，没有经过任何校正，当冰川面积大于 0.5 km^2 时，平均面积误差也在 $\pm 10\%$ 以内，但错误明显多于类型 3；

5: 冰川面积从 Landsat MSS 卫星像片或 1:20 万比例尺的地形图上测量获得, 没有经过其他验证, 冰川面积精度最低, 且存在明显的错误。

B. 冰川空间位置精度

影响冰川空间位置精度的要素既包括原始参考资料的类型和比例尺, 也包括数据化过程。世界冰川编目是在 19 世纪 70 年代组织发起的, 根据当时的资料条件, 世界冰川编目规范推荐使用地形图和航空相片作为冰川编目参考资料, 要求地形图的比例尺不得小于 1: 25 万。中国冰川编目大部分都是基于 1: 5 万和 1: 10 万地形图完成的, 总共使用了超过 2000 幅地形图和 342 000 幅 1950s-1980s 拍摄的航片, 以及 200 余幅 1970s-1980s 的 Landsat MSS 卫星图片。图 2 虽然是对冰川面积进行的评价, 但也一定程度上可以反映冰川编目参考资料的空间位置精度。

由于几何校正精度较低, 数字化的冰川分布图存在较多的问题, 在具体应用时需要注意:

1) 由于几何校正精度低和地图参照系的不同, 数字化冰川分布图不能与现有的 DEM 数据(如 SRTM 和 ASTER GDEM) 进行很好的匹配, 不能直接利用没有经过二次校正的数字化冰川分布图分析冰川的高度分布特征;

2) 冰川分布图无论在原始绘制的过程中还是在数字化的过程都引入了一些误差, 这些误差对冰川变化分析影响显著。建议不要直接使用数字化的冰川分布图与其他时期的冰川分布图进行比较, 可使用同时代的地形图、航片或卫星图像校正后再进行比较。

C. 冰川编目代表的冰川状态时间

冰川编目所代表的冰川状态时间对于冰川变化分析十分重要, 冰川编目中有两项时间, 分别是地形图的出版年代和航空相片的拍摄时间, 根据冰川编目规范, 航空相片的拍摄时间更能代表冰川编目时的冰川状态时间。因为冰川编目所使用的数据源不统一, 很难用单一时间来简单说明第一次冰川编目所代表的冰川状态时间, 图 3 是根据数据汇总而来的冰川编目所采用的航空相片的分布图, 从图中可以看出主要的航空相片分布在 1963、1964、1968、1969、1970、1974、1976 和 1980 年, 其中根据 1968-1970 的航片调查的冰川占全部冰川的 40%。

此外, 冰川编目参考资料信息是对五级流域内的所有冰川进行统一说明的。对于较小的流域, 这些信息基本能反映所有冰川的状态时间, 但对较大的流域, 由于涉及的数据源较多, 可能无法准确描述冰川的状态时间。建议在进行冰川变化分析时进行进一步的确认。

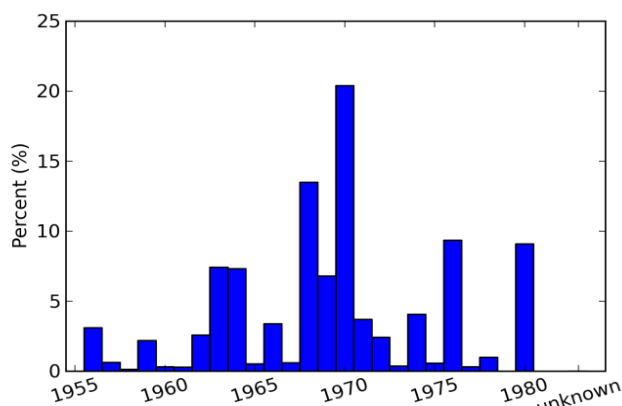


图 3 冰川编目代表的冰川状态时间分布图

V. 结论

经过系列质量控制的中国第一次冰川编目数据库能够较好的反映《中国冰川目录》的成果水准, 可满足冰川资源评估和冰川水资源评估的各种需求。

数字化冰川分布图和《中国冰川目录》中冰川面积存在差异, 建议在使用过程中将后者作为冰川的真实面积, 其面积误差基本都在 10% 以内。

数字化冰川分布图在绘制和数字化过程中引入了一系列的误差, 建议不要直接使用第一次冰川编目数据库中的数字化冰川分布图进行冰川变化分析。

第一次冰川编目获得的各种冰川参数与当前基于遥感和 GIS 方法获得的冰川参数有诸多不同, 在进行冰川变化分析时, 不能想当然的进行直接比较, 需要认真理解这些参数在定义和测量方法上的差异。

致谢

本文得到基金委项目(40925004)和科技部基础性工作专项(2006FY110200)的联合资助。

REFERENCES

- [1] W. Haeberli, "Glacier fluctuations and climate change detection - operational elements of a worldwide monitoring strategy," WMO Bulletin, vol. 44, pp. 23-31, 1995.
- [2] M. B. Dyurgerov and M. F. Meier, "Twentieth century climate change: Evidence from small glaciers," Proceedings of the National Academy of Sciences, vol. 97, pp. 1406-1411, 2000.
- [3] Yafeng. Shi, Concise Glacier Inventory of China. Shanghai: Shanghai Popular Science Press, 2008.
- [4] Yafeng. Shi, Chaohai. Liu and Ersi. Kang, "The Glacier inventory of China," Annals of Glaciology, vol. 50, pp. 1-4, 2009.
- [5] Lizong. Wu and Xin. Li, China Glacier Information System. Beijing: Ocean Press of China, 2004. (In Chinese)
- [6] F. Müller, T. Cafilisch and G. Müller, "Instructions for Compilation and Assemblage of Data for a World Glacier Inventory," Temporal Technical Secretariat for the World Glacier Inventory (TTS/WGI/IAHS(ICS)/UNESCO), ETH Zürich 2007-12-19 1977