

青藏高原积雪数据集 (V1.0)

评估报告

冯爱霞

国家气象信息中心资料服务室
2020 年 02 月

目录

一、数据集基本情况	1
(一) 数据源	1
(二) 时空范围	1
(三) 数据内容	1
二、数据完整性分析	1
二、数据质量评估	3
(一) 数据质量控制方案	3
1、质量控制码及含义	4
2、质量控制内容	4
3、质量控制方法	4
(二) 质量控制结果	5
1、数据的完整性	5
2、允许值范围检查结果	7
3、主要变化范围检查结果	7
4、内部一致性检查	7
5、质量控制的总体结果	8
参考文献	9

一、数据集基本情况

（一）数据源

目前我国已建成包含中国基本、基准气象站、一般气象站在内的 2479 个地面气象站。气象要素包括气压、气温、相对湿度、积雪、冻土、地温等。本数据集来源于《中国地面气象要素日值数据集》中的积雪要素。

（二）时空范围

- （1）空间范围：20° N-40° N，75° E-105° E；
- （2）时间范围：1951 年 01 月-2018 年 12 月；
- （3）站点数目： 20° N-40° N，75° E-105° E 空间范围内 340 个国家级地面站，其地理分布如图 1 所示。

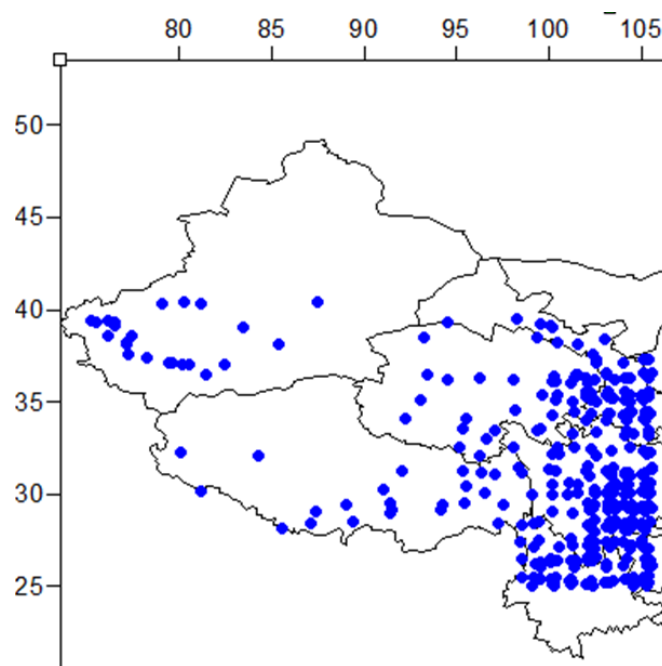


图 1 青藏高原积雪观测站空间分布

（三）数据内容

青藏高原积雪数据资料包含 2 个要素：日积雪深度和日雪压。根据《地面气象观测规范》要求，在符合观测雪深的日子，每天 08 时进行雪深观测，雪深不足 0.5cm 记 0；若雪深达 5cm 或以上时，进行雪压观测^[1]。

二、数据完整性分析

用实有率评估积雪深度和雪压的完整性。统计量的计算方法如下：

$$\text{实有率} = \frac{\sum_{i=1}^N \text{实有观测数据量}_i}{\sum_{i=1}^N \text{数据总量}_i} \times 100\%$$

实有观测数据量_i表示数据集中第 i 站或者第 i 年某要素项数据非缺测或无观测任务或空白，即非“32766”、“-32766”、“32744”、“-32744”的数据量；数据总量表示_i第 i 站或者第 i 年某要素应观测的数据总量。积雪数据集分别从积雪要素站点的实有率和逐年的实有率展示积雪 2 各要素的时空完整性。积雪深度和雪压的各站实有率如图 2 和图 3 所示，从图中可知除新疆几个站外，其他站的雪深和雪压实有率都在 90%以上。

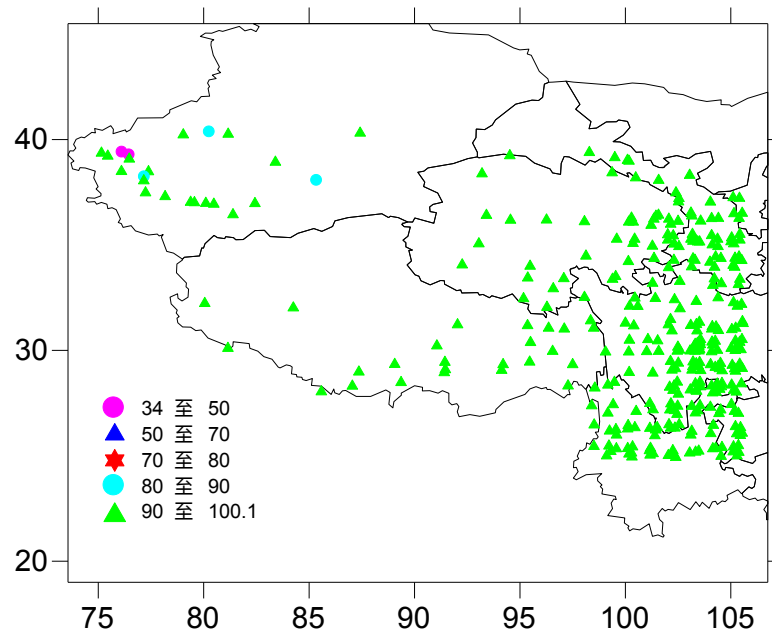


图 2 青藏高原积雪深度各站的实有率

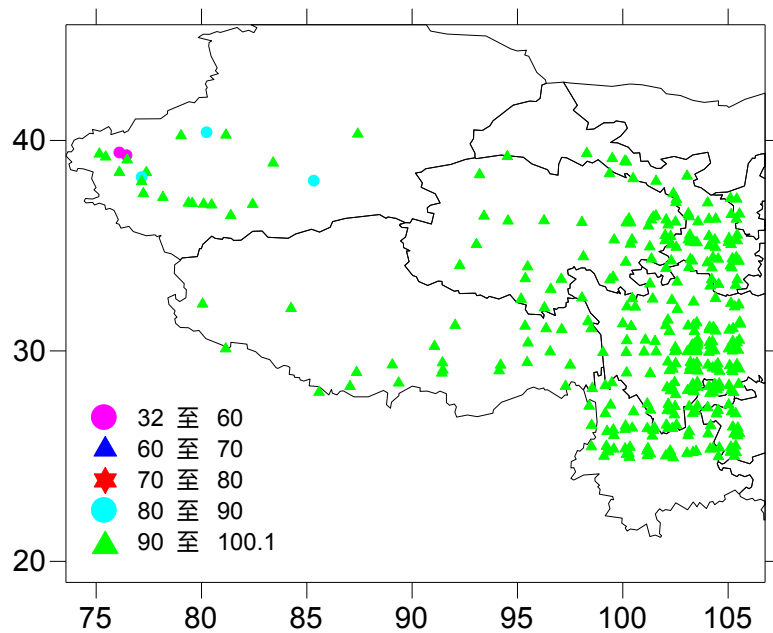


图3 青藏高原雪压各站的实有率

积雪深度和雪压的逐年实有率如图4所示，自1960年后雪深和雪压的实有率都分别在99%和98%以上。

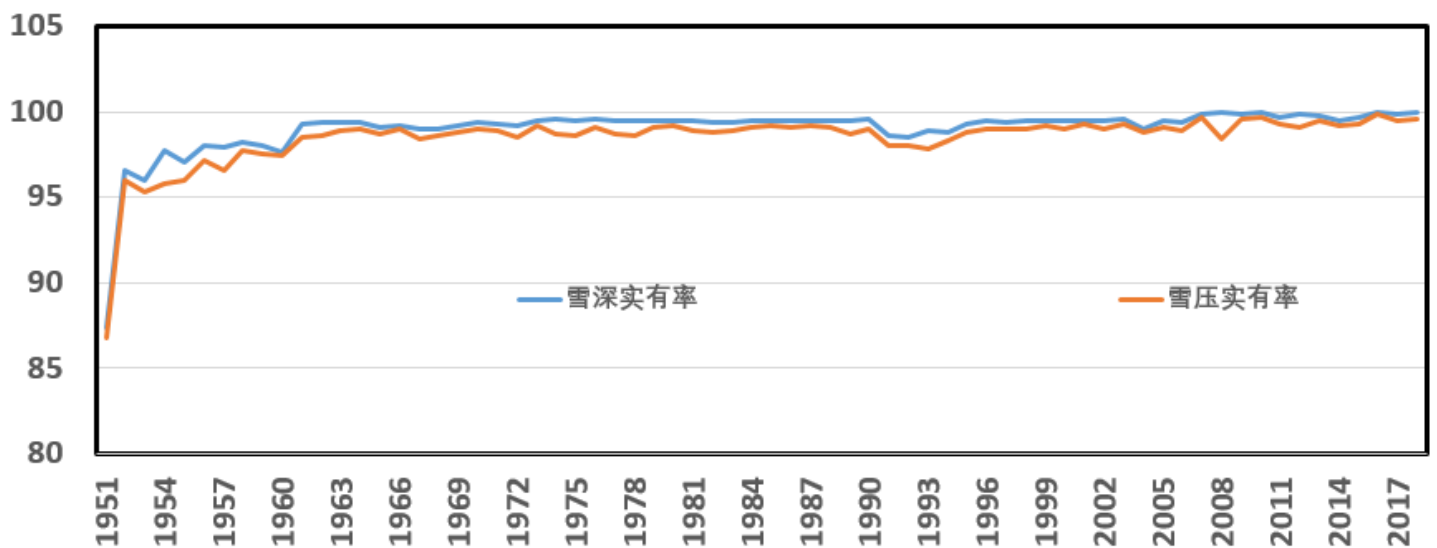


图4 1951-2018年雪深和雪压的逐年实有率

从积雪深度和积雪雪压的各站实有率和逐年实有率可知，积雪数据集的完整性较好。

二、数据质量评估

(一) 数据质量控制方案

1、质量控制码及含义

在对数据进行质量控制的过程中，随着质量控制过程的进行，需对被检验数据进行设置或者修改质量控制码，简称 QC 码，含义如下表。

表 1 数据质量控制码

质量控制码	含义
0	数据正确
1	数据可疑
2	数据错误
8	数据缺测或无观测
9	数据未进行质量控

2、质量控制内容

针对积雪的 2 个要素进行质量控制的内容包括缺测检查、允许值范围检查、主要变化范围检查、内部一致性检查、人工核查与更正。

3、质量控制方法

3.1 缺测检查

检查积雪 2 个要素是否为缺测数据。若为缺测数据，不再进行其他的检查，赋予该数据最终 QC 码为 8；反之赋予该数据初始 QC 码为 9，继续参与后续的质量检查。

3.2 允许值范围检查

检查积雪 2 个要素是否超过允许值范围^[2]。若超出允许值范围（2 个要素的允许值范围见表 2，不再进行其他的检查，赋予该数据最终 QC 码为 2；

表 2 积雪 2 个要素的允许值范围

要素	允许上限值	允许下限值	单位
----	-------	-------	----

积雪深度	300	0	cm
雪压	500	0	g/cm ²

3.3 主要变化范围检查

计算 2 个要素各个台站每日的标准差，以及逐站逐要素逐日的距平值。若距平值介于 3 倍标准差与 5 倍标准差之间则判识为可疑，若距平值大于 5 倍标准差则判识为错误^[3]。

3.4 内部一致性检查

通过上述步骤检查的数据参与内部一致性检查。

对于积雪，若雪深 $\geq 5\text{cm}$ ，则雪压应大于 0，否则判识雪深和雪压均可疑；若雪深小于 5cm 时，雪压应为 0，否则判识雪深和雪压均可疑。

3.5 人工核查与更正

对于上述步骤检查的疑误数据进行人工核查，核查后明确为错误的数据判识为错误数据且改其质控码为 2，明确为正确的数据改判其质控码为 0，无法明确质量信息的数据保留原检测结果。

（二）质量控制结果

1、数据的完整性

数据的缺测率可以表征数据的完整性，本评估报告从站点的多年平均缺测率和逐年的高原站点平均缺测率来揭示青藏高原积雪雪深和雪压的时空完整性情况。青藏高原各个站点雪深和雪压的多年平均缺测率如图 5 和图 6 所示。从图中可知除新疆几个站外，其他站的雪深和雪压缺测率都在 10%以下。

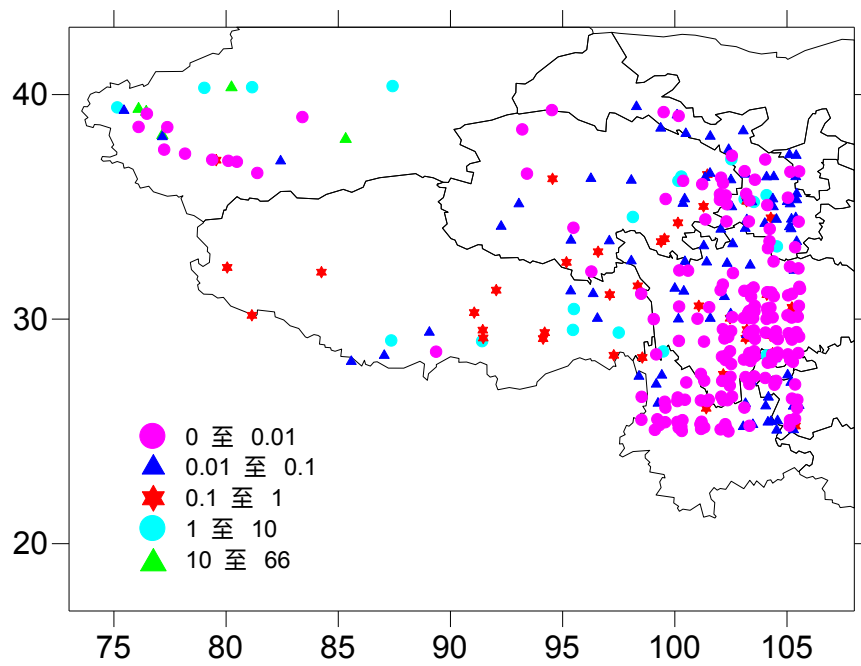


图5 青藏高原积雪深度各站的缺测率

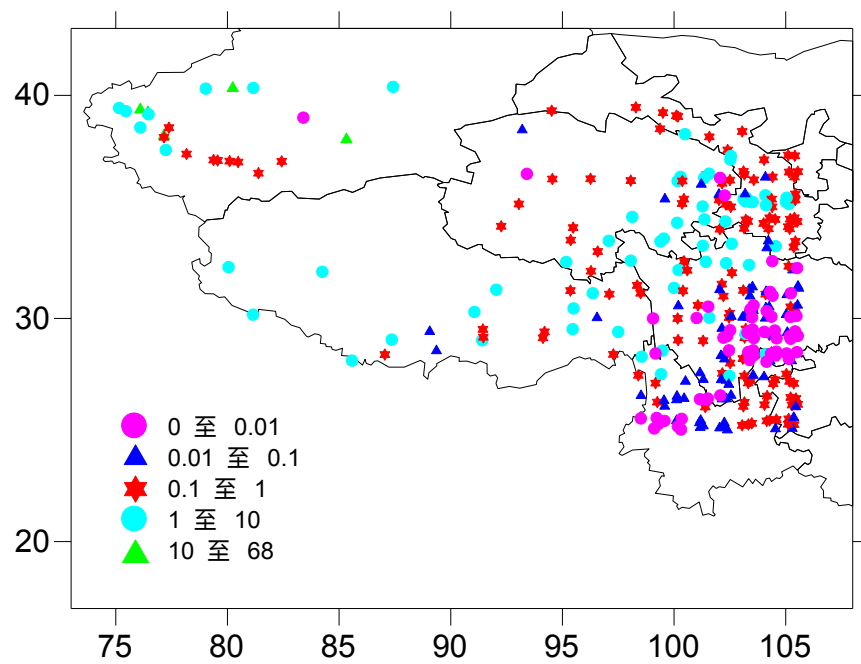


图6 青藏高原雪压各站的缺测率

青藏高原雪深和雪压逐年平均缺测率如图7所示。从图中可知，雪深缺测率自1960年以后都在1%以下，雪压则在2%以下，且雪深的缺测率一直小于雪压的缺测率。从积雪深度和雪压的各站缺测率和逐年缺测率可知，积雪数据集的完整性较好。

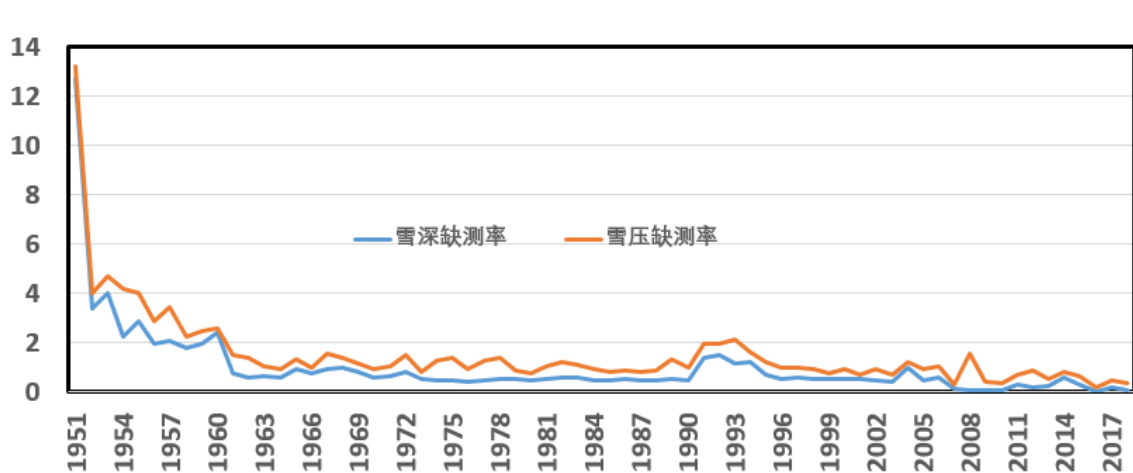


图 7 青藏高原雪深和雪压逐年的缺测率

2、允许值范围检查结果

根据积雪深度范围值 $[0, 300]$ ，积雪记录总数 7244397 条，未检查出错误数据。

根据积雪雪压范围值 $[0, 500]$ ，亦未检查出错误数据。

3、主要变化范围检查结果

根据积雪深度的距平值在 $(3\rho, 5\rho]$ 的数据共计 27231 条，占比 0.38%；积雪深度的距平值大于 5ρ 的数据共计 23805 条，占比 0.33%；雪压距平值在 $(3\rho, 5\rho]$ 的数据共计 1841 条，占比 0.03%；雪压的距平值大于 5ρ 的数据共计 5717 条，占比 0.08%。

4、内部一致性检查

对于积雪，若雪深 $\geq 5\text{cm}$ ，则雪压应大于 0，记录数据中根据此判识，数据均无可疑记录；若雪深小于 5cm 时，雪压应为 0，判识雪深和雪压均可疑的数据共计 4 条，具体如下：

表 5 雪压与雪深不一致记录的详细信息

站号	省份	站名	海拔高度（米）	年	月	日	雪深(厘米)	雪压(克/平方厘米)
52787	甘肃	乌鞘岭	3045.1	1955	4	5	2.0	3.0
52787	甘肃	乌鞘岭	3045.1	1955	4	25	2.0	12.0
56029	青海	玉树	3716.9	1955	3	18	2.0	3.0
56029	青海	玉树	3716.9	1955	3	28	1.0	2.0

经查阅较早的地面观测规范^[4]，可知在 1955 年雪压是不观测量，因此判断这 4 条雪压为错误数据，雪深判识为可疑数据。

5、质量控制的总体结果

经过质量控制，积雪深度和雪压的缺测、可疑以及错误情况如表 6 所示。

表 6 雪压与雪深数据质量情况一览表

要素	缺测记录数	缺测率	可疑记录数	可疑率	错误记录数	错误率
雪深	50563	0.7%	27235	0.38%	23810	0.33%
雪压	85654	1.2%	1846	0.03%	5732	0.08%

参考文献

- [1] 中国气象局. 地面气象观测规范. 北京: 气象出版社, 2003.
- [2] 江慧. 中国地面气象要素日值数据集评估报告. 内部文件.
- [3] 任芝花, 刘小宁, 杨文霞. 2005. 极端异常气象资料的综合性质量控制与分析, 气象学报, 63 (4): 526-533.
- [4] 中央气象局. 地面气象观测规范. 北京: 气象出版社, 1954.