

# 青藏高原土壤水分自动观测日数据集

## 评估报告

国家气象信息中心

2020 年 3 月

## 1. 概述

青藏高原是世界上海拔最高的高原，平均海拔超过 4000 m (Gao et al. 2016)，素有“世界屋脊”和“亚洲水塔”之称(Yang et al. 2003, Zhang et al. 2019)，广泛分布的冰川、冻土不仅是巨大的陆地淡水储存库，同时也是亚洲诸多大江大河的发源地(Piao et al. 2010, Yang et al. 2011)。在青藏高原上多种淡水存在的方式中，土壤含水量对于供应植被生长、调节局地小气候、反映局地水文水资源特征以及指示区域生态环境状况具有重要意义(Stich et al. 2003, Seneviratne et al. 2010)。长期以来，多个科研院所在青藏高原上建立、运行了多个野外观测台站。同时，中国气象局也从 2010 年开始逐步建立自动土壤水分观测站，进行逐小时土壤水分数据观测，目前观测站点已达 2000 余个。采用有效的质控方案对其进行质量控制和评估，建立青藏高原土壤水分自动观测日数据集，对于气候变化研究、生态环境监测以及政府决策等方面具有重大科学价值和现实需求。

## 2. 基本信息

### 2.1 数据源

本数据集包括青藏高原地区（25–40 °N，105 °E 以西）土壤水分自动站观测数据。观测要素为土壤体积含水量、土壤相对湿度、土壤重量含水率、土壤有效水分贮存量。要素层深为 10cm。本文分别从数据的预处理、质量控制方案、质量评估等三个方面对数据集进行了评估。

### 2.2 时空范围

数据集包括 2010–2018 年青藏高原地区（25–40 °N，105 °E 以西）土壤水分站共 284 个站的日值资料。

## 3. 数据预处理及统计方案

本数据集的预处理流程主要包括以下几个方面。（1）数据清洗。对收集到的数据进行数据清洗，包括对台站元数据的检查和修正、以及观测数据中的非法字符清洗、空数据文件清除等程序无法识别的各类错误。（2）规范名称。按照统一的青藏高原计划项目分析产品编码规则，对数据集中的数据文件进行了规范

的命名（详见数据集说明文档 SURF\_TIB\_SOIL\_DAY\_DOCU\_C.docx）。（3）缺测处理。四个要素都缺测的时次，进行了数据剔除，不出现在数据集中。

土壤水分日值资料统计方法：采用每天各小时土壤体积含水量（包括正确和可疑数据）进行算数平均得到土壤体积含水量日值（如果每天缺测或者质控码标记为错误 6 个小时以上数据，当日土壤体积含水量按缺测处理）。

## 4. 数据质量控制

本节简要介绍了土壤水分小时数据的质控方案，青藏高原土壤水分自动观测日数据集是基于正确和可疑的小时数据统计而来。本质控方案主要以土壤体积含水量为质控对象，其它三个要素数据首先直接采用对应层次的土壤体积含水量质控结果，最后进行界限值检查。

### 4.1 质控码含义

数据正确： QC 码=0

数据可疑： QC 码=1

数据错误： QC 码=2

数据缺测： QC 码=8

## 4.2 质控流程图

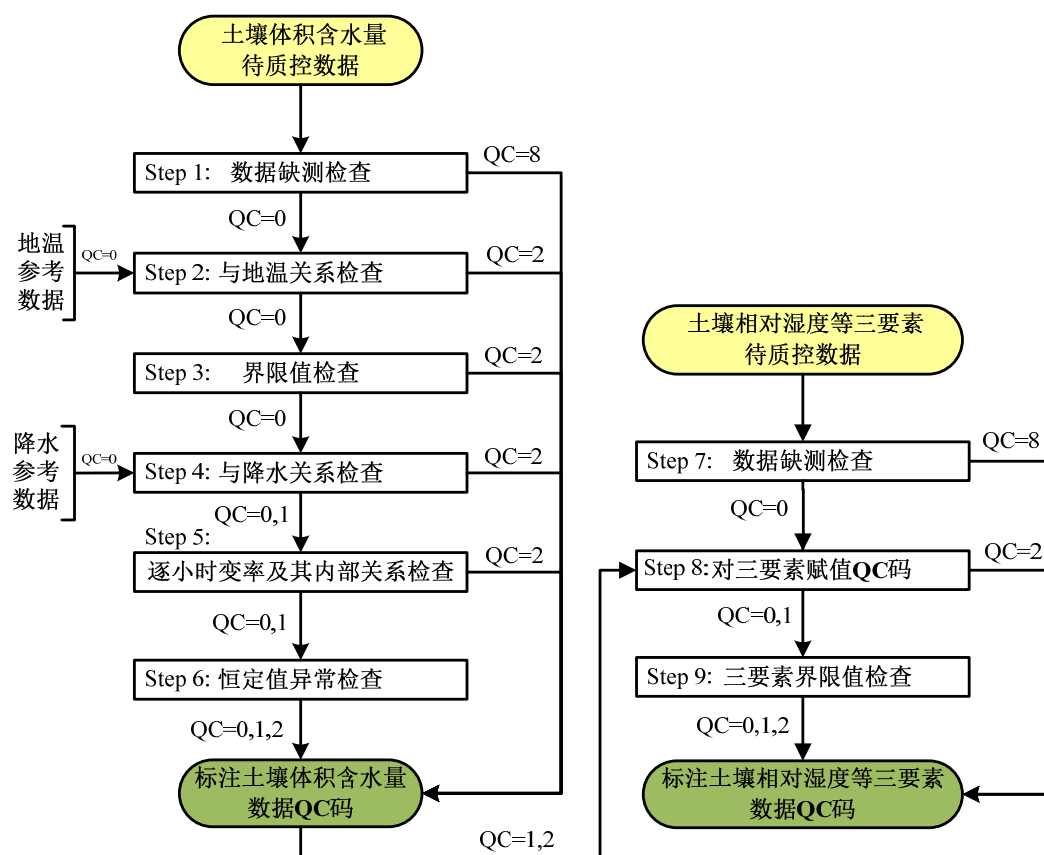


图 1 质量控制流程图

质控流程主要包括 9 个步骤（图 1），其中第 1、7 步为数据缺测检查，第 8 步为质控码赋值操作，其余各步为数据质量控制检查，最后对每个数据标注质量控制码。对于待质控数据，首先设置其对应的初始 QC 码=0，当数据缺测时 QC 码=8。然后根据质控流程的先后顺序进行逐步检查，并根据检查结果决定是否改变 QC 码。

## 4.2 相关参数说明

t: 表示质控时次；

L: 表示土壤层号， $L=1, 2, \dots, 20$ ，分别对应各观测土壤层；

RR(t) : 表示质控时次土壤水分观测站对应的小时降水量（单位：mm）；

Ts(L,t) : 表示质控时次土壤水分观测站第 L 层土壤对应的土壤温度（单位：

摄氏度)。

$Q(L,t)$ ：表示第  $L$  层土壤  $t$  时次的土壤体积含水量（单位：0.01g/cm<sup>3</sup>）；

$\Delta Q(L,t)$ ：表示第  $L$  层土壤  $t$  时次土壤体积含水量逐小时变率，通过  $\Delta Q(L,t)=Q(L,t)-Q(L,t-1)$  计算（单位：0.01g/cm<sup>3</sup>）；

$R(L,t)$ ：表示第  $L$  层土壤  $t$  时次土壤相对湿度（单位：% ，取整数）。

### 4.3 质量控制方案

#### 4.3.1 与地温关系检查

质控时次的地温数据 QC 码=0 时，进行本项检查。

检查方法为：对于某一确定时次，若某台站某层土壤其对应的地温数据  $\leq 0^{\circ}\text{C}$  时，则对应的土壤体积含水量判为错误，QC 码=2。

业务实施时，建议地温数据的选取采用如下方法：

以土壤水分观测站为中心，在 30Km 范围内选取距离最近的地面观测站对应地温数据（表 1）。

表 1 土壤对应地温数据选取方法

| 土壤水分观测深度 | 对应地温数据 |
|----------|--------|
| 0-10     | 5cm 地温 |

#### 4.3.2 界限值检查

本项检查适用于所有土壤层的土壤体积含水量。

检查方法为：对于某一确定时次，若某台站某层土壤其相应的土壤体积含水量观测值超过其对应的界限值范围（见表 2），则该数据判为错误，QC 码=2。

表 2 土壤体积含水量界限值范围

| 要素      | 理论下界                                    | 理论上界   |
|---------|---|--|
| 土壤体积含水量 | $\frac{w_k^{ij}}{1.5} \times \rho^{ij}$ | $\left[ \left( \frac{\rho^{ij}}{\rho_s} \right)^{\frac{1}{\rho_w}} \right] / \left( \frac{1}{\rho_w} \right)^{\frac{1}{\rho_w}} \times 100$<br><br>或 60*（0.01g/cm <sup>3</sup> ） |

\*注：若土壤体积含水量理论上界计算值大于 60（0.01g/cm<sup>3</sup>），则取 60 为上界。

表 2 中各参数说明：

$\rho^{ij}$ ：表示第 i 台站第 j 层土壤的土壤容重（单位：g/cm<sup>3</sup>）。

$w_k^{ij}$ ：表示第 i 台站第 j 层土壤的凋萎湿度（单位：% ，重量含水率表示）。

$\rho_w$ ：表示水的密度，方案采用定值 1.0 g/cm<sup>3</sup>。

$\rho_s$ ：表示土粒密度，对于各站各层土壤均采用恒定值 2.8 g/cm<sup>3</sup>。

### 4.3.3 与降水关系检查

#### （1）逐小时降水数据选取

与降水的关系检查包括 2 种方法，在方案实施时均需采用逐小时降水数据作为辅助数据。逐小时降水数据选取方法如下：

以自动土壤水分观测站为中心，在 10Km 范围内与距离最近的地面自动站进行匹配，若匹配成功则以地面观测站的逐小时降水数据作为土壤水分观测站对应的逐小时降水数据；若 10Km 范围内无地面自动站站，则不进行本项检查。

#### （2）表层土壤与降水关系检查

本项检查仅适用于 0~10cm 土壤层（L=1）的土壤体积含水量。

质控时次的小时降水数据 QC 码=0 时，进行本项检查。

对于表层土壤，其检查方法如下：

当  $RR(t)=0\text{mm}$ ，且  $\Delta Q(L,t)>0.5$ （0.01g/cm<sup>3</sup>）时，则该质控时次及前一时次的土壤体积含水量判为可疑，QC 码=1；

当  $0\text{mm}<RR(t)\leq 8.0\text{mm}$  且  $\Delta Q(L,t)>[RR(t-1)+RR(t)]$  时，则该质控时次及前一时次的土壤体积含水量判为可疑，QC 码=1；

当  $RR(t)>8.0\text{mm}$  且  $\Delta Q(L,t)>RR(t)$  时，则该质控时次及前一时次的土壤体积含水量判为可疑，QC 码=1。

### 4.3.4 逐小时变率及其内部关系检查

逐小时负变率检查：对于第 L 层土壤，当 $\Delta Q(L,t) < -0.5$ （0.01g/cm<sup>3</sup>）且前一时次土壤相对湿度满足关系： $R(L,t-1) < 98$ （%）时，则该质控时次及前一时次的土壤体积含水量判为可疑，QC 码=1。

#### 4.3.5 恒定值异常检查

本项检查会触发对历史土壤体积含水量质控码进行修改的操作，当质控码发生修改时，其对应时次的土壤相对湿度、土壤重量含水率、土壤有效水分贮存量三个要素质控码也应做相应修改，质控码修改规则：QC=2 的数据不可被修改；QC=0 或 1 的数据可被修改为 1 或 2。

对于某一台站的第 L 层土壤，若 $\Delta Q(L,t) \neq 0.0$  时，则进行本项检查。

从质控时次的前一时次（t-1）开始向前推，若连续多个时次土壤体积含水量恒定不变（连续时次总数记为 N），则对于该层土壤的观测数据，根据以下方法依次进行检查：

当  $Q(L,t-N)$  或  $Q(L,t-1)$  对应 QC 码=2 时，则 t-N 至 t-1 时次范围内体积含水量均判为错误，QC 码=2，恒定值异常检查结束；

当  $Q(L,t-N)$  或  $Q(L,t-1)$  对应 QC 码=1 时，则 t-N 至 t-1 时次范围内体积含水量均判为可疑，QC 码=1，恒定值异常检查结束；

当  $Q(L,t-N)$  及  $Q(L,t-1)$  对应 QC 码=0 时，取 t-N 至 t-1 时次范围内上相邻层（L-1）及下相邻层（L+1）土壤相对湿度数据作为参考数据序列，分别用  $R_i^{up}$ 、 $R_i^{dn}$  表示（ $i=1,2, \dots, N$ ）。

对于序列  $R_i^{up}$ ，首先对其进行由小到大排序；其次，令  $m=[N/2]$ ，其中符号[] 表示向下取整；那么，令  $A_{up}$  为序列  $R_j^{up}$ （ $j=1,2, \dots, m$ ）的算术平均值，令  $B_{up}$  为序列  $R_j^{up}$ （ $j=m+1, m+2, \dots, N$ ）的算术平均值，即：
$$A_{up} = \sum_{i=1}^m R_i^{up} / m, B_{up} = \sum_{i=m+1}^N R_i^{up} / (N-m)$$
。

对于序列  $R_i^{dn}$ ，同样可以得到  $A_{dn}$  和  $B_{dn}$ 。

令  $C_{max}$  为  $R(L,t-1)$  与数组  $[A_{up}, B_{up}, A_{dn}, B_{dn}]$  的差值的绝对值的最大值，

即：Cmax=max(|R(L,t-1)-[Aup, Bup, Adn, Bdn]|)。

判定方法：若 R(L,t-1)，Cmax 及连续时次数 N 满足表 3 任意一组条件，则 t-N 至 t-1 时次范围内质控层（L 层）土壤体积含水量均判为可疑，QC 码=1，恒定异常值检查结束。

表 3 恒定值检查参数范围

| 序号 | R(L,t-1)的范围 | Cmax 的范围 | N 的范围 |
|----|-------------|----------|-------|
| 1  | (0%， 50%]   | >10%     | ≥60   |
| 2  | (50%， 100%] | >20%     | ≥60   |
| 3  | (100%， +∞]  | >10%     | ≥60   |

#### 4.3.6 土壤相对湿度等三要素界限值检查

本项检查适用于所有土壤层的土壤相对湿度、土壤重量含水率、土壤有效水分贮存量三个要素。当质控时次（t 时次）土壤相对湿度等三个要素采用对应层次的土壤体积含水量质控结果后，且质控码为 0 或 1 时进行本项检查（表 4）。

表 4 土壤相对湿度等三要素判定条件

| 要素        | 界值范围          | 对应 QC 码 |
|-----------|---------------|---------|
| 土壤相对湿度    | >150% 且 ≤180% | QC=1    |
|           | <7% 或 >180%   | QC=2    |
| 土壤重量含水率   | >40% 且 ≤50%   | QC=1    |
|           | <1.2% 或 >50%  | QC=2    |
| 土壤有效水分贮存量 | >40mm 且 ≤50mm | QC=1    |
|           | <0mm 或 >50mm  | QC=2    |

检查方法为：对于某一确定时次，若某台站某层土壤其相应的土壤相对湿度、土壤重量含水率、土壤有效水分贮存量满足表 5 所列条件，则该数据判为可疑或错误。



## 5. 质量评估

### 5.1 评估指标

用数据缺测率指标评估青藏高原土壤水分自动站观测资料的数据完整性。统计方法如下：假设有  $n$  个站进行观测资料数据质量统计，其中第  $i$  个站 ( $i=1, 2, \dots, n$ ) 在观测运行期间的数据个数为应有数据个数  $i$ ，对通过质量检查的数据个数  $i$ 、缺测数据个数  $i$  进行统计，则该站观测资料数据统计结果为：

$$\text{数据缺测率}_i = \frac{\text{缺测数据个数}_i}{\text{应有数据个数}_i} \times 100\% \quad (1)$$

$n$  个站观测资料数据统计结果为：

$$\overline{\text{数据缺测率}} = \frac{\sum_{i=1}^n \text{缺测数据个数}_i}{\sum_{i=1}^n \text{应有数据个数}_i} \times 100\% \quad (2)$$

### 5.2 评估结果

#### 4.2.1 数据完整性评估

本文采用两种方法评估数据完整性，第一种：假如将 2010-2018 年共 3287 天都定为应观测数据日，那么 2010-2018 年青藏高原地区土壤体积含水量、相对湿度、重量含水率、有效水分贮存量平均缺测率分别为 37.5%、39.9%、39.6%、44.5%。显然，这样计算缺测较高，由于部分台站开始观测时间较晚或者冬季寒冷地区台站不进行土壤水分观测，所以这些数据作为应观测数据会增大平均缺测率。第二种（本文主要采用以下方法计算数据完整性）：将 2010-2018 年共 3287 天中有土壤水分观测的日期（四个要素中，至少有一个要素不为缺测）定为应观测数据日，那么 2010-2018 年青藏高原地区土壤体积含水量、相对湿度、重量含水率、有效水分贮存量平均缺测率分别为 0.0%、4.1%、3.7%、10.9%。

#### 4.2.2 数据质量评估

对逐小时的土壤水分四个要素数据进行了详细的质量控制,每个数据标有质控码,在统计各要素日值的过程中,只选取了质量较为可靠(正确和可疑)的逐小时数据,因此计算出的日值数据质控码均为正确数据。

## 参考文献

- Gao, Y. H., D. J. Cooper and X. X. Ma (2016). Phosphorus additions have no impact on plant biomass or soil nitrogen in an alpine meadow on the Qinghai-Tibetan Plateau, China. *Applied Soil Ecology*, 106: 18-23.
- Piao, S., P. Ciais, Y. Huang, Z. Shen, S. Peng, J. Li, L. Zhou, H. Liu, Y. Ma, Y. Ding, P. Friedlingstein, C. Liu, K. Tan, Y. Yu, T. Zhang and J. Fang (2010). The impacts of climate change on water resources and agriculture in China. *Nature*, 467(7311): 43-51.
- Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky and A. J. Teuling (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99(3-4): 125-161.
- Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. O. Kaplan, S. Levis, W. Lucht, M. T. Sykes, K. Thonicke and S. Venevsky (2003). Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, 9(2): 161-185.
- Yang, M. X., T. D. Yao, X. H. Gou, T. Koike and Y. Q. He (2003). The soil moisture distribution, thawing-freezing processes and their effects on the seasonal transition on the Qinghai-Xizang (Tibetan) plateau. *Journal of Asian Earth Sciences*, 21(5): 457-465.
- Zhang, Q., K. K. Fan, V. P. Singh, C. Q. Song, C. Y. Xu and P. Sun (2019). Is Himalayan-Tibetan Plateau "drying"? Historical estimations and future trends of surface soil moisture. *Science of the Total Environment*, 658: 374-384.