

中国气象科学研究院

庐山云雾观测数据集建设项目

雨量筒

数据使用说明文档

中国气象科学研究院

成都信息工程大学

2021 年 12 月

目录

1 概述.....	1
2 文件头信息说明.....	1
3 存储格式说明.....	3
3.1 nc 格式存储说明.....	3
3.1.1 维度信息.....	4
3.1.2 变量和属性信息.....	4
3.1.3 nc 存储示例.....	6
3.2 csv 格式存储说明.....	7
3.2.1 文件头描述信息.....	7
3.2.2 要素代码.....	7
3.2.3 csv 存储示例.....	8
4 数据读取说明.....	8
4.1 nc 格式文件读取.....	9
4.1.1 组（groups）的定位与读取.....	9
4.1.2 文件头信息读取.....	10
4.1.3 观测要素信息读取.....	11
4.2 csv 格式文件读取.....	11
4.2.1 文件头信息读取.....	11
4.2.2 观测要素信息读取.....	12
4.3 Station_level 和质量控制码说明.....	14
4.3.1 Station_level 说明.....	14
4.3.2 质量控制码说明.....	14

1 概述

雨量筒数据集包括 netCDF4 格式数据和 csv 文本数据两种，数据文件内容包括文件头和数据实体两部分，数据实体包括观测数据和相应的质量控制信息。先存储文件头，即描述信息和要素代码，再存储数据实体，即观测数据和质量控制信息。一个数据文件存放着设备当天的观测数据。

2 文件头信息说明

雨量筒的文件头包括描述信息和要素代码两部分，具体信息见表 1。

表 1 雨量筒文件头信息

序号	要素代码	代码全称	要素名称	单位	备注
1	Station_name	Station name	站名	—	描述信息
2	Country	Country	国家	—	描述信息
3	Province	Province	省份	—	描述信息
4	City	City	地市	—	描述信息
5	County	County	区县	—	描述信息
6	Station_ID	Station identity	区站号	—	描述信息
7	LAT	Latitude	纬度	°（度）	描述信息
8	LON	Longitude	经度	°（度）	描述信息
9	ALT	Altitude	测站海拔高度	m（米）	描述信息
10	Station_type	Station type	测站类型	—	描述信息
11	Station_level	Station level	测站级别	—	描述信息
12	Admi_code_CHN	Administrative area code of China	行政区代码	—	描述信息
13	Mete_data_code	Meteorological data code	资料代码	—	描述信息

14	Manufacturer_model	Manufacturer and model	厂家代码	-	描述信息
15	Software_version	Software version	软件版本	-	描述信息
16	Prec_sens_HGT	Rain gauge height	雨量筒距地面高度	m (米)	描述信息
17	Data_level	Data level	数据级别	-	描述信息
18	Timezone	Timezone	时区	-	描述信息
19	Time_resolution	Time resolution	时间分辨率	s (秒)	描述信息
20	Obse_begi_DT	Observing beginning datetime	观测数据起始时间	yyyy-mm-dd hh:mm:ss	描述信息
21	Obse_end_DT	Observing ending datetime	观测数据终止时间	yyyy-mm-dd hh:mm:ss	描述信息
22	Data_crea_DT	Data creating datetime	数据创建时间	yyyy-mm-dd hh:mm:ss	描述信息
23	Dataset_version	Dataset version	数据集版本	-	描述信息
24	Datetime	Datetime	资料时间	yyyy-mm-dd hh:mm:ss	要素代码
25	Datetime_Temp	Datetime of temperature	温度时间	yyyy-mm-dd hh:mm:ss	要素代码
26	Temp	Temperature	温度	℃ (摄氏度)	要素代码
27	Datetime_Prec	Datetime of precipitation	降雨量时间	yyyy-mm-dd hh:mm:ss	要素代码
28	Prec	Precipitation	降雨量	mm (毫米)	要素代码
29	Datetime_oclock	Date and o' clock of temperature maximum and cumulative precipitation	整点时间	yyyy-mm-dd hh:mm:ss	要素代码
30	Temp_MAX	Temperature maximum	最大温度	℃ (摄氏度)	要素代码
31	Prec_cumu	Cumulative precipitation	累积降雨量	mm (毫米)	要素代码
32	Q_Temp	Quality control code of temperature	温度质控码	-	要素代码
33	Q_Prec	Quality control code of precipitation	降雨量质控码	-	要素代码
34	Q_Temp_MAX	Quality control code of temperature	最大温度质控码	-	要素代码

		maximum			
35	Q_Prec_cumu	Quality control code of cumulative precipitation	累积降雨量 质控码	-	要素代码

3 存储格式说明

3.1 nc 格式存储说明

基于 netCDF4.0 标准对文件头信息和数据实体按照树形目录分组存储，树形目录结构如图 1 所示。具体地，依据 netCDF4.0 特性，对文件头要素信息和观测要素信息进行分组（groups），共分为两个大组，分别是 file_information（文件头信息）和 observational_information（观测要素信息）；其中 file_information（文件头信息）又包含 station（站点信息）、instrument（设备信息）以及 data（数据信息）三个组。

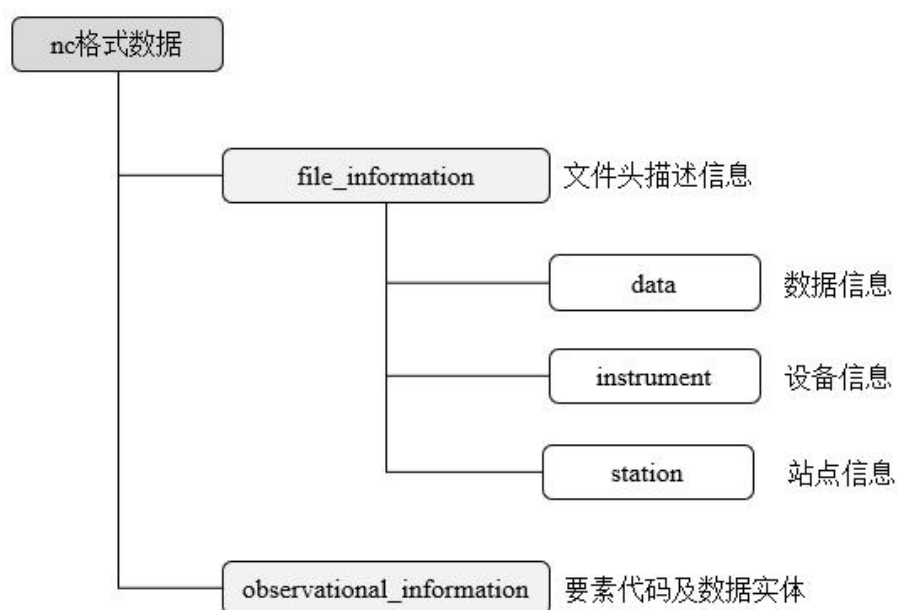


图 1 雨量筒数据的 nc 格式存储的树形目录结构

3.1.1 维度信息

雨量筒数据进行 nc 存储时的维度信息见表 2。

表 2 雨量筒 nc 存储的维度信息

序号	维名称	描述	值	备注
1	Datetime	时间	UNLIMITED (观测记录随时间的增加而增加)	/

3.1.2 变量和属性信息

雨量筒数据进行 nc 存储时的变量和属性信息见表 3。

表 3 雨量筒 nc 存储的变量和属性信息

序号	变量名	维度	数据类型	组信息
1	Station_name	1×1	string	/file_information/station
2	Country	1×1	string	/file_information/station
3	Province	1×1	string	/file_information/station
4	City	1×1	string	/file_information/station
5	County	1×1	string	/file_information/station
6	Station_ID	1×1	string	/file_information/station
7	LAT	1×1	float	/file_information/station
8	LON	1×1	float	/file_information/station
9	ALT	1×1	ushort	/file_information/station
10	Station_type	1×1	ubyte	/file_information/station
11	Station_level	1×1	string	/file_information/station
12	Admi_code_CHN	1×1	string	/file_information/station
13	Mete_data_code	1×1	string	/file_information/instrument/
14	Manufacturer_model	1×1	string	/file_information/instrument/
15	Software_version	1×1	string	/file_information/instrument/
16	Prec_sens_HGT	1×1	float	/file_information/instrument/
17	Data_level	1×1	string	/file_information/data/
18	Timezone	1×1	string	/file_information/data/
19	Time_resolution	1×1	string	/file_information/data/

20	Obse_begi_DT	1×1	string	/file_information/data/
21	Obse_end_DT	1×1	string	/file_information/data/
22	Data_crea_DT	1×1	string	/file_information/data/
23	Dataset_version	1×1	string	/file_information/data/
24	Datetime	Datetime ×1	string	/observational_information/
25	Datetime_Temp	Datetime ×1	string	/observational_information/
26	Temp	Datetime ×1	float	/observational_information/
27	Datetime_Prec	Datetime ×1	string	/observational_information/
28	Prec	Datetime ×1	float	/observational_information/
29	Datetime_oclock	Datetime ×1	string	/observational_information/
30	Temp_MAX	Datetime ×1	float	/observational_information/
31	Prec_cumu	Datetime ×1	float	/observational_information/
32	Q_Temp	Datetime ×1	ubyte	/observational_information/
33	Q_Prec	Datetime ×1	ubyte	/observational_information/
34	Q_Temp_MAX	Datetime ×1	ubyte	/observational_information/
35	Q_Prec_cumu	Datetime ×1	ubyte	/observational_information/

nc 存储时所用数据类型的信息见表 4。

表 4 nc 存储数据类型说明

数据类型	存储长度(单位: bit)	存储数据范围	精度
byte	8	[-128, 127]	—
ubyte	8	[0, 255]	—
short	16	[-32768, 32767]	—
ushort	16	[0, 65535]	—
int	32	[-2147483648, 2147483647]	—
uint	32	[0, 4294967295]	—
int64	64	[-9223372036854775808, 9223372036854775808]	—
uint64	64	[0, 18446744073709551615]	—
float	32	[-3.40E+38, 3.40E+38]	7 位
double	64	[-1.79E+308, 1.79E+308]	16 位
string	—	—	—

3.1.3 nc 存储示例

图 2 是在 HDFView 3.1.0 软件中打开雨量筒数据的 nc 存储格式文件后所显示的信息，根据分组信息以树状目录结构对 nc 存储格式文件中的数据进行展示。



图 2 雨量筒数据文件的 nc 格式存储示例

3.2 csv 格式存储说明

先存放文件头描述信息和要素代码，再按照时间顺序依照要素代码的既定顺序对数据实体进行逐行存储，各数据项间用“,”间隔。同时，文件头描述信息、观测要素与数据实体间用换行符进行区分，第 1 行为文件头信息，第 2 行为观测要素信息，第 3 行及之后为数据实体，逐行存储结构如图 3 所示。



图 3 雨量筒数据的 csv 文本格式的逐行存储结构

3.2.1 文件头描述信息

Lushan cloud and fog experiment station,China,Jiangxi,Jiujiang,Lushan scenic area,LSYWZ,29.57,115.97,1080,1,015,360400,PRE (Precipitation),HOB0 onset RG3-M,HOB0ware,1.2,LX,UTC+8,-,2020-01-01 01:00:00,2020-01-01 23:59:35,2021-12-12 00:22:14,1.0

3.2.2 要素代码

Datetime_Temp,Temp,Q_Temp,Datetime_Prec,Prec,Q_Prec,Datetime_oclock,Temp_MAX,Pre
ec_cumu,Q_Temp_MAX,Q_Prec_cumu

3.2.3 csv 存储示例

图 4 是在 Windows 操作系统自带的记事本软件（Notepad）中打开雨量筒数据的 csv 存储格式文件后所显示的信息，根据文件头描述信息、要素代码和数据实体逐行数据内容进行展示。

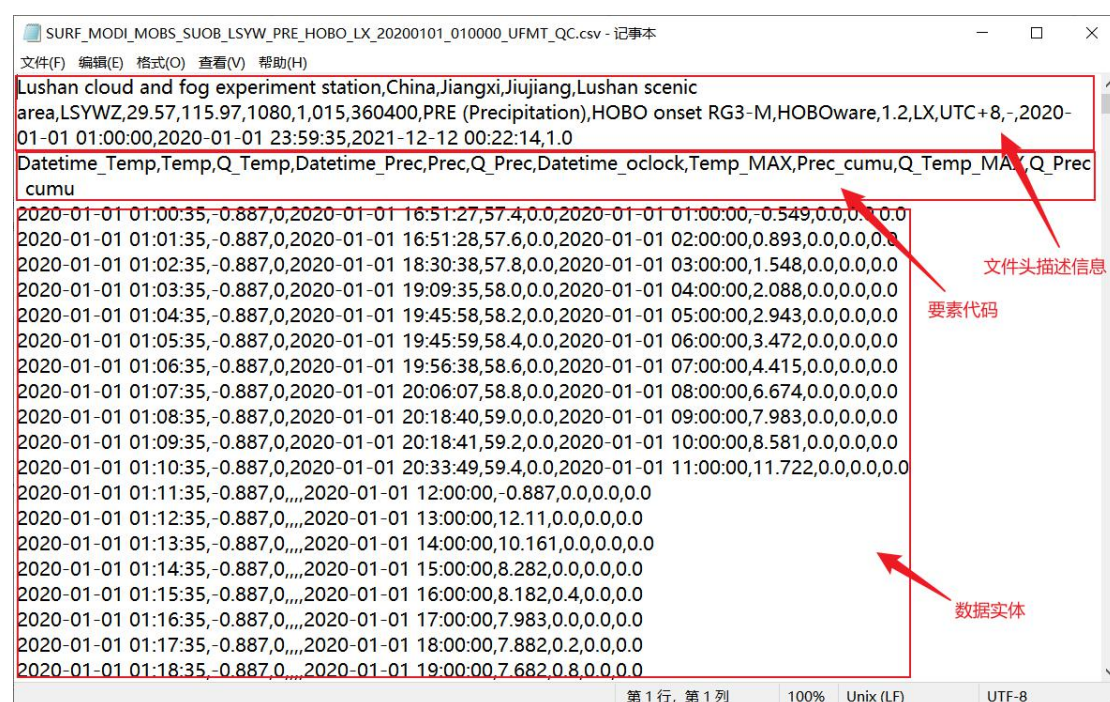


图 4 雨量筒数据文件的 csv 格式存储示例

4 数据读取说明

以 Python 语言为例对雨量筒数据集的 nc 格式数据和 csv 文本数据的读取使用进行说明，其中的示例代码可按从前往后的顺序运行，运行环境及配置信息如下：

- 语言环境：Python 3.8.12
- 运行环境：Windows 10 专业版 21H2

- IDE 环境: Visual Studio Code
- Python 工具包: pandas、numpy、netcdf4

其中 CSV 读取的 `filereader` 包是专门为本数据集编写, 在导入该工具包时, 应将此包复制到项目根目录中。

4.1 nc 格式文件读取

雨量筒数据集的 nc 存储, 是基于 netCDF4.0 按照文件头信息 (`file_information`) 和 观测要素信息 (`observational_information`) 进行分组 (`groups`) 并以树状目录结构进行数据的存储, 其中, 文件头信息 (`file_information`) 又分为了站点信息 (`station`)、设备信息 (`instrument`) 和数据信息 (`data`) 三个组。因此, 在进行数据的读取使用时, 也依据分组信息对文件头描述信息、观测要素信息和数据实体进行读取。

4.1.1 组 (`groups`) 的定位与读取

示例代码:

```
import netCDF4 as nc

# 打开文件名为“nc_demo.nc”的 nc 格式存储数据文件
nc_obj = nc.Dataset('nc_demo.nc')
# 查看当前状态的 groups 信息
print(nc_obj.groups.keys())
# 定位到 file_information 组
file_group = nc_obj.groups['file_information']
# 分别定位到 station 组、instrument 组以及 data 组
station_group = file_group.groups['station']
instrument_group = file_group.groups['instrument']
```

```
data_group = file_group.groups['data']
# 定位到 observational_information 组
obs_group = nc_obj.groups['observational_information']
# 查看 file_information 组下的分组
print(file_group.groups.keys())
```

示例代码运行结果：

```
dict_keys(['file_information', 'observational_information'])
dict_keys(['station', 'instrument', 'data'])
```

4.1.2 文件头信息读取

示例代码：

```
# 查看变量名
print(station_group.variables.keys())
print(instrument_group.variables.keys())
print(data_group.variables.keys())
```

示例代码运行结果：

```
dict_keys(['Station_name', 'Country', 'Province', 'City', 'County', 'Station_ID', 'LAT', 'LON',
'ALT', 'Station_type', 'Station_level', 'Admi_code_CHN'])

dict_keys(['Mete_data_code', 'Manufacturer_model', 'Software_version',
'Prec_sens_HGT'])

dict_keys(['Data_level', 'Timezone', 'Time_resolution', 'Obse_begi_DT', 'Obse_end_DT',
'Data_crea_DT', 'Dataset_version'])
```

示例代码：

```
# 查看 Station_name 信息
station_name_var = station_group.variables['Station_name']
print(station_name_var[:])
print(station_name_var.long_name)
print(station_name_var.units)
```

示例代码运行结果：

```
Lushan cloud and fog experiment
station Station name
```

-

4.1.3 观测要素信息读取

示例代码：

```
# 查看变量名
print(obs_group.variables.keys())
# 查看 Temp 信息
datetime_var = obs_group.variables['Temp']
print(datetime_var[:])
print(datetime_var.long_name)
print(datetime_var.units)
```

示例代码运行结果：

```
dict_keys(['Datetime', 'Datetime_Temp', 'Temp', 'Q_Temp', 'Datetime_Prec', 'Prec',
'Q_Prec', 'Datetime_oclock', 'Temp_MAX', 'Q_Temp_MAX', 'Prec_cumu', 'Q_Prec_cumu'])

masked_array(data=[7.28 , 7.28 , 7.381, ..., 5.552, 5.552, 5.552], mask=False,
fill_value=1e+20, dtype=float32)
'Temperature'
'°C'
```

4.2 csv 格式文件读取

4.2.1 文件头信息读取

示例代码：

```
from filereader import CSVReader

reader = CSVReader(r'./csv_demo.csv')
data = reader.read()
print(data['header'])
```

示例代码运行结果：

```
{'Station_name': 'Lushan cloud and fog experiment station',
'Country': 'China',
'Province': 'Jiangxi',
'City': 'Jiujiang',
```

```
'County': 'Lushan scenic area',  
'Station_ID': 'LSYWZ',  
'LAT': '29.57',  
'LON': '115.97',  
'ALT': '1080',  
'Station_type': '1',  
'Station_level': '015',  
'Admi_code_CHN': '360400',  
'Mete_data_code': 'PRE (Precipitation)',  
'Manufacturer_model': 'HOBO onset RG3-M',  
'Software_version': 'HOBOWare',  
'Prec_sens_HGT': '1.2',  
'Data_level': 'LX',  
'Timezone': 'UTC+8',  
'Time_resolution': '-',  
'Obse_begi_DT': '2020-01-01 01:00:00',  
'Obse_end_DT': '2020-01-01 23:59:35',  
'Data_crea_DT': '2021-12-12 00:22:14',  
'Dataset_version': '1.0'}
```

4.2.2 观测要素信息读取

示例代码：

```
print(data['obs']['temp'])
```

示例代码运行结果：

```
      Datetime_Temp      Temp Q_Temp
0    2018-07-01 01:00:58  20.519      0
1    2018-07-01 01:01:58  20.519      0
2    2018-07-01 01:02:58  20.519      0
3    2018-07-01 01:03:58  20.519      0
4    2018-07-01 01:04:58  20.519      0
...
1375 2018-07-01 23:55:58  21.187      0
1376 2018-07-01 23:56:58  21.187      0
1377 2018-07-01 23:57:58  21.187      0
1378 2018-07-01 23:58:58  21.187      0
1379 2018-07-01 23:59:58  21.187      0

[1380 rows x 3 columns]
```

示例代码：

```
print(data['obs']['pre'])
```

示例代码运行结果：

```
      Datetime_Prec      Prec Q_Prec
0    2018-07-01 02:49:33   94.4    0.0
1    2018-07-01 02:55:12   94.6    0.0
2    2018-07-01 02:57:49   94.8    0.0
3    2018-07-01 03:02:00   95.0    0.0
4    2018-07-01 03:07:48   95.2    0.0
..
66   2018-07-01 11:14:21  107.6    0.0
67   2018-07-01 11:18:43  107.8    0.0
68   2018-07-01 11:58:28  108.0    0.0
69   2018-07-01 11:58:29  108.2    0.0
70   2018-07-01 12:11:26   94.2    0.0

[71 rows x 3 columns]
```

示例代码：

```
print(data['obs']['max'])
```

示例代码运行结果：

	Datetime_oclock	Temp_MAX	Prec_cumu	Q_Temp_MAX	Q_Prec_cumu
0	2018-07-01 01:00:00	20.519	0.0	0	0
1	2018-07-01 01:01:00	20.519	0.0	0	0
2	2018-07-01 01:02:00	20.519	0.0	0	0
3	2018-07-01 01:03:00	20.519	0.0	0	0
4	2018-07-01 01:04:00	20.519	0.0	0	0
...
1375	2018-07-01 23:55:00	21.187	0.0	0	0
1376	2018-07-01 23:56:00	21.187	0.0	0	0
1377	2018-07-01 23:57:00	21.187	0.0	0	0
1378	2018-07-01 23:58:00	21.187	0.0	0	0
1379	2018-07-01 23:59:00	21.187	0.0	0	0

[1380 rows x 5 columns]

4.3 Station_level 和质量控制码说明

4.3.1 Station_level 说明

代码 015 表示地面观测站中的其他气象站类别。其中，01 表示地面观测站（站网），5 表示其他气象站（站台级别）。

4.3.2 质量控制码说明

数据质量控制码的取值及含义见表 5。

表 5 质量控制码的标识/代码表

质量控 制码	描述	含义
0	数据正常	通过质量控制，未发现数据异常；或数据虽异常，但最终确认数据正确
1	数据可疑	通过质量控制，发现数据异常，且未明确数据正确还是错误
2	数据错误	通知质量控制，确认数据错误
3	数据为订正	原数据明显偏离真值，但在一定范围内可参照使用。在原始数据

	值	基础上通过偏差订正等方式重新获取的更正数据
4	数据为修改值	原数据因错误或缺测而完全不可用，通过与原数据完全无关的替代方式重新获取的更正数据
5	预留	
6	预留	
7	无观测任务	按规定，台站无相应要素数据观测任务
8	数据缺测	该项数据应观测，但因各种原因数据缺测
9	数据未做质量控制	该数据未进行质量控制

注：质控码 0、3、4 均当可信使用