

中国气象科学研究院

庐山云雾观测数据集建设项目

雾滴谱仪

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.....	8
3.2.1 文件头描述信息.....	8
3.2.2 要素代码.....	8
3.2.3 csv 存储示例.....	9
4 数据读取说明.....	9
4.1 nc 格式文件读取.....	10
4.1.1 组（groups）的定位与读取.....	10
4.1.2 文件头信息读取.....	11
4.1.3 观测要素信息读取.....	12
4.2 csv 格式文件读取.....	12
4.2.1 文件头信息读取.....	12
4.2.2 观测要素信息读取.....	13
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.....	15

1 概述

雾滴谱仪数据集包括 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

雾滴谱仪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 雾滴谱仪文件头信息

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	FSD_sens_HGT	Fog monitor height	雾滴谱仪距地面高度	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	Sample_area	Sample area	采样面积	mm ² (平方毫米)	描述信息
21	Channel_count	Channel count	分档数	-	描述信息
22	Size_each_bin	Size of each bin	每档尺度	μm (微米)	描述信息
23	Obse_begi_DT	Observing beginning datetime	观测数据起始时间	yyyy-mm-dd hh:mm:ss	描述信息
24	Obse_end_DT	Observing ending datetime	观测数据终止时间	yyyy-mm-dd hh:mm:ss	描述信息
25	Data_crea_DT	Data creating datetime	数据创建时间	yyyy-mm-dd hh:mm:ss	描述信息
26	Dataset_version	Dataset version	数据集版本	-	描述信息
27	Datetime	Datetime	资料时间	yyyy-mm-dd hh:mm:ss	要素代码
28	Temp_ambi	Ambient temperature	环境温度	°C (摄氏度)	要素代码
29	Temp_reco	Recovery temperature	恢复温度	°C (摄氏度)	要素代码
30	Reje_DOF	Reject particles that fall outside the FM 100' s depth of field	景深外粒子个数	-	要素代码
31	Numb_part_chan	Particles number per channel	分档粒子个数	-	要素代码
32	Numb_conc	Number concentration	数浓度	cm ⁻³ (个/立方厘米)	要素代码
33	Volu_conc	Volume concentration	体积浓度	μm ³ cm ⁻³ (立方微米/立方厘米)	要素代码
34	LWC	Liquid water	液态水含量	g m ⁻³ (克/立	要素代码

		content		方米)	
35	MVD	Median volume diameter	体积中值直径	μm (微米)	要素代码
36	ED	Effective diameter	有效直径	μm (微米)	要素代码
37	TAS	True air speed	真实的空气流速	m s^{-1} (米/秒)	要素代码
38	Status	Status	状态	-	要素代码
39	Q_data	Quality control code of data	数据质控符	-	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。

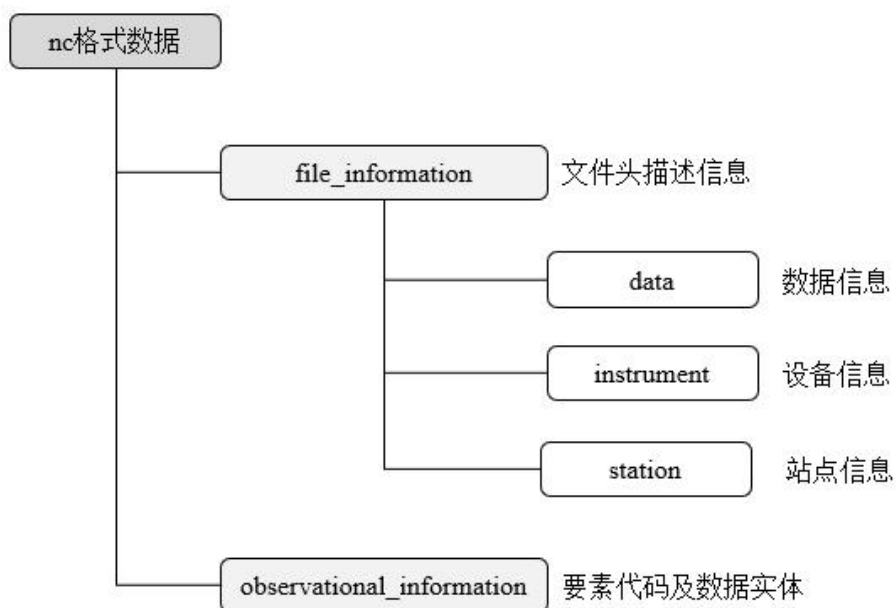


图 1 雾滴谱仪数据的 nc 格式存储的树形目录结构

3.1.1 维度信息

雾滴谱仪数据进行 nc 存储时的维度信息见表 2。

表 2 雾滴谱仪 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)	/
2	Dime_Numb_part_ch	分档数	20	2 4 6 8 10 12 14 16 18 20 23 26 29 32 35 38 41 44 47 50

3.1.2 变量和属性信息

雾滴谱仪数据进行 nc 存储时的变量和属性信息见表 3。

表 3 雾滴谱仪 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	string	/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	FSD_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	ubyte	/file_information/data/
20	Sample_area	1×1	float	/file_information/data/
21	Channel_count	1×1	ubyte	/file_information/data/
22	Size_each_bin	20×1	ubyte	/file_information/data/
23	Obse_begi_DT	1×1	string	/file_information/data/
24	Obse_end_DT	1×1	string	/file_information/data/
25	Data_crea_DT	1×1	string	/file_information/data/
26	Dataset_version	1×1	string	/file_information/data/
27	Datetime	Datetime ×1	string	/observational_information/
28	Temp_ambi	Datetime ×1	float	/observational_information/
29	Temp_reco	Datetime ×1	float	/observational_information/
30	Reje_DOF	Datetime ×1	ushort	/observational_information/
31	Numb_part_chan	Datetime ×	ushort	/observational_information/

		Dime_Numb_pa rt_ch		
32	Numb_conc	Datetime ×1	float	/observational_information/
33	Volu_conc	Datetime ×1	float	/observational_information/
34	LWC	Datetime ×1	float	/observational_information/
35	MVD	Datetime ×1	float	/observational_information/
36	ED	Datetime ×1	float	/observational_information/
37	TAS	Datetime ×1	float	/observational_information/
38	Status	Datetime ×1	ubyte	/observational_information/
39	Q_data	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 软件中打开雾滴谱仪数据的 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内容进行展示。



图 2 雾滴谱仪数据文件的 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 雾滴谱仪数据的 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,FSD (Fog droplet distribution),DMT FM-100,PADS3 (Old version) / PADS4 (New version),0.8,LX,UTC+8,1,0.24,20,2 4 6 8 10 12 14 16 18 20 23 26 29 32 35 38 41 44 47 50,2020-01-02 08:21:43,2020-01-02 23:59:59,2021-12-12 00:20:28,1.0

3.2.2 要素代码

Datetime,Temp_ambi,Temp_reco,Reje_DOF,Numb_conc,Volu_conc,LWC,MVD,ED,TAS,Status,

- 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

雾滴谱仪数据集的 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(r'./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',
'FSD_sens_HGT'])

dict_keys(['Data_level', 'Timezone', 'Time_resolution', 'Sample_area', 'Channel_count',
'Size_each_bin', '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_ambi 信息
datetime_var = obs_group.variables['Temp_ambi']
print(datetime_var[:])
print(datetime_var.long_name)
print(datetime_var.units)
```

示例代码运行结果：

```
dict_keys(['Datetime', 'Temp_ambi', 'Temp_reco', 'Reje_DOF', 'Numb_conc', 'Volu_conc',
'LWC', 'MVD', 'ED', 'TAS', 'Status', 'Q_data', 'Numb_part_chan'])

masked_array(data=[ nan, nan, nan, ..., 9.475428, 9.524253, 9.470342], mask=False,
fill_value=1e+20, dtype=float32)
'Ambient 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': 'FSD (Fog droplet distribution)',  
'Manufacturer_model': 'DMT FM-100',  
'Software_version': 'PADS3 (Old version) / PADS4 (New version)',  
'FSD_sens_HGT': '0.8',  
'Data_level': 'LX',  
'Timezone': 'UTC+8',  
'Time_resolution': '1',  
'Sample_area': '0.24',  
'Channel_count': '20',  
'Size_each_bin': ['2', '4', '6', '8', '10', '12', '14', '16', '18', '20', '23', '26', '29', '32', '35', '38',  
'41', '44', '47', '50'],  
'Obse_begi_DT': '2020-01-02 08:21:43',  
'Obse_end_DT': '2020-01-02 23:59:59',  
'Data_crea_DT': '2021-12-12 00:20:28',  
'Dataset_version': '1.0'}
```

4.2.2 观测要素信息读取

示例代码：

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


示例代码运行结果：

	Datetime	Temp_ambi	Temp_reco	Reje_DOF	Numb_conc	Volu_conc	\
0	2020-01-02 08:21:43				[]	Nan	
1	2020-01-02 08:21:44				[]	Nan	
2	2020-01-02 08:21:45				[]	Nan	
3	2020-01-02 08:21:46				[]	Nan	
4	2020-01-02 08:21:47				[]	Nan	
...
50410	2020-01-02 23:59:55	9.474321	9.560284	2875.0	[]	73.190605	
50411	2020-01-02 23:59:56	9.523148	9.609124	1762.0	[]	44.535122	
50412	2020-01-02 23:59:57	9.475428	9.560284	2313.0	[]	73.158531	
50413	2020-01-02 23:59:58	9.524253	9.609124	3033.0	[]	87.087265	
50414	2020-01-02 23:59:59	9.470342	9.560284	4011.0	[]	103.727341	
	LWC	MVD	ED	TAS	Status	Q_data	Numb_part_chan
0	NULL	Nan	Nan	Nan	Nan	Nan	8
1	NULL	Nan	Nan	Nan	Nan	Nan	8
2	NULL	Nan	Nan	Nan	Nan	Nan	8
3	NULL	Nan	Nan	Nan	Nan	Nan	8
4	NULL	Nan	Nan	Nan	Nan	Nan	8
...
50410	NULL	0.004598	7.536495	6.810345	13.098276	1	0
50411	NULL	0.002194	6.636115	5.772817	13.099407	1	0
50412	NULL	0.002193	5.222948	4.714823	13.013767	1	0
50413	NULL	0.004033	7.775458	6.07491	13.014891	1	0
50414	NULL	0.005966	8.694445	7.07424	13.398085	1	0
[50415 rows x 13 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	数据为订正值	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