

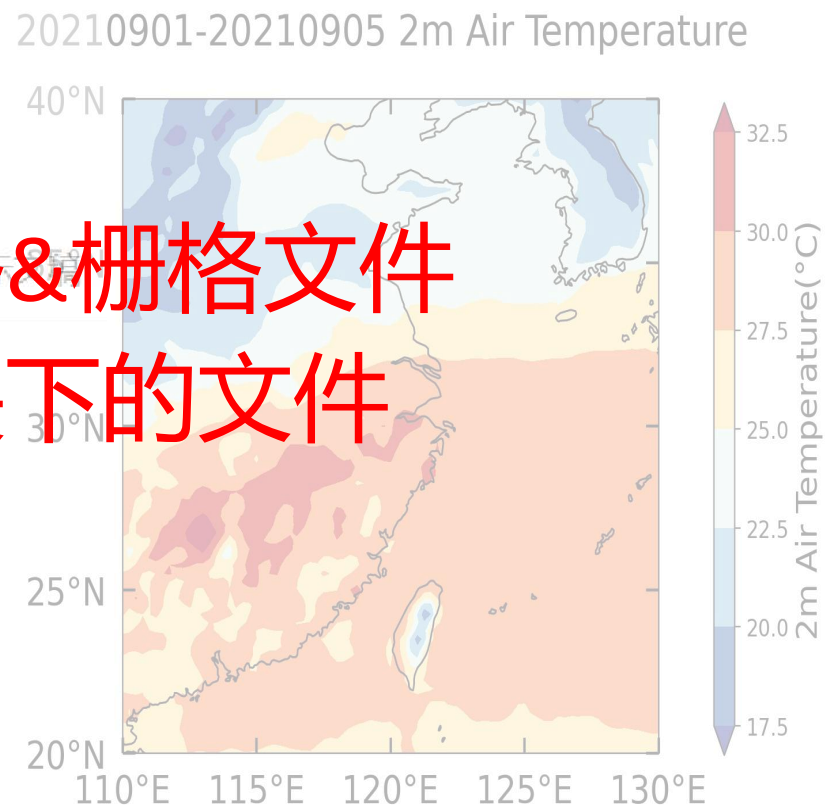
此电脑 > Data (D:) > bllibili > Python绘图技巧3

名称	修改日期	类型
----	------	----

day20210901	2022/8/10 19:21	文件夹
day20210902	2022/8/10 19:21	文件夹
table	2022/8/10 19:28	文件夹
air2m20210901.nc	2022/8/9 0:13	NC 文件
air2m20210902.nc	2022/8/9 0:14	NC 文件
air2m20210903.nc	2022/8/9 0:14	NC 文件
air2m20210904.nc	2022/8/9 0:14	NC 文件
air2m20210905.nc	2022/8/9 0:15	NC 文件
Python绘图技巧3: 批量读取文件.ps1x	2022/8/10 22:3	

Python绘图技巧3: 批量读取文件

批量读取文本文件&表格&栅格文件
批量读取多个子文件夹下的文件



➤ Python实现批量读取文件

Python中有很多库&函数都可以实现批量读取文件

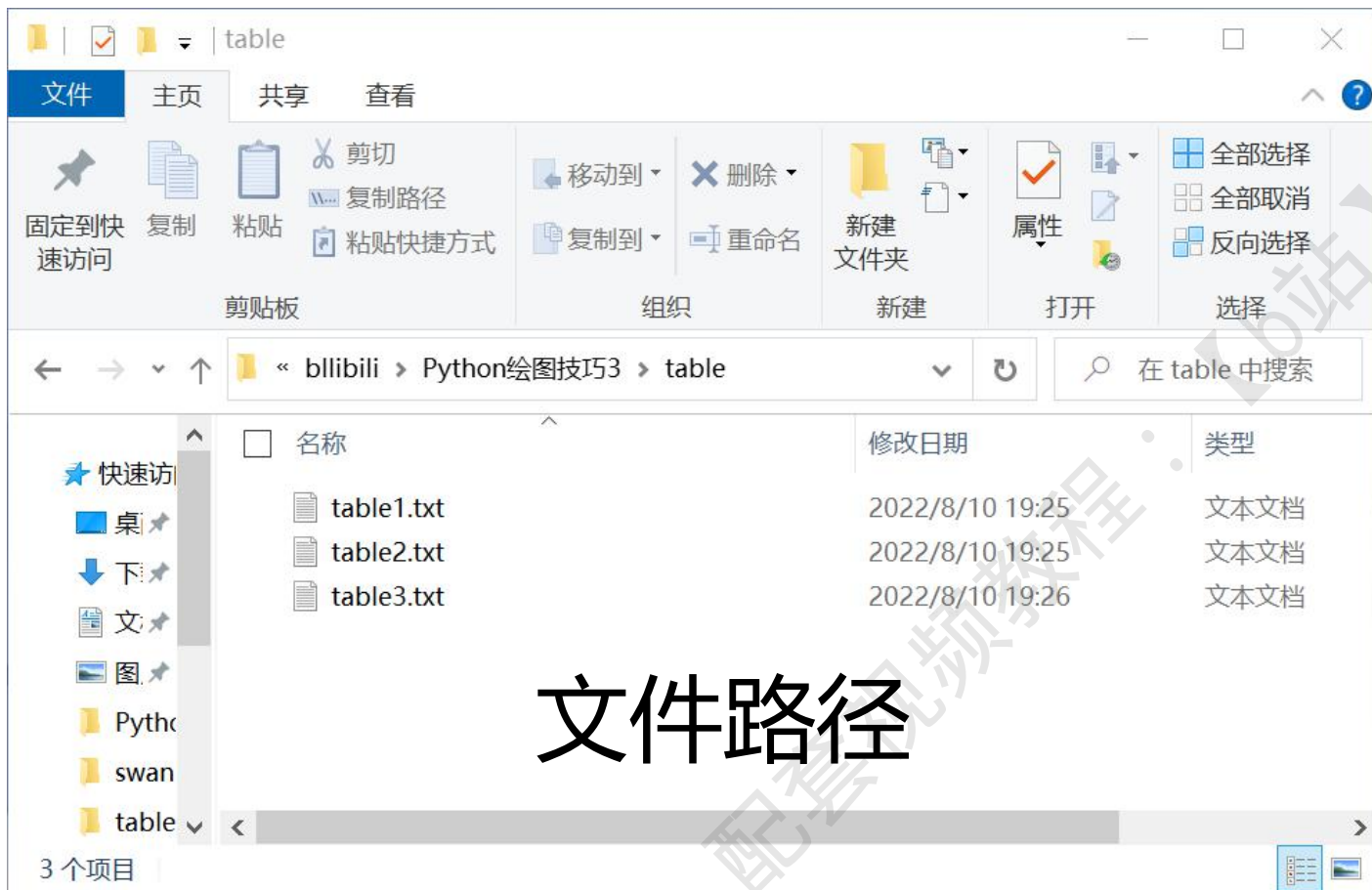
os

pathlib

up自己平时做数据分析&可视化常见的一些文件分类：

- 文本文件 .txt/.dat等 使用f.read()/readlines()等函数读取
最简单的文本文件
- 表格文件 .txt/.csv/.xls/.xlsx等 使用pandas相关函数读取
通常为变量的一维时间序列 (多个变量→二维表格)
- 栅格文件 .nc/.hdf/.grd等 使用netCDF4/xarray等第三方库
通常为变量的空间分布/时空分布 (二维及以上)

➤ 批量读取文本文件（简单例子）



方法1: os

```
import os
```

```
filepath = 'D:/bilibili/Python绘图技巧3/table'
filelist = os.listdir(filepath)
print('filelist: ', filelist)
```

获取文件列表

```
for file in filelist:
    with open(filepath + '/' + file) as f:
        data = f.read()
        print(data)
```

for循环挨个读取文件

```
filelist: ['table1.txt', 'table2.txt', 'table3.txt']
11111111111111111111
22222222222222222222
33333333333333333333
```

方法2: pathlib

```
from pathlib import Path
```

```
filepath = Path('D:/bilibili/Python绘图技巧3/table')
filelist = list(filepath.glob('*'))
print(filelist)
```

```
for file in filelist:
    with open(file) as f:
        data = f.read()
        print(data)
```

```
[WindowsPath('D:/bilibili/Python绘图技巧3/table/table1.txt'),
 ('D:/bilibili/Python绘图技巧3/table/table3.txt')]
11111111111111111111
22222222222222222222
33333333333333333333
```

这里输出路径截图没法截全

*注意一下这里的filelist的区别, os.listdir()获取的是文件名字, filepath.glob获取的是文件路径+名字

*with函数的作用: (1) 每次自己关闭文件 (2) 可进行异常处理

批量获取文件内容

```
from pathlib import Path

filepath = Path('D:/bilibili/Python绘图技巧3/table')
filelist = list(filepath.glob('*'))
```

```
my_table = []
```

新建空列表

```
for file in filelist:
    with open(file) as f:
        data = f.read()
        my_table.append(data)
```

在for循环挨个读取文件
时候，添加数据

```
print(my_table)
```

```
['11111111111111111111', '2222222222222222222222', '333333333333333333333333']
```

同样的，也可以在这个for
循环里进行一些批量数据
处理的操作（如果需要）

➤ 批量读取表格&栅格文件

• 文件信息

根据上期视频《气象科研绘图8：使用ECWMF数据绘制2m气温空间分布图》-绘图要求 修改

变量：2m温度场

区域范围：20N-40N，110E-130E，空间分辨率 $0.25^{\circ} \times 0.25^{\circ}$

时间范围：~~2021年9月1日~~，0点-23点，时间分辨率1h

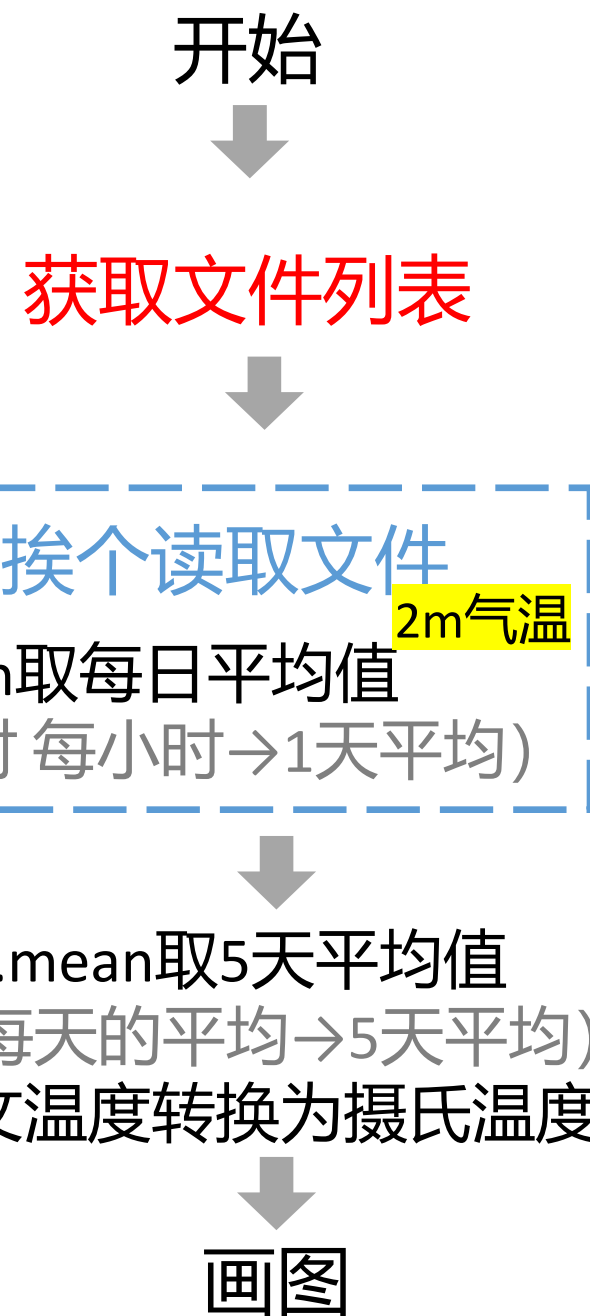
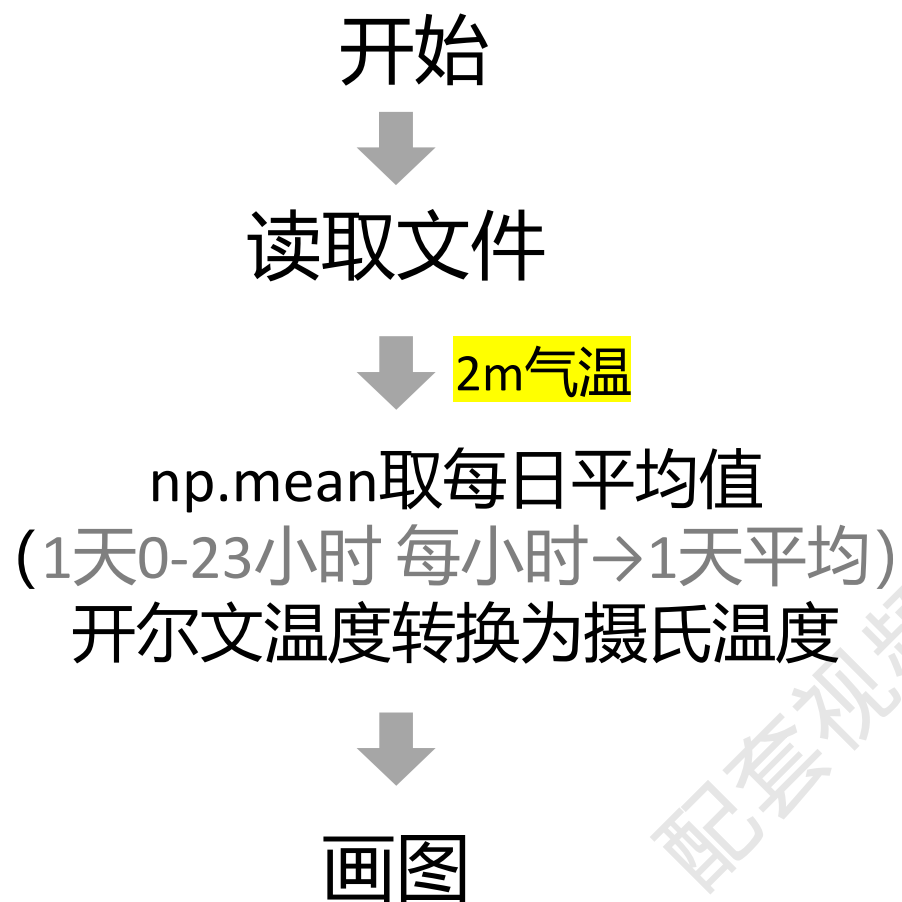
2021年9月1日-2021年9月5日

• 绘图要求

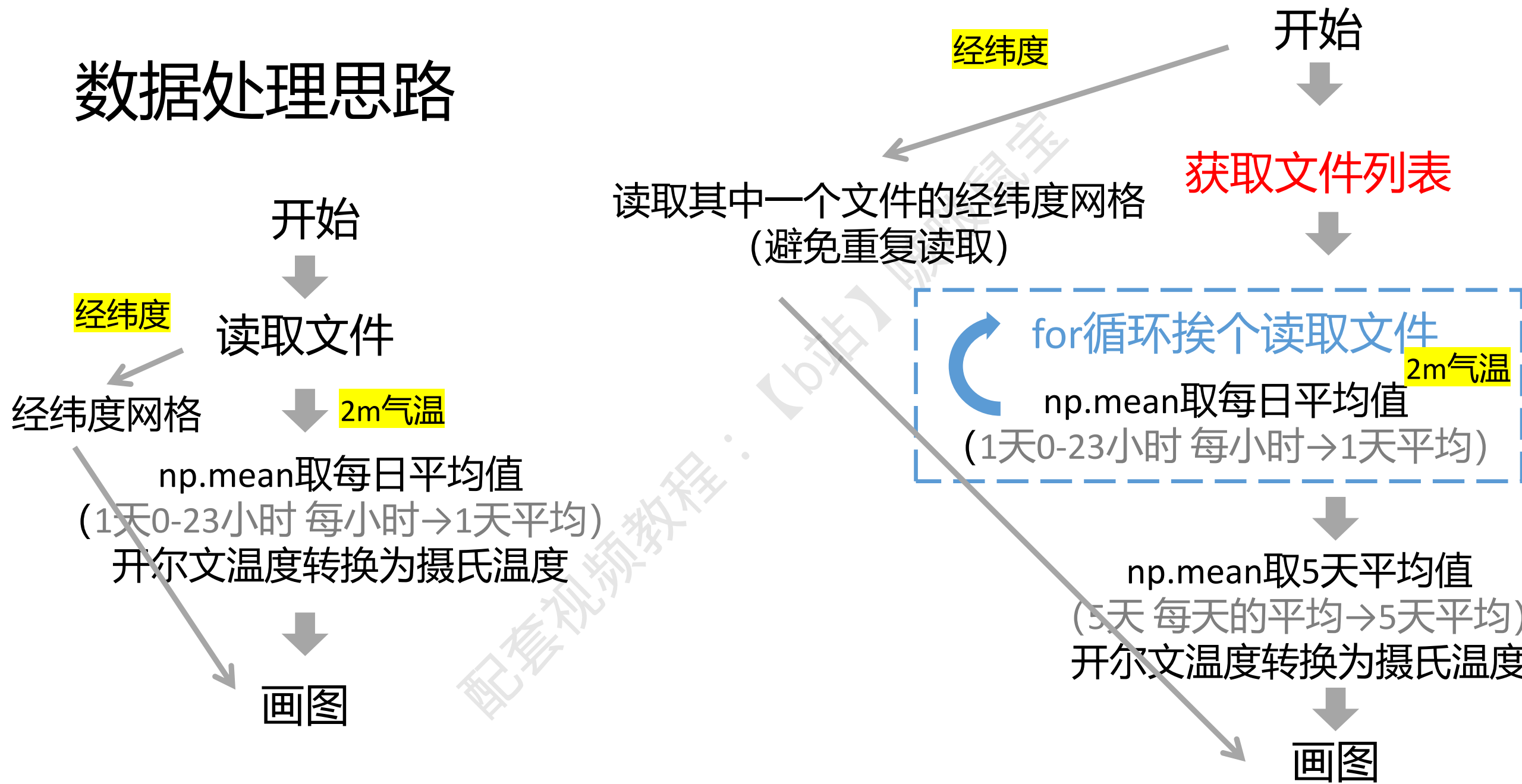
绘制该区域~~2021年9月1日~~平均2m气温空间分布，要求单位为摄氏温度

2021年9月1日-2021年9月5日

数据处理思路



数据处理思路



批量读取表格&栅格文件

获取文件列表

函数不需要更换，进行相应的路径&文件名
字修改即可

for循环挨个读取文件

将open(file)换成打开表格&栅格文件的函数，f.read()换成相应的提取数据的函数

```
filepath = Path('D:/bilibili/Python绘图技巧3')  
filelist = list(filepath.glob('*nc'))
```

```
print(filelist)
```

只读取后缀.nc文件

```
[WindowsPath('D:/bilibili/Python绘图技巧3/air2m20210901.nc'),  
 WindowsPath('D:/bilibili/Python绘图技巧3/air2m20210903.nc'),  
 WindowsPath('D:/bilibili/Python绘图技巧3/air2m20210905.nc')]
```

```
#for循环
```

```
t2m_all = []
```

```
for file in filelist:
```

```
    with xr.open_dataset(file) as f:
```

```
        t2m = f['t2m'].data
```

```
        t2m_day = np.mean(t2m, axis=0) #对时间维度作平均
```

```
        t2m_all.append(t2m_day)
```

up自己平时做数据分析&可视化常见的一些文件分类：

- 文本文件 .txt/.dat等 使用f.read()/readlines()等函数读取
最简单的文本文件
- 表格文件 .txt/.csv/.xls./xlsx等 使用pandas相关函数读取
通常为变量的一维时间序列（多个变量→二维表格）
- 栅格文件 .nc/.hdf/.grd等 使用netCDF4/xarray等第三方库
通常为变量的空间分布/时空分布（二维及以上）

批量读取表格&栅格文件

获取文件列表

函数不需要更换，进行相应的路径&文件名
字修改即可

for循环挨个读取文件

将open(file)换成打开表格&栅格文件的函数，f.read()换成相应的提取数据的函数

```
filepath = Path('D:/bilibili/Python绘图技巧3')  
filelist = list(filepath.glob('*.nc'))  
  
print(filelist)
```

```
[WindowsPath('D:/bilibili/Python绘图技巧3/air2m20210901.nc'),  
 WindowsPath('D:/bilibili/Python绘图技巧3/air2m20210903.nc'),  
 WindowsPath('D:/bilibili/Python绘图技巧3/air2m20210905.nc')]
```




```
#for循环  
t2m_all = []  
for file in filelist:  
    with xr.open_dataset(file) as f:  
        t2m = f['t2m'].data  
        t2m_day = np.mean(t2m, axis=0) #对时间维度作平均  
        t2m_all.append(t2m_day)
```

添加数据处理：np.mean取每日平均值
(1天0-23小时 每小时→1天平均)


数据处理

np.mean取5天平均值
(5天 每天的平均→5天平均)
开尔文温度转换为摄氏温度



```
#对时间维度作平均 (5天每天的→5天平均)
t2m_mean = np.mean(t2m_all, axis=0)

#开尔文温度转换成摄氏温度
t2m_mean = t2m_mean - 273.15
```

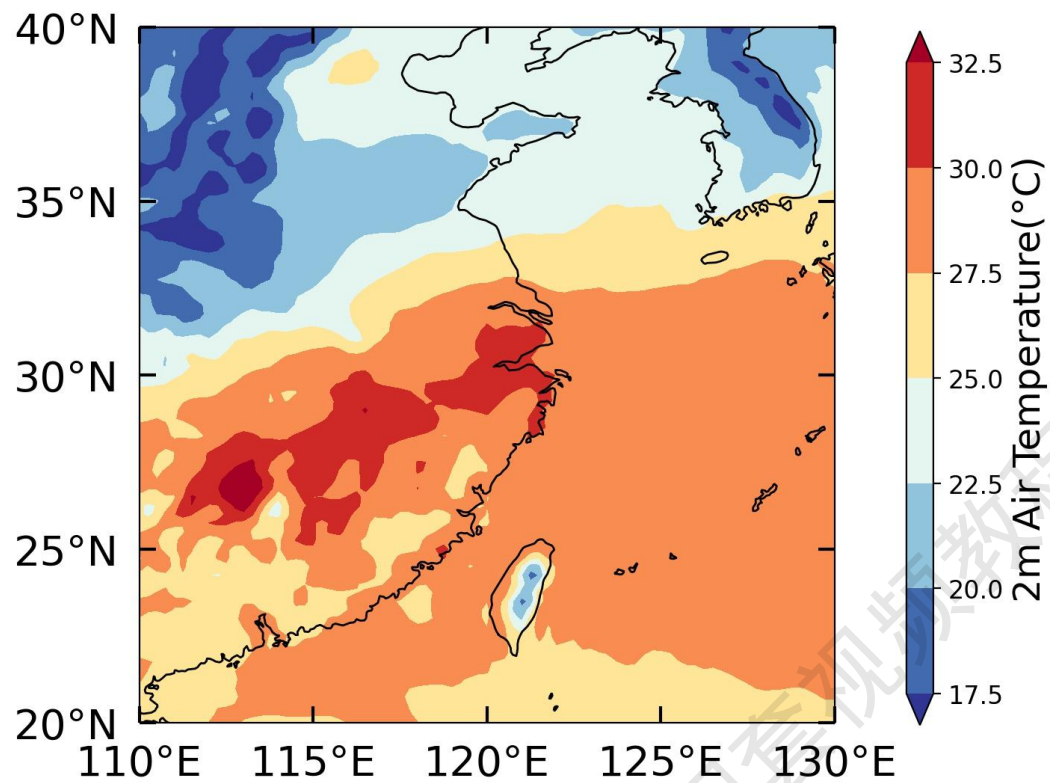


画图

配套视频教程：

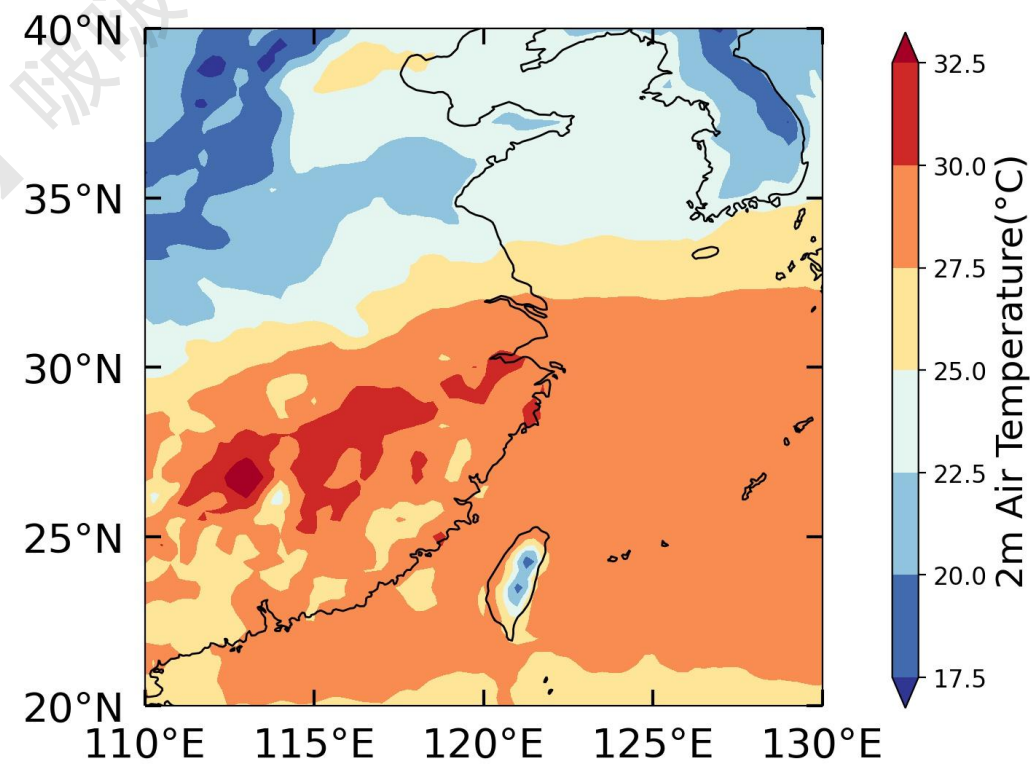
气象科研绘图8 (上期视频)

20210901 2m Air Temperature



本期视频

20210901-20210905 2m Air Temperature



➤ 补充：批量读取多个子文件夹下的文件

此电脑 > Data (D:) > bllibili > Python绘图技巧3

名称	修改日期	类
day20210901	2022/8/10 19:21	文件夹
day20210902	2022/8/10 19:21	文件夹
table	2022/8/10 19:26	文件夹
air2m20210901.nc	2022/8/9 0:13	NC 文件
air2m20210902.nc	2022/8/9 0:14	NC 文件
air2m20210903.nc	2022/8/9 0:14	NC 文件
air2m20210904.nc	2022/8/9 0:14	NC 文件
air2m20210905.nc	2022/8/9 0:15	NC 文件
Python绘图技巧3: 批量读取文件.pptx	2022/8/10 21:31	PPTX 演示文稿

此电脑 > Data (D:) > bllibili > Python绘图技巧3 > day20210901

名称	修改日期
air2m20210901.nc	2022/8/9 0:13

此电脑 > Data (D:) > bllibili > Python绘图技巧3 > day20210902

名称	修改日期
air2m20210902.nc	2022/8/9 0:14

pathlib

提示：可以通过更改filepath.glob()中的参数实现批量读取多个子文件夹下的文件

获取文件列表

```
filepath = Path('D:/bilibili/Python绘图技巧3')  
filelist = list(filepath.glob('day*/*.nc'))  
print(filelist)
```

```
[WindowsPath('D:/bilibili/Python绘图技巧3/day20210901/ai  
902.nc')]
```

这里输出路径截图没法截全

➤ 数据&代码获取

1.和鲸社区 关注账号【啵啵鼠宝】

本项目链接：

<https://www.heywhale.com/mw/project/62f3b8f1d7bf29827ee4c9bc>



啵啵鼠宝 LV3 几秒前 · 发布了 项目

Python绘图技巧3：批量读取nc文件
掌握python批量读取文件的各种方法

2.github仓库【Boboshubao】

仓库链接：<https://github.com/orange-Nan/Boboshubao>