

代码内容审核建议 7 (2.2.2)

1. 此处的字母表示，略微费解，可以在例如前面补充一下

Landsat 文件名格式: LXSS\_LLLL\_PPPRRR\_YYYYMMDD\_yyyymmdd\_CC\_TX,

例如: LC08\_L1TP\_023031\_20180310\_20180320\_01\_T1

“S=卫星。案例为 08”可改为 “SS=卫星编号。案例为 08”

- Landsat文件名所包含的信息

例如: LC08\_L1TP\_023031\_20180310\_20180320\_01\_T1

L=Landsat。案例为L

X=传感器, “C”代表OLI/TIRS组合, “O”代表OLI, “T”代表TIRS, “E”代表ETM+, “T”代表TM, “M”代表MSS。案例为C, 即OLI/TIRS组合传感器

S=卫星。案例为08

LLLL=处理矫正水平 (Processing correction level), 由地面控制点、数字高程和传感器收集的数据确定。“L1TP” (Precision Terrain) 为1级精度地形矫正产品, 如果参考数据不充足, 则处理为系统的、地形矫正的L1GT(Systematic Terrain)和L1GS(Systematic)产品。案例为L1TP

PPP=WRS path 世界参考系统 (the Worldwide Reference), 是Landsat卫星数据的全球定位系统。可以指定path, row查询世界上任何部分的卫星图像。案例为023

RRR=WRS row。案例为031

YYYYMMDD=获取年份。案例为20180310

yyymmdd=处理年份。案例为20180320

CC=采集编号 (01, 02) 。案例为01

TX=采集类别 (“RT”为实时, “T1”为层1 (Tier1) 包含最高数据质量; “T2”为层2, 因为缺失数据精度较T1低) 。案例为T1

对于遥感影像的处理, 即为对raser (栅格数据) 的处理, 可以使用rasterio, earthpy, GDAL等库读写, 定义或重定义投影坐标系统, 裁切与融合, 数据处理及可视化等。

更进一步，此处的等号(“=”)是不合适的，建议修改为表格，

这里以前三个为例，其他的，请根据如下修改，**“【】”内的文字为疑问和建议**

| 占位符 | 代表意义  | 备注  | 举例中的值               |
|-----|---|---|---------------------|
| L   | Landsat 的缩写, 代表此类型数据集                                   | <b>【应该是固定值吧, 统一为 L, 可以在此处点名】</b>  | L                   |
| X   | 传感器类型, 可选值为“C”, “O”, <b>“T”</b> , “E”, <b>“T”</b> , “M” | “C”代表 OLI/TIRS 组合, “O”代表 OLI, <b>“T”代表 TIRS</b> , “E”代表 ETM+, <b>“T”代表 TM</b> , “M”代表 MSS。<br><b>【此处出现两次 T, 请确认修改,</b><br><br><b>或者是, T, M 指的是 E 中的 T 和 M? 则可以修改为下面的格式】</b> | C, 即 OLI/TIRS 组合传感器 |
| X   | 传感器类型, 可选值为“C”, “O”, <b>“T”</b> , “E”                   | “C”代表 OLI/TIRS 组合, “O”代表 OLI, <b>“T”代表 TIRS</b> , “E”代表 ETM+ ( <b>“T”代表 TM</b> , “M”代表 MSS)<br><b>【或者是, T, M 指的是 E 中的 T 和 M? 】</b>  | C, 即 OLI/TIRS 组合传感器 |
| SS  | 卫星编号  | <b>【可添加卫星编号取值范围】</b>  | 08                  |

2. 此处应该将通道数和代表的颜色光谱意义分离

“4、 3 、 2 Red、 Green 、 Blue” 可修改为“(4,3,2) (Red, Green, Blue)”  
建议将 RGB 列名改为“通道编号”，存储类似格式的数据：“(4,3,2)”  
并引入新的一列，列名为“代表波段名”，存储类似格式的数据“(Red, Green, Blue)”  
举例

| 通道编号    | 代表波段名              | 主要用途  |
|---------|--------------------|-------|
| (4,3,2) | (Red, Green, Blue) | 自然真彩色 |

将不同的波段合成显示，可以有不同的显示结果，方便不同目的性的观察，通常如下

| R、 G、 B                     | 主要用途        |
|-----------------------------|-------------|
| 4、 3 、 2 Red、 Green 、 Blue  | 自然真彩色       |
| 7、 6 、 4 SWIR2、 SWIR1 、 Red | 城市          |
| 5、 4、 3 NIR、 Red 、 Green    | 标准假彩色图像， 植被 |
| 6、 5、 2 SWIR1、 NIR 、 Blue   | 农业          |
| 7、 6 、 5 SWIR2、 SWIR1 、 NIR | 穿透大气层       |
| 5、 6 、 2 NIR、 SWIR1 、 Blue  | 健康植被        |
| 5、 6 、 4 NIR、 SWIR1 、 Red   | 陆地/水        |
| 7、 5 、 3 SWIR2、 NIR 、 Green | 移除大气影响的自然表面 |
| 7、 5 、 4 SWIR2、 NIR 、 Red   | 短波红外        |
| 6、 5 、 4 SWIR1、 NIR 、 Red   | 植被分析        |

```
from rasterio.plot import plotting_extent
extent=plotting_extent(w_180310_array[0], w_180310_raster_prof["transform"])

plt.rcParams.update({'font.size': 12})
fig, axs = plt.subplots(1,2,figsize=(25, 12))
ep.plot_rgb(
    w_180310_array,
    rgb=[3,2,1],
    stretch=True,
    extent=extent,
    str_clip=0.5,
    title="w_180310_B4,B3,B2",
    ax=axs[0]
)

ep.plot_rgb(
    w_180310_array,
    rgb=[5,4,1], #4/3/2 5/4/1 波段对应文件，为波段号减1
    stretch=True,
    extent=extent,
    str_clip=0.5,
    title="w_180310_B6,B5,B2",
    ax=axs[1]
)
```

后面的备注有误，  
“[3,2,1]”对应的是“B4， B3， B2”  
“[5,4,1]”对应的是“B6， B5， B2”

3. 协方差矩阵表示方法修改：建议将字母含义表示和数值分开，这样看，太混乱了

示例修改： $\begin{bmatrix} (x(i)|x(i)) & \\ & \end{bmatrix} = \begin{bmatrix} 0.3637 & \\ & \end{bmatrix}$

量，即投影点方差越高，误差越低。

使得投影点方差为最大的新坐标轴是由特征点的协方差求得， $\begin{bmatrix} 0.36372533(x(i)|x(i)) & 0.55326727(y(j)|x(i)) \\ 0.55326727(x(i)|y(j)) & 1.10198785(y(j)|y(j)) \end{bmatrix}$ ，其中 $x(i)|x(i)$ 及 $y(j)|y(j)$ 是各个特征自身的方差；而 $y(j)|x(i)$ 和 $x(i)|y(j)$ 则是特征之间的协方差，其值相同。所要寻找新的坐标系是使得特征点到新坐标系其中一个轴的投影的方差为最大，相对到其它轴投影方差则次第减小，则说明轴之间的方差即 $y(j)|x(i)$ 和 $x(i)|y(j)$ 趋近于0，其它轴的投影点方差也相对较小。 $\begin{bmatrix} 3.11020298e-04(x(i)|x(i)) & -4.96686164e-18(y(j)|x(i)) \\ -4.96686164e-18(x(i)|y(j)) & 2.73202237e+00(y(j)|y(j)) \end{bmatrix}$ ，从计算的结果也能观察到这一特点，其中只有 $y(j)|y(j)$ 的方差最大，其它的均趋于0。轴之间的协方差基本为0，基本不存在相关性；即当协方差为0时，表示两个特征基本完全独立。而为了让协方差为0，在确定第2个轴（基）时只能在与第一个轴（基）正交方向上选择，因此新坐标系的两个轴是正交垂直的。

此外，“次第减小”这个说法是否正确，请确认

4. 为了打印美观，可以打印 `print()`pandas 格式的数据时，将提示信息和打印数据分开即，

```
print("iris covariance:")
print(COV_iris)
```

```
import numpy as np
import pandas as pd
import plotly.express as px
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()

iris_df=px.data.iris()
feature_selection=["sepal_length","sepal_width","petal_width","species"]
COV_iris=iris_df[feature_selection].cov() #先从3维度讲起，方便打印观察，["sepal_length","sepal_width","petal_width"]
print("iris covariance",COV_iris)
f, ax = plt.subplots(figsize=(9,8))
sns.heatmap(COV_iris, annot=True, linewidths=.5, ax=ax)

fig=px.scatter_3d(iris_df, x='sepal_length', y='sepal_width', z='petal_width',color='species')
fig.show()
```

| iris covariance | sepal_length | sepal_width | petal_width |
|-----------------|--------------|-------------|-------------|
| sepal_length    | 0.685694     | -0.039268   | 0.516904    |
| sepal_width     | -0.039268    | 0.188004    | -0.117981   |
| petal_width     | 0.516904     | -0.117981   | 0.582414    |

此外，需要注意的是，计算协方差矩阵时，字符型的数据将不会被考虑到计算中，因此虽然选了四个特征，但是协方差矩阵是 3\*3 的，这个需要点明

建议在计算协方差矩阵 COV\_iris 之前，打印一下 `iris_df[feature_selection].head()`

```
In[19]: iris_df[feature_selection].head()
Out[19]:
   sepal_length  sepal_width  petal_width species
0           5.1           3.5           0.2  setosa
1           4.9           3.0           0.2  setosa
2           4.7           3.2           0.2  setosa
3           4.6           3.1           0.2  setosa
4           5.0           3.6           0.2  setosa

In[20]: COV_iris
Out[20]:
   sepal_length  sepal_width  petal_width
sepal_length    0.685694   -0.039268    0.516904
sepal_width    -0.039268    0.188004   -0.117981
petal_width     0.516904   -0.117981    0.582414
```

```
iris_mapping_N=np.matmul(iris_df[feature_selection[:3]].to_numpy(),N_f_matrix.T).astype(float) #线性映射
species_colors_={'setosa':'coral', 'versicolor':'lightblue', 'virginica':'wheat'}
iris_3Features['color_']=iris_3Features.species.apply(lambda row:species_colors_[row])
axs[0].scatter(iris_mapping_N[:,0],iris_mapping_N[:,1],iris_mapping_N[:,2],c=iris_3Features.color_,)
axs[1].scatter(iris_mapping_N[:,0],iris_mapping_N[:,1],iris_mapping_N[:,2],c=iris_3Features.color_,)
```

后面进行线性映射运算时，也对特征进行了进一步的选择，可考虑添加注释

5. 大小写问题【Landsat 算专有名词么，是否应该是 The landsat series of satellites】-

### 2.2.2.1 Landsat地理信息读取，裁切，融合，打印显示

从1970s年代，地球资源卫星 (the Landsat series of satellites) 几乎不间断的提供全球范围地球表面中分辨率多光谱遥感影像数据。不间断的数据记录，增强了应用数据分析的潜力，可以帮助各领域的研究者在人口不断增长，城市不断发展下，关注粮食、水、森林等自然资源，应用其无与伦比的影像质量、覆盖率、细节价值分析城市的发展与自然平衡的问题(美国地质调查局USGS-earthexplorer,United States Geological Survey)。

注，下述mermaid图表代码，在JupyterLab下无法显示，但是在通常的markdown下显示。