Chapter 5 : Data with Pandas
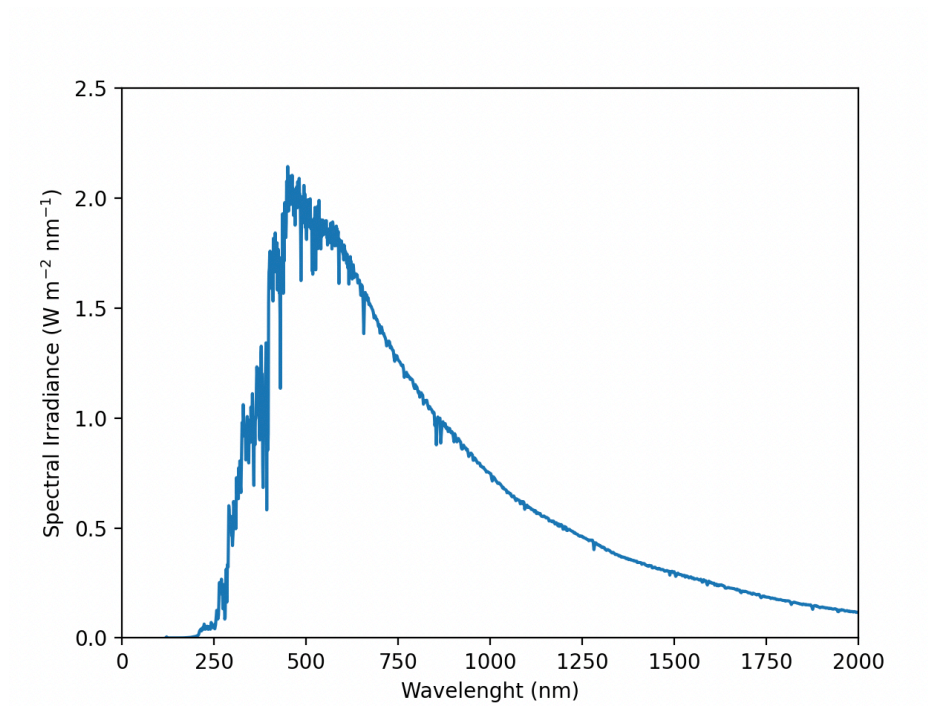
- pandas is a fast, powerful, flexible tools in Python for data analysis and data manipulation tools.

- To call pandas in python, we need to do:

  import pandas as pd

- As we learn in data arrays, python read data in. rows not column. This is a bit tricky if we are dealing with large amount of data for example 100 x 100 data.

- pandas is the solution to read data from large files and make the computation easier.

- pandas can read various format of data including text file, excel, csv.

- The format to read filename:

  - o   for excel : pd.read_excel('filename.xlsx')
  - o   for csv or text file : pd.read_csv('filename.txt')

- If we have data with header but we want to skip or ignore the header, we can add option Header=None when reading the data.

- In text file, in order to make the pandas read column efficiently, we can set the separation as sep = '\s+'

- In our first example, we are going to read a solar spectral data stored in excel. This data contained two columns, which are wavelength in nm and spectral irrandiance in W m^{-2} nm^{-1} from observation with ~1697 rows of data. This data is an example of black body radiation data.

  The program:

```
1    #This program shows how to read data from excel file
2    #The data given in this example is a Standard Solar Spectra
3
4    #import libraries
5    import matplotlib.pyplot as mpl
6    import numpy as np
7    import pandas as pd
8
9    #first read the excel file using pandas
10   nama=['wave','irr']
11   df=pd.read_excel('solar_spectra.xlsx',header=0,names=nama)
12   #print(df['wavelength'],df['irradiance'])
13   x= df['wave']
14   y=df['irr']
15
16   print(y)
17   mpl.plot(x,y)
18   mpl.xlim(0,2000)
19   mpl.ylim(0.0,2.5)
20   mpl.xlabel('Wavelenght (nm)')
21   mpl.ylabel('Spectral Irradiance (W m$^{-2}$ nm$^{-1}$)')
22   mpl.show()
23
24
```

The output:

We can include the formulation of black body radiation in the plot to fit the data.