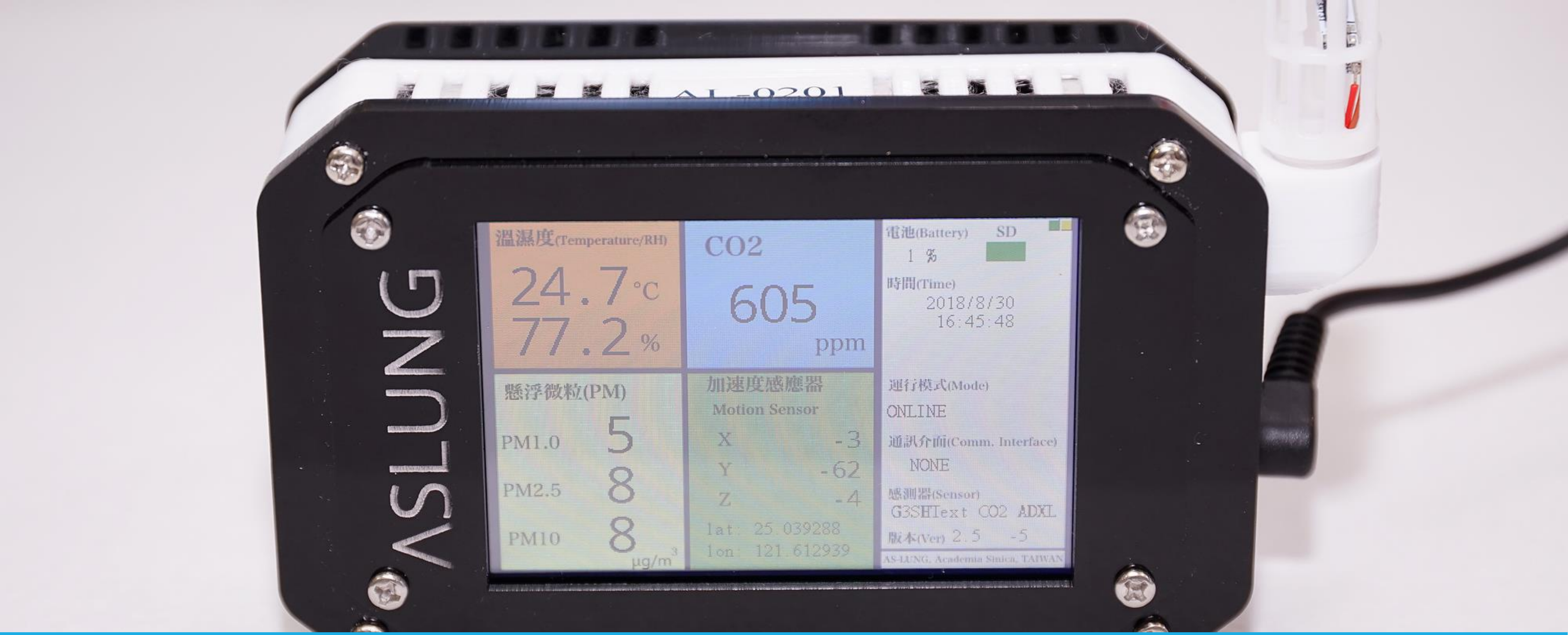


Advanced Institute on Health Investigation and Air Sensing for Asian Pollution (AI on Hi-ASAP)

Shih-Chun Candice LUNG & Chun-Hu Liu 2020/10/05

Outline

- Get calibration factor from reference instrument
  - Rational of calefaction factor
  - How to run python code to set calibration factor
- Data cleaning
  - What data cleaning tasks need to be done
  - How to run python code to do data cleaning

Get calibration factor from reference instrument

Rationale of calibration factor

AS-Lung compare with reference instrument

Reference instrument

AS-Lung

Rename the filename to fit the python code

Rename the filename of AS-Lung

Ex: 2020-09-27.csv ➔AL-0001_2020-09-27.csv

The filename should be contain AS-Lung ID (AL-0001 or AL0001)

Rename the filename of reference standard

Ex: Grimm026.xls ➔standard_Grimm026.xls

The filename should be contain “standard” and with excel data file

Rename the column name to fit the python code

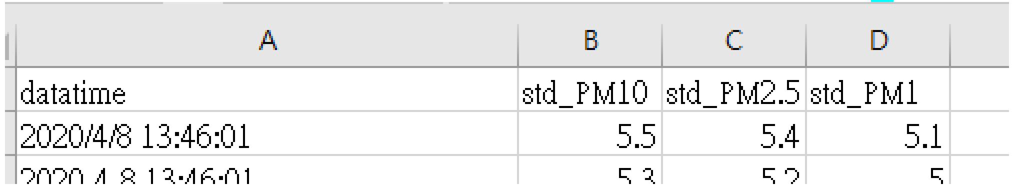
Data format of AS-Lung

Do not modify the data format of AL-Lung

Data format of reference instrument

Rename the column name of time and PM

Modify the data time format of time column



A	B	C	D
datatime	std_PM10	std_PM2.5	std_PM1
2020/4/8 13:46:01	5.5	5.4	5.1
2020.4.8 13:46:01	5.3	5.2	5

Select regression model

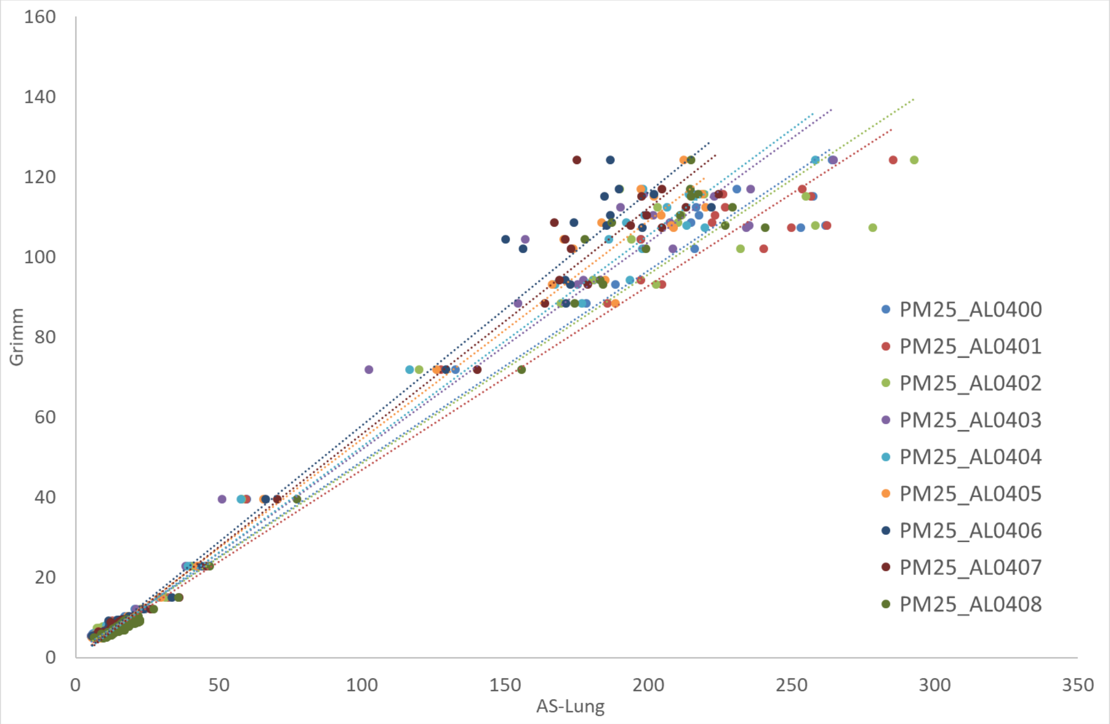
Simple linear regression or Two segments regression

Calibration factor

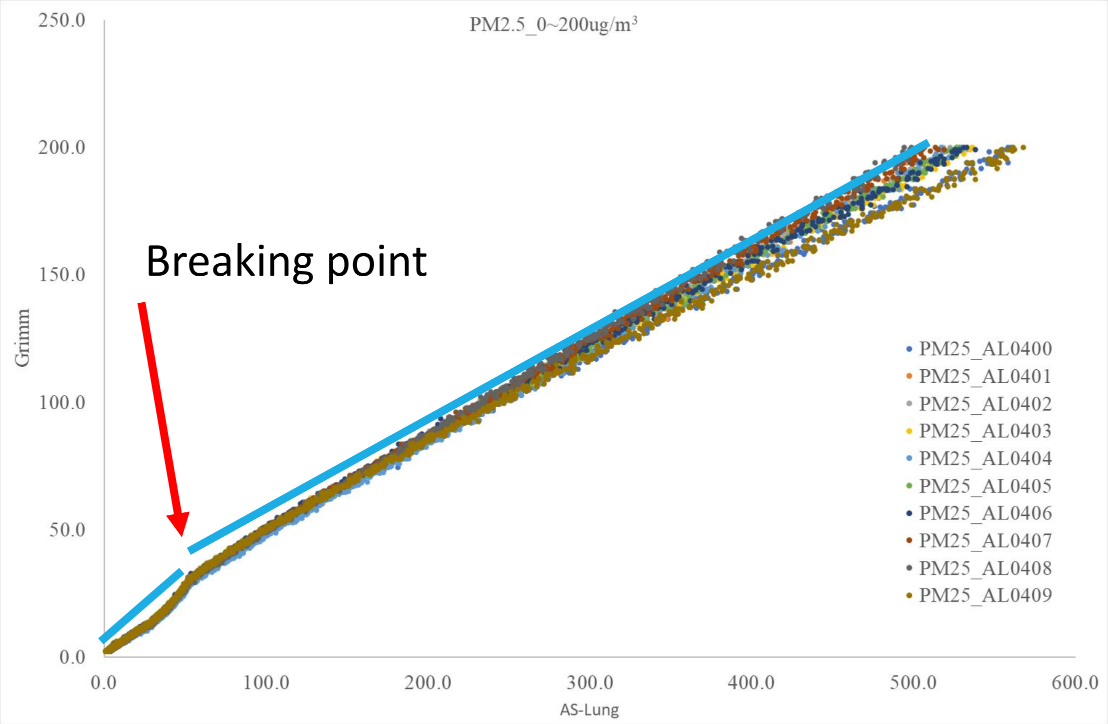
Factors of PM_1 and $PM_{2.5}$ will be reported since PM10 of G3 sensor is not reliable.

Regression model

Simple linear regression



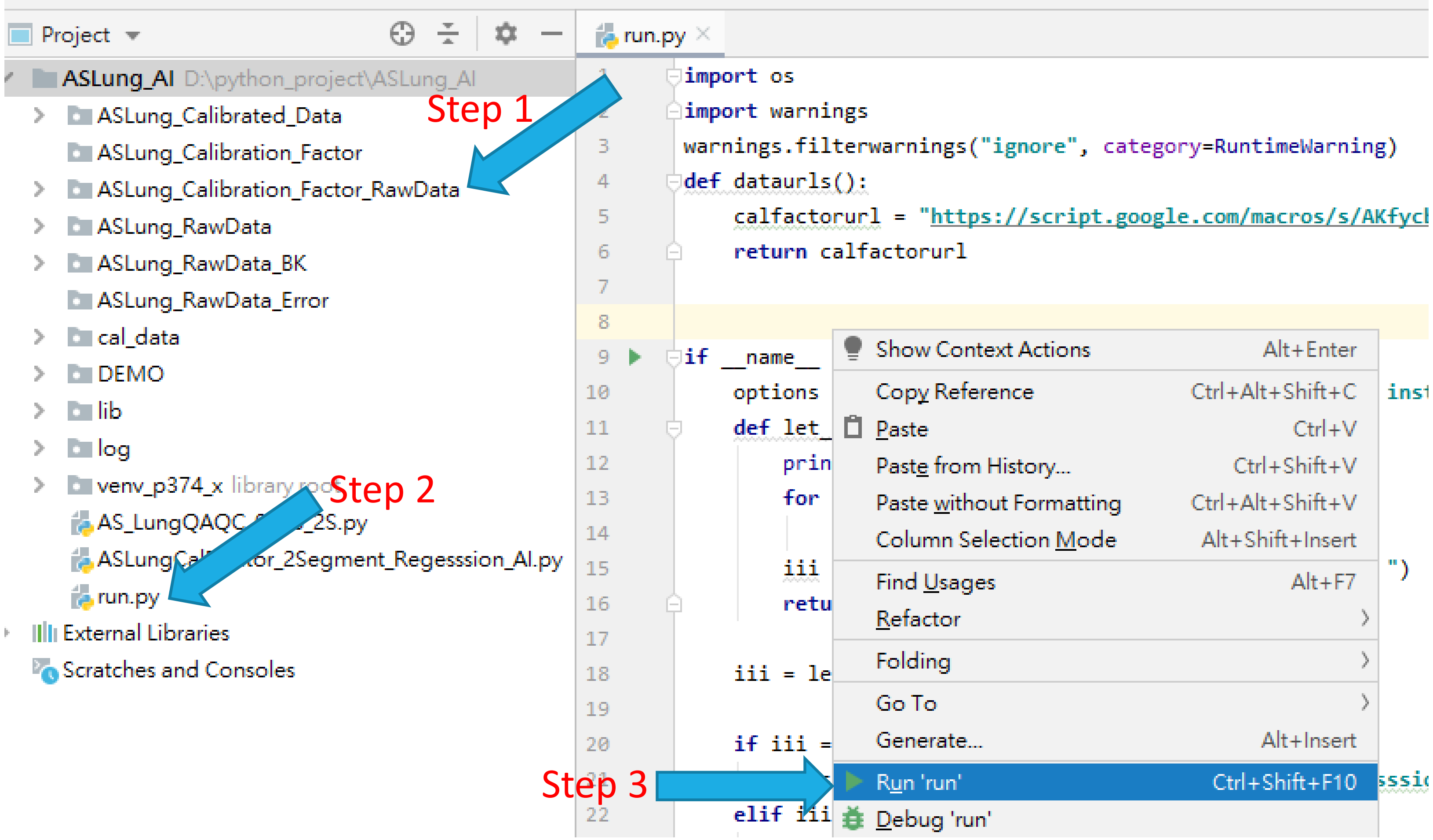
Two segments regression



When run the python code, we can select regression model.

How to run the Python Code (1)

1. Open PyCharm -> project -> run.py



```
import os
import warnings
warnings.filterwarnings("ignore", category=RuntimeWarning)
def dataurls():
    calfactorurl = "https://script.google.com/macros/s/AKfycl
    return calfactorurl


if __name__
    options                                              inst
    def let_
        prin
        for

        iii                                              ")
        retu

    iii = le

    if iii =                                             ssi(

    elif iii
```

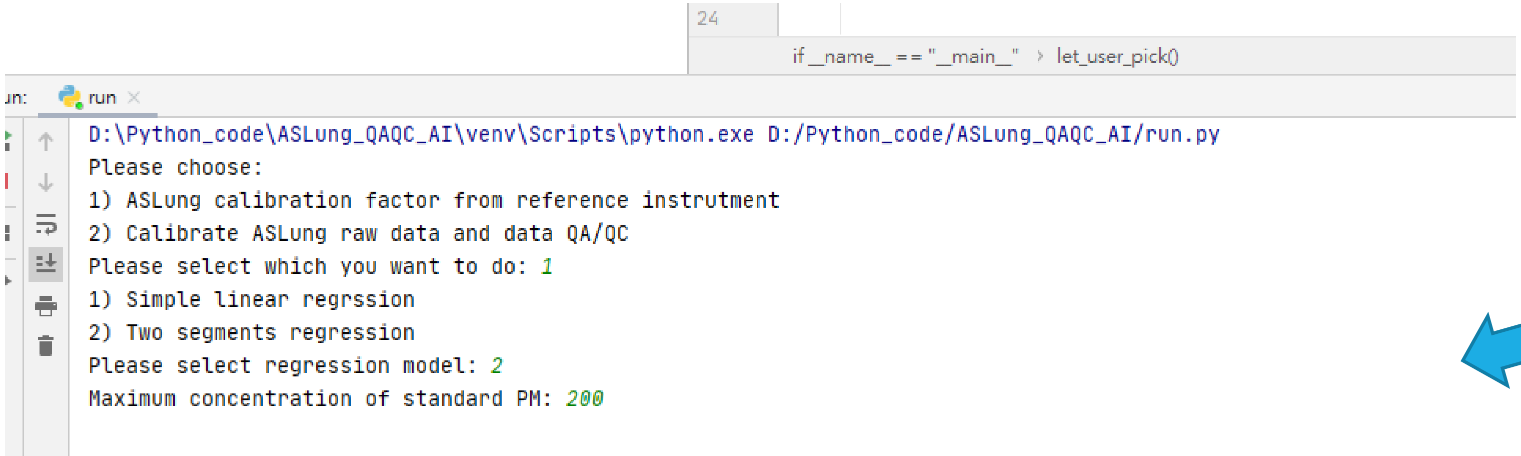
Step 1:
Copy raw data to the folder of “ASLung_Calibration_Factor_RawData”

Step 2:
Open run.py (double click)

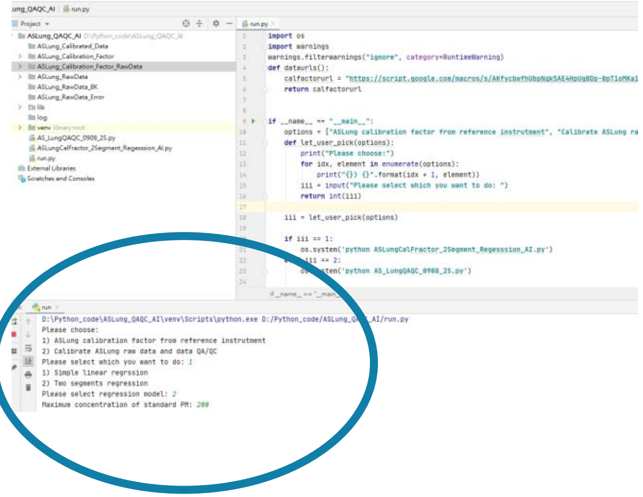
Step 3:
Run the python code

How to run the Python Code (2)

2. Select 1) ASLung calibration factor from reference instrument to run the python code



```
D:\Python_code\ASLung_QAQC_AI\venv\Scripts\python.exe D:/Python_code/ASLung_QAQC_AI/run.py
Please choose:
1) ASLung calibration factor from reference instrutment
2) Calibrate ASLung raw data and data QA/QC
Please select which you want to do: 1
1) Simple linear regrssion
2) Two segments regression
Please select regression model: 2
Maximum concentration of standard PM: 200
```



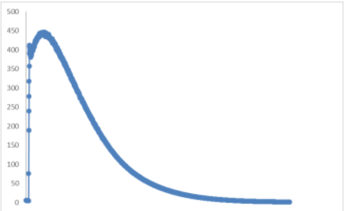
```
import os
import warnings
warnings.filterwarnings("ignore", category=RuntimeWarning)
def dataurls():
    calfactorurl = "https://script.google.com/macros/s/AKfycbwfhUbpNqk5AE4HpUg0Dp-8pTloMKa1
    return calfactorurl


if __name__ == "__main__":
    options = ["ASLung calibration factor from reference instrutment", "Calibrate ASLung ra
    def let_user_pick(options):
        print("Please choose:")
        for idx, element in enumerate(options):
            print("{}) {}".format(idx + 1, element))
        iii = input("Please select which you want to do: ")
        return int(iii)

    iii = let_user_pick(options)

    if iii == 1:
        os.system('python ASLungCalFractor_2Segment_Regesssion_AI.py')
    elif iii == 2:
        os.system('python AS_LungQAQC_0908_25.py')
```

3. Select Regression model and input PM value after “Maximum concentration of standard PM”

3.1 Maximum concentration of standard PM is the highest concentration of the regression.

3.2 We use 200 (chamber) or 150 (hood) as maximum value.

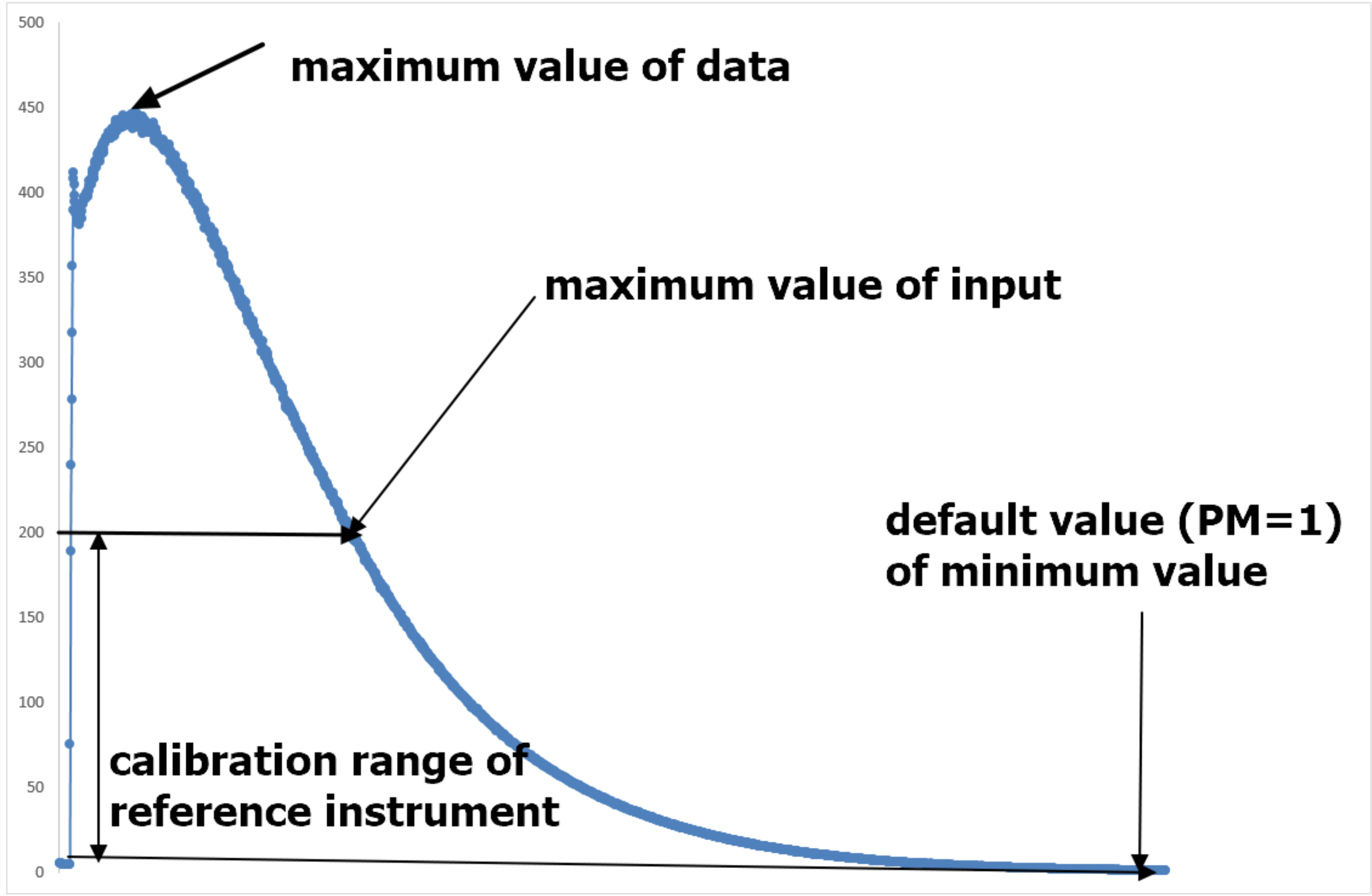
3.2 If max value is 200, the calibration range is from 1 to 200.



4. The python code will automatically average data to 1 minutes.

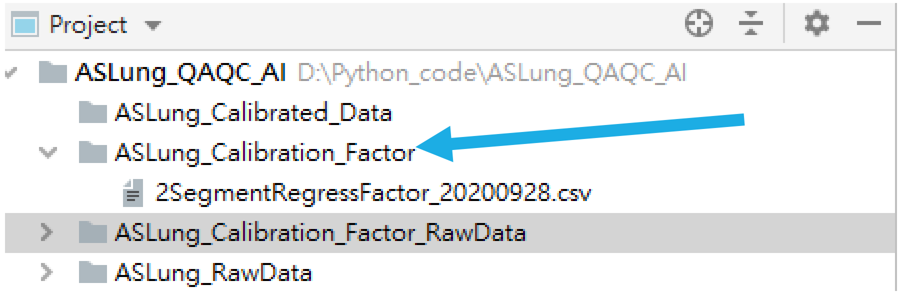
5. The python code also select the calibration range from 1 to max value of input after the maximum value of the data set

How to run the Python Code (2)



How to run the Python Code (3)

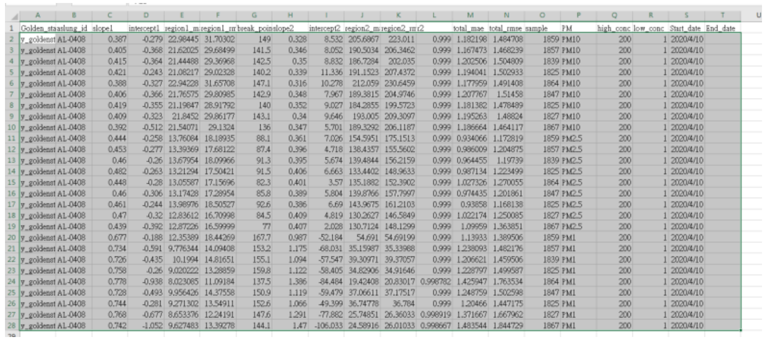
6. When finish, the file will be save in the folder of “ASLung_Calibration_Factor “



Project

- ASLung_QAQC_AI D:\Python_code\ASLung_QAQC_AI
  - ASLung_Calibrated_Data
  - ASLung_Calibration_Factor
    - 2SegmentRegressFactor_20200928.csv
  - ASLung_Calibration_Factor_RawData
  - ASLung_RawData

7. Open the file, copy the calibration factor and past them to the google sheet.



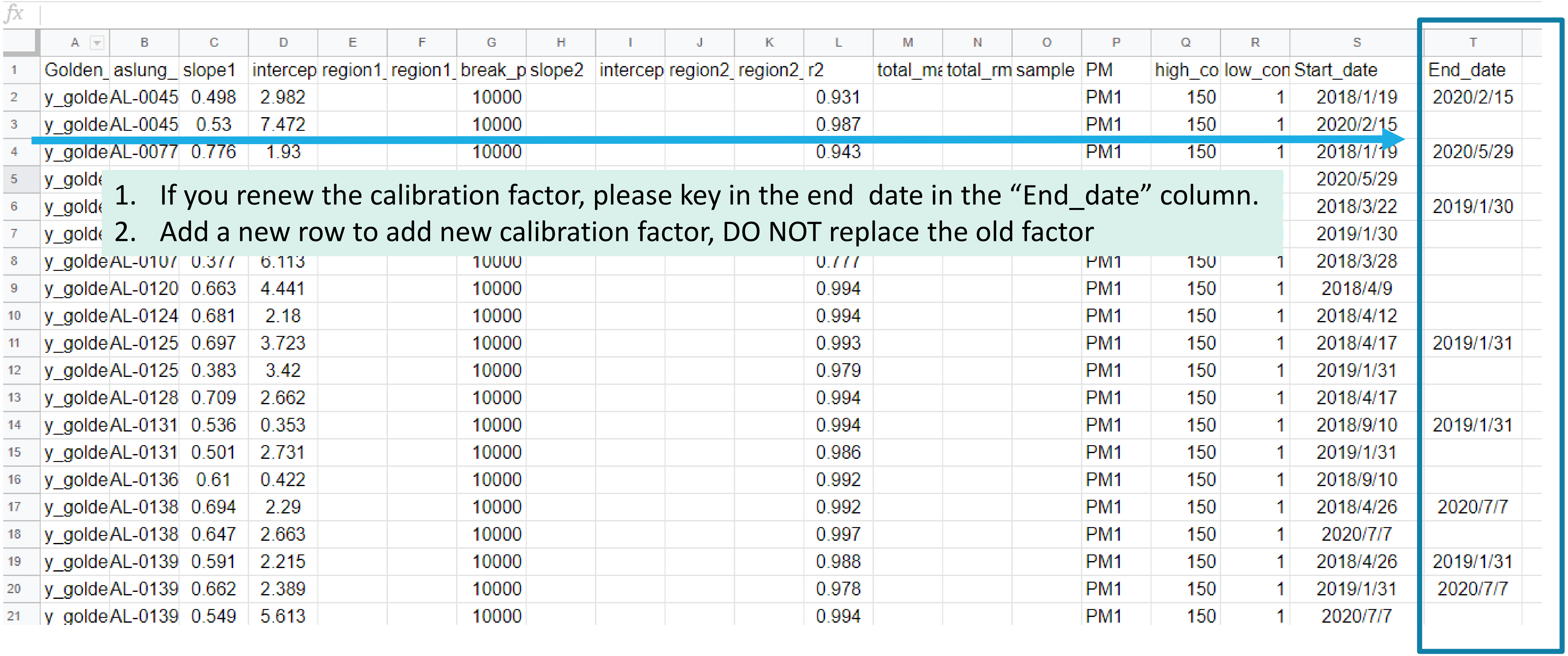
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Golden_sta	aslung_id	slope1	intercept1	region1_m	region1_rr	break_poin	slope2	intercept2	region2_m	region2_rr	r2	total_mae	total_rmse	sample	PM	high_conc	low_conc	Start_date	End_date
2	y_goldenst	AL-0408	0.387	-0.279	22.98445	31.70302	149	0.328	8.532	205.6867	223.011	0.999	1.182198	1.484708	1859	PM10	200	1	2020/4/10	
3	y_goldenst	AL-0408	0.405	-0.368	21.62025	29.68499	141.5	0.346	8.052	190.5034	206.3462	0.999	1.167473	1.468239	1857	PM10	200	1	2020/4/10	
4	y_goldenst	AL-0408	0.415	-0.364	21.44488	29.36968	142.5	0.35	8.832	186.7284	202.035	0.999	1.202506	1.504809	1839	PM10	200	1	2020/4/10	
5	y_goldenst	AL-0408	0.421	-0.243	21.08217	29.02328	140.2	0.339	11.336	191.1523	207.4372	0.999	1.194041	1.502933	1825	PM10	200	1	2020/4/10	
6	y_goldenst	AL-0408	0.388	-0.327	22.94228	31.65708	147.1	0.316	10.278	212.059	230.6459	0.999	1.177959	1.491408	1864	PM10	200	1	2020/4/10	
7	y_goldenst	AL-0408	0.406	-0.366	21.76575	29.80985	142.9	0.348	7.967	189.3815	204.9746	0.999	1.207767	1.51458	1847	PM10	200	1	2020/4/10	
8	y_goldenst	AL-0408	0.419	-0.355	21.19847	28.91792	140	0.352	9.027	184.2855	199.5723	0.999	1.181382	1.478489	1825	PM10	200	1	2020/4/10	
9	y_goldenst	AL-0408	0.409	-0.323	21.8452	29.86177	143.1	0.34	9.646	193.005	209.3097	0.999	1.195263	1.48824	1827	PM10	200	1	2020/4/10	
10	y_goldenst	AL-0408	0.392	-0.512	21.54071	29.1324	136	0.347	5.701	189.3292	206.1187	0.999	1.186664	1.464117	1867	PM10	200	1	2020/4/10	
11	y_goldenst	AL-0408	0.444	-0.258	13.76084	18.18935	88.1	0.361	7.026	154.5951	175.1513	0.999	0.934066	1.172819	1859	PM2.5	200	1	2020/4/10	
12	y_goldenst	AL-0408	0.453	-0.277	13.39369	17.68122	87.4	0.396	4.718	138.4357	155.5602	0.999	0.986009	1.204875	1857	PM2.5	200	1	2020/4/10	
13	y_goldenst	AL-0408	0.46	-0.26	13.67954	18.09966	91.3	0.395	5.674	139.4844	156.2159	0.999	0.964455	1.19739	1839	PM2.5	200	1	2020/4/10	
14	y_goldenst	AL-0408	0.482	-0.263	13.21294	17.50421	91.5	0.406	6.663	133.4402	148.9633	0.999	0.987134	1.223499	1825	PM2.5	200	1	2020/4/10	
15	y_goldenst	AL-0408	0.448	-0.28	13.05587	17.15696	82.3	0.401	3.57	135.1882	152.3902	0.999	1.027326	1.270055	1864	PM2.5	200	1	2020/4/10	
16	y_goldenst	AL-0408	0.46	-0.306	13.17428	17.28954	85.8	0.389	5.804	139.8766	157.7997	0.999	0.974435	1.201861	1847	PM2.5	200	1	2020/4/10	
17	y_goldenst	AL-0408	0.461	-0.244	13.98976	18.50527	92.6	0.386	6.69	143.9675	161.2103	0.999	0.93858	1.168138	1825	PM2.5	200	1	2020/4/10	
18	y_goldenst	AL-0408	0.47	-0.32	12.83612	16.70998	84.5	0.409	4.819	130.2627	146.5849	0.999	1.022174	1.250085	1827	PM2.5	200	1	2020/4/10	
19	y_goldenst	AL-0408	0.439	-0.392	12.87226	16.59999	77	0.407	2.028	130.7124	148.1299	0.999	1.09959	1.363851	1867	PM2.5	200	1	2020/4/10	
20	y_goldenst	AL-0408	0.677	-0.188	12.35389	18.44269	167.7	0.987	-52.184	54.691	54.69199	0.999	1.13933	1.389506	1859	PM1	200	1	2020/4/10	
21	y_goldenst	AL-0408	0.734	-0.591	9.776344	14.09408	153.2	1.175	-68.031	35.15987	35.33988	0.999	1.238093	1.482176	1857	PM1	200	1	2020/4/10	
22	y_goldenst	AL-0408	0.726	-0.435	10.1994	14.81651	155.1	1.094	-57.547	39.30971	39.37057	0.999	1.206621	1.459506	1839	PM1	200	1	2020/4/10	
23	y_goldenst	AL-0408	0.758	-0.26	9.020222	13.28859	159.8	1.122	-58.405	34.82906	34.91646	0.999	1.228797	1.499587	1825	PM1	200	1	2020/4/10	
24	y_goldenst	AL-0408	0.778	-0.938	8.023085	11.09184	137.5	1.386	-84.484	19.42408	20.83017	0.998782	1.425947	1.763534	1864	PM1	200	1	2020/4/10	
25	y_goldenst	AL-0408	0.728	-0.493	9.956426	14.37558	150.9	1.119	-59.479	37.06611	37.17517	0.999	1.248759	1.502598	1847	PM1	200	1	2020/4/10	
26	y_goldenst	AL-0408	0.744	-0.281	9.271302	13.54911	152.6	1.066	-49.399	36.74778	36.784	0.999	1.20466	1.447175	1825	PM1	200	1	2020/4/10	
27	y_goldenst	AL-0408	0.768	-0.677	8.653376	12.24191	147.6	1.291	-77.882	25.74851	26.36033	0.998919	1.371667	1.667962	1827	PM1	200	1	2020/4/10	
28	y_goldenst	AL-0408	0.742	-1.052	9.627483	13.39278	144.1	1.47	-106.033	24.58916	26.01033	0.998667	1.483544	1.844729	1867	PM1	200	1	2020/4/10	

This google sheet only for demo and course, all the research group has individual google sheet

https://docs.google.com/spreadsheets/d/1yuvjPvsr1sEzm_pXpZWpnMxEC3WbsKn65V-apohp86E/edit#gid=0

The format between csv file and google sheet are the same, DONOT change the data format

Google sheet



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Golden_	aslung_	slope1	intercep	region1_	region1_	break_p	slope2	intercep	region2_	region2_	r2	total_ma	total_rm	sample	PM	high_co	low_con	Start_date	End_date
2	y_golde	AL-0045	0.498	2.982			10000					0.931				PM1	150	1	2018/1/19	2020/2/15
3	y_golde	AL-0045	0.53	7.472			10000					0.987				PM1	150	1	2020/2/15	
4	y_golde	AL-0077	0.776	1.93			10000					0.943				PM1	150	1	2018/1/19	2020/5/29
5	y_golde																		2020/5/29	
6	y_golde																		2018/3/22	2019/1/30
7	y_golde																		2019/1/30	
8	y_golde	AL-0107	0.377	6.113			10000					0.777				PM1	150	1	2018/3/28	
9	y_golde	AL-0120	0.663	4.441			10000					0.994				PM1	150	1	2018/4/9	
10	y_golde	AL-0124	0.681	2.18			10000					0.994				PM1	150	1	2018/4/12	
11	y_golde	AL-0125	0.697	3.723			10000					0.993				PM1	150	1	2018/4/17	2019/1/31
12	y_golde	AL-0125	0.383	3.42			10000					0.979				PM1	150	1	2019/1/31	
13	y_golde	AL-0128	0.709	2.662			10000					0.994				PM1	150	1	2018/4/17	
14	y_golde	AL-0131	0.536	0.353			10000					0.994				PM1	150	1	2018/9/10	2019/1/31
15	y_golde	AL-0131	0.501	2.731			10000					0.986				PM1	150	1	2019/1/31	
16	y_golde	AL-0136	0.61	0.422			10000					0.992				PM1	150	1	2018/9/10	
17	y_golde	AL-0138	0.694	2.29			10000					0.992				PM1	150	1	2018/4/26	2020/7/7
18	y_golde	AL-0138	0.647	2.663			10000					0.997				PM1	150	1	2020/7/7	
19	y_golde	AL-0139	0.591	2.215			10000					0.988				PM1	150	1	2018/4/26	2019/1/31
20	y_golde	AL-0139	0.662	2.389			10000					0.978				PM1	150	1	2019/1/31	2020/7/7
21	y_golde	AL-0139	0.549	5.613			10000					0.994				PM1	150	1	2020/7/7	

1. If you renew the calibration factor, please key in the end date in the “End_date” column.
2. Add a new row to add new calibration factor, DO NOT replace the old factor

Data cleaning

What data cleaning tasks need to be done

AS-Lung monitoring data

There are two ways toe get raw data
1.data in SD card or 2.Data from database

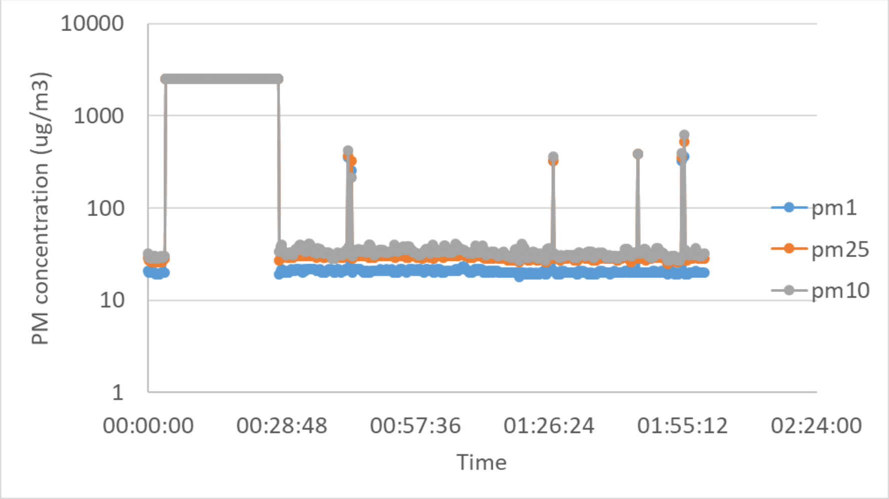
Data cleaning criteria

Step of data cleaning

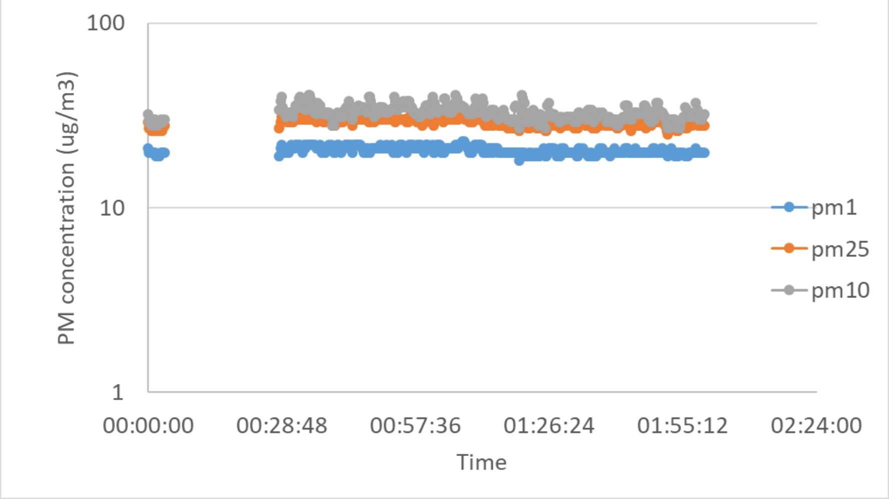
1. PM < 1
2. PM > 50 and PM1=PM2.5=PM10
3. Remove ghost peak ▶

Set PM value as NaN
NaN is equal null

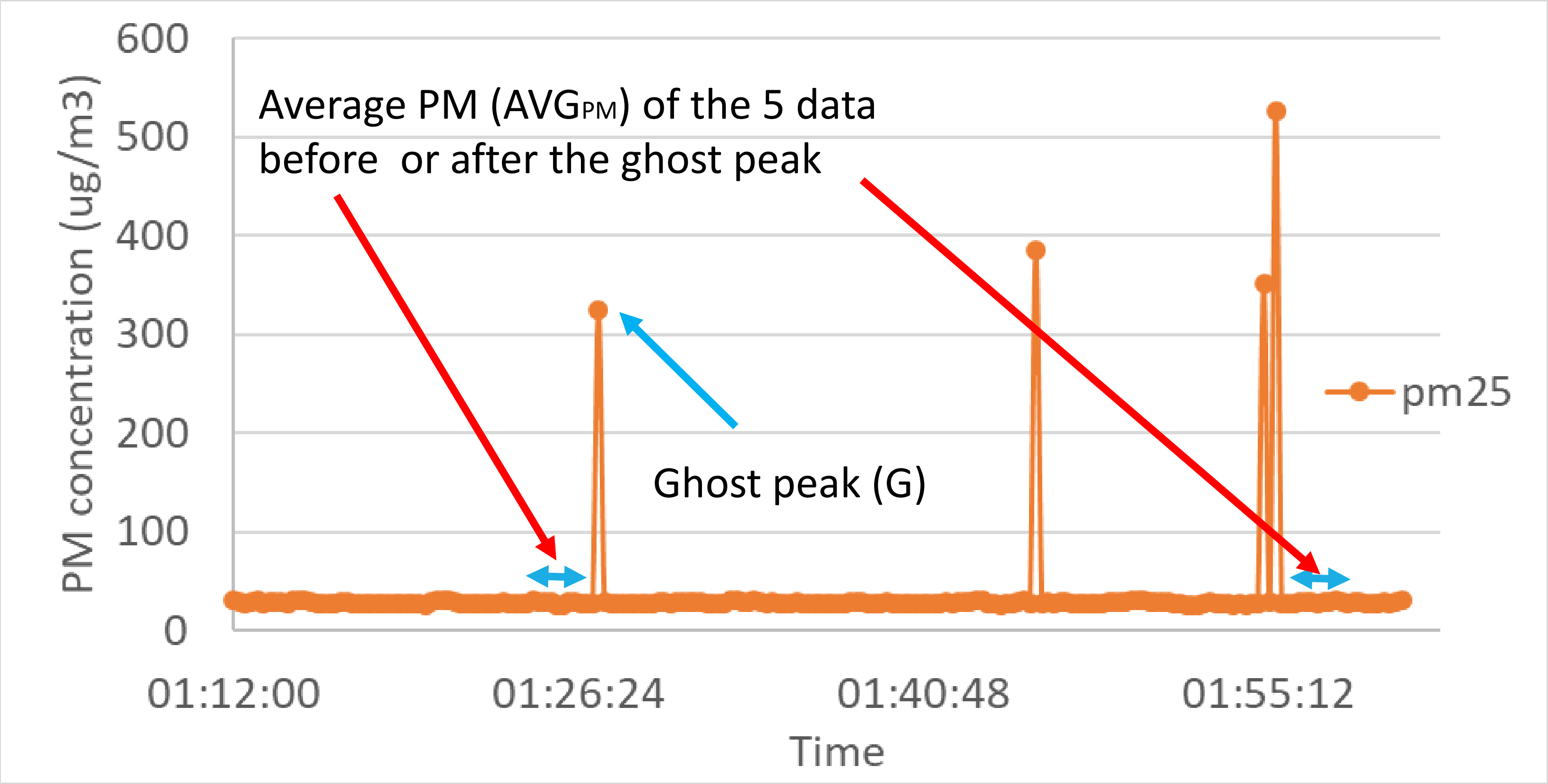
Before



After



Remove ghost peak



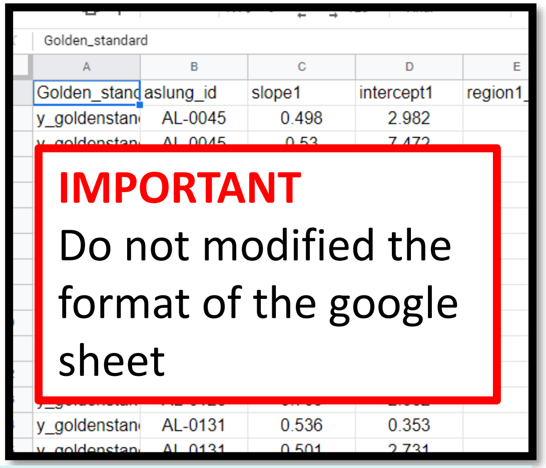
Average PM (AVG_{PM}) of the 5 data before or after the ghost peak

Ghost peak (G)

$$\frac{G}{AVG_{PM}} > 10$$

What data cleaning tasks need to be done

4. Remove data of temperature, humidity and CO2 when values are less than 1
5. Get calibration factor from google drive and calibrate AS-Lung data

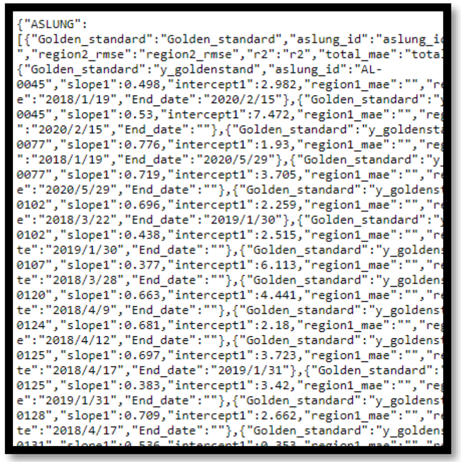


Golden_standard

A	B	C	D	E
Golden_stand	aslung_id	slope1	intercept1	region1_
y_goldenstan	AL-0045	0.498	2.982	
y_goldenstan	AL-0045	0.53	7.472	
y_goldenstan	AL-0131	0.536	0.353	
y_goldenstan	AL-0131	0.501	2.731	

IMPORTANT
Do not modified the format of the google sheet

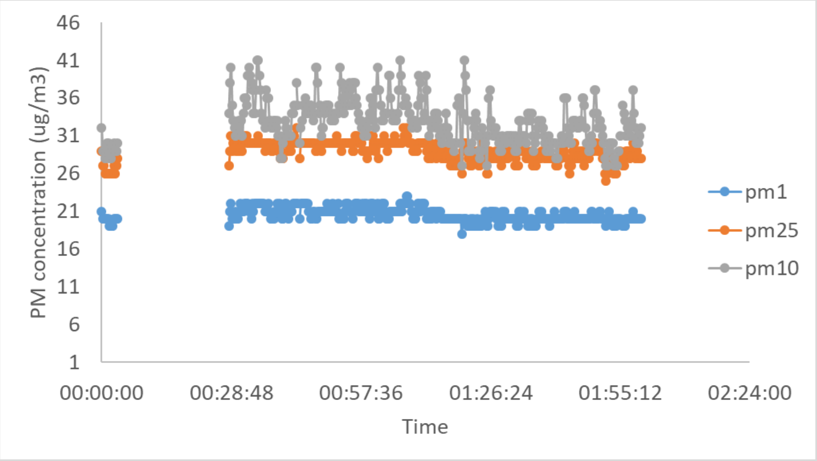
Calibration factor in google sheet



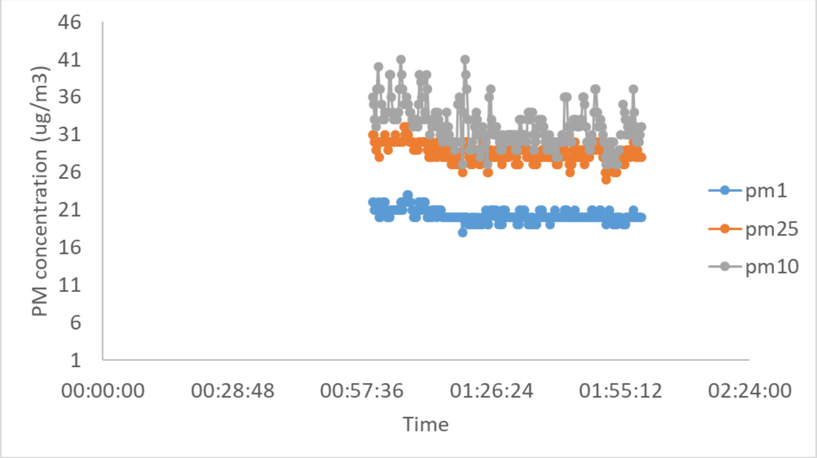
```
{"ASLUNG":
[{"Golden_standard":"Golden_standard","aslung_id":"aslung_i
","region2_rmse":"region2_rmse","r2":"r2","total_mae":"tota
{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0045","slope1":0.498,"intercept1":2.982,"region1_mae":"","r
e":"2018/1/19","End_date":"2020/2/15"},{"Golden_standard":"
0045","slope1":0.53,"intercept1":7.472,"region1_mae":"","re
":"2020/2/15","End_date":""},{"Golden_standard":"y_goldenst
0077","slope1":0.776,"intercept1":1.93,"region1_mae":"","re
":"2018/1/19","End_date":"2020/5/29"},{"Golden_standard":"y
0077","slope1":0.719,"intercept1":3.705,"region1_mae":"","r
e":"2020/5/29","End_date":""},{"Golden_standard":"y_goldens
0102","slope1":0.696,"intercept1":2.259,"region1_mae":"","r
e":"2018/3/22","End_date":"2019/1/30"},{"Golden_standard":"
0102","slope1":0.438,"intercept1":2.515,"region1_mae":"","r
te":"2019/1/30","End_date":""},{"Golden_standard":"y_golden
0107","slope1":0.377,"intercept1":6.113,"region1_mae":"","r
te":"2018/3/28","End_date":""},{"Golden_standard":"y_golden
0120","slope1":0.663,"intercept1":4.441,"region1_mae":"","r
te":"2018/4/9","End_date":""},{"Golden_standard":"y_goldens
0124","slope1":0.681,"intercept1":2.18,"region1_mae":"","re
e":"2018/4/12","End_date":""},{"Golden_standard":"y_goldens
0125","slope1":0.697,"intercept1":3.723,"region1_mae":"","r
te":"2018/4/17","End_date":"2019/1/31"},{"Golden_standard":
0125","slope1":0.383,"intercept1":3.42,"region1_mae":"","re
e":"2019/1/31","End_date":""},{"Golden_standard":"y_goldens
0128","slope1":0.709,"intercept1":2.662,"region1_mae":"","r
te":"2018/4/17","End_date":""},{"Golden_standard":"y_golden
0131","slope1":0.536,"intercept1":0.353,"region1_mae":"","r
```

Calibration factor API from google sheet

Before

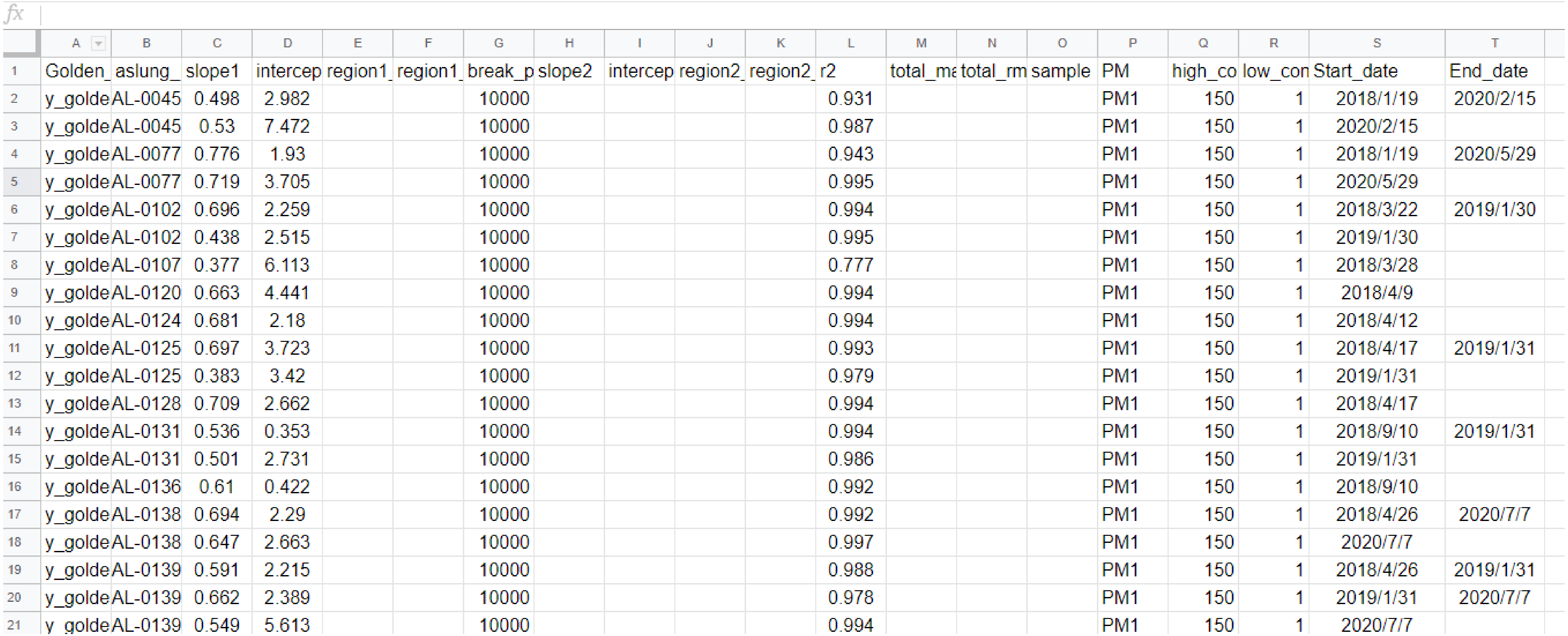


After



6. If calibrated $PM_1 > PM_{2.5}$, $PM_1 = PM_{2.5}$
7. If the missing data is more than 1/3 in an hour, the python code will auto remove all the data in the hour

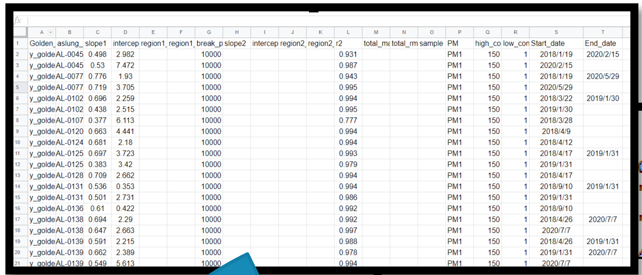
Google sheet



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Golden_	aslung_	slope1	intercep	region1_	region1_	break_p	slope2	intercep	region2_	region2_	r2	total_ma	total_rm	sample	PM	high_co	low_con	Start_date	End_date
2	y_golde	AL-0045	0.498	2.982			10000					0.931				PM1	150	1	2018/1/19	2020/2/15
3	y_golde	AL-0045	0.53	7.472			10000					0.987				PM1	150	1	2020/2/15	
4	y_golde	AL-0077	0.776	1.93			10000					0.943				PM1	150	1	2018/1/19	2020/5/29
5	y_golde	AL-0077	0.719	3.705			10000					0.995				PM1	150	1	2020/5/29	
6	y_golde	AL-0102	0.696	2.259			10000					0.994				PM1	150	1	2018/3/22	2019/1/30
7	y_golde	AL-0102	0.438	2.515			10000					0.995				PM1	150	1	2019/1/30	
8	y_golde	AL-0107	0.377	6.113			10000					0.777				PM1	150	1	2018/3/28	
9	y_golde	AL-0120	0.663	4.441			10000					0.994				PM1	150	1	2018/4/9	
10	y_golde	AL-0124	0.681	2.18			10000					0.994				PM1	150	1	2018/4/12	
11	y_golde	AL-0125	0.697	3.723			10000					0.993				PM1	150	1	2018/4/17	2019/1/31
12	y_golde	AL-0125	0.383	3.42			10000					0.979				PM1	150	1	2019/1/31	
13	y_golde	AL-0128	0.709	2.662			10000					0.994				PM1	150	1	2018/4/17	
14	y_golde	AL-0131	0.536	0.353			10000					0.994				PM1	150	1	2018/9/10	2019/1/31
15	y_golde	AL-0131	0.501	2.731			10000					0.986				PM1	150	1	2019/1/31	
16	y_golde	AL-0136	0.61	0.422			10000					0.992				PM1	150	1	2018/9/10	
17	y_golde	AL-0138	0.694	2.29			10000					0.992				PM1	150	1	2018/4/26	2020/7/7
18	y_golde	AL-0138	0.647	2.663			10000					0.997				PM1	150	1	2020/7/7	
19	y_golde	AL-0139	0.591	2.215			10000					0.988				PM1	150	1	2018/4/26	2019/1/31
20	y_golde	AL-0139	0.662	2.389			10000					0.978				PM1	150	1	2019/1/31	2020/7/7
21	y_golde	AL-0139	0.549	5.613			10000					0.994				PM1	150	1	2020/7/7	

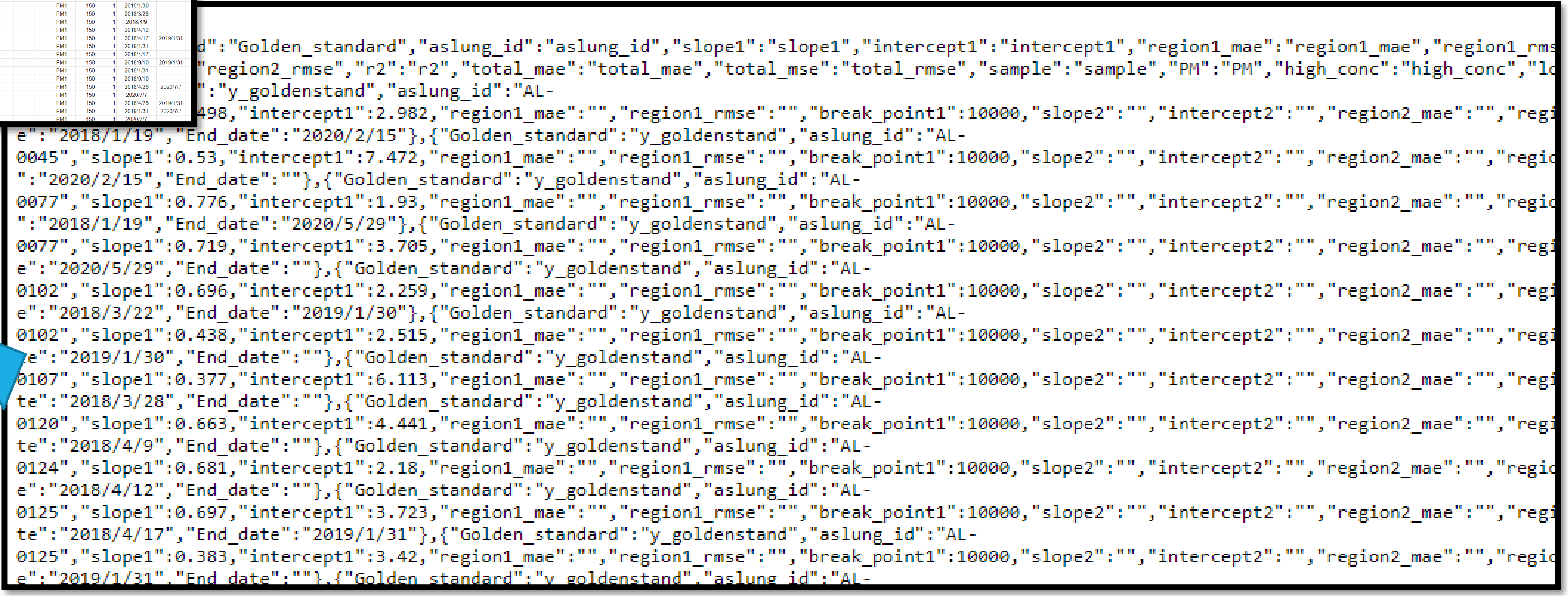
Calibration factor API

https://script.google.com/macros/s/AKfycbwfhUbpNqk5AE4HpUg0Dp-0pT1oMKa1mxLzWWAXb3dlnhTYRN8/exec



fx

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Golden_	aslung_	slope1	intercep	region1_	region1_	break_p	slope2	intercep	region2_	region2_	r2	total_ma	total_rm	sample	PM	high_co	low_con	Start_date	End_date
2	y_golde	AL-0045	0.498	2.982			10000					0.931				PM1	150	1	2018/1/19	2020/2/15
3	y_golde	AL-0045	0.53	7.472			10000					0.987				PM1	150	1	2020/2/15	
4	y_golde	AL-0077	0.776	1.93			10000					0.943				PM1	150	1	2018/1/19	2020/5/29
5	y_golde	AL-0077	0.719	3.705			10000					0.995				PM1	150	1	2020/5/29	
6	y_golde	AL-0102	0.696	2.259			10000					0.994				PM1	150	1	2018/3/22	2019/1/30
7	y_golde	AL-0102	0.438	2.515			10000					0.995				PM1	150	1	2019/1/30	
8	y_golde	AL-0107	0.377	6.113			10000					0.777				PM1	150	1	2018/3/28	
9	y_golde	AL-0120	0.663	4.441			10000					0.994				PM1	150	1	2018/4/9	
10	y_golde	AL-0124	0.681	2.18			10000					0.994				PM1	150	1	2018/4/12	
11	y_golde	AL-0125	0.697	3.723			10000					0.993				PM1	150	1	2018/4/17	2019/1/31
12	y_golde	AL-0125	0.383	3.42			10000					0.979				PM1	150	1	2019/1/31	
13	y_golde	AL-0128	0.709	2.662			10000					0.994				PM1	150	1	2018/4/17	
14	y_golde	AL-0131	0.536	0.353			10000					0.994				PM1	150	1	2018/9/10	2019/1/31
15	y_golde	AL-0131	0.501	2.731			10000					0.986				PM1	150	1	2019/1/31	
16	y_golde	AL-0136	0.61	0.422			10000					0.992				PM1	150	1	2018/9/10	
17	y_golde	AL-0138	0.694	2.29			10000					0.992				PM1	150	1	2018/4/26	2020/7/7
18	y_golde	AL-0138	0.647	2.663			10000					0.997				PM1	150	1	2020/7/7	
19	y_golde	AL-0139	0.591	2.215			10000					0.988				PM1	150	1	2018/4/26	2019/1/31
20	y_golde	AL-0139	0.662	2.389			10000					0.978				PM1	150	1	2019/1/31	2020/7/7
21	y_golde	AL-0139	0.549	5.613			10000					0.994				PM1	150	1	2020/7/7	

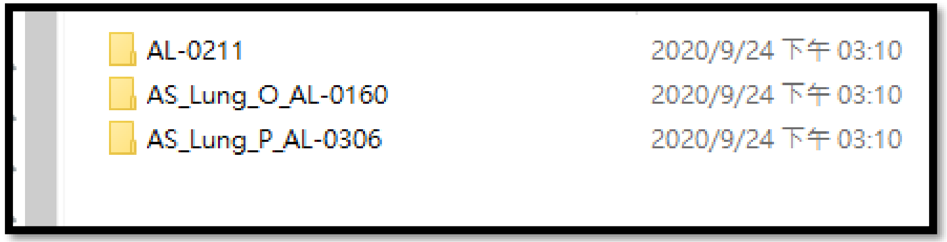


```
d":"Golden_standard","aslung_id":"aslung_id","slope1":"slope1","intercept1":"intercept1","region1_mae":"region1_mae","region1_rms
"region2_rmse","r2":"r2","total_mae":"total_mae","total_mse":"total_rmse","sample":"sample","PM":"PM","high_conc":"high_conc","lo
":"y_goldenstand","aslung_id":"AL-
498,"intercept1":2.982,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regi
e":"2018/1/19","End_date":"2020/2/15"},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0045","slope1":0.53,"intercept1":7.472,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regio
":"2020/2/15","End_date":""},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0077","slope1":0.776,"intercept1":1.93,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regio
":"2018/1/19","End_date":"2020/5/29"},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0077","slope1":0.719,"intercept1":3.705,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regi
e":"2020/5/29","End_date":""},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0102","slope1":0.696,"intercept1":2.259,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regi
e":"2018/3/22","End_date":"2019/1/30"},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0102","slope1":0.438,"intercept1":2.515,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regi
te":"2019/1/30","End_date":""},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0107","slope1":0.377,"intercept1":6.113,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regi
te":"2018/3/28","End_date":""},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0120","slope1":0.663,"intercept1":4.441,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regi
te":"2018/4/9","End_date":""},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0124","slope1":0.681,"intercept1":2.18,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regio
e":"2018/4/12","End_date":""},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0125","slope1":0.697,"intercept1":3.723,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regi
te":"2018/4/17","End_date":"2019/1/31"},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
0125","slope1":0.383,"intercept1":3.42,"region1_mae":"","region1_rmse":"","break_point1":10000,"slope2":"","intercept2":"","region2_mae":"","regio
e":"2019/1/31","End_date":""},{"Golden_standard":"y_goldenstand","aslung_id":"AL-
```

We create calibration factor API from google sheet. Keep the format of the google sheet is very important.

Data prepare

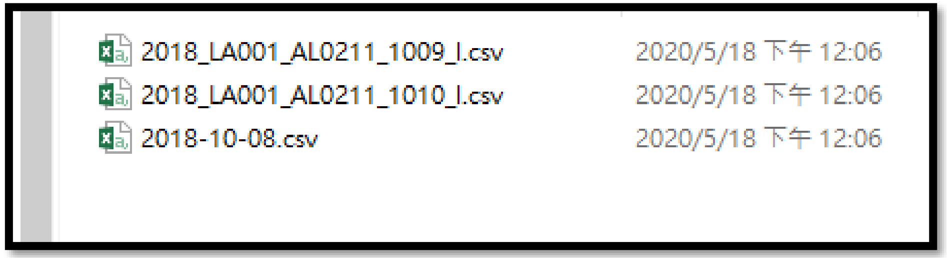
1.Folder name of the data set should be contain AS-Lung ID, EX:AL-0001



AL-0211 2020/9/24 下午 03:10
AS_Lung_O_AL-0160 2020/9/24 下午 03:10
AS_Lung_P_AL-0306 2020/9/24 下午 03:10

AS-Lung ID only
Or
Lab ID before AS-Lung ID ,
separate by “_”

2.Rename the filename is optional

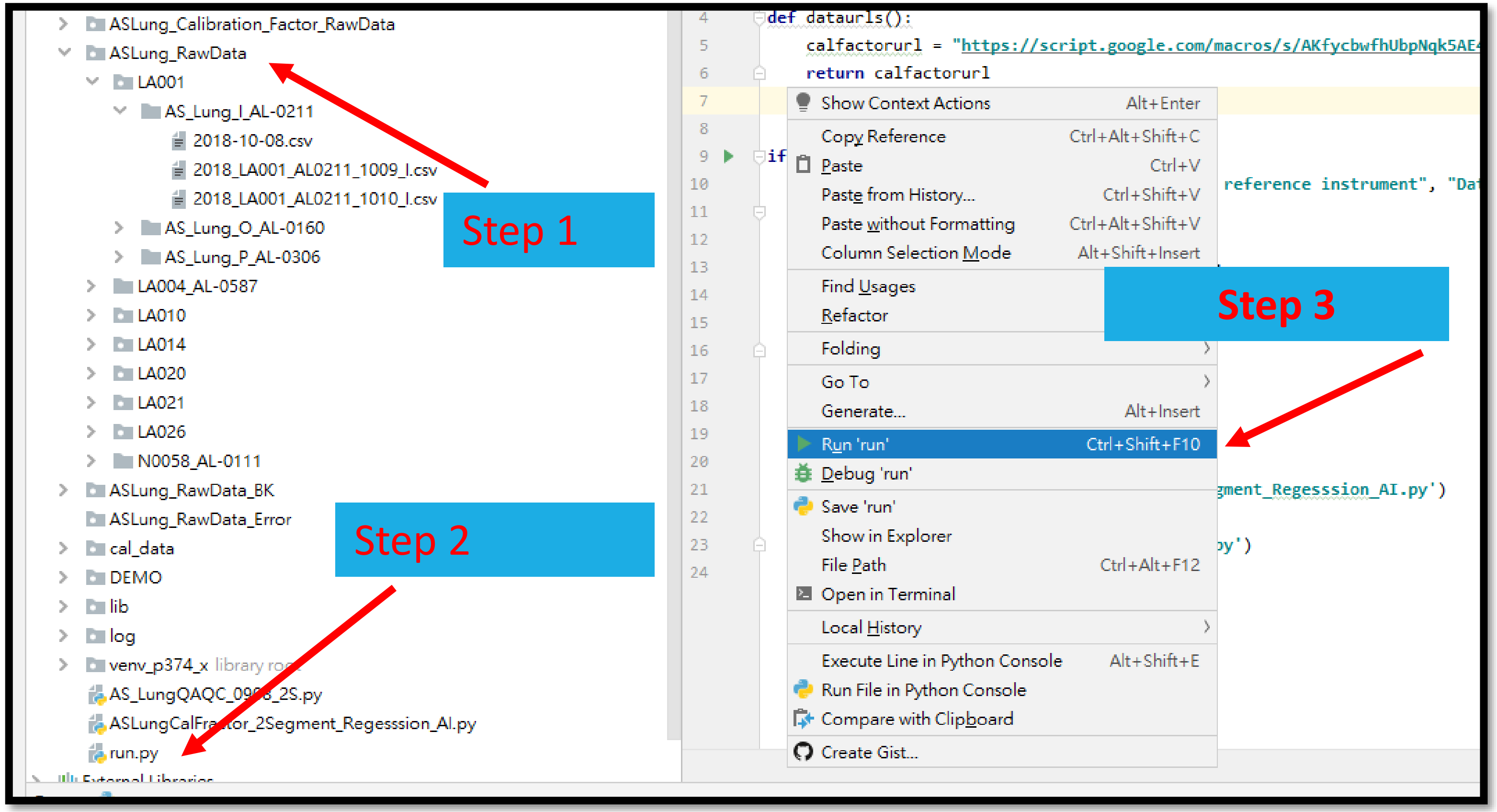


2018_LA001_AL0211_1009_l.csv 2020/5/18 下午 12:06
2018_LA001_AL0211_1010_l.csv 2020/5/18 下午 12:06
2018-10-08.csv 2020/5/18 下午 12:06

Keep the original filename
Or
Rename the filename
it is optional

How to run data cleaning

3. Open PyCharm -> project -> run.py



```
def dataurls():
    calfactorurl = "https://script.google.com/macros/s/AKfycbwfhUbpNqk5AE4
    return calfactorurl
```

Step 1

Step 2

Step 3

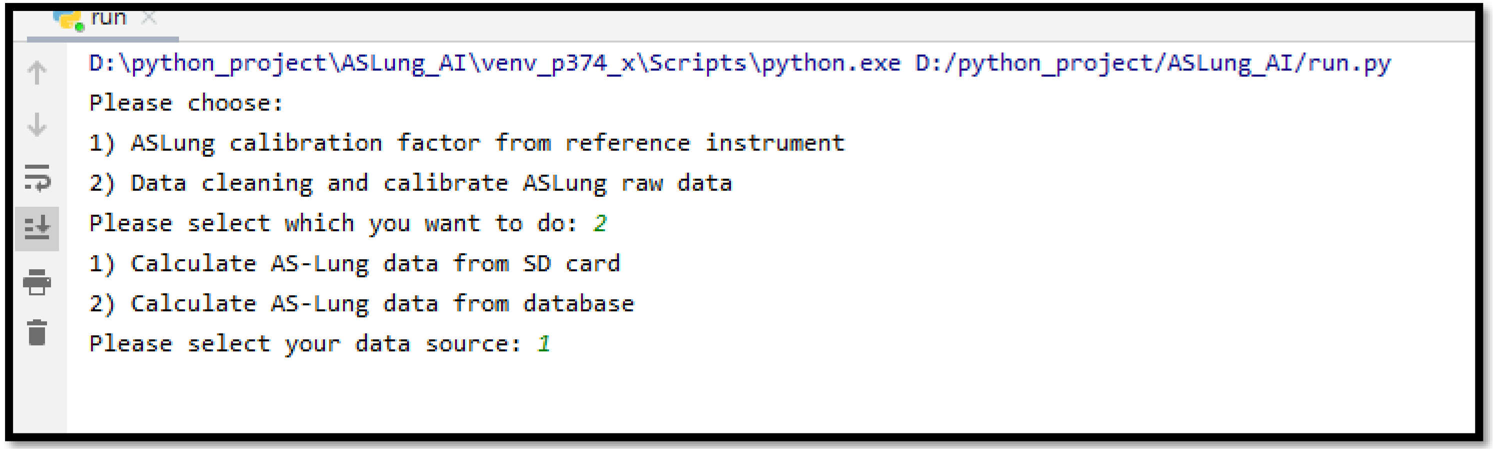
Step 1:
Copy raw data to the folder of "ASLung_RawData"

Step 2:
Open run.py (double click)

Step 3:
Run the python code

How to run data cleaning

4. Select 2) Data cleaning and calibrate ASLung raw data

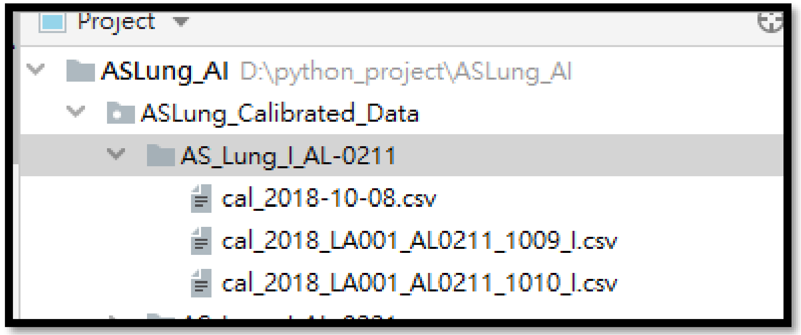


```
D:\python_project\ASLung_AI\venv_p374_x\Scripts\python.exe D:/python_project/ASLung_AI/run.py
Please choose:
1) ASLung calibration factor from reference instrument
2) Data cleaning and calibrate ASLung raw data
Please select which you want to do: 2
1) Calculate AS-Lung data from SD card
2) Calculate AS-Lung data from database
Please select your data source: 1
```

5. Select your data source, SD card or database

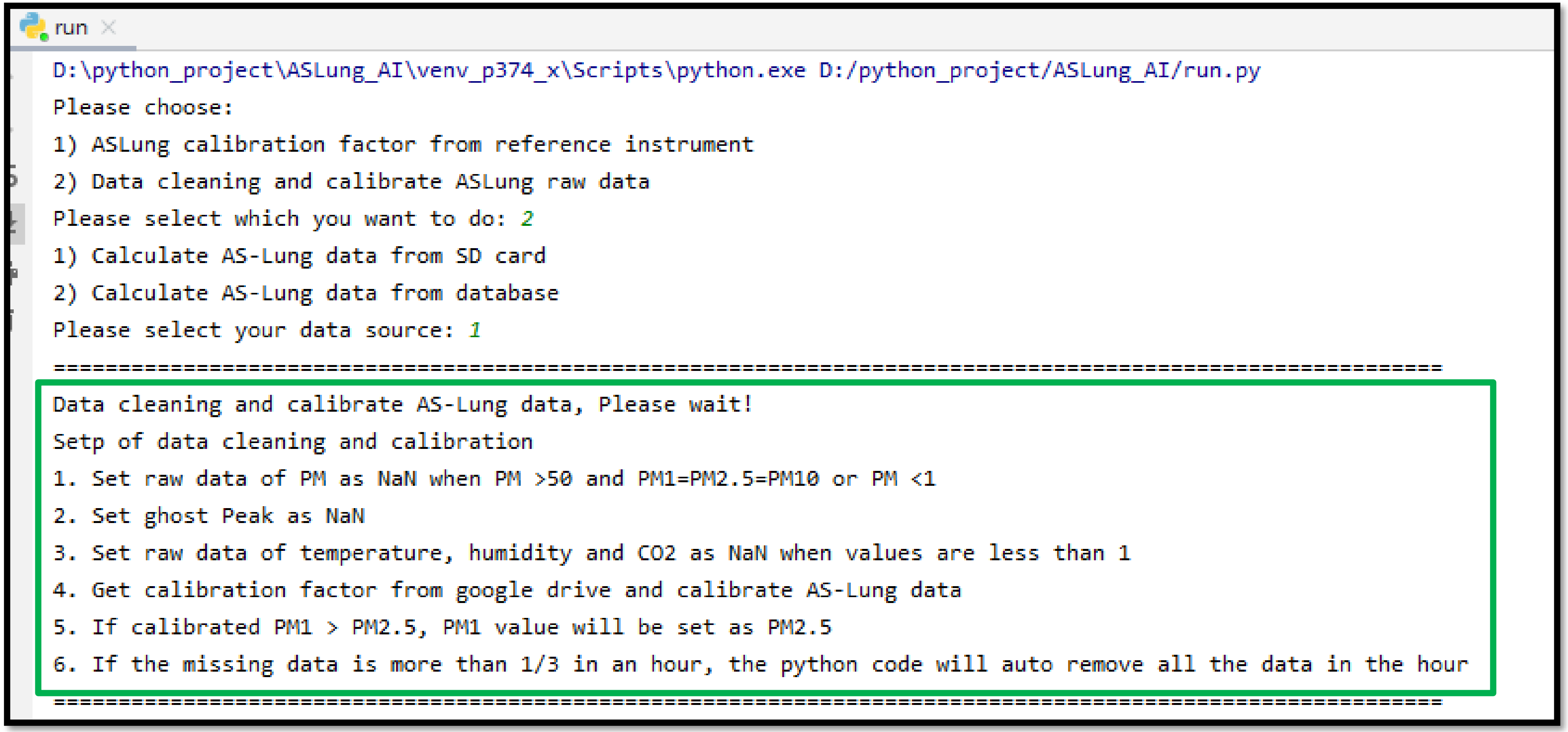
6 . Calibrated data will be save in the folder of

“ASLung_Calibrated_Data”, the filename will add prefix of “cal_” in the front of original filename



```
ASLung_AI D:\python_project\ASLung_AI
  ASLung_Calibrated_Data
    AS_Lung_I_AL-0211
      cal_2018-10-08.csv
      cal_2018_LA001_AL0211_1009_I.csv
      cal_2018_LA001_AL0211_1010_I.csv
```

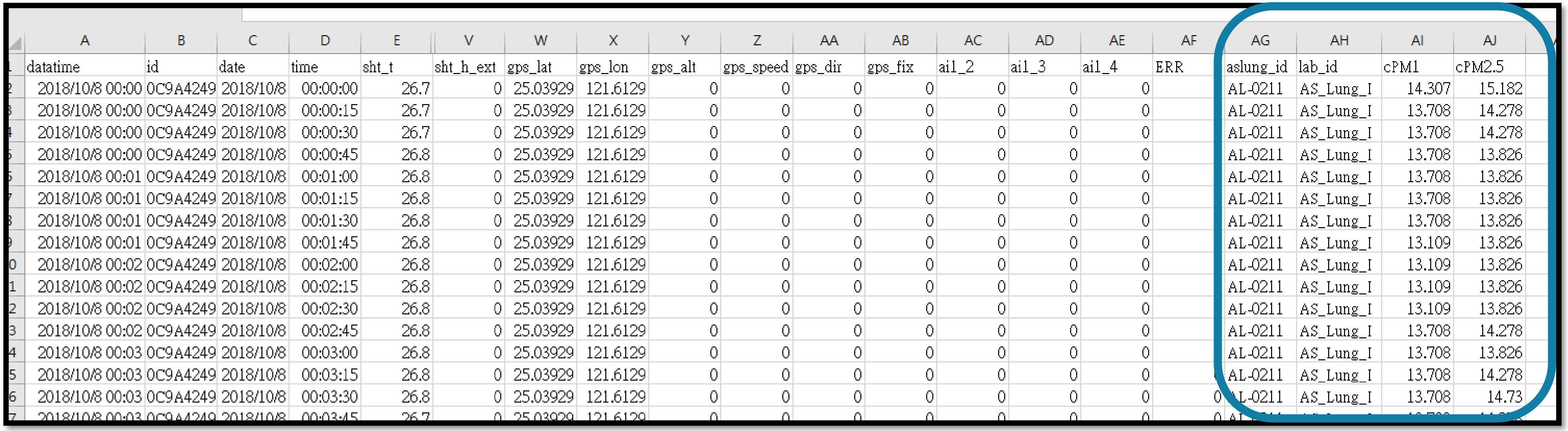
Step of data cleaning will remind again when run the python code



```
D:\python_project\ASLung_AI\venv_p374_x\Scripts\python.exe D:/python_project/ASLung_AI/run.py
Please choose:
1) ASLung calibration factor from reference instrument
2) Data cleaning and calibrate ASLung raw data
Please select which you want to do: 2
1) Calculate AS-Lung data from SD card
2) Calculate AS-Lung data from database
Please select your data source: 1
=====================================================================================================
Data cleaning and calibrate AS-Lung data, Please wait!
Setp of data cleaning and calibration
1. Set raw data of PM as NaN when PM >50 and PM1=PM2.5=PM10 or PM <1
2. Set ghost Peak as NaN
3. Set raw data of temperature, humidity and CO2 as NaN when values are less than 1
4. Get calibration factor from google drive and calibrate AS-Lung data
5. If calibrated PM1 > PM2.5, PM1 value will be set as PM2.5
6. If the missing data is more than 1/3 in an hour, the python code will auto remove all the data in the hour
=====================================================================================================
```

Open CSV file after calibrated

Open the calibrated csv file, AS-Lung ID, lab id , calibrated PM_1 and $PM_{2.5}$ will add in the end column of the data set

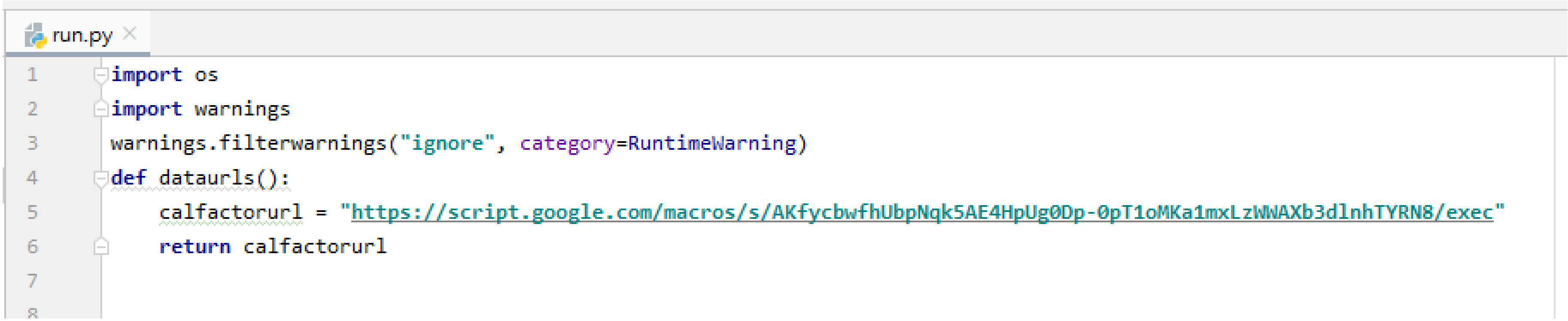


	A	B	C	D	E	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1	datatime	id	date	time	sht_t	sht_h_ext	gps_lat	gps_lon	gps_alt	gps_speed	gps_dir	gps_fix	ai1_2	ai1_3	ai1_4	ERR	aslung_id	lab_id	cPM1	cPM2.5
2	2018/10/8 00:00	0C9A4249	2018/10/8	00:00:00	26.7	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	14.307	15.182
3	2018/10/8 00:00	0C9A4249	2018/10/8	00:00:15	26.7	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.708	14.278
4	2018/10/8 00:00	0C9A4249	2018/10/8	00:00:30	26.7	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.708	14.278
5	2018/10/8 00:00	0C9A4249	2018/10/8	00:00:45	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.708	13.826
6	2018/10/8 00:01	0C9A4249	2018/10/8	00:01:00	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.708	13.826
7	2018/10/8 00:01	0C9A4249	2018/10/8	00:01:15	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.708	13.826
8	2018/10/8 00:01	0C9A4249	2018/10/8	00:01:30	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.708	13.826
9	2018/10/8 00:01	0C9A4249	2018/10/8	00:01:45	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.109	13.826
0	2018/10/8 00:02	0C9A4249	2018/10/8	00:02:00	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.109	13.826
1	2018/10/8 00:02	0C9A4249	2018/10/8	00:02:15	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.109	13.826
2	2018/10/8 00:02	0C9A4249	2018/10/8	00:02:30	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.109	13.826
3	2018/10/8 00:02	0C9A4249	2018/10/8	00:02:45	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.708	14.278
4	2018/10/8 00:03	0C9A4249	2018/10/8	00:03:00	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0		AL-0211	AS_Lung_I	13.708	13.826
5	2018/10/8 00:03	0C9A4249	2018/10/8	00:03:15	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0	0	AL-0211	AS_Lung_I	13.708	14.278
6	2018/10/8 00:03	0C9A4249	2018/10/8	00:03:30	26.8	0	25.03929	121.6129	0	0	0	0	0	0	0	0	AL-0211	AS_Lung_I	13.708	14.73
7	2018/10/8 00:03	0C9A4249	2018/10/8	00:03:45	26.7	0	25.03929	121.6129	0	0	0	0	0	0	0	0	AL-0211	[illegible]	[illegible]	[illegible]

How to update calibration factor API link

1.Open run.py

2.Go to line 5, find **calfactorurl = “xxxxxxxxxxx”**



```
1  import os
2  import warnings
3  warnings.filterwarnings("ignore", category=RuntimeWarning)
4  def dataurls():
5      calfactorurl = "https://script.google.com/macros/s/AKfycbwfhUbpNqk5AE4HpUg0Dp-0pT1oMKa1mxLzWWAXb3dlnhTYRN8/exec"
6      return calfactorurl
7
8
```

3. Past the new link after **calfactorurl=“**

You can get your google sheet and calibration factor API link from “GoogleSheetAndAPI.xlsx

Thanks!