

# ECWMF

## ESoWC

### Challenge 24 Air Quality from OpenAQ Quality Control implementation

## Milestone 3

### Pecos Quality Control Classification of Measurements and Stations

## User Manual

### Interaction

- 1. Choose OpenAQ dataset to import**
- 2. Choose Lower bound of values, higher bound of values, expected timestep of measurements and increment on measurements**
- 3. Get a report of all outliers identified in test\_result.csv**

### Python Script

1. Open Python Script in your independent development environment. This could be using anaconda or other.
2. Import the python libraries that are required either using conda install or pip install commands. These are the required libraries:

Openaq (<http://dhhagan.github.io/py-openaq/installing.html>)

Pandas ([https://pandas.pydata.org/pandas-docs/stable/getting\\_started/install.html](https://pandas.pydata.org/pandas-docs/stable/getting_started/install.html))

Pecos (<https://pecos.readthedocs.io/en/stable/installation.html>)

matplotlib.pyplot ([https://pandas.pydata.org/pandas-docs/stable/getting\\_started/install.html](https://pandas.pydata.org/pandas-docs/stable/getting_started/install.html))

e.g. Install using: pip install openaq

3. It firstly imports OpenAQ dataset and a user can choose which Country to import. The default setting is India.
  - a. To change the country get the country code i.e. IN for India
  - b. Edit the script on line 25 on '#Step 0 Choose the measurement country to import'
  - c. Edit where is states country='IN' to country='Your choice'

4. a. To choose the parameter change the parameter setting on line 25 in '#Step 0 Choose the measurement country to import'.

b. It is defaulted to 'pm25' though it can be changed to another parameter by changing: **parameter='pm25'** to **parameter='your\_choice'** it should be one of

```
'o3', 'no2', 'so2', 'pm10', 'pm25'
```

5. In the next step it creates the Pecos monitoring object
6. In step 5 it checks for missing data
7. A range of acceptable values can be chosen in '# Step 6 Choose acceptable value range and Check data for expected ranges'
  - a. To choose these by editing it on step 6 on line 50 to choose a lower and higher bound. Change 0 and 100 to your lower and high bound: **pm.check\_range([0, 100], 'value')** to **pm.check\_range([your\_lower\_bound, higher\_bound], 'value')**
8. In Step 7 you can choose expected frequency of the timestamp between measurements. This could be 900 which is 15 minutes or any multiple. It can be changed on line 42 by editing it: **#pm.check\_timestamp(900)** to **pm.check\_timestamp(your\_choice)**

(It is recommended to commented it out because it finds many non monotonic timestamps and can be used by removing the # from it )
9. To change the increment of measurements change it on line 54 on '#Step 8 Choose acceptable increment from measurements of 15 minutes and check for abrupt changes between consecutive time steps' . Edit it from 10 to your choice.  
**pm.check\_increment([None, 10], 'value')** to **pm.check\_increment([None, your\_choice], 'value')**
10. In # Step 11 its write the Pecos test results and reports the outliers to test\_results.csv and monitoringreport.html.