

ECWMF

ESoWC Challenge 24 Air Quality from OpenAQ Quality Control implementation

Milestone 3

Pecos Quality Control of Air Quality Measurements

User Manual

Interaction

1. Choose OpenAQ dataset to import (Cities, Country, Select Stations or Latitude/Longitude Centre and Radius)
2. Choose the sources of OpenAQ dataset (csv download, ndjson download or API)
3. Choose Lower bound of values, higher bound of values, expected timestep of measurements and increment on measurements
4. Get a report of all outliers identified in test_result.csv

Python Script

1. Choose to 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)

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

matplotlib.pyplot (https://pandas.pydata.org/pandas-docs/stable/getting_started/install.html)

seaborn (<https://seaborn.pydata.org/#>)

csv

e.g. Install using: pip install openaq

or

conda install matplotlib

3. Run the python script to test it and evaluate the results

It produces these reports:

4. Choose what selection of the OpenAQ dataset to import
 - a. Cities – Select every station in chosen City
 - b. Country – Select every station in chosen country

“Milestone3_Pecos_EveryStationCountry_Import_OpenAQ_Station_QC_Completed.py”

- c. One Station – The Chosen station

“Milestone3_Pecos_QualityControl_OpenAQStationimportOpenAQAPIdataset_Completed.py”

- d. Latitude / Longitude Coordinates and Radius – The Latitude, longitude of a centre and radius around it. It selects every station in radius

“Milestone3_Pecos_QualityControl_Coordinates_OpenAQ_Deployed.py”

5. Depending on what decision was for step 4 choose source of import of OpenAQ dataset and open the python script for it

Importing from API

Every selection from step 4 is available

Importing from CSV

Only selection Step 4 d is available

Importing from ndsjson

Only selection Step 4 d is available

6. It firstly imports OpenAQ dataset and a user can choose which some selections

Depending on the selection of Step 4 change selection to your selection

- a. Cities :
- b. Country:
- c. Selection of stations

- d. One Station:
- e. Latitude / Longitude Coordinates and Radius:

In Country 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 it states country='IN' to country='Your choice'

- 7. a. To choose the parameter change the parameter setting

In Country 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'
```

- 8. Choose the time Schedule by editing the Dt_Begin and Dt_End

These are in format (25,1,2020)

- 9. In the next steps it creates the Pecos monitoring object. These can be edited in the method. "Milestone3_Pecos_Complete_QualityControl_One_OpenAQStation" It maybe different put shall begin with Milestone_Pecos_Quality_Control

- 10. Go through the method to find where to edit the Pecos Quality control search criteria. In step 5 it checks for missing data

- 11. 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')**

- 12. 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)

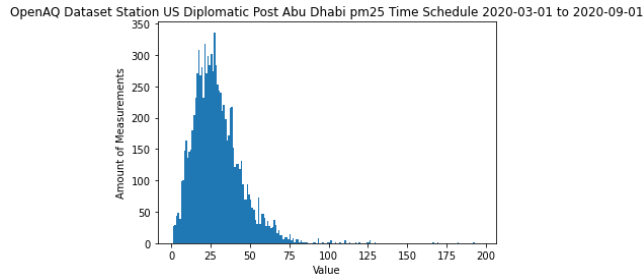
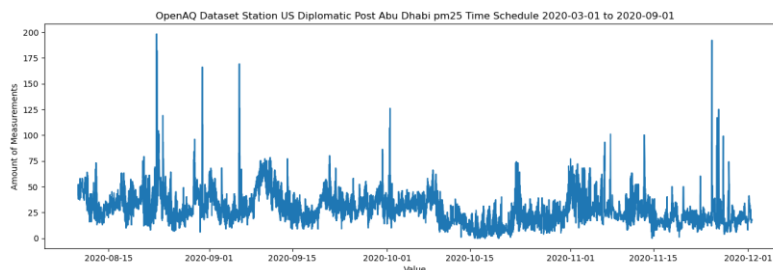
13. To change the increment of measurements change it on line 54 on 'Step 8 Choose acceptable increment from measurements of 15 minutes and check for abrupt changes between consecutive time steps' . Edit it from 10 to your choice.

`pm.check_increment([None, 10], 'value')` to `pm.check_increment([None, your_choice], 'value')`

14. In # Step 11 its write the Pecos test results and reports the outliers to `test_results.csv` and `monitoringreport.html`.

15. Next process the python script when completed

16. It firstly prints Line Plots and Histograms for Stations in the OpenAQ dataset selection



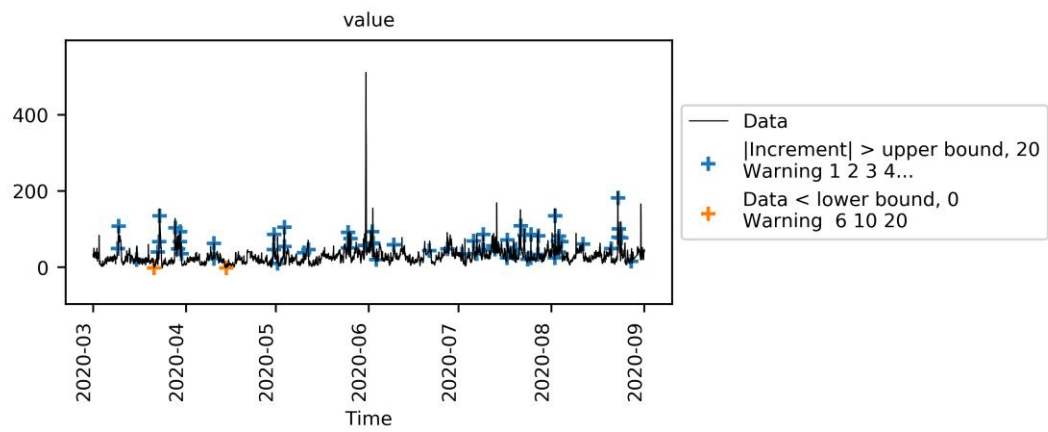
17. It next prints out the Outliers

18. It next prints the monitoring reports per Station and results with the graph:

`Monitoring_Report_Station.html` – Report with Performance Metric and Outliers

`OpenAQ_Results_Station_Station.csv` – OpenAQ Dataset

`Test_results_Station.csv` – Outliers



19. Next it prints Nearest Highway to Mean of Dataset in a Scatter graph