# ECWMF

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

## Milestone 3

**Pecos Quality Control of Air Quality Measurements**

1. **Choose OpenAQ dataset to import (Cities, Country, Select Stations or Latitude/Longitude Centre and Radius)**

2. **Choose Lower bound of values, higher bound of values, expected timestep of measurements, acceptable stagnant measurements, upper bound on averaged measurements and maximum increment on measurements**

3. **Get a report of all outliers identified in test_result.csv**

**Python Script**

1. Choose 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:

   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)

   numpy (https://numpy.org/install/)

   Datetime (https://docs.python.org/3/library/datetime.html)

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

   csv  (https://docs.python.org/3/library/csv.html)

   e.g. Install using:  pip install openaq

   or

   conda install matplotlib

3. Run the python script to test it and evaluate the results or go to step 4 to choose your own parameters

OpenAQ Dataset:

It imports a downloaded OpenAQ CSV from Milestone 1 that is included in the Github download. This can be changed to your own downloaded OpenAQ CSV when processed the python script from Milestone 1 and copy the address printed.

i.e. "OpenAQ_Dataset Unique selection pm25 CoordinateCentreandRadius 2020-03-01 to 2020-09-01.csv" or other

It produces these reports for every station from the OpenAQ dataset:

1 OpenAQDataset QC 'parameter' iteration 'iteration' 'OpenAQStation' Monitoring Report.html

The page with the Outlier and Graph from Pecos Quality Control

2 Test_Results OpenAQDataset QC 'parameter' iteration 'iteration' 'OpenAQStation' Results .csv

The Outliers from Pecos Quality Control

3 OpenAQDataset QC 'parameter' iteration 'iteration' 'OpenAQStation' Metrics.csv

The OpenAQ dataset Metrics from the Pecos Quality Control

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_OpenAQStationimportOpenAQ APIdataset_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.  Step 1 Choose your downloaded OpenAQ dataset and the QC iteration

    a To import your downloaded OpenAQ dataset change the parameter in Step 1. It is on line 561.

    Change 'OpenAQDatasetSelected'

     to your OpenAQ dataset

    b Choose an iteration that is used in the results to determine the iteration of the Pecos Quality Control on your downloaded OpenAQ dataset. It is on line 568

    Change 'iteration_OpenAQStations'

    To your iteration

    c When you have chosen the Coordinates and Radius in Milestone 1 to download your OpenAQ dataset you can add in the Coordinate and Radius. It is on line 574 and 576

6.  Step 4 Determine the parameter of your OpenAQ dataset. The step 2 and 3 print out the attributes of the OpenAQ dataset and the unique OpenAQ Stations

    a.  When your OpenAQ dataset has multiple parameters and you only want one parameter in the Pecos Quality Control

    b.  To choose the parameter change the parameter settings from the default to your selection. It is on the line 631

        Change 'Parameter_Default'

        To your selection from these

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

    c.  The parameter needs to be changed to determine only one parameter. It the line 633

        Change  'OnlyOneParameter'

        To 1

        It allows only one parameter to be applied to the Pecos Quality Control

7.  Step 5 Choose the time Schedule by editing the Dt_Begin and Dt_End

    a.  It finds the begin and end date for your OpenAQ dataset.

    b.  It can be changed to different begin and end that is within the begin and end of your OpenAQ dataset

        Change 'ChooseTimeScheduleNoorYes'

        To 1

It is allows choice of begin and end

   c. The begin and end can be selected

Change 'dt_begin' and 'dt_end'

To begin and end

These are in format (25,1,2020)

8. Step 6 Allows to choose removing the negative measurements and -999.99 measurements

   a. The measurements -999.99 are those not measured. These can be removed. It is the line 722

Change 'Remove_Neg'

To 1

It removes -999.99 measurements

   b. The negative measurements can be removed.

Change 'Remove_Negative_Measurements'

To 1

It removes negative measurements

In Steps 7 and 8 It formats the OpenAQ dataset for the Pecos Quality Control

In the next steps 9 to 10 it graph the OpenAQ dataset. It creates the Pecos monitoring object. The Pecos Quality control can be edited in the method. "Milestone3_Pecos_Complete_QualityControl_One_OpenAQStation" It maybe different put shall begin with Milestone3_Pecos_Quality_Control

9. Go through the method to find where to edit the Pecos Quality control search criteria. There are steps within it.

10. In Step 4 you can choose expected frequency of the timestamp between measurements. This is criteria 1. This could be 900 which is 15 minutes or any multiple. It can be changed on line 42 by editing it:

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 )

In step 5 it checks for missing data

11. The criteria 2 is 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: from the Line 223

LowerBound = **Your Choice**

HigherBound = **Your Choice**

12. The criteria 3 is the acceptable stagnant measurements and the expected change in the measurements that determine the measurements are not stagnant. It done over a timeschedule. These can be edited

Change 'DeltaTimeSchedule'

To your time schedule and 900 is 15 mins

Change DeltaLowerBound

To your choice of lower bound is often None

Change DeltaHigherBound

To your choice of expected minimum change in measurement over time schedule

13. The criteria 4 is maximum increment. To change the increment of measurements change 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.

Increment_Increase = **Your Choice**

Increment_Decrease = **Your Choice**

14. The criteria 5 is the upper and lower bound over normalised OpenAQ measurements by applying average and standard deviation over a chosen time schedule. These can be edited

Change 'UpperBoundOutlier'

To your choice of upper bound

Change 'LowerBoundOutlier'

To your choice of lower bound

Change 'TimeSchedule'

To your choice of time schedule

15. To find out more about these quality control criteria go to https://pecos.readthedocs.io/en/latest/qc_tests.html

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

17. Next process the python script when completed

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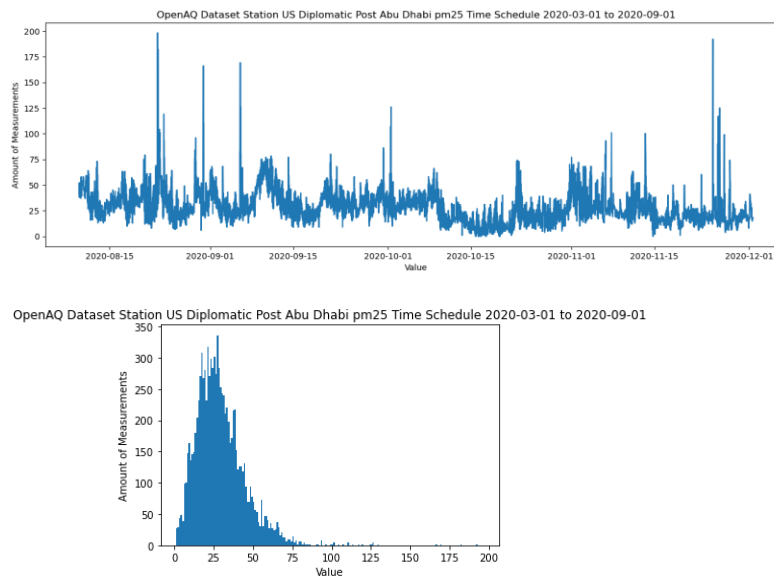




19. It next prints out the Outliers

20. 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

These are:

> 1 OpenAQDataset QC 'parameter' iteration 'iteration' 'OpenAQStation' Monitoring Report.html
>
> The page with the Outlier and Graph from Pecos Quality Control
>
> 2 Test_Results OpenAQDataset QC 'parameter' iteration 'iteration' 'OpenAQStation' Results .csv
>
> The Outliers from Pecos Quality Control
>
> 3 OpenAQDataset QC 'parameter' iteration 'iteration' 'OpenAQStation' Metrics.csv
>
> The OpenAQ dataset Metrics from the Pecos Quality Control

value

400

200

0

2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
2020-09

Time

Data

|Increment| > upper bound, 20
Warning 1 2 3 4...

Data < lower bound, 0
Warning  6 10 20