Air pollution analysis

Introduction

Many studies suggest the quality of air has been significantly improving in the last years in the majority of the world regions. However, air quality still creates a significant problem in Europe, especially in some densely populated urban areas and during certain weather conditions. Several reports observe the serious impact of the air pollution on the people's health and many analysis and models have been tested to understand and finally reduce the problem.

The air pollution primarily results from:

- Natural processes (soil erosion, volcano eruptions)
- Human activity, which includes three major sources:
  - Industry pollution
  - Traffic pollution (air exhaust, brake and tire wear, dust resuspension from roads, air and sea traffic)
  - House heating

In addition to the increasing level of certain chemical compounds (CO, SO2, NOx, BC, etc.) in the air, a dangerous type of pollution are small particles suspended in the atmosphere generated by a variety of human activities – Particulate Matter (PM) or Particulates. PM is a type of air pollution that can travel long distances in the atmosphere and causes a wide range of diseases and a significant reduction of life expectancy in most of the population of Europe.

Objectives of the activity

A professional air pollution analysis is very complex: it requires advanced conceptual models, analytical tools, access to many data sets, profound domain knowledge, considerable research experience and multidisciplinary knowledge. From this perspective obtaining scientifically significant results in a short period by high school teams is very difficult.

However, due to the availability of publicly accessible data and relatively easy interpretation of commonly known dependences (e.g. population density on air pollution level, wind direction on pollution in a given place), this topic is a good training area for inexperienced young scientists who want to develop their scientific

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 710577

analytical and reasoning skills. Working on this problem the student will have an opportunity to exercise the basic steps of real scientific research:

- Search, select and validate information on a given topic (air pollution problem).
- Identify data sources for analyses; understand the data structure and data meaning.
- Formulate hypotheses and verify them based on available data sets and analytical methods.
- Analyze of different data sets using basic statistical concepts.
- Closely review work of their colleagues and agree on common interpretation.
- Understand pollution measurements data and correlation with other factors (demographical, geographical, weather, economic).
- Communicate in the team, defending one's own point of view and working out a consensus.
- Generalize different findings and draw conclusions.

Scope of the work and expected results

The students are expected to perform limited analysis of air pollution level in selected European areas based on publicly available data bases; identify, describe and try to explain characteristic regularities observed; study how other factors (like population density, wind direction, air temperature, geographical conditions etc.) influence the level of different chemical compounds and dust in the air. The analysis should include a few regions/places located in at least 3 different European countries. Observations related to separate areas should be compared, explained and generalized. Long and short period trends (daily, annual, etc.) are expected to be identified and correlated with other factors. The student may also try to investigate how air pollution travel within some areas, or/and between countries (depending on the wind speed and direction). Finally, the analysis could be supplemented by additional information (press report on pollution, publications, similar analysis) about the regions/cities the teams will consider.

Suggested structure of the report

- Short, general introduction to the air pollution problem (why it is important, which areas/countries are especially endangered, what measure can be taken to limit negative consequences etc.)
- Areas/countries and which types of data the team has decided to include in its analysis and why
- Short description the technical aspects of your analysis (type of analysis, software tools, scope of data used)
- Hypotheses confirmed, rejected; characterize results obtained for separate areas.
- Final conclusions and summary of observation.

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 710577

Hints on tools and analytical approach

The teams may use any analytical tools, however open source and open formats are preferred. Additionally, the team should provide precise information which method or algorithm was used to obtain a given result. It is highly recommended to use spreadsheets and other popular software packages like R. Although more advanced and accurate statistical concepts may also be utilized, analysis based only on the Pearson correlation coefficient (PCC) (https://en.wikipedia.org/wiki/Pearson_correlation_coefficient) are expected to be sufficient to discover the majority of patterns in the data for these analysis purposes. PCC is commonly implemented in many software tools including popular spreadsheets and open software packages like R.

Information on how to interpret the PCC values are available for example here:

http://www.dummies.com/education/math/statistics/how-to-interpret-a-correlation-coefficient-r/

An example of PCC calculation in R for eruption duration and waiting time for Old Faithful geyser in the USA Yellowstone National Park are provided here

http://www.r-tutor.com/elementary-statistics/numerical-measures/correlation-coefficient

In Excel PCC is calculated by the CORREL function:

https://support.office.com/en-us/article/CORREL-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92

and explained in numberless examples in the Internet for instance:
https://www.youtube.com/watch?v=78xBn-RcUtI

https://www.youtube.com/watch?v=sGlsdHD-lcA

There is also a dedicated R package to analyze air pollution:
http://davidcarslaw.github.io/openair/

Charts (Graphs) e.g. in spreadsheet and maps are also recommended as tools to present the project results.

Recommended data sets to be used in the task

The data sets that are primarily intended to be analyzed are referenced below

https://openaq.org/#/?_k=exn3it – community collected air quality measurements which (as for Dec. the 21st 2017) – collected 142,273,175 measurements from 8,061 locations in 61 countries). The data includes: PM2.5, PM10, ozone (O3), sulfur dioxide (SO2), nitrogen dioxide (NO2), carbon monoxide (CO), and black carbon (BC)

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 710577

Additionally, data sources which storing the following measurements may be useful:

- Wind speed and direction
- Temperature
- City population (population density on a given area)
- Rain
- Air humility / dew point temperature
- Atmospheric pressure

However, the competition participants primarily work on any publicly available data sets providing in their report references to the data sources. Additionally, the data sets used in the analysis should be also publicly available at the time the team reports are assessed (spring 2018).

Other data sets you might use:

- All EU countries are obliged to monitor air pollution by the CAFE directive. Monitoring results are public, and in some countries (in Poland) are accessible by WWW. This data can contain meteorological data.
- You can use this database: https://www.ogimet.com/resynops.phtml.en . METAR format is a format used to send meteorological data for airplane pilots, this is one of (many) websites that store and publish METAR data. If there is an airport in your city weather reports probably will be published on this site.

Examples of simple types of air quality analysis

- Select two areas: a typical rural and an urban one (high population density) and compare the level of pollution in the same period: observe and explain differences (in terms both overall pollution values and the specific types of chemical compounds dominating in these areas.
- Select one area and observe the annual trend of a pollution level. Perform the same analysis for different pollution components.
- Correlate the level of pollution for a given area with the air pressure. Then carry out a similar analysis for wind speed and direction (represented or by quantified directions, or first transform the direction value by using the sine/cosine function.
- Consider two nearly located areas; Analyze the pollution values in both of them; observe if under certain wind directions the pollution travel from one area to the second one.

References

1. https://en.wikipedia.org/wiki/Air_pollution
2. http://www.euro.who.int/__data/assets/pdf_file/0006/189051/Health-effects-of-particulate-matter-final-Eng.pdf

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 710577

3. http://www.euro.who.int/__data/assets/pdf_file/0010/157969/e96194.pdf?ua=1 Executive summary and pages: 171-4
4. http://www.euro.who.int/__data/assets/pdf_file/0006/78657/E88189.pdf
5. https://doi.org/10.1016/S1352-2310(03)00221-8 "Characterizing seasonal variations and spatial distribution of ambient PM10 and PM2.5 concentrations based on long-term Swiss monitoring data"

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT NO. 710577