Data science seminars and challenge

Development of a validation aid for air quality microsensor data

Ronald Phlypo: responsable du master Sigma
Sana Louhichi: Co-responsable du master M2 SIAM
Stéphane Socquet : Directeur Production Atmo AuRA, sujet porté par l'équipe Innovation et développement (Steve Micallef)

Atmo Auvergne-Rhône-Alpes (www.atmo-auvergnerhonealpes.fr), an observatory approved by the French Ministry of Energy Transition to monitor air quality in the Auvergne-Rhône-Alpes region, produces daily measurement data from nearly 90 reference stations spread throughout the region. These stations are equipped with automatic analysers that produce continuous hourly (and quarter-hourly) data on a number of regulated pollutants in the air.

These regulatory measuring devices are subject to regular calibration and preventive maintenance in order to guarantee their quality and reliability. Because of their metrology, the reliability of these measuring devices is particularly high. In addition, the data is validated daily by Atmo's teams. Data validation (or invalidation) marks the data with a status code: data marked (A, P , O or R) is considered usable and transmitted.

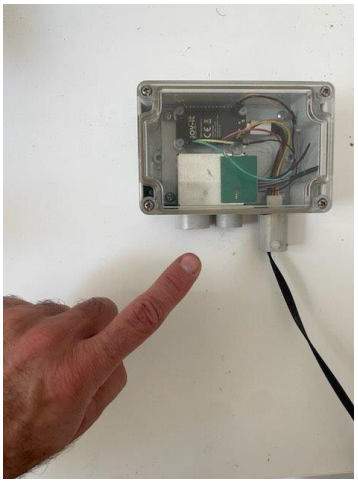
Micro-sensor technology

In recent years, the technological development of portable micro monitoring stations has opened up the possibility of miniaturising air quality measurement and making it more accessible to the general public. These innovative microstations make it possible to set up new networks of measurement points distributed throughout the country, which would augment the historical monitoring system.

The data supplied by these devices could eventually make it possible to estimate air quality more precisely, in real time, and would encourage the emergence of new services for citizens.

However, initial studies and operations show that micro-sensors are much less reliable than the 90 regulatory measurement stations. The intrinsic variability of these micro-sensors, the difficulty of calibration, and the diversity of acquisition conditions (typology, exposure, weather conditions, speed) mean that the measurements cannot be used as they stand, and require a major effort to validate the data. It has also become apparent that the accuracy of the measurements is highly dependent on the chemical composition of the atmosphere: the geographical location therefore plays a key role in the accuracy of the sensor's parameters.

To date, micro-sensors are validated on a weekly basis, by invalidating extremely high peaks. This validation is far from optimal: a certain number of invalid data are not detected. Atmo needs to optimise its validation, and is therefore looking for technical data processing solutions that could improve the validation performance of micro-sensor data.



Statutory air quality measurement station (left) and a microsensor (right) both measuring fine particles PM2.5

The proposed challenge is to devise data processing tools and methods for automatically validating micro-sensor data in real time, based on regulatory reference data collocated with certain micro-sensors.

The ultimate aim would be to provide air quality experts with a decision-making aid. This could involve proposing a selection of data considered to be atypical or invalid, which a person could decide on a daily basis (using a dashboard, for example).

The subject will involve 3 main stages:

1- Study of the variability of the sensor compared to a colocated regulatory analyser
2- Invalidation of microsensor data in comparison with a co-located analyser
3- Invalidation of micro-sensors throughout the network, and in particular sensors that are not co-located with a regulatory analyser

To meet this challenge, it will be necessary to take into account variables that complement air quality data: geographical location, the environment of each measurement (urban site, proximity to roads, proximity to industry, rural area, etc.), meteorological parameters and correlations between parameters. For example, meteorology has a major influence on pollutant levels.

The selected technical proposal will have to be adapted to a real situation and will constitute a technical development lead for Atmo.

The following are available for work:

- 2 years of PM2.5 fine particle data measured by 40 microsensors at hourly intervals in the Grenoble area (data to be validated).
- 2 years of technical data measured by the 40 micro-sensors: temperature and relative humidity.
- 2 years series of measurements at regulatory stations in the Grenoble area on the pollutants nitrogen dioxide, ozone and particulate matter (PM10 and PM2.5).
- 2 years of meteorological measurements (Atmo AuRA and météoFrance)
- Documentary resources and access to Atmo AuRA experts to answer questions

Ressources :
Drive : 2024_DATASciences_Challenge (https://airrhonealpes-my.sharepoint.com/:f:/r/personal/atmoaura_atmo-aura_fr/Documents/Partage/D21-

Innovation%20et%20nouvelles%20technologies/2024_DATASciences_Challenge?csf=1&web=1&e=iAfzwe)

- Corporate video: https://youtu.be/Ec5Q1kxnIhE
- Access to data via API:
  o pollutants, technical parameters, meteorology, geographical location
  o API Atmo AuRA: https://api.atmo-aura.fr/documentation
  o Internship report by J Dutroncy (Campus numérique in the alps)
  o Hugo Cordier's internship report (INSA Lyon)
- Data validation guide