MEESE — Atmospheric and Air Pollution
Handout 2

May 14, 2020

1 Introduction

Instead of solving a highly constrained 'problem', this second handout is an **open exercise** designed to stimulate your curiosity and invite you to investigate **real** datasets relevant for Air Pollution applications. These datasets include (i) pollutant concentration measurements, (ii) weather measurements and (iii) vessel traffic from Tarragona metropolitan area. These datasets should be thought as a *sandbox* for thinking about urban air pollution. The goal is not to obtain a consistent analysis and state-of-the-art final product but to demonstrate your capacity to (i) establish hypothesis, (ii) process data to validate them, (iii) present your results and (iv) identify potential pitfalls in your procedure.

2 Objective

In order to proceed, you need first to identify a question you want to address. A few examples of such questions could be:

1. Road traffic during the COVID-19 lockdown decreased significantly in Tarragona. Did this impact somehow the local pollution levels in the city?

2. How strong is the relation between local atmospheric stability as established by Pasquill classification and air quality in Tarragona?

3. Is it possible to establish a relation between vessel traffic intensity at the Port of Tarragona and air quality in the city?

4. Does it exist a significantly relation between local pollution levels, weather and Port activity in Tarragona? In order words, provided weather and Port traffic intensity, can I predict pollution levels in the city?

5. Does wind direction and velocity impact local pollution levels in Tarragona? What about radiation?

Of course, you may come up with your own questions. In case you need additional data to address your own problem, contact me and we'll find out, if available, how easy is to get them.

Once you know what question you plan to address, you must communicate it by email to **alexandre.fabregat@urv.cat**. You can be requested to change your choice to ensure a fair number of people addressing each problem.

3   Materials and methods

Take into account that:

1. You might not need all datasets in your problem.

2. Pollutant concentration measurements have been retrieved from a station located in Tarragona (*Parc de la Ciutat*) while weather measurements come from a station located in the city whereabouts (*Complex Educatiu*) and cloud cover data come from the station at the Airport of Reus (see Fig. 1 for reference). Despite this lack of *spatial collocation*, we will assume here that all data can be assimilated to the pollutant data station location. Although this is not necessarily consistent, potential solutions would highly increase the complexity of the analysis and are beyond the scope of this exercise.

3. Note that the variable that 'connects' each dataset is time. Use the variable **dt_h** (hour index) to *join* separate weather and pollutant measurement datasets. Regardless of the value itself, two observations with the same value of **dt_h** have been taken at the same time!

4. Note that while pollutant concentration and weather data have a nominal **hourly sampling rate** (i.e., a record per hour), cloud cover is available at a **daily sampling rate** (i.e., a record per day). For sake of simplicity, assume that the cloud cover value is constant over the entire 24-hour period of the corresponding day.

5. Note that there might be missing data in any dataset. These missing data is due to malfunction of the equipment, data corruption...
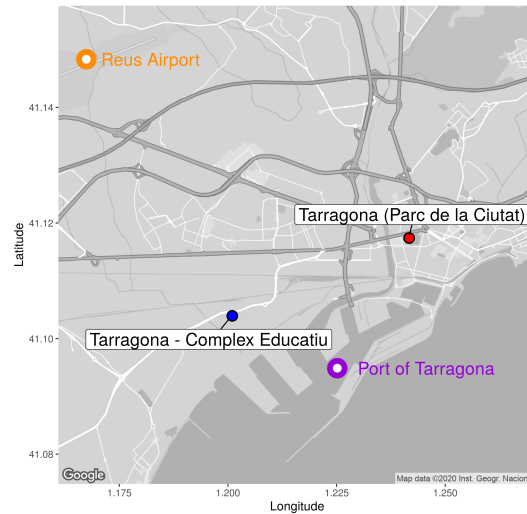

Figure 1: Map showing the location of (i) the pollutant concentration measurement station — Parc de la Ciutat (red dot), (ii) the weather measurement station — Complex Educatiu (blue dot), (iii) the cloud cover station (Reus Airport) and (iv) the Port of Tarragona.

The materials provided for this handout are comprised of four (4) datasets:

3.1 Pollution measurements

File name	pollutant_dataset.csv	Hourly pollutant concentration for several compounds
Number of rows	79268	
Variables	poll_codi	Pollution station code (unique value just for reference)
	contaminant	Pollutant name: H2S, NO, NO2, NOX, PM1, PM10, PM2.5, SO2, CO, O3
	data	measurement date (hourly)
	Concentracio	concentration in $\mu g\,m^{-3}$
	dt_h	Hour index: use it to combine -merge- datasets!

Table 1: Pollution database

3.2   Weather measurements

File name	weather_dataset.csv	Hourly data on several weather variables
Number of rows	11118	
Variables	meteo_codi	Weather station code (unique value just for reference)
	date	measurement date (hourly)
	wind_vel	Wind velocity in $\mathrm{m\,s^{-1}}$
	wind_dir	concentration in degrees (0 pointing North)
	temp	Temperature in Celsius
	hum_rel	Relative humidity in %
	press	Pressure in mbar
	precipita	Precipitation in mm
	insol	Insolation in $\mathrm{W\,m^{-2}}$
	dt_h	Hour index: use it to combine -merge- datasets!

Table 2: Weather database

3.3 Port of Tarragona vessel traffic

File name	vessel_database.csv	Hourly number of docked vessels
Number of rows	11608	
Variables	num_vess	Number of total docked vessels
	dt_h	Hour index: use it to combine -merge- datasets!

Table 3: Port of Tarragona vessel database

3.4 Cloud cover

File name	pollutant_dataset.csv	Watch out! Daily sampling instead of hourly!
Number of rows	456	
Variables	cloud_cover	Ranging from 0 to 8
	date	measurement date (daily!)

Table 4: Cloud cover database

Additionally, location (latitude and longitude) of the pollutant station (*pollutant_station.csv*), the weather station (*weather_station.csv*), the Airport of Reus and the Port of Tarragona are provided (*infrastructures_station.csv*).

Data can be retrieved by clicking on "Clone or Download" and then "Download ZIP" from here:

https://github.com/alexfabregat/MEESE_2020_Handout_2.

4   Deliverable

The deliverable must be handed out by **June 5th** and must consist of a 6 pages max report with, at least, the following section:

- Introduction: where you introduce the problem/question you addressed.

- Materials and methods: where you explain what data you used and how you processed it.

- Analysis and results: where you present your findings.

- Discussion: where you discuss any issue/problem/observation/... you found during your project.

Please, write a clear, concise and neat report.