

README

Air Pollution Impacts of Maritime Traffic: A New Method based on Observational Data and Randomization-Based Inference

Léo Zabrocki, Marion Leroutier and Marie-Abèle Bind

Contact: leo.zabrocki@psemail.eu and marion.leroutier@psemail.eu

In this document, we explain how our replication material is organized. We first display the general organization of the material and then detail how each sub-folder is structured. Should you have any questions or find coding errors, do not hesitate to contact us!

1 General Organization

The replication material is organized as follows:

- [1.hourly_analysis](#) contains the files to reproduce the analysis at the hourly level.
- [2.daily_analysis](#) contains the files to reproduce the analysis at the daily level.
- [3.toy_example](#) contains the files to reproduce the toy example for understanding randomization inference.

All folders contain the raw data, the scripts to clean and merge theme, and the scripts used in the analysis. R codes are contained in R Markdown files: when they are compiled, they produce nicely formatted .html files which display the code, its explanation and its output. We took great care to detail the code as much as possible.

The full replication of our study presents however two important caveats. First, we were not allowed to share the weather data from Météo-France. We therefore

added as small amount of noise to the original data. Researchers who reproduce our analysis will get different results. They can nonetheless easily check our coding procedure. Second, the matching procedure at the hourly level is computationally demanding. We had to rent an Amazon Web Services virtual computer (EC2 t3.2xlarge) to run the matching algorithm.

2 Hourly Analysis

- [1.data](#) contains all the raw and clean data for the analysis.
 - [1.raw_data](#) contains all the raw data to build the main data used in the analysis. The folder [3.weather_data](#) contains the real weather data for which a small amount of noise was added.
 - [2.data_for_analysis](#) contains the main data, its codebook and the outputs of the matching procedures. All scripts of [2.scripts](#) should be run to create these data.
- [2.scripts](#) contains all the scripts for the analysis.
 - [1.data_wrangling](#) contains the script to clean and merge all raw datasets. If you run this code, please make sure to change the name of the raw weather dataset.
 - [2.eda](#) contains the script to carry out an exploratory data analysis.
 - [3.matching_analysis](#) contains all the scripts to run our matching and randomization-based inference procedures.
 - [4.custom_ggplot2_theme](#) contains our custom ggplot2 theme.
- [3.outputs](#) contains all the figures produced by the codes.

3 Daily Analysis

The folder of the daily analysis is organized in the same manner as the hourly analysis folder.

4 Toy Example

This folder contains the data, the codes and the outputs to reproduce the toy example found in the supplementary materials.