A Joint Bayesian Space-Time Model to Integrate Spatially Misaligned Air Pollution Data in R-INLA

C. Forlani, M. Cameletti, E. Krainski, S. Bhatt and M. Blangiardo*

This repository contains the code to reproduce the results presented in the paper *A Joint Bayesian Space-Time Model to Integrate Spatially Misaligned Air Pollution Data in R-INLA* (Forlani C., Bhatt S., Cameletti M., Krainski E., Blangiardo M.), under review.

We are trying to improve the predictions of NO_2 concentration integrating data from two deterministic models called AQUM and PCM. The observations from the monitoring stations represent our response variable.

We want to model these data jointly in order to take into account the uncertainty associated with the deterministic model outputs.

The paper can be found at .

Data

The data workspace can be requested directly to the main author.

The workspace contains the following objects:

- `aqum`: daily output of the AQUM model for 495 locations and 1823 days
- `pcm`: annual output of the PCM model for 44117 locations and 5 years
- `final_dataset`: daily observations of NO_2 concentration for 126 monitors and 1826 days with associated site type
- `boundary`: boundary of the study area
- `london.shape`: boundary of Greater London
- `england.shape`: boundary of England
- `shape`: combines boundary of study area and geographical boundary of England
- `coordinates.aqum`: coordinates of the 495 AQUM locations
- `coordinates.pcm`: coordinates of the 44117 PCM locations
- `monitors`: coordinates of the 126 monitoring stations
- `monitors_val`: list of the 10 sets of monitors used for validation
- `pred.grid`: locations where we want to make predictions and corresponding site type
- `pred.grid.aqum`: values of AQUM interpolated at the prediction grid nodes, for each day
- `pred.grid.pcm`: values of PCM interpolated at the prediction grid nodes, for each year
- `roads.major`: shapefile of motorways over the study area

Note that all coordinates are in UTM projection so that distances are computed in meters. Results do not change when rescaling UTM coordinates to Kilometers.

Please see the paper for data description.

How to run the code

To run the code you just need to run the script `00_main.R` which contains instructions for data praparation and sources all the other scripts.

These can also be run manually setting `data_id` (1 to 6) and `formula_id` (1 to 9) once the data are loaded.

Notes on the notation

Please see the paper for details on the model specification for all the models under comparison.

The following are differences in the notation from the paper to the code:

- $y_1(s,t)$ is referred to as `pcm`
- $y_2(s,t)$ is referred to as `aqum`
- $y_3(s,t)$ is referred to as `y`
- $z_1(s)$ is referred to as `psi.field`, the copy for y as `psi.field.copy`, the copy for AQUM as `psi.field.aqum.copy`
- $z_2(t)$ is referred to as `csi.field`, and the copy for y as `csi.field.copy`
- The prior for $\lambda_{1,2}$ is called `lambda.aqum.space.prior`
- The prior for $\lambda_{1,3}$ is called `lambda.pcm.prior`
- The prior for $\lambda_{2,3}$ is called `lambda.aqum.time.prior`