

## Milestone M2 German Dataset

The Agrimonia dataset includes satellite data, model output and in situ measurements with different spatial and temporal resolutions from national and international agencies. For the air quality dataset in Lower Saxony, we followed the same preprocessing strategy as for the Agrimonia-Lombardy dataset (see Fassò et al. 2022). Moreover, the dataset is structured in the same ways, such that both datasets can easily be combined and the results can be compared. To be precise, the dataset includes five dimensions: air quality (AQ), weather and climate (WE), pollutants' emissions (EM), livestock (LI) and land and soil characteristics (LA). Because geostatistical methods can use neighbouring territory information for improving the overall predictive capability close to the borders, we additionally included the bordering German regions by applying a  $0.3^\circ$  buffer over the regional borders as shown in Figure 1.

Map of Lower Saxony by NUTS-2 regions

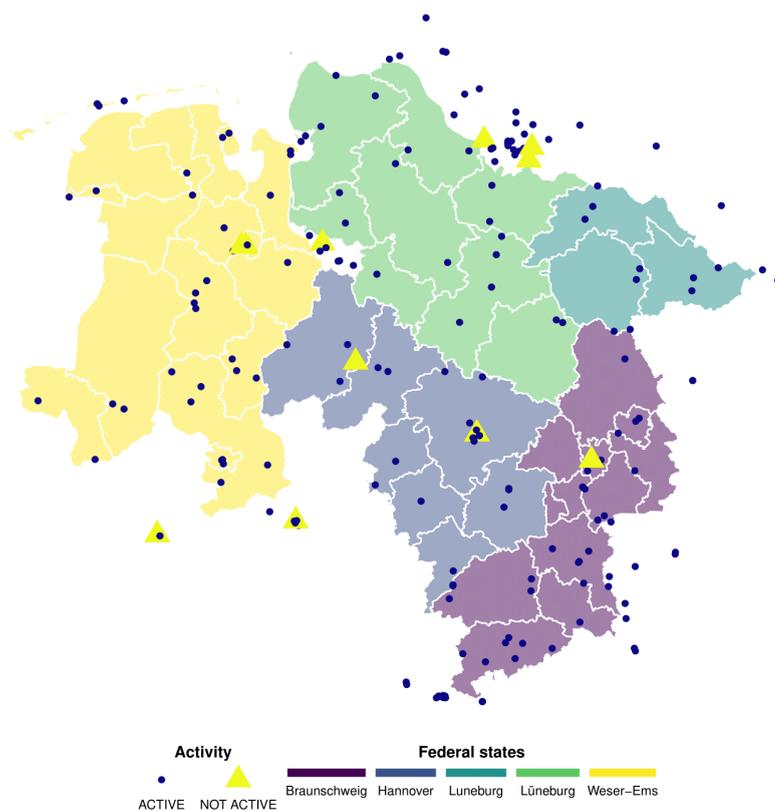


Figure 1: Sampling stations in Lower Saxony

The AQ data are pollutant concentrations [ $\mu\text{g}/\text{m}^3$ ] sampled at  $S = 141$  ground-level monitoring stations, irregularly located across Lower Saxony. The coordinates of the sampling stations are shown in Figure 1. For Lower Saxony, AQ data was obtained from the federal environmental agency in Germany (Umweltbundesamt) as well as from European Environmental Agency (EEA). An overview of the weekly and daily patterns can be obtained from Figure 2.

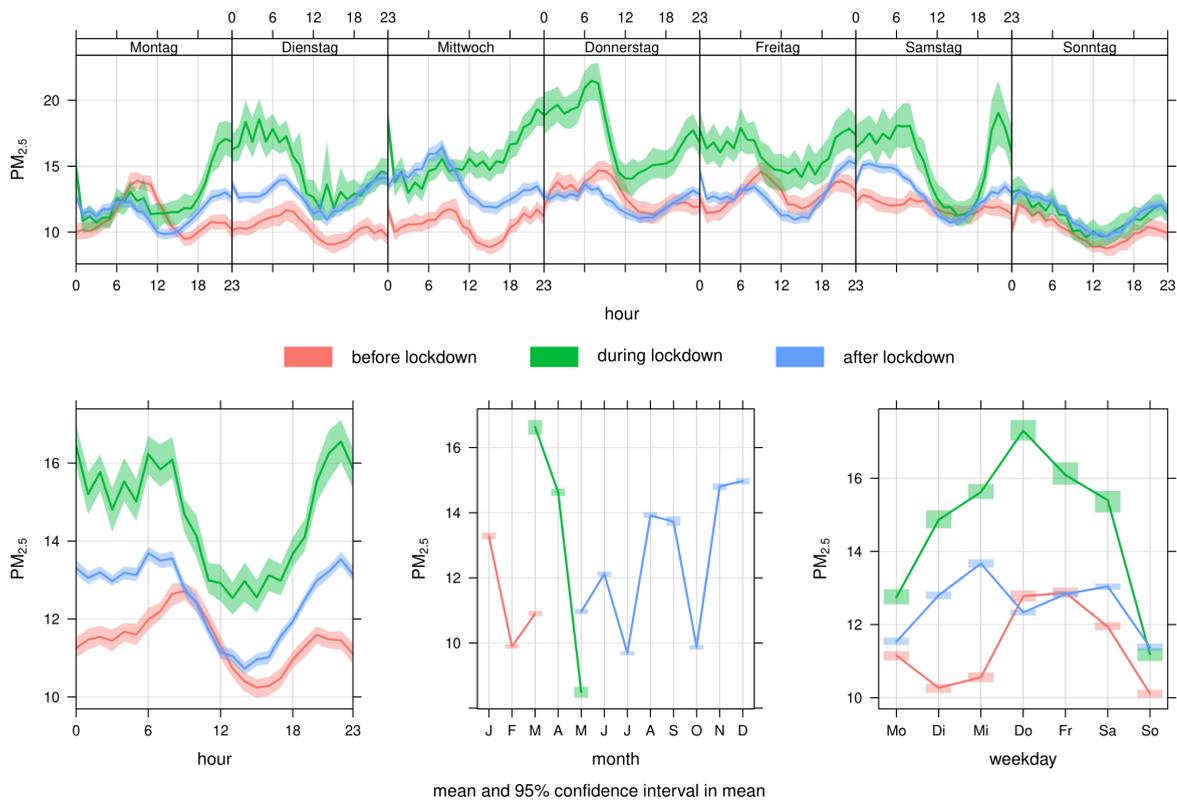


Figure 2: Weekly and daily patterns of PM<sub>2.5</sub> concentrations before, during, and after the lockdown period.

Meteorological data are obtained from the Copernicus Climate Change Service ([climate.copernicus.eu](https://climate.copernicus.eu)) through the ERA5 datasets containing the numerical model output computed by the European Centre for Medium-Range Weather Forecasts (ECMWF). ERA5 is the fifth generation ECMWF reanalysis of the global climate for the past decades. The reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of atmospheric science.

Information about livestock is obtained from the Regional Statistical office of Lower Saxony (Statistisches Landesamt Niedersachsen). For Lower Saxony, the density of bovines is available from 2017-2021. Eventually, land use information is obtained from the Corine Land Cover (CLC) datasets based on the classification of satellite images produced by the national teams of the participating countries (44 classes in the hierarchical 3-level CLC nomenclature).

The complete dataset is summarised in the attached metadata file `Meta_Data_v0_1 (GE).xlsx`.

### References

Fassò, A., Rodeschini, J., Fusta Moro, A., Shaboviq, Q., Maranzano, P., Cameletti, M., & Otto, P. (2022). Agrimonia: a dataset on livestock, meteorology and air quality in the Lombardy region, Italy. *arXiv preprint arXiv:2210.10604*.