

Spatiotemporal Air Pollution Modelling in the Barcelona Life Study Cohort (BiSC)

Alan Domínguez, MPH
alan.dominguez@isglobal.org

September 29th, 2023

ISGlobal Barcelona
Institute for
Global Health



Barcelona Life Study Cohort (BiSC)

- ▶ Research project aimed at understanding how Barcelona's **air pollution** affects the **development of the baby during pregnancy**.
- ▶ **Accurate estimate** of the exposure of pregnant mothers to environmental pollution (air pollution, noise, stress, importance of physical activity and green spaces).
- ▶ We aim to know better the **mediating function of the placenta**.
- ▶ BiSC is funded by **ERC (AirNB project)** and **HEI (Frontier project)**.



Traffic-Related Air Pollution



Traffic-Related Air Pollution

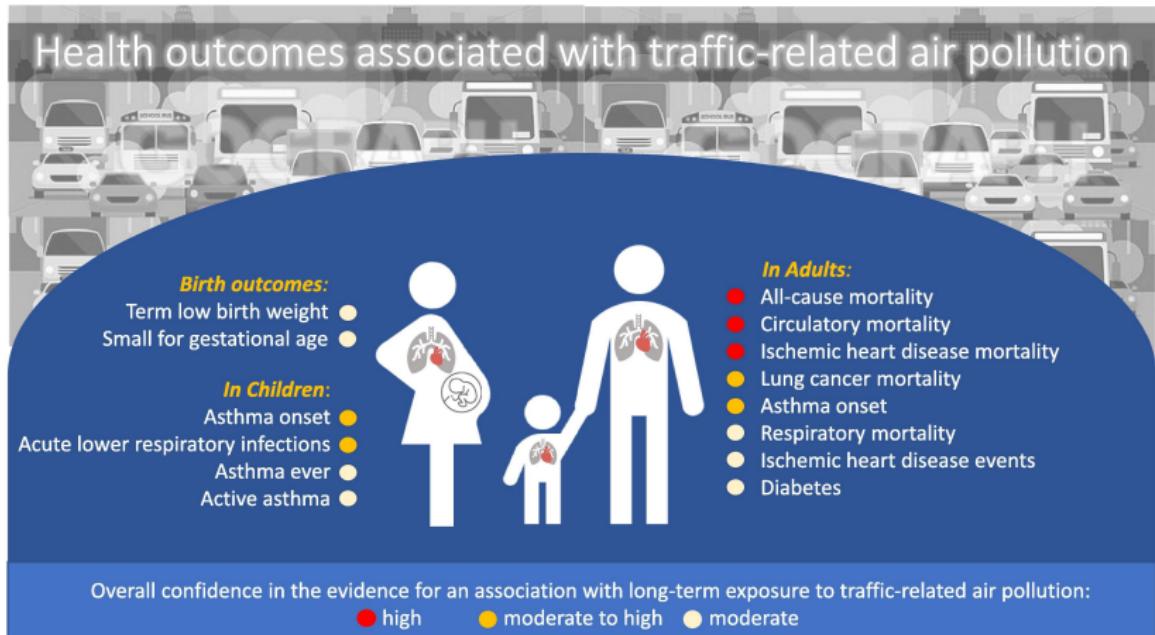


Figure: Health Effect Institute (HEI) review (2022)

Traffic-Related Air Pollution

- ▶ Air Pollution stands as the **primary environmental risk factor** for human health.
- ▶ Plays a significant role in the **global burden of disease**.
- ▶ Exposure to ambient air pollution has been linked with a **extensive array of health outcomes**.
- ▶ **Urbanizations** plays a pivotal role, **increasing exposure** to air pollution.

Air Pollution Exposure Assessment

- ▶ To enable **air pollution health effects studies**, study populations need to be **assigned exposures estimates**.
- ▶ Modelling air pollution is a challenging task due to their **complexity**.
- ▶ Difficulty to assess **air pollution variability** (both spatial and temporal).
- ▶ Different approaches have been proposed over the **last years** to **improve predictive performance**.
- ▶ Difficulty when we use inputs data from **different data sources**.
- ▶ **Machine learning** have been used to **model non-linearities** and **interaction effects** between predictor variables.

Objectives

- ▶ To **developed and compared** LUR, DM, and Hybrid spatiotemporal models to **estimate outdoor air pollution exposure** at residential addresses for the Barcelona Life Study Cohort (BiSC).
- ▶ To obtain **accurate estimates** for NO_2 , $PM_{2.5}$, BC , and $PM_{2.5}$ constituents (Fe, Zn, Cu, Sb) for use in **prenatal exposure assessment** for the BiSC cohort study.

Study domain: Barcelona Metropolitan Area

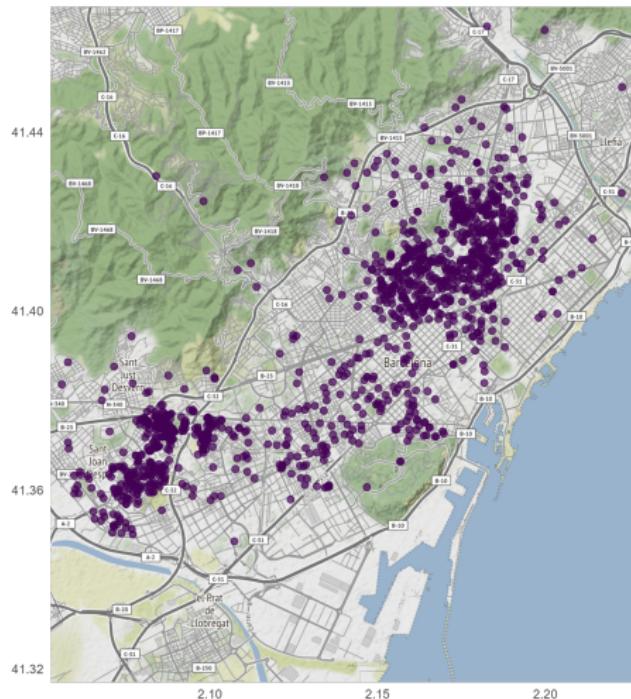


Figure: Spatial distribution of the BiSC participants across Barcelona (2019-2021)

Workplace and Home addresses for the BiSC participants

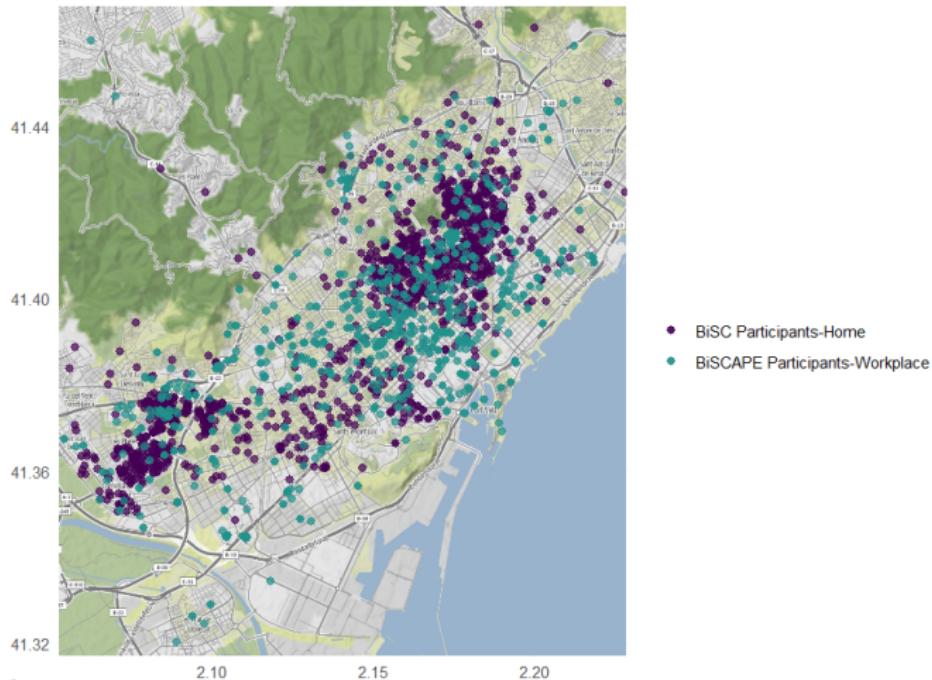
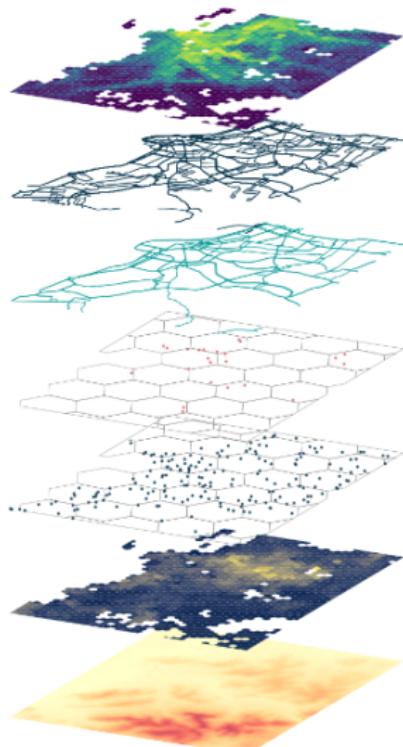


Figure: Spatial distribution of the Workplace and Home addresses for the BiSC participants across Barcelona (2019-2021)

Data Overview



- ▶ **Air pollution measurements**
- ▶ **Traffic variables** (road length, major road length, distance to road, distance to major road, traffic intensity, traffic intensity only major roads, traffic intensity in nearest road, traffic in nearest major road).
- ▶ **Land-use characteristics** (residential, industry, natural, urban green, road density).
- ▶ **Urban configuration** (population, building density, imperiousness, slope, building heights, street bearings, low emission zones, street width, street canyon, sampler height).
- ▶ **Air pollution and meteorological monitoring stations**
- ▶ **Traffic count measurements**
- ▶ **Dispersion model estimates**

Air Pollution measurements (Sampling Campaigns)

- ▶ Air pollution was measured in two different campaigns.
- ▶ Sampling campaign was measured in BiSC Home participants at week 12 and 32 of pregnancy and at 34 ESCAPE's sites.

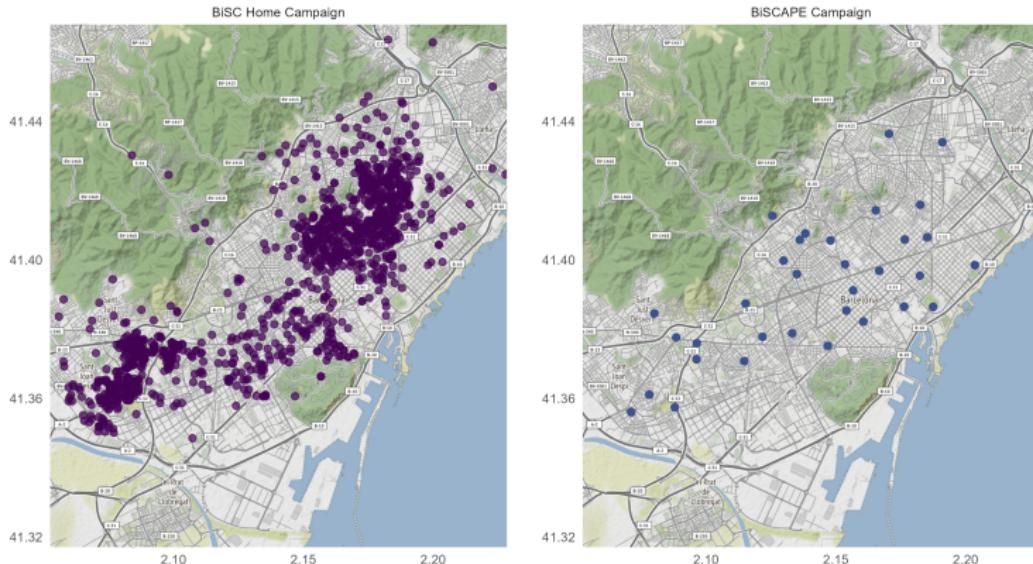


Figure: Air Pollution measurements campaign across Barcelona (2019-2021)

Air Quality and meteorological stations

- ▶ We use data from monitoring stations from Barcelona (2018 - 2022).
- ▶ Different monitoring stations for NO_2 , $PM_{2.5}$, and NO_x for BC model.

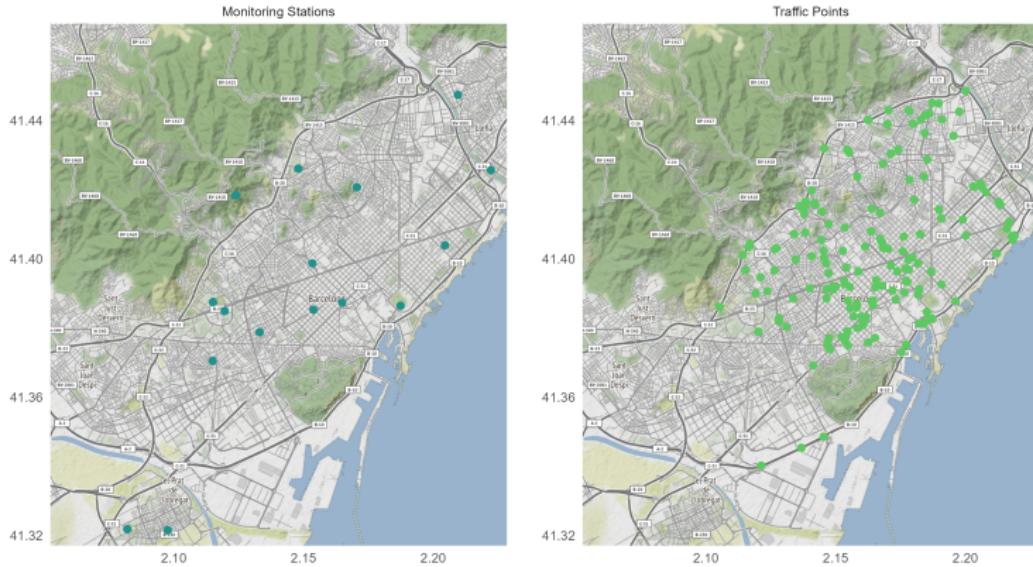


Figure: Monitoring stations and traffic point measurements, Barcelona (2018 - 2022)

Traffic, land use, urban configuration variables

- ▶ Spatial variables that better represent the pollutant concentration distribution across the study area were computed at different buffer sizes (25m, 50m, 100m, 300m, 500m, and 1000m).

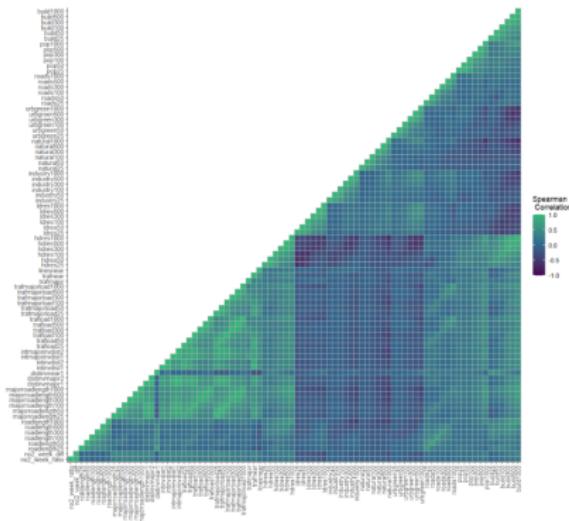
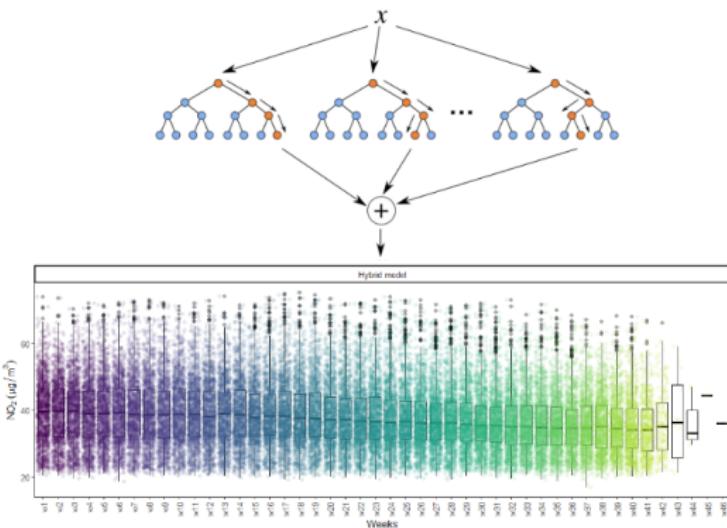


Figure: Correlation matrix for traffic, land use, and urban configuration variables

Random Forest to model Outdoor exposure to Air Pollution

- ▶ Supervised learning algorithm that uses ensemble learning (bagging).
- ▶ Each tree is built using a bootstraps sample of the data.
- ▶ Each node of the tree is split according to a subset of randomly chosen predictors.
- ▶ Final predictions is obtained averaging the outputs of each tree.



Workflow diagram for Predicting Outdoor Air Pollution

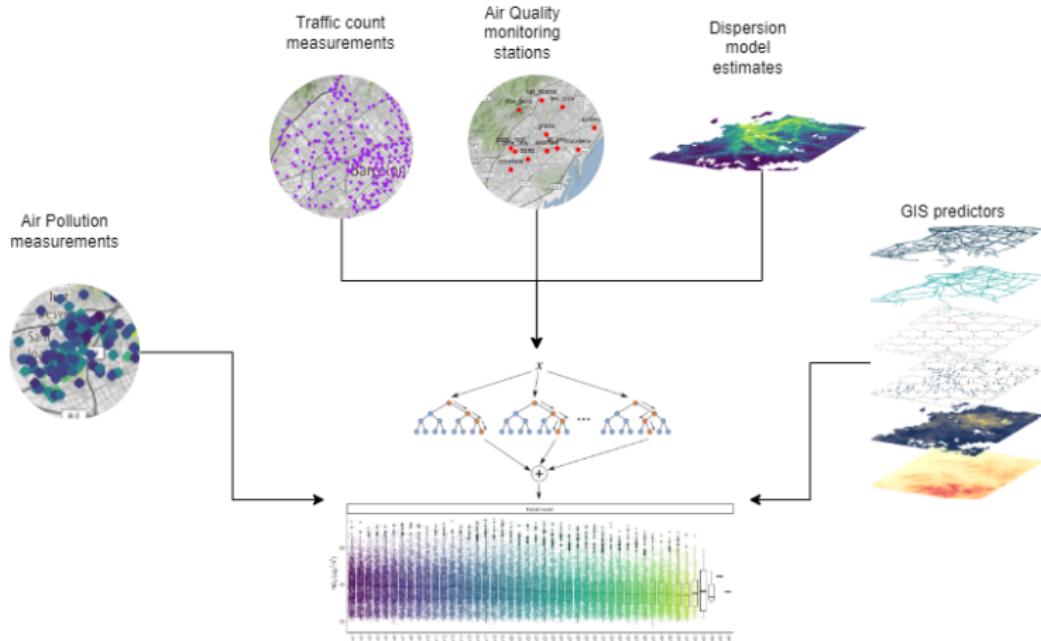


Figure: Workflow for building the Hybrid models

Model performance

- ▶ **How much of the spatial and temporal variation in outdoor air pollution is explained by the model?**
- ▶ We compare model predictions with measured levels of air pollution
- ▶ Do we care how well the model explains variations throughout the study area: We use $10 - CV$ to assess spatiotemporal performance of the model.
- ▶ The entire training data set was randomly split into 10 subsets. In each round of CV, we use nine subsets for model training and to make predictions for the held-out test subset

Results of Model Validation: NO_2 , $PM_{2.5}$, BC

- ▶ NO_2 : **0.64** $R^2(10 - cv)$
- ▶ $PM_{2.5}$: **0.66** $R^2(10 - cv)$
- ▶ BC : **0.84** $R^2(10 - cv)$

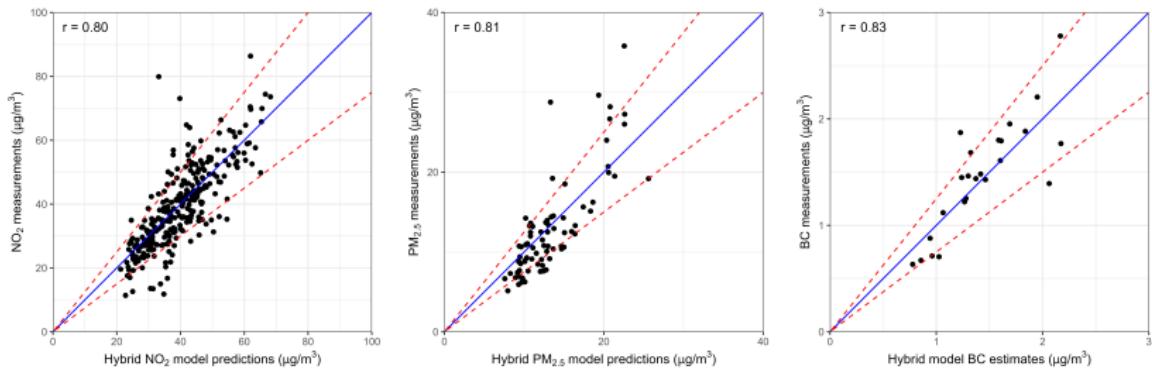


Figure: Correlation between measurements and predictive values for NO_2 , $PM_{2.5}$, BC

Results of Model Validation: *Fe, Cu, Zn*

- ▶ *Fe*: **0.54** $R^2(10 - cv)$
- ▶ *Cu*: **0.70** $R^2(10 - cv)$
- ▶ *Zn*: **0.42** $R^2(10 - cv)$

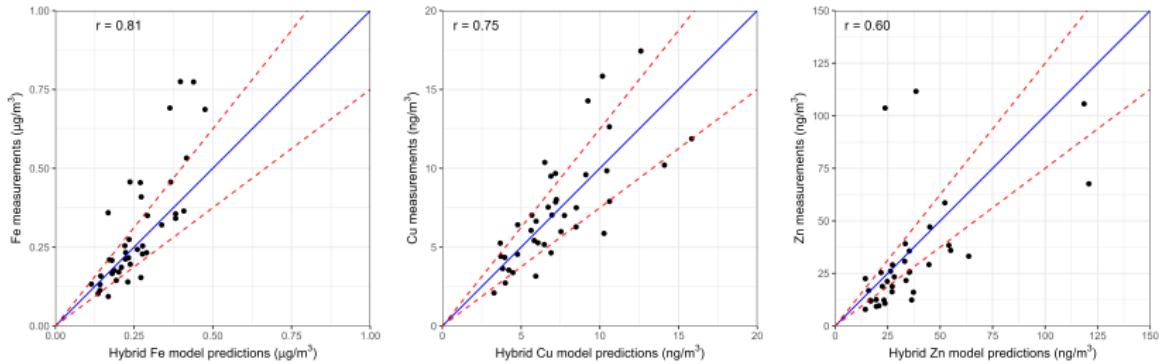


Figure: Correlation between measurements and predictive values for *Fe*, *Cu*, *Zn*.

Random Forest variable importance

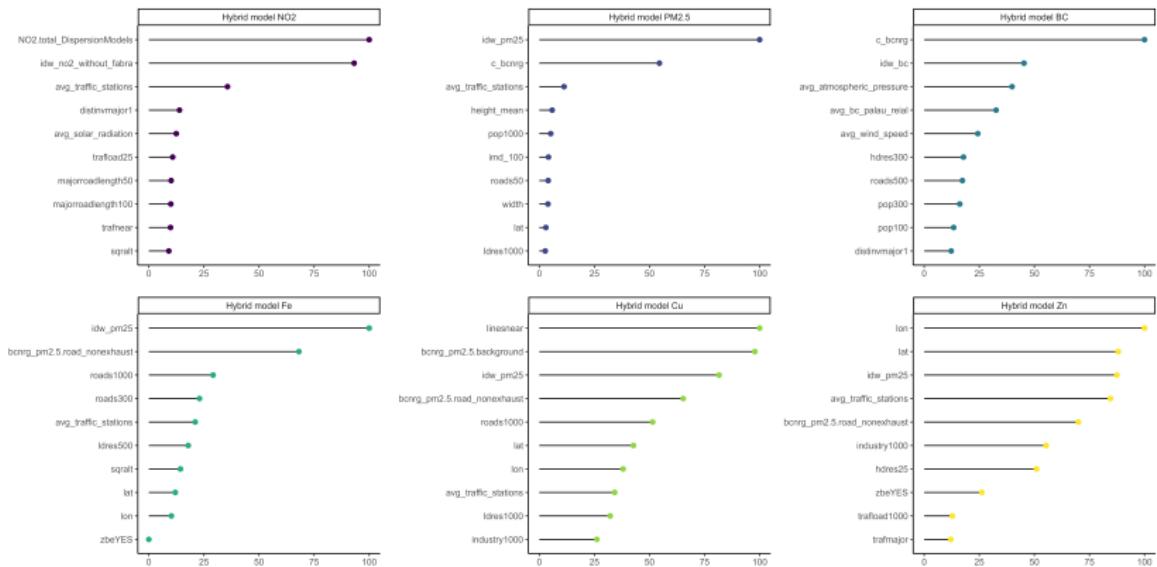


Figure: Hybrid model variable importance for NO_2 , $PM_{2.5}$, BC , Fe , Cu , Zn .

Birthweight and Traffic-related Air Pollution

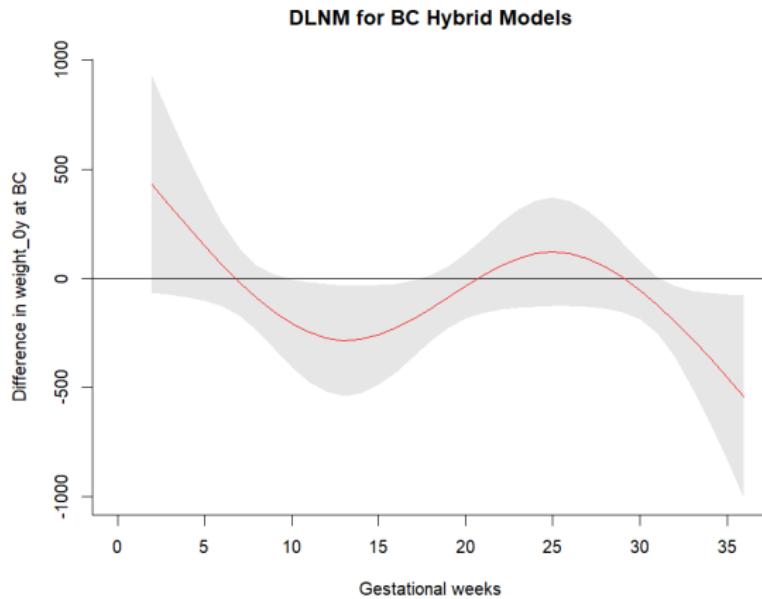


Figure: Distributed lag non-linear model for *BC* exposure and Birthweight

Conclusions

- ▶ **Good performance metrics** for the hybrid models for NO_2 , $PM_{2.5}$, BC , and Cu
- ▶ **Predictions show good temporal and spatial variability** across the different pollutants.
- ▶ Despite the **wide coverage** of NO_2 measurements, the hybrid model **does not show a big improvement** on the performance metrics.
- ▶ **Preliminary results** in associations analysis, between health outcomes and traffic-related air pollution exposure show **stronger associations** and **narrow confidence intervals** for **hybrid models** compared to LUR and DM.
- ▶ **Target-oriented CV** could be useful to **check the performance** of the models in **different settings** and **scenarios**.

Acknowledgements

