**COVID-19-INDIA**

Data and code to support Cleary *et al.* Comparing lagged impacts of mobility changes and environmental factors on COVID-19 waves in rural and urban India: a Bayesian spatiotemporal modelling study

This study describes a Bayesian spatiotempotal framework incorporating distributed lag non-linear models (DLNMs) using the INLA package in R (ref) to explore the lagged impact of climate, human mobility and stringency of government interventions on COVID-19 in India during two pandemic waves. The approach is similar to the modelling framework described by [Ge *et al.,* 2022](https://www.nature.com/articles/s41467-022-30897-1) who built a Bayesian framework to explore the daily reduction in the Rt of COVID-19 associated with different variants, non-pharmaceutical interventions and vaccinations by district and over time. Our modelling approach describes how incorporating DLNMs into a Bayesian framework can be used to explore the lagged impacts of drivers on weekly COVID-19 transmission, accounting for spatial and temporal correlations and random effects, similar to the methodology described by [Lowe](https://www.thelancet.com/pdfs/journals/lanplh/PIIS2542-5196(20)30292-8.pdf) *[et al.](https://www.thelancet.com/pdfs/journals/lanplh/PIIS2542-5196(20)30292-8.pdf)* [2021](https://www.thelancet.com/pdfs/journals/lanplh/PIIS2542-5196(20)30292-8.pdf) in their Bayesian spatiotemporal study on dengue risk in Brazil.

A description of each file and folder is provided below.

**Folders**

**mobility\_network\_analysis\_git:** Contains data and R scripts for analysing Google COVID-19 Mobility Research Dataset. This dataset was analysed by researchers at the University of Southampton, UK as per the terms of the data sharing agreement. We also defined mobility reductions and communities of population movements between administrative level II units, i.e. districts, across the country for five periods during the pandemic.

* **1\_India\_mobility\_v4.R**: for aggregating Google mobility dataset from S2 cell level to district level by origin and destination. For comparability of changes in mobility across districts, aggregated flows were further standardised using pre-pandemic mean baseline levels of mobility for the first eight weeks of 2020 (December 29, 2019 – February 22, 2020).
* **2\_Network\_modularity\_v2.R**: to detect the connectivity and community structure of population movement networks in the country. In the context of travel networks, a community refers to a group of areas that are more closely connected internally than with other areas in the network.

**wave1:** Contains data and R scripts to run models for exploring the lagged impacts of drivers on COVID-19 transmission during the initial pandemic wave in India (July 19th 2020 and November 29th 2020)

**delta\_wave:** Contains data and R scripts to run models for exploring the lagged impacts of drivers on COVID-19 transmission during the delta pandemic wave in India (March 7 to July 3, 2021)

Each folder also contains some or all of the following subfolders.

* **Rcode**: Folder containing R scripts for running models, visualising data and model outputs and carrying out spatial and temporal cross-validation.
* **data:** a folder containing weekly COVID-19 and covariate data, as well as one- and two-week lagged datasets (.csv), a grid file for facet\_geo() plots (.csv), and shapefiles for three administrative levels in India (.shp, .shx, .dbf, .prj).
* **figs:** a folder to save the figures generated by the R scripts.
* **output:** a folder to save the model outputs. This also contains a subfolder named **preds**, with the R output files containing results of spatial (preds\_gid1) and temporal (preds\_week) cross-validation.

**R code for modelling:**

**1\_load\_packages\_data.R:** R script to prepare data and lagged variables for INLA-DLNM modelling

### **2\_visualise\_model data.R:** R script to visualise and explore spatial variation in mobility, covid, climate, population demographics and stringency of government interventions

### **3\_run\_models.R:** R script to run INLA models of increasing complexity.

### **4\_run\_models\_leave\_one\_week\_out\_orig.R:** R script for running spatial and temporal cross-validation on INLA-DLMN models.

### **5\_model\_output.R:** R script to visualise INLA-DLMN model output.

### **6\_DLNM\_output.R:** R script to visually explore exposure-lag-response relationships from DLNM INLA model output.

### **7\_load\_packages\_data\_lag.R:** R script to prepare data and lagged variables for INLA-DLNM modelling using lagged data.

### **8\_run\_models\_lag.R:** R script to run INLA models of increasing complexity using lagged data

### **9\_run\_models\_leave\_one\_week\_out\_lag.R:** R script for running spatial and temporal cross-validation on INLA-DLMN models using lagged data.

### **10\_model\_output\_lag.R:** R script to visualise INLA-DLMN model output using lagged data.