AIMS African Institute for Mathematical Sciences NEXT EINSTEIN INITIATIVE

IMPERIAL

A Bayesian Multisource Fusion Model for Spatiotemporal $PM_{2.5}$ in an Urban Setting

From London to Dakar

Abi I. Riley

12/06/2025

Introduction

Project Aims:

- To develop a Bayesian multisource air pollution model for exposure assessment in Greater London, using monthly data (2014 to 2019) at fine spatial resolution.

- To study the long-term effects of outdoor air pollution on adolescent mental health.

MRC
Centre for Environment & Health

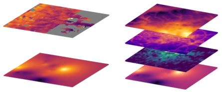
UK RI Medical Research Council

Imperial College London

1

Processing

Data Preparation and Gap Filling



2

Modelling

Bayesian Spatiotemporal Model

For monitoring sites $\mathbf{s}$ and month t:

$$Y(\mathbf{s}, t) = \eta(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$
$$\eta(\mathbf{s}, t) = \beta X(\mathbf{s}, t) + \omega(\mathbf{s}, t)$$

Comparing different modelling assumptions and specifications.

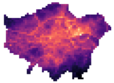
3

Prediction

Prediction Across Entire Domain

For unmonitored locations $\mathbf{s^*}$ at time t:

$$f(\mathbf{Y}(\mathbf{s}^*, t) \mid \mathbf{Y})$$

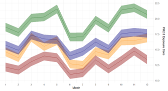


4

Exposure

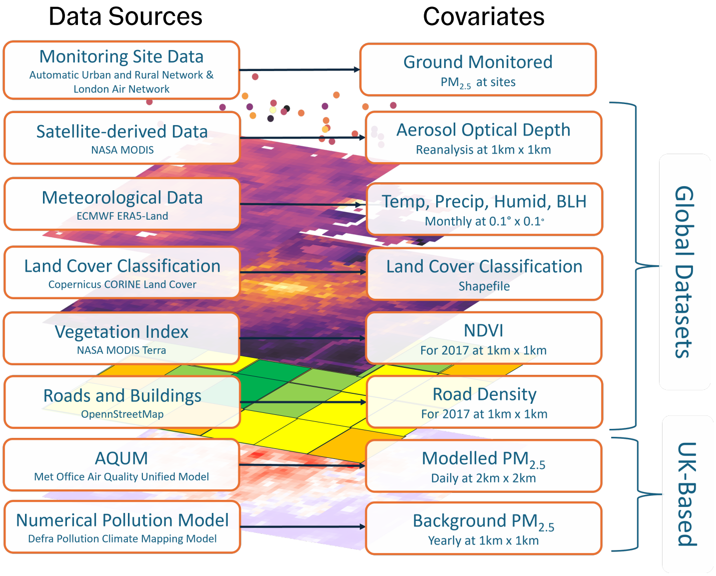
Estimating Exposure

Extract exposure for each participant



Abi Riley a.riley21@imperial.ac.uk

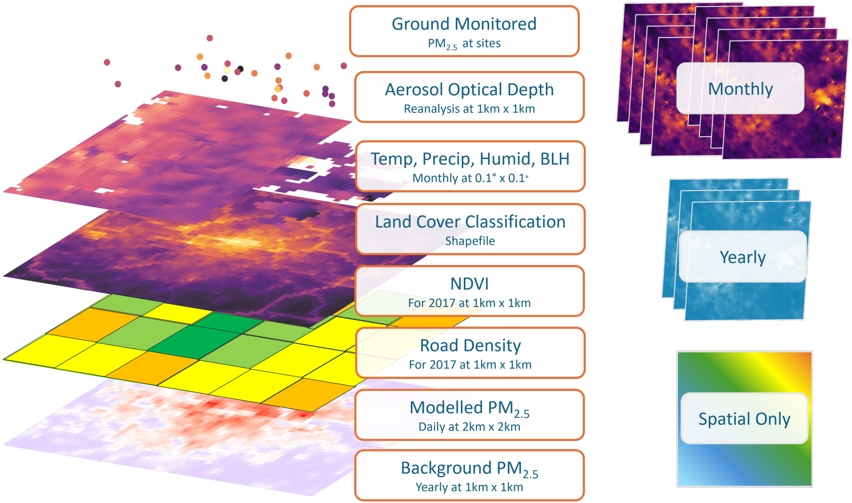
Data



Abi Riley a.riley21@imperial.ac.uk

Data Fusion

Spatial and Temporal Misalignment



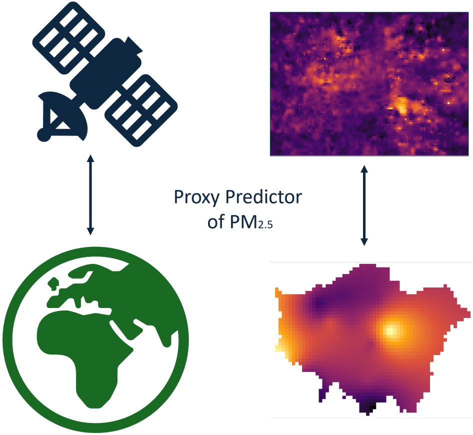
Abi Riley a.riley21@imperial.ac.uk

Satellite-derived Datasets

Processing and Interpolation

Aerosol Optical Depth

- Global NASA product MCD19A2
- Daily 1km x 1km grids
- Proxy measurement for particulate matter
  Christopher, S. et. al, 2020



Abi Riley a.riley21@imperial.ac.uk

Methods

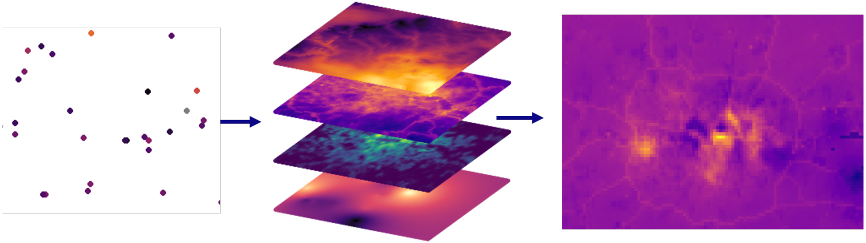
Bayesian Spatiotemporal Model

A Bayesian Hierarchical Model

For pollution concentrations $\mathrm{Y}(\mathbf{s}, \mathrm{t})$:

$$\mathrm{Y}(\mathbf{s}, \mathrm{t}) = \eta(\mathbf{s}, \mathrm{t}) + \epsilon(\mathbf{s}, \mathrm{t})$$
$$\eta(\mathbf{s}, \mathrm{t}) = \mu(\mathbf{s}, \mathrm{t}) + \omega(\mathbf{s}, \mathrm{t})$$

with random error $\epsilon(\mathbf{s}, \mathrm{t}) \sim \mathrm{N}(0, \sigma_\epsilon^2)$ and the spatiotemporal SPDE model $\omega(\mathbf{s}, \mathrm{t})$. Cameletti, M. et. al, 2019



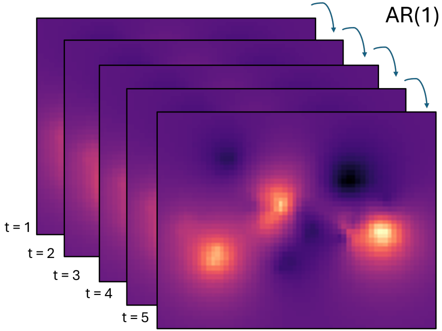
Abi Riley a.riley21@imperial.ac.uk

Methods
Spatiotemporal AR(1) Model

Spatiotemporal Process Term
A mean-zero stationary Gaussian process evolving in time according to an order 1 autoregressive process:

$$\omega(\mathbf{s}, t) = a\omega(\mathbf{s}, t-1) + u(\mathbf{s}, t)$$
$$\omega(\mathbf{s}, 1) \sim N(0, \sigma_\omega^2/(1-a^2))$$

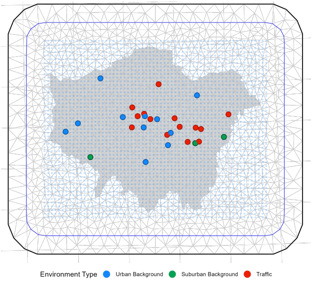
for $t = 1, \ldots, T$.



Abi Riley a.riley21@imperial.ac.uk

Methods

Spatiotemporal Model using INLA-SPDE



Gaussian Field and Covariance function

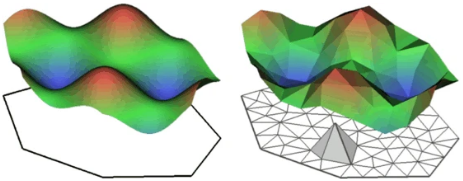
Let $u(\mathbf{s}, t)$ be a zero-mean Gaussian field with temporally independent Matérn covariance function

$$\mathrm{Cov}(u(\mathbf{s}_i), u(\mathbf{s}_j)) = \sigma_u^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa d)^{\nu} K_{\nu} (\kappa d)$$

where $d = ||\mathbf{s}_i - \mathbf{s}_j||$.

Stochastic Partial Differential Equations (SPDE) Approach

An approximation method using Gaussian Markov Random Fields and a discredited mesh of spatial points. Creating an n-dimensional precision matrix for the correlations between points, Σ.



Model 1: Baseline Spatiotemporal Model

- Based on Cameletti, M. et. al, 2019
- Linear additive covariates, β

$$\eta(\mathbf{s}, \mathrm{t}) = \beta_0 + \beta_1 \mathrm{BG}(\mathbf{s}) + \beta_2 \mathrm{PCM}(\mathbf{s}, \mathrm{t}) + \beta_3 \mathrm{AQUM}(\mathbf{s}, \mathrm{t}) + \omega(\mathbf{s}, \mathrm{t}) \tag{1}$$

Model 1: Baseline Spatiotemporal Model

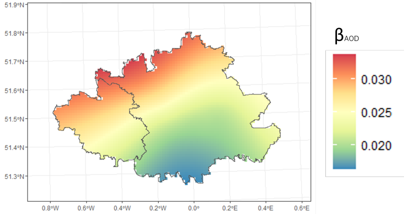
- Based on Cameletti, M. et. al, 2019
- Linear additive covariates, $\boldsymbol{\beta}$

$$\eta(\mathbf{s},\mathrm{t}) = \underbrace{\beta_0}_{\text{Intercept}} + \beta_1 \underbrace{\mathrm{BG}(\mathbf{s})}_{\text{Background Indicator}} + \beta_2 \underbrace{\mathrm{PCM}(\mathbf{s},\mathrm{t})}_{\text{Yearly Pollution Model}} + \beta_3 \underbrace{\mathrm{AQUM}(\mathbf{s},\mathrm{t})}_{\text{Coarse Pollution Model}} + \underbrace{\omega(\mathbf{s},\mathrm{t})}_{\text{SPDE Term}} \tag{2}$$

Model 2: Spatially-Varying Coefficients Model

Using a spatial SPDE model with a covariate-weighted projection matrix:

$$\beta_{\mathsf{k}}(\mathbf{s}) \sim \mathsf{N}(0, \sigma^2_\beta \Sigma)$$



- Allowing covariate effects to vary spatially
- Non-linear and non-stationary spatial process
- To capture complex physical and chemical dynamics

$$\eta(\mathbf{s},\mathsf{t}) = \beta_0 + \beta_1\mathsf{BG}(\mathbf{s}) + \beta_2\mathsf{PCM}(\mathbf{s},\mathsf{t}) + \beta_3(\mathbf{s})\mathsf{AQUM}(\mathbf{s},\mathsf{t}) + \omega(\mathbf{s},\mathsf{t}) \qquad (3)$$

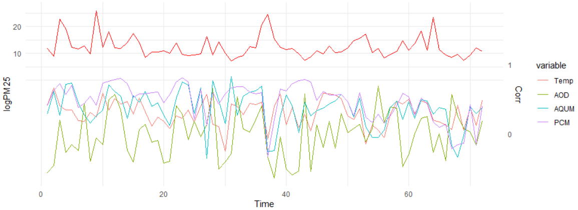
Abi Riley a.riley21@imperial.ac.uk

Model 3: Time-Varying Coefficients Model

Time-varying coefficient for NDVI, $\beta(t)$, based on Franco-Villoria, M., et. al, 2019

- Strong seasonality across variables
- AR1 model in time
- Complex time interactions
- Cyclic monthly relationships

$$\beta_4(1) \sim \mathrm{N}(0, \sigma_\beta^2(1-\phi_\beta^2)^{-1})$$
$$\beta_4(t) = \phi_\beta \; \beta_4(t-1) + \epsilon_t$$



$$\eta(\mathbf{s}, t) = \beta_0 + \beta_1 \mathrm{BG}(\mathbf{s}) + \beta_2 \mathrm{PCM}(\mathbf{s}, t) + \beta_3(\mathbf{s}) \mathrm{AQUM}(\mathbf{s}, t) + \beta_4(t) \mathrm{NDVI}(\mathbf{s}, t) + \omega(\mathbf{s}, t) \quad (4)$$

Abi Riley a.riley21@imperial.ac.uk

Model 4: Mixed Coefficients Model

Spatially-varying coefficient for AQUM, $\beta(\mathbf{s})$ and Time-varying coefficient for NDVI, $\beta(\mathrm{t})$

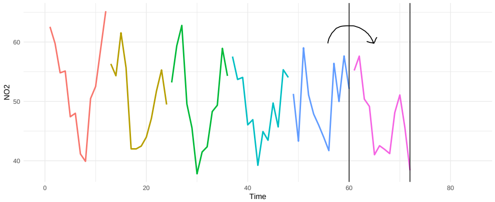
$$\eta(\mathbf{s},\mathrm{t}) = \beta_0 + \beta_1 \mathrm{BG}(\mathbf{s}) + \beta_2 \mathrm{PCM}(\mathbf{s},\mathrm{t}) + \beta_3(\mathbf{s}) \mathrm{AQUM}(\mathbf{s},\mathrm{t}) + \beta_4(\mathrm{t}) \mathrm{NDVI}(\mathbf{s},\mathrm{t}) + \omega(\mathbf{s},\mathrm{t}) \tag{5}$$

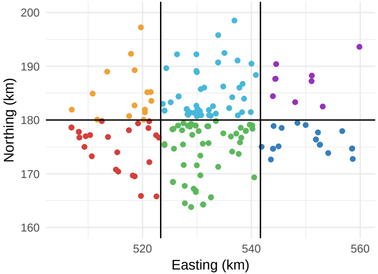
Model Selection

Model Fit

To compare model goodness-of-fit we use:

- Fitted vs Observed R^2
- Posterior Predictive Model Choice Criterion Gelfand, A. et. al, 1998





Predictive Performance

To compare predictions we use k-fold cross-validation methods:

- Block spatial folds
- Temporal forecasting

Comparing:

- Predicted vs Observed R^2
- Root Mean Squared Error (RMSE)
- Bias
- 95% Coverage (Cov)

Abi Riley a.riley21@imperial.ac.uk

Results
Comparing Model Statistics

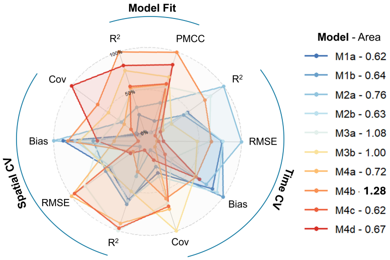
	Model Fit	
Model	R^2	PMCC
1	0.94	194
2	0.94	189
3	0.94	188
4	0.94	185

	Spatial Cross-Validation				Temporal Cross-Validation			
Model	R^2	RMSE	Bias	Cov	R^2	RMSE	Bias	Cov
1	0.69	0.22	-0.024	0.91	0.71	0.19	0.003	0.89
2	0.69	0.22	-0.023	0.91	0.72	0.19	0.004	0.89
3	0.70	0.21	-0.027	0.91	0.72	0.19	0.012	0.90
4	0.68	0.22	-0.024	0.91	0.72	0.19	-0.007	0.90

Results

Comparing Model Statistics

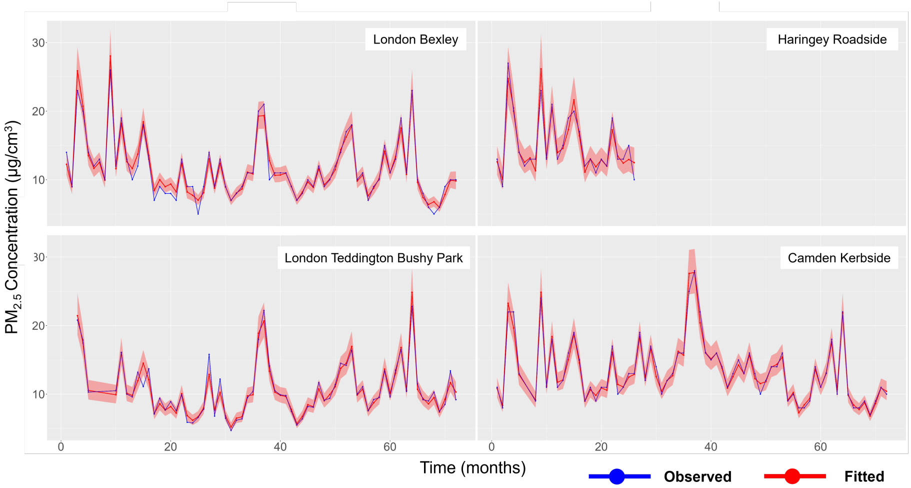
	Model Fit	
Model	R^2	PMCC
1	0.94	194
2	0.94	189
3	0.94	188
4	0.94	185



	Spatial Cross-Validation				Temporal Cross-Validation			
Model	R^2	RMSE	Bias	Cov	R^2	RMSE	Bias	Cov
1	0.69	0.22	-0.024	0.91	0.71	0.19	0.003	0.89
2	0.69	0.22	-0.023	0.91	0.72	0.19	0.004	0.89
3	0.70	0.21	-0.027	0.91	0.72	0.19	0.012	0.90
4	0.68	0.22	-0.024	0.91	0.72	0.19	-0.007	0.90

Abi Riley a.riley21@imperial.ac.uk

Results
Model Fit

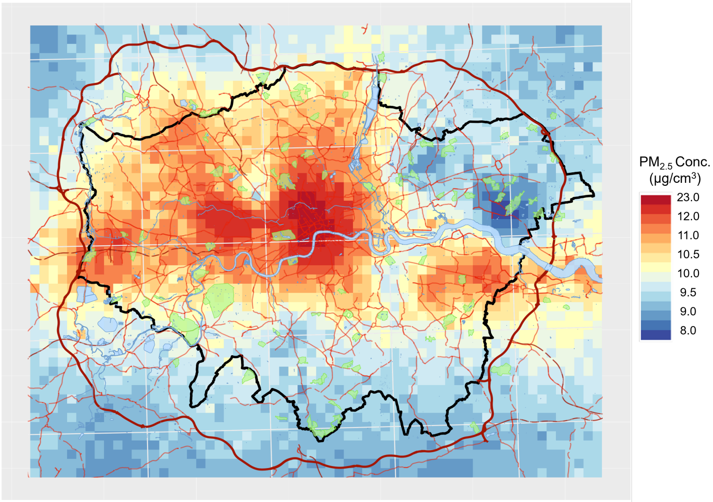


Abi Riley a.riley21@imperial.ac.uk

Results

Model Prediction

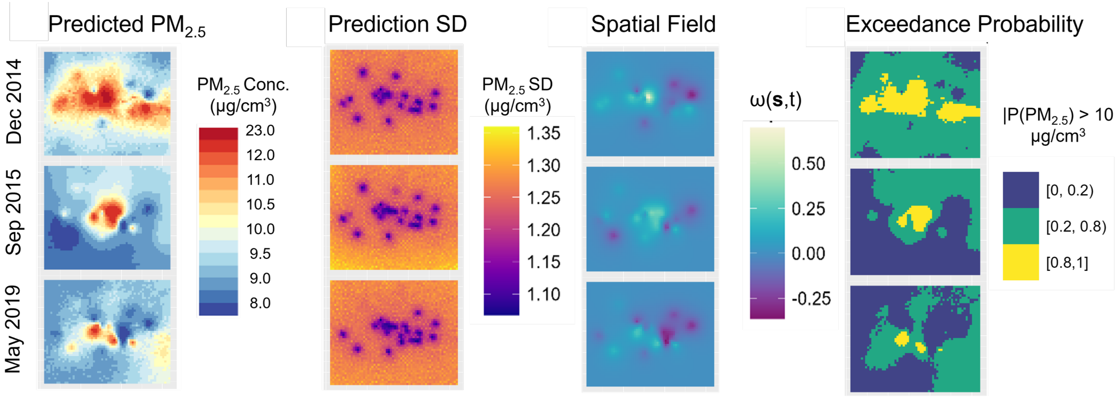
- Captures overall temporal trends and seasonality
- Shows spatial features across London:
  - Major Roads, e.g. M25
  - Heathrow and City airports
  - Large greenspaces



Abi Riley a.riley21@imperial.ac.uk

Results

Posterior Maps



Abi Riley a.riley21@imperial.ac.uk

What have we learnt?

Bayesian hierarchical approach
The Bayesian framework readily includes measurement error and produces posterior predictive distributions.

Multisource data
The models combine many different sources of data, especially calibrating the observed monitoring site data with the PCM annual background and the AQUM low-spatial resolution model.
However, AOD was not found to be useful over London.

Model Complexity
The ’best’ chosen model includes some of the more complex model options, but still balances model performance with computational load.

Outputs

Paper Under Review

The International EnvironMetrics Society - TIES

A Bayesian Multisource Fusion Model for Spatiotemporal $PM_{2.5}$ in an Urban Setting

Abi I. Riley[1] | Marta Blangiardo[1,2] | Frédéric B. Piel[1,2] | Andrew Beddows[3] | Sean Beevers[3] | Gary W. Fuller[3] | Paul Agnew[4] | Monica Pirani[1]

[1]MRC Centre for Environment & Health, Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom, W12 0BZ

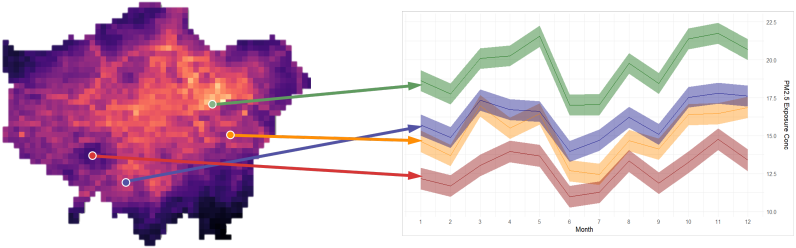
[2]UK Small Area Health Statistics Unit, Department of Epidemiology and Biostatistics, School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom, W12 0BZ

[3]Environmental Research Group, School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom, W12 0BZ

Abstract

Airborne particulate matter ($PM_{2.5}$) is a major public health concern in urban environments, where population density and emission sources exacerbate exposure risks. We present a novel Bayesian spatiotemporal fusion model to estimate monthly $PM_{2.5}$ concentrations over Greater London (2014-2019) at 1km resolution. The model integrates multiple $PM_{2.5}$ data sources, including outputs from two atmospheric air quality dispersion models and predictive variables, such as vegetation and satellite aerosol optical depth, while explicitly modelling a latent spatiotemporal field. Spatial misalignment of the data is addressed through an upscaling approach to predict across the entire area. Building on stochastic partial differential equations (SPDE) within the integrated nested Laplace approximations (INLA) framework, our method introduces spatially- and temporally-varying coefficients to flexibly calibrate datasets and capture fine-scale variability. Model

Application
Exposure Estimation



- $PM_{2.5}$ concentrations
- High-resolution monthly data 2014-2019
- With model uncertainty

This model will be used to estimate the monthly exposures of a cohort of children in London and study the potential adverse effects on their mental health and well-being.

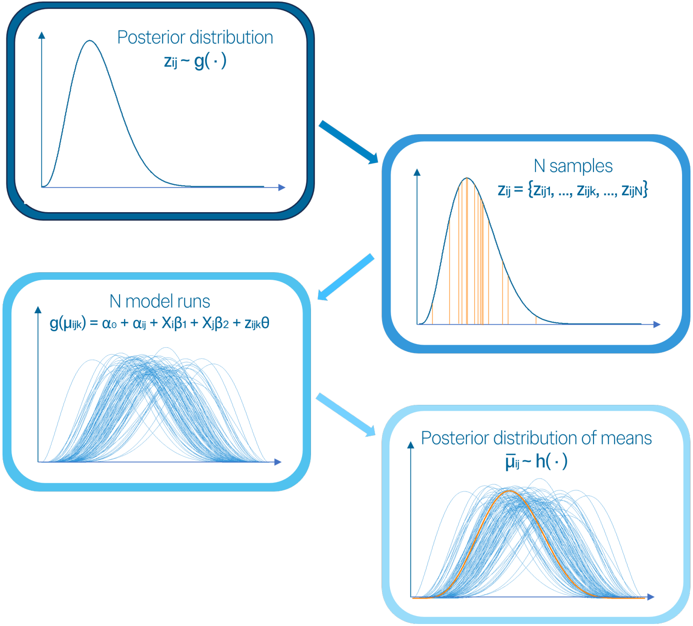
Application
Uncertainty Propagation

For participant i, Y_i :

$$Y_i|\mu_i, \boldsymbol{\xi} \sim f(y|\mu_i, \boldsymbol{\xi})$$
$$g(\mu_i) = \alpha_0 + \alpha_i + X_i\boldsymbol{\beta}_1 + X_j\boldsymbol{\beta}_2 + z_i\theta$$

for:

- α_i : individual random intercept
- X_i : covariates
- z_i : exposure estimate



Abi Riley a.riley21@imperial.ac.uk

Bibliography

Christopher, S., Gupta, P. (2020). Global Distribution of Column Satellite Aerosol Optical Depth to Surface PM2.5 Relationships. Remote Sensing, 12(12), 1985. doi.org/10.3390/rs12121985

Cameletti, M., Gómez-Rubio, V., & Blangiardo, M. (2019). Bayesian modelling for spatially misaligned health and air pollution data through the INLA-SPDE approach. Spatial Statistics, 31. doi.org/10.1016/J.SPASTA.2019.04.001

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., & Rue, H. (2018). Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. doi.org/10.1201/9780429031892

Franco-Villoria, M., Ventrucci, M., & Rue, H. (2019). A unified view on Bayesian varying coefficient models. Electronic Journal of Statistics, 13(2), 5334–5359. doi.org/10.1214/19-EJS1653

Gelfand, A., & Ghosh, S. Model choice: A minimum posterior predictive loss approach, Biometrika, Volume 85, Issue 1, March 1998, Pages 1–11, doi.org/10.1093/biomet/85.1.1

AIMS | African Institute for Mathematical Sciences NEXT EINSTEIN INITIATIVE

IMPERIAL

a.riley21@imperial.ac.uk

linkedin.com/in/abiriley

Thank you. Any questions?

Supervised by: Monica Pirani, Marta Blangiardo, Fred Piel and James Kirkbride.