

Multivariate Bayesian Modeling with INLA

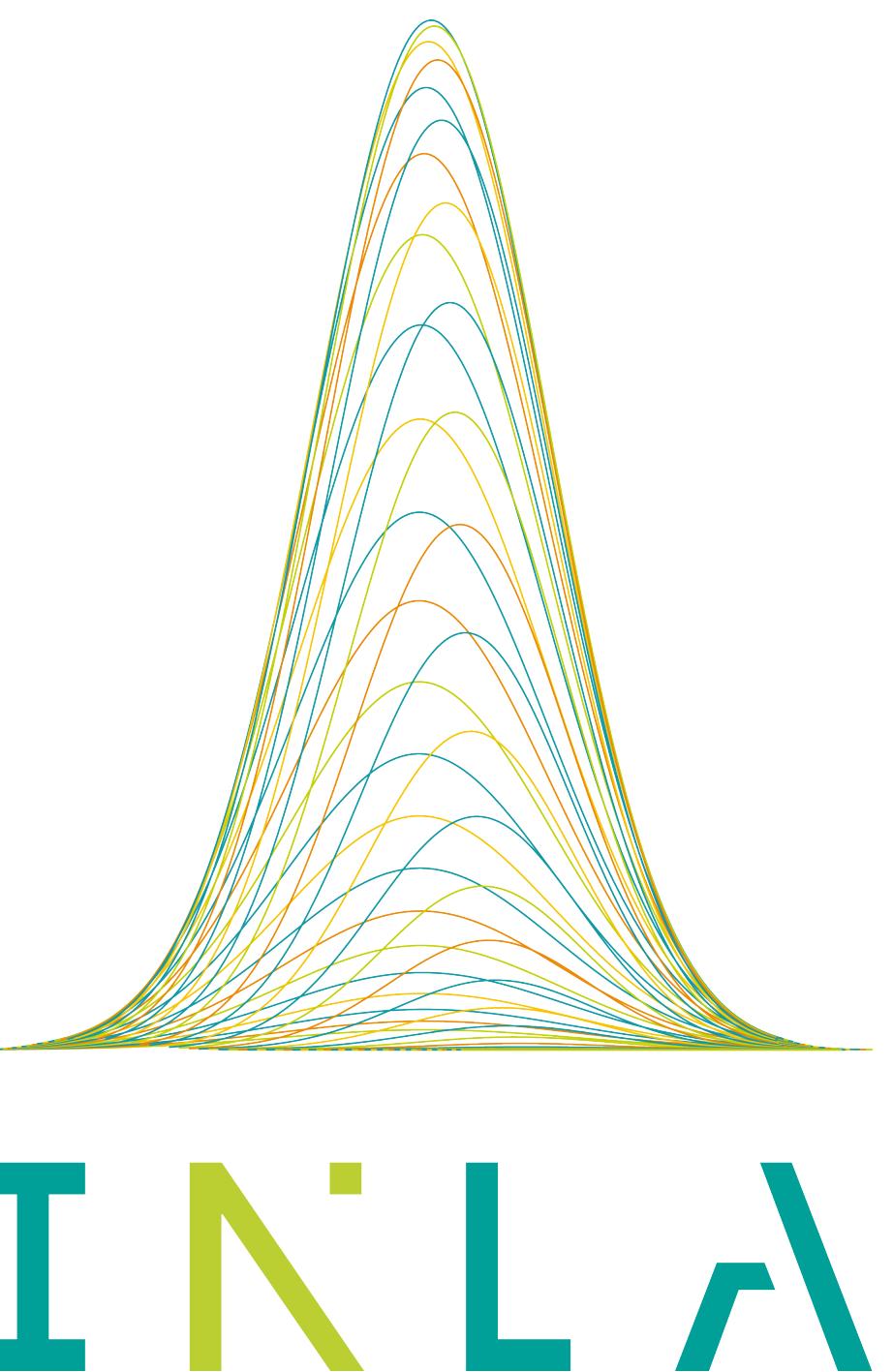
Lisa Gaedke-Merzhäuser

Bayesian Computational Statistics and Modeling Group

KAUST

Ethiopian Public Health Institute

October 17, 2025





Multivariate Modeling

Many applications many different modeling approaches

- Epidemiology

- Ecology

- Environmental Sciences

How to construct a good model?

- Start with individual model for each process

- Then combine them



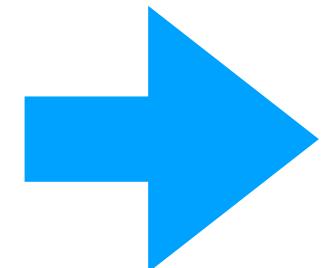
Multivariate Modeling

- Some observations arise jointly or we want to model them jointly

$$\text{Process 1: } \eta_1 = B_1 x_1$$

$$\text{Process 2: } \eta_2 = B_2 x_2$$

...



$$\eta = BX$$

- Each univariate process has their own unknown parameters
- But we also want to link the processes, to have

- Shared latent or covariate effects
- Correlation between processes



Linear Models of Coregionalization (LMC)

- Common approach for spatial and spatio-temporal processes
- Represents response variable as a linear combination of independent univariate latent GPs

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 & 0 & 0 \\ 0 & \mathbf{B}_2 & 0 \\ 0 & 0 & \mathbf{B}_3 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix}$$

- Combined through a coregionalization matrix, containing
 - scale parameters σ_i and coupling terms λ_i



Case Study on large-scale Air Pollution Model

- >We want jointly model Ozone, PM2.5 and PM10
- Consider region over northern Italy in 2022
- Gridded observation data available from Copernicus Atmosphere Monitoring Service (CAMS)



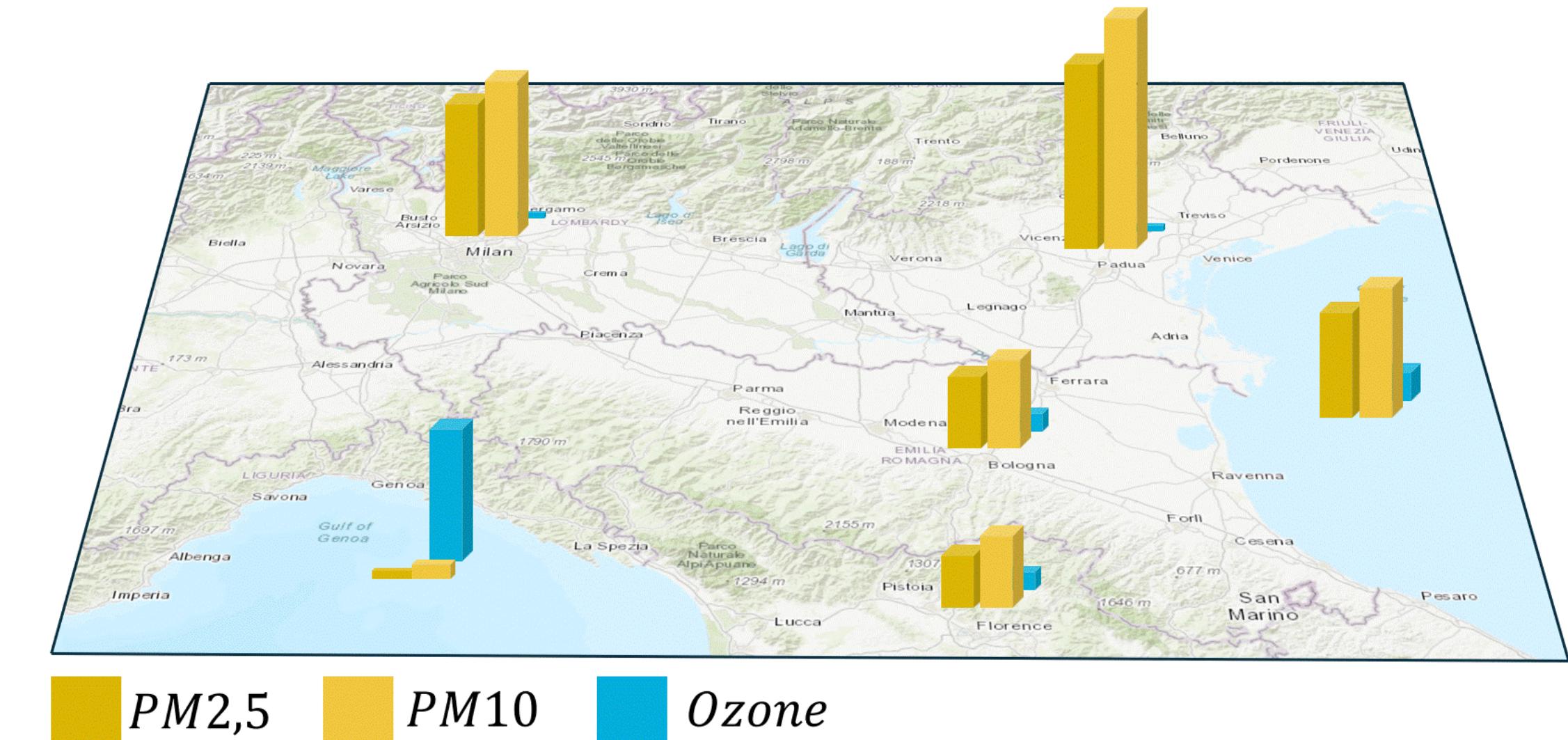


Case Study on large-scale Air Pollution Model

- We want jointly model Ozone, PM2.5 and PM10

- Consider region over northern Italy in 2022

- Gridded observation data available from Copernicus Atmosphere Monitoring Service (CAMS)





Defining our Multivariate Model

- Trivariate model: PM2.5, PM10, Ozone
- Each univariate model contains spatio-temporal field
- Spatial mesh: 4210 mesh nodes
- Temporal mesh: 48 days
- Covariates: Altitude, Intercept

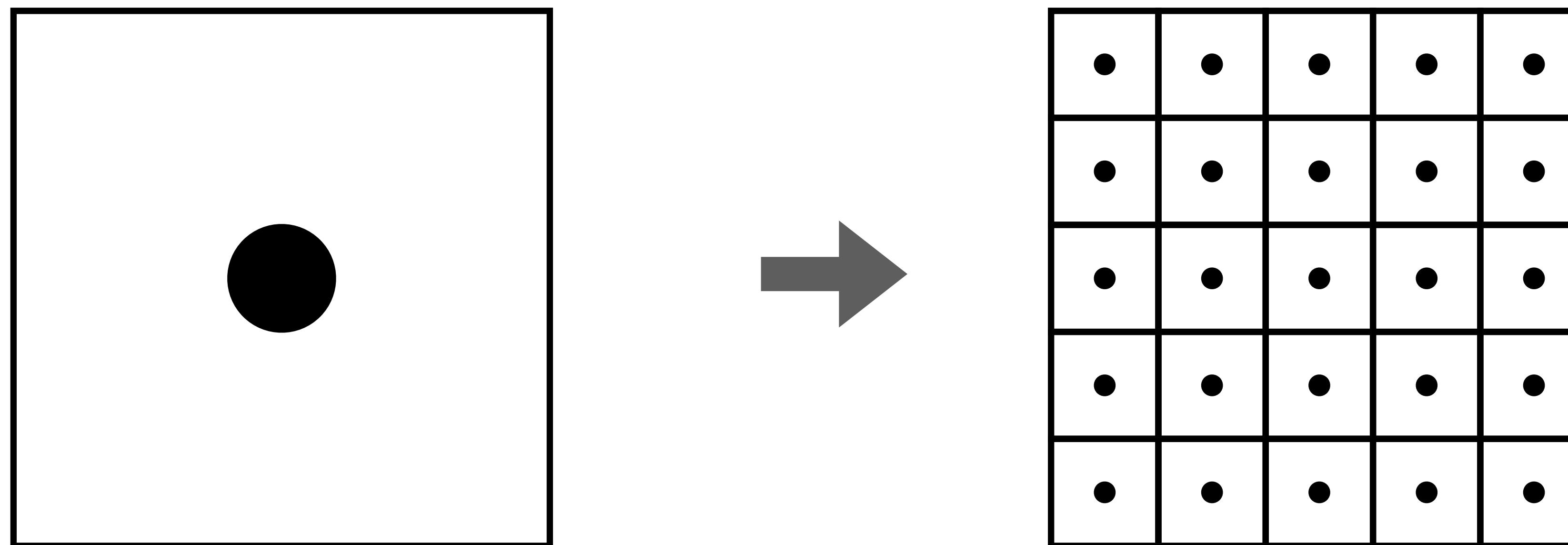




Defining our Multivariate Model

We want to perform spatial down-scaling

- Decrease spatial resolution from 0.1 to 0.02 degrees



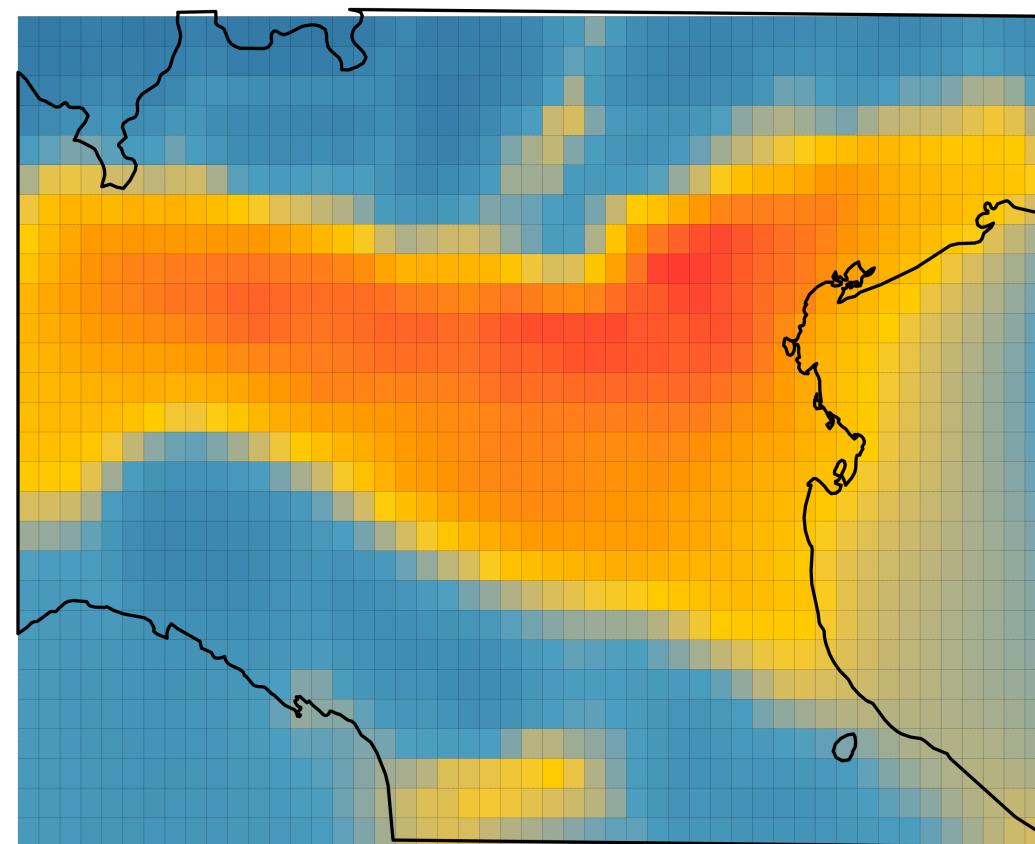


Defining our Multivariate Model

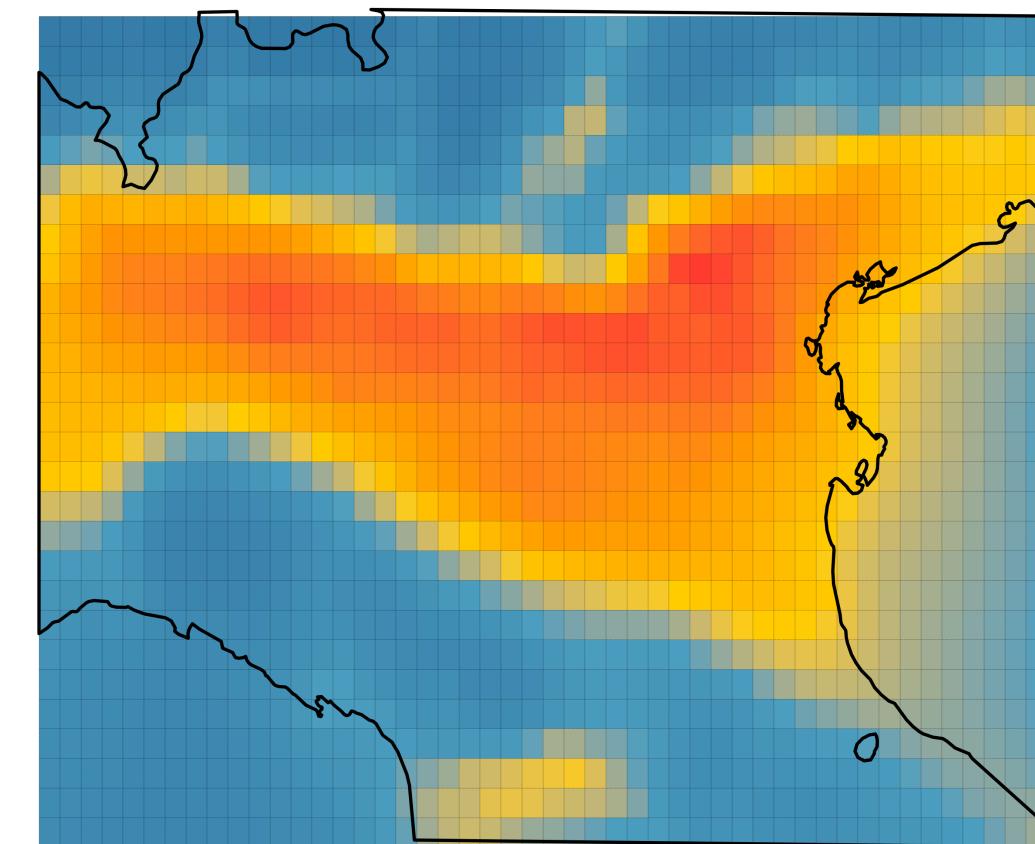
- We want to perform spatial down-scaling

- Decrease spatial resolution from 0.1 to 0.02 degrees

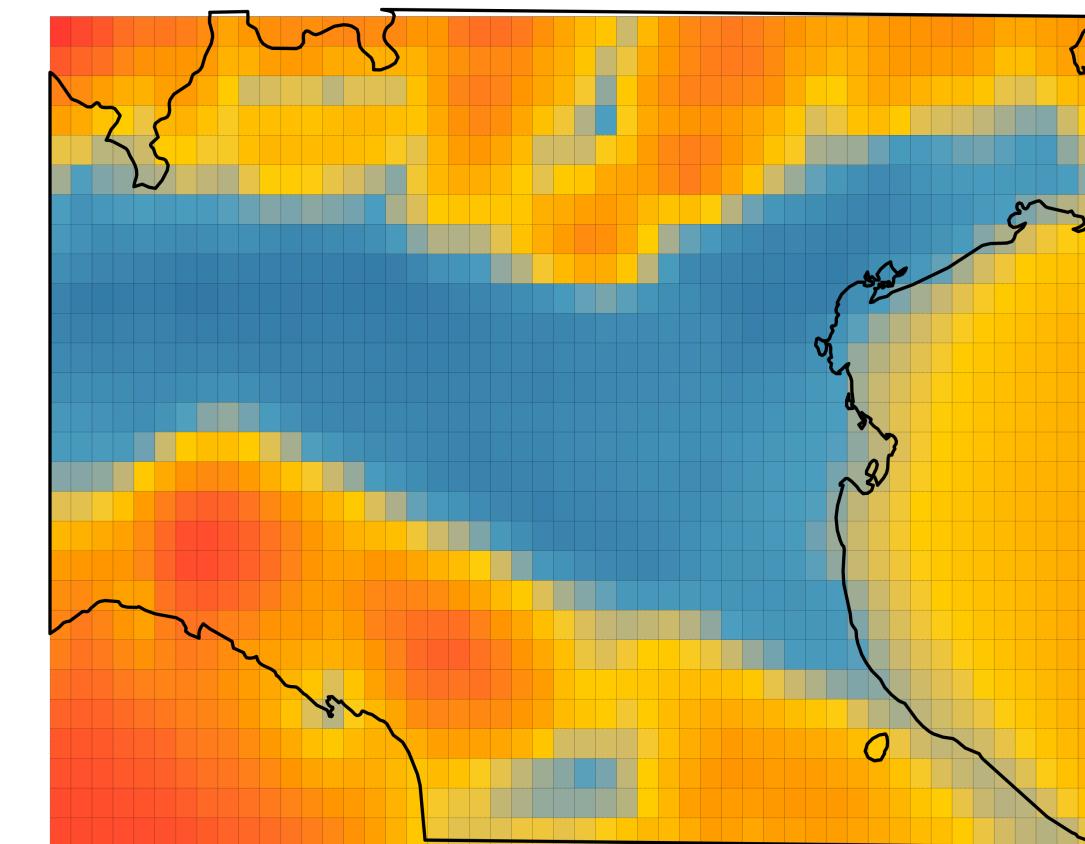
- Data averaged over entire time period



PM10



PM2.5



Ozone



How to define this in INLA?



Posterior Results

What are our correlation coefficients?

$$\begin{array}{ll} \text{PM2.5} & \boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \end{bmatrix} = \begin{bmatrix} \sigma_1 \mathbf{I} & 0 & 0 \\ \lambda_1 \sigma_1 \mathbf{I} & \sigma_2 \mathbf{I} & 0 \\ (\lambda_3 + \lambda_1 \lambda_2) \sigma_1 \mathbf{I} & \lambda_2 \sigma_2 \mathbf{I} & \sigma_3 \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{B}_1 & 0 & 0 \\ 0 & \mathbf{B}_2 & 0 \\ 0 & 0 & \mathbf{B}_3 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} \\ \text{PM10} \\ \text{Ozone} \end{array}$$

$$\lambda_1 = 0.97$$

→ PM2.5 and PM10 are strongly positively correlated

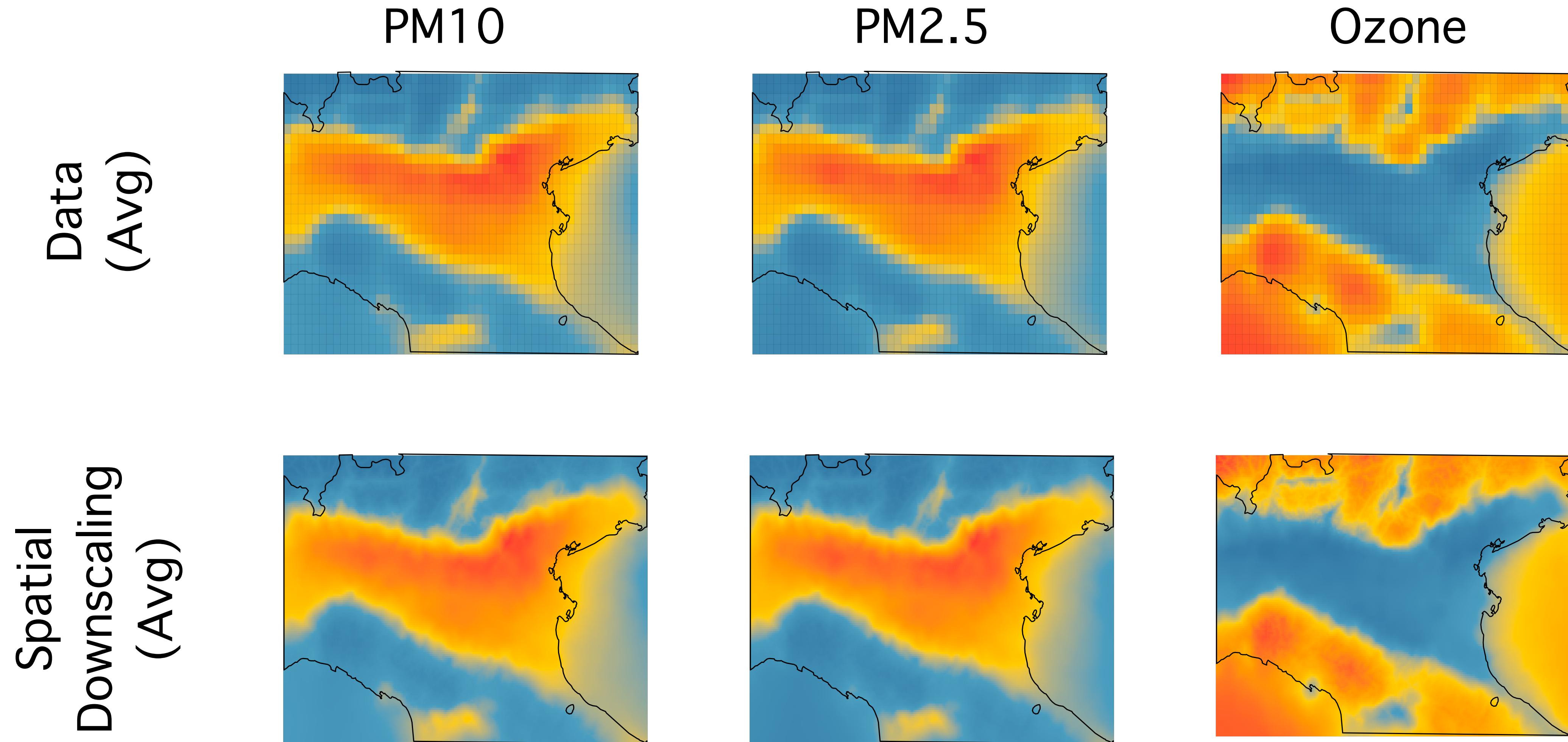
$$\lambda_2 = -0.63$$

→ PM2.5/PM10 and Ozone are negatively correlated

$$\lambda_3 = -0.61$$



Posterior Results

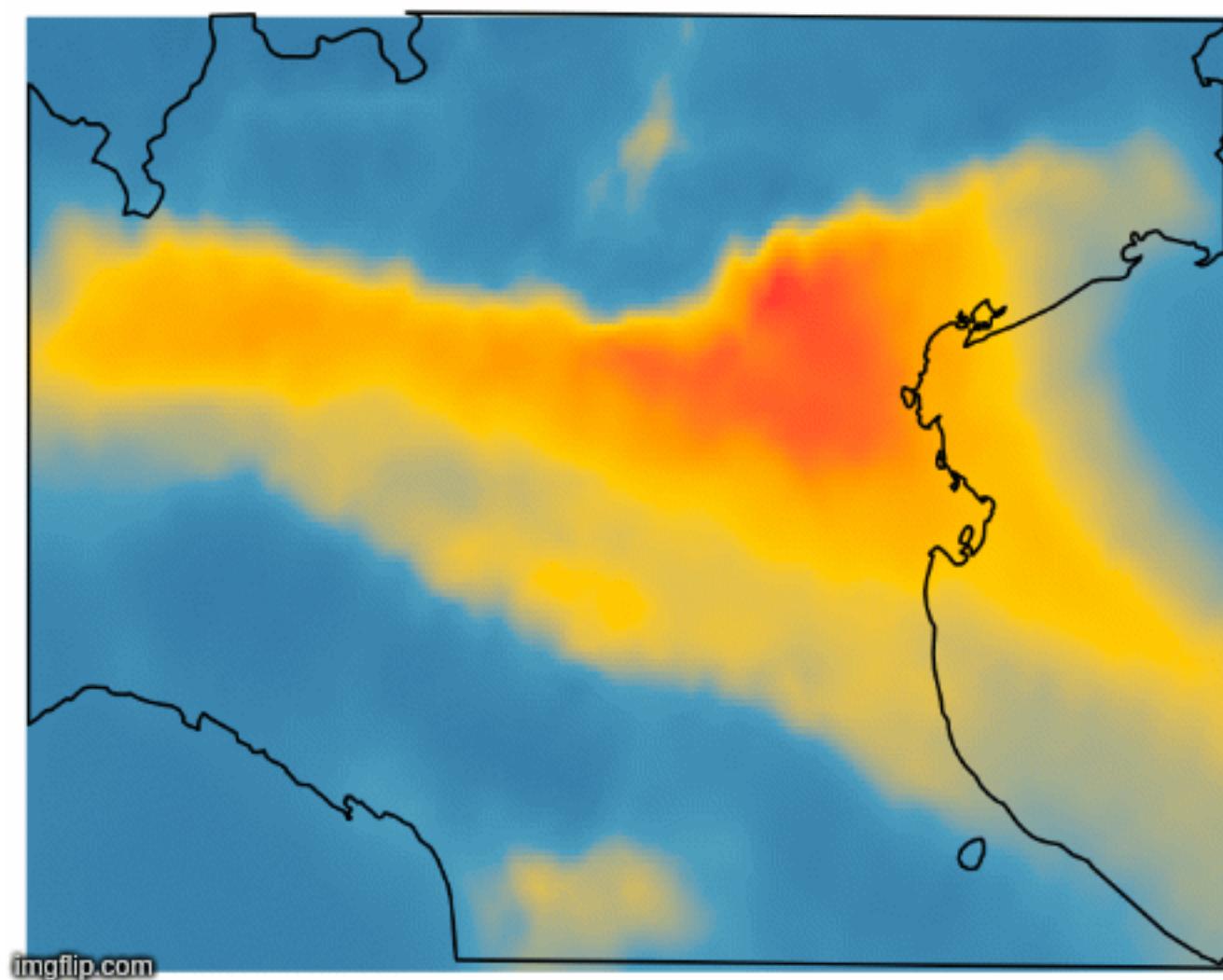




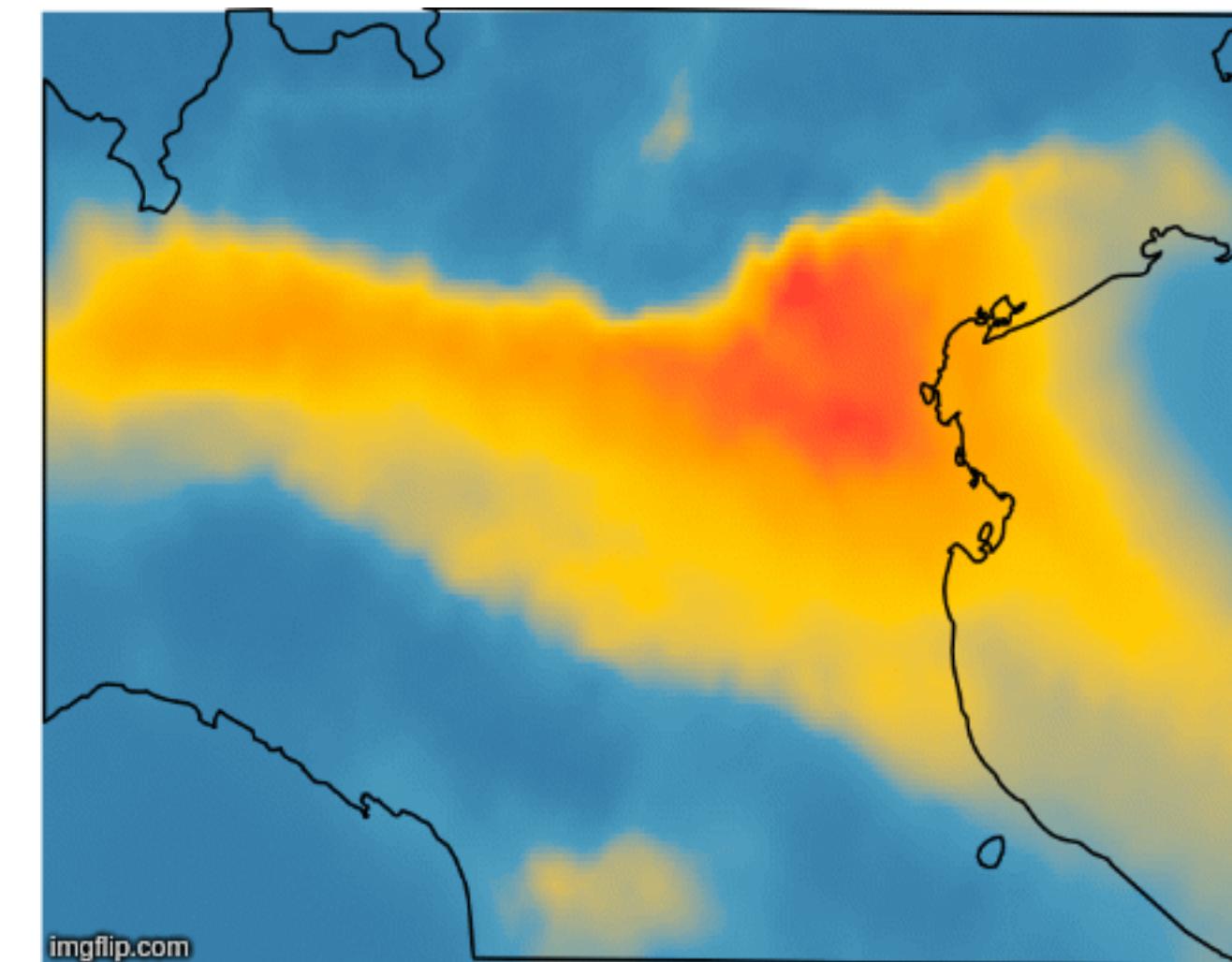
Posterior Results

Spatial
Downscaling
Over Time

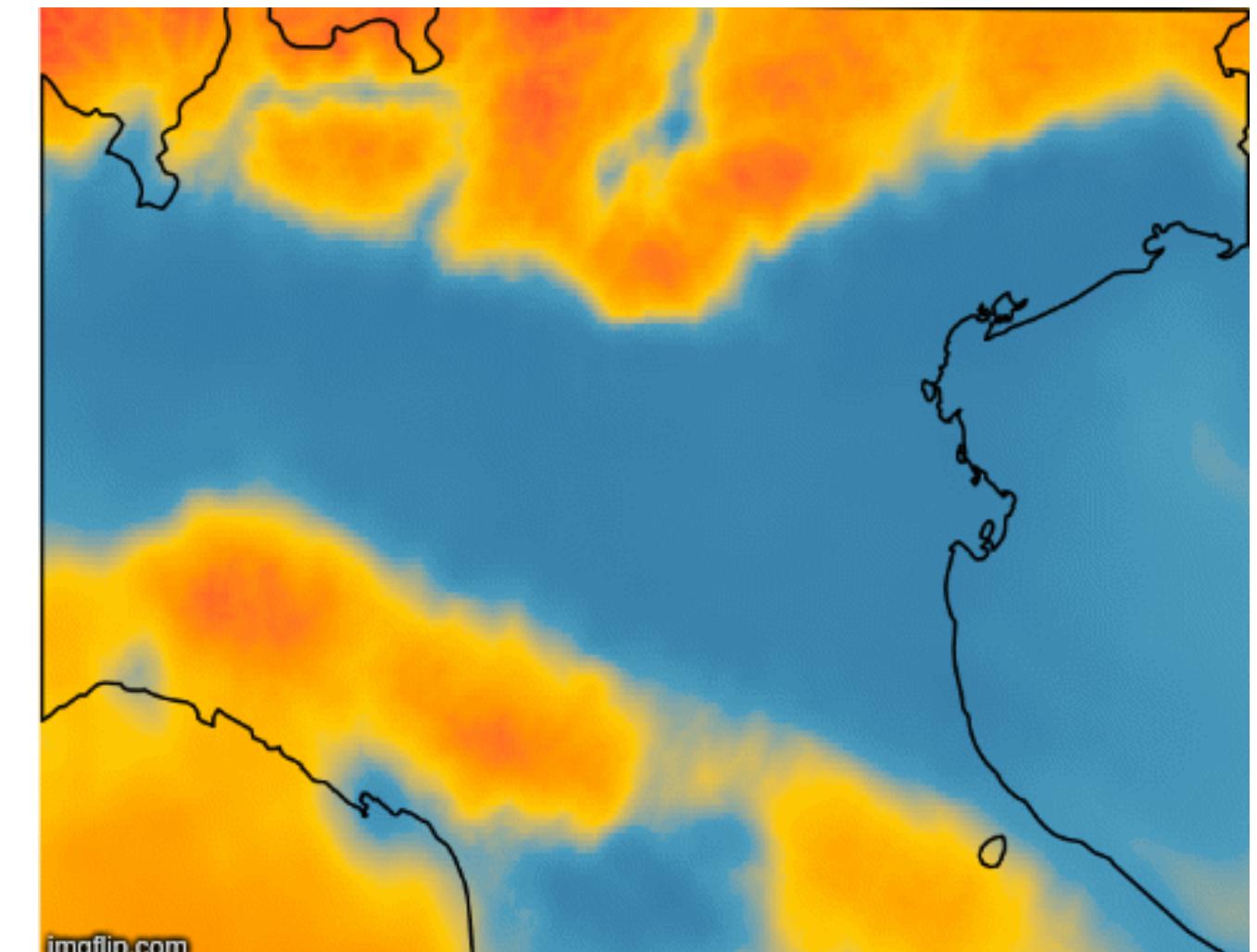
PM10



PM2.5

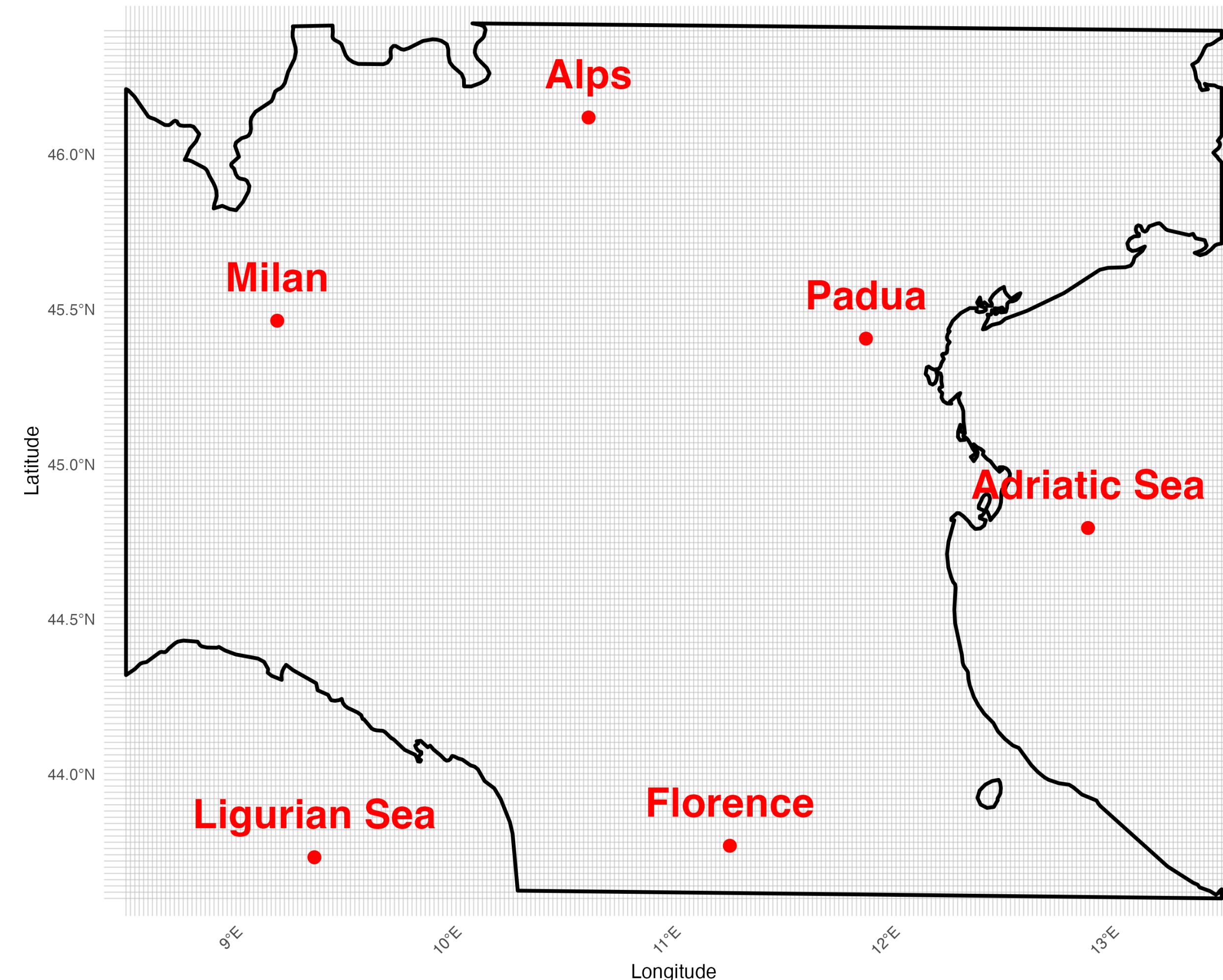


Ozone





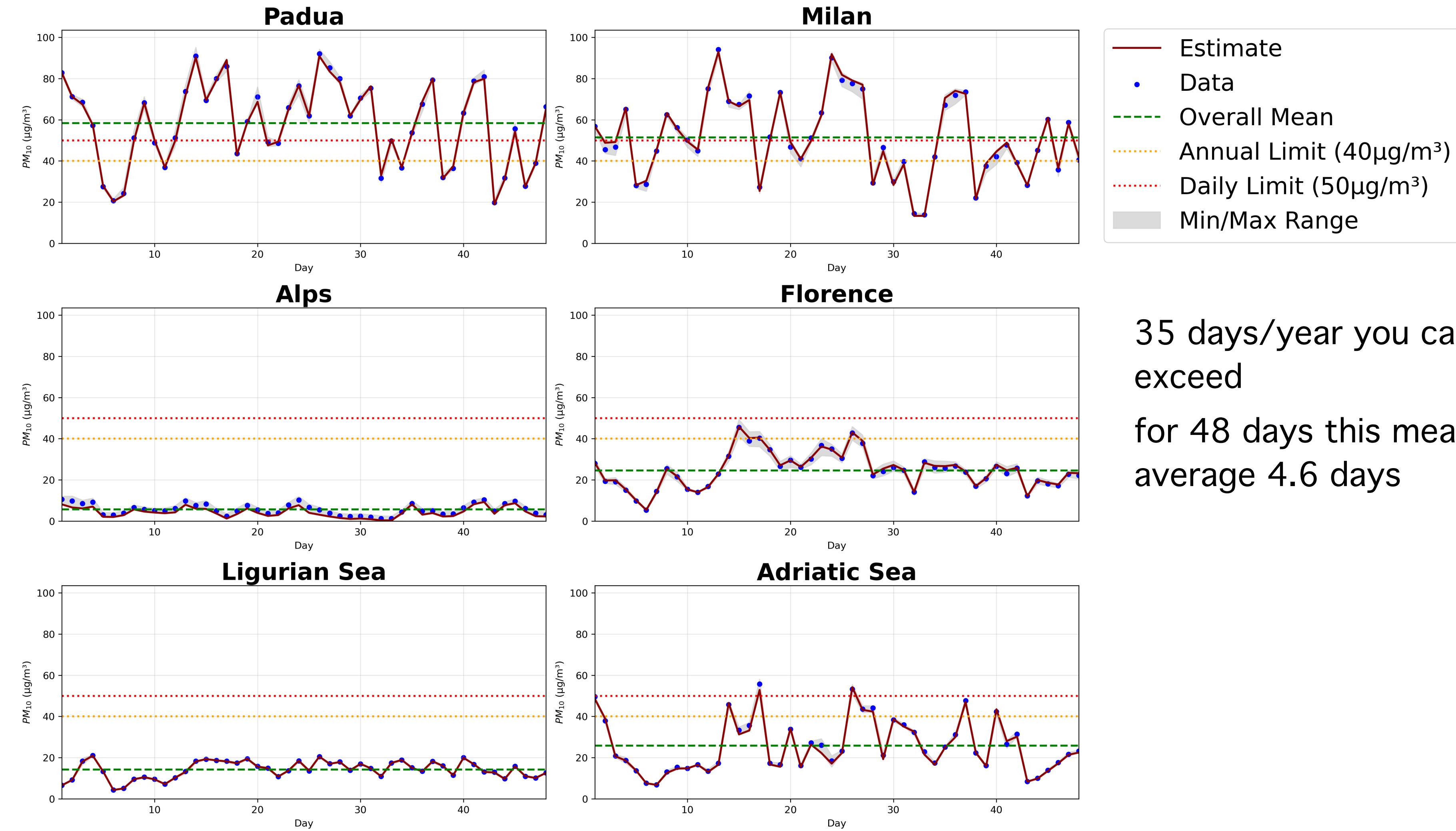
Posterior Results





Posterior Results

PM₁₀ Concentration by City

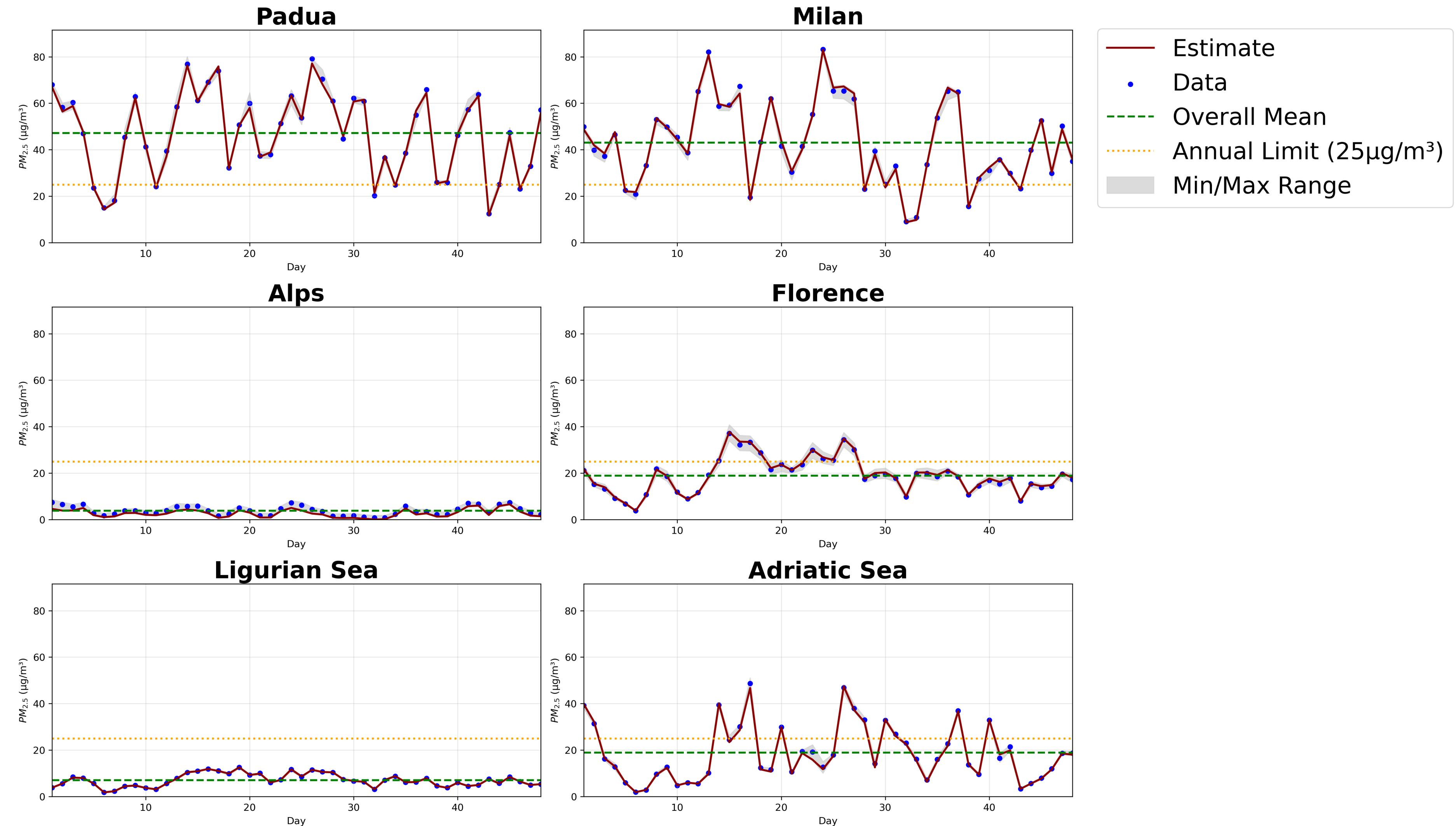


35 days/year you can exceed
for 48 days this means on average 4.6 days



Posterior Results

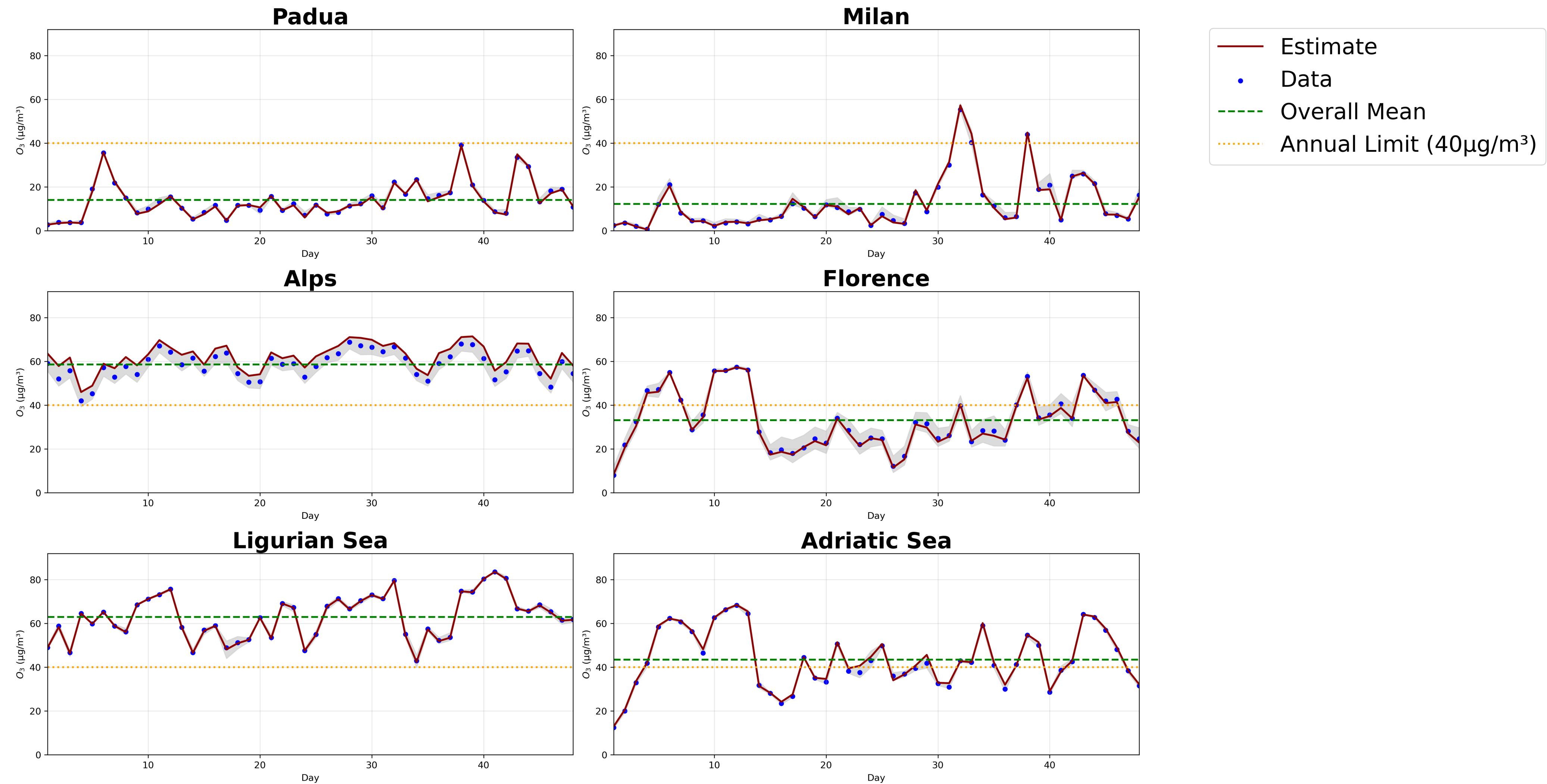
PM_{2.5} Concentration by City





Posterior Results

Ozone Concentration by City





General Considerations

Increasing variance / uncertainty for each added process

- Not recommended to combine more than 3 processes

Performance considerations

- Resulting model has a VERY large number of latent parameters (>1 mio)
- NOT possible to run on regular laptop, needs to be run on supercomputer
- We ran this model on 20 GPUs and it took about multiple hours



Thank you very much for your attention.

Questions?