Ozone Pollution in Lombardy

A spatio–temporal Bayesian model for O_3 pollution

Matteo Coppellotti, Alessandro Ganelli, Francesco M. Mancinelli,
Leonardo Marchesin, Beatrice Sisti, Alessandro Venanzi
Tutored by Michela Frigeri and Prof. Alessandra Guglielmi

The data

- Hourly $[O_3]$ recorded by ARPA
- 51 stations across Lombardy
- Data collected in 2010-2022

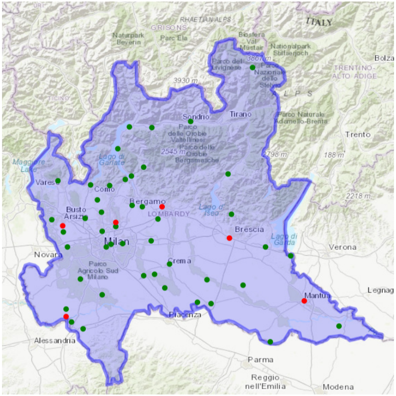


Figure: ARPA monitoring stations

Quantities of interest

- Bayesian modeling of two quantities under study:
  - **#days in each month** with at least one hour overcoming the threshold of **180 mg/m**^{3}
  - **#days in each month** with at least one 8-hours moving average overcoming the threshold of **120 mg/m**^{3}
- Propose and compare different Bayesian modeling frameworks
- Identify similarities and differences in parameters' posterior estimates under these two scenarios

Dealing with missing values

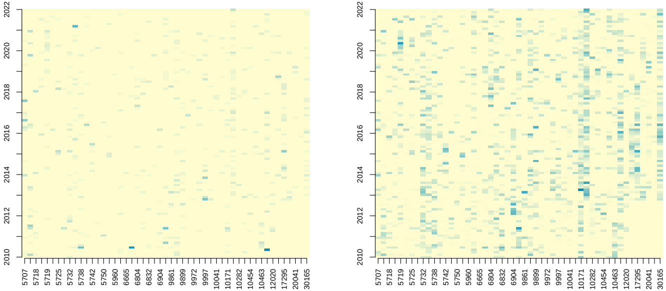
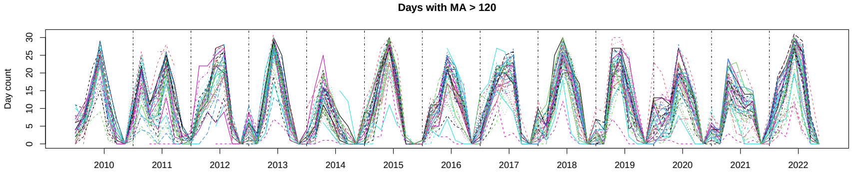


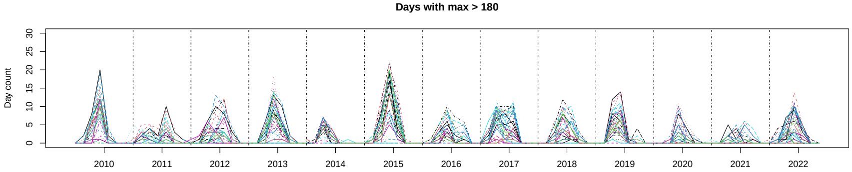
Figure: Daily and hourly missing data, respectively

Dealing with missing values

- A month with six or more days of missing data is considered missing in its entirety.
- For the **180 dataset**: a day is considered not valid if it lacks 6 or more hours of data and the reported values don't overcome 180.
- For the **120 dataset**: the moving average is considered valid if there are at least 6 actual measured values.

Quantities of interest





Covariates

- Station altitude
- Zone: Urban, Industrial, Rural
- Population density
- Weather
  - Temperature
  - Wind
  - Precipitations
  - Solar radiations

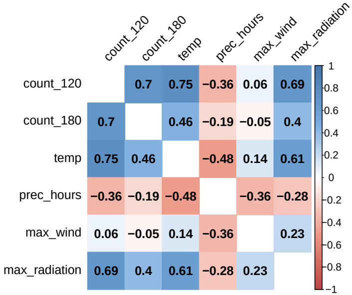


Figure: Correlation between some of the most informative weather variables

Covariates

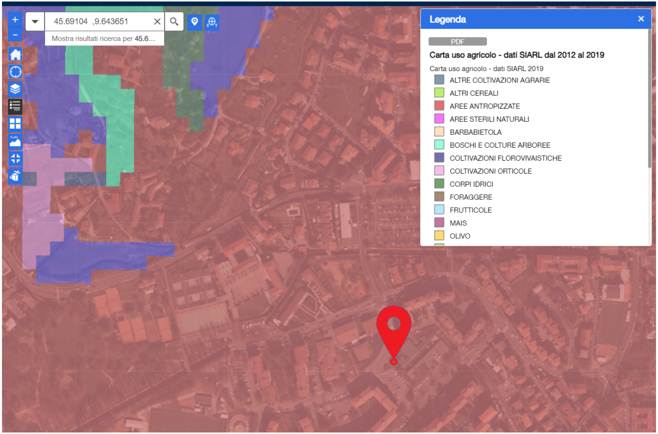
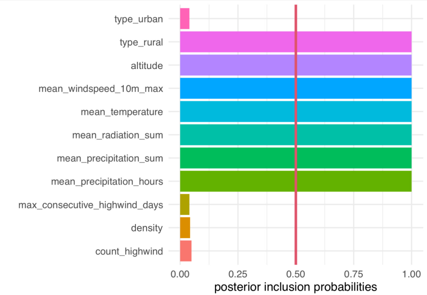


Figure: Station 17297, Bergamo, *Geoportale della Lombardia*

Covariate selection



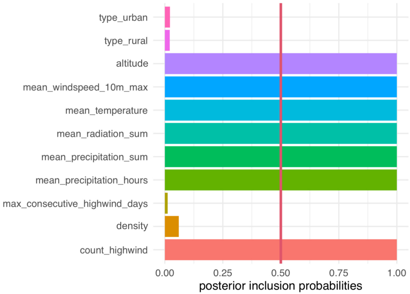


Figure: MPM for *Model 180* and *Model 120*.

120 dataset: final model

$$n_{ijk} \mid \lambda_{ijk} \overset{ind}{\sim} Poi(\lambda_{ijk})$$

$$log(\lambda_{ijk}) = \underline{x}_{ijk}^T \underline{\beta} + \xi_j + \eta_k + \mathrm{w}_k + \gamma_i \cdot \delta_{\{i==\mathrm{July}\}}$$

$$\underline{\beta} \mid \sigma_\beta^2 \sim \mathcal{N}_p(\underline{0}, \sigma_\beta^2 \mathbf{I})$$

$$\underline{\xi} = (\xi_{2010}, \ldots, \xi_{2022}) \mid \sigma_\xi^2 \sim \mathcal{N}_p(\underline{0}, \sigma_\xi^2 \mathbf{I})$$

$$\underline{\eta} = (\eta_1, \ldots, \eta_{45}) \mid \sigma_\eta^2 \sim \mathcal{N}_p(\underline{0}, \sigma_\eta^2 \mathbf{I})$$

$$\mathcal{H} : (\mathcal{H})_{ij} = exp(\frac{1}{\rho} \cdot \mathrm{dist}(\mathrm{s}_i, \mathrm{s}_j))$$

$$\gamma_1, \cdots, \gamma_{45} \mid \sigma_\gamma^2 \overset{\mathrm{iid}}{\sim} \mathcal{N}_p(0, \sigma_\gamma^2)$$

$$\sigma_\beta^2, \sigma_\xi^2, \sigma_\gamma^2, \sigma^2 \overset{\mathrm{iid}}{\sim} inv - gamma(4, 2)$$

$$i = \mathrm{April}, \ldots, \mathrm{October} \quad j = 2010, \ldots, 2022 \quad k = 1, \ldots, 45$$

120 dataset: final model

The total number of predictions over the monthly maximum is: 26

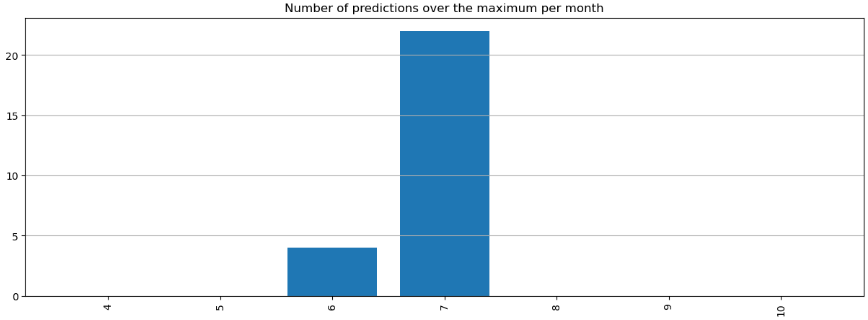


Figure: Predictions over the maximum by month.

120 dataset: posterior predictive performance

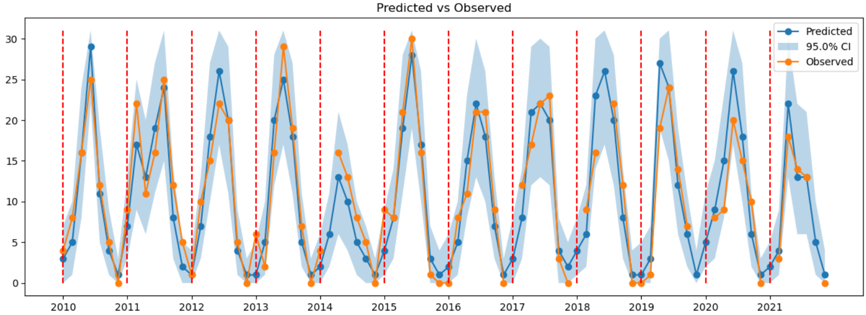


Figure: Posterior predictive median and 95% CIs vs observed values.

180 dataset: ZIP spatial model

$$n_{ijk} \mid \lambda_{ijk}, \theta_i \overset{ind}{\sim} \theta_i \cdot 0 + (1-\theta_i) \cdot \{Poi(\lambda_{ijk})\}$$

$$log(\lambda_{ijk}) = \underline{x}_{ijk}^T \underline{\beta} + \xi_j + \eta_k$$

$$\underline{\beta} \mid \sigma_\beta^2 \sim \mathcal{N}_p(\underline{0}, \sigma_\beta^2 \mathbf{I})$$

$$\underline{\xi} = (\xi_{2010}, \dots, \xi_{2022}) \mid \sigma_\xi^2 \sim \mathcal{N}_p(\underline{0}, \sigma_\xi^2 \mathbf{I})$$

$$\underline{\eta} = (\eta_1, \dots, \eta_{45}) \mid \sigma^2 \sim \mathcal{N}_p(\underline{0}, \sigma^2 \cdot \mathcal{H})$$

$$\mathcal{H} : (\mathcal{H})_{ij} = exp(\frac{1}{\rho} \cdot \mathrm{dist}_e(\mathrm{s}_i, \mathrm{s}_j))$$

$$\sigma_\beta^2, \sigma_\xi^2, \sigma^2 \overset{\text{iid}}{\sim} inv - gamma(4, 2)$$

$$\theta_1, \dots, \theta_7 \overset{\text{iid}}{\sim} \mathcal{B}\mathrm{ern}(\frac{1}{2})$$

$i = \text{April}, \dots, \text{October} \quad j = 2010, \dots, 2022 \quad k = 1, \dots, 45$

180 dataset: posterior predictive performance

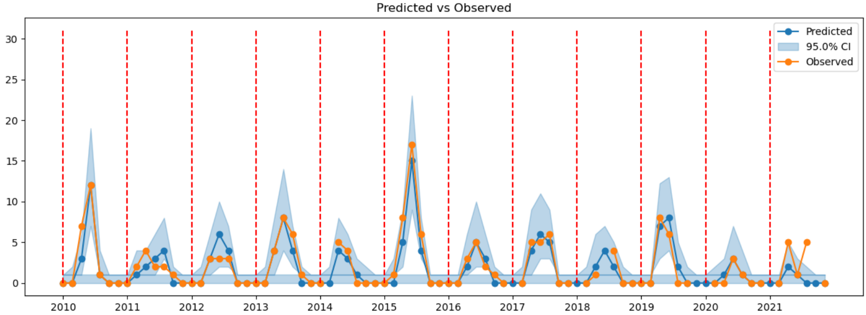


Figure: Posterior predictive median and 95% CIs vs observed values.

Point estimate for eta

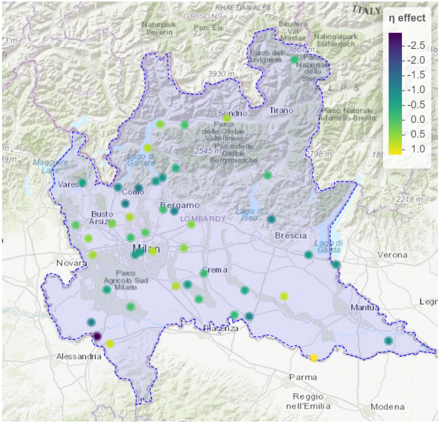


Figure: 180 model

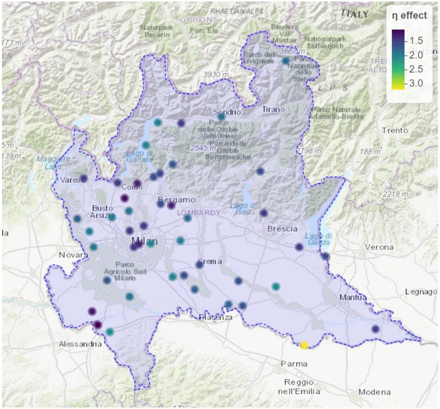


Figure: 120 model

Binomial model

$$n_{ijk} \,|\, \theta_{ijk} \overset{ind}{\sim} Bin(N_i, \theta_{ijk})$$

$$logit(\theta_{ijk}) = x_{ijk}^T \beta + \gamma_i + \xi_j + w_k,$$

$$i = 1, \ldots, 7, \;\; j = 1, \ldots, 13, \;\; k = 1, \ldots, 45$$

$$\gamma_i \;|\; \sigma_\gamma^2 \overset{iid}{\sim} \mathcal{N}(\mu_\gamma, \sigma_\gamma^2) \text{ for } i = 1, \ldots, 7$$

$$\xi_j \;|\; \sigma_\xi^2 \overset{iid}{\sim} \mathcal{N}(\mu_\xi, \sigma_\xi^2) \text{ for } j = 1, \ldots, 13.$$

$$w \;|\; \Sigma, \sigma_s^2 \sim \mathcal{N}(0, \Sigma + \sigma_s^2 H)$$

$$H_{ij} = \exp\left(\frac{1}{\rho} \text{ dist}(s_i, s_j) \right)$$

$$\beta \;\sim \mathcal{N}(0, \mathbf{I}); \mu_\gamma, \mu_\xi \overset{iid}{\sim} \mathcal{N}(0, 1)$$

$$\sigma_{\gamma,i}^2, \sigma_{\xi,j}^2, \sigma^2, \sigma_{s,k}^2 \overset{iid}{\sim} \text{Inv-Gamma}(4, 2)$$

95% Marginal posterior Credible Intervals for β

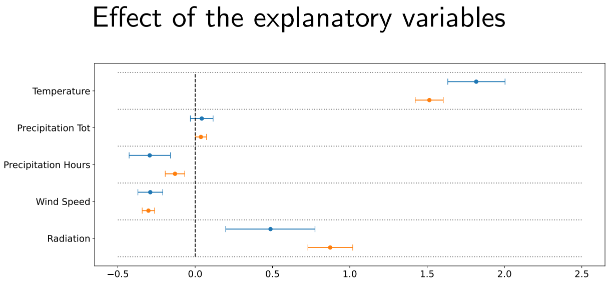


Figure: Orange: 120 model, Blue: 180 model

95% Marginal posterior Credible Intervals for ξ

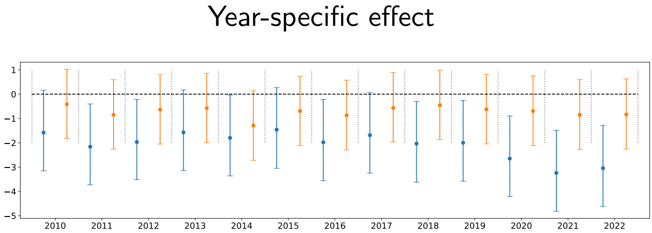


Figure: Orange: 120 model, Blue: 180 model

References

Istituto Nazionale di Statistica.
www.istat.it, 2023.

Konstantinos Fokianos.
Truncated poisson regression for time series of counts.
Scandinavian Journal of Statistics, 2001.

Geoportale Regione Lombardia.
www.geoportale.regione.lombardia.it, 2023.

Ming-Hung Tsai and Tsair-Wei Lin.
Modeling data with a truncated and inflated poisson distribution.
Statistical Methods and Applications, 2017.

Patrick Zippenfenig.
Open-meteo.com weather api, 2023.