Ozone concentration in Lombardy

Barta M., Lenzi L., Makishima N., Mtibaa F., Trottier S., Villa R.

February 2024

Table of Contents

Table of Contents

POLITECNICO MILANO 1863

The problem

Ozone : a threat to the environment and health

Respiratory illness :

- Asthma
- Lung issues
- Stroke

Around 20,000 people die prematurely in Europe each year.

Plant absorption :

- Impact on crops
- Food supply and economical loss

POLITECNICO
MILANO 1863

Mapping the Ozone levels in Lombardy

Known elevation factors :

- High temperatures
- Solar radiation

Known reduction factors :

- Rain
- Humidity

Still hard to predict the area of contamination :

- Wind-driven
- Dependent on elevation

POLITECNICO
MILANO 1863

Table of Contents

POLITECNICO MILANO 1863

Explorative analysis of the data

The location of interest for our study : Lombardy

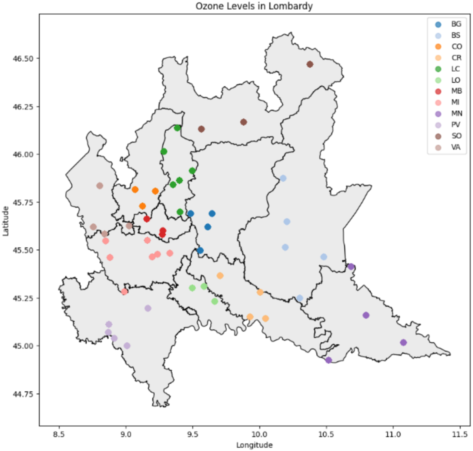
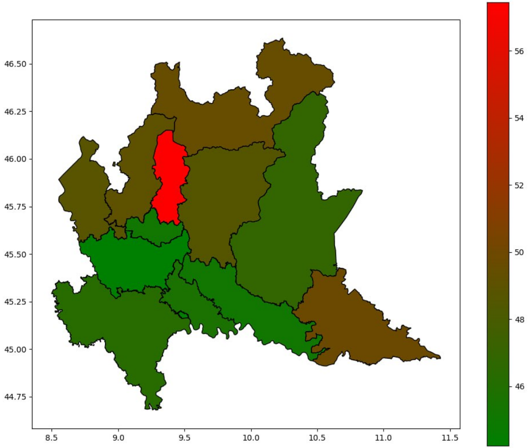


Figure: Ozone sensors in Lombardy

POLITECNICO MILANO 1863

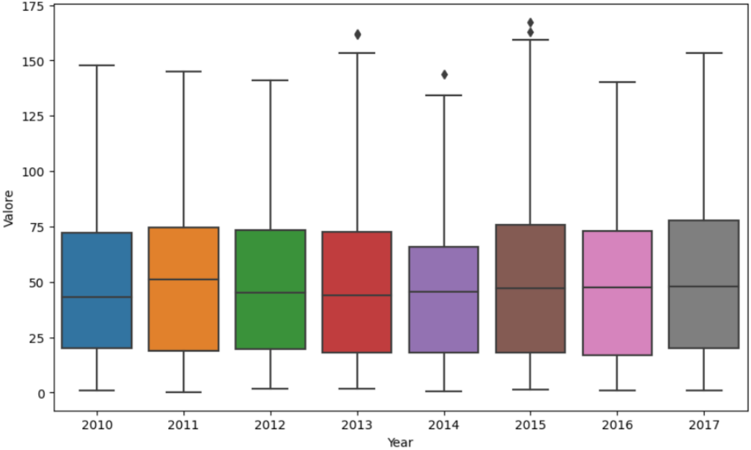
Overview of mean ozone levels in Lombardy (2010-2017)



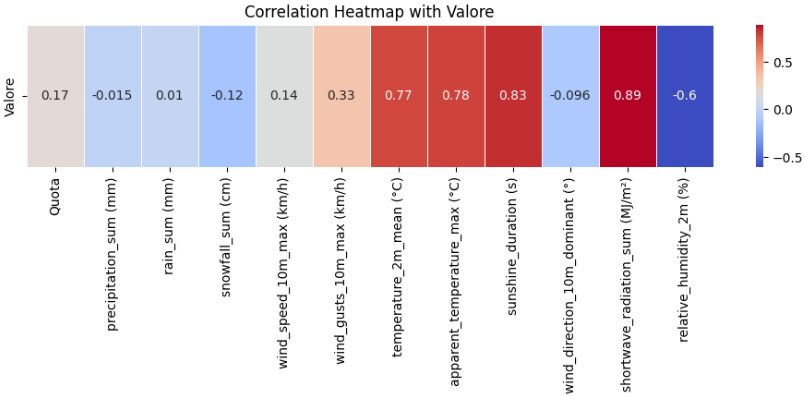
POLITECNICO
MILANO 1863

Explorative analysis of the data

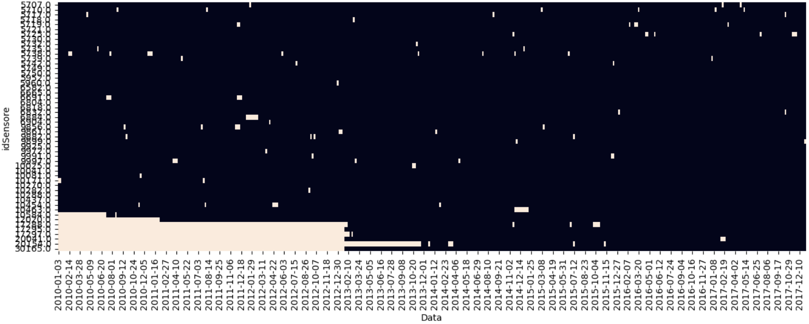
Boxplot of ozone levels over years (2010-2017)



Exploration of meteorological effects



Missing data by sensor and time



POLITECNICO
MILANO 1863

Table of Contents

POLITECNICO
MILANO 1863

The variables and their meaning

Main variables:

- $O_l(s_i, w)$: True Ozone level during year l at place s_i, for week w.
- $Z_l(s_i, w)$: Measured Ozone level.
- $x_{lj}(s_i, w)$: Measured meteorological variable j in the p of interest.
- $\delta_{lj}(s_i, w) = x_{lj}(s_i, w) - x_{lj}(s_i, w-1)$

First assumptions:

- $Z_l(s_i, w) = O_l(s_i, w) + \epsilon_l(s_i, w)$

$\{\epsilon_l(s_i, w), \forall(l, i, w)\}|\sigma_\epsilon \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$

- $\delta_{lj}(s_i, w)$: spatially correlated, temporally independent.

We also assume a correlation due to the weather parameters.

POLITECNICO MILANO 1863

A model of interest

An autoregressive model[3]

$$O_l(\mathbf{s}, w) = \rho O_l(\mathbf{s}, w-1) + \xi_l \mathbf{1} + \delta_l(\mathbf{s}, w)\boldsymbol{\beta} + \eta_l(\mathbf{s}, w)$$

- $0 < \rho < 1$ is the autoregressive factor.
- ξ_l is the global annual intercept of year l.
- $\delta_l(\mathbf{s}, w)\boldsymbol{\beta}$ is the weather adjustment term.
- $\eta_l(\mathbf{s}, w)$ is a spatially correlated error term.

The initial condition

$$O_l(\mathbf{s}, 1) = \mu_l \mathbf{1} + \gamma_l(\mathbf{s})$$

- $\gamma_l(\mathbf{s})$ is the regional effect in year l for all sites s_i.

POLITECNICO
MILANO 1863

A model of interest

Priors for the different terms of the model

- $\rho \sim N(0,1)$
- $\xi_1, ..., \xi_L \overset{iid}{\sim} N(0,1)$
- $\eta_1(\mathbf{s},1), ..., \eta_L(\mathbf{s},1), ..., \eta_1(\mathbf{s},W), ..., \eta_L(\mathbf{s},W) | \sigma_\eta, \phi_\eta \overset{iid}{\sim} N(0, \Sigma_\eta)$,

with $\Sigma_\eta(i,j) = \sigma_\eta^2 exp(-\phi_\eta d_{ij})$

- $\mu_1, ..., \mu_L \overset{iid}{\sim} N(0,1)$
- $\gamma_1(\mathbf{s}), ..., \gamma_L(\mathbf{s}) | \sigma_\gamma, \phi_\gamma \overset{iid}{\sim} N(0, \Sigma_\gamma)$,

with $\Sigma_\gamma(i,j) = \sigma_\gamma^2 exp(-\phi_\gamma d_{ij})$

- $\boldsymbol{\beta} \sim N(0, I_p)$
- $\forall (l,w) \in [1,L] \times [1,W], \delta_l(\mathbf{s},w) | A, \boldsymbol{\phi} \overset{iid}{\sim} N(0, \Sigma_\delta)$

with $A = (\mathbf{a}_1, ..., \mathbf{a}_p), \Sigma_\delta(i,j) = \sum_{k=1}^p \exp(-\phi_k d_{ij}) \mathbf{a}_k^T \mathbf{a}_k$

- $\phi_1, ..., \phi_p \overset{iid}{\sim} U(0.001, 0.1)$
- $\forall (i,j) \in [1,\text{p}]^2, a_{i,j} \overset{iid}{\sim} N(0,1)$
- $\frac{1}{\sigma_\epsilon^2}, \frac{1}{\sigma_\eta^2}, \frac{1}{\sigma_\gamma^2} \overset{iid}{\sim} Gamma(a,b)$

POLITECNICO MILANO 1863

Table of Contents

POLITECNICO MILANO 1863

A reduced model

$$O_l(\mathbf{s}, w) = \rho O_l(\mathbf{s}, w-1) + \xi_l \mathbf{1} + \delta_l(\mathbf{s}, w)\boldsymbol{\beta} + \eta_l(\mathbf{s}, w)$$

The initial condition is:

$$O_l(\mathbf{s}, 1) = \mu_l \mathbf{1} + \gamma_l(\mathbf{s})$$

The priors are :

- $\rho \sim N(0,1)I(0 < \rho < 1)$
- $\xi_1, ..., \xi_L \overset{iid}{\sim} N(0,1)$
- $\eta_1(\mathbf{s},1), ..., \eta_L(\mathbf{s},1), ..., \eta_1(\mathbf{s},W), ..., \eta_L(\mathbf{s},W)|\sigma_\eta, \phi_\eta \overset{iid}{\sim} N(0, \Sigma_\eta)$, with $\Sigma_\eta(i,j) = \sigma_\eta^2 exp(-\phi_\eta d_{ij})$
- $\mu_1, ..., \mu_L \overset{iid}{\sim} N(0,1)$
- $\gamma_1(\mathbf{s}), ..., \gamma_L(\mathbf{s})|\sigma_\gamma, \phi_\gamma \overset{iid}{\sim} N(0, \Sigma_\gamma)$, with $\Sigma_\gamma(i,j) = \sigma_\gamma^2 exp(-\phi_\gamma d_{ij})$
- $\boldsymbol{\beta} \sim N(0, I_p)$
- $\frac{1}{\sigma_\epsilon^2}, \frac{1}{\sigma_\eta^2}, \frac{1}{\sigma_\gamma^2} \overset{iid}{\sim} Gamma(a,b)$, where $a = 2$ and $b = 1$

POLITECNICO MILANO 1863

The steps of the method

- Split the data in a training/verification part on a shorter time period.
- Train a model with all the covariates.
- Look at the posterior distribution of β.
- Remove the covariates whose coefficient contains 0 in its 95% confidence interval.
- Retrain a reduced model with the remaining covariates
- Compare the results of both models on the verification part of the data.

POLITECNICO MILANO 1863

Selection of the covariates

Precipitation_sum (mm)	(-0.599079, 1.17779)
Rain_sum (mm)	(-1.14169, 0.609446)
Snowfall_sum (cm)	(-0.240372, 0.14824)
Wind_speed_10m_max (km/h)	(-0.0158849, 0.0385158)
Temperature_2m_mean (°C)	(-0.0573153, 0.413884)
Apparent_temperature_max (°C)	(-1.13284, -0.622585)
Sunshine_duration (s)	(0.0111516, 0.177066)
Wind_gusts_10m_max (km/h)	(0.0147535, 0.0742081)
Wind_direction_10m_dominant (°)	(-0.0230213, 0.00380585)
Shortwave_radiation_sum (MJ/m^2)	(-0.313512, 0.0424594)
Relative_humidity_2m (%)	(-0.170991, -0.0886358)

Table: Confidence intervals

POLITECNICO MILANO 1863

Table of Contents

POLITECNICO
MILANO 1863

Model Comparison

Comparison between 3 models

Model 1

- Apparent_temperature_max, Sunshine_duration, Wind_gusts_10m_max, Relative_humidity_2m.
- VMSE : 50169.4225.

Model 2

- Temperature_2m_mean, Apparent_temperature_max, Sunshine_duration, Shortwave_radiation_sum.
- VMSE : 167018.2634.

Model 3

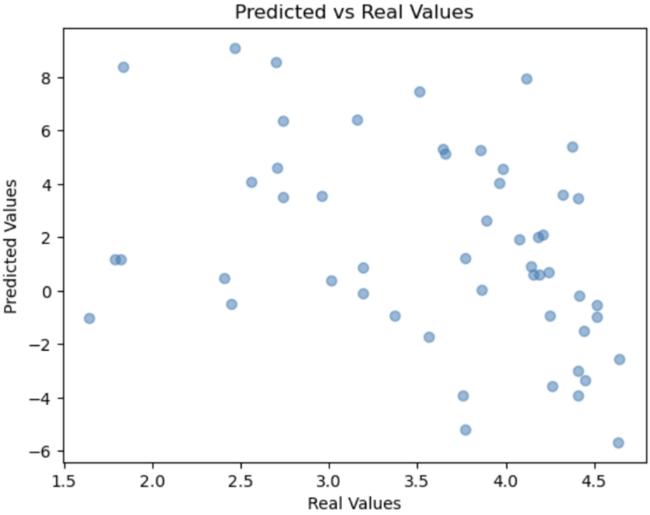
- Apparent_temperature_max, Wind_speed_10m_max, Relative_humidity_2m
- VMSE : 20.2133.

Table of Contents

POLITECNICO MILANO 1863

Predictions obtained with the third trained model

Comparison with some known ground truth



Predictions obtained with the third trained model

Ozone levels predicted at a new location over a year

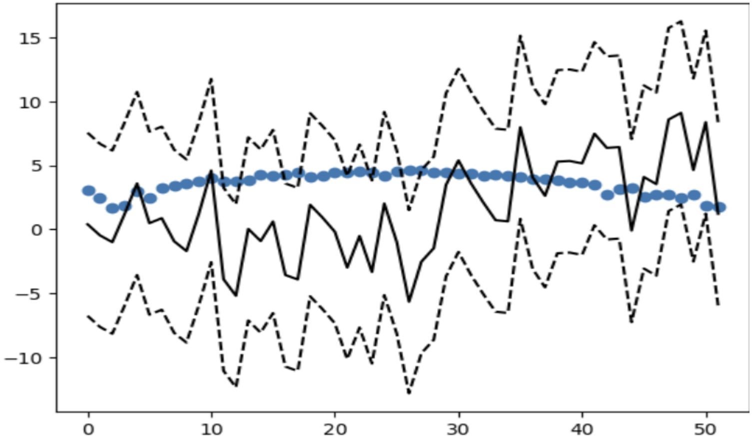


Figure: Prediction for a location in the validation set

Table of Contents

POLITECNICO
MILANO 1863

Conclusion

- Importance of the covariates
- Computational issues
- Possible improvements

Thank you for your attention !

References

[1] Hussein A Al-Amery and Osama T Al-Taai. "The ozone effect on shortwave solar radiation in the atmosphere over Iraq". In: *AIP Conference Proceedings*. Vol. 2290. 1. AIP Publishing. 2020.

[2] Amandine Chevalier et al. "Influence of altitude on ozone levels and variability in the lower troposphere: a ground-based study for western Europe over the period 2001–2004". In: *Atmospheric Chemistry and Physics* 7.16 (2007), pp. 4311–4326.

[3] Sujit K Sahu, Alan E Gelfand, and David M Holland. "High-resolution space–time ozone modeling for assessing trends". In: *Journal of the American Statistical Association* 102.480 (2007), pp. 1221–1234.

POLITECNICO MILANO 1863