Vehicles entering the gates for Area C in Milan

Marco Vivian, Margherita Brogi,
Rossana Cometa, Ardiana Prekazi,
Alessandra Sala, Nick Usubelli

Bayesian Statistics - Politecnico di Milano

Milan, February 15, 2024

Overview

Spatio - temporal analysis of vehicle entries in Area C in Milan

Figure: Area C gates in Milan

Data

After merging data, we selected or created the following variables:

- **data**: dates ranging from January 1st, 2023, to December 31st, 2023;
- **numero_transiti**: the number of passages in Area C;
- **festivo**: a binary variable where 0 indicates a weekday and 1 indicates weekends and holidays;
- **longitude**: the longitudinal coordinate;
- **latitude**: the latitudinal coordinate;
- **nome varco**: the name of the gate;
- **precipitazioni**: the total precipitation recorded in millimeters;
- **av_temp**: the average temperature recorded in degrees Celsius;
- **estate**: a binary variable that is 1 if the day falls in summer, 0 otherwise;
- **inverno**: a binary variable that is 1 if the day falls in winter, 0 otherwise;

The model

Spatio-temporal regression model in which the response variable Y takes on count values $\rightarrow$ **Poisson model**

Y_{it}: number of entrances in gate i on day t

First simple model:

$$Y_{it}|\lambda_{it} \overset{\text{ind}}{\sim} Pois(\lambda_{it})$$
$$log(\lambda_{it}) = f(t) + c_i$$

where:

- the spatial component is given by the gate index $i = 1, .., 40$;
- the temporal component $t = 1, ..., 365$ represent the days in 2023;
- $f(t)$ is a time dependent function following the trend of Y_{it};
- c_i is a gate-specific factor

Modeling $\mathbf{c_i}$:

CARBayesST allows to implement spatio-temporal generalised linear mixed models for areal unit data:

$$Y_{it}|\lambda_{it} \overset{\text{ind}}{\sim} Pois(\lambda_{it})$$
$$log(\lambda_{it}) = O_{it} + \phi_{it}$$

where $i, i = 1, ..., 40$, $t, t = 1, ..., 365$, O_{it} is the offset for gate i and day t, and ϕ_{it} is a latent component for areal unit i and time period t encompassing one or more sets of spatio-temporally autocorrelated random effects.

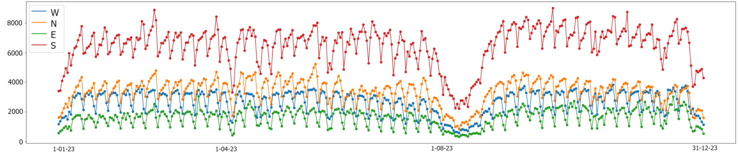


Figure: Area C entries time series

Spatial autocorrelation is controlled by a symmetric non-negative KxK, $K = 40$, proximity matrix **W** $= [w_{ij}]$:

1. $\mathbf{W_1}$: $w_{ij} = 1$ if the gates i and j are adjacent, $w_{ij} = 0$ otherwise;
2. $\mathbf{W_2}$: decreasing values starting from 1 for adjacent gates, 0.5 if gate i is one gate away from j, 0.25 if gate i is two gates away from j and so forth;
3. $\mathbf{W_3}$: values according to the Euclidean distance between gates.

Most suitable model for ϕ: first-order temporal autoregressive process to estimate the evolution of the spatial structure in the data over time.

	W_1	W_2	W_3
DIC	138040.004	137866.878	137831.428
WAIC	135937.018	135386.108	135198.682
LMPL	-72195.480	-71979.841	-71778.294
Log-likelihood	-58053.025	-57867.418	-57789.947

Table: Model fit criteria

STAN model

$$Y_{it}|\lambda_{it} \overset{\text{ind}}{\sim} Pois(\lambda_{it})$$
$$log(\lambda_{it}) = f(t) + \beta_i X_t + c_i$$

- $\mathbf{X_t}$ values of weather conditions on day t
- $\beta_\mathbf{i}$ coefficient that takes into account the effect of $\mathbf{X_t}$ on the different gates
- $\mathbf{f}(t) = a_0 sin(\omega_1 t) + b_0 cos(\omega_1 t) + a_1 sin(\omega_1 t) + b_1 cos(\omega_2 t)$, where
  - $\boldsymbol{\omega}_1 = 2\pi/7$ weekly periodicity
  - $\boldsymbol{\omega}_2 = 2\pi/226$ seasonal periodicity
  - $a_0, b_0, a_1, b_1 \overset{\text{iid}}{\sim} N(0, 6)$
- $\mathbf{c}_i \overset{\text{ind}}{\sim} N(log(\mu_c), 0.3)$ where μ_c are the values estimated using *CARBayesST*
- $\beta_\mathbf{i} \overset{\text{iid}}{\sim} N(0, 1)$

Preliminary analysis on β_i

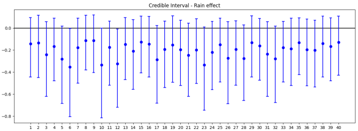


Figure: CI of $\beta_{\mathbf{pioggia}}$

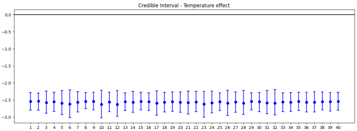


Figure: CI of $\beta_{\mathbf{temperatura}}$

From the $\beta_{\mathbf{pioggia}}$ and $\beta_{\mathbf{temperatura}}$ credible intervals we can see how they are all very similar to each other, meaning that rain has similar influence among the different gates on the number of vehicles entries. The same holds for $\beta_{\mathbf{temperatura}}$.

Traceplots of the weather effects

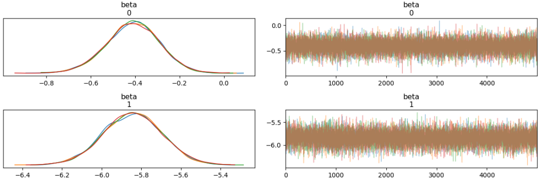


Figure: Traceplots of $\beta_{\mathbf{pioggia}}, \beta_{\mathbf{temperatura}}$

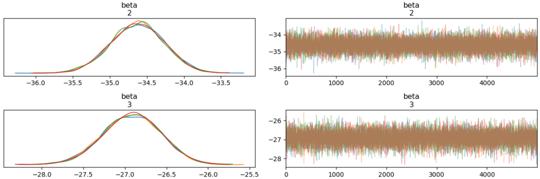


Figure: $\beta_{\mathbf{estate}}, \beta_{\mathbf{inverno}}$

Traceplots of c_i

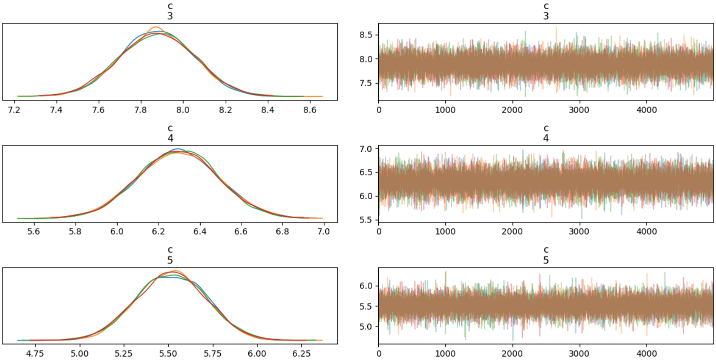


Figure: Traceplots of c_i

Univariate clustering on c_i

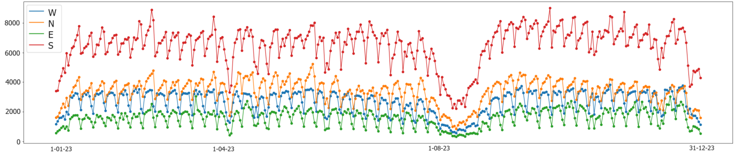


Figure: Area C entries time series

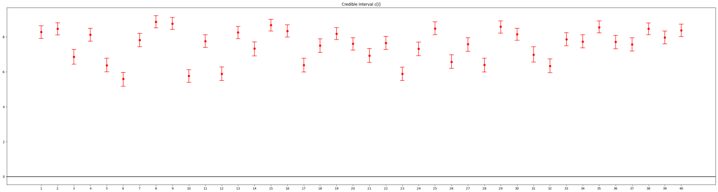


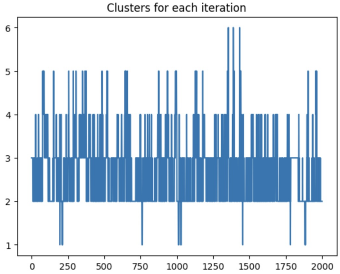
Figure: Posterior 95% credible intervals of gate-specific intercepts $\mathbf{c_i}$

Dirichlet process mixture model with Stick breaking construction:

$$c_i \overset{iid}{\sim} \sum_{k=1}^{C} w_k N(\mu_k, \sigma_k^2) \qquad i = 1, \ldots, 40$$

$$w_k = v_k \prod_{i=1}^{k-1} (1 - v_i) \qquad k = 2, \ldots, C$$

$$w_1 = v_1$$



Results

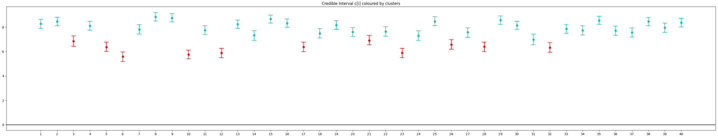


Figure: Posterior 95% credible intervals of gate specific intercepts after clustering

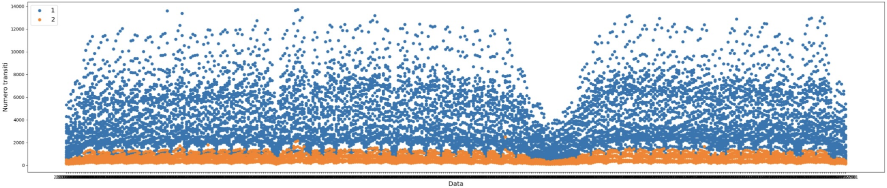


Figure: Area C entries time series after clustering

Interpretation and conclusions

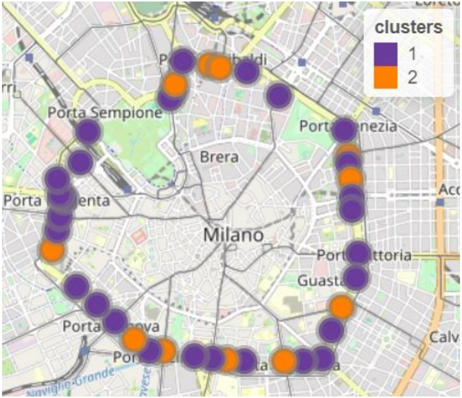


Figure: Area C gates according to the clusters

Thank you!

References:

1. David B. Hitchcock, "Bayesian Count Regression Models", *STAT535: Chapter 12*, 2022.
2. D. Lee, A. Rushworth, G. Napier,, "Spatio-Temporal Areal Unit Modeling in R with Conditional Autoregressive Priors Using the CARBayesST Package", 2018.
3. "Bayesian Hierarchical Poisson Regression Model for Overdispersed Count Data", *SAS/STAT Examples*.
4. Gary L. Rosner, Purushottan W. Laud, Wesley O. Johnson , "Bayesian Thinking in Biostatistics", Chapter 9 (pages 241-264), 2021.
5. M. Frigeri, A. Guglielmi, "Spatio-temporal models for particulate matter in the Po valley", 2022.

Tutors: Michela Frigeri, Alessandro Carminati