A Bayesian model for data flow: BikeMi

Andrea De Gobbis, Lorenzo Ghilotti, Giorgio Meretti

February 19, 2020

Politecnico di Milano

What we are working on

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

The BikeMi stations net in Milan is rather large and travelled every day.

The two perspective of the problem

Our work is structured in two parts:

Global model

The total volume of travels in a specific day Y_t without considering the graph structure. This results in a single time series.

Network model

For each node, the analysis focuses on in and out bikes flow for specific time slots. The dimensionality is much higher.

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Model

$$Y_t | \lambda_t \overset{ind}{\sim} \mathrm{Po}(\lambda_t)$$

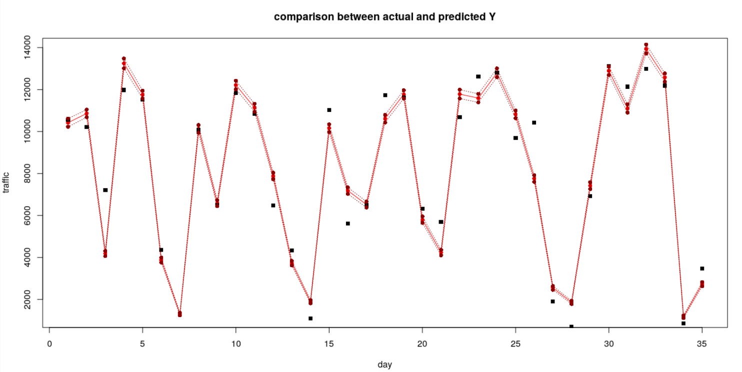
$$\log \lambda_t = \boldsymbol{\beta}^T \mathbf{z}_t$$

Prior

$$\beta_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\beta^2)$$

Loglinear GLM: Poisson likelihood

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti



Negative Binomial distribution

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

$$X \sim \mathrm{NegBin}(p, r)\ if:$$

$$f(k; p, r) = \frac{\Gamma(k+r)}{k!\Gamma(r)} p^k (1-p)^r, \qquad k = 0, 1, 2...$$

$$\mathbb{E}[X] = \mu = \frac{r(1-p)}{p} \qquad Var(X) = \frac{r(1-p)}{p^2}$$

Moreover:

$$p = \frac{r}{r+\mu}$$

Loglinear GLM: NegBin likelihood (1)

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Model

$$Y_t | p_t, r \overset{ind}{\sim} \text{NegBin}(p_t, r)$$

$$p_t = \frac{r}{r + \mu_t}$$

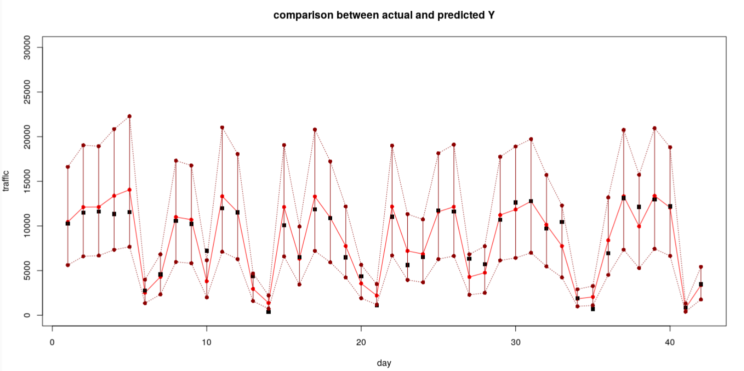
$$\log \mu_t = \boldsymbol{\beta}^T \mathbf{z}_t$$

Priors

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

$$r \sim \mathcal{U}(0, 50)$$

Loglinear GLM: NegBin likelihood (1)



Loglinear GLM: NegBin likelihood (2)

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Model

$$Y_{tj}|p_{tj}, r_j \overset{ind}{\sim} \text{NegBin}(p_{tj}, r_j) \qquad j = 1 : wday, j = 2 : wend$$

$$p_{tj} = \frac{r_j}{r_j + \mu_{tj}}$$

$$\log \mu_{tj} = \beta_1 + \beta_2 R_t + \beta_3 T_t + \theta_j$$

Priors

$$\beta_i|\tau_i \overset{ind}{\sim} \mathcal{N}(0, \tau_i)\ i = 1, 2, 3$$

$$\tau_i \overset{ind}{\sim} Gamma(2, 10)$$

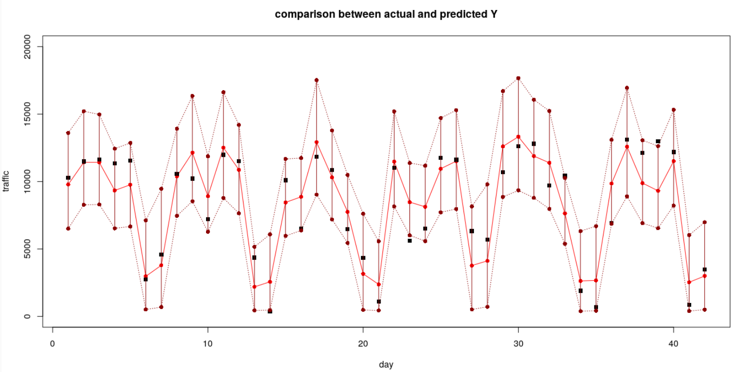
$$\theta_j|\tilde{\tau}_j \overset{ind}{\sim} \mathcal{N}(0, \tilde{\tau}_j)\ j = 1, 2$$

$$\tilde{\tau}_j \overset{ind}{\sim} Gamma(2, 10)$$

$$r_j \overset{ind}{\sim} Unif(0, 50)\ j = 1, 2$$

Loglinear GLM: NegBin likelihood (2)

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti



Ingredients of a BSTS

Main ingredients of BSTS model:

- locally linear trend μ_t,
- periodicity γ_t,
- autoregressive component ρ_t,
- regression $\boldsymbol{\beta}^T \mathbf{z}$.

Ingredients of a BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Main ingredients of BSTS model:

- locally linear trend μ_t,
- periodicity γ_t,
- autoregressive component ρ_t,
- regression $\beta^T \mathbf{z}$.

Ingredients of a BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Main ingredients of BSTS model:

- locally linear trend μ_t,
- periodicity γ_t,
- autoregressive component ρ_t,
- regression $\beta^T \mathbf{z}$.

Ingredients of a BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Main ingredients of BSTS model:

- locally linear trend μ_t,
- periodicity γ_t,
- autoregressive component ρ_t,
- regression $\boldsymbol{\beta}^T \mathbf{z}$.

Standard BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Model

$$Y_t = \mu_t + \gamma_t + \rho_t + \boldsymbol{\beta}^T \mathbf{z}_t + \tau_\epsilon^{-\frac{1}{2}} \tilde{\epsilon}_t$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + \tau_\eta^{-\frac{1}{2}} \tilde{\eta}_t$$

$$\delta_t = \delta_{t-1} + \tau_v^{-\frac{1}{2}} \tilde{v}_t$$

$$\gamma_t = \sum_{i=1}^{S-1} \gamma_{t+i-S} + \tau_w^{-\frac{1}{2}} \tilde{w}_t$$

$$\rho_t = \alpha \rho_{t-1} + \tau_u^{-\frac{1}{2}} \tilde{u}_t$$

$$\tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{v}_t, \tilde{w}_t, \tilde{u}_t \overset{iid}{\sim} \mathcal{N}(0, 1)$$

Priors

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{0}, \tau_b \mathbf{I}) \qquad \tau_b \quad hpm$$

$$\alpha \sim \mathcal{N}(a, \tau_a) \qquad a, \tau_a \ hpm$$

$$\tau_\epsilon \sim Unif(a_\epsilon, b_\epsilon) \qquad a_\epsilon, b_\epsilon \ hpm$$

$$\tau_\eta \sim Unif(a_\eta, b_\eta) \qquad a_\eta, b_\eta \ hpm$$

$$\tau_v \sim Unif(a_v, b_v) \qquad a_v, b_v \ hpm$$

$$\tau_w \sim Unif(a_w, b_w) \qquad a_w, b_w \ hpm$$

$$\tau_u \sim Unif(a_u, b_u) \qquad a_u, b_u \ hpm$$

Initial Conditions

$$\mu_0 \sim \mathcal{N}(m, \tau_\eta) \qquad m \ hpm$$

$$\delta_0 \sim \mathcal{N}(d, \tau_v) \qquad d \ hpm$$

$$\boldsymbol{\gamma}_{0:(2-S)} \sim \mathcal{N}_{S-1}(\mathbf{g}, \tau_w \mathbf{I}) \qquad \mathbf{g} \ hpm$$

$$\rho_0 \sim \mathcal{N}(r, \tau_u) \qquad r \ hpm$$

Independence

$$\{\tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{u}_t, \tilde{w}_t, \tilde{v}_t, \mu_0, \delta_0, \boldsymbol{\gamma}_{0:(-S+2)}, \rho_0,$$

$$\boldsymbol{\beta}, \alpha, \tau_\epsilon, \tau_\eta, \tau_v, \tau_w, \tau_u\}$$

$$family\ of\ \perp\!\!\!\perp\ real\ random\ variables$$

Hyperparameters for BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

We fix the hyperparameters as follows:

	τ_ϵ	τ_η	τ_υ	τ_w	τ_u
a	$1\text{e}{-7}$	$1\text{e}{-5}$	$1\text{e}{-5}$	$1\text{e}{-5}$	$1\text{e}{-5}$
b	$1\text{e}{-4}$	$1\text{e}{-4}$	$1\text{e}{-4}$	$1\text{e}{-4}$	$1\text{e}{-3}$

m and g according to frequentist stationarization

	τ_α	τ_β	r	d
value	$1\text{e}{-5}$	$1\text{e}{-5}$	0	0

4 rules for a proper evaluation

1. performance is subject to correctness of modelling assumptions and proper diagnostics,
2. use codified Beyesian tools to evaluate the worth of a model from a predictive viewpoint,
3. in absence of strong evidence for a format let interpretability be the deciding factor,
4. in case of a draw in the previous points, select the simplest paradigm.

Some diagnostics

Andrea De Gobbis, Lorenzo Ghilotti, Giorgio Meretti

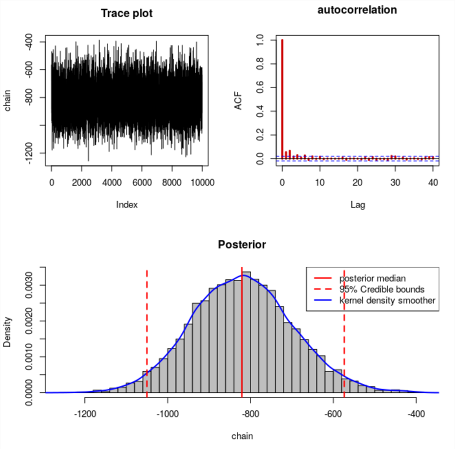


Figure 1: Diagnostics for β_1 samples

Prediction

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

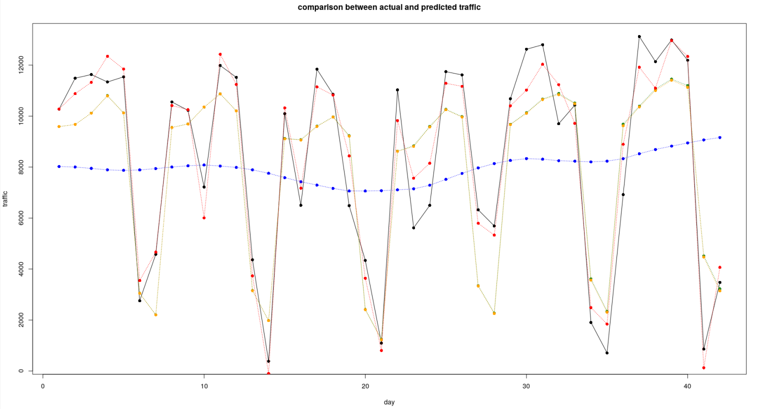


Figure 2: Sum of trend (blue), periodicity (green), autoregression (yellow) and regression (red)

Prediction

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

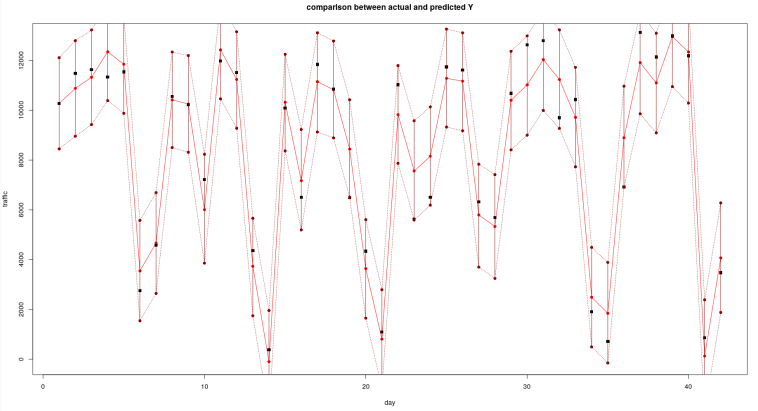


Figure 3: 90% credibility intervals in dark red, in black the true responses y

Without temperature

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

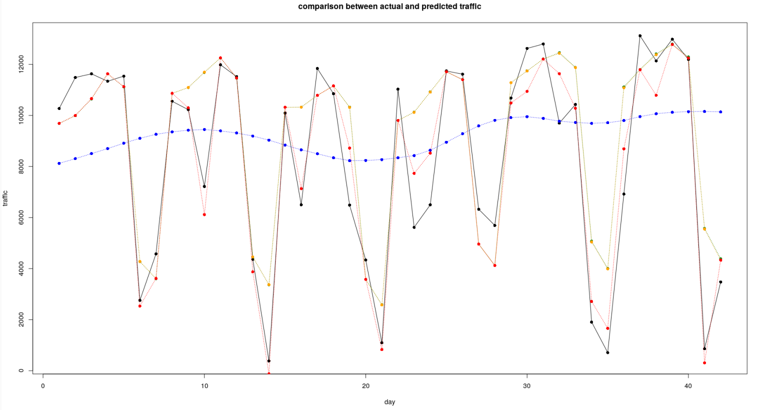


Figure 4: Sum of trend (blue), periodicity (green), autoregression (yellow) and regression (red)

Robust version

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

We apply the lag operator

$\mathcal{L}_t(\cdot) = \frac{1}{2}(\cdot)_{t-1} + \frac{1}{3}(\cdot)_{t-S} + \frac{1}{6}(\cdot)_{t-2S}$ to the trend.

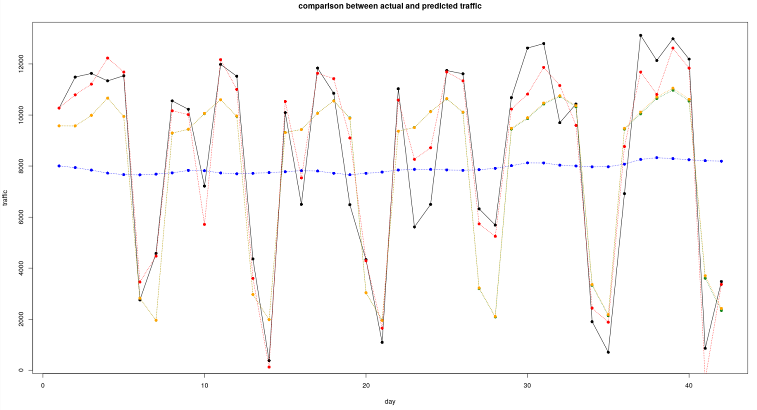


Figure 5: Cumulative action of trend (blue), periodicity (green), autoregression (yellow) and regression (red) in robust model

Evaluation of multiple BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

According to the criteria we have the following results:

	standard	no temperature	robust
Mean standardized residuals	0.5809041	0.6066627	0.6031869
Mean tail probabilities	0.2979543	0.298591	0.2908414
$elpd_{loo}$	-362.6	-365.4	-364.1

Future prediction

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

We try to predict one week ahead

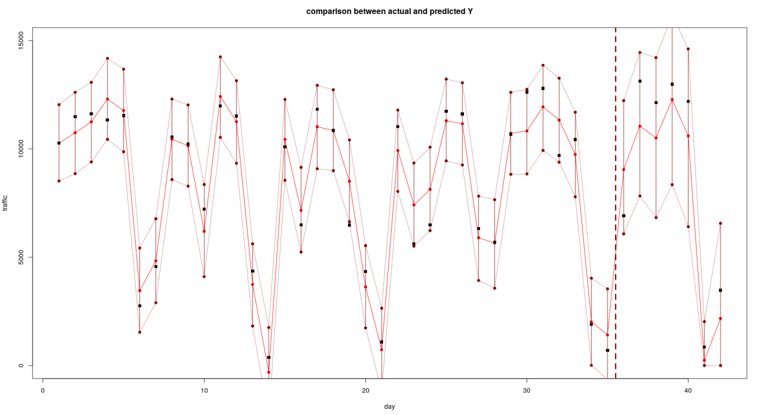


Figure 6: Predicted traffic with the model

Standard BSTS with time zones

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Model

$$Y_{F(t-1)+h} = \mu_t + \gamma_t + \chi_{F(t-1)+h} + $$
$$+ \boldsymbol{\beta}^T \mathbf{z}_{th} + \tau_\epsilon^{-\frac{1}{2}} \tilde{\epsilon}_{F(t-1)+h}$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + \tau_\eta^{-\frac{1}{2}} \tilde{\eta}_t$$

$$\delta_t = \delta_{t-1} + \tau_v^{-\frac{1}{2}} \tilde{v}_t$$

$$\gamma_t = \sum_{i=1}^{S-1} \gamma_{t+i-S} + \tau_w^{-\frac{1}{2}} \tilde{w}_t$$

$$\chi_{F(t-1)+h} = \sum_{j=1}^{F-1} \chi_{F(t-1)+h+j-F} + $$
$$+ \tau_u^{-\frac{1}{2}} \tilde{u}_{F(t-1)+h}$$

$$\tilde{\epsilon}_{F(t-1)+h}, \tilde{\eta}_t, \tilde{v}_t, \tilde{w}_t, \tilde{u}_{F(t-1)+h} \overset{iid}{\sim} \mathcal{N}(0,1)$$

Initial Conditions

$$\mu_0 \sim \mathcal{N}(m, \tau_\eta) \qquad m \;\; hpm$$
$$\delta_0 \sim \mathcal{N}(d, \tau_v) \qquad d \;\; hpm$$
$$\boldsymbol{\gamma}_{0:(2-S)} \sim \mathcal{N}_{S-1}(\mathbf{g}, \tau_w \mathbf{I}) \qquad \mathbf{g} \;\; hpm$$
$$\boldsymbol{\chi}_{0:(2-F)} \sim \mathcal{N}_{F-1}(\mathbf{c}, \tau_u \mathbf{I}) \qquad \mathbf{c} \;\; hpm$$

Priors

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{0}, \tau_b \mathbf{I}) \qquad \tau_b \;\; hpm$$
$$\tau_\epsilon \sim Unif(a_\epsilon, b_\epsilon) \qquad a_\epsilon, b_\epsilon \;\; hpm$$
$$\tau_\eta \sim Unif(a_\eta, b_\eta) \qquad a_\eta, b_\eta \;\; hpm$$
$$\tau_v \sim Unif(a_v, b_v) \qquad a_v, b_v \;\; hpm$$
$$\tau_w \sim Unif(a_w, b_w) \qquad a_w, b_w \;\; hpm$$
$$\tau_u \sim Unif(a_u, b_u) \qquad a_u, b_u \;\; hpm$$

Independence

$$\{\tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{w}_t, \tilde{v}_t, \mu_0, \delta_0, \boldsymbol{\gamma}_{0:(-S+2)},$$
$$\boldsymbol{\chi}_{0:(2-F)}, \boldsymbol{\beta}, \tau_\epsilon, \tau_\eta, \tau_v, \tau_w, \tau_u\}$$
$$family\ of\ \perp\!\!\!\perp\ real\ random\ variables$$

Division in time zones

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

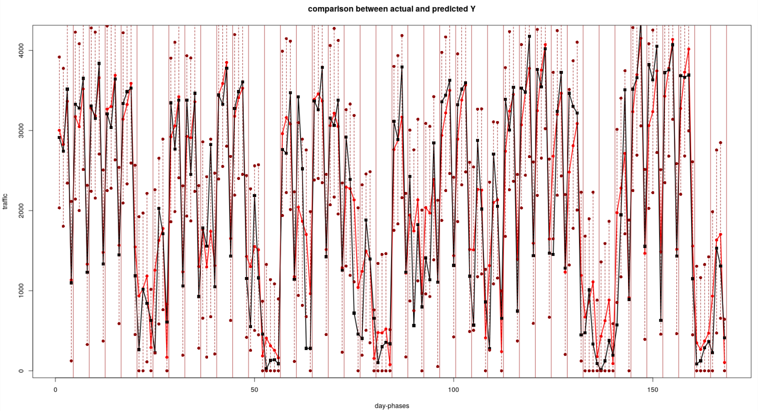


Figure 7: Global time series per slot

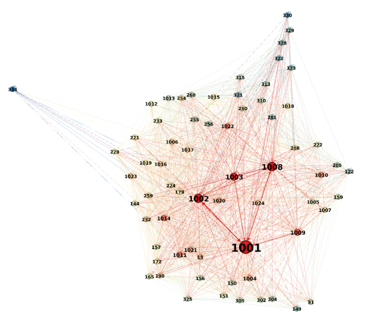
Network models

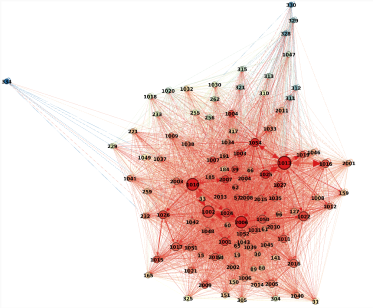
In the network models we deal with the flows of bicycles inward and outward a cluster of station, given a time interval.
The problem is the with the very high dimensionality of the data, and the computational weight due to the high number of parameters.
To shrink the model we first introduced a prior clusterization with a modified DBSCAN and later using simplifying hypothesis.

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

The euclidian distance isn't versatile enough

- We fed the classic DBSCAN a different semidistance, $smd(x, y) = \frac{p(y)+p(x)}{2}||x - y||$.
- $p(x) = g(||x - x_0||)$ where x_0 are the coordinates of the Duomo, and $g(u) = \max\{\beta, e^{(\frac{u}{\alpha})^2 \log(1-\gamma)} + \gamma\}$
- We fitted the parameters in a way to minimize the autorings, and to break down the biggest clusters


1001
1002
1003
1008



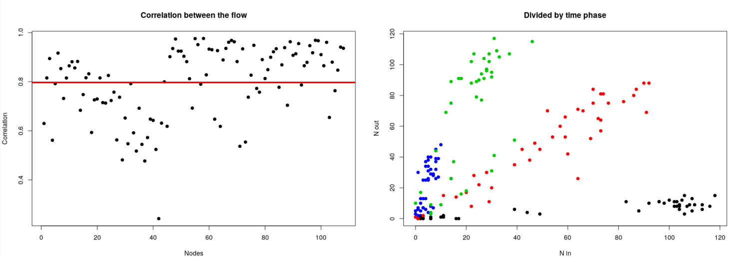
Preprocessing

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

```
> ins[1:8,1:8]
  15 19 30 31 33 39 46 54
1  1 17 11  1  2 27 14  5
2 10 11 18  7  2 32 12  8
3 23 13 14 12  1 71  9 17
4  7  4  7  5  0 13  8  4
5  8 26 22  2  5 31 17  8
6 18  9 17  8  8 53 13  5
7 21 14 24 11  1 77 10 11
8  4  8  6 10  1 13  8 11
```

The two flows are structured as 168 x 109 matrices where the columns are the nodes and the rows are the evolving time

Correlations between the flows in the same node



Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

The likelihood is a Bivariate Poisson, a distribution over $\mathbb{N} \times \mathbb{N}$.

$$\begin{cases} Y_j = X_0 + X_j \quad j = 1, 2 \\ X_j \sim \text{Po}(\lambda_j) \quad j = 0, 1, 2 \end{cases} \tag{1}$$

In this way $\text{Cov}(Y_1, Y_2) = \text{Var}(X_0) = \lambda_0$. And marginally $Y_j \sim \text{Po}(\lambda_0 + \lambda_j)$.

Some simplifying hypothesis

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Two way mixed effect

$$\beta_{ith} \cdot \mathbf{x} \Rightarrow (\alpha_t + \gamma_h) \cdot \mathbf{x}$$

Offsets

$$\log(\frac{\lambda_{ith}}{O_i}) = (\alpha_t + \gamma_h) \cdot \mathbf{x}$$

$$\Rightarrow \log(\lambda_{ith}) = \log(O_i) + (\alpha_t + \gamma_h) \cdot \mathbf{x}$$

The offsets model

$$\begin{cases} \log(\lambda_{ith}^{(j)}) = (\theta^{(j)} + \alpha_t^{(j)} + \gamma_h^{(j)}) \cdot \mathbf{x} + \log(O_i) \\ \theta^{(j)} | a, B \overset{iid}{\sim} \mathcal{N}(a, B) \\ \alpha_t^{(j)} | \Sigma \overset{iid}{\sim} \mathcal{N}(0, \Sigma) \\ \gamma_h^{(j)} | \Omega \overset{iid}{\sim} \mathcal{N}(0, \Omega) \\ a \sim \mathcal{N}(a_0, A_0) \\ B \sim invWishart(B_0, p+1) \\ \Sigma, \Omega \overset{iid}{\sim} invWishart(C_0, p+1) \end{cases} \tag{2}$$

Where $\alpha_t^{(j)}, \beta_h^{(j)}, \theta^{(j)}, a \in \mathbb{R}^3$, $\Sigma, \Omega, B \in \mathbb{R}^{3\times3}$ are the parameters, $A_0, B_0, C_0 \in \mathbb{R}^{3\times3}, \quad a_0 \in \mathbb{R}^3$ are the hyperparameters, and x are the covariates.
The offsets are the the average of the traffic of the first 12 days.

The parametric offsets (PO) model

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

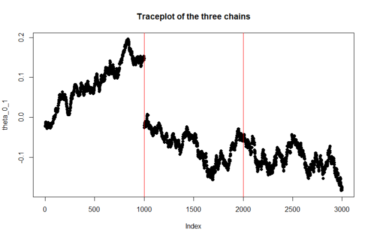
The offsets are now grouped with the other parameters.

$$\begin{cases} \log(\lambda_{ith}^{(j)}) = (\theta^{(j)} + \alpha_t^{(j)} + \beta_h^{(j)}) \cdot \mathbf{x} + \Phi_i \\ \theta^{(j)} | a, B \overset{iid}{\sim} \mathcal{N}(a, B) \\ \alpha_t^{(j)} | \Sigma \overset{iid}{\sim} \mathcal{N}(0, \Sigma) \\ \beta_h^{(j)} | \Omega \overset{iid}{\sim} \mathcal{N}(0, \Omega) \\ \Phi_i \overset{iid}{\sim} \mathcal{N}(\Phi_0, F_0) \\ a \sim \mathcal{N}(a_0, A_0) \\ B \sim invWishart(B_0, p+1) \\ \Sigma, \Omega \overset{iid}{\sim} invWishart(C_0, p+1) \end{cases} \tag{3}$$

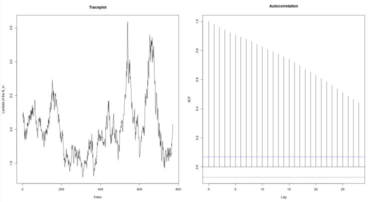
Where $\alpha_t^{(j)}, \beta_h^{(j)}, \theta^{(j)}, a \in \mathbb{R}^3$, $\Sigma, \Omega, B \in \mathbb{R}^{3\times 3}$, $\Phi_i \in \mathbb{R}$ are the parameters, $A_0, B_0, C_0, F_0 \in \mathbb{R}^{3\times 3}, \quad a_0 \in \mathbb{R}^3, \Phi_0 \in \mathbb{R}$ are the hyperparameters, and $\mathbf{x}$ are the covariates.

Posterior with JAGS

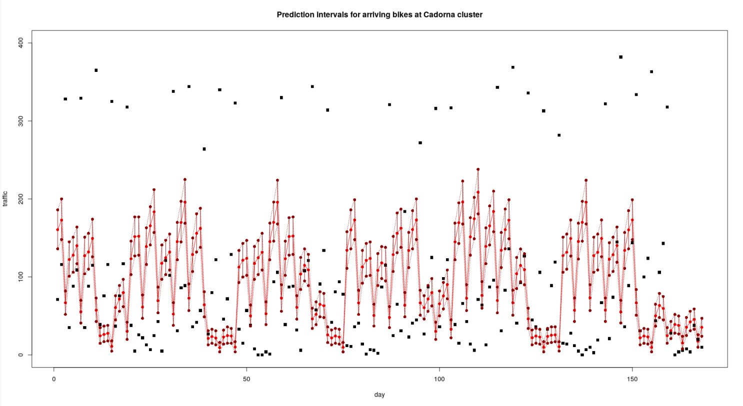
Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti



	N^{IN}	N^{OUT}
Offsets	41.0%	62.2%
PO	39.7%	38.6%



The results are unsatisfactory, the iterations are not enough to stabilize the MCMC and the autocorrelation is too high. The predictive power is low.


Prediction intervals for arriving bikes at Cadorna cluster
traffic
day
0
100
200
300
400
0
50
100
150