A Bayesian model for data flow: BikeMi

Andrea De Gobbis, Lorenzo Ghilotti, Giorgio Meretti

January 8, 2020

Politecnico di Milano

What we are doing

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

The BikeMi stations net in Milan

Two prospective of the problem

We followed two distinct paths:

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

Network model: dividing in the different nodes and analysing the flow of bikes in the net. The dimensionality is much higher.

Poisson model

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Day by day Poisson:

$$\begin{cases} Y_t \sim \mathrm{Po}(Y_t|\lambda_t) \\ \lambda_t = \exp\{\alpha + \boldsymbol{\beta} \cdot \mathbf{x}_t\} \\ \alpha \sim \mathcal{N}(0, \sigma_\alpha^2) \\ \beta_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\beta^2) \end{cases} \tag{1}$$

Poisson model

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

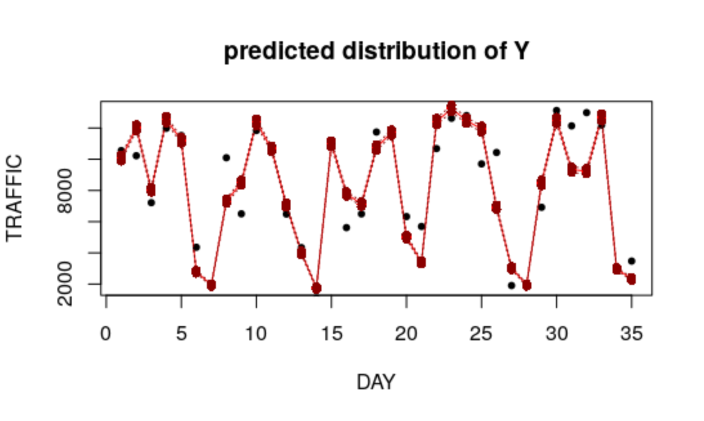
With covariates $\mathbf{x}_t$:

- Y_{t-1} volume on the previous day
- Y_{t-7} volume on the same weekday of the previous week
- W_t dummy for weekday / weekend
- R_t, R_{t-1} dummies for rain in the current and previous day
- T_t mean temperature for the day
- S_t, M_t dummies for Saturday and Monday

Predictive distribution of the poisson model

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Only 3 of 35 in the 90% credible interval



Time series model: BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

$$\begin{cases} \mathbf{X}(t) = \mathbf{f}(\mathbf{X}(1:(t-1))) + \boldsymbol{\epsilon}_1(t) \\ \mathbf{Y}(t) = \mathbf{g}(\mathbf{X}(t)) + \boldsymbol{\epsilon}_2(t) \end{cases} \tag{2}$$

with suitable initial conditions and priors

Time series model: BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Elementary model with precisions

$$\begin{cases} Y_t = \mu_t + \gamma_t + \frac{1}{\sqrt{\tau_\epsilon}}\tilde{\epsilon}_t \\ \\ \mu_t = \mu_{t-1} + \delta_{t-1} + \frac{1}{\sqrt{\tau_\eta}}\tilde{\eta}_t \\ \delta_t = \delta_{t-1} + \frac{1}{\sqrt{\tau_v}}\tilde{v}_t \\ \gamma_t = \sum_{i=1}^{S-1} \gamma_{t+i-S} + \frac{1}{\sqrt{\tau_w}}\tilde{w}_t \\ \\ \tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{v}_t, \tilde{w}_t, \overset{iid}{\sim} \mathcal{N}(0,1) \end{cases} \tag{3}$$

Time series model: BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Elementary model priors

$$\begin{cases} \mu_0 \sim \mathcal{N}(m, \tau_m) \\ \delta_0 \sim \mathcal{N}(d, \tau_d) \\ \boldsymbol{\gamma}_{0:(2-S)} \sim \mathcal{N}_{S-1}(\mathbf{g}, \tau_g \mathbf{I}) \\ \\ \tau_* \sim Gamma(a_*, b_*),\ with \ast = \{\epsilon, \eta, v, w\} \\ \\ \{\tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{v}_t, \tilde{w}_t, \mu_0, \delta_0, \boldsymbol{\gamma}_{0:(-S+2)}, \tau_\epsilon, \tau_\eta, \tau_v, \tau_w\} \\ independent. \end{cases} \tag{4}$$

Time series model: BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Complete model with precisions

$$\begin{cases} Y_t = \mu_t + \gamma_t + \rho_t + \boldsymbol{\beta}^T \mathbf{z}_t + \frac{1}{\sqrt{\tau_\epsilon}} \tilde{\epsilon}_t \\ \\ \mu_t = \mu_{t-1} + \delta_{t-1} + \frac{1}{\sqrt{\tau_\eta}} \tilde{\eta}_t \\ \delta_t = \delta_{t-1} + \frac{1}{\sqrt{\tau_v}} \tilde{v}_t \\ \gamma_t = \sum_{i=1}^{S-1} \gamma_{t+i-S} + \frac{1}{\sqrt{\tau_w}} \tilde{w}_t \\ \rho_t = \alpha \rho_{t-1} + \frac{1}{\sqrt{\tau_u}} \tilde{u}_t \\ \\ \tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{v}_t, \tilde{w}_t, \tilde{u}_t \overset{iid}{\sim} \mathcal{N}(0,1) \end{cases} \tag{5}$$

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Complete model priors

$$
\begin{cases}
\mu_0 \sim \mathcal{N}(m, \tau_m) \\
\delta_0 \sim \mathcal{N}(d, \tau_d) \\
\boldsymbol{\gamma}_{0:(2-S)} \sim \mathcal{N}_{S-1}(\mathbf{g}, \tau_g \mathbf{I}) \\
\rho_0 \sim \mathcal{N}(r, \tau_r) \\
\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{0}, \tau_b \mathbf{I}) \\
\alpha \sim \mathcal{N}(a, \tau_a) \\
\tau_* \sim Gamma(a_*, b_*), \ with * = \{\epsilon, \eta, v, w, u\} \\
\{\tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{u}_t, \tilde{w}_t, \tilde{v}_t, \mu_0, \delta_0, \boldsymbol{\gamma}_{0:(-S+2)}, \rho_0 \, \boldsymbol{\beta}, \alpha, \tau_\epsilon, \tau_\eta, \tau_v, \tau_w, \tau_u\} \\
independent.
\end{cases} \tag{6}
$$

Time series model: BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Robust model with precisions

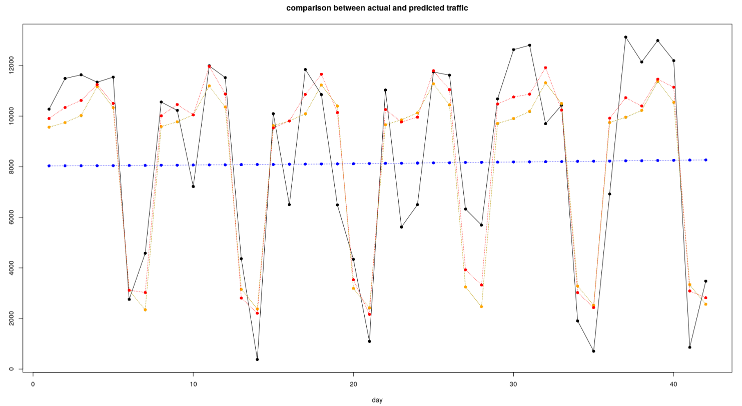
$$
\begin{cases}
Y_t = \mu_t + \gamma_t + \rho_t + \boldsymbol{\beta}^T \mathbf{z}_t + \frac{1}{\sqrt{\tau_\epsilon}}\tilde{\epsilon}_t \\
\\
\mu_t = avgpred(\mu_t) + avgpred(\delta_t) + \frac{1}{\sqrt{\tau_\eta}}\tilde{\eta}_t \\
\delta_t = avgpred(\delta_t) + \frac{1}{\sqrt{\tau_v}}\tilde{v}_t \\
\gamma_t = \sum_{i=1}^{S-1} \gamma_{t+i-S} + \frac{1}{\sqrt{\tau_w}}\tilde{w}_t \\
\rho_t = \alpha\rho_{t-1} + \frac{1}{\sqrt{\tau_u}}\tilde{u}_t \\
\\
\tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{v}_t, \tilde{w}_t, \tilde{u}_t \overset{iid}{\sim} \mathcal{N}(0,1)
\end{cases}
\tag{7}
$$

where $avgpred(\phi_t) = \frac{1}{2}\phi_{t-1} + \frac{1}{3}\phi_{t-S} + \frac{1}{6}\phi_{t-2S}$
and same priors as before

Time series model: BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

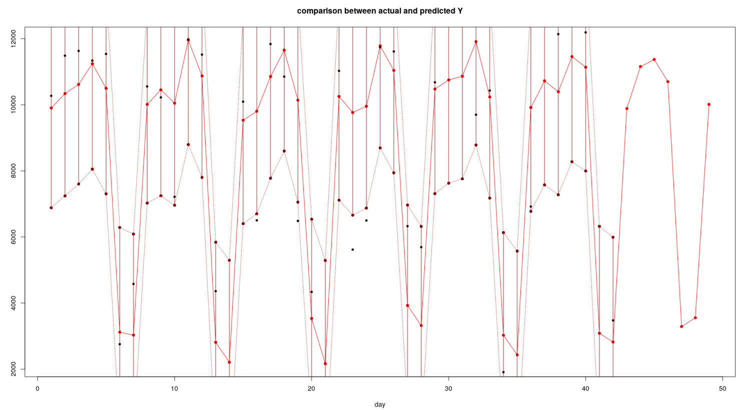
Instability of the errors, τ_ϵ explodes, bad autocorrelation



Time series model: BSTS

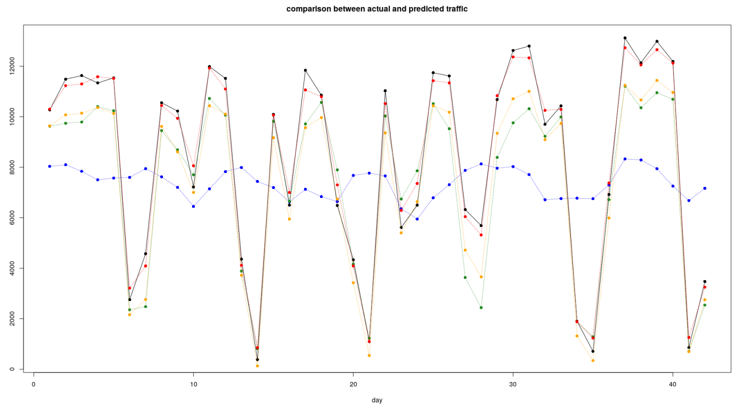
Andrea De Gobbis, Lorenzo Ghilotti, Giorgio Meretti

Instability of the errors, τ_ϵ explodes, bad autocorrelation



Time series model: BSTS

Blocking the maximum variance, τ_* under control, partial mixing of state variables

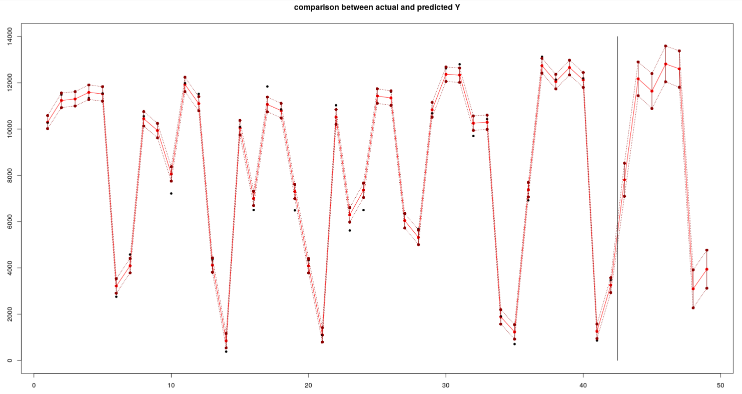


Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Time series model: BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Smaller induced variability, attempt of prediction with real weather.



Time series model: BSTS

Comparison truncated-variance robust BSTS vs Poisson.

Observations inside credibility intervals:

- Poisson 3/35
- BSTS 24/42

WAIC:

- Poisson -0.0171
- BSTS -0.0160

Time series model: BSTS

Andrea De Gobbis, Lorenzo Ghilotti, Giorgio Meretti

Robust model for time zones with precisions

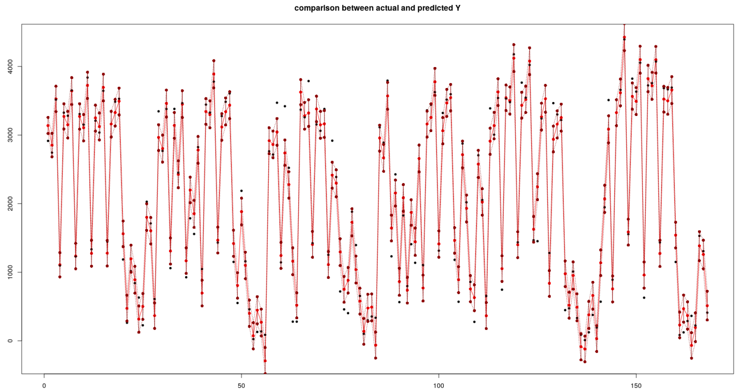
$$
\begin{cases}
Y_{th} = \mu_t + \gamma_t + \rho_t + \chi_k + \boldsymbol{\beta}^T \mathbf{z}_t + \frac{1}{\sqrt{\tau_\epsilon}} \tilde{\epsilon}_{th} \\
\\
\mu_t = avgpred(\mu_t) + avgpred(\delta_t) + \frac{1}{\sqrt{\tau_\eta}} \tilde{\eta}_t \\
\delta_t = avgpred(\delta_t) + \frac{1}{\sqrt{\tau_v}} \tilde{v}_t \\
\gamma_t = \sum_{i=1}^{S-1} \gamma_{t+i-S} + \frac{1}{\sqrt{\tau_w}} \tilde{w}_t \\
\rho_t = \alpha \rho_{t-1} + \frac{1}{\sqrt{\tau_u}} \tilde{u}_t \\
\chi_k = \sum_{i=1}^{F-1} \chi_{k+i-F} + \xi \delta_{t(k)=6,7} + \frac{1}{\sqrt{\tau_\zeta}} \tilde{\zeta} \\
\\
\tilde{\epsilon}_t, \tilde{\eta}_t, \tilde{v}_t, \tilde{w}_t, \tilde{u}_t, \tilde{\zeta} \overset{iid}{\sim} \mathcal{N}(0,1)
\end{cases}
\tag{8}
$$

where $avgpred(\phi_t) = \frac{1}{2}\phi_{t-1} + \frac{1}{3}\phi_{t-S} + \frac{1}{6}\phi_{t-2S}$, $h = mod(k,4)+1$ and priors of the same class as before

Time series model: BSTS

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Robust truncated variance model for time zones



Two prospective of the problem

We followed two distinct paths:

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

Network model: dividing in the different nodes and analysing the flow of bikes in the net. The dimensionality is much higher.

Network model

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

For every (i, j) edge of the graph and $t \in 1 : 42$ we have $Y_{ij}(t)$ the number of travels from node i to j at day t.

Problem:

more than 4 million variables $\Rightarrow$

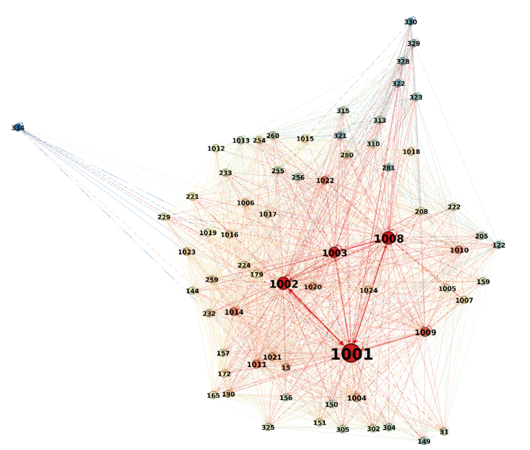
Computationally untreatable

Solutions:

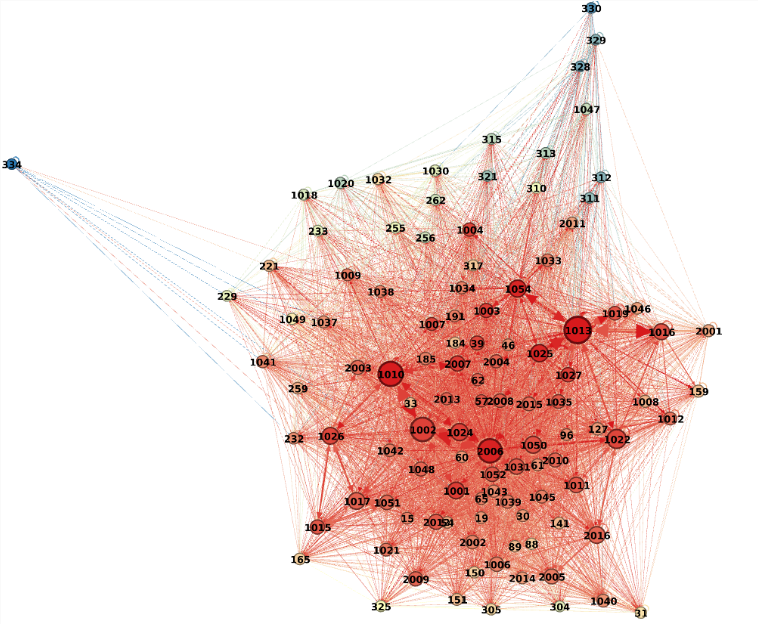
- clusterization through DBSCAN
- simplification of the variables

Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Algorithm to divide the nodes into initial clusters. We introduced a modified version with a weight to break up the bigger clusters, minimizing the autorings presence.



From 334 nodes to 140.



Andrea De Gobbis,
Lorenzo Ghilotti,
Giorgio Meretti

Modeling the flux

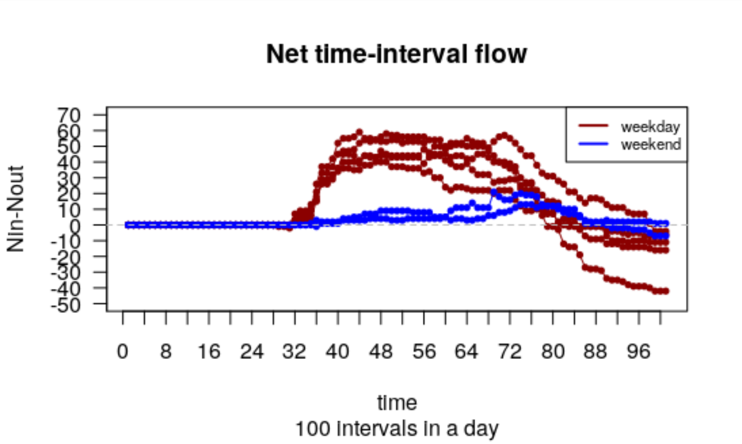
To further simplify the analysis we focus on the bikes arriving and departing from each node $N_i^{IN}(\Delta t)$ and $N_i^{OUT}(\Delta t)$ in the time interval Δt.

We are interested in the smallest Δt as possible but this would increase the number of variables $\Rightarrow$ consider them as functional data.

$$V_i(t) = \lim_{\Delta t \to 0} \frac{N_i^{IN}(\Delta t) - N_i^{OUT}(\Delta t)}{\Delta t}$$

$$\Phi_i(t) = \int_0^t V_i(u)\,\mathrm{d}u$$

We can analyse when new bikes should be brought to which station.



Andrea De Gobbis, Lorenzo Ghilotti, Giorgio Meretti