Probabilistic forecasts for bike-sharing systems

N. Gast (Inria)
joint work with G. Massonnet (Inria / Mines Nantes), D. Reijsbergen (U. Edinburgh), M. Tribastone (IMT Lucca)

May 24, 2017

Bike-sharing systems (BSS)

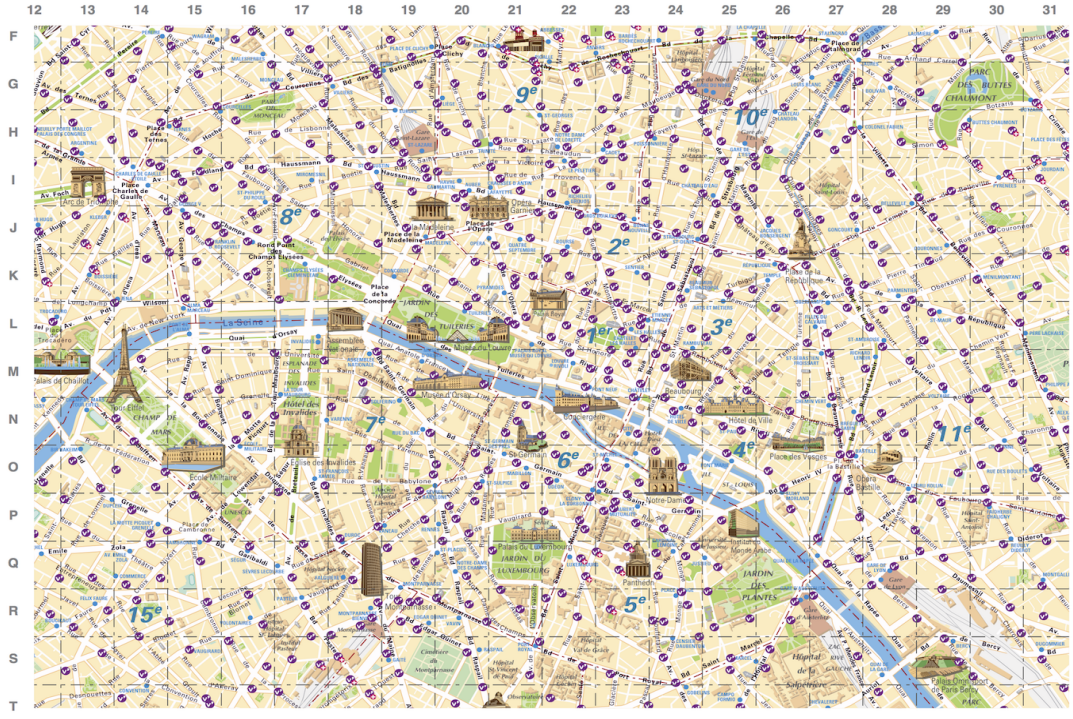
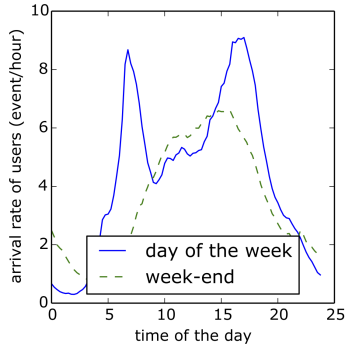


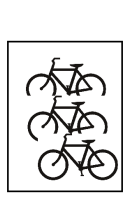
Figure: Vélib' stations in the centre of Paris

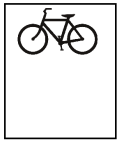
- Each station has a given number of parking slots.
- Users enter the system by picking up a bike at a station and making a trip to another station, where they drop the bike on an available parking spot.

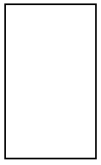
Forecasting : what and why?

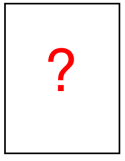
What is forecasting :

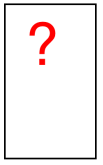


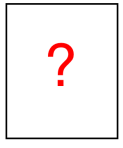


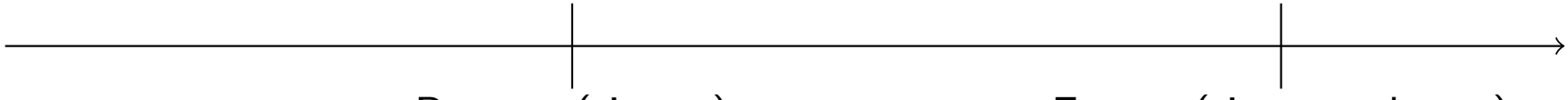












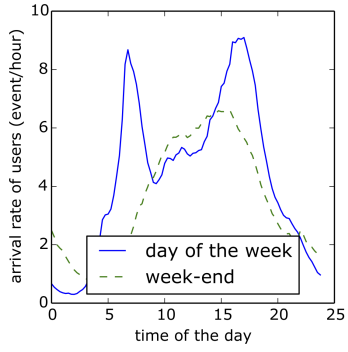
Historical data

Present (time t)

Future (time t+$x$hours)

Forecasting : what and why?

What is forecasting :



Why forecasting :

- Operator perspective (rebalancing)
- User perspective (will I find a bike?) This talk

Outline

Vélib' data

Vélib' Data (Paris) : availability at stations + trips info from September 2013 to December 2014

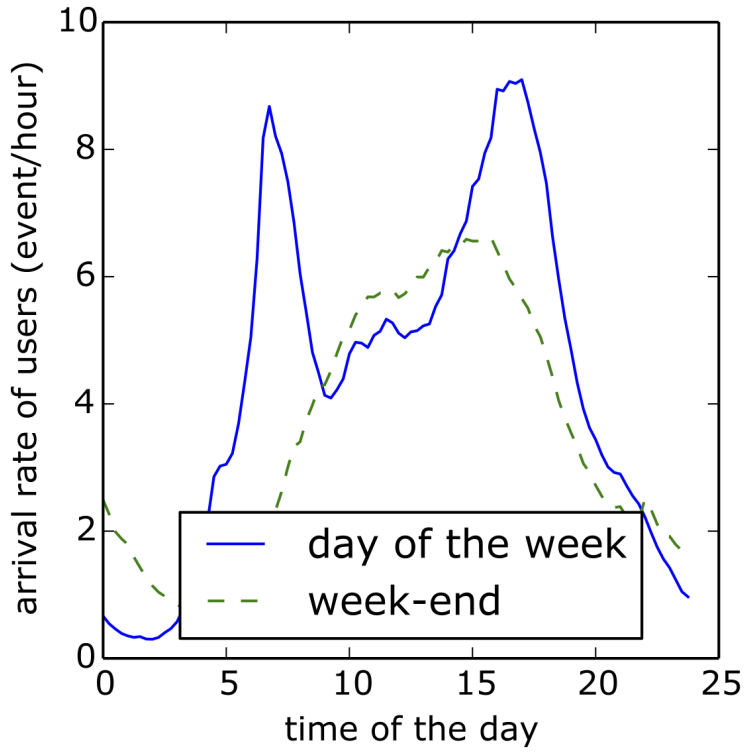


Figure: Evolution of the average departure rate from *Vélib'* stations during the day

First attempt: deterministic forecasts

Output of forecasting method : $X_t(t+h)$.

Examples

- **Last-Value (LVP)** : availability at $t+h$ is equal to the availability at t.
- **Historical (HP)** : distribution of the bikes availability at $t+h$, based on historical observations.
- **Machine learning tools** (ARIMA, Bayesian network,...)

Is this good enough?

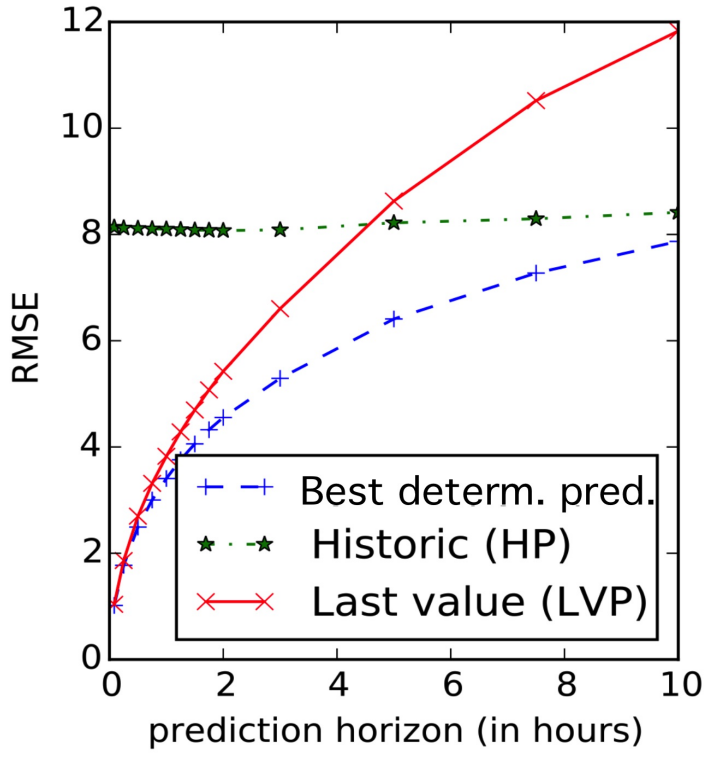


Figure: Comparison of the RMSEs for different predictors.

Average error :

- 3 bikes for $h = 30$ min
- 5 bikes for $h = 2$h.

Is this relevant for users?

(a) Empty station

(b) Full station

Figure: Shortage and blocking stations

Predict blocking situtations

- Warn the users
- Rebalance the system
- Improve traffic flow

→ Need for good forecasting tools

Outline

Output of forecasting method :

$$\textbf{Prob}\left(X_t(t+h) = k \middle| \text{present and historical data} \right),$$

for $k \in \{0, \ldots, \text{capacity}\}$.

- How to evaluate a probabilistic predictor?
  - Scoring rule
  - False positive / false negative

Queueing network representation

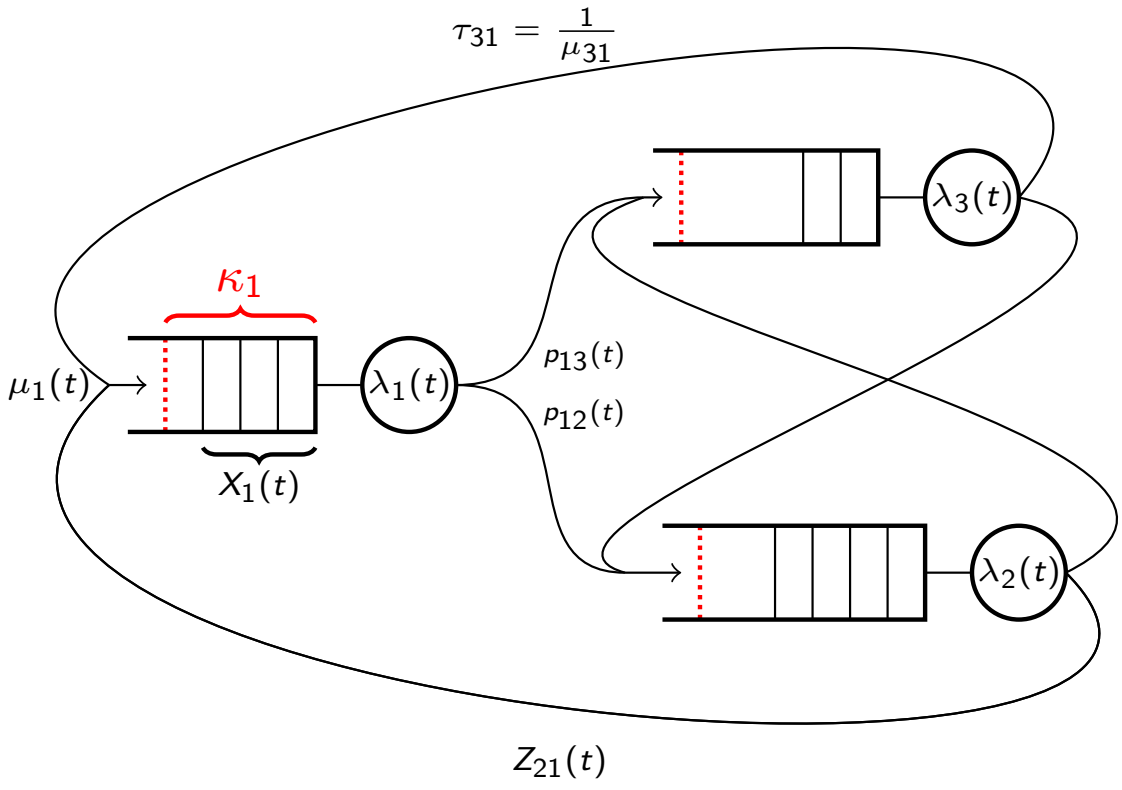
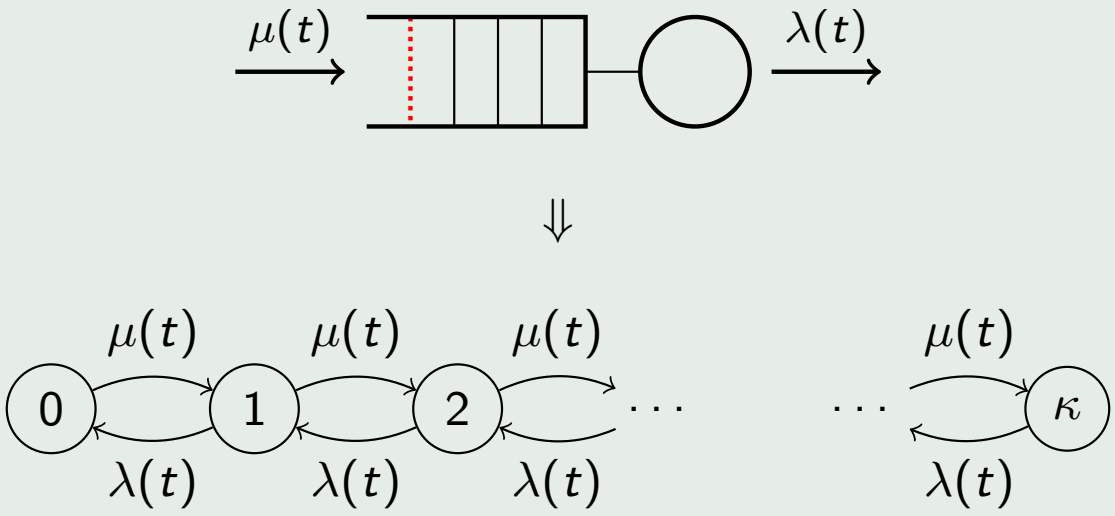


Figure: A BSS network with 3 stations

- Moment-based Probabilistic Prediction of Bike Availability for Bike-Sharing Systems. / Feng, Cheng; Hillston, Jane; Reijsbergen, Daniel. (QEST 2016)
- Probabilistic Forecasts of Bike-Sharing Systems for Journey Planning. Nicolas Gast; Guillaume Massonnet; Daniel Reijsbergen; Mirco Tribastone (CIKM 2015)

Mean-field analysis → independence of the stations



Probabilistic predictor

$$p(j|i,t,h) = \exp\left(\int_0^h Q(t+s)ds\right)_{i,j}$$

where $Q(t)$ is the kernel of the Markov chain at time t.

Application to *Vélib'* Data

We consider four trip recommendation predictors :

- Queueing model
- Historic predictor
- Last-value
- Always-go

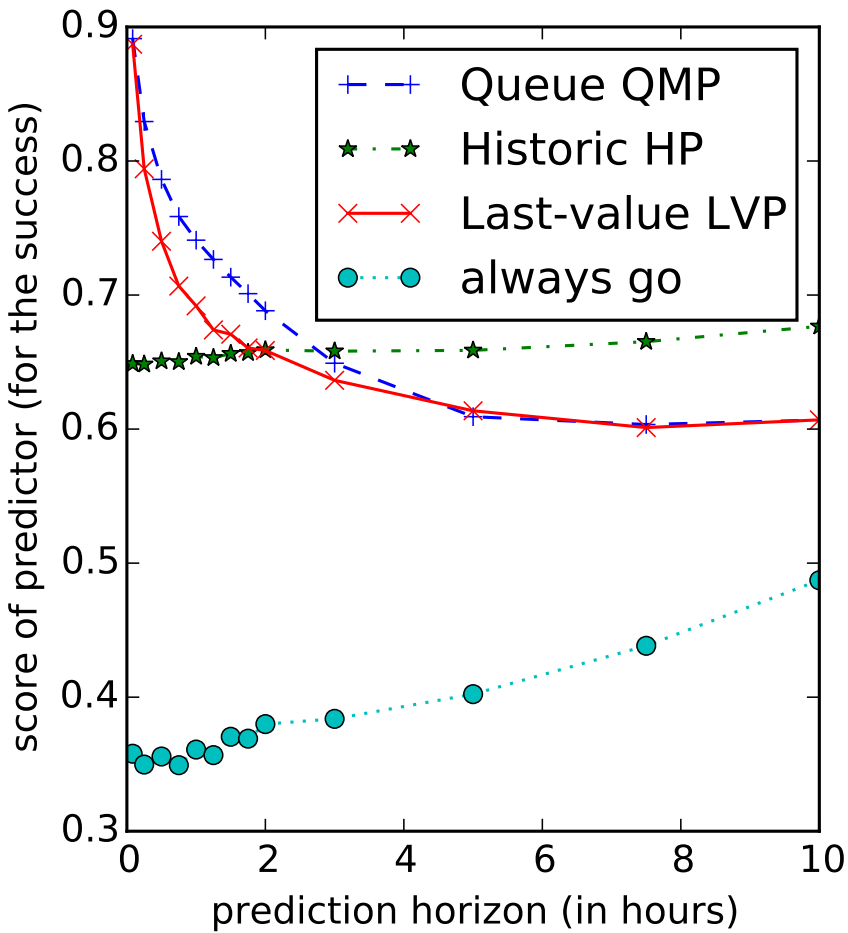
Metric: Scoring rule GO/NO-GO

- 1 if prediction correct
- 0 if trip could have been made but was not recommended
- -5 if trip could not been made but has been recommended

Application to *Vélib'* Data

Metric: Scoring rule GO/NO-GO

- 1 if prediction correct
- 0 if trip could have been made but was not recommended
- -5 if trip could not been made but has been recommended



Outline

What's important?

- Deterministic predictors : not well-suited for users.
- Stochasticity of the system must be included in forecasting tools.
- Observation of the current state : useful for horizon of 2 to 5 hours but not really useful after.