

Statistical Analysis of a Bike Sharing Transportation System

J. Raimbault^{1,2}

¹Erasmus Mundus Master in Complex Systems Science, Ecole Polytechnique

²LVMT, Ecole Nationale des Ponts et Chaussées

Therapeutic Evaluation and Complex Systems

Oral Presentation

December 18, 2013

Outline

- 1 Introduction
- 2 Data collection
- 3 Statistical analysis

Bike-sharing systems

- New flexible and ecological transportation system ([DeMaio, 2009]) ? Complementarity to other urban transportation modes ([Midgley, 2009])
- Understanding the mechanisms is necessary for its good management (ex optimizing redistribution process) but also has intrinsic value (urban life patterns)
- Many top-down approaches: statistical models ([Borgnat et al., 2009b, Borgnat et al., 2009a],[Michau et al., 2011]) or data-mining analysis ([O'Brien et al., 2013],[Vogel et al., 2011, Kaltenbrunner et al., 2010])

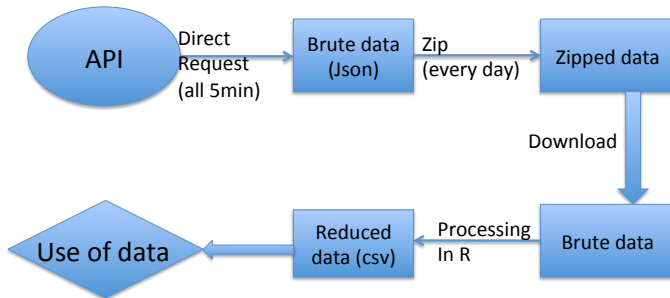
Objectives of the project

- Study of Paris' system (Vlib) [Nair et al., 2013]
- Statistical analysis of a large set of real data, general and with specific purposes (e. g. parametrisation of an ABM), using existing or new methods
- At the beginning, apply TE evaluation to relocation system but many issues.

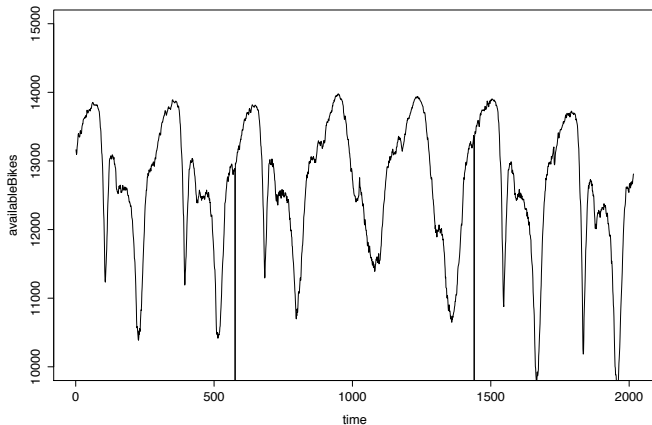
Type and origin of data

- Public data provided by the operator in real time. Problem: need a constantly running collection data process, and only docking station status (limited information).
- Why not ask full travel data to operator ? Independant and open research ([Banos, 2013]), reporting bias (in [Nair et al., 2013] results are not presented complete because company did not want for commercial reasons). We do a compromise, and see if we can however have good results.
- Also risk of unconscious spin in the description of results [Boutron et al., 2010]

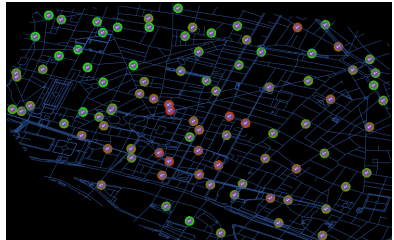
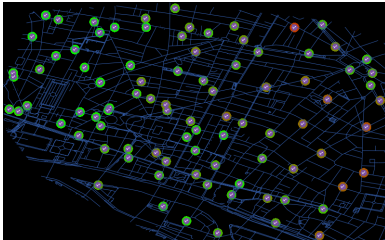
Data collection process



Visualisation: mobility patterns



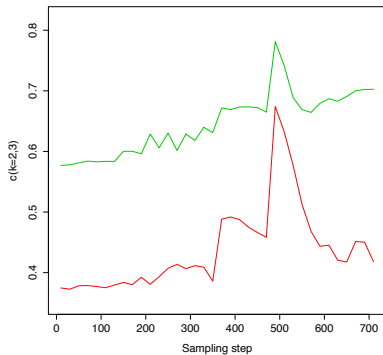
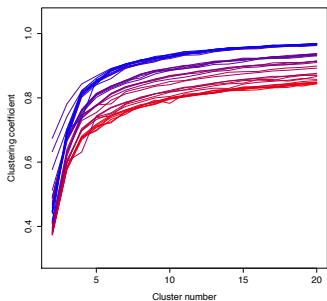
Visualisation: heatmaps



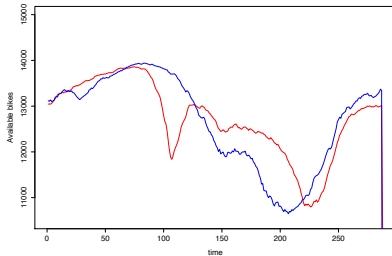
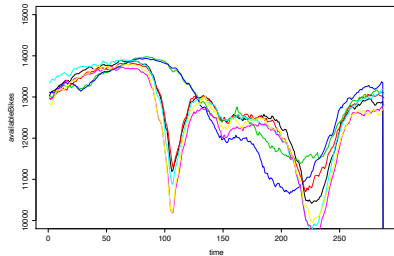
Extraction of patterns: clustering time-series

- Aim: extract typical use patterns (more characteristic: difference week/weekends)
- First sampling of time series, then kmeans ([Warren Liao, 2005]) on sampled series of all stations for a day: gives a reduced representation of each day
- Clustering on days to isolate patterns

Clustering process: role of sampling



Clustering: results

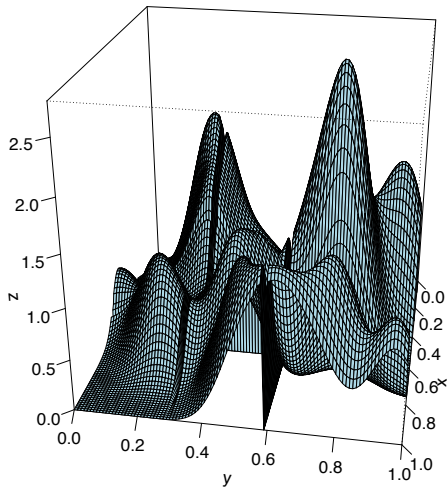


Inference of Origin/destinations in urban mobility

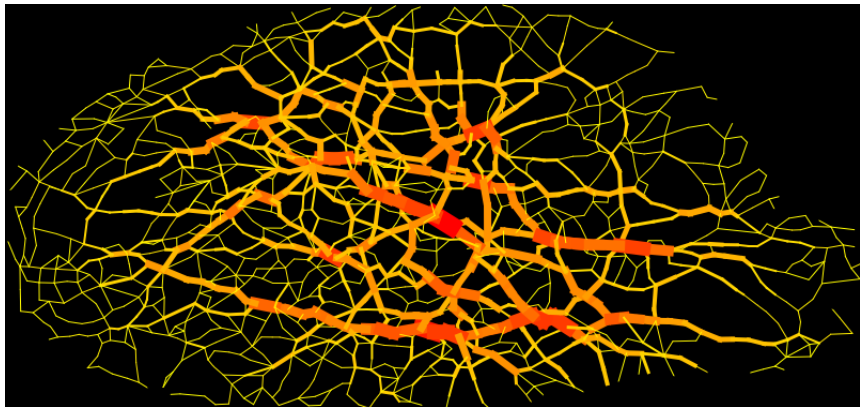
- Core of the parametrisation of the ABM, but also problem with its intrinsic value ([Leurent, 2006])
- Gaussian kernels non-parametric estimation ([Tsybakov, 2004]) with package kernlab ([Karatzoglou et al., 2004]). With $(d_i(t))$ real arrivals at $(\vec{x}_i(t))$, $D(t)$ spatial field is given by

$$[D(t)](\vec{x}) = \frac{1}{K} \sum_i d_i(t) \cdot \exp\left(\frac{\|\vec{x} - \vec{x}_i\|}{2\sigma^2}\right)$$

Example



Mapping cumulated flows



Use of TE methods ?

- Idea: evaluate effect of redistribution procedure
- Docking stations are individuals, a treatment is a given area (day with redistribution gives treated, without redistribution is control (but for a similar day; use of clustering ?)). Then do a meta-analysis on all areas.
- Problem: not even implementable; problem of finding redistributed area, size of areas, etc

Conclusion

- Unfortunately did not go so far as expected. However good results and powerful parametrisation for the ABM
- We can argue that these data were “enough” but still claim for a broader opening (since yesterday: www.data.gouv.fr !)

References I



Banos, A. (Décembre 2013).

Pour des pratiques de modélisation et de simulation libérées en Géographie et SHS.

PhD thesis, UMR CNRS Géographie-Cités, ISCIPIF.



Borgnat, P., Abry, P., and Flandrin, P. (2009a).

Modélisation statistique cyclique des locations de vélo'v à lyon.

In *XXI^e colloque GRETSI (traitement du signal et des images), Dijon (FRA), 8-11 septembre 2009*. GRETSI, Groupe d'Etudes du Traitement du Signal et des Images.






Borgnat, P., Abry, P., Flandrin, P., Rouquier, J.-B., et al. (2009b).




Studying lyon's vélo'v: a statistical cyclic model.

In *European Conference on Complex Systems 2009*.

References II

-  Boutron, I., Dutton, S., Ravaud, P., and Altman, D. G. (2010). Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA: the journal of the American Medical Association*, 303(20):2058–2064.
-  DeMaio, P. (2009). Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation*, 12(4):41–56.
-  Kaltenbrunner, A., Meza, R., Grivolla, J., Codina, J., and Banchs, R. (2010). Urban cycles and mobility patterns: Exploring and predicting trends in a bicycle-based public transport system. *Pervasive and Mobile Computing*, 6(4):455–466.

References III

-  Karatzoglou, A., Smola, A., Hornik, K., and Zeileis, A. (2004). kernlab-an s4 package for kernel methods in r.
-  Leurent, F. (2006).
Modélisation du trafic, des déplacements sur un réseau et de l'accessibilité aux activités grâce au transport.
PhD thesis, Université Paris Dauphine-Paris IX.
-  Michau, G., Robardet, C., Merchez, L., Jensen, P., Abry, P., Flandrin, P., and Borgnat, P. (2011).
Peut-on attraper les utilisateurs de vélo'v au lasso.
In *Proceedings of the 23e Colloque sur le Traitement du Signal et des Images. GRETSI-201*, pages 46–50.

References IV



Midgley, P. (2009).

The role of smart bike-sharing systems in urban mobility.
JOURNEYS, 2:23–31.



Nair, R., Miller-Hooks, E., Hampshire, R. C., and Bušić, A. (2013).

Large-scale vehicle sharing systems: Analysis of vélib'. *International Journal of Sustainable Transportation*, 7(1):85–106.



O'Brien, O., Cheshire, J., and Batty, M. (2013).

Mining bicycle sharing data for generating insights into sustainable transport systems.
Journal of Transport Geography.

References V



Tsybakov, A. B. (2004).

Introduction to nonparametric estimation. (introduction à l'estimation non-paramétrique.).



Vogel, P., Greiser, T., and Mattfeld, D. C. (2011).

Understanding bike-sharing systems using data mining:
Exploring activity patterns.

Procedia-Social and Behavioral Sciences, 20:514–523.



Warren Liao, T. (2005).

Clustering of time series data—a survey.

Pattern Recognition, 38(11):1857–1874.

Questions

?