

Le modèle DFM pour une analyse comparative des systèmes de vélos en libre service

Charles Bouveyron, Etienne Côme, Julien Jacques

▶ **To cite this version:**

Charles Bouveyron, Etienne Côme, Julien Jacques. Le modèle DFM pour une analyse comparative des systèmes de vélos en libre service. 48èmes Journées de Statistique organisée par la Société Française de Statistique, 2016, Montpellier, France. hal-01383938

HAL Id: hal-01383938
https://hal.science/hal-01383938

Submitted on 15 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le modèle DFM pour une analyse comparative des systèmes de vélos en libre service

Charles Bouveyron [1] & Etienne Côme [2] & Julien Jacques [3]

[1] *Laboratoire MAP5, UMR CNRS 8145,*
Université Paris Descartes & Sorbonne Paris Cité
charles.bouveyron@parisdescartes.fr
[2] *Laboratoire GRETTIA, IFSTTAR*
etienne.come@ifsttar.fr
[3] *Laboratoire ERIC, Université Lumiére Lyon 2*
julien.jacques@univ-lyon2.fr

Résumé. Les systèmes de partage de vélos (BSS) sont devenus un moyen de transport intermodal durable et sont maintenant proposés dans de nombreuses villes à travers le monde. La plupart des BSS offrent également un libre accès à leurs données, en particulier les rapports d'état en temps réel sur leurs stations de vélos. L'analyse de la masse de données générées par ces systèmes est d'un intérêt particulier pour les fournisseurs de BSS désireux de mettre à jour les structures et leurs politiques de développement. Ce travail a été motivé par l'intérêt pour l'analyse et la comparaison de plusieurs BSS européens pour identifier les modes de fonctionnement communs et de proposer des solutions pratiques pour éviter les problèmes potentiels. Notre approche repose sur l'identification des patterns communs entre et au sein des systèmes. A cette fin, une méthode de clustering, appelée FunFEM, pour les séries chronologiques (ou plus généralement les données fonctionnelles) est développée. Elle est basée sur un modèle de mélange fonctionnel qui permet le regroupement des données dans un sous-espace fonctionnel discriminant. Ce modèle présente l'avantage dans ce contexte d'être parcimonieux et de permettre la visualisation des groupes dans un espace quantitatif. Des expériences numériques confirment la bonne tenue de FunFEM, en particulier par rapport aux méthodes de l'état de l'art. L'application de FunFEM aux données BSS de JCDecaux et de Transport for London Initiative nous permet d'identifier 10 tendances générales, y compris quelques patterns pathologiques, et de proposer des stratégies d'amélioration basées sur la comparaison du système. La visualisation des données en cluster dans le sous-espace discriminant se révèle être particulièrement instructif sur l'efficacité du système. La méthodologie proposée est mise en $\frac{1}{2}$uvre dans un paquet pour le logiciel R, nommé `funFEM`, qui est disponible sur le CRAN. Le paquet fournit également un sous-ensemble des données analysées dans ce travail.

Mots-clés. Clustering, données fonctionnelles, modèle de mélange, sous-espace

Abstract. Bike sharing systems (BSSs) have become a means of sustainable intermodal transport and are now proposed in many cities worldwide. Most BSSs also provide

open access to their data, particularly to real-time status reports on their bike stations. The analysis of the mass of data generated by such systems is of particular interest to BSS providers to update system structures and policies. This work was motivated by interest in analyzing and comparing several European BSSs to identify common operating patterns in BSSs and to propose practical solutions to avoid potential issues. Our approach relies on the identification of common patterns between and within systems. To this end, a model-based clustering method, called FunFEM, for time series (or more generally functional data) is developed. It is based on a functional mixture model that allows the clustering of the data in a discriminative functional subspace. This model presents the advantage in this context to be parsimonious and to allow the visualization of the clustered systems. Numerical experiments confirm the good behavior of FunFEM, particularly compared to state-of-the-art methods. The application of FunFEM to BSS data from JCDecaux and the Transport for London Initiative allows us to identify 10 general patterns, including pathological ones, and to propose practical improvement strategies based on the system comparison. The visualization of the clustered data within the discriminative subspace turns out to be particularly informative regarding the system efficiency. The proposed methodology is implemented in a package for the R software, named `funFEM`, which is available on the CRAN. The package also provides a subset of the data analyzed in this work.

Keywords. Clustering, functional data, mixture model, sub-space

Bibliographie

[1] C. Bouveyron, E. Côme and J. Jacques, *The discriminative functional mixture model for a comparative analysis of bike sharing systems*, The Annals of Applied Statistics, vol. 9 (4), pp. 1726-1760, 2015.

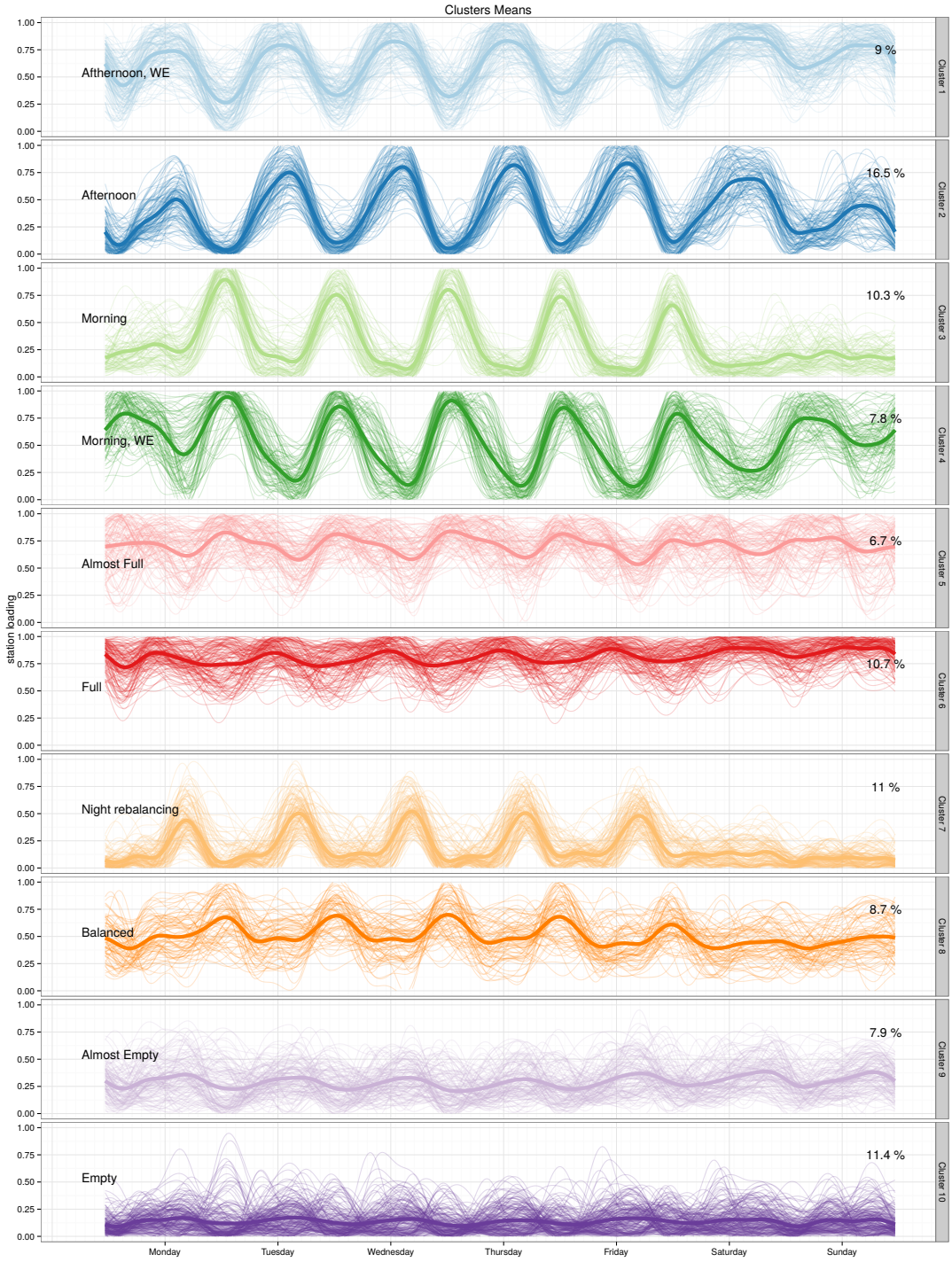


Figure 1: Profils moyens des 10 groupes identifiés sur les données de Paris par la méthode FunFEM ainsi que quelques échantillons de courbes de chacun des groupes.

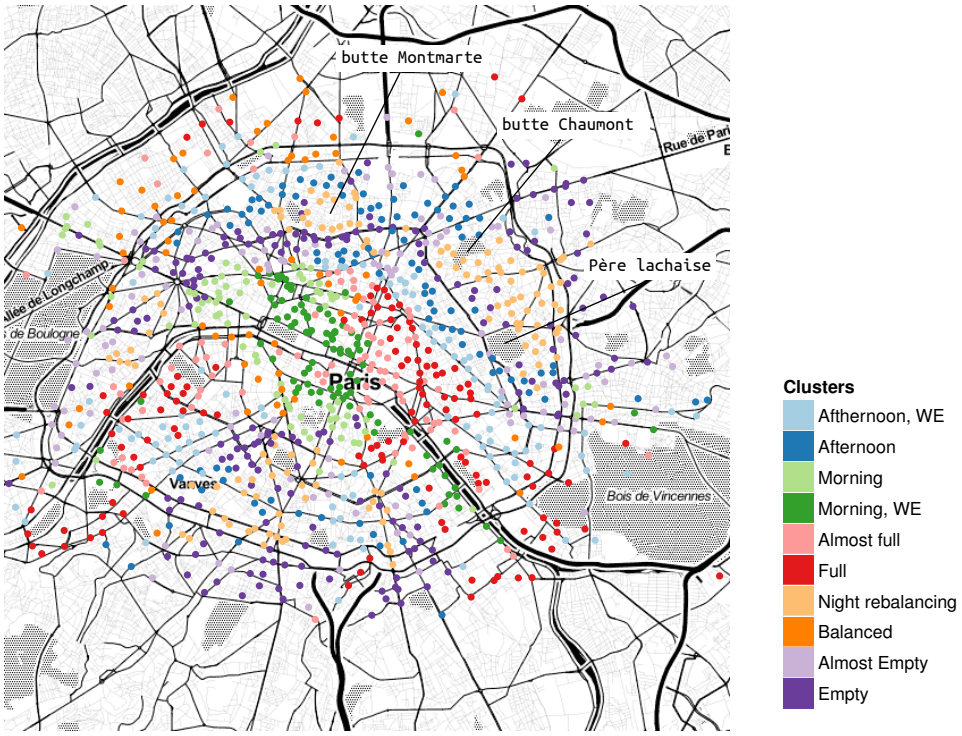


Figure 2: Représentation spatiale du résultat du clustering pour les stations parisiennes.