

Traces numériques de mobilité comment suivre la piste ?

NTIC & transport

Etienne Côme, Latifa Oukhellou
COSYS/GRETTIA

3 mars 2014



IFSTAR

Plan

- 1 Contexte et positionnement
- 2 Un exemple avec les traces VLS, et le Vélib
 - Clustering de stations avec des profils d'usage similaires
 - Clustering de matrices O-D dynamiques
 - Travaux actuels & futurs
- 3 Quelques exemples de visualisations

Contexte

Latifa Oukhellou (DR), Etienne Côme (CR),
Yufei Han (Post-Doc) et Andry Randriamanamihaga (Doc)

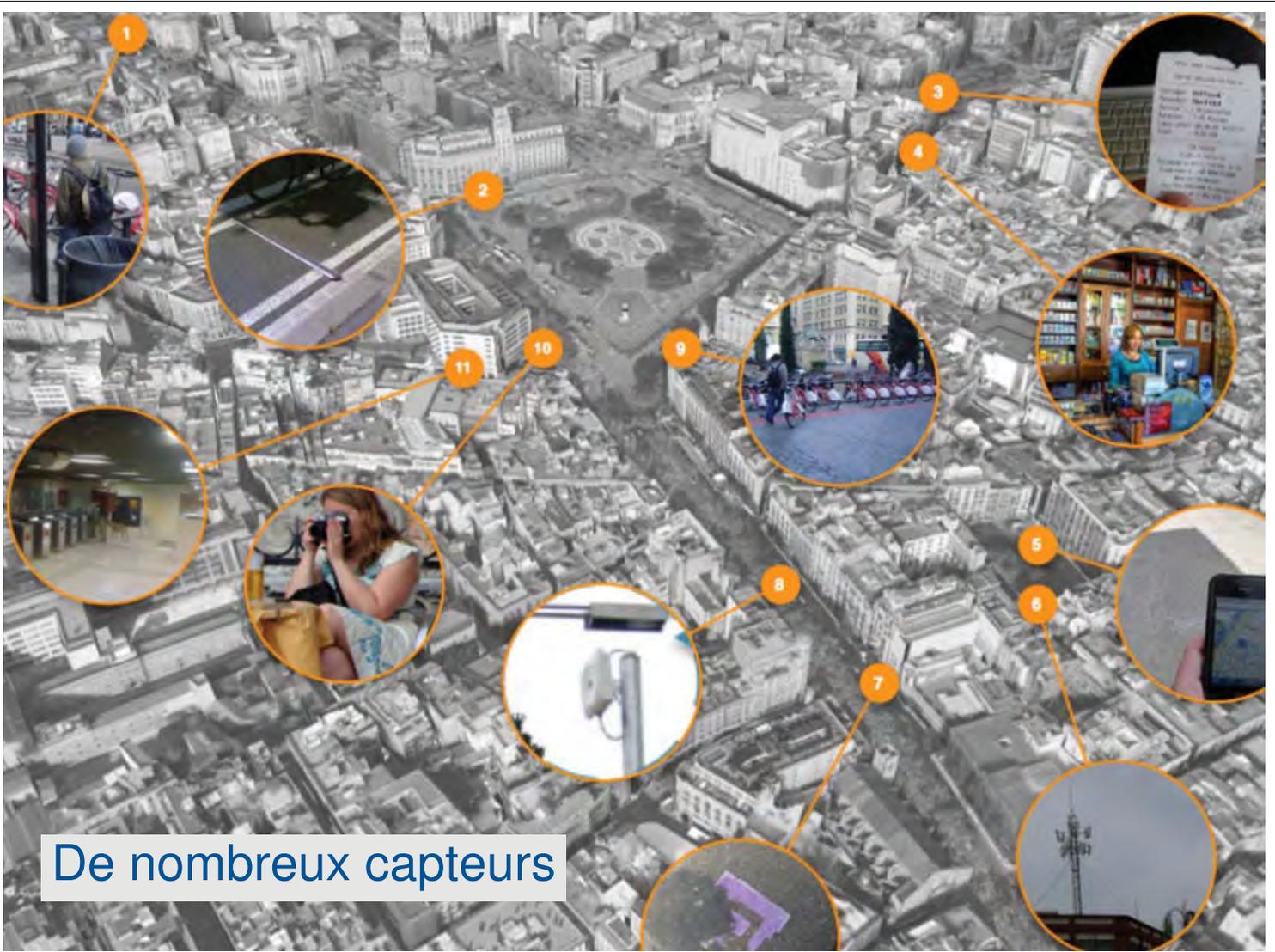
- GRETTIA, COSYS
- Analyse de données, **Clustering**, Apprentissage statistique,...
- Fouille de données de mobilité : ex VLS (Vélib', Velov,...), GPS, bluetooth, billettique

Terrain d'expérimentation

- Données billettique (projet Mobilletic)
- Données VLS

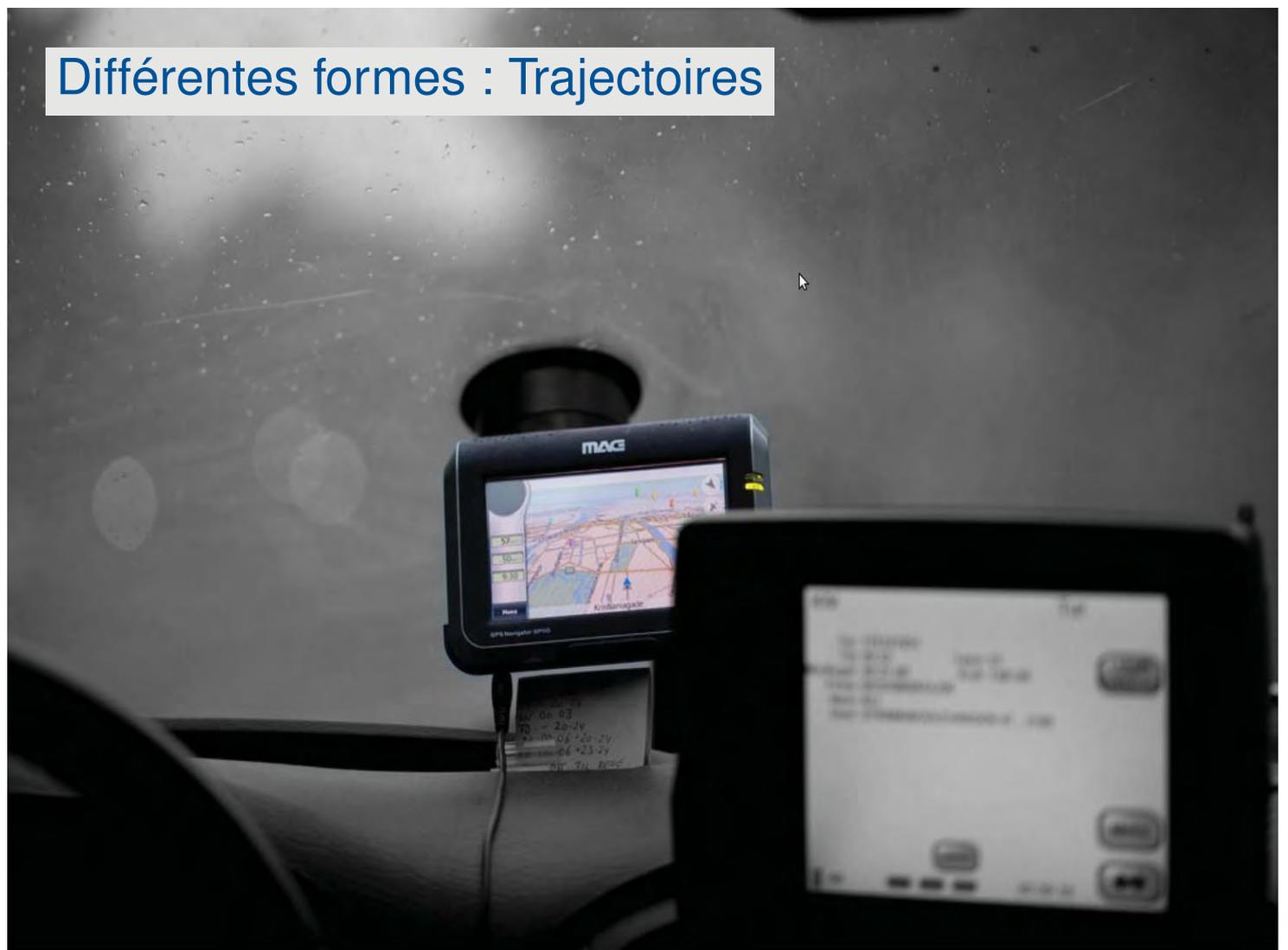
Traces numériques de mobilité





De nombreux capteurs

Différentes formes : Trajectoires



Différentes formes : Points



Différentes formes : Origines-Destinations





Nécessite de fouiller !!

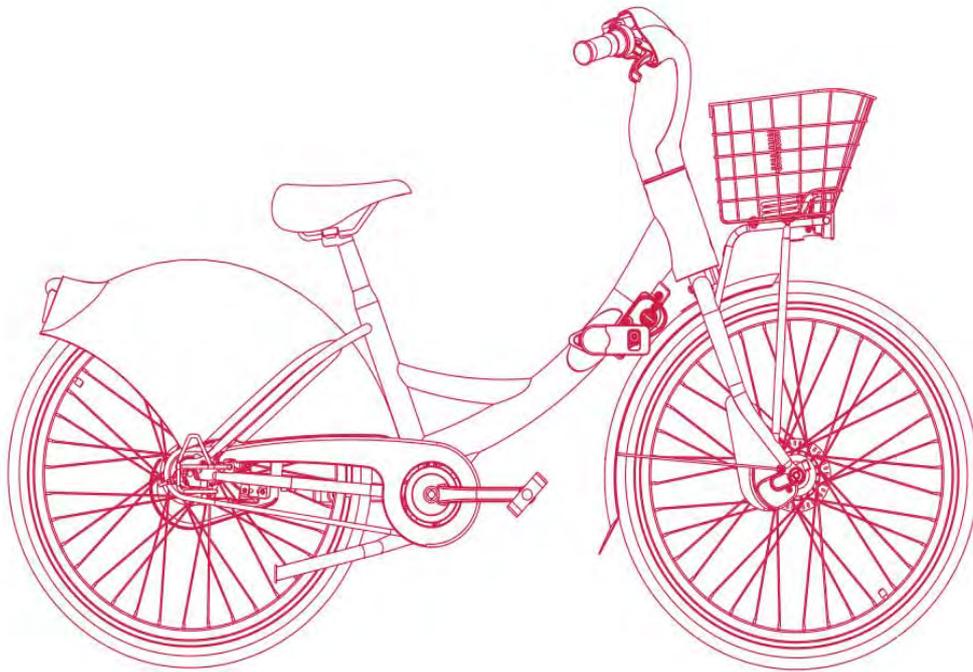
Contexte et positionnement

Contexte

Problématique

- Traces numériques de plus en plus nombreuses
- Générée en particulier lors de nos déplacements
⇒ Pertinentes pour l'étude des mobilité
- Usage détourné + Volume important
⇒ Outils de traitement et d'analyse
- Complémentarité / substitution enquêtes classiques

Un exemple avec les traces VLS, et le Vélib



Un exemple avec les traces VLS, et le Vélib

Exemple : VLS (Vélib)

Traces numériques ?

- Données d'occupation des stations (captation depuis juillet 2013)
- **Données des trajets** effectués sur 5 mois de 2011 (Mairie de Paris et JC. Decaux)
- Données **temporelles** : dates et heures de départ et d'arrivée
- Données **spatiales** : stations de départ et d'arrivée
- Type d'abonnement (longue ou courte durée)

Exemple : VLS (Vélib)

Objectifs "Opérationnels"

- Eviter le problème d'une station vide au départ et pleine au retour
⇒ Optimisation des politiques de re-déploiements des vélos
- Concevoir de nouveaux systèmes, ou étendre un système existant
⇒ Positionnement et dimensionnement des stations en fonction des données socio-économiques et géographiques, ...

Objectifs "Traitements"

- Rechercher des motifs d'usage récurrents du Vélib', modèles spatio-temporels
- Construire des modèles de prédiction d'usage, estimation de la demande
- Lire la ville et son rythme au travers de données de mobilité

Effets temporels

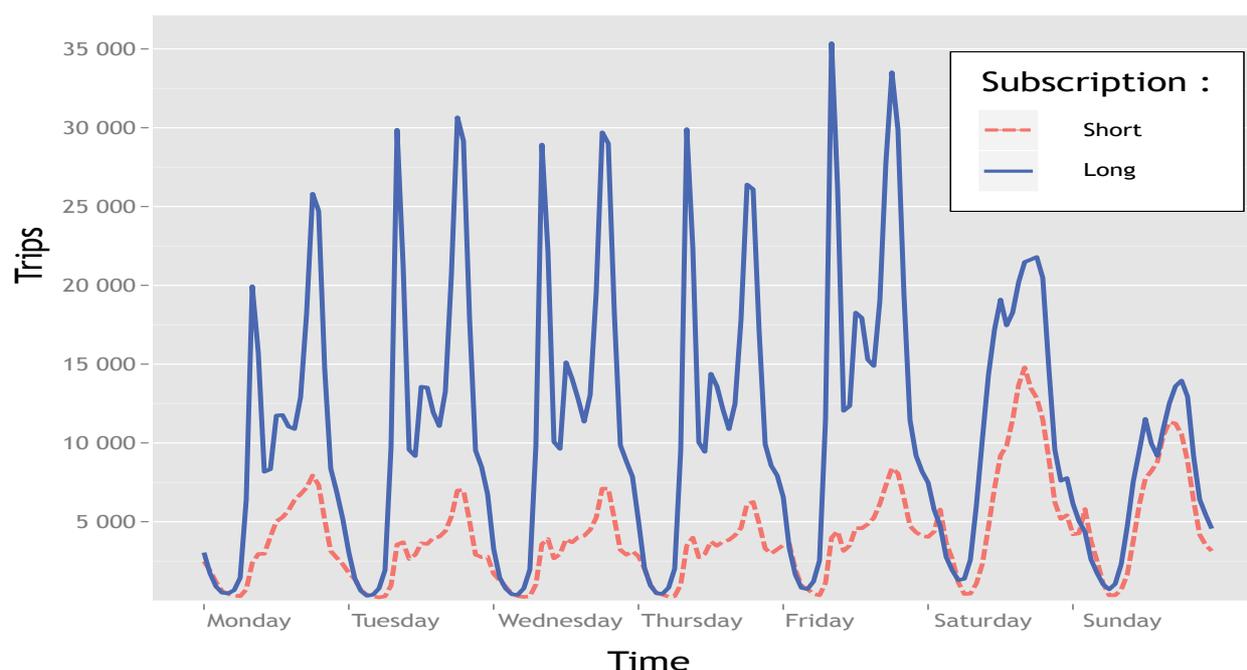


FIGURE 1: Nombre de trajets / heure (courte / longue abonnements)

Effets spatiaux

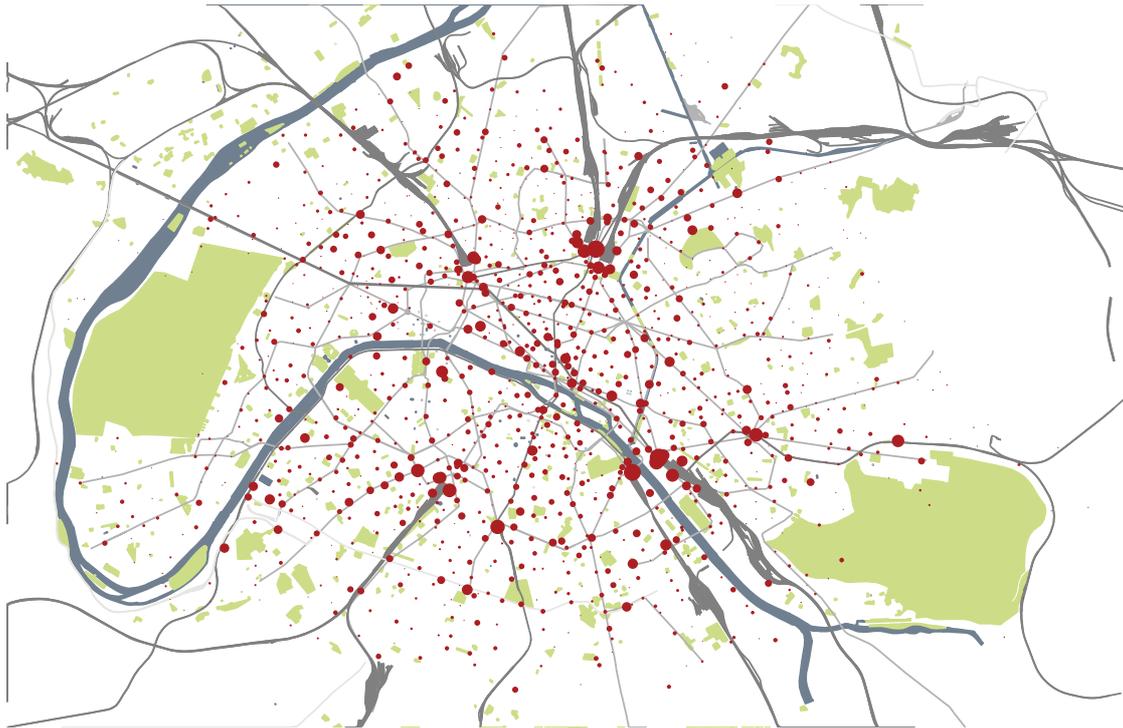


FIGURE 2: trajets entrants entre [6h,7h] en semaine

Analyse exploratoire : laisser parler les données

Méthodologie générale

- Utiliser des algorithmes de clustering (regroupement automatique) pour trouver des formes d'usage type du Vélib'
- Croiser les groupes ou clusters trouvés avec des données géographiques et socio-économiques de la ville
⇒ Facteurs influents sur l'usage du système VLS.

Deux pistes de travail :

- 1 Grouper les stations avec des profils d'usage similaires
→ Clustering de stations ou de couples O/D
- 2 Segmenter / Résumer la dynamique temporelle du système
→ Adaptation de LDA à la fouille de matrices O/D dynamiques

Clustering de stations avec des profils d'usage similaires

Regroupement de stations à partir de leurs profils d'usage temporels

Objectifs :

- Stations décrites par les dynamiques de flux entrants et sortants
- Prise en compte des jours de semaine / week end
- modèle pour données de comptage
- Croiser les résultats avec d'autres variables explicatives : population, emplois, loisirs, ...
- Analyse réalisée avec 8 groupes (bon compromis : interprétations/attache aux données)

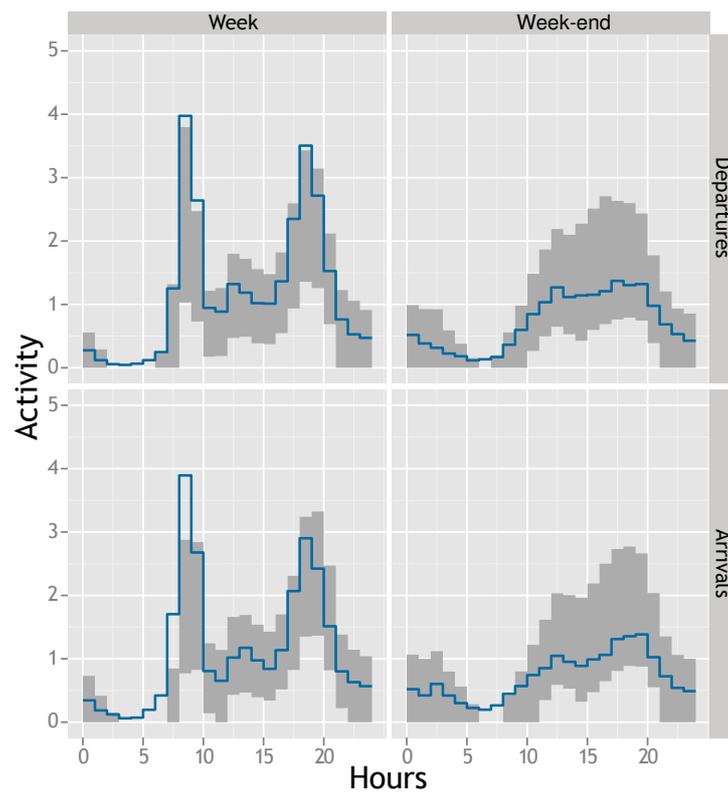
Related work

[Borgnat et al., 2011], [Froehlich et al., 2009], [Lathia et al., 2012]...

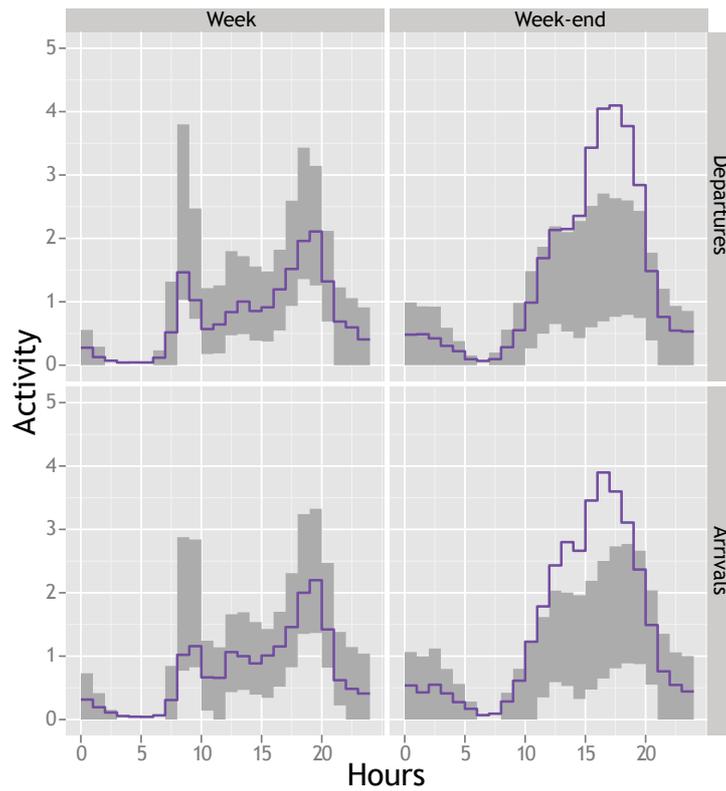
Pôles multimodaux



Pôles multimodaux



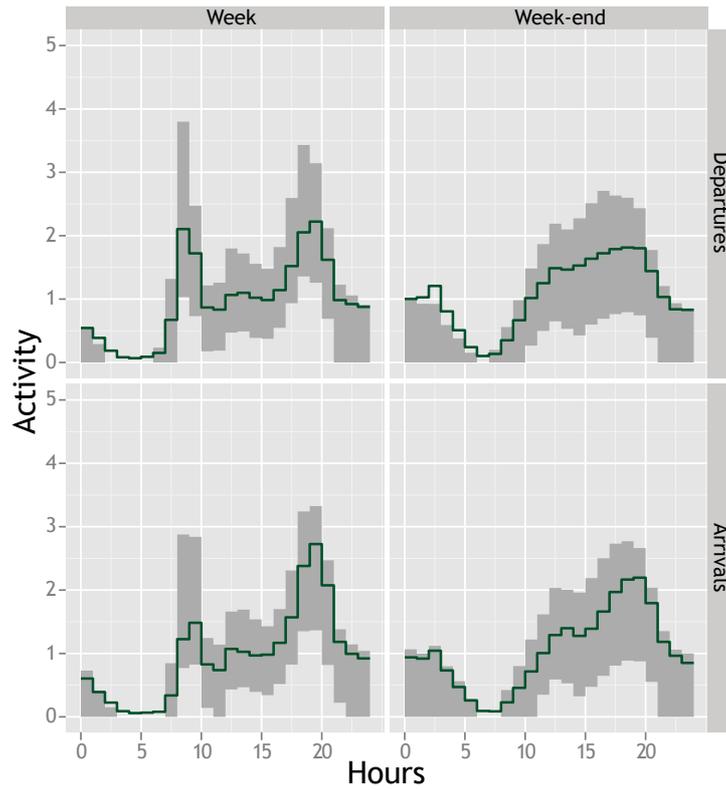
Parcs



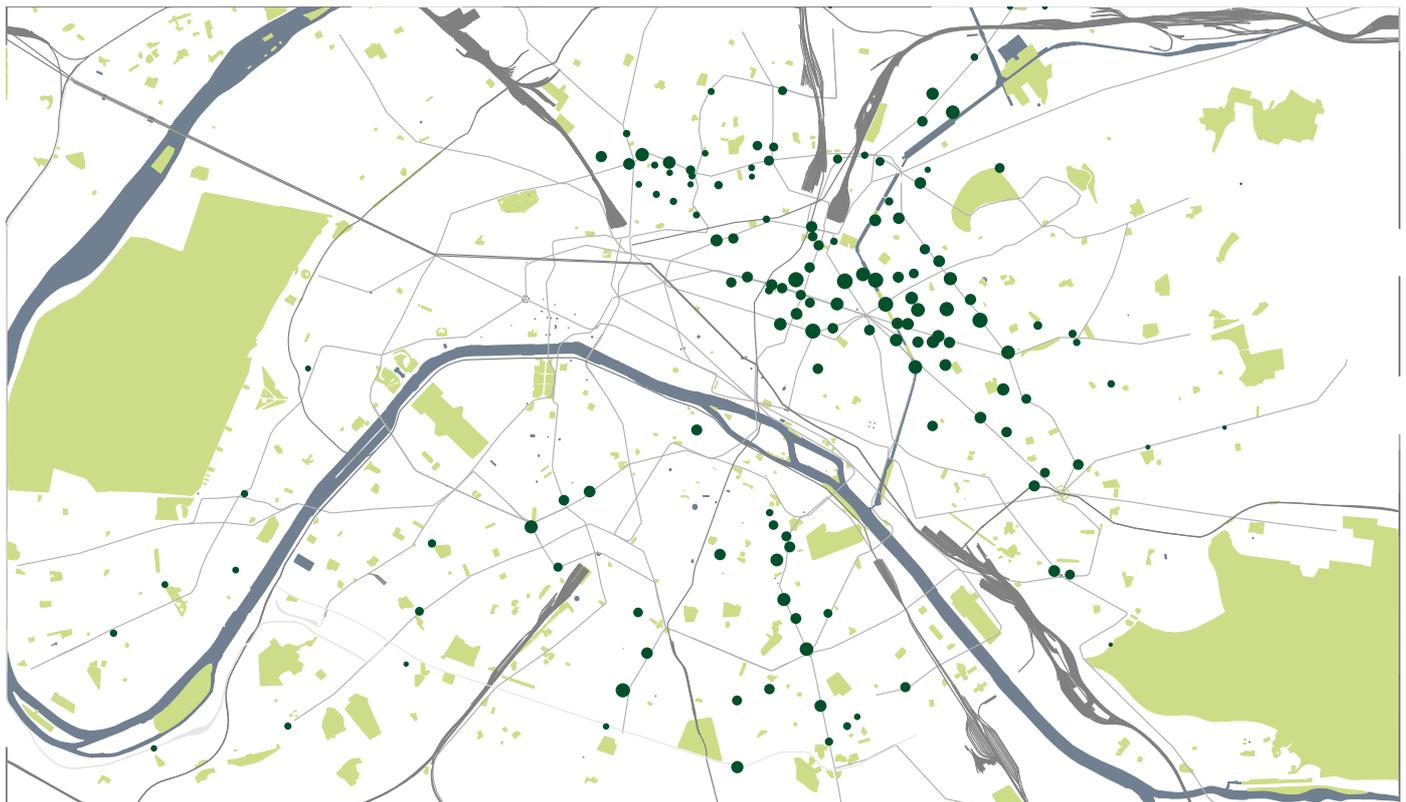
Parcs



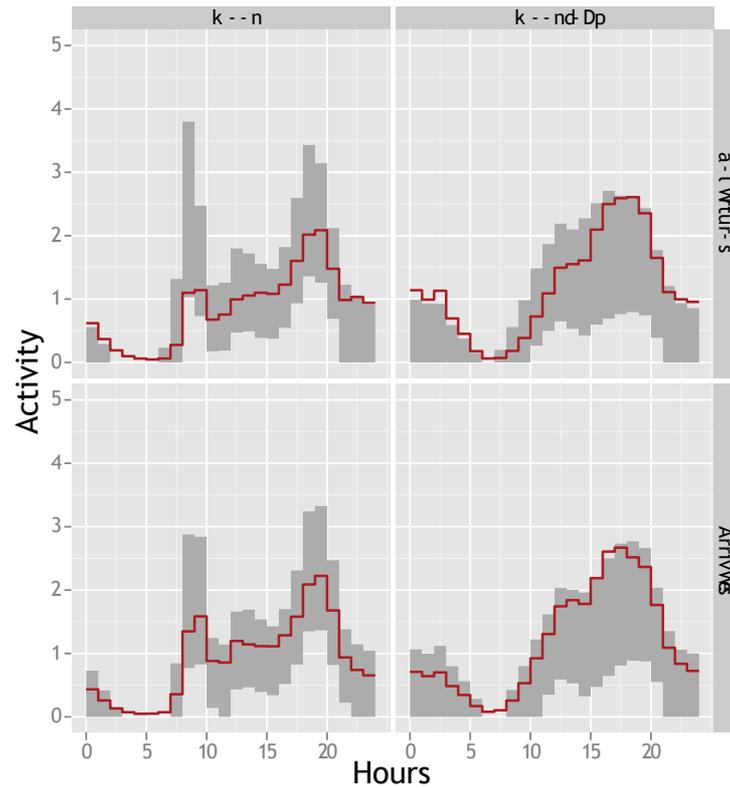
Sorties nocturnes



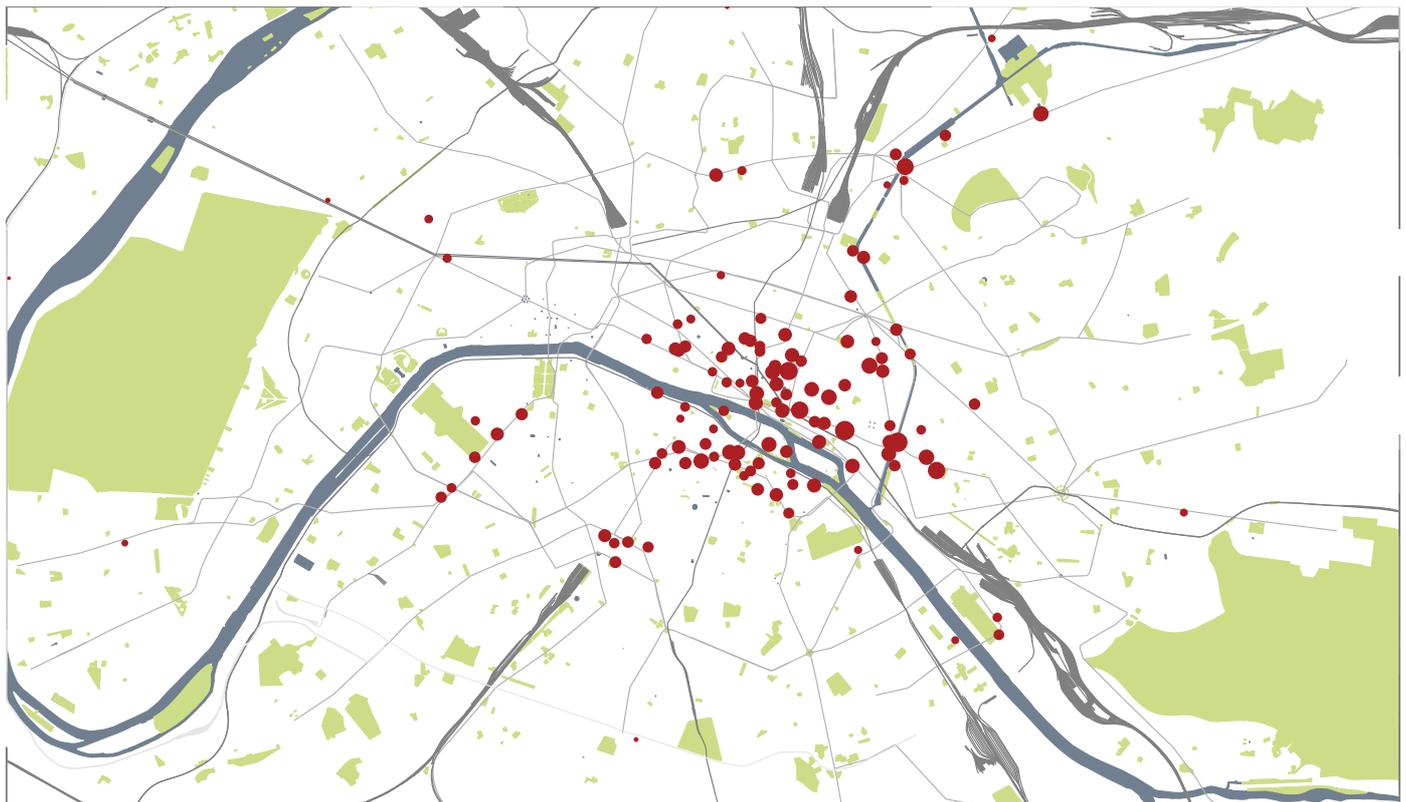
Sorties nocturnes



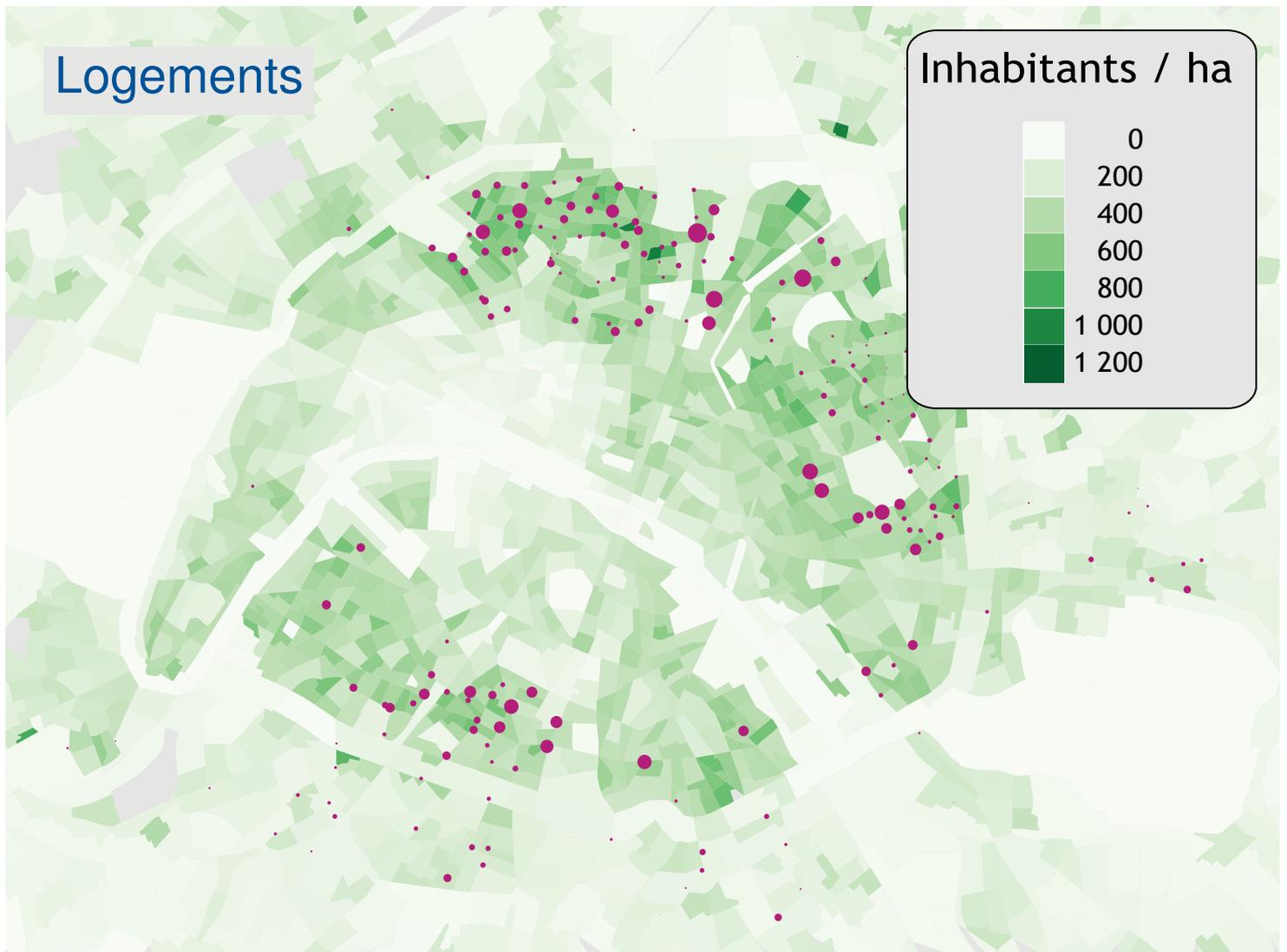
Sorties nocturnes et week-end



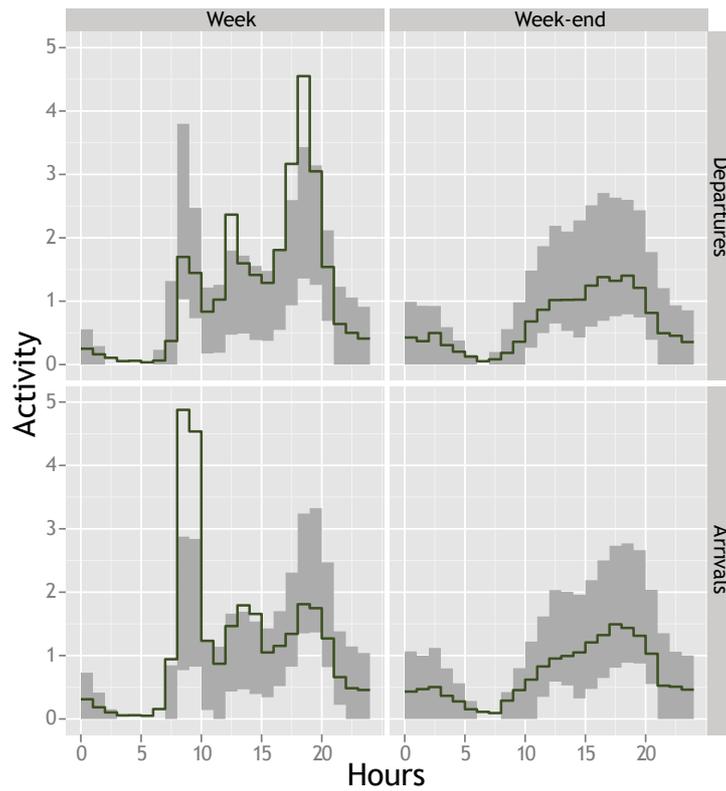
Sorties nocturnes et week-end



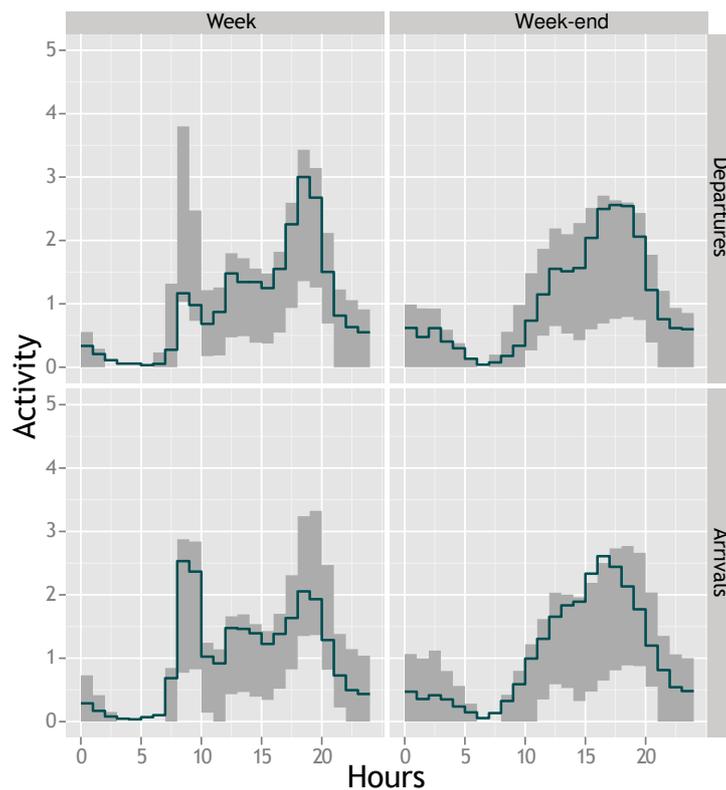
Logements



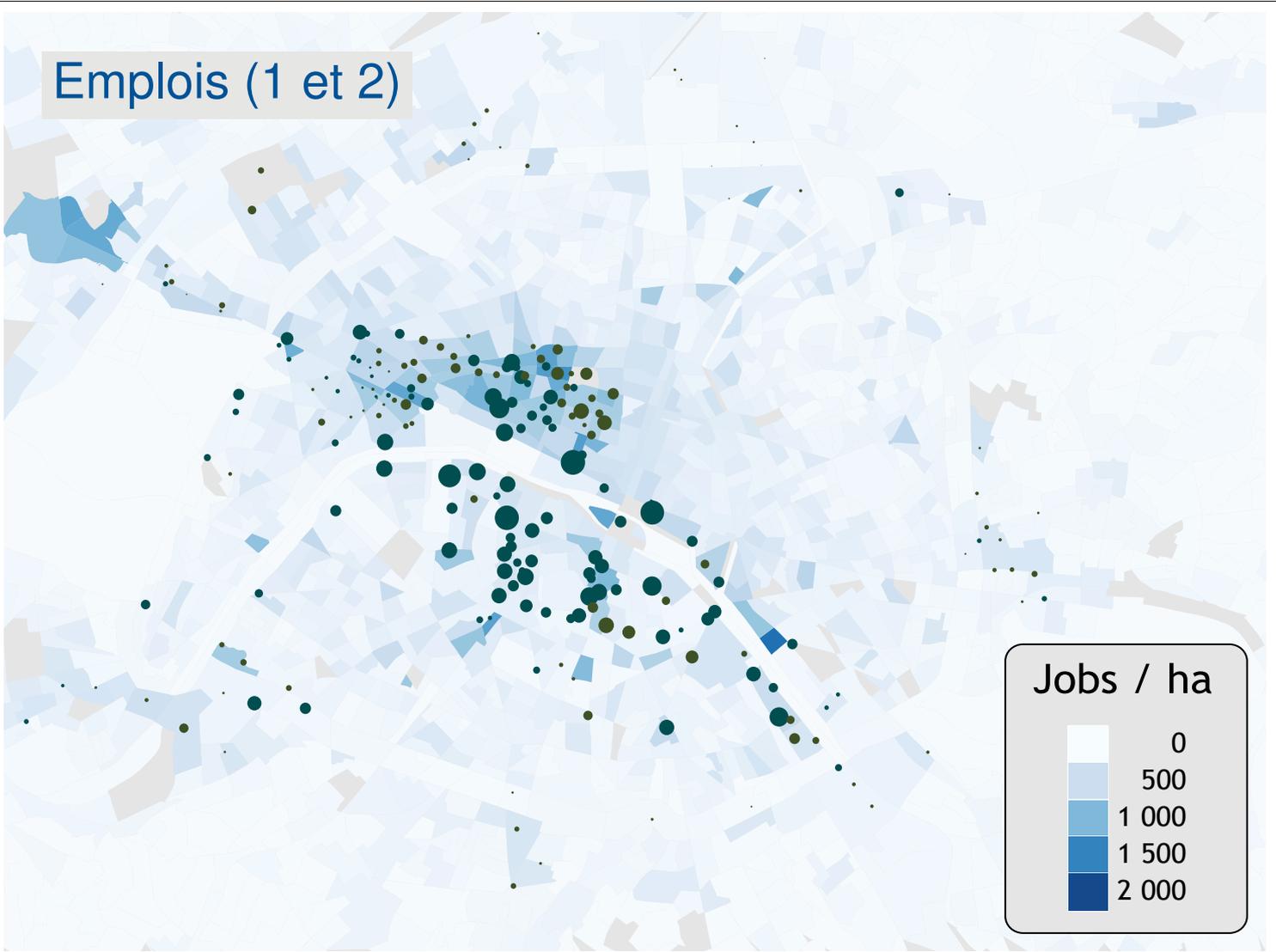
Emplois(1)



Emplois (2)



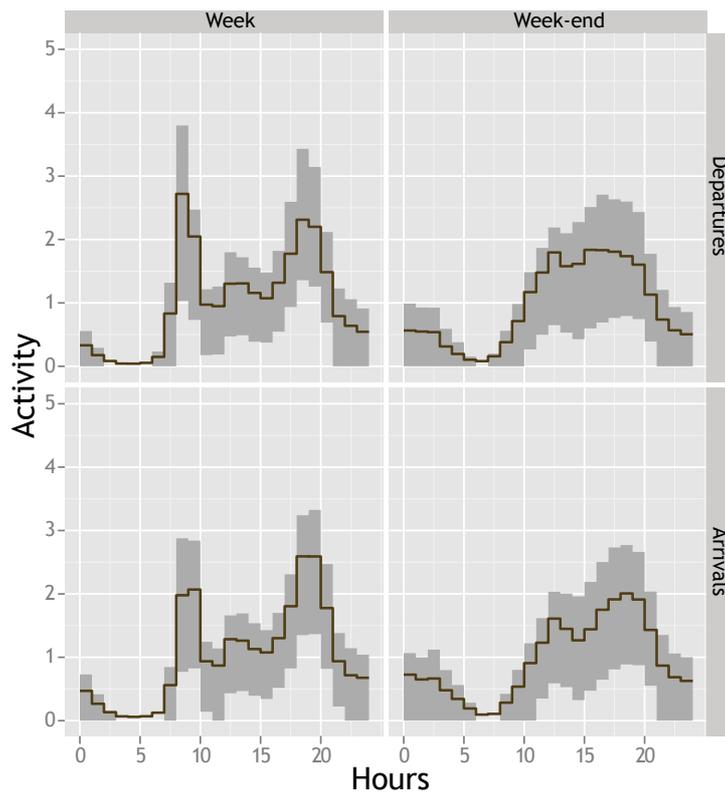
Emplois (1 et 2)



Un exemple avec les traces VLS, et le Vélib

Clustering de stations avec des profils d'usage similaires

Usage mixte



Usage mixte



Croisement données population-emplois-services

	hab/ha	emp/ha	serv/ha	com/ha
	162	237	4.2	3.7
Sorties (1)	367	189	6.3	4.4
Sorties (2)	261	322	7.7	6.9
Parcs	172	90	2	1.7
Gares	209	206	2.4	1.8
Logements	375	108	3.8	2.7
Emplois(1)	138	409	4.5	2.8
Emplois(2)	157	456	5.7	5.6
Moyennes	301	163	3.8	2.8

TABLE 1: Comparaison des moyennes de densités de population, d'emplois, de services et de commerces pour les différentes groupes de stations.

Conclusions sur le clustering de stations

Résultats

- Des stations bien différenciées en termes d'usage
- Interprétation aisée des groupes de stations
- Densités de population, d'emplois et d'équipements explicatives des groupes de stations
- Profils temporels des clusters interprétables et informatifs

Clustering de matrices O-D dynamiques

Clustering de matrices O-D dynamiques

Objectifs

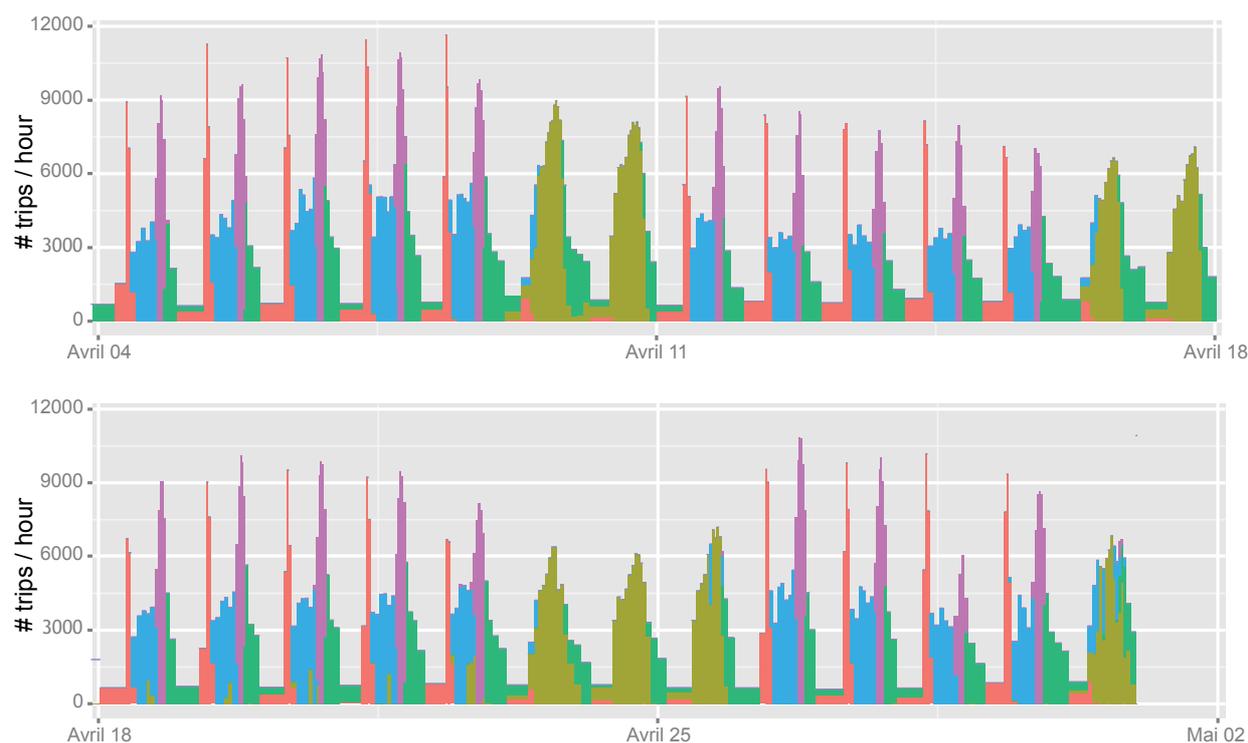
- Représentation des données : matrices OD dynamiques
⇒ Recherche de stationnarités et de points de changement dans la dynamique des matrices OD
- Modèle utilisé “Latent Dirichlet Allocation”

Résultats

- Segmentation temporelle et cycles
- Distribution spatiale des flux (OD de référence) / segment
⇒ Permet de caractériser le déséquilibre du réseau / segment

Représentation synthétique de la dynamique du réseau

Interprétation temporelle



Interprétation temporelle

Remarques

- La cyclostationnarité est clairement visible
- Faible mélange entre les différentes activités latentes
- Interprétation temporelle des 5 "activités" obtenues :
 - 1 Domicile ↔ Travail
 - 2 Déjeuner
 - 3 Travail ↔ Domicile
 - 4 Loisirs nocturnes
 - 5 Loisirs

Analyse spatiale

Quelles stations

génèrent des départs

durant une activité a // comportement moyen du système ?

Quelles stations

génèrent des arrivées

durant une activité a // comportement moyen du système ?

Comment les stocks évoluent

balance des vélos

(vélos arrivant - vélos partant)
durant chacune des activités ?

"Domicile-Travail", stations avec un fort flux entrant

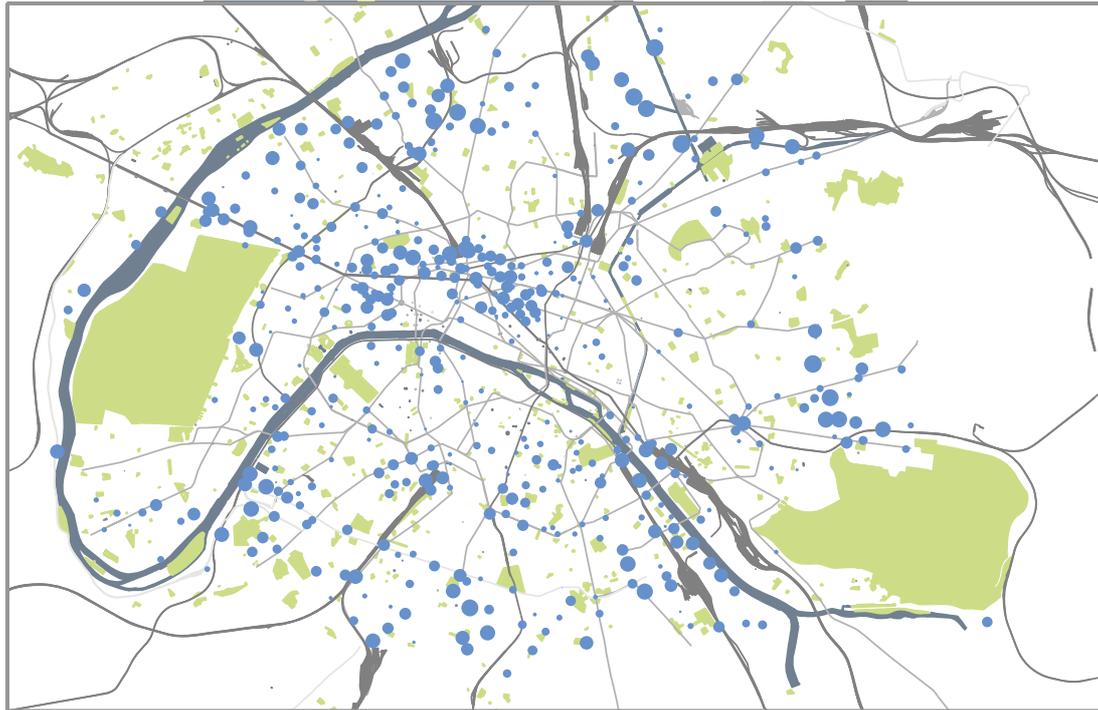


FIGURE 3: Spécificité entrante / stations

"Loisirs", stations avec un fort flux sortant

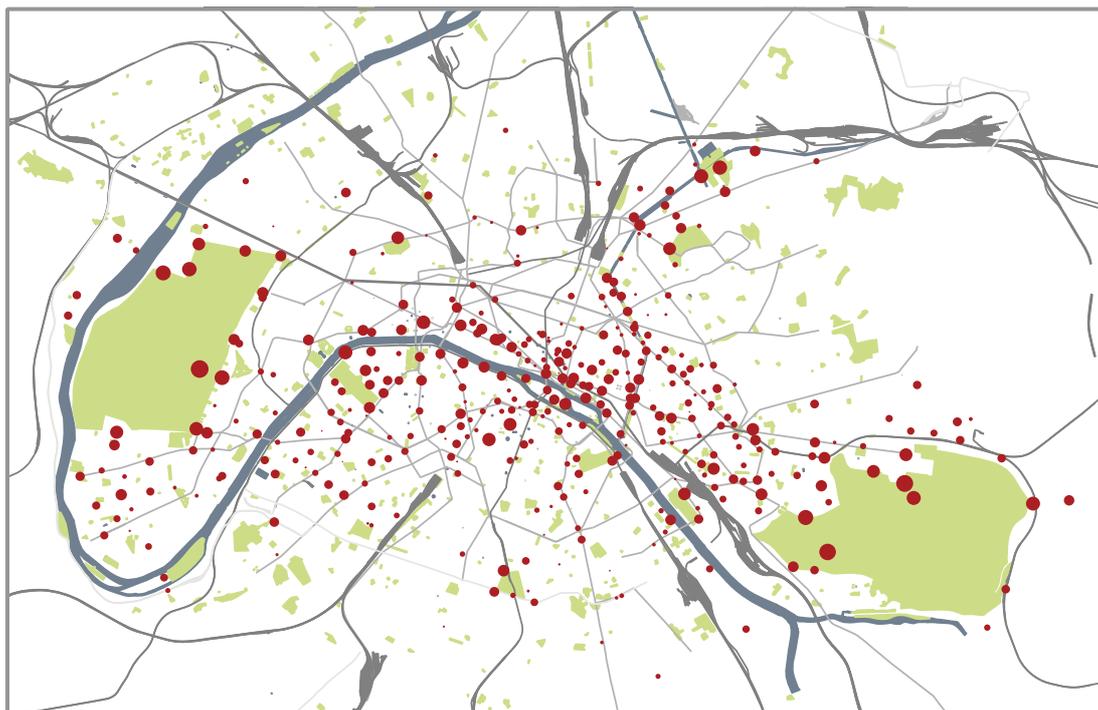


FIGURE 4: Spécificité sortante / stations

Interprétation spatiale : déséquilibre du réseau



FIGURE 5: Activité latente "Domicile→Travail", déséquilibre du réseau pour $N_{dep} = 10\,000$

"Travail→Domicile"



FIGURE 6: Activité latente "Travail→Domicile", déséquilibre du réseau pour $N_{dep} = 10\,000$

"Déjeuner"



FIGURE 7: Activité latente "Déjeuner", déséquilibre du réseau pour $N_{dep} = 10\,000$

"Loisirs nocturnes"



FIGURE 8: Activité latente "Loisirs nocturne", déséquilibre du réseau pour $N_{dep} = 10\,000$

"Loisirs"



FIGURE 9: Activité latente "Loisirs", déséquilibre du réseau pour $N_{dep} = 10\,000$

Résultats

- Meilleure compréhension de la dynamique du réseau
- Modèle synthétique de la dynamique
- Mise en évidence des cycles
- Appliqué aussi à des données Bluetooth (TRB)

Limites

- Prise en compte uniquement de la demande satisfaite
- Pas de liens explicites avec les données socio-économiques et géographiques

Travaux actuels & futurs

Travaux actuels & futurs

Estimation/Prédiction de la demande (satisfaite + satisfaite)

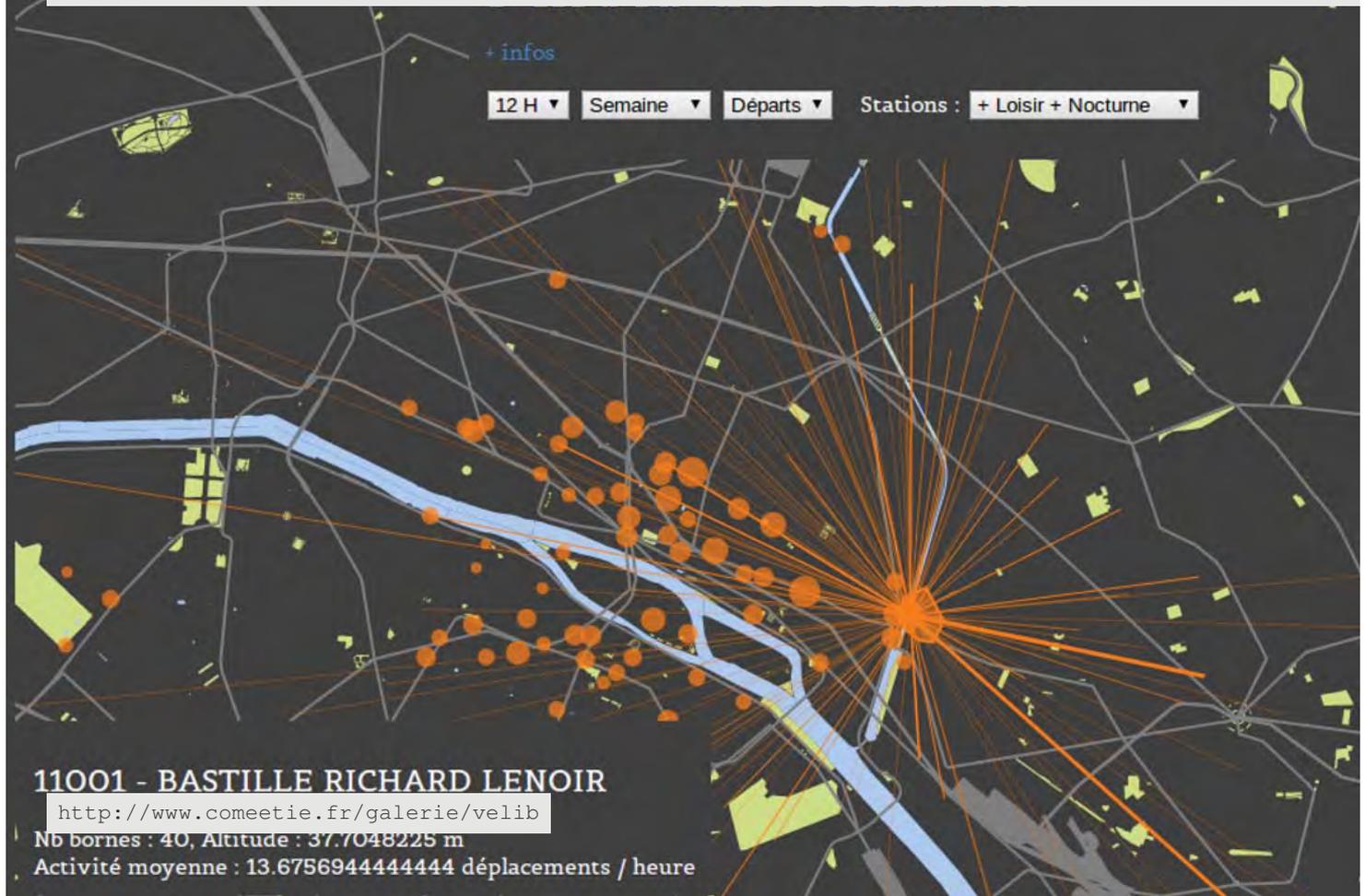
Objectifs :

- prédire l'évolution des stocks à courts termes
- fournir des outils d'aide au dimensionnement
→ pour l'extension d'un réseau → pour l'implantation d'un nouveau réseau

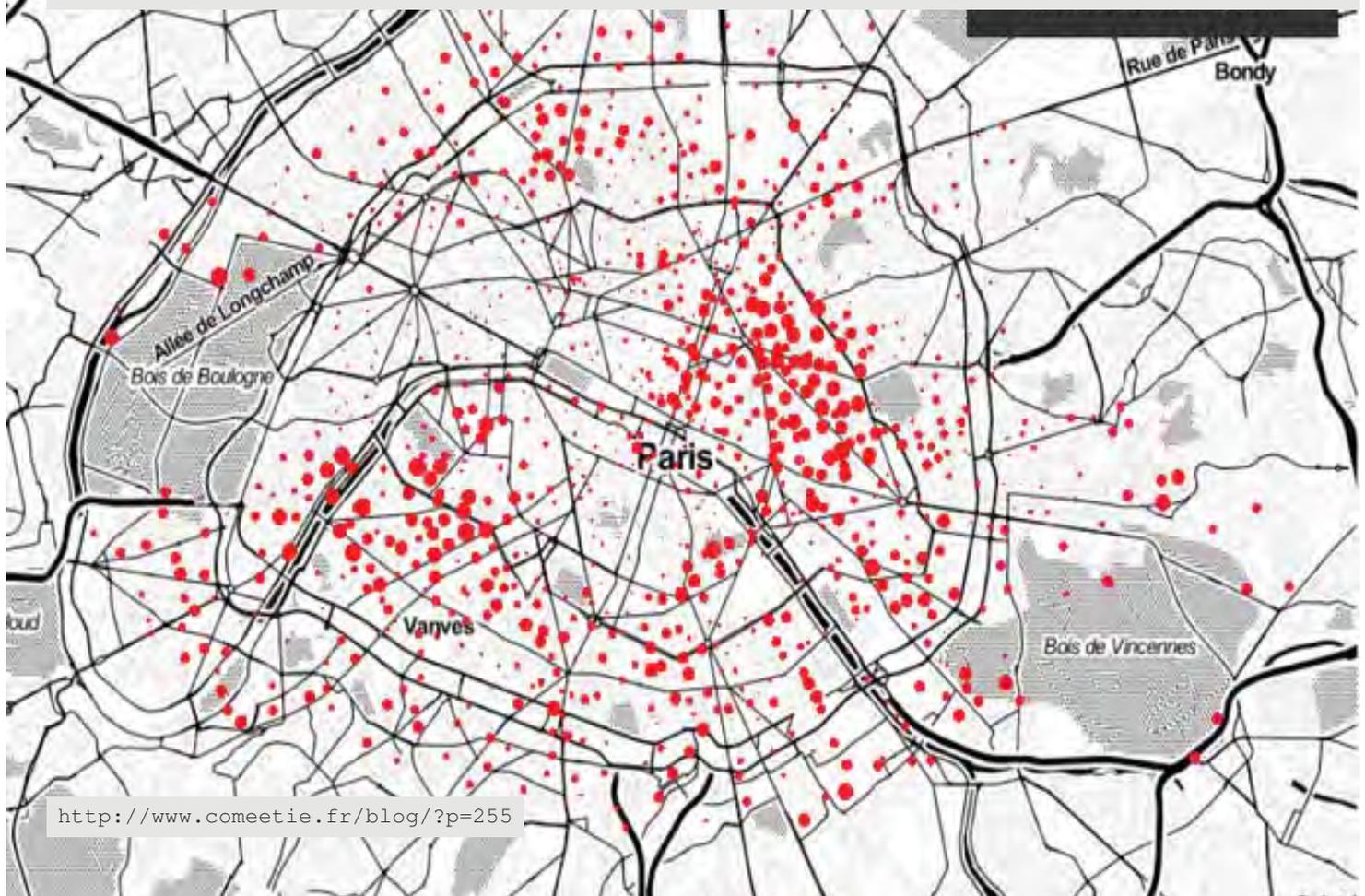
Travaux :

- source d'information contextuelle (socio-éco, transport,...)
- modèle de régression pour données de comptages
- prise en compte des effets liés à la saturation des stations
- extension LDA semi-paramétrique avec prise en compte des données contextuelles

Quelques exemples de visualisations pour se détendre

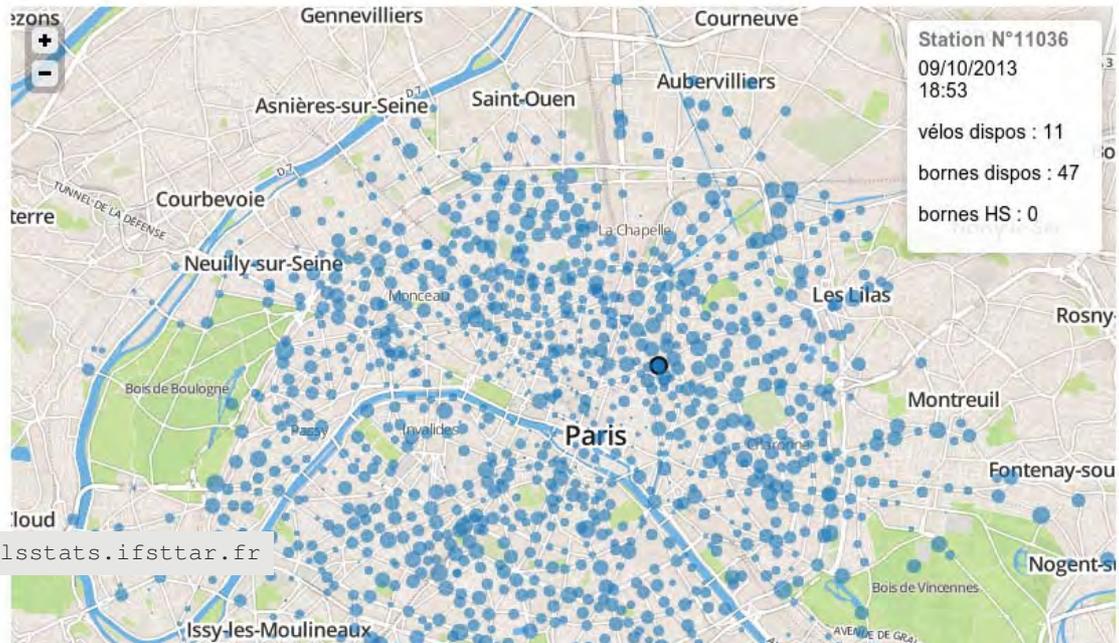
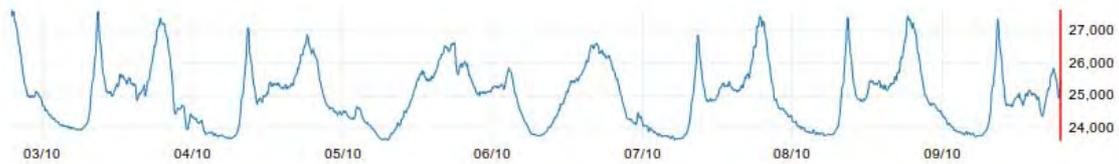


Quelques exemples de visualisations pour se détendre



Quelques exemples de visualisations pour se détendre

Enfin, naviguez dans l'historique en utilisant les flèches gauche et droite de votre clavier ou en utilisant les boutons ci dessous:



Quelques exemples de visualisations

Publications

- 1 E. Côme, L. Oukhellou. *Model-based count series clustering for Bike-sharing system usage mining, a case study with the Vélib' system of Paris*. Accepted in ACM TIST.
- 2 A. Randriamanahaga, E. Côme, L. Oukhellou, G. Govaert. *Clustering the Vélib' origin-destinations flows by means of Poisson mixture models*. In Proceedings of the European Symposium on Artificial Neural Networks, 2013. + Neurocomputing
- 3 E. Côme, A. Randriamanahaga, L. Oukhellou. *Spatio-temporal analysis of Dynamic Origin-Destination data using Latent Dirichlet Allocation. Application to the Vélib' Bike Sharing System of Paris..* In Proceedings of TRB 2014 (accepted)
- 4 H. Yufei, E. Côme, L. Oukhellou. *Toward Bicycle Demand Prediction of Large-Scale Bicycle-Sharing System*. In Proceedings of TRB 2014 (accepted)
- 5 P-A Laharotte, R. Billot, E. Côme, L. Oukhellou, A. Nantes, N-E El Faouzi. *Spatio-temporal clustering Analysis of Bluetooth OD (B-OD). Application to a large urban network*. In Proceedings of TRB 2014 (submitted)

Merci pour votre attention !

latifa.oukhellou@ifsttar.fr
etienne.come@ifsttar.fr (@comeetie)
mobilletic.ifsttar.fr
vlsstats.ifsttar.fr
www.comeetie.fr

Ifsttar

Centre de Marne-la-Vallée
Batiment le "Bienvenue"
14-20 Bd Newton Cité Descartes, Champs sur Marne
F-77447 Marne la Vallée Cedex 2

Tél. +33 (0)1 81 66 87 19



References

References



Borgnat, P., Abry, P., Flandrin, P., Robardet, C., Rouquier, J.-B. and Fleury, E. (2011).
Shared Bicycles in a City : A Signal processing and Data Analysis Perspective.
Advances in Complex Systems, 14, 1–24.



Froehlich, J., Neumann, J. and Oliver, N. (2009).
Sensing and Predicting the Pulse of the City through Shared Bicycling.
In *Proceedings of the 21st International Joint Conference on Artificial Intelligence* pp. 1420–1426,.



Lathia, N., Ahmed, S. and Capra, L. (2012).
Measuring the impact of opening the London shared bicycle scheme to casual users.
Transportation Research Part C : Emerging Technologies 22, 88–102.