

Fouille de données de mobilité

Thomas Devogele

Université François Rabelais (Tours)

thomas.devogele@univ-tours.fr

Laurent Etienne

Ecole Navale (Brest)

Laurent.etienne@ecole-navale.fr



La fouille de donnée (classique)

● Définition

- **Extraction automatique** ou semi-automatique de **connaissances cachées**, potentiellement utiles, à partir de données stockées dans des **grandes bases de données**.
- connaissances cachées
 - Relation
 - La présence de A entraîne souvent la présence de B
 - Motifs (pattern)
 - Comportement type
- à la croisée
 - de l'intelligence artificielle
 - de la statistique
 - Des bases de données

Les différentes phases

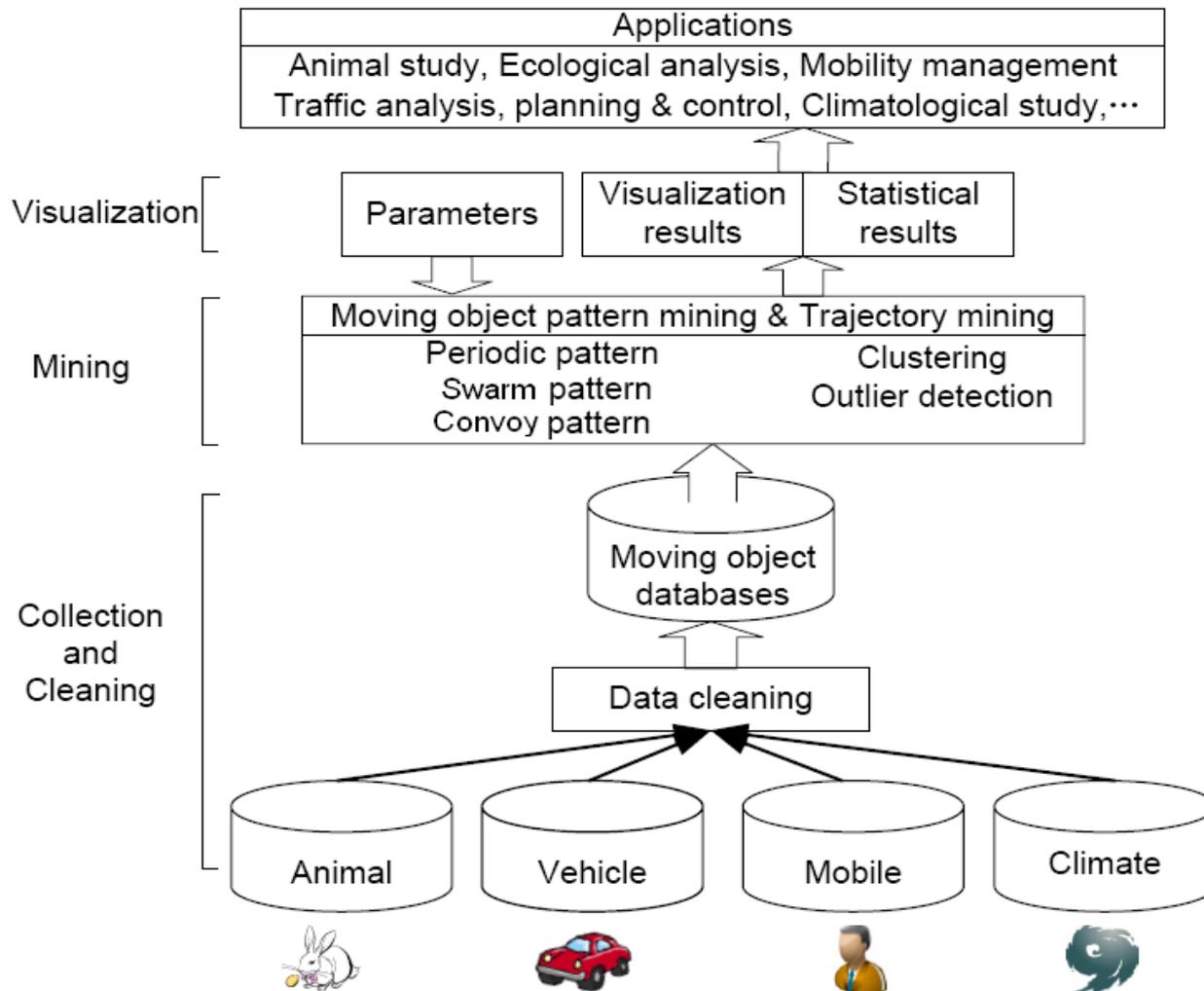
● Phases

- Acquisition
- Préparation des données
- Groupement de trajectoires (Clustering)
- Définition de motifs (pattern)
- Association à une classe (Classification)
- Détection de trajectoires inhabituelles
- Visualisation

● Remarque

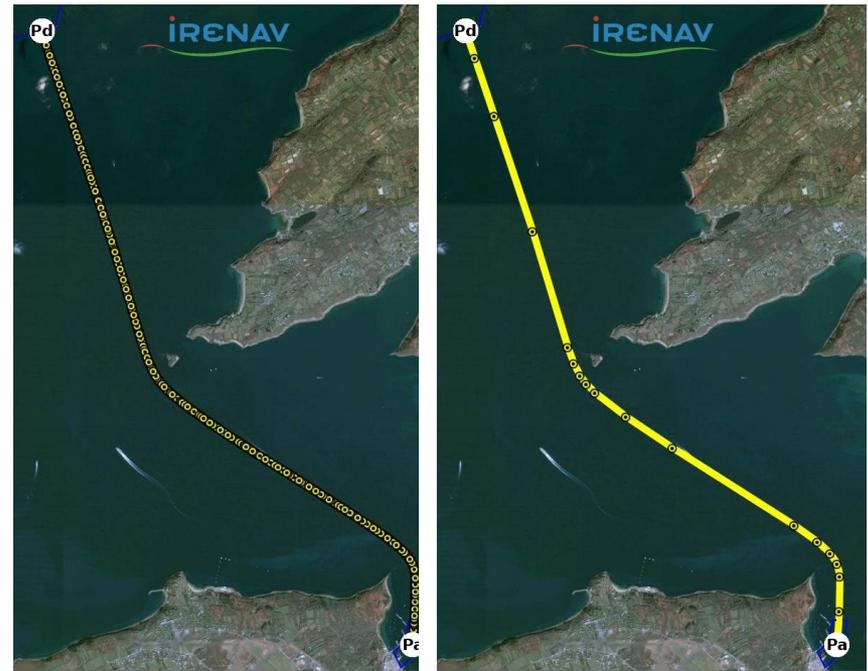
- Contexte applicatif fort

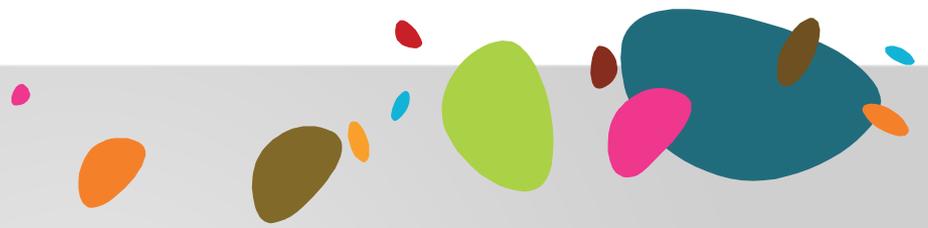
Fouille de données de mobilité (MoveMine)



Préparation des données

- Filtrage des données incohérentes
 - Erreur de mesure
 - Données manquantes
- Echantillonnage
 - Généralisation
 - Douglas et Peucker Spatio-temporelle
 - Rééchantillonnage si pas assez de données
 - Echantillonnage au même pas
 - Toutes les positions sont interpolées aux mêmes estampilles temporelles





Clustering et similarité



Clustering classiques

- En entrée

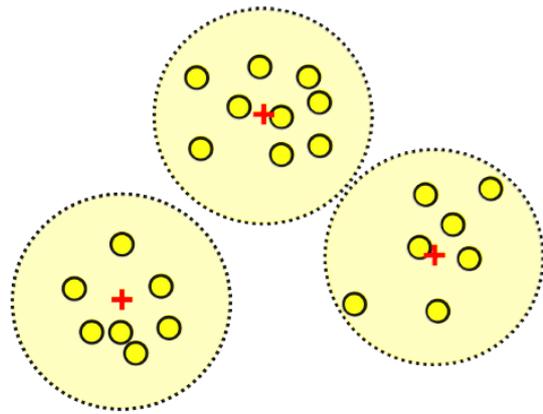
- un **ensemble d'individus** chacun ayant un certain nombre d'attributs, et une mesure de similarité entre eux deux à deux,

- En sortie

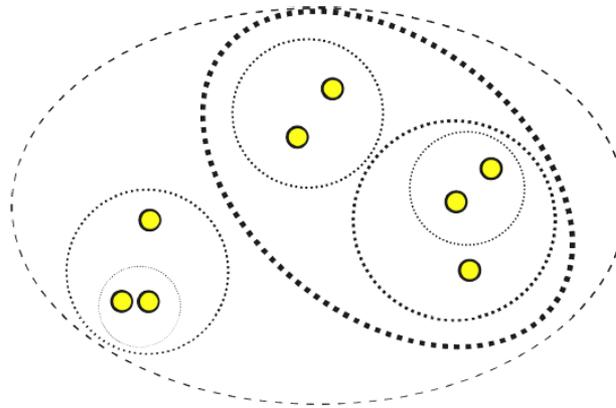
- des **classes** telles que :
 - **Similarité** forte entre individus d'une même classe
 - **Similarité** faible entre individus de classes différentes



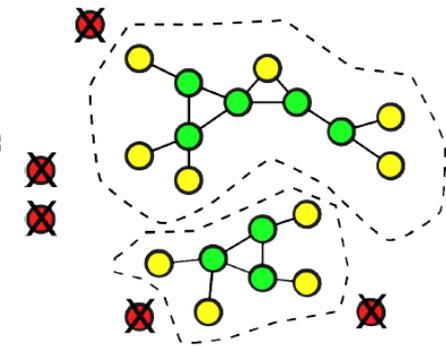
Méthodes classique de clustering



(a) *K*-means



(b) Hierarchical

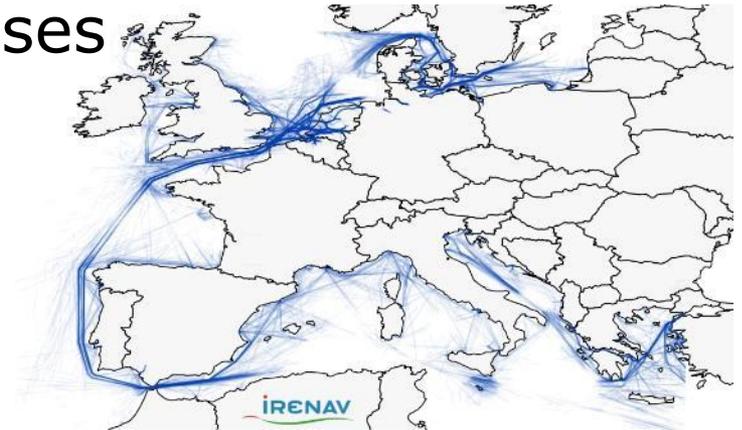


(c) Density-based

Besoin de notion de **similarité**

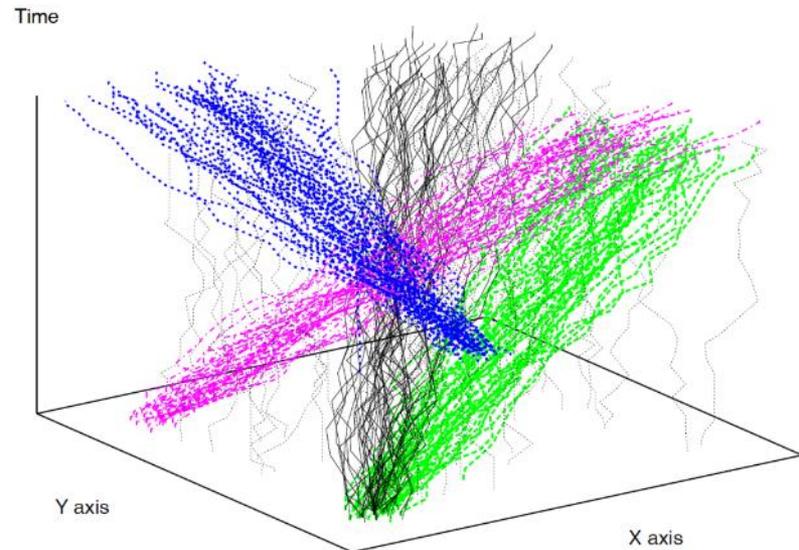
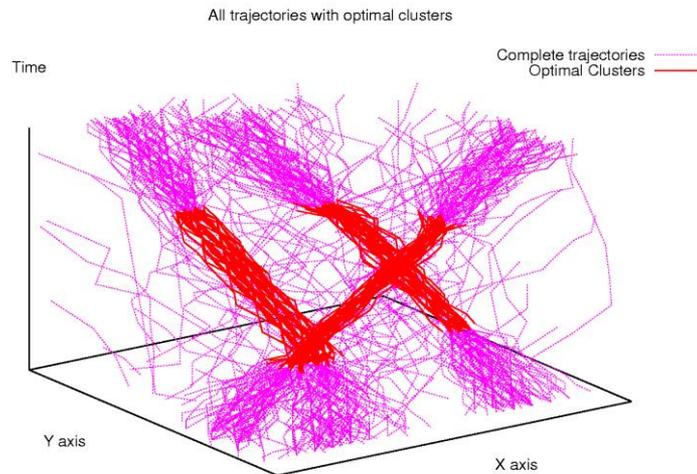
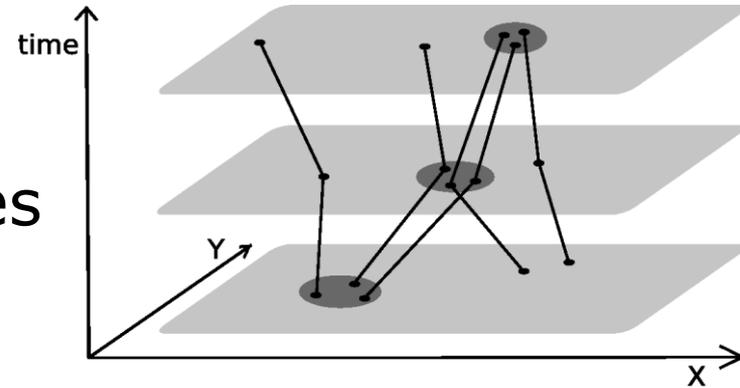
Clustering de trajectoires

- Détecter quoi ?
 - Des mouvements de groupes
 - Nuée (Swarm), Troupeau (Flock), Convoi (Convoy),
 - Des trajectoires suivant la même routes
 - Des routes classiques d'un grand nombre d'individus
 - Des mouvements courant à certaines périodes
 - Des mouvements de leader/followers
 - Des zones de passage denses



Clustering de trajectoires

- Regrouper quoi ?
 - Des positions
 - De bouts de trajectoires
 - Des trajectoires



Clustering de trajectoires

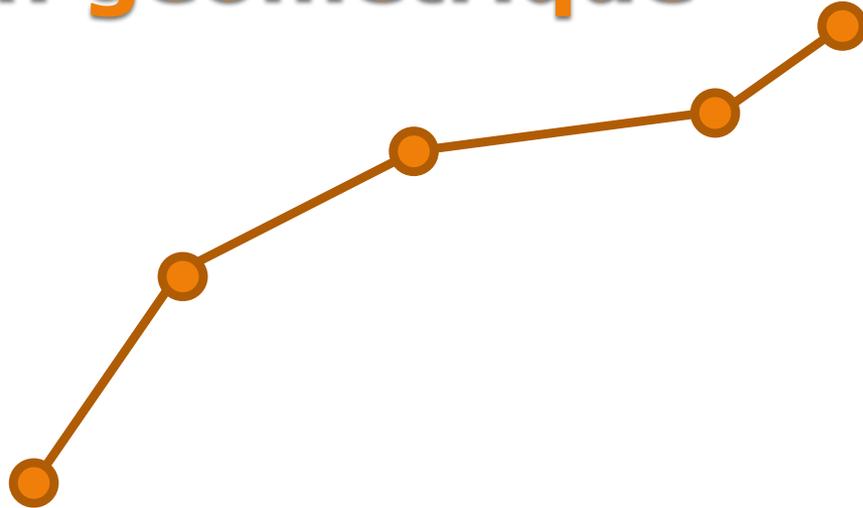
- Selon quels critères ?
 - **Proximité spatiale**
 - Distance entre trajectoires (lignes)
 - **Proximité spatiale et temporelle**
 - Position au « même instant »
 - **Densité**
 - Zones denses qui s'agrègent par voisinage

Problème de la similarité

- Distance
 - Spatiale
 - Temporelle
 - Spatio-temporelle (max de distances normalisées ?)
- À partir de
 - Représentation géométrique
 - Représentation symbolique
- Problème de **granularité** spatiale ou temporelle
 - Trajectoire avec des positions toutes les secondes et trajectoire avec des positions toutes les minutes
- Problème **d'échantillonnage**
 - Position à des instants différents
- Intégration de « **réseaux** » (contraintes fortes)
 - Physique (route, chemin de fer)
 - Graphe entre des POI

Représentation géométrique

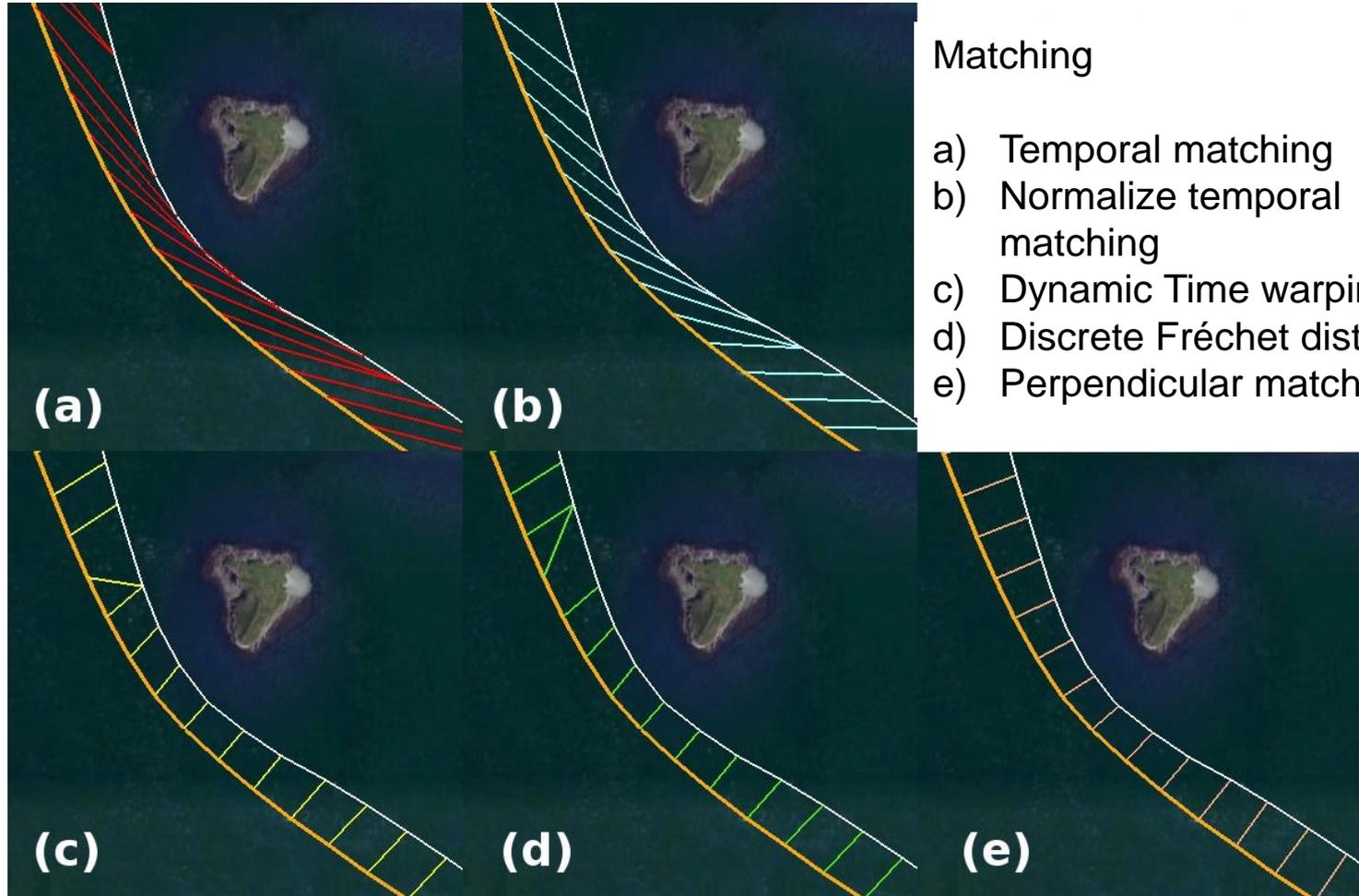
- **Objet**
 - Id, type, ...
- **Position**
 - Id, x, y, t
- **Trajectoire**
 - Suite ordonnée de positions d'un même objet
 - Problème
 - Quand commence et finit une « trajectoire » ?



Distances spatiales

- Distance entre des lignes (ensemble de points ordonnées)
 - Distance points ayant la même estampille temporelle (absolue ou relative)
 - Dynamic Time Warpping (DTW)
 - Distance de Fréchet discrète

Different matching processes



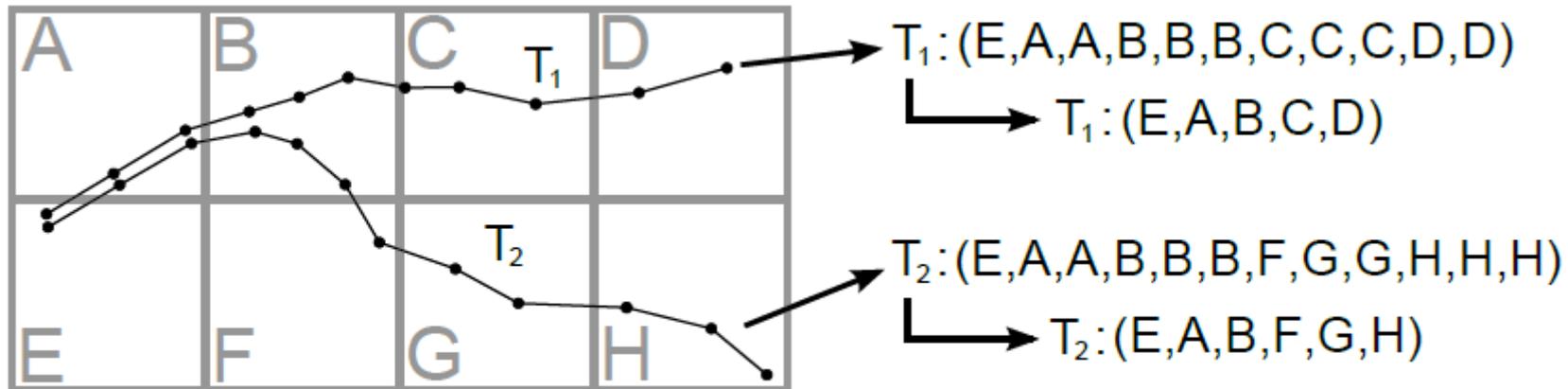
Matching

- a) Temporal matching
- b) Normalize temporal matching
- c) Dynamic Time warping
- d) Discrete Fréchet distance
- e) Perpendicular matching

Distance temporelle

- Si les Trajectoires Tr1 et Tr2 ont des route proches ($A \rightarrow B \rightarrow C$)
 - Mais **périodes (p) différentes**
 - Tr1.p : [8h15 – 9h30]
 - Tr2.p : [12h35 – 15h]
 - Mais **vitesses (v) différentes**
 - Tr1.v : 90km/h entre A et B puis 50 km/h entre B et C
 - Tr2.v : 20 km/h entre A et B puis 80 km/h entre B et C
- Comment mesurer la similarité (S)
 - entre périodes $S(\text{Tr1.p}, \text{Tr2.p})$
 - important pour convoie
 - entre vitesses $S(\text{Tr1.v}, \text{Tr2.v})$
 - DTW possible
 - Important pour détecter des comportements différents
- Actuellement peu ou mal pris en compte
 - Essentiellement
 - Comparaison des positions aux mêmes instants

Représentation symbolique



(a)

	A	B	C	D	E	F	G	H
A	0	0.5	1	1	0.5	0.7	1	1
B	0.5	0	0.5	1	0.7	0.5	0.7	1
⋮								

(b)

Distance d'édition

E	A	B	-	C	D
			-	×	×
E	A	B	F	G	H

$$\text{EditDistance}(T_1, T_2) = 1 + 0.5 + 0.5 = 2$$

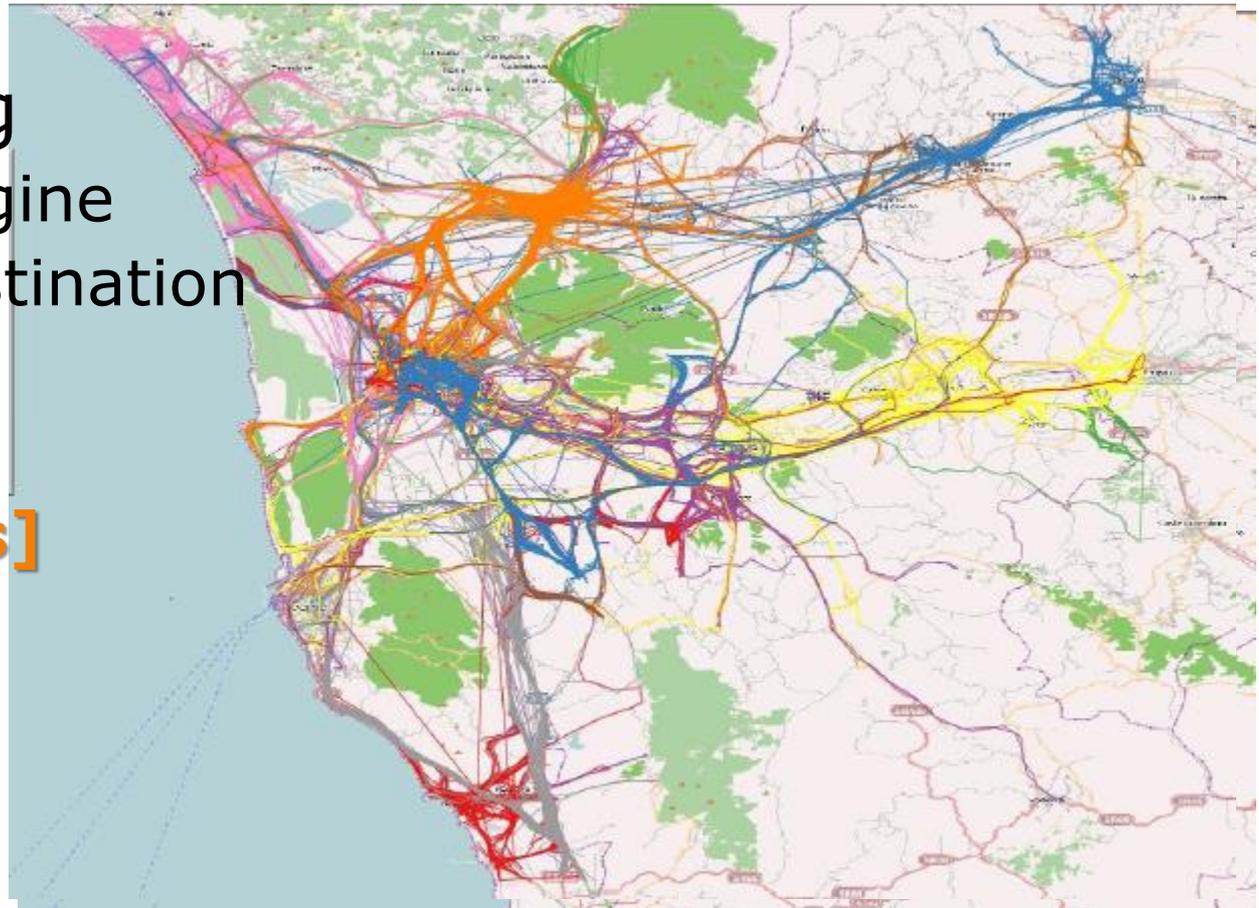
(c)

[Höferlin et al, 2011]

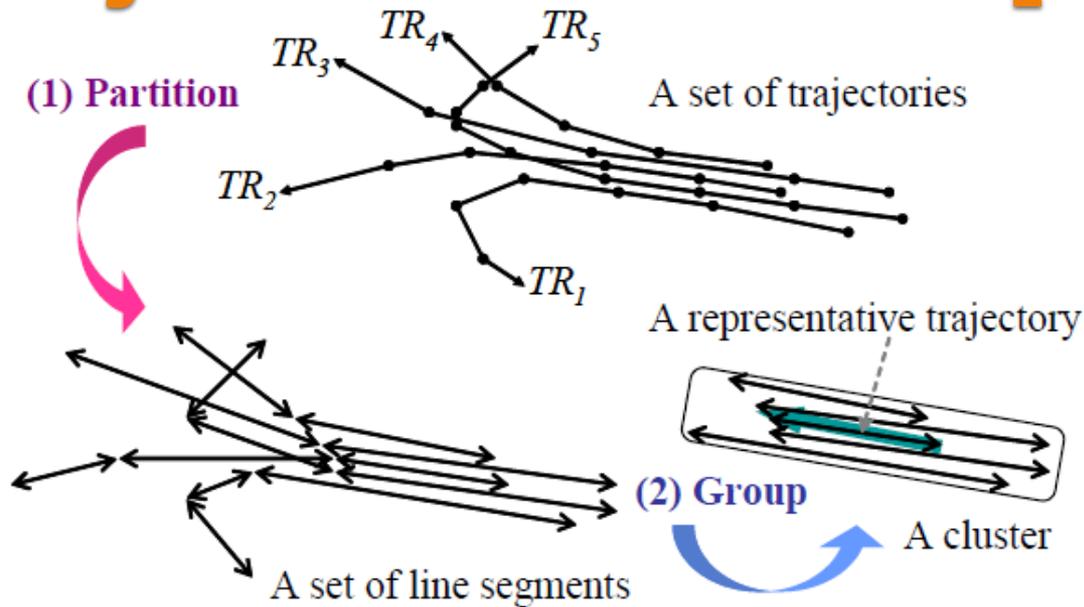
Figure 4: Processing example of two trajectories. Vector quantization and multiplicity removal (a), substitution cost matrix (b), edit distance applied to the trajectories T_1 and T_2 (c).

Exemple de Groupement de trajectoires : M-Atlas

- Clustering
 - Zone Origine
 - Zone Destination
 - densité
- [M-Atlas]

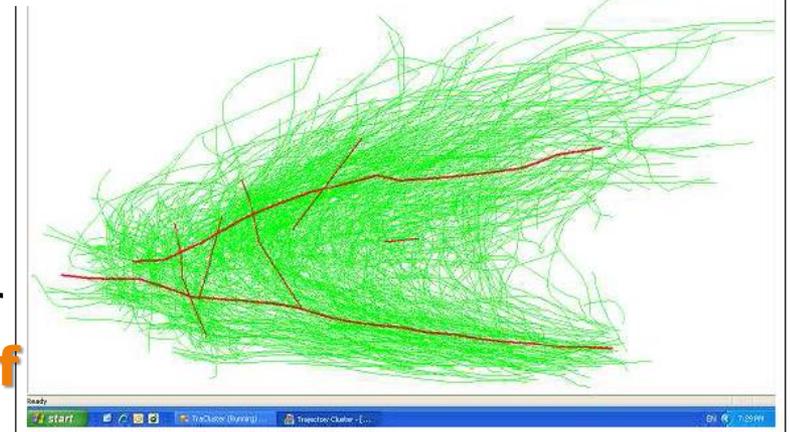


Exemple de Groupement de trajectoires : Traclus [Lee 2007]



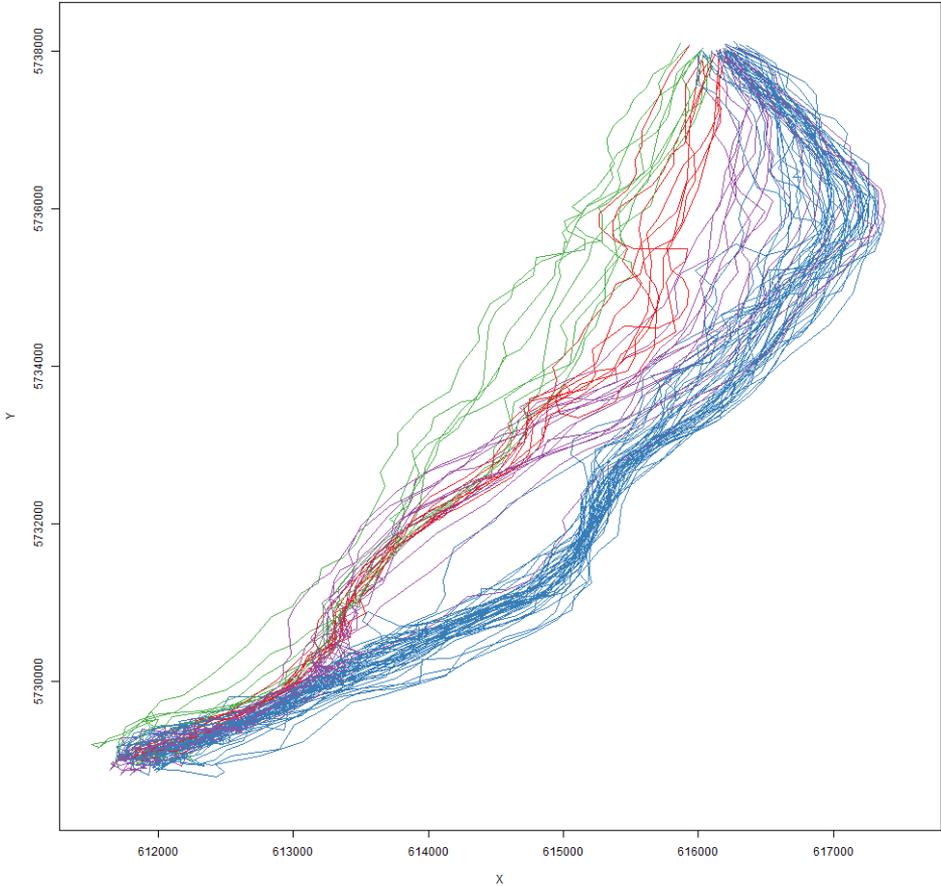
trajectories: thin green (bright in a B&W printing) lines
 representative trajectories: thick red (dark in a B&W printing) lines

- Simplification, partition
- Mesure de similarité
 - (longueur, angle, distance...)
- Regroupement
- Représentation de chaque cluster par un élément significatif : **Motif**



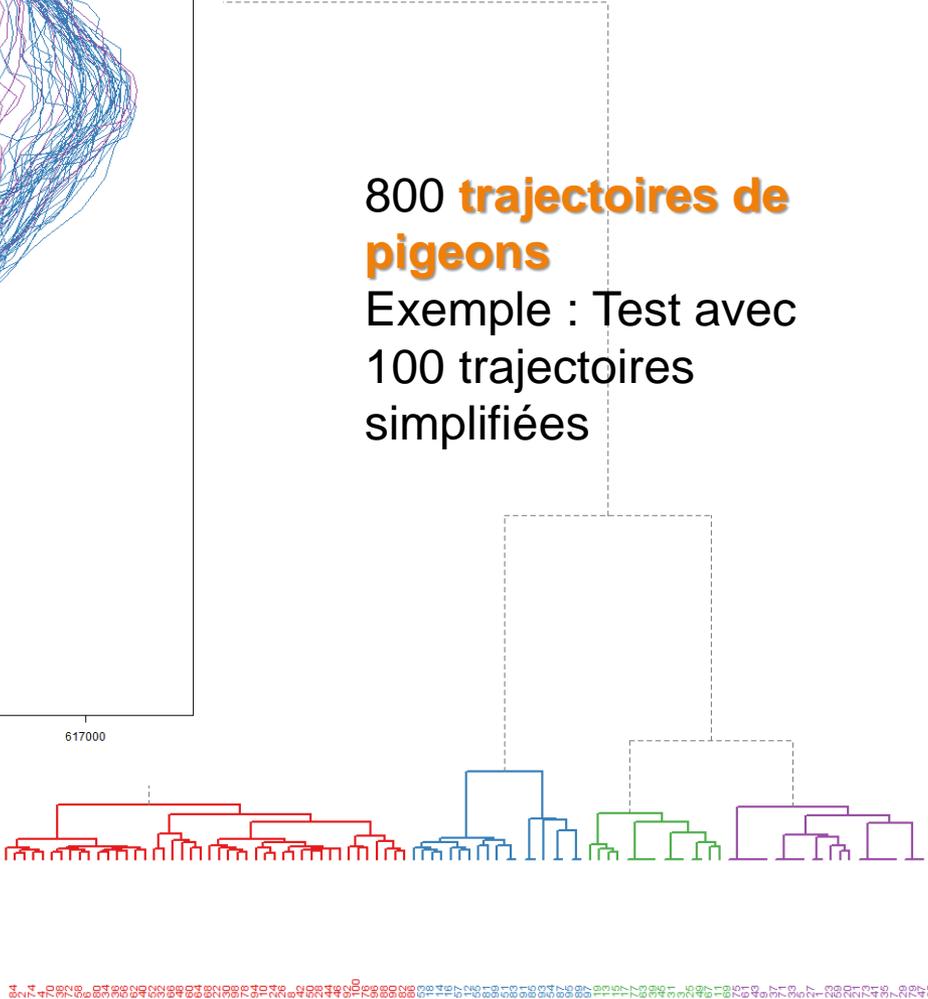
Classification hiérarchique Fréchet

Clusters using Fréchet distance

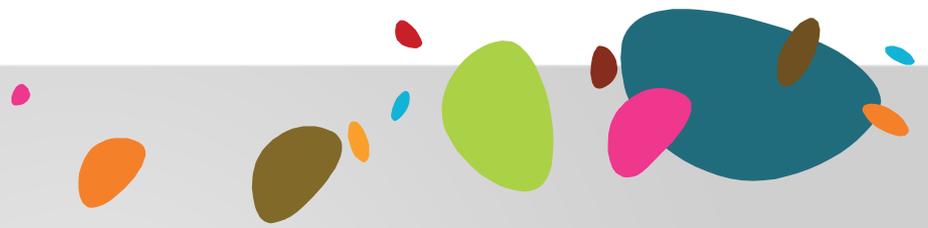


Fréchet distance dendrogram (4 groups)

800 **trajectoires de pigeons**
Exemple : Test avec 100 trajectoires simplifiées



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100



motifs

patterns



Définition de motifs (pattern)

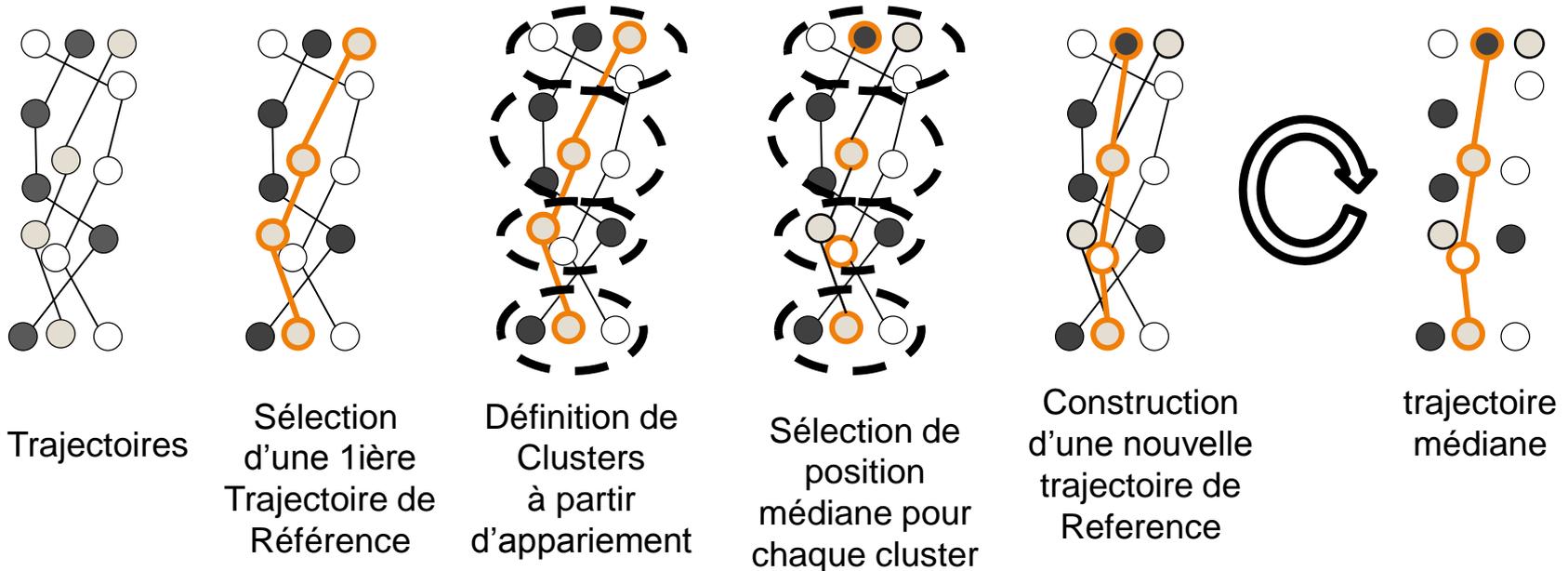
- Un motif d'un cluster de trajectoires
 - **résumer l'information** contenue dans un cluster
 - Centralité
 - moyenne, médiane, médoïde
 - Dispersion
 - écart type, quartile ...
 - Emprise
 - enveloppe convexe ...
- Pourquoi définir un motif ?
 - Pour faciliter la visualisation de gros jeux de données
 - Pour comparer des motifs
 - Pour comparer une trajectoire et un motif
 - Pour extraire des comportements d'un ensemble de trajectoires proches

Motifs

- Postulat
 - **Des objets de même type et d'un même cluster ont des comportements similaires quand ils se déplacent**
- Définition de motifs spatio-temporels pour aller d'un lieu A à un lieu B
 - **Approche médiane et non moyenne**
- Deux motifs définis
 - **Trajectoire médiane**
 - Pigeons
 - **Boîte à moustache spatio temporelle**
 - navires

Trajectoire médiane

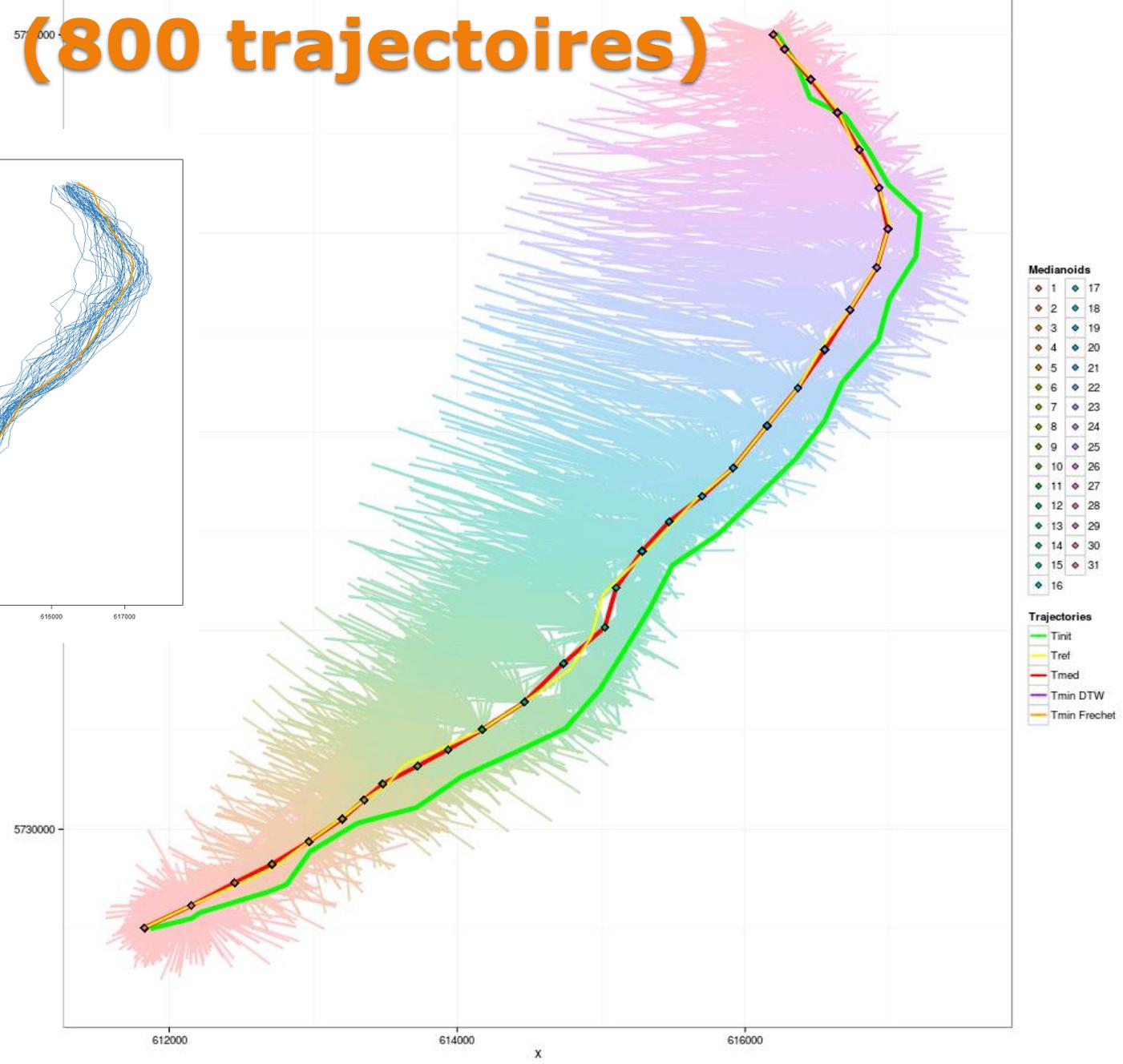
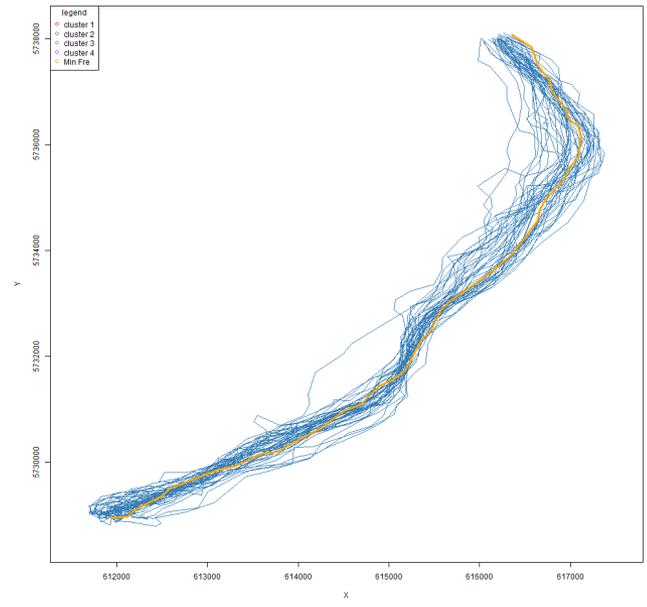
Principe



- **Appariement** basé sur la distance de Fréchet
- Basé sur des **positions réelles**
- Méthode **robuste** : outliers, boucles, volumes de données...
- Ajout d'une **estampille temporelle** relatif (durée depuis le départ)

Pigeons (800 trajectoires)

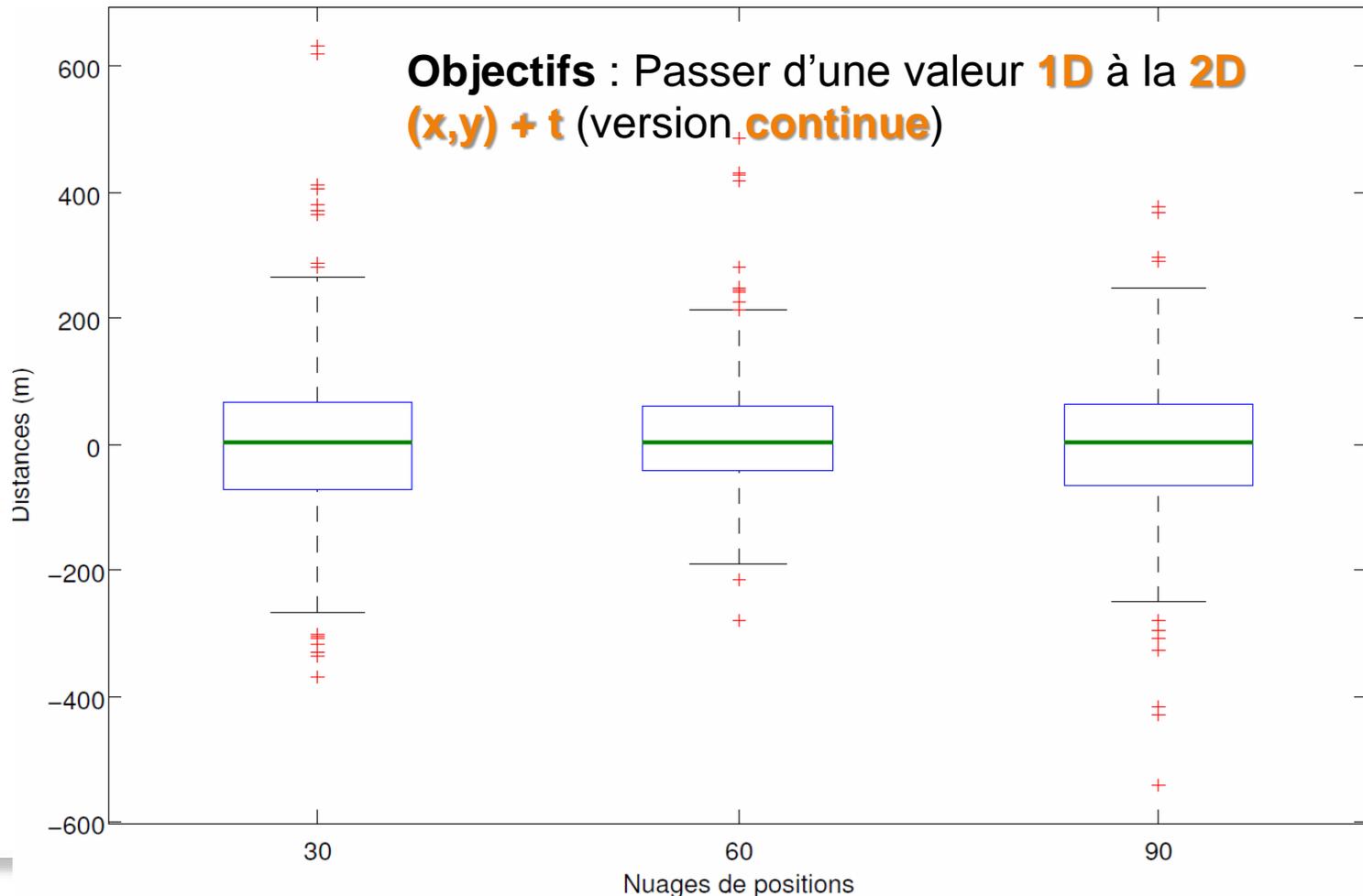
Trajectory Clusters 2



Boîte à moustache spatio-temporelle

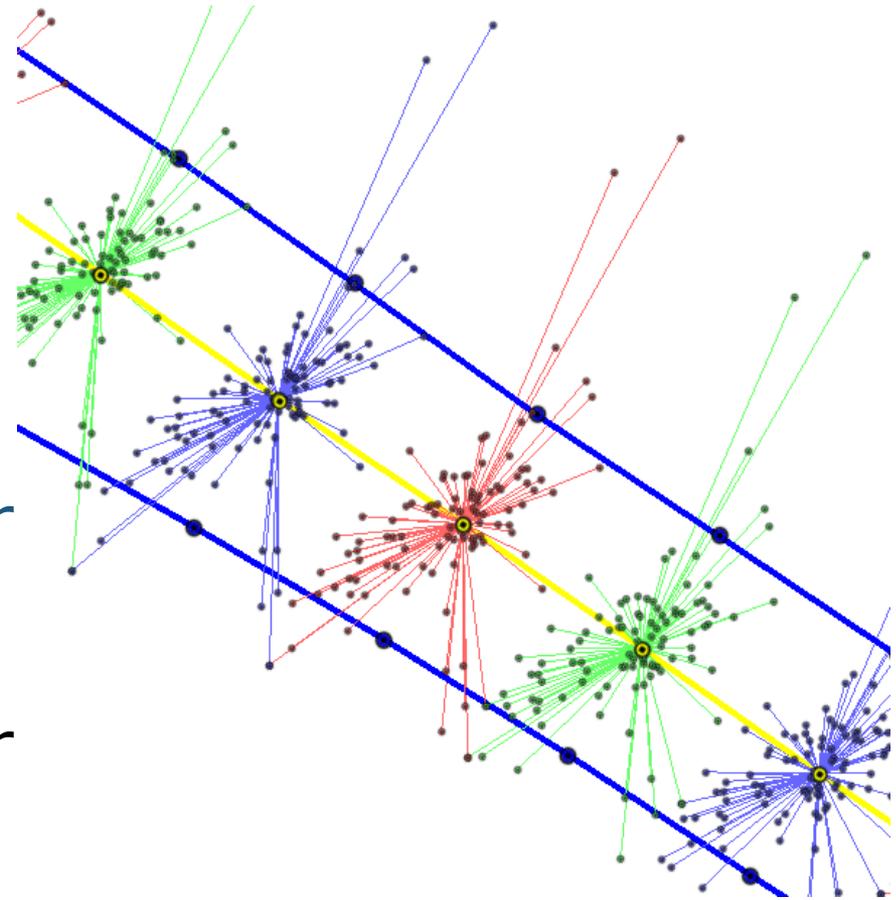
● Box plots classique (Boîte à moustache)

nuages à 3, 6, 9 km après le départ : distance à la médiane

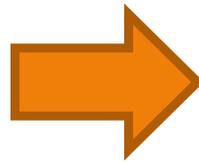


Boîte à moustache spatio-temporelle

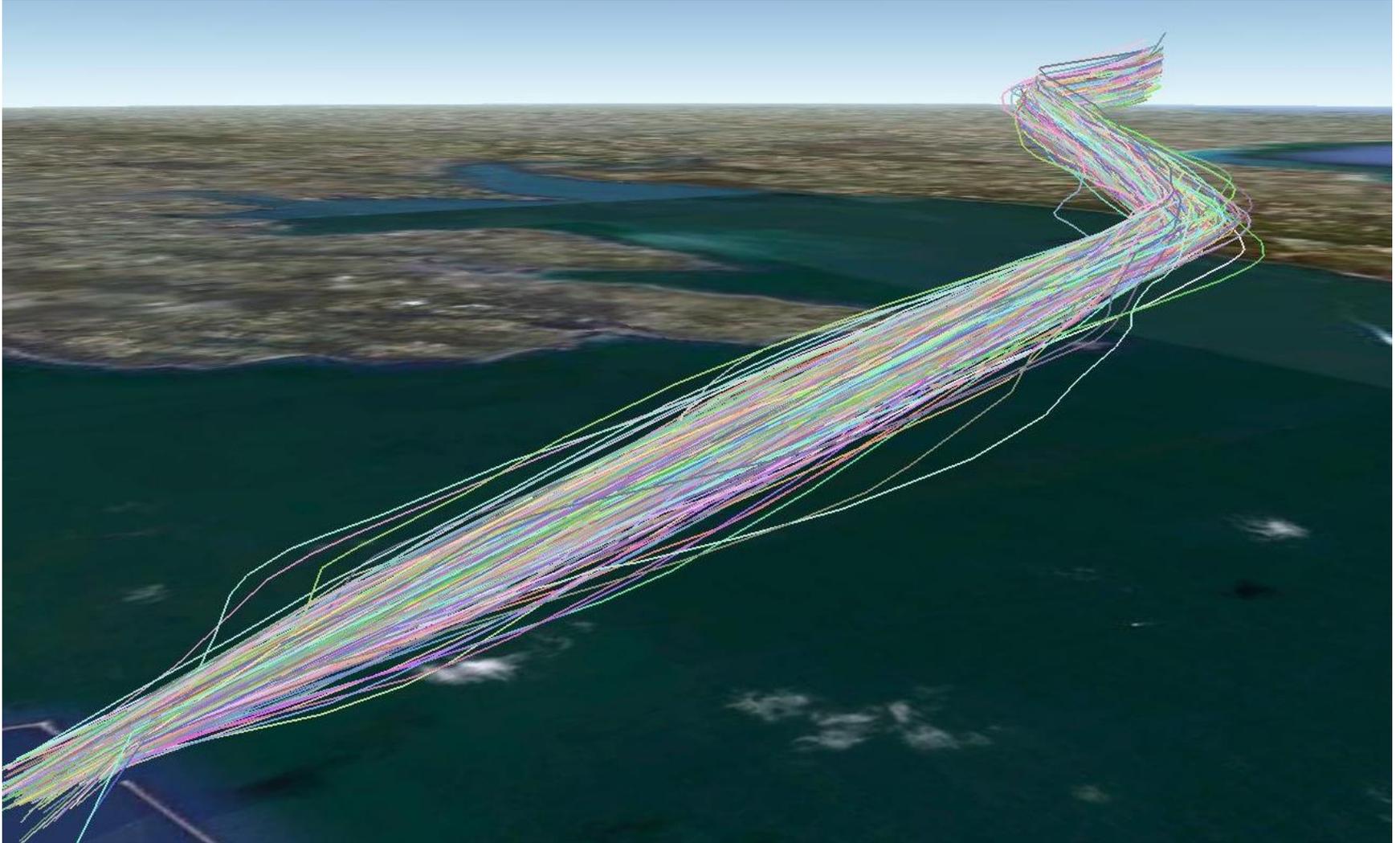
- A partir des nuages de positions de chaque position médiane
 - Définition d'un couloir spatial
 - Choix du 9^{ème} décile
 - Définition d'un couloir temporel
 - Choix du 9^{ème} décile



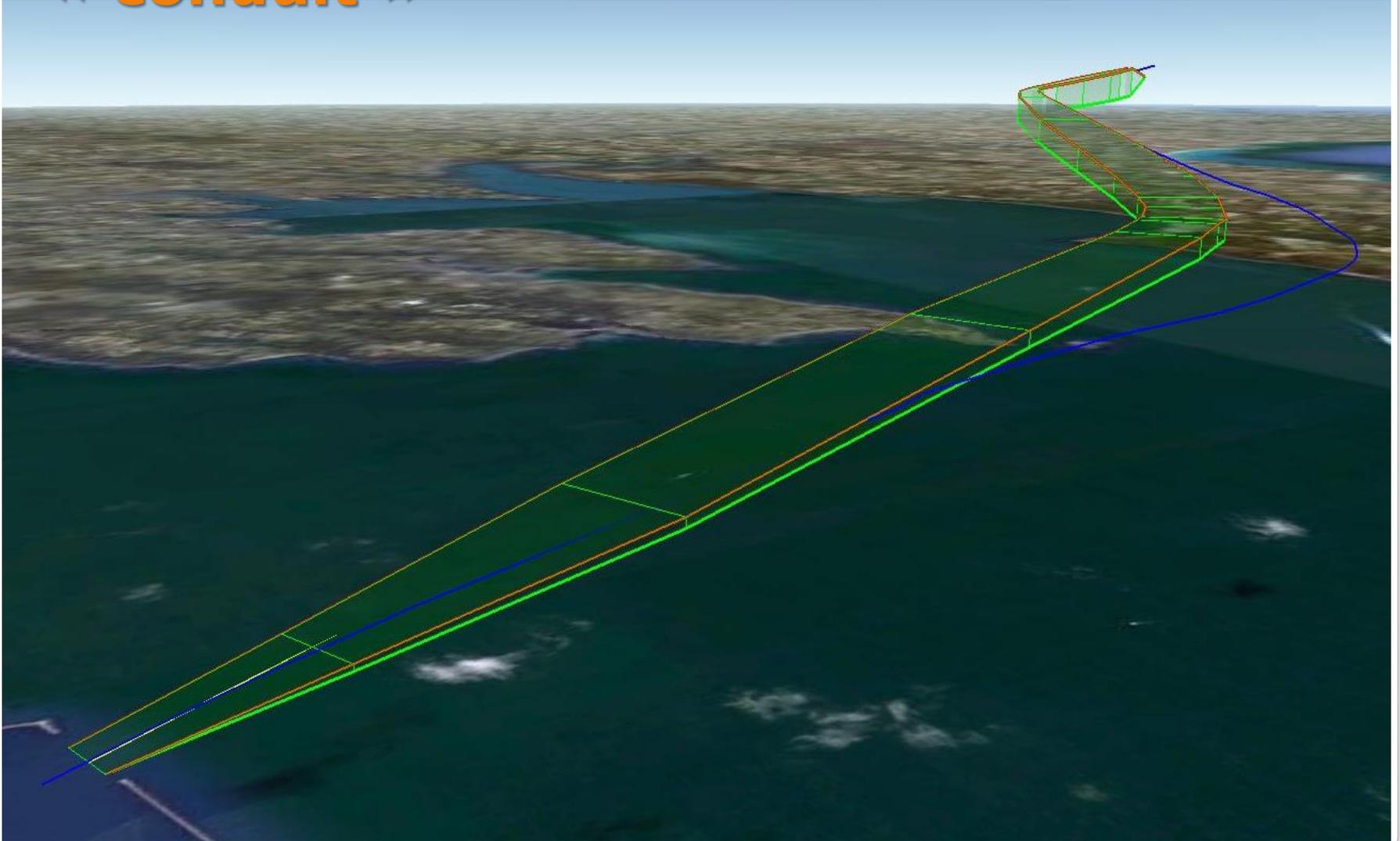
Boîte à moustache spatiale

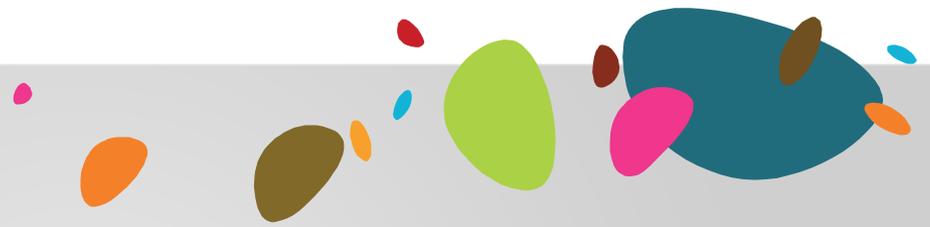


Représentation 3 D (2D + t)



Boîte à moustache spatio-temporelle « conduit »





Classification et détection d'outliers

Association à une classe
Détection de trajectoires inhabituelles

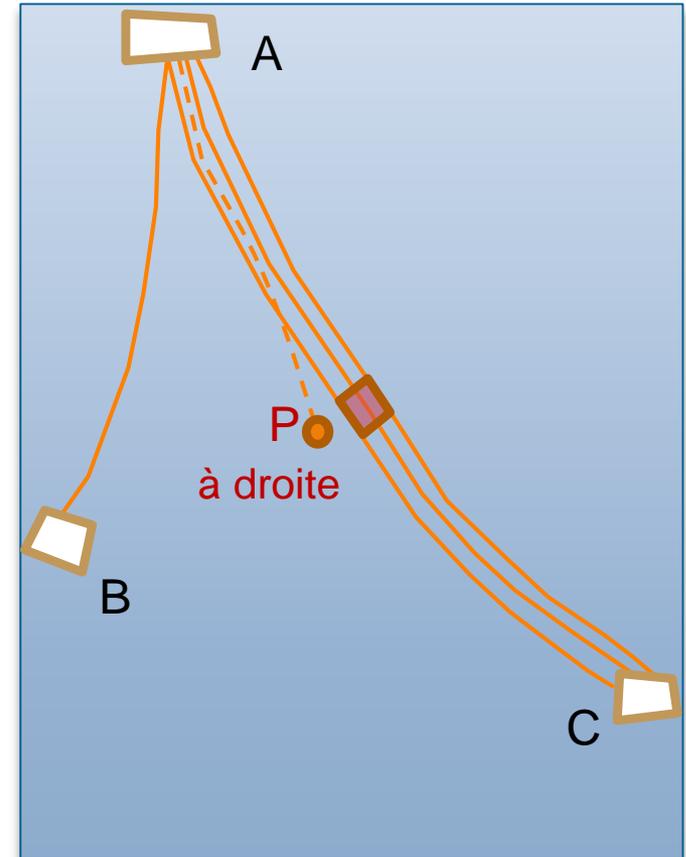


Classification

- Une nouvelle trajectoire, position
 - La classification consiste à l'associer à un
 - Cluster
 - Motif
 - La qualifier
 - De préférence en temps réel

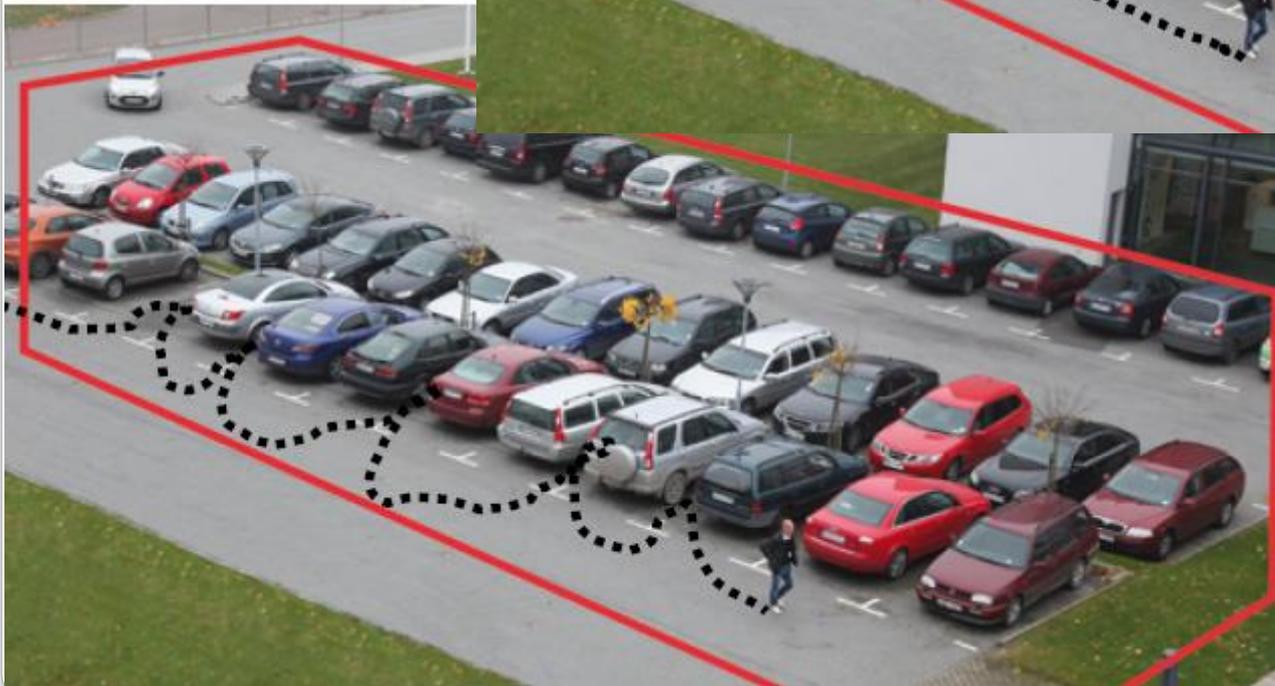
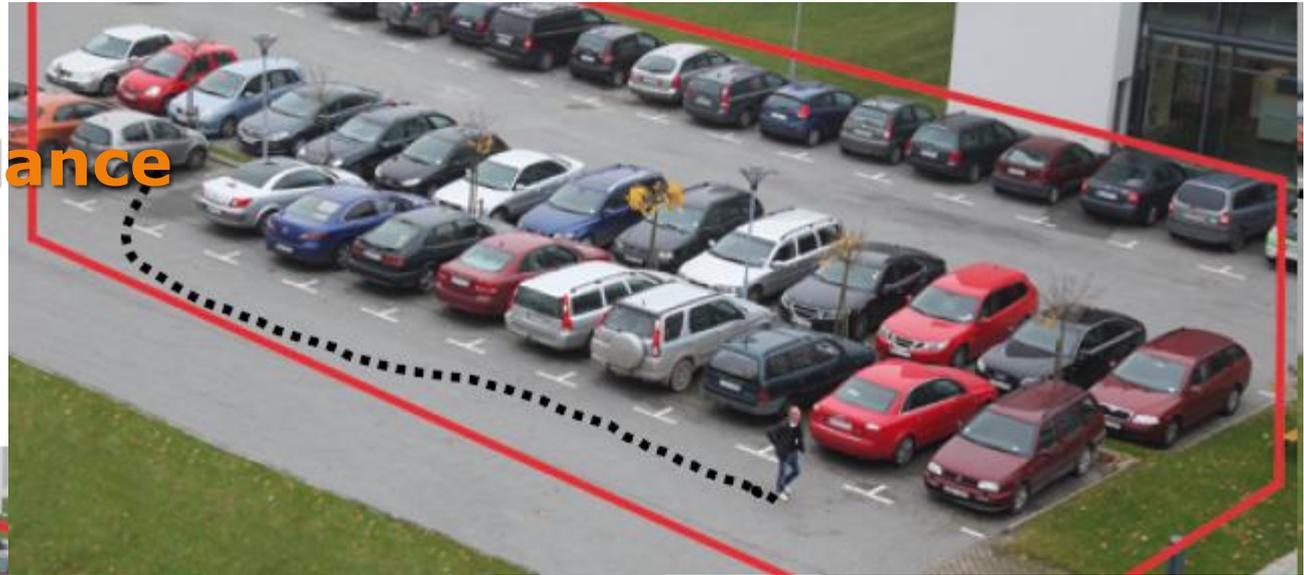
Classification de position: Qualification de positions

- Objectif
 - Pour chaque nouvelle position P d'un objet O, détecter à la volée si elle est habituelle ou non
- Méthode
 - Attacher P à un motif
 - Rechercher la dernière zone du graphe traversée : A
 - Apparier la trajectoire partielle entre A et P avec une trajectoire médiane partant de A : AC
 - Partitionner l'espace en fonction de la route type et du t relatif
 - **Qualification** de la position
 - 5 qualifications possibles :
 - Normale
 - En retard
 - En avance
 - À droite
 - À gauche



Détection d'outliers

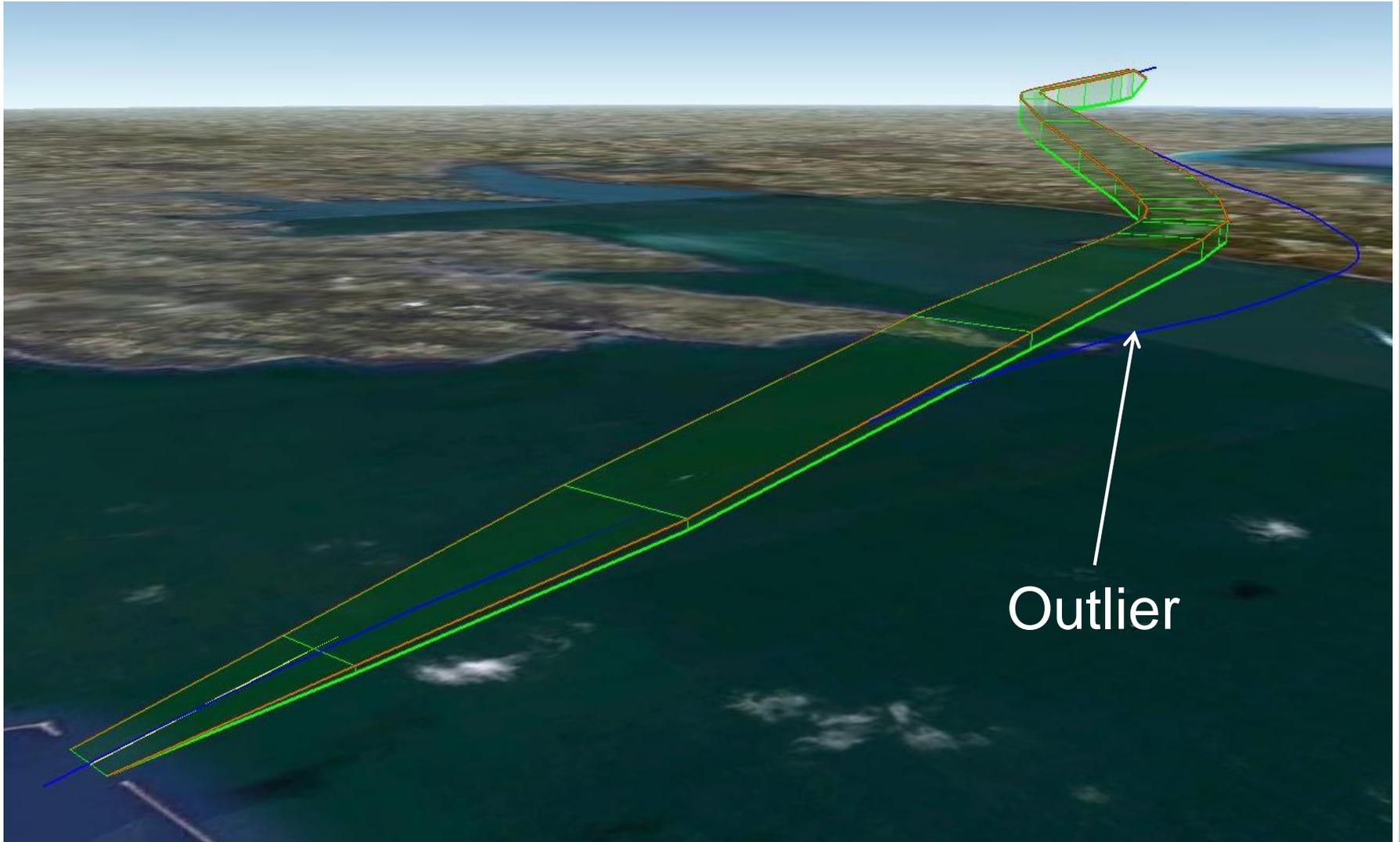
- Exemple
 - Surveillance vidéo

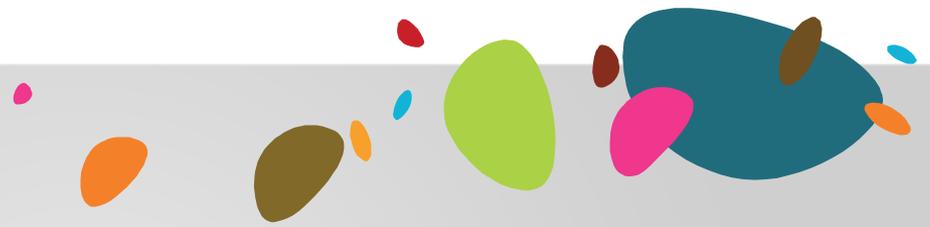


Détection d'outliers

- Besoin de mesures de similarité entre un motif et une trajectoire
- Grand nombre de mesures possibles
 - Besoin de techniques pour les agréger
 - Logique floue

Détection d'outliers





Conclusion



Fouille de trajectoires

- Présentation des travaux essentiellement en fouille de flux de trajectoires
 - Clusters, motifs, outlier
- **Perspectives**
 - Déplacements relatifs
 - Évitements, rapprochements, synchronisation ...
 - Extraction de comportements (profil)
 - Piéton → Piéton qui flâne, piétons qui téléphone, Piéton qui fait du lèche-vitrine, coureur à pieds ...
 - Etude de l'appropriation de l'espace par les objets mobiles
 - Prédiction
 - Etudes des ruptures de comportement

Conclusion

● Perspectives plus long terme

- Analyse des déplacements multi-modaux
 - complexe
- Prise en compte de phénomènes physiques
 - pluie ↔ piétons
 - marée ↔ navires
- Représentation visuelle multi-échelles
 - Clusters de clusters
 - Différents niveaux
 - Véhicule > Véhicule lourd > Bus
 - cubes de géovisualisation (Spatial & Temporal OLAP)

Jeux de données

- MoveBank

- https://www.movebank.org/panel_embedded_movebank_webapp

- Essentiellement des trajectoires d'animaux

- Chorochronos

- <http://www.chorochronos.org/>

- Jeux de données et algorithmes

- Edinburgh Pedestrian Database

- <http://homepages.inf.ed.ac.uk/rbf/FORUMTRACING/>

- 92 000 trajectoires de piétons sur une place

plateformes

- M-Atlas

- <http://www.m-atlas.eu/>

- MoveMine

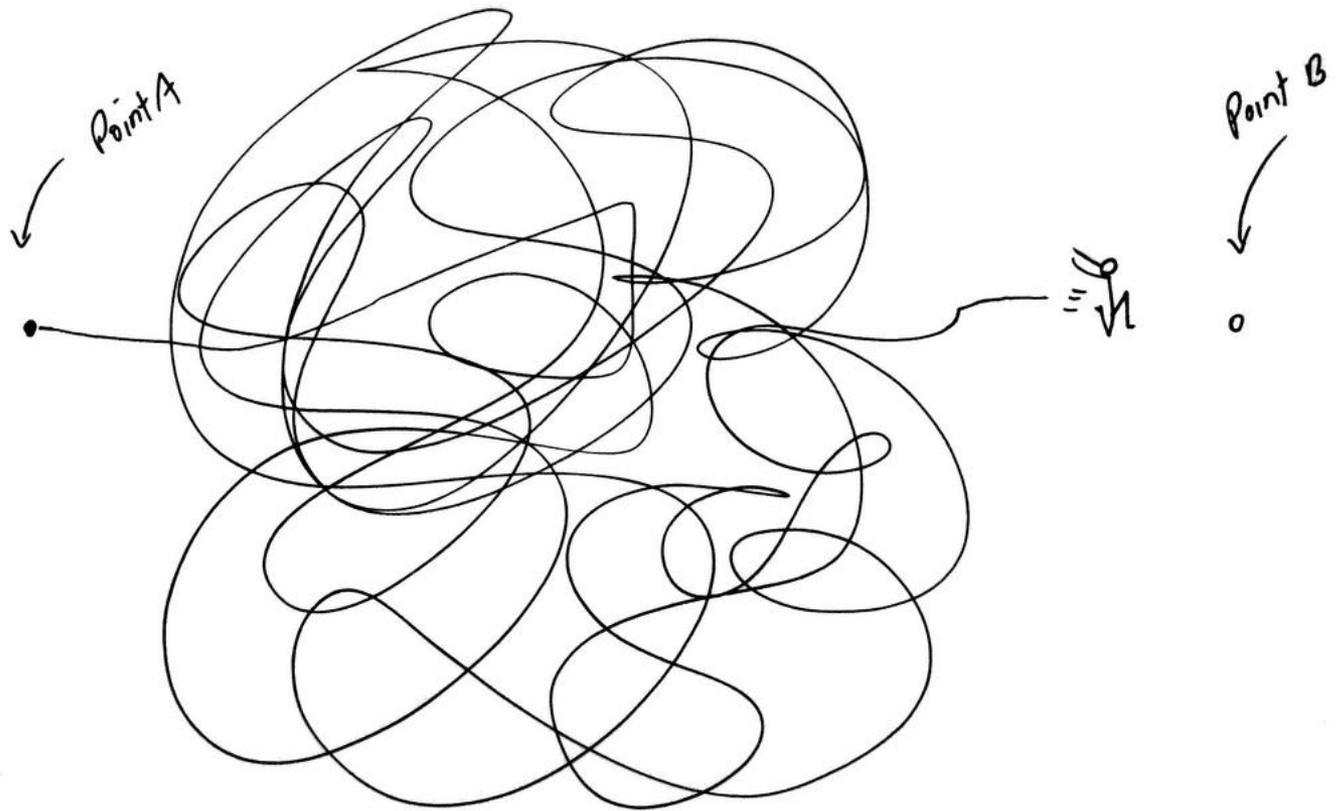
- <http://web.engr.illinois.edu/~klei2/movemine/>

- Multi-Move

- <http://www.lirmm.fr/~phan/multimove.jsp>

Références

- **Data Mining: Concepts and Techniques**, de Han, Kamber et Pei, Morgan Kaufmann (2011)
- **Spatio-temporal clustering**, de S. Kisilevich, F. Mansmann, M. Nanni, S. Rinzivillo (2010), *Data Mining and Knowledge Discovery Handbook*, pp 855-874
- **Mobility data**, de Renso, Spaccapietra et Zimányi, Cambridge University Press (2013)
- **Mobility Data Management and Exploration**, de Pelekis et Theodoridis, Springer (2014)
- **Clustering de trajectoires d'objets mobiles**, Mohamed Khalil EL MAHRSI, EGC 2010



**MERCI DE VOTRE
ATTENTION**