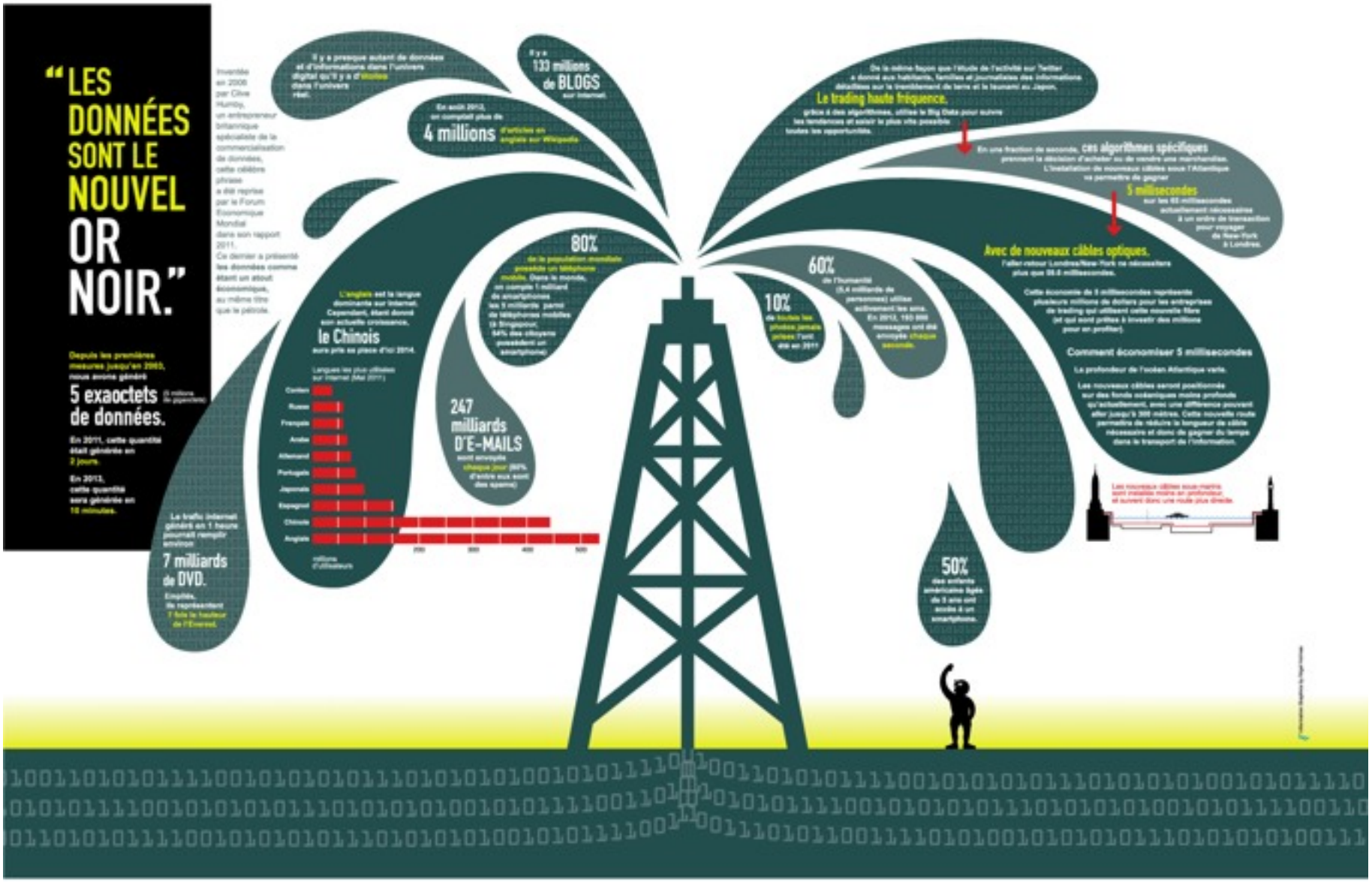
INSTITUT de MATHEMATIQUES de TOULOUSE

cimi Centre International de Mathématiques et d'Informatique TOULOUSE

Mathématiques et



Infographie EMC 2012

1970s KO Analyse des données (multivariate analysis)

•1980s MO IA, réseaux de neurones, Stat. fonctionnelle

•1990s GO 1er changement de paradigme : **Data Mining** Modèles prédictifs et données pré-acquises

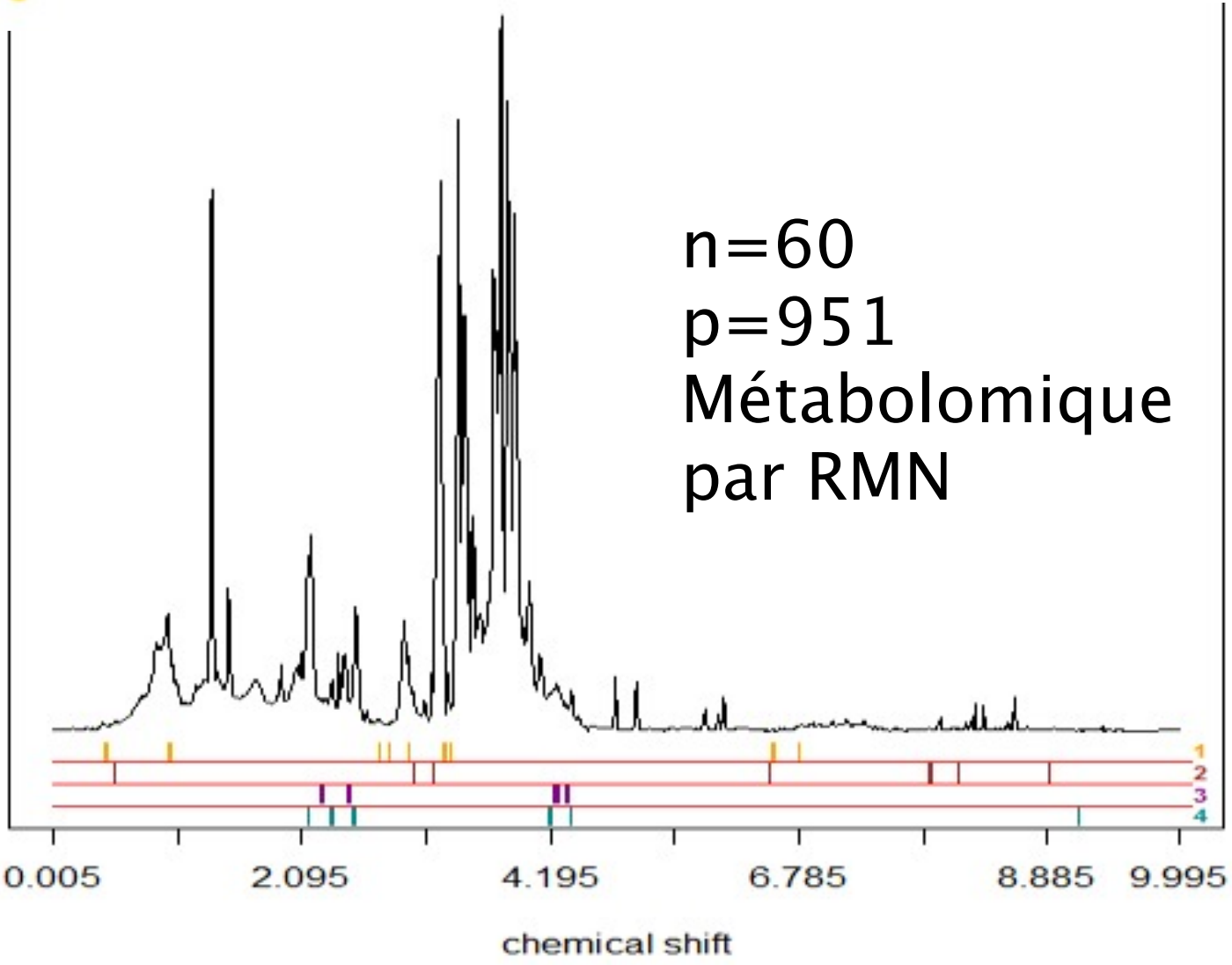
•2000s TO 2ème chang. de paradigme : **Apprent. Statistique** p>>n , parcimonie, Biais^2 + Variance

•2010s PO 3ème chang. de paradigme : **Big Data**

E(approximation)+E(estimation)+E(optimisation)

Apprentissage Statistique et omiques

- p>>n (p#10^4, n#10)
- Tests multiples
- Sélection de modèles par pénalisation
- Sélection de variables (biomarqueurs)
- SVM, boosting, random forest…
- Inférence de graphes


n=60
p=951
Métabolomique
par RMN
1
2
3
4
0.005
2.095
4.195
6.785
8.885
9.995
chemical shift

Applications du Big Data

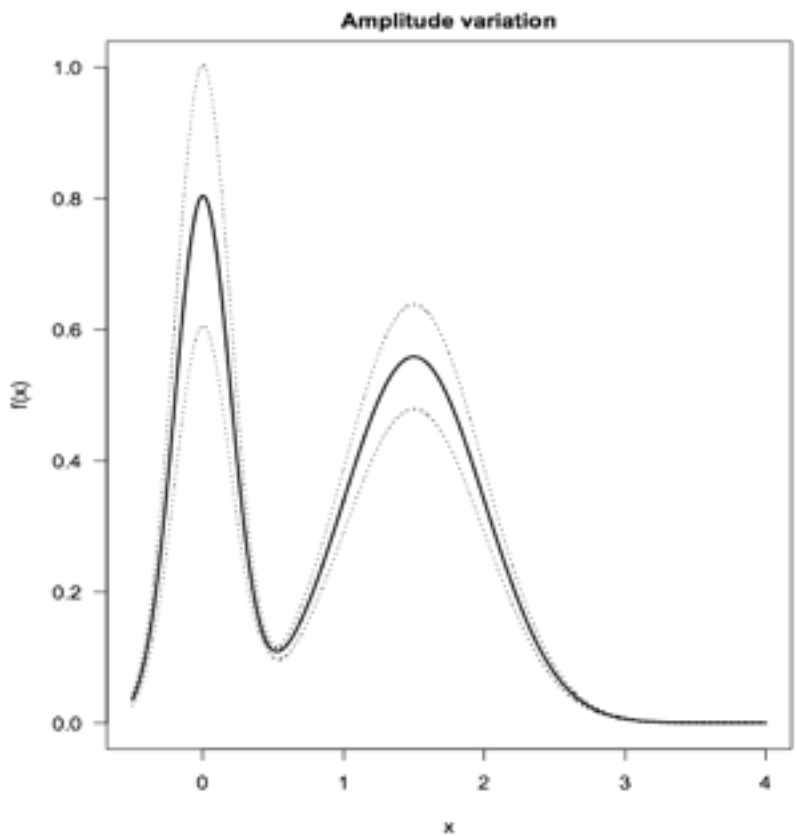
- E-commerce (1000 merci, tinyclues…)
- Géolocalisation (Médiamobile, Datasio…)
- Industries (Total, Geosys…)

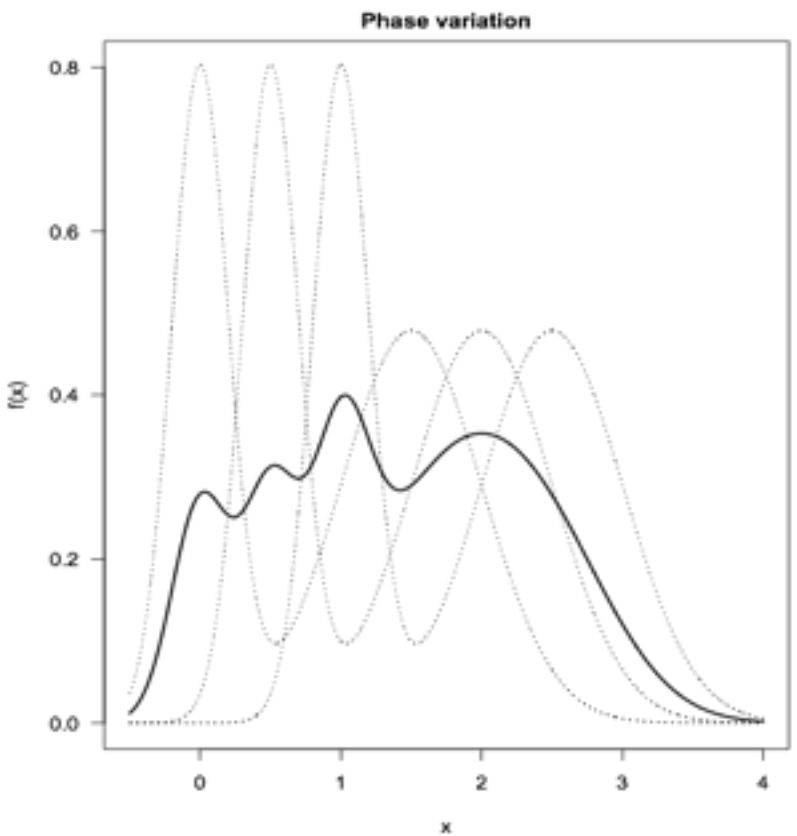
- Structures complexes des données industrielles

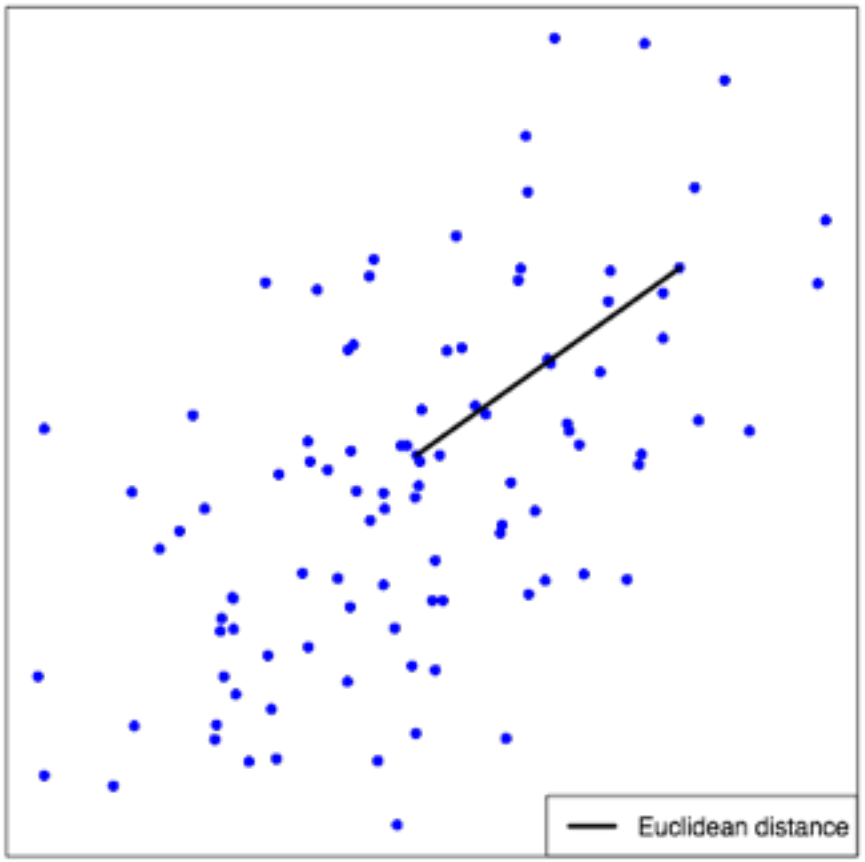
  graphes, signaux, images, fonctions…
- Vitesse d'acquisition (VVV)

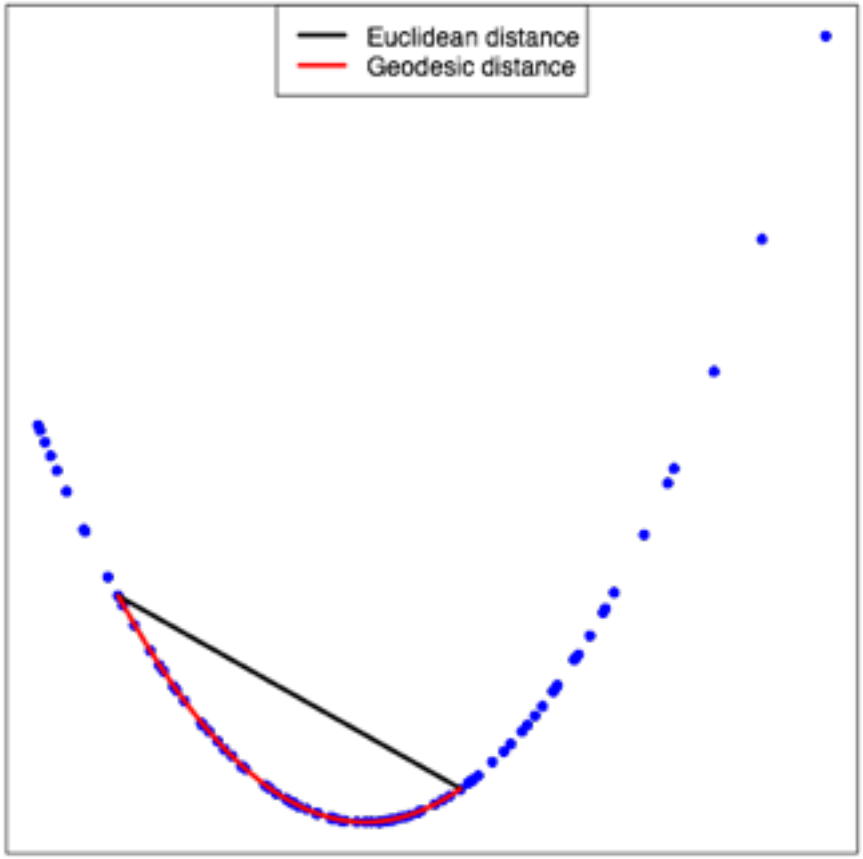
Problèmes mathématiques et/ou Informatiques

- E(Approximation)+E(Estimation) +E(Optimisation)
- Déterminer La structure mathématique « géométrie » ou distance adaptée
- Trouver des invariants ou comportements reproductibles
- Vitesse d'acquisition vs. Décision séquentielle ou adaptative

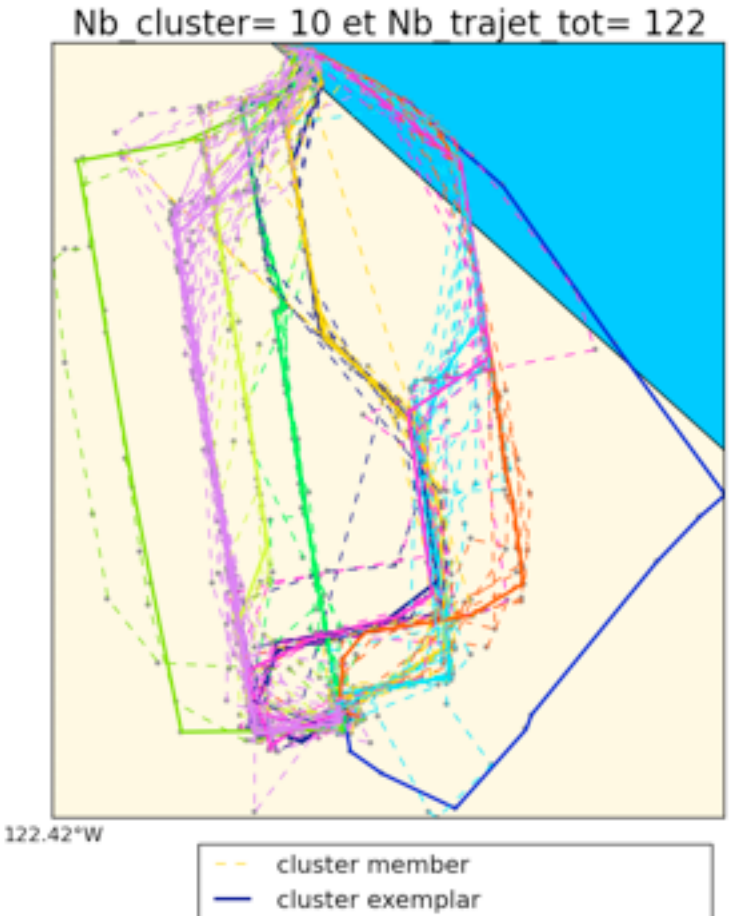

Amplitude variation
1.0
0.8
0.6
0.4
0.2
0.0
f(x)
0
1
2
3
4
x


Phase variation
0.8
0.6
0.4
0.2
0.0
f(x)
0
1
2
3
4
x

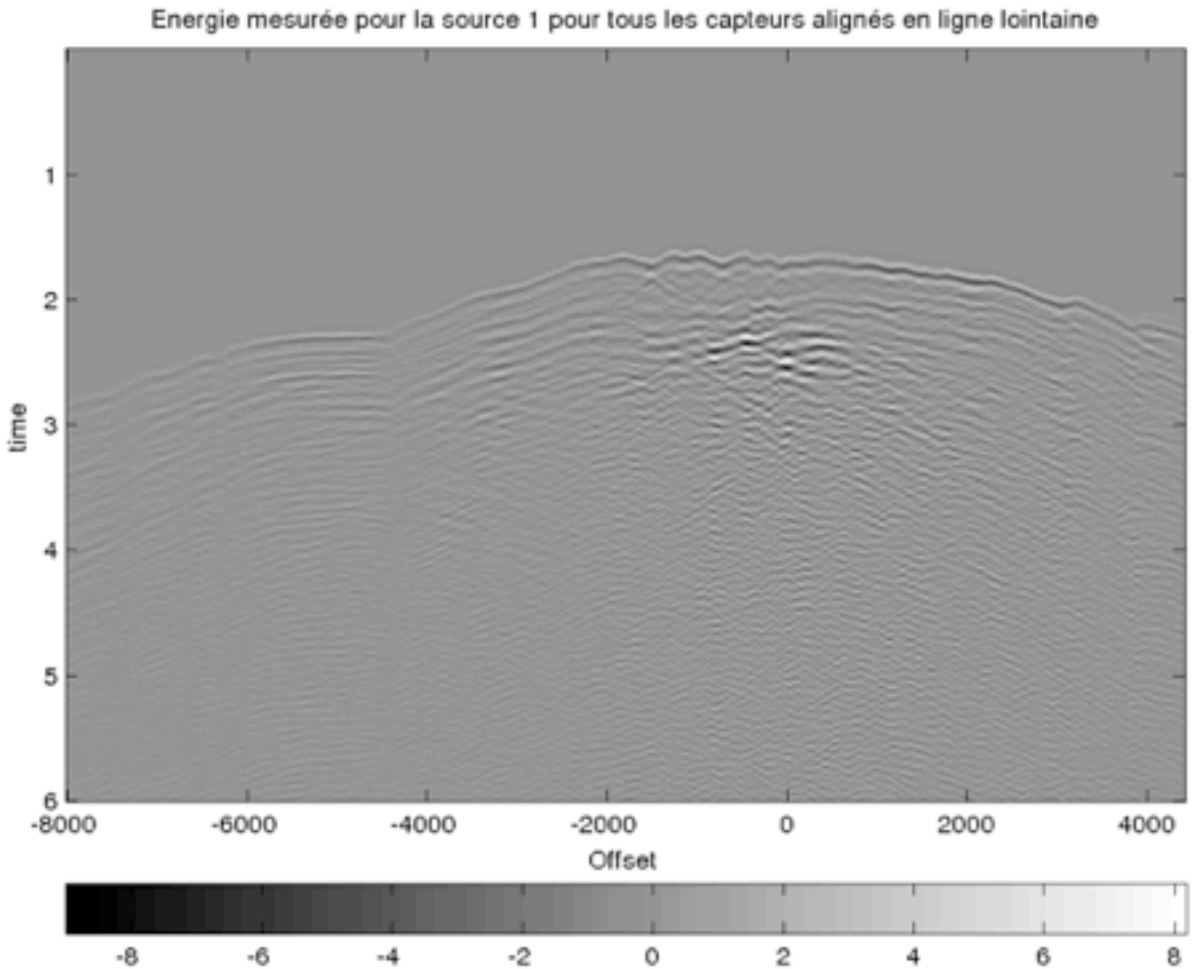

Euclidean distance


Euclidean distance
Geodesic distance

Géolocalisation des taxis à SF



Caractérisation de simulations de tirs



Systèmes de recommandation multi agents

O'Reilly

Mathématiques utilisées 1/2

- **Géométrie** : trouver le bon espace ou un modèle de déformations
- **Statistique** : estimer la distance géodésique
- **Théorie des graphes** : estimation de la distance en trouvant les « chemins minimaux » entre les points
- **Optimisation** : calculer la distance minimale

Mathématiques utilisées 2/2

- **Clustering** : trouver des groupes similaires au sens de la distance
- **Statistique** : étudier la variabilité, plans d'expérience dynamiques, numériques
- **Apprentissage** (on line): prévoir l'appartenance au groupe, apprentissage renforcé
- **Probabilités** : analyse de graphes, matrices aléatoires, chaines de Markov…

Conclusion très provisoire

- Beaucoup de sous-disciplines des mathématiques et d'interactions
- Connexions Mathématiques / Informatique
- Problème industriel = problème de recherche