

Modèle statistique du trafic urbain sur un réseau: application à l'estimation à partir de données GPS



Projet Mobile Millennium
<http://traffic.berkeley.edu>

Aude Hofleitner
Thèse Grettia/UC Berkeley
<http://eecs.berkeley.edu/~aude>

Données et systèmes d'information du trafic

Traditionnellement:

Données:

- Infrastructure dédiée (boucles, radars)
- Coût d'installation et de maintenance
- Limitation de la couverture du réseau

Systèmes d'information:

- Centres d'appel, PMV, internet
- Peu de disponibilité pendant le trajet
- Peu de capacités de prédiction



Données et systèmes d'information du trafic

Traditionnellement:

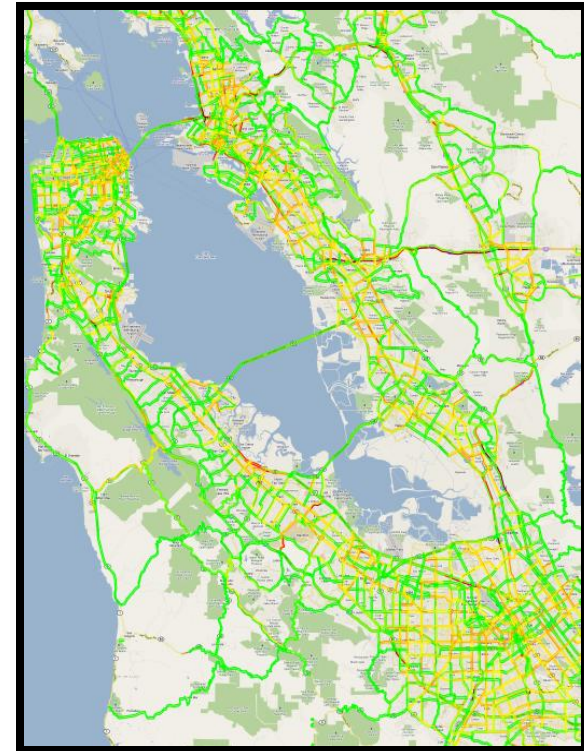
Données:

- Infrastructure dédiée (boucles, radars)
- Coût d'installation et de maintenance
- Limitation de la couverture du réseau



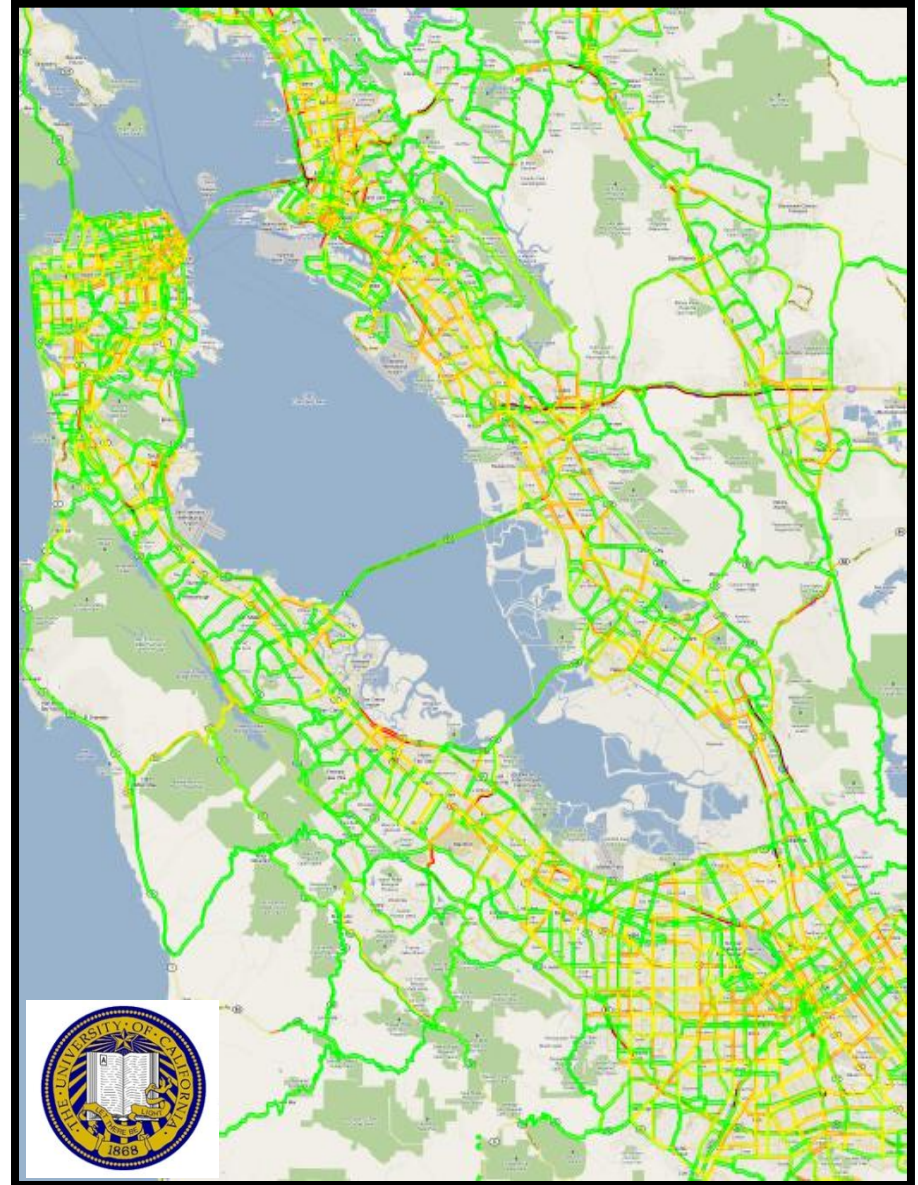
Systèmes d'information:

- Centres d'appel, PMV, internet
- Peu de disponibilité pendant le trajet
- Peu de capacités de prédiction



Nouvelle génération:

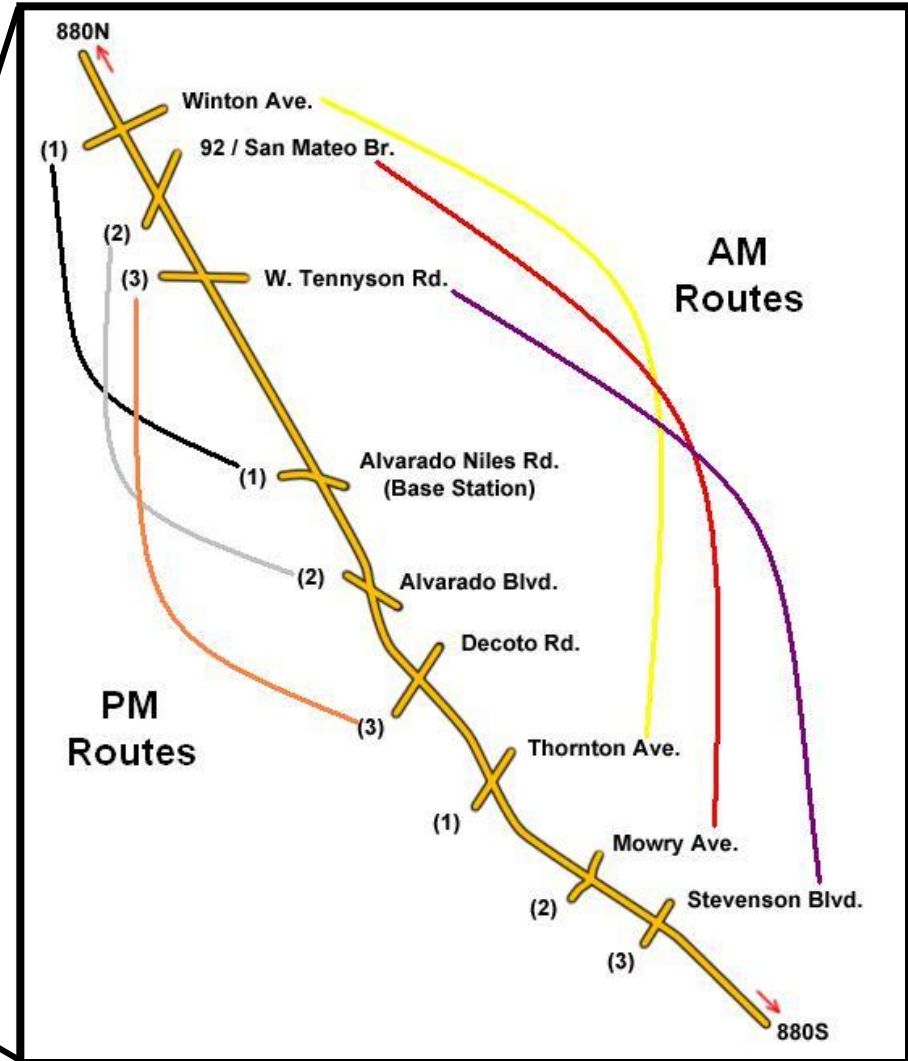
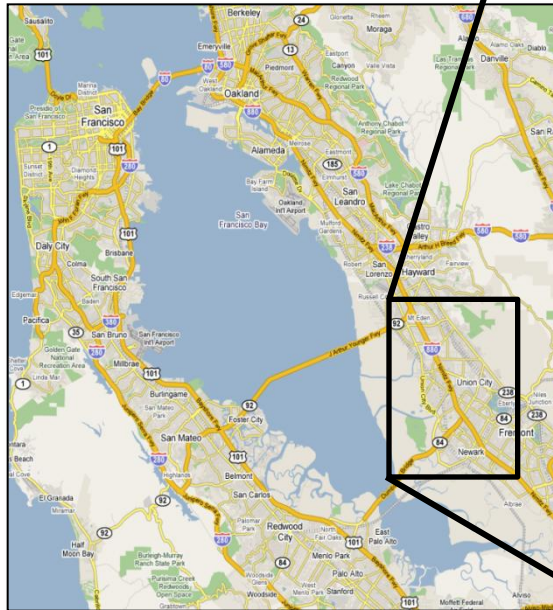
- Utilisation du GPS et capacités de communication
- Information en temps réel sur téléphones portables,
- Estimation et prédiction sur l'ensemble du réseau
- Intégration avec des systèmes de choix d'itinéraires.



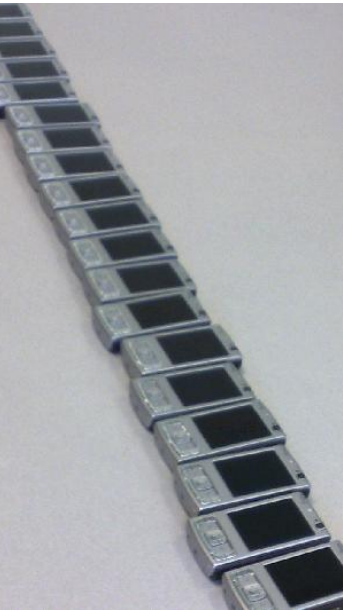
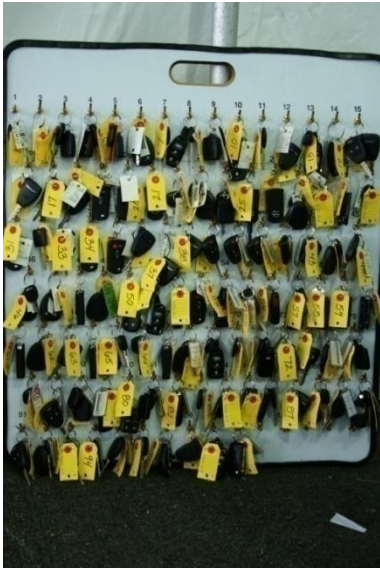
Réseau autoroutier (Mobile Century)

Preuve expérimentale de la possibilité d'utiliser des téléphones portables pour estimer le trafic

- 8 février 2008
- I80, Union City, Californie
- 100 véhicules, 165 étudiants
- 10 heures de test
- Réseau de 16 kilomètres
- Taux de pénétration: 2% - 5%



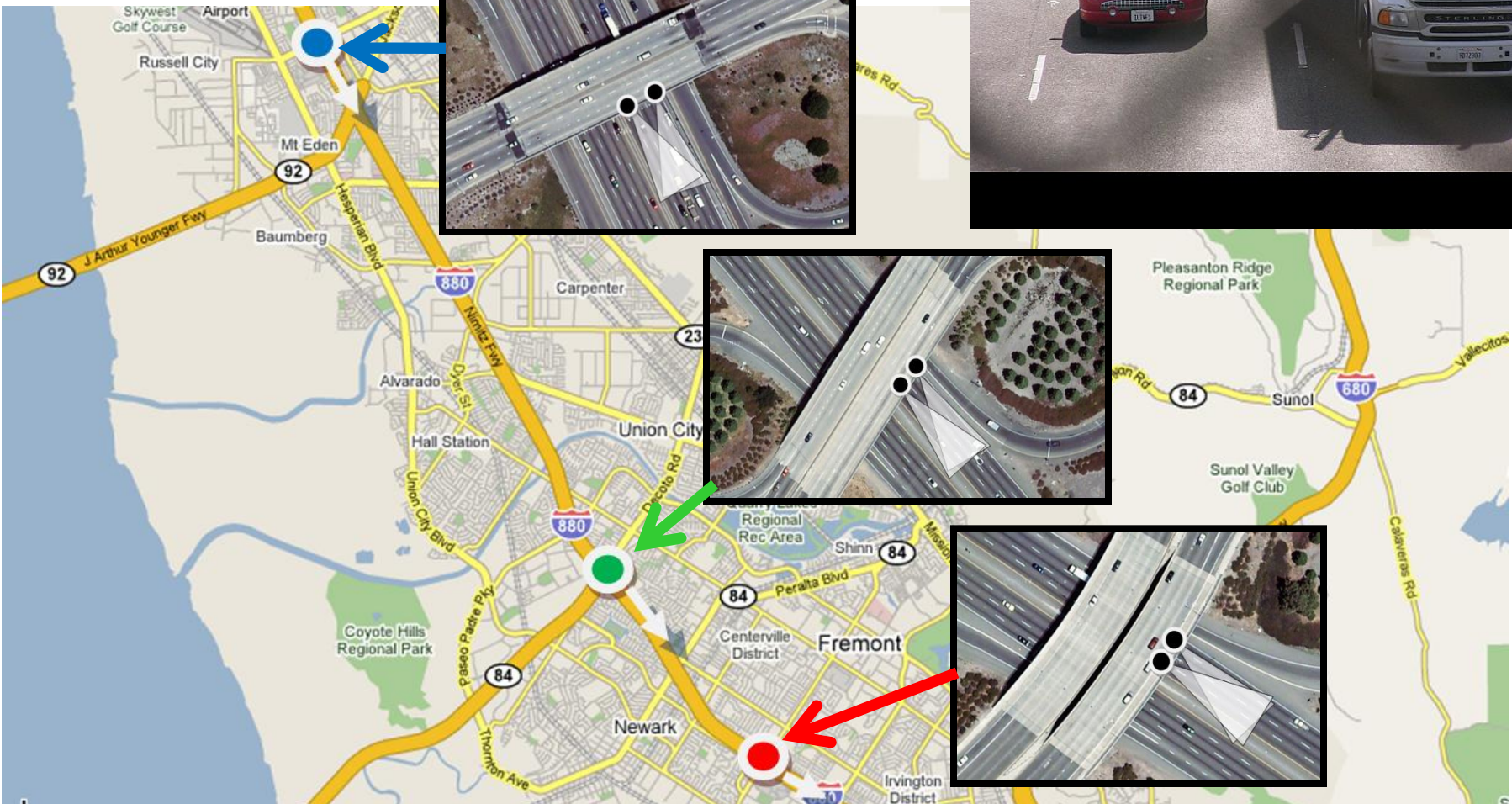
Un aperçu de *Mobile Century*



Validation du test grâce à des caméras

Données vidéo collectées:

- Comptes de véhicules
- Temps de parcours



Un aperçu des données et résultats

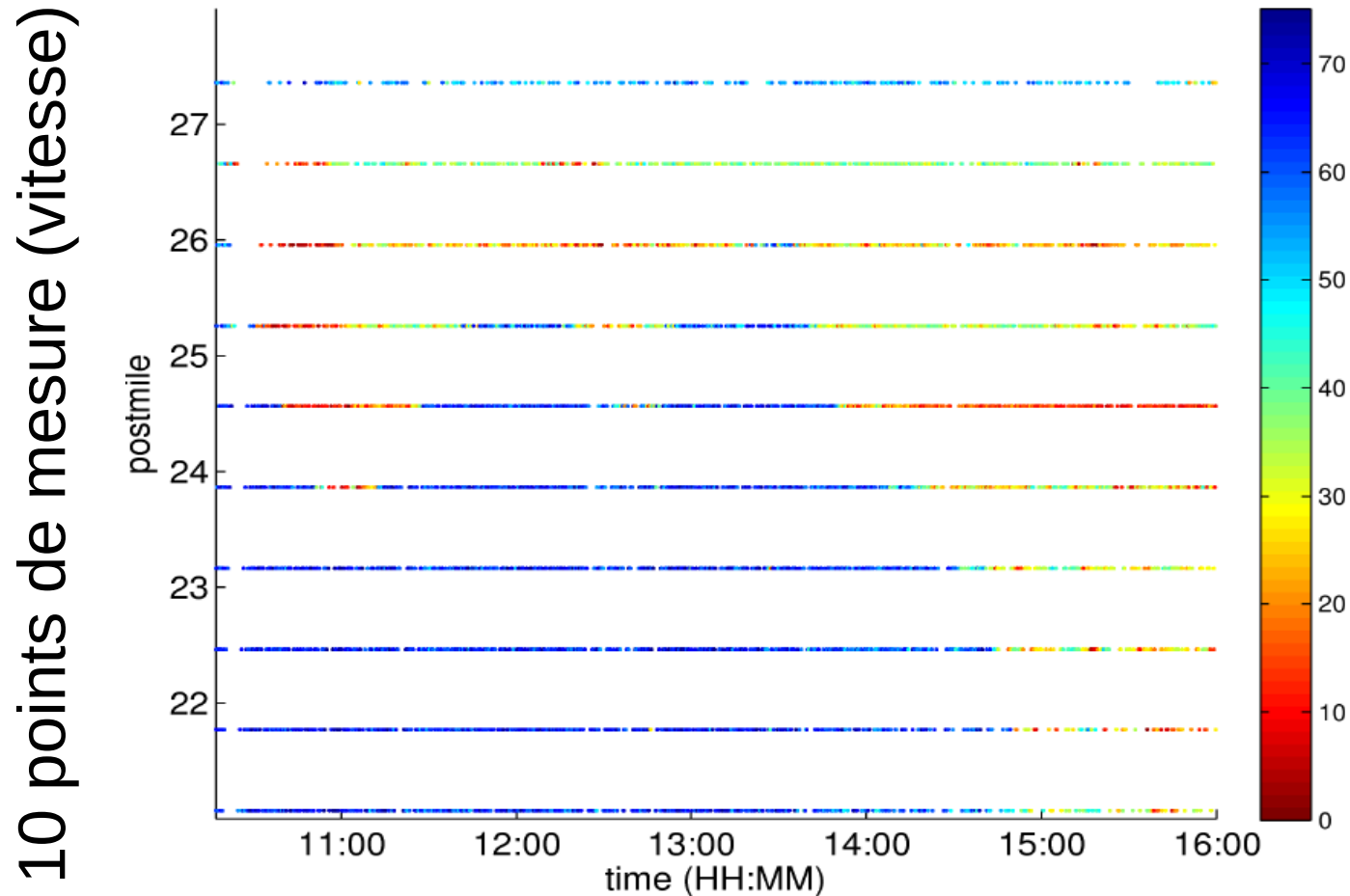




Reconstruction de l'état du trafic

Ensemble de Kalman filter

- Modèle physique: LWR (*Lighthill-Whitham-Richards*)
- Assimilation de données

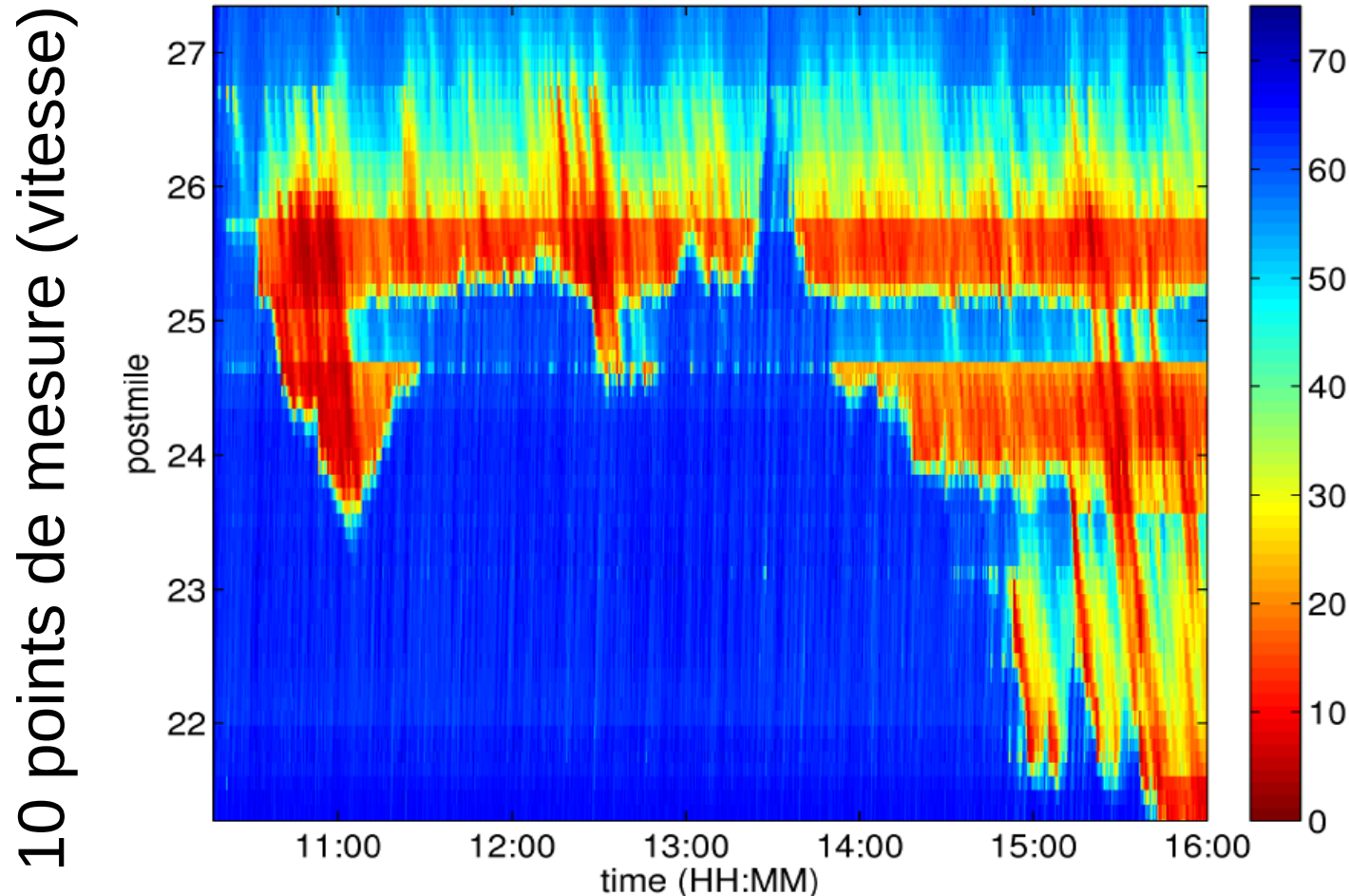


[Work, Blandin, Tossavainen, Piccoli, Bayen, *AMRX*, 2010]

Reconstruction de l'état du trafic

Ensemble de Kalman filter

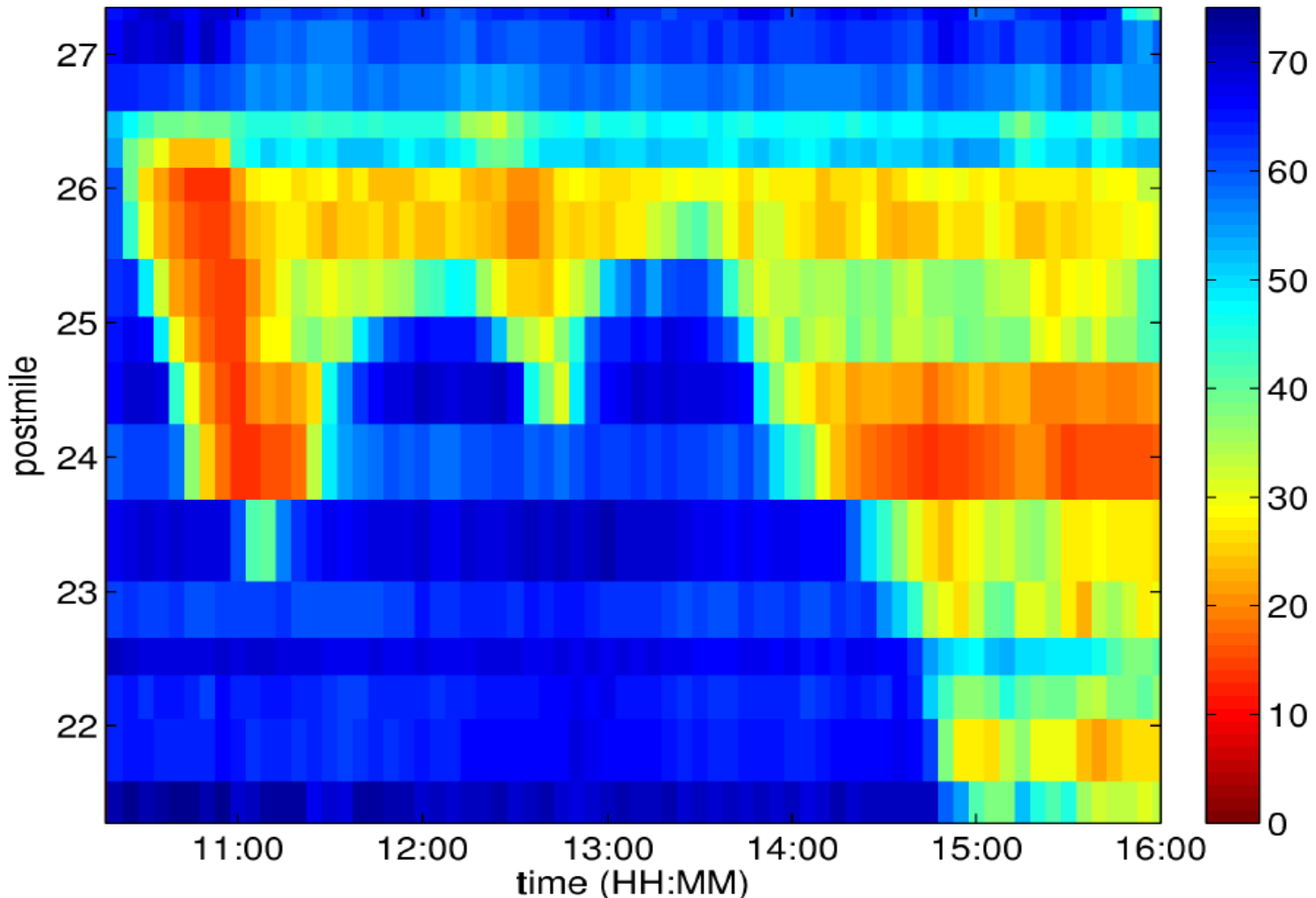
- Utilise le modèle physique pour intégrer les données
- L'interpolation ne fournit pas de résultats aussi précis



[Work, Blandin, Tossavainen, Piccoli, Bayen, *AMRX*, 2010]



Données provenant des détecteurs magnétiques

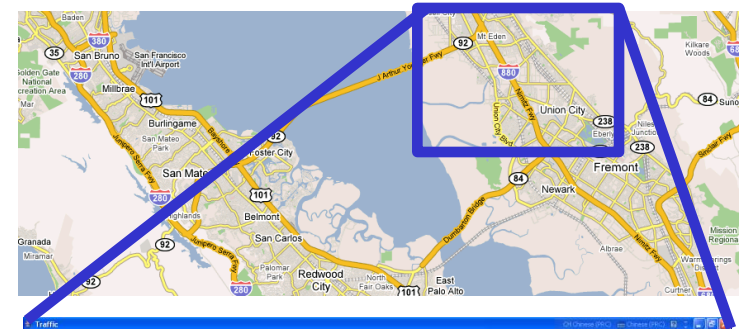
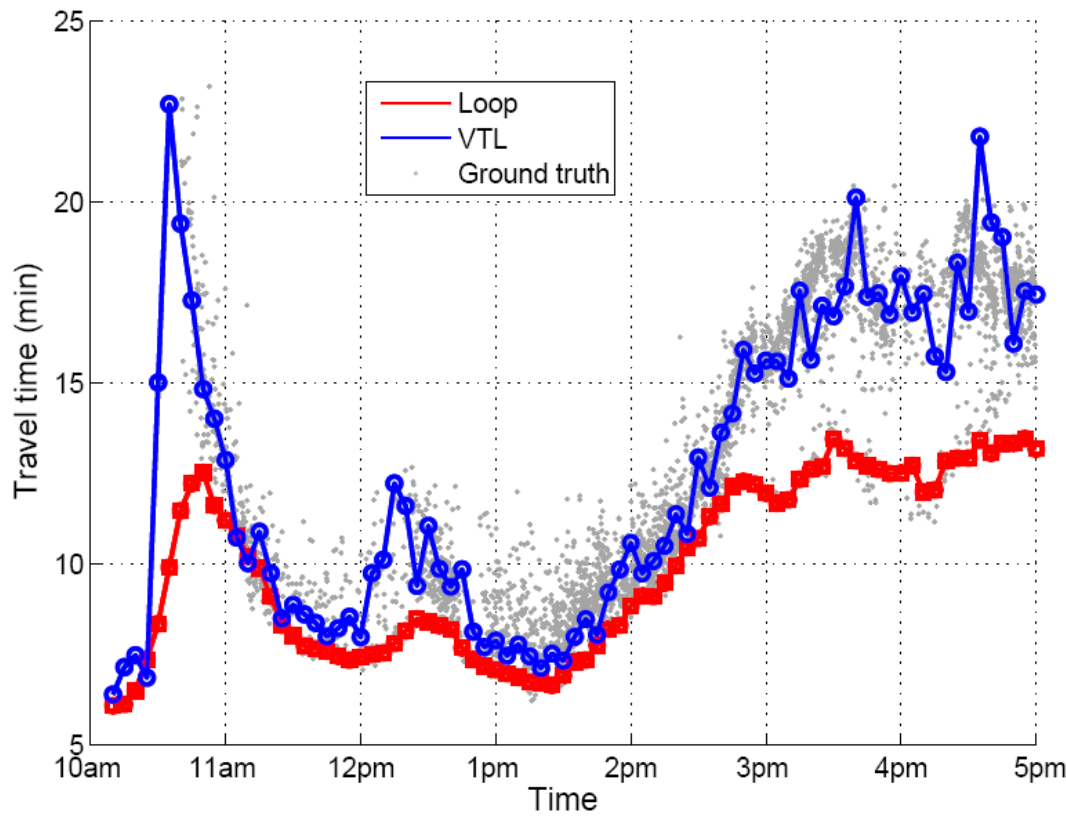


[Work, Blandin, Tossavainen, Piccoli, Bayen, *AMRX*, 2010]

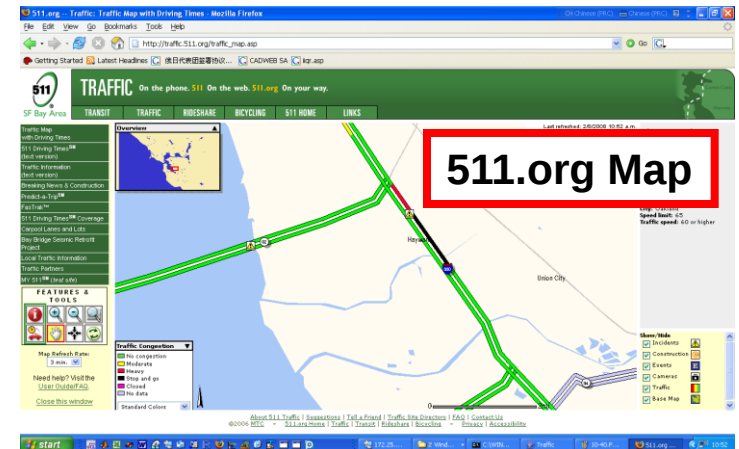
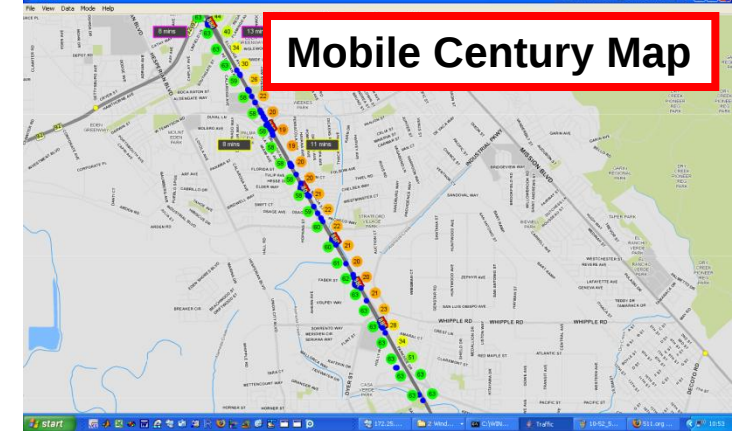
Validation utilisant les données vidéo

Estimation des temps de parcours

- Possible à faibles taux de pénétration (2%)
- Amélioration des estimations.
- Validation grâce à des caméras



Mobile Century Map



2^{ème} étape: réseau urbain

Des différences importantes avec le *trafic autoroutier*:

- **Feux de signalisations:** différences de temps de parcours en fonction du moment d'entrée dans l'arc (par rapport au cycle)
- Piétons
- Parking en double file
- Intersections

➔ **Modélisation statistique de la dynamique du trafic sur le réseau**

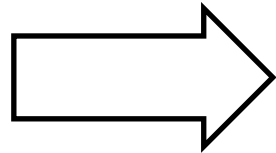
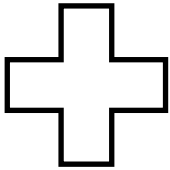
- Probabilité de distribution des temps de parcours sur les arcs
- Dynamique à l'échelle du réseau



Un modèle d'estimation du trafic urbain

Modélisation statistique de la
dynamique du trafic

Modèle



**Estimation/
prédiction**

Données

Mesures fournissant des informations
sur l'état du système et/ou les
paramètres du modèle.

Le Web 2.0 pour l'estimation du trafic

Emergence de l'internet mobile et des instruments de mesures sur téléphones portables

- Communication: GSM, GPRS, WiFi, bluetooth
- Mesures: GPS, accéléromètre,...

Les smartphones et le Web 2.0

- Utilisation du réseau de téléphones portables pour l'estimation du trafic
- Mesures des conditions de trafic et communication avec une unité centrale
 - ➔ Intérêts pour le particulier: information en temps réel
 - ➔ Intérêts pour les agences publiques: meilleure capacité de contrôle et d'optimisation des infrastructures.



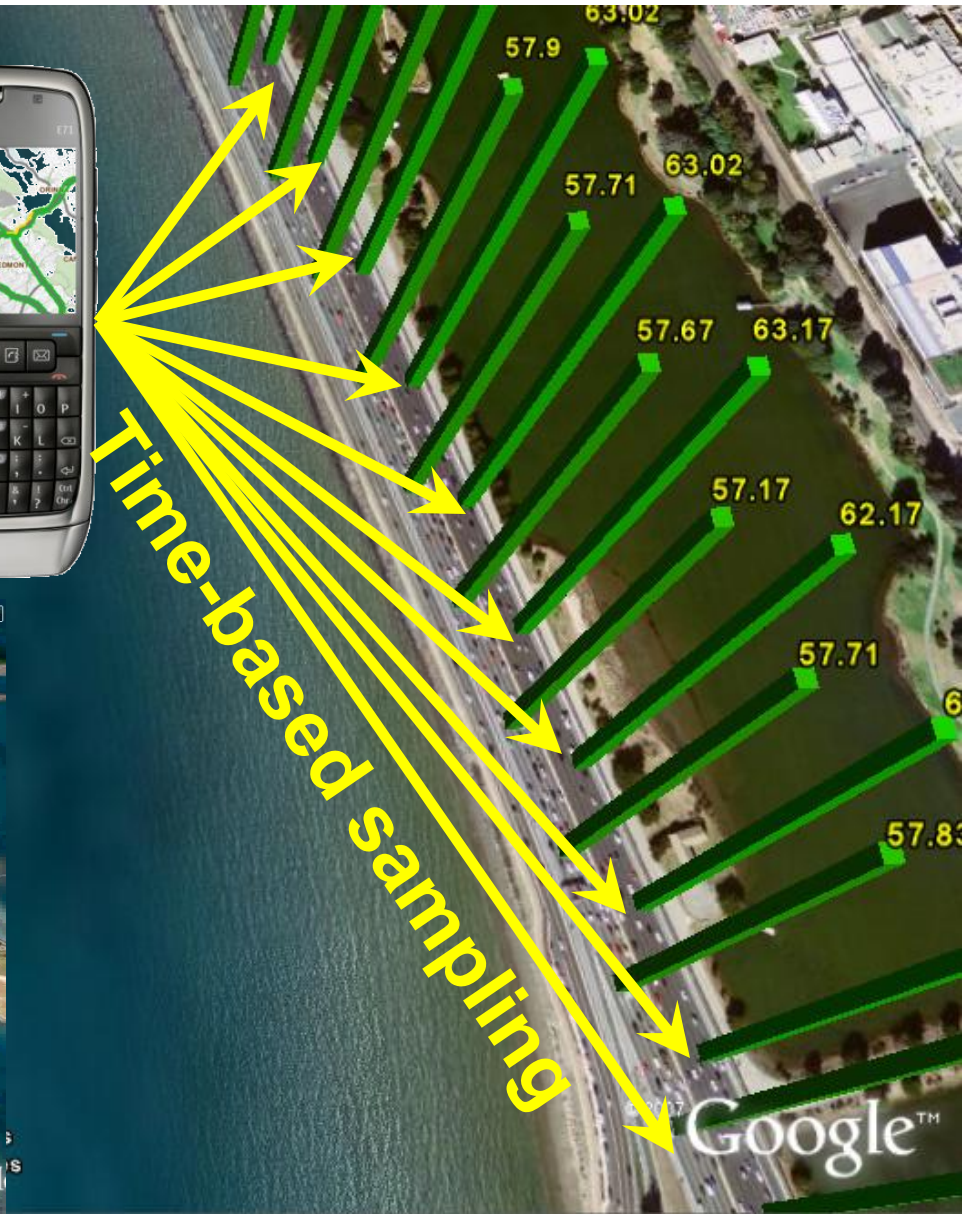
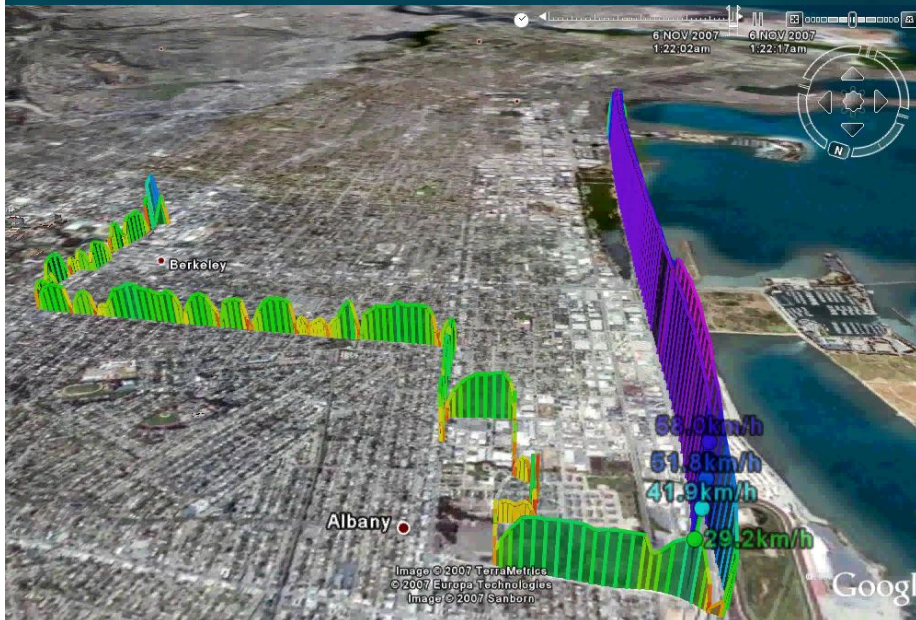
Information et vie privée... quel compromis?

Information disponible:

- Localisation précise
- Suivi de véhicules
- Informations personnelles
- ...



Time-based sampling



Une solution: Virtual Trip Lines (VTLs)



Mesures

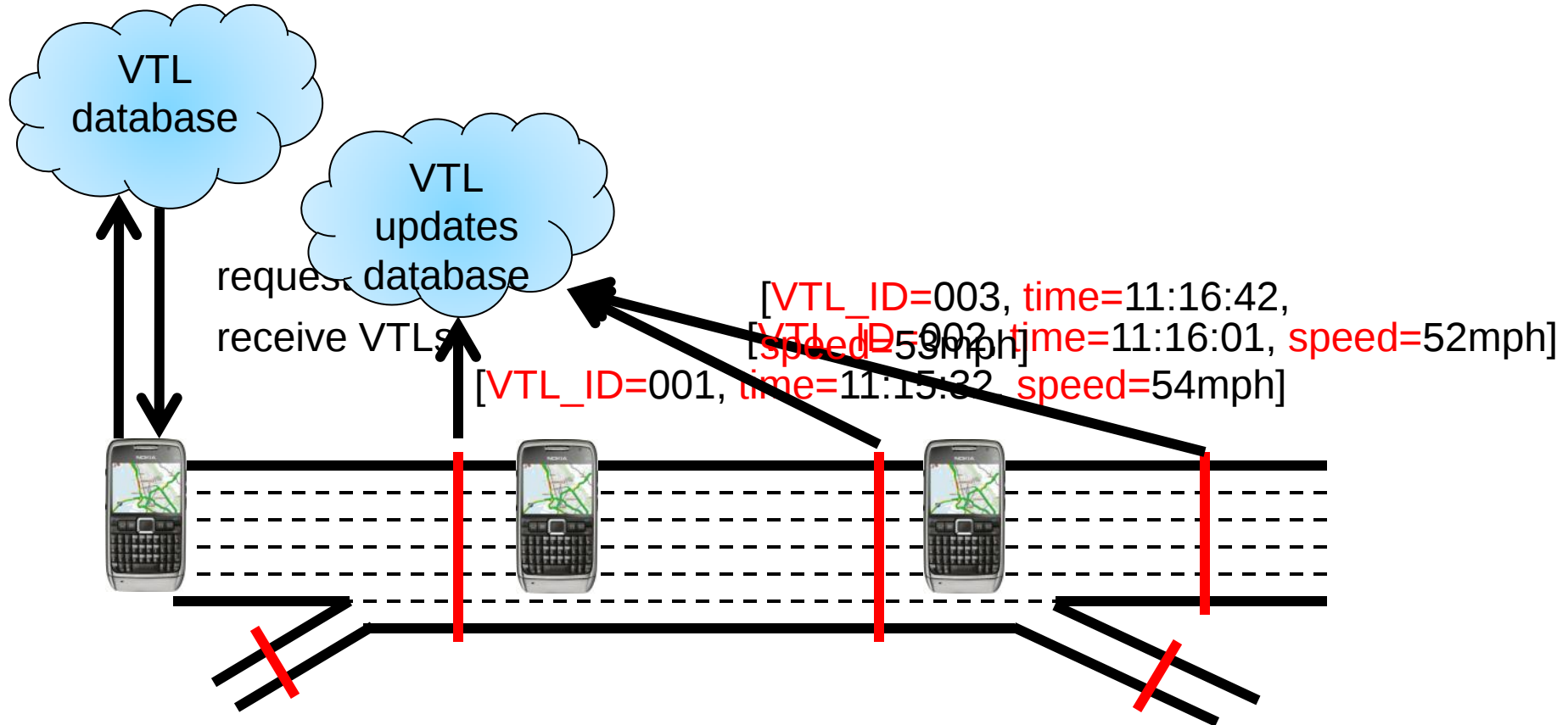
La collecte de données est basée sur:

- *Virtual Trip Lines* (VTLs), marqueurs “virtuels” représentant une location géographique dans des zones prédéfinies.
- Collectent de façon probabiliste, des informations sur les véhicules
- Données anonymes envoyées au serveur central



En pratique...

1. Le téléphone télécharge les positions des VTLs dans son voisinage (automatique)
2. La trajectoire a-t-elle croisé un VTL?
 - a) Si oui, envoi des informations de manière anonyme



Déploiement national de “l’infrastructure virtuelle”

- 450,000 VTLs dans le nord de la Californie, 4.5 millions sur l’ensemble du pays
- 5,000 utilisateurs du logiciel dans le nord de la Californie (disponible pour Nokia et Blackberry)

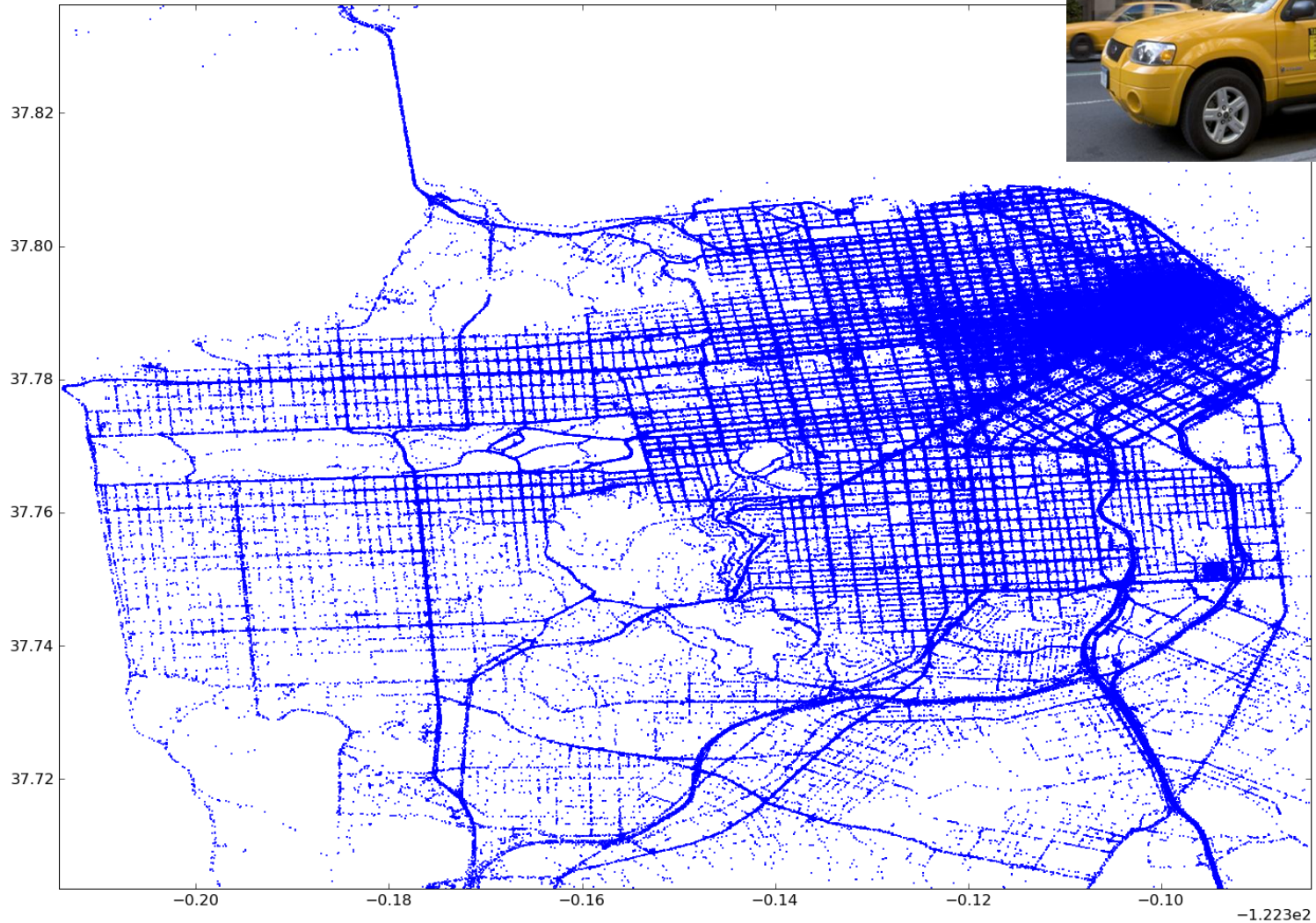




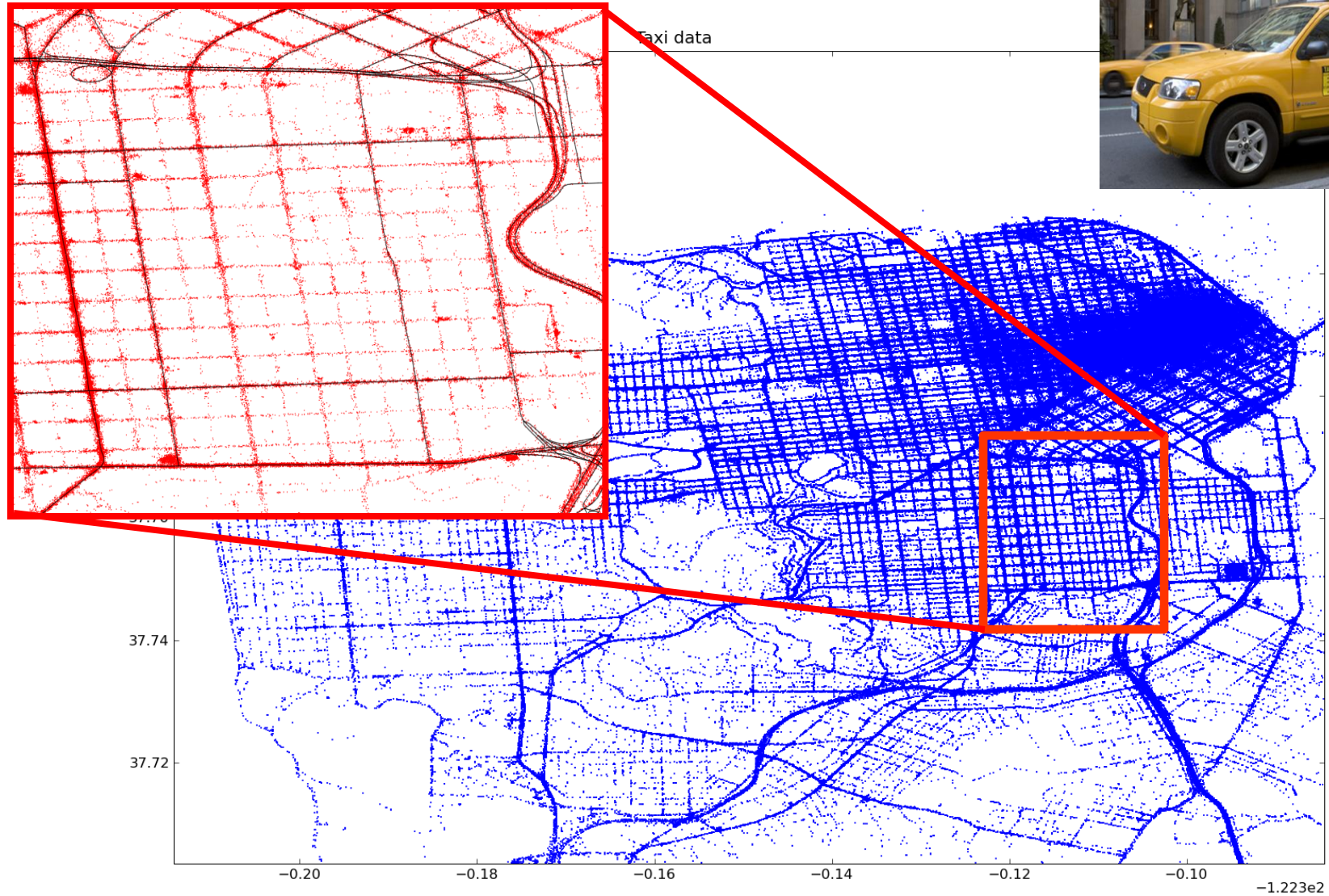
Autres sources de données (*cabspotting*, SF taxis)

2010-03-29 04:00:02.0

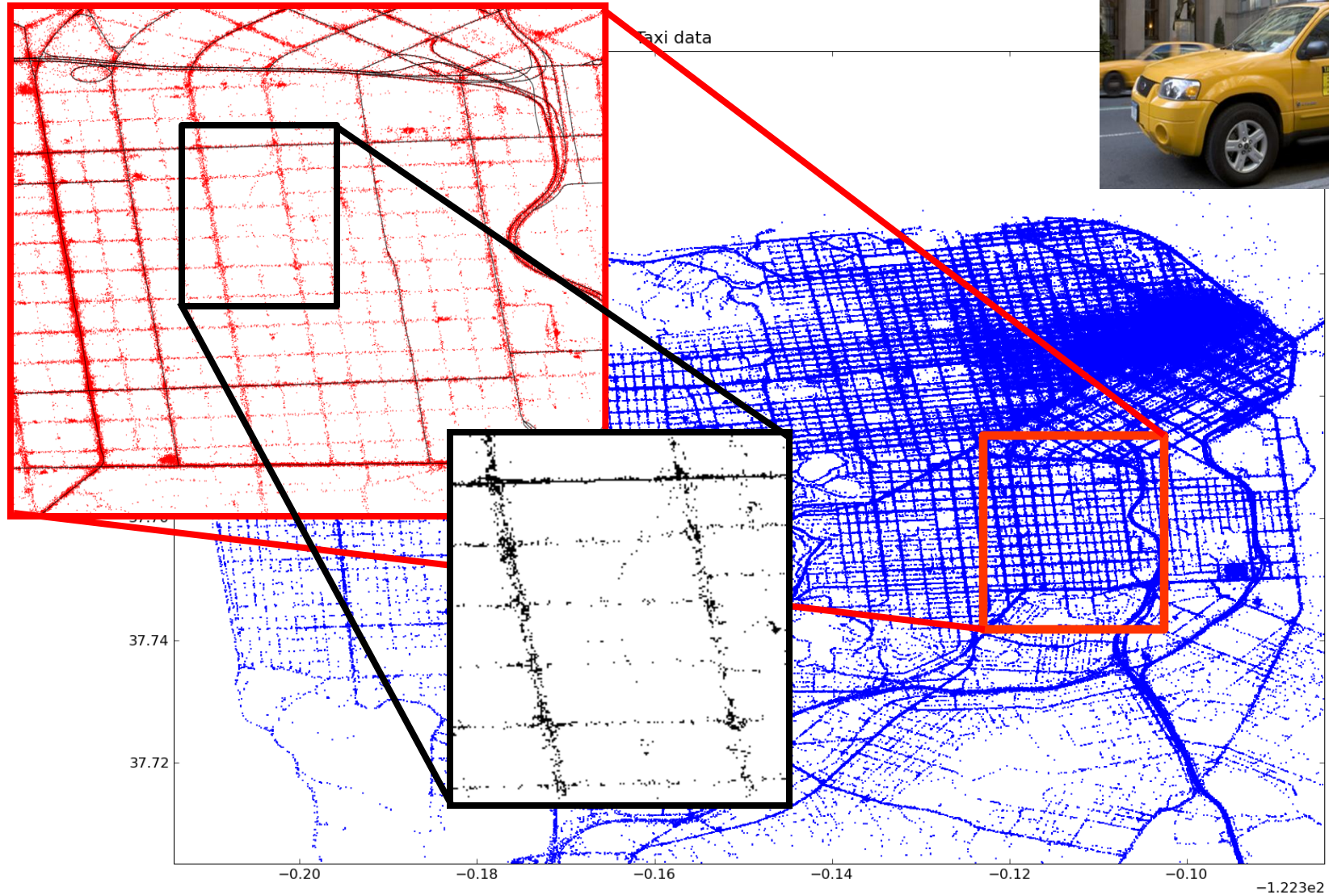
Données reçues en une journée (SF taxis)



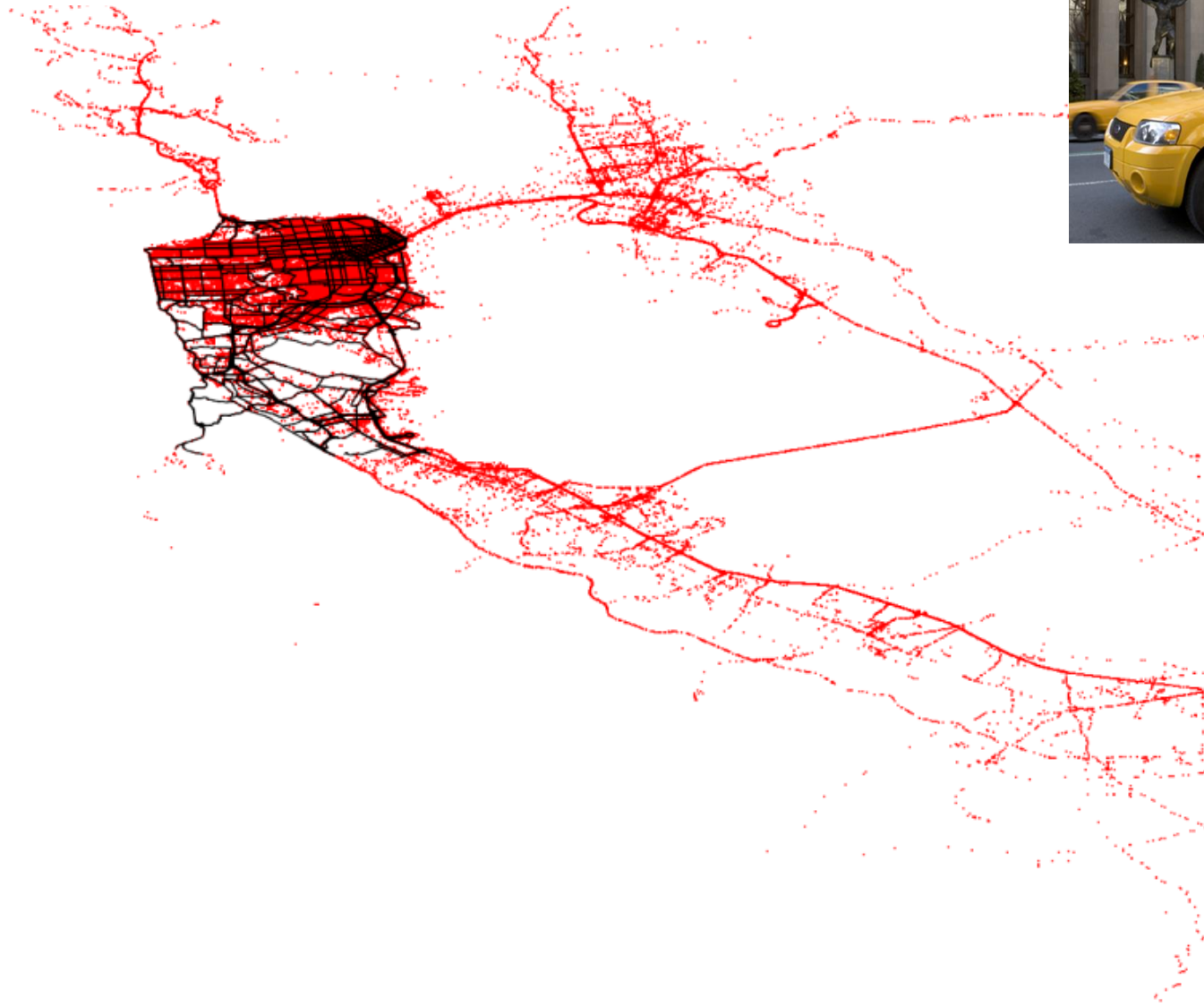
Données reçues en une journée (SF taxis)



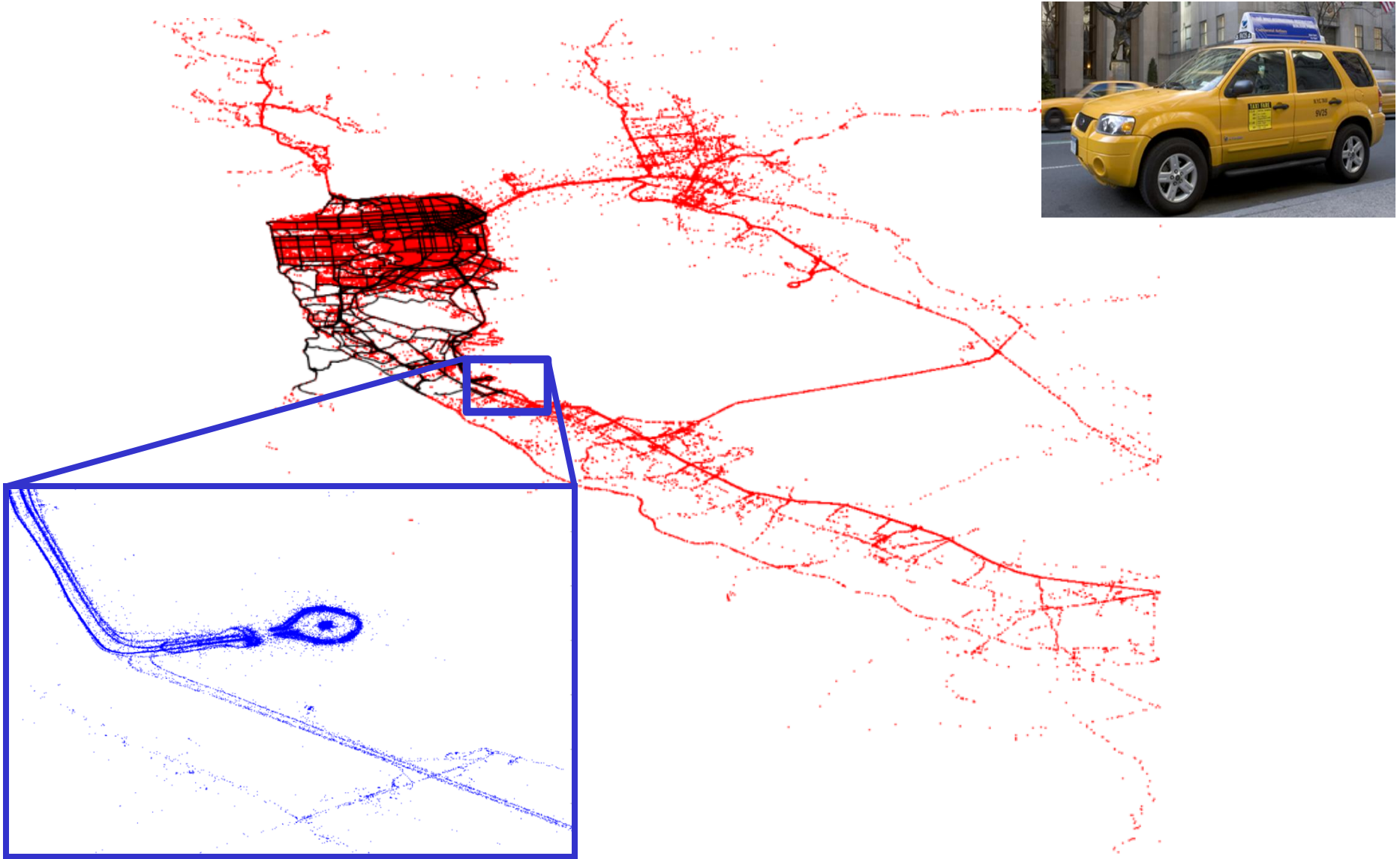
Données reçues en une journée (SF taxis)



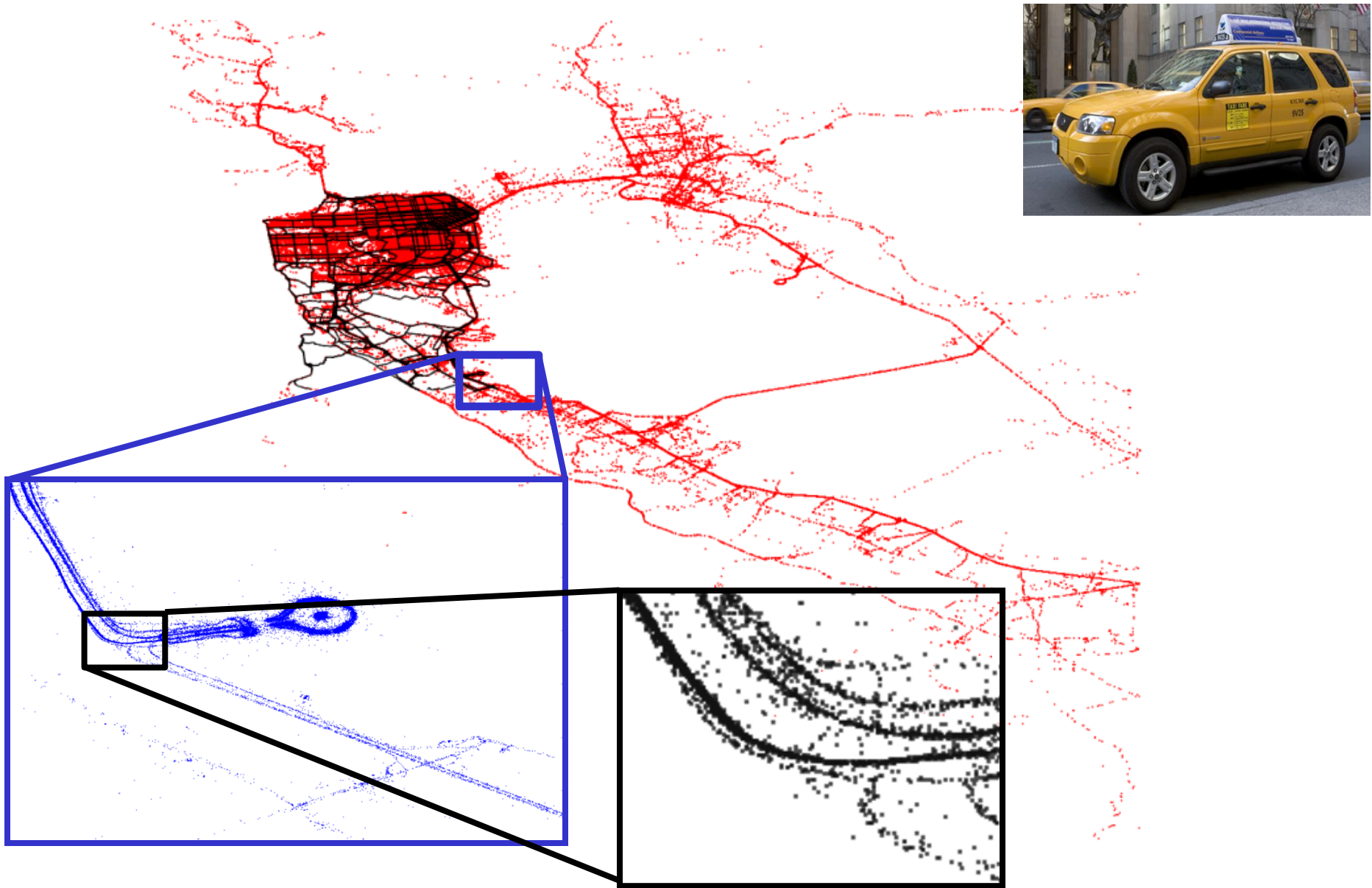
Données reçues en une journée (SF taxis)



Données reçues en une journée (SF taxis)

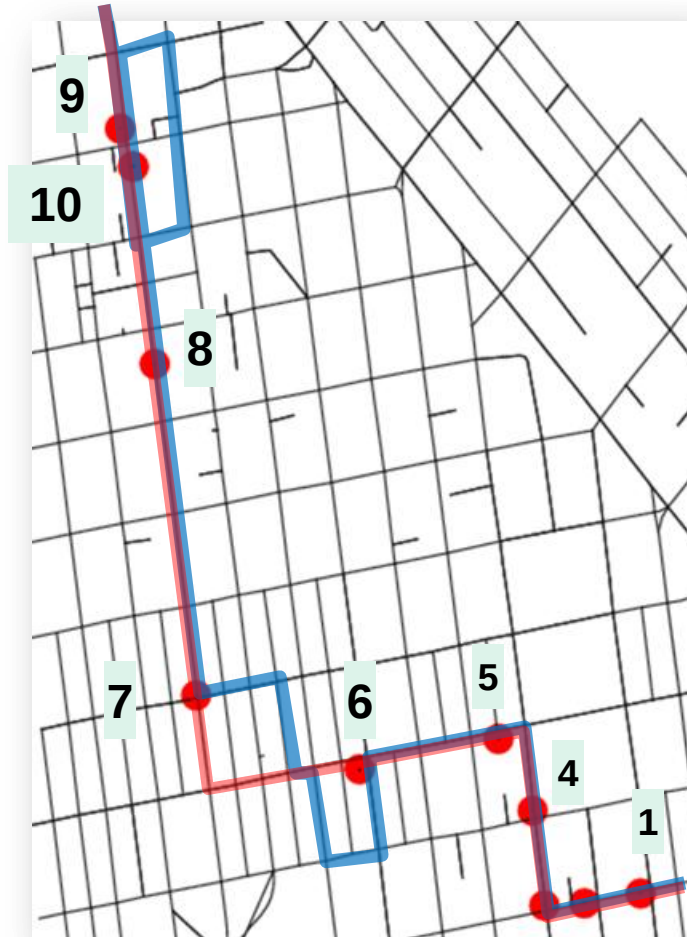


Données reçues en une journée (SF taxis)



Utilisation des données GPS

Nouvelle source de données, nouvelles méthodes d'exploitation:



La reconstruction de la trajectoire est source d'erreurs:

- aux intersections (point 6)
- lorsque le véhicule est arrêté (point 9)

Trajectoire

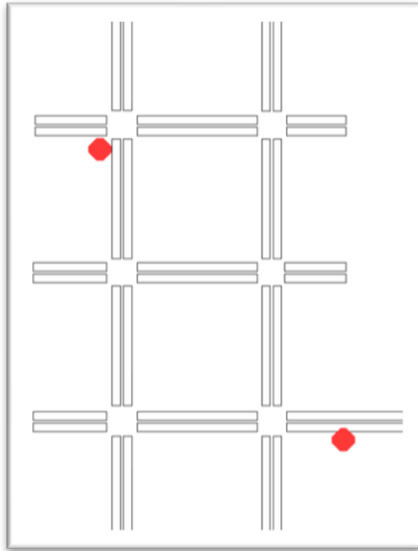
Reconstruction naïve

● GPS point

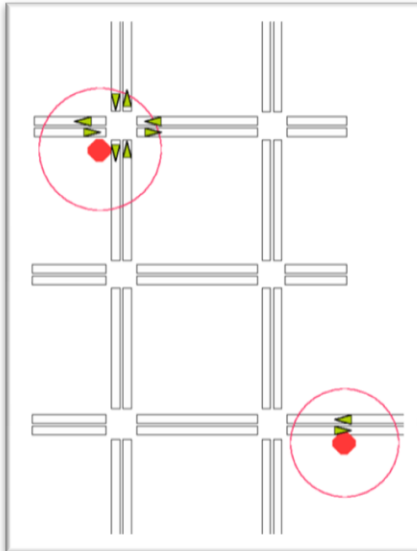
Utilisation des données GPS

Nouvelle source de données, nouvelles méthodes d'exploitation:

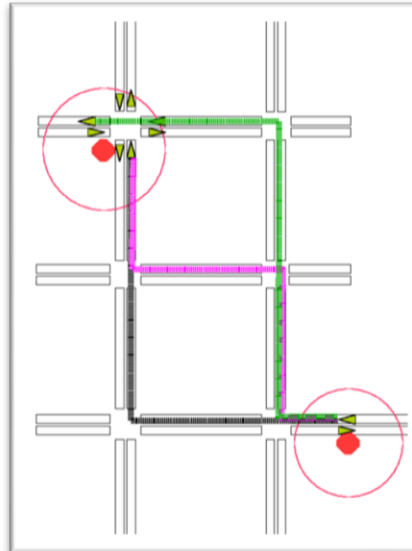
1 - Données



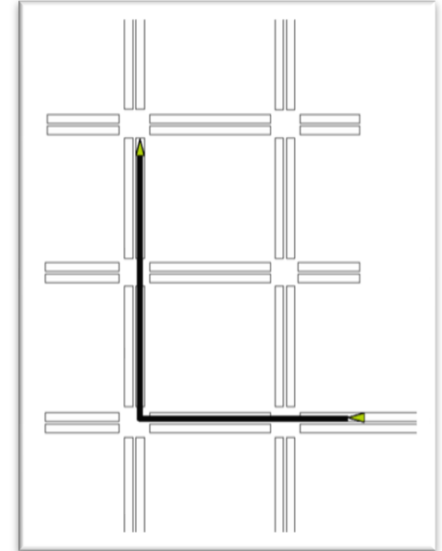
2 – Projection sur le réseau



3 – Recherche des chemins possibles pour chaque projection



4 – Choix du “meilleur” chemin



Algorithme de champs markovien combinant un modèle d'erreurs GPS avec un modèle de trafic.



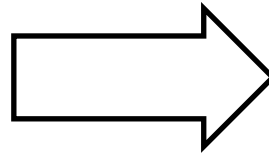
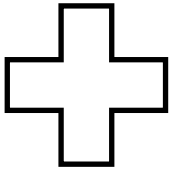
Utilisation des données GPS

2010-03-29 04:00:02.0

L'estimation du trafic aujourd'hui

Modélisation statistique de la
dynamique du trafic

Modèle



**Estimation/
prédiction**

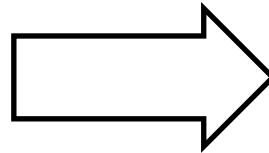
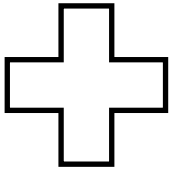
Données

Mesures fournissant des informations
sur l'état du système et/ou les
paramètres du modèle.

L'estimation du trafic aujourd'hui

Modélisation statistique de la
dynamique du trafic

1. **Probabilité de
distribution des temps
de parcours sur les
arcs du réseau**
2. **Dynamique du trafic à
l'échelle du réseau**



**Estimation/
prédiction**

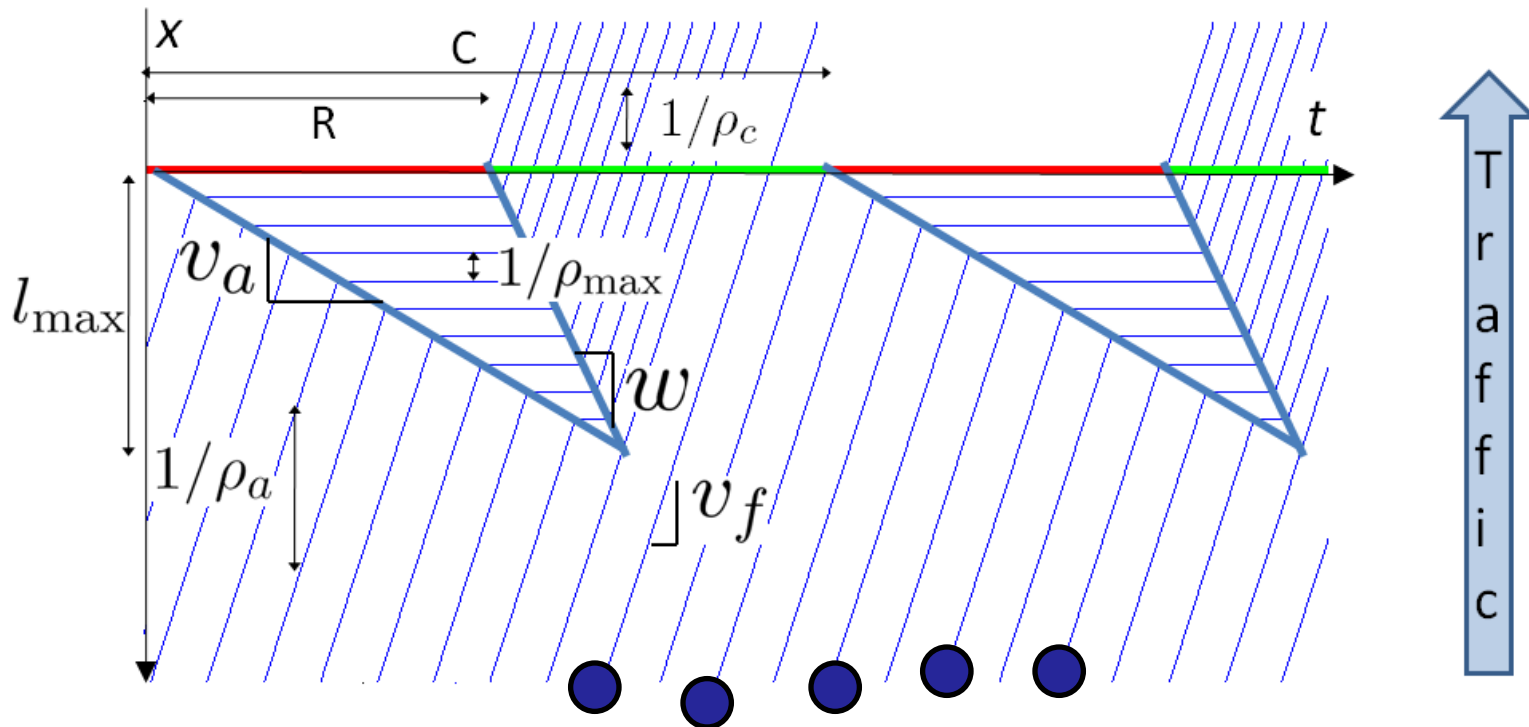
Données

Mesures fournissant des informations
sur l'état du système et/ou les
paramètres du modèle.

Trafic sur un arc du réseau

Hypothèses simplificatrices

- Diagramme fondamental triangulaire (vitesse libre v_f , densité critique ρ_c , densité maximale $\rho_{\max}$)
- Taux d'arrivée constants (densité ρ_a)
- Stationarité
 - Pas de variation de la longueur de la queue
 - Evolution périodique



Résultats classiques de théorie du trafic

Formation et dissipation des queues (Rankine Hugoniot)

Formation: $v_a = \frac{\rho_a v_f}{\rho_{\max} - \rho_a}$

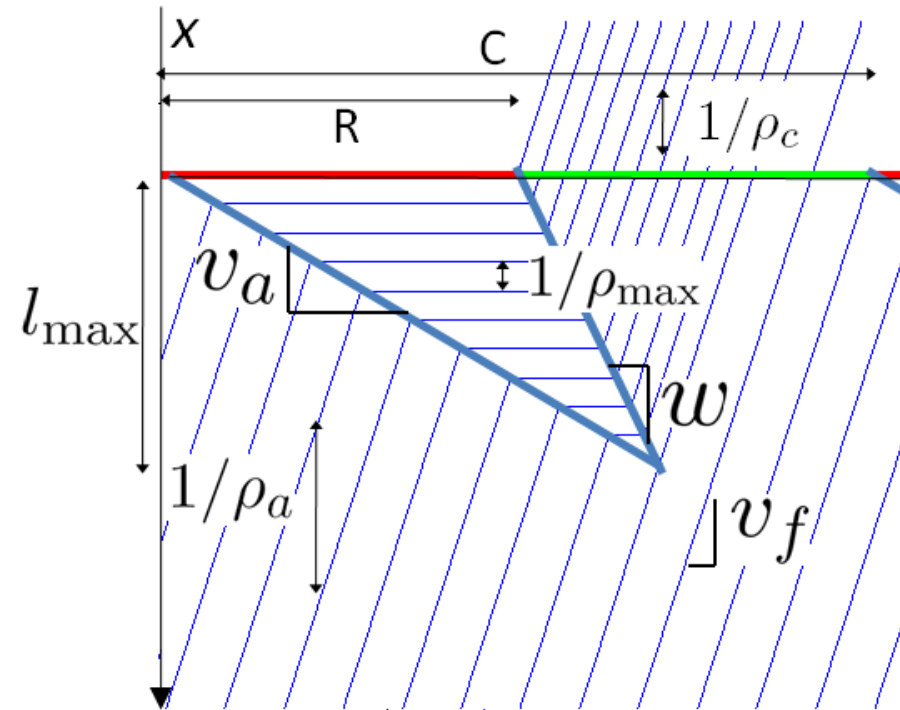
Dissipation: $w = \frac{\rho_c v_f}{\rho_{\max} - \rho_c}$

Longueur de la queue

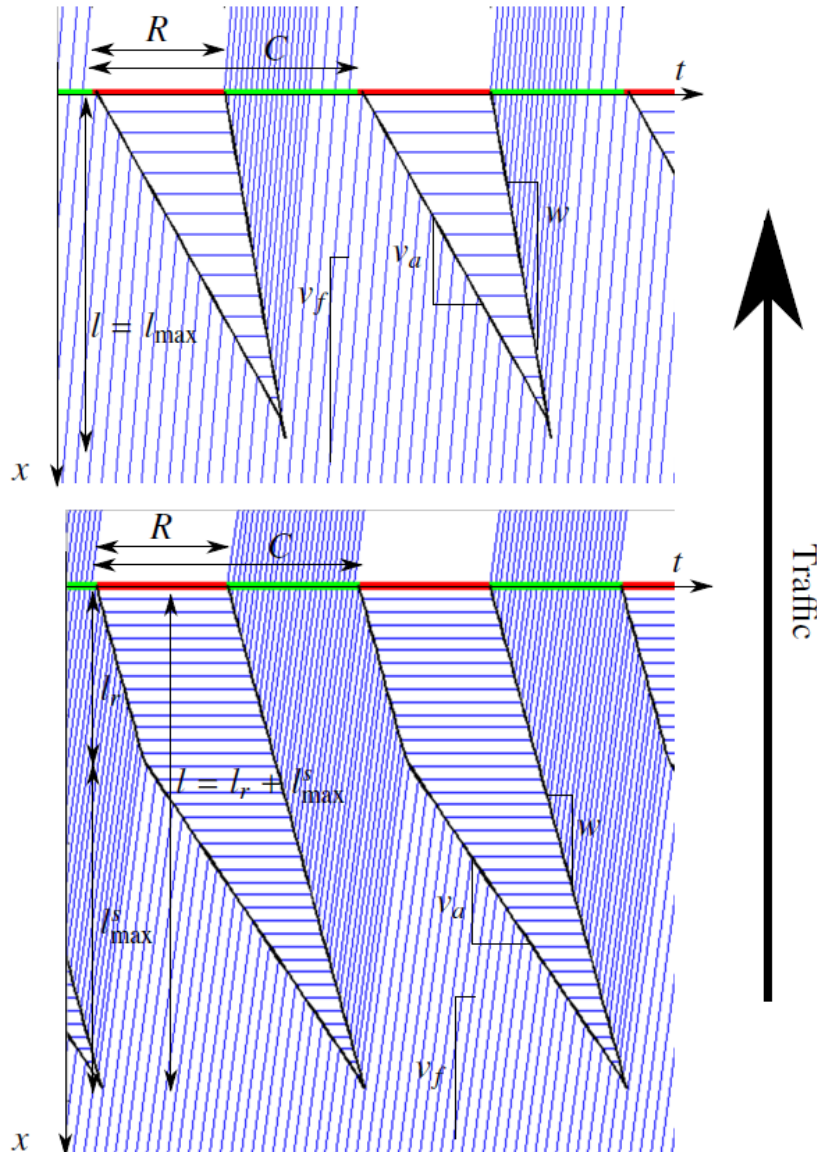
$$l_{\max} = R \frac{w v_a}{w - v_a} = R \frac{v_f}{\rho_{\max}} \frac{\rho_c \rho_a}{\rho_c - \rho_a}.$$

Queue à saturation:

$$\rho_a^s = \frac{C-R}{C} \rho_c. \quad l_{\max}^s = v_f \rho_c (C - R) / \rho_{\max}$$



Régime non-saturé/congestionné



Paramètres du réseau

- Temps de rouge R
- Temps de cycle C
- Pas (inverse de la vitesse libre) ρ_f , distribution φ^p
- Queue de saturation $l_{\max}^s$

Etat de congestion: longueur de la queue:

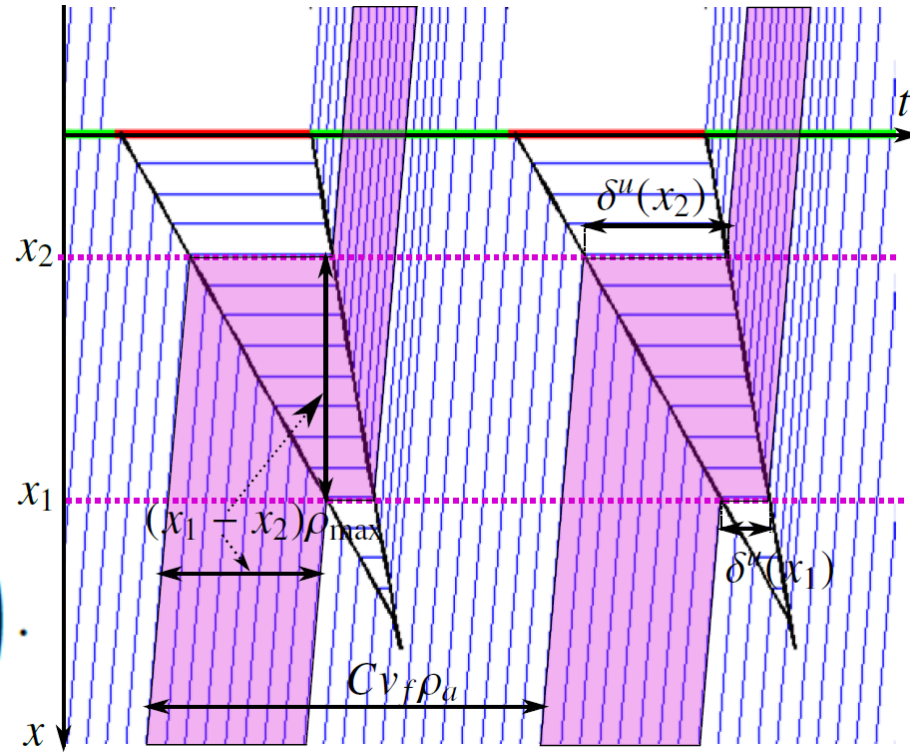
- Régime non saturé: $l_{\max}$
- Régime congestionné: $l_r + l_{\max}^s$

Remarque: Le régime congestionné est un *régime critique* où le nombre de véhicules entrant dans l'arc est le nombre de véhicules pouvant sortir de l'arc en un cycle (distance $l_{\max}^s$).

Pdf des délais entre x_1 et x_2 (régime non-saturé)

- Véhicules *ne s'arrêtant pas* entre x_1 et x_2
 → Délai null
- Véhicules *s'arrêtant* entre x_1 et x_2
 → Délai uniforme entre un délai minimal (délai en x_1) et un délai maximal (délai en x_2)

Délai en x : $\delta^u(x) = R \left(1 - \frac{\min(x, l_{\max})}{l_{\max}} \right)$.

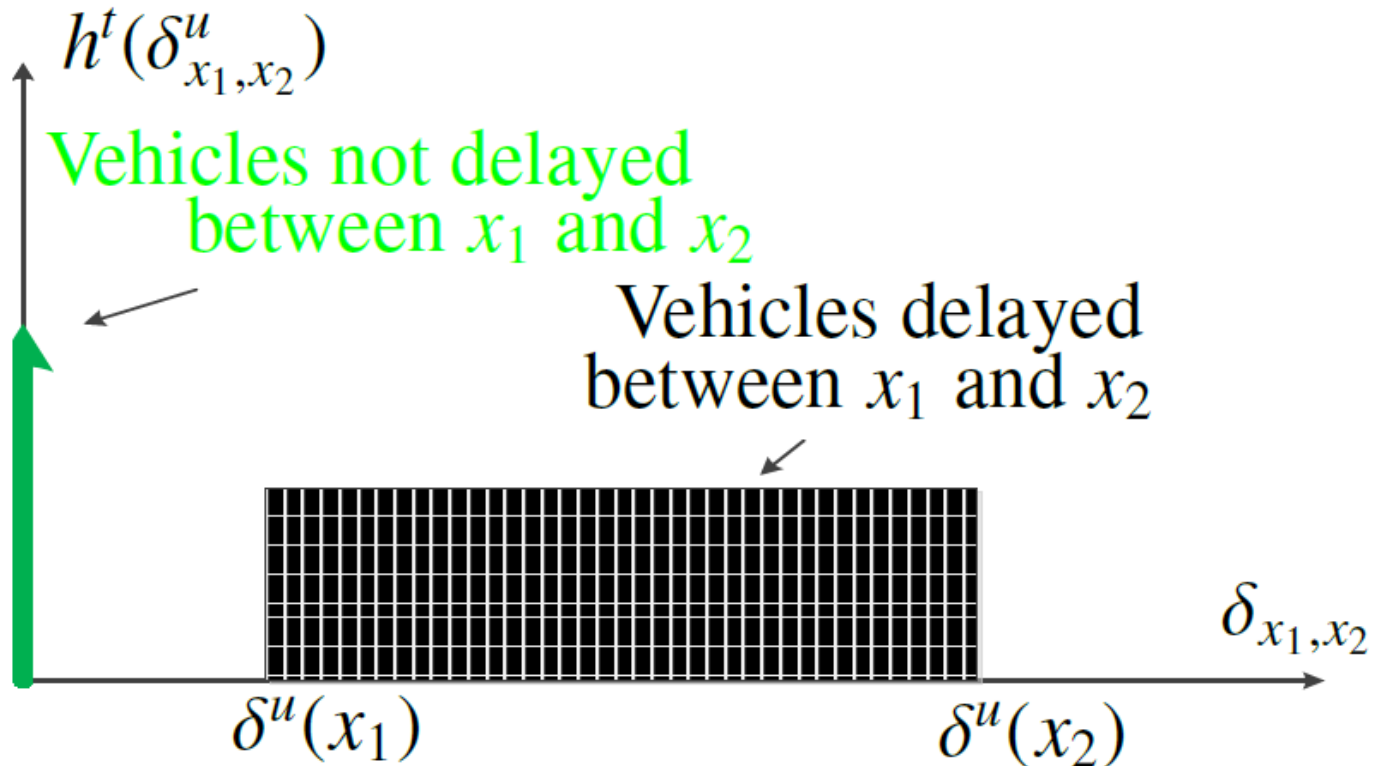


Fraction des véhicules avec délai entre x_1 et x_2 (sur l'ensemble des véhicules entrant dans la durée d'un cycle)

$$\eta_{x_1, x_2}^u = \frac{\min(x_1, l_{\max}) - \min(x_2, l_{\max})}{l_{\max}} \left(\frac{R}{C} + \left(1 - \frac{R}{C} \right) \frac{l_{\max}}{l_{\max}^s} \right).$$

Pdf des délais entre x_1 et x_2 (régime non-saturé)

$$h^t(\delta_{x_1, x_2}) = (1 - \eta_{x_1, x_2}^u) \text{Dir}_{\{0\}}(\delta_{x_1, x_2}) + \frac{\eta_{x_1, x_2}^u}{\delta^u(x_2) - \delta^u(x_1)} \mathbf{1}_{[\delta^u(x_1), \delta^u(x_2)]}(\delta_{x_1, x_2})$$



Pdf des délais envoyés par les véhicules

Le véhicule envoie ses positions x_1 et x_2 , successivement à t_1 et t_2

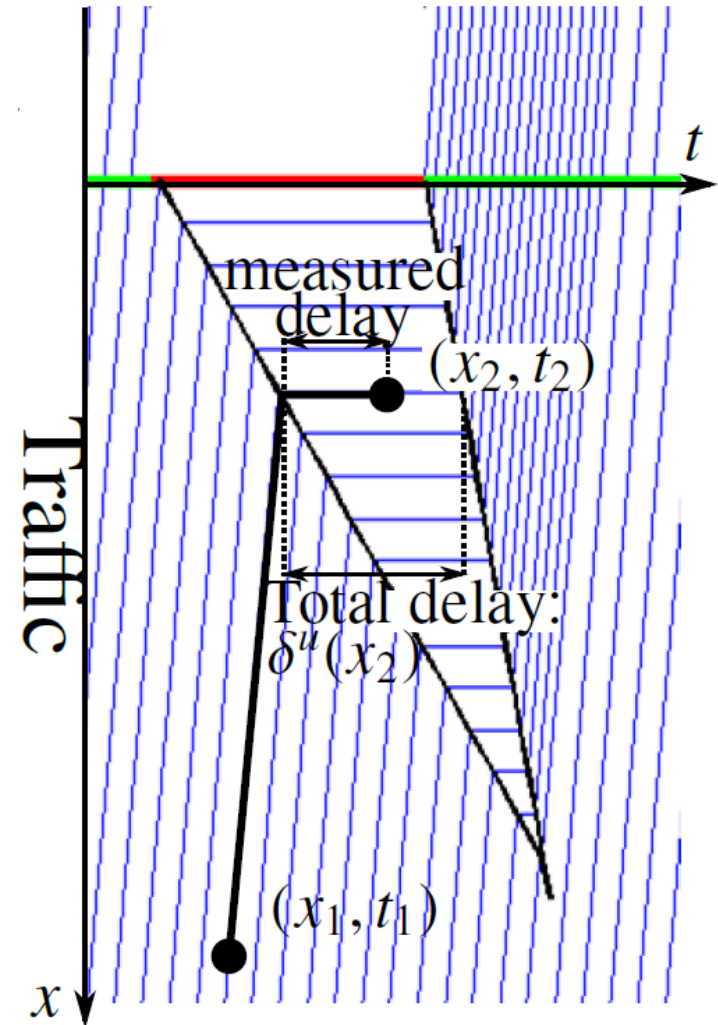
Le véhicule peut être arrêté quand il envoie ses positions

- Véhicule arrêté en x_1 (probabilité ξ_1),
délai compris entre 0 et $\delta^u(x_1)$
- Véhicule arrêté en x_2 (probabilité ξ_2),
délai compris entre 0 et $\delta^u(x_2)$
- Véhicule arrêté ni en x_1 , ni en x_2 ,
probabilité $\mathcal{P}(\bar{s}_{x_1}, \bar{s}_{x_2})$, on est
dans le cas précédent

La probabilité d'envoyer sa position x_i en étant arrêté, ξ_i , dépend du ratio entre le délai et le cycle

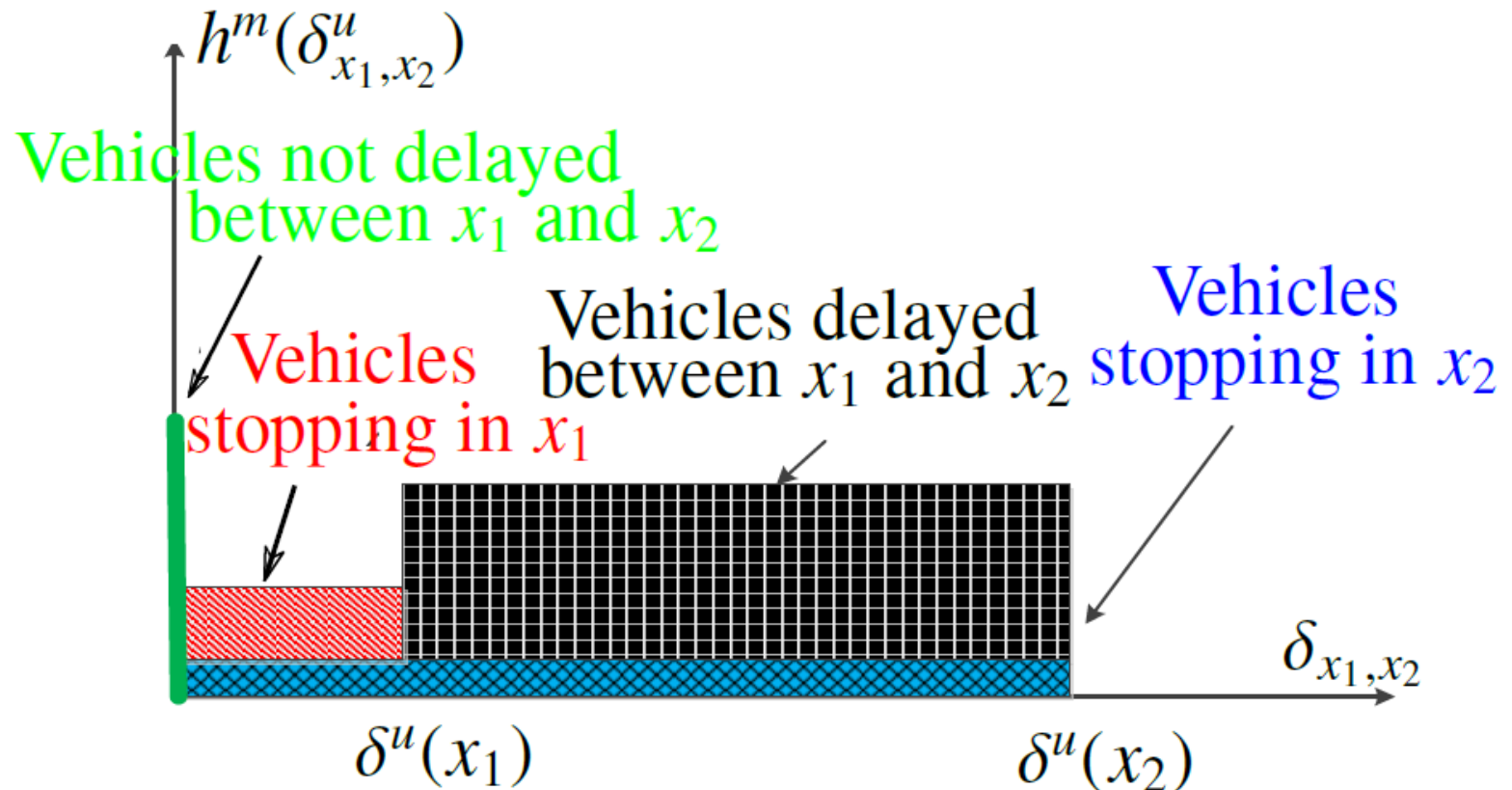
Cas particulier: $x_1 = x_2$:

Le véhicule est arrêté en x_1 et en x_2



Pdf des délais envoyés par les véhicules

$$h^m(\delta_{x_1, x_2}) = \mathcal{P}(\bar{s}_{x_1}, \bar{s}_{x_2}) h^t(\delta_{x_1, x_2}) + \frac{\zeta_{x_1}}{\delta^u(x_1)} \mathbf{1}_{[0, \delta^u(x_1)]}(\delta_{x_1, x_2}) + \frac{\zeta_{x_2}}{\delta^u(x_2)} \mathbf{1}_{[0, \delta^u(x_2)]}(\delta_{x_1, x_2})$$



Pdf des délais entre x1 et x2 (régime congestionné)

- Entre deux arrêts, les véhicules avancent de $l_{\max}^s$

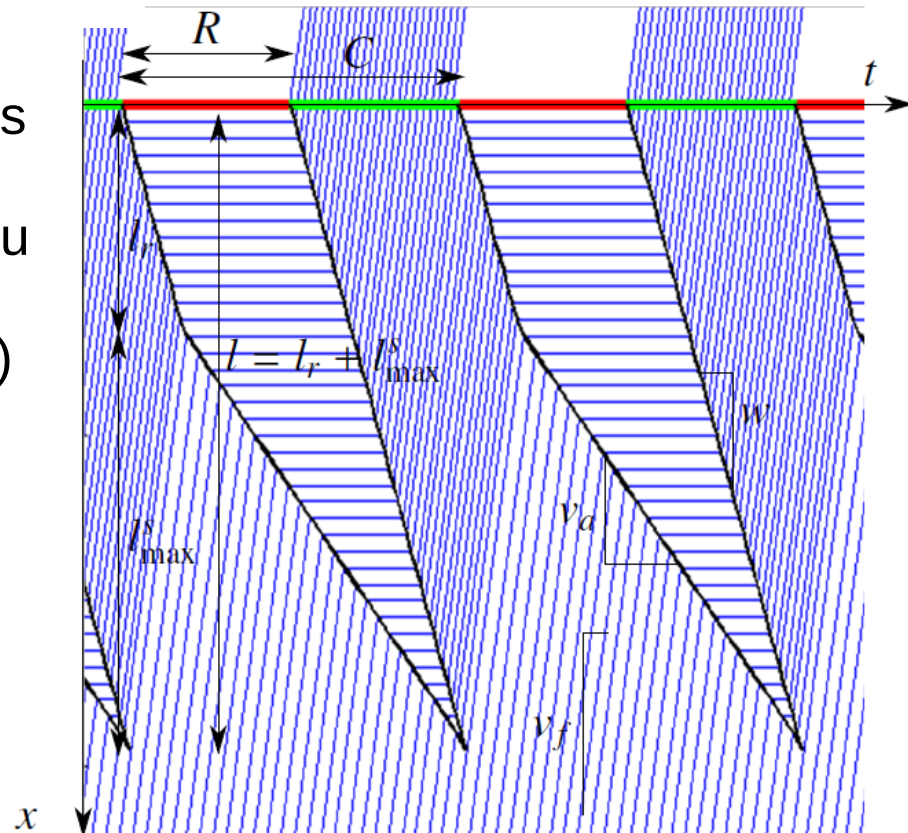
$$\text{Délai en } x: \delta^c(x) = \begin{cases} R & \text{if } x \leq l_r \\ R \frac{l_r + l_{\max}^s - x}{l_{\max}^s} & \text{if } x \in [l_r, l_r + l_{\max}^s] \\ 0 & \text{if } x \geq l_r + l_{\max}^s \end{cases}$$

Les pdf de délais dépendent des positions relatives de x1 et x2 par rapport à

- l_r (longueur de la queue quand le feu passe au rouge)
- $l_r + l_{\max}^s$ (longueur totale de la queue)

Nombre maximal d'arrêt dans la queue l_r entre x1 et x2

$$n_s = \left\lceil \frac{\min(x_1, l_r) - \min(x_2, l_r)}{l_{\max}^s} \right\rceil.$$

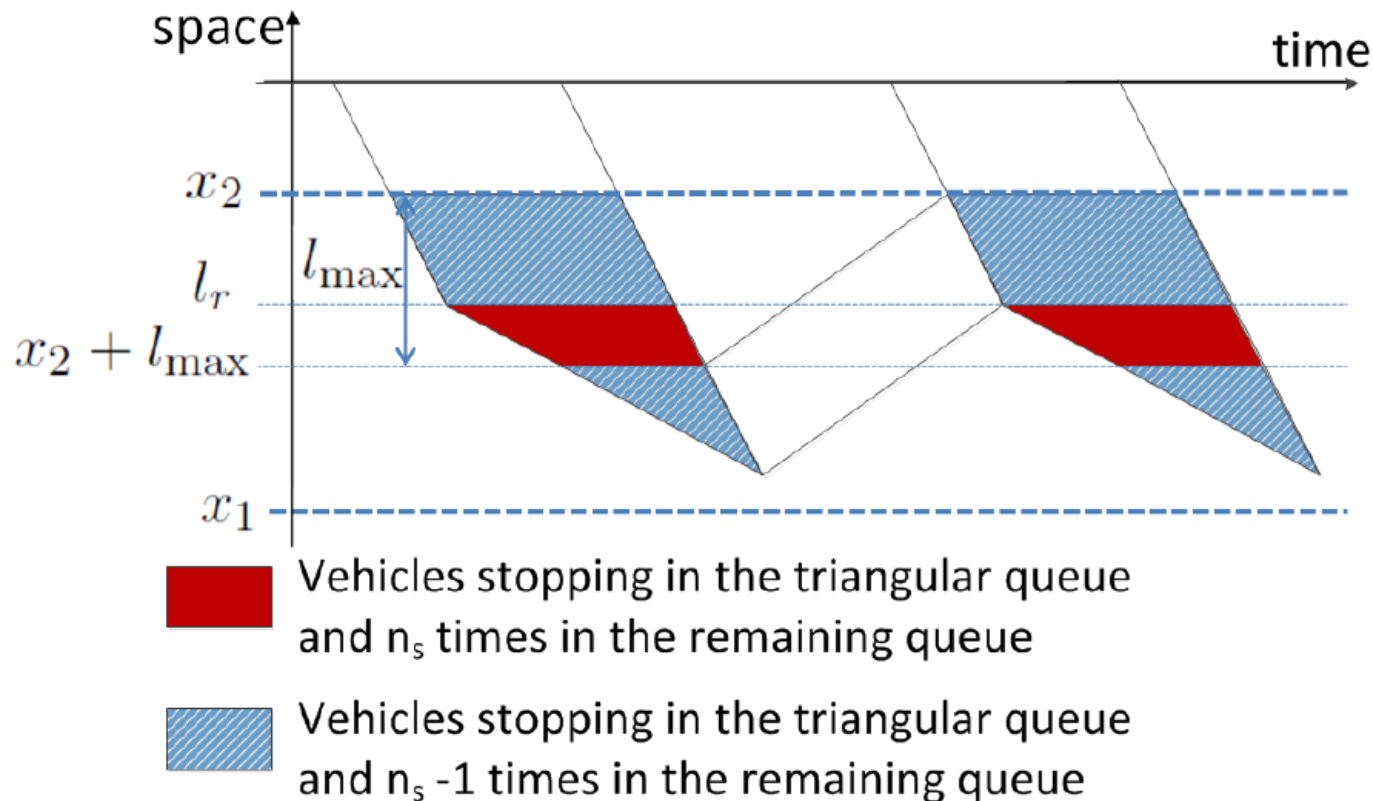


Pdf des délais entre x_1 et x_2 (régime congestionné)

- Cas Numéro 1: $x_1 > l_r + l_{\max}^s$, $x_2 < l_r$

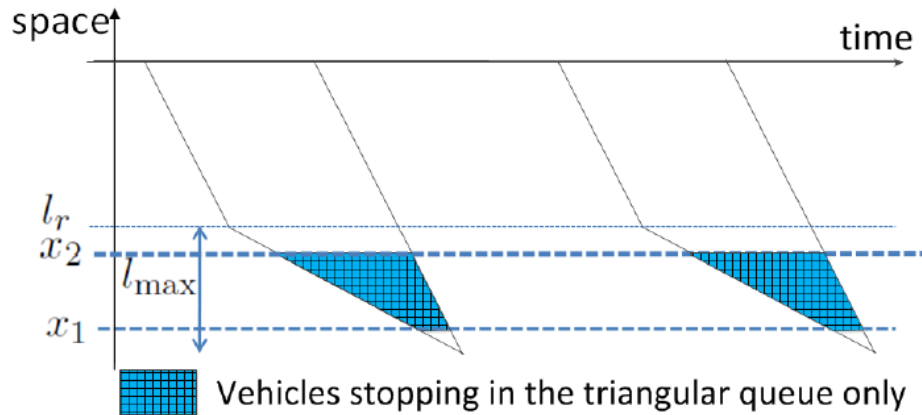
$$x_c = x_2 + n_s l_{\max}^s$$

$$h^t(\delta_{x_1, x_2}) = \frac{1}{\delta_{\max} - \delta_{\min}} \mathbf{1}_{[\delta_{\min}, \delta_{\max}]}(\delta_{x_1, x_2}), \quad \begin{aligned} \delta_{\min} &= \delta^c(x_c) + (n_s - 1)R \\ \delta_{\max} &= \delta^c(x_c) + n_s R \end{aligned}$$



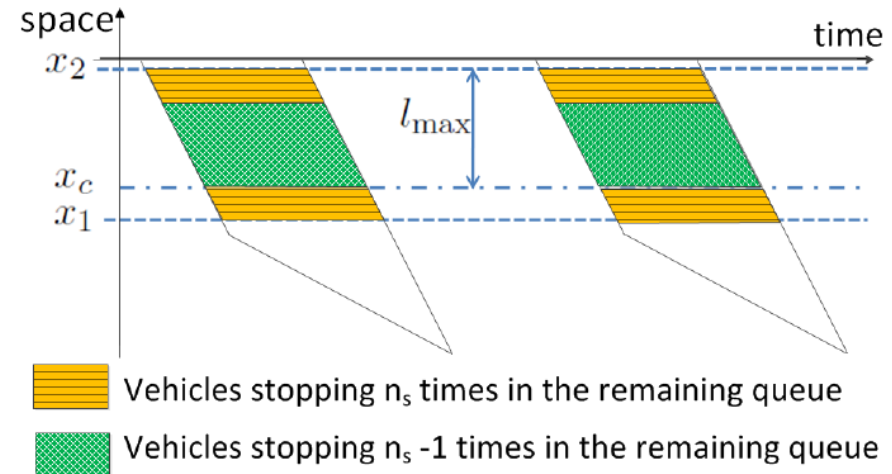
Pdf des délais entre x_1 et x_2 (régime congestionné)

- Cas Numéro 2: $x_1 > l_r$, $x_2 >$



- Similaire au cas non-saturé: x_1 et x_2 sont avant la queue l_r .
 - Véhicules ne s'arrêtant pas:
 - délai nul
 - Véhicules s'arrêtant
 - Délai uniforme entre le délai en x_1 et le délai en x_2

- Cas Numéro 3: $x_1 < l_r$, $x_2 < l_r$



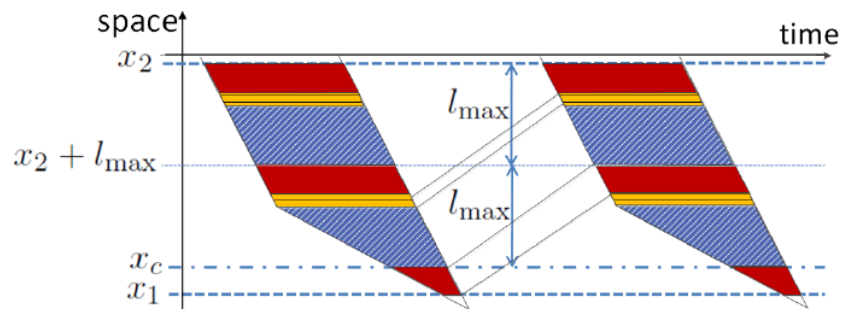
- Deux masses de probabilité
 - Véhicules s'arrêtant n_s fois:
 - délai $n_s R$
 - Véhicules s'arrêtant $n_s - 1$ fois:
 - délai $(n_s - 1) R$



Pdf des délais entre x_1 et x_2 (régime congestionné)

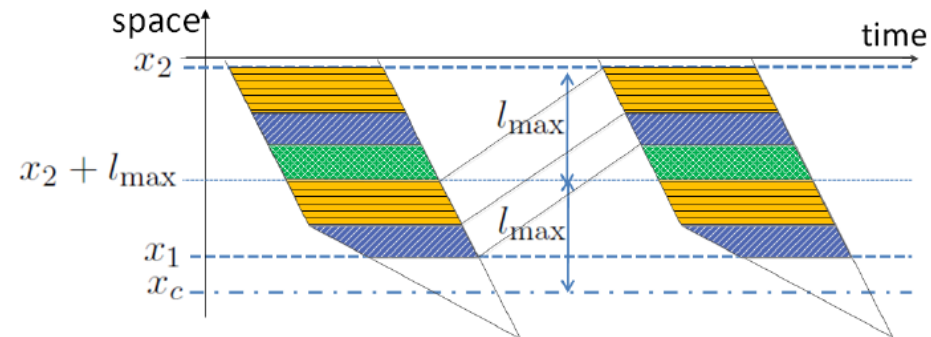
-Cas Numéro 4: $x_1 > l_r$, $x_2 < l_r$ $x_c = x_2 + n_s l_{\max}^s$

-Cas 4a: $x_c \leq x_1$



-  Vehicles stopping in the triangular queue and n_s times in the remaining queue
-  Vehicles stopping in the triangular queue and $n_s - 1$ times in the remaining queue
-  Vehicles stopping n_s times in the remaining queue

Cas 4b: $x_c \geq x_1$



-  Vehicles stopping in the triangular queue and $n_s - 1$ times in the remaining queue
-  Vehicles stopping n_s times in the remaining queue
-  Vehicles stopping $n_s - 1$ times in the remaining queue

➔ Mélanges entre des délais avec distribution uniforme et masse de probabilité



Pdf des temps de parcours entre x_1 et x_2

- Délais δ_{x_1, x_2} : distribution h
- Temps de parcours libre $y_{f; x_1, x_2}$: distribution φ_{x_1, x_2}^y

$$p_f \sim \varphi^p(p_f) \Rightarrow \varphi_{x_1, x_2}^y(y_{f; x_1, x_2}) = \varphi^p\left(\frac{y_{f; x_1, x_2}}{x_1 - x_2}\right) \frac{1}{x_1 - x_2}$$

- Temps de parcours $y_{x_1, x_2} = \delta_{x_1, x_2} + y_{f; x_1, x_2}$
(somme de variables aléatoires indépendantes)

$$g^s(y_{x_1, x_2}) = \left(h^s * \varphi_{x_1, x_2}^y\right)(y_{x_1, x_2})$$



Mélange de distributions log-concaves

Si le pas à une distribution log-concave, les pdf de temps de parcours entre points x_1 et x_2 sont des mélange de distributions log-concaves avec au plus 5 modes.

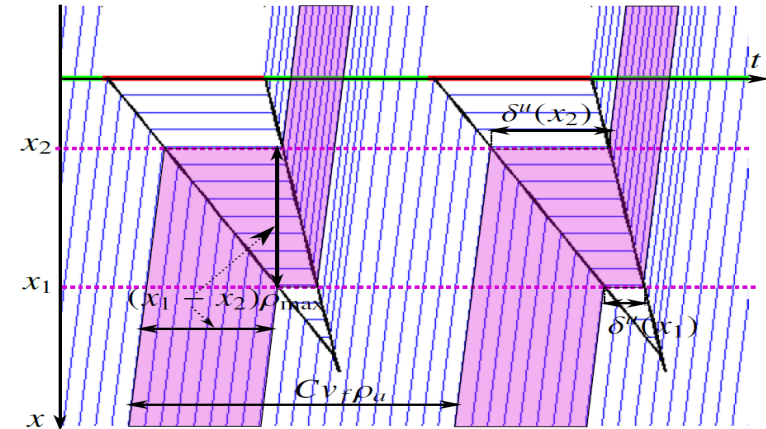
Chaque mode du mélange correspond à un « type » de délai.

- *Les distributions de délais sont des mélanges avec au plus 5 modes. Chaque mode a une distribution log-concave*
- *La log-concavité est fermée sous l'opération de convolution [Prékopa, 1973]*

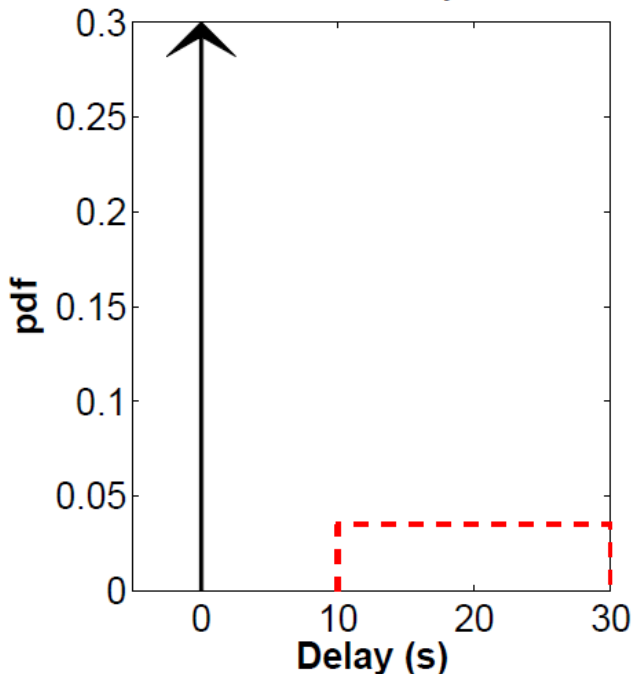
Illustration dans le régime non-saturé

Mélange à 2 modes

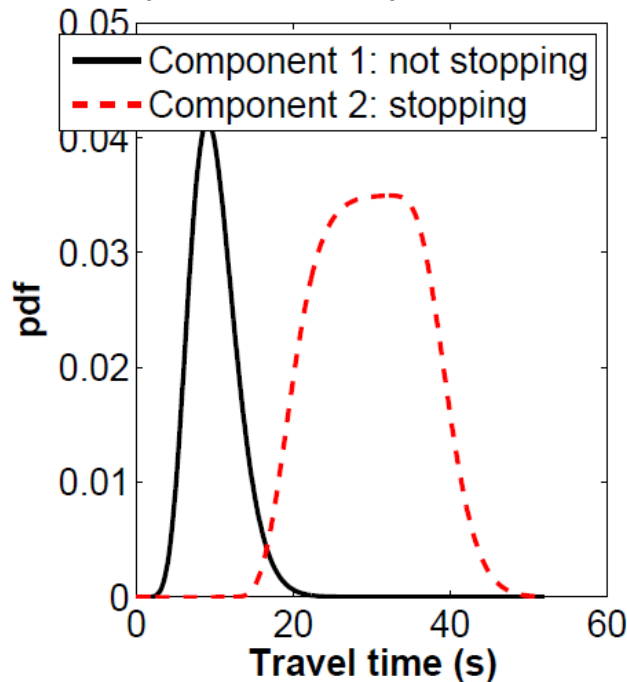
- Véhicules sans délai
- Véhicules avec délai uniforme



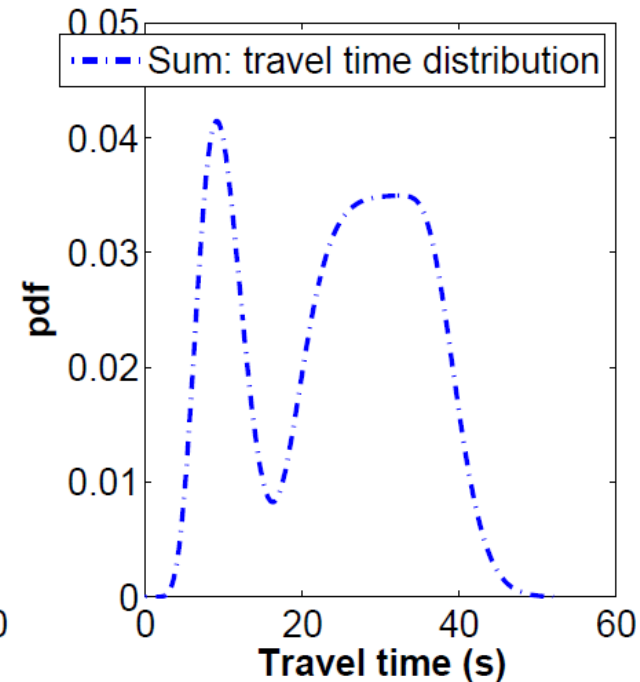
Pdf of delays



Components of the pdf of travel times



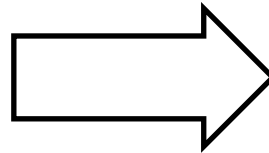
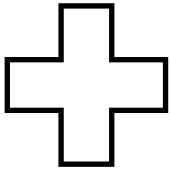
Pdf of travel times



L'estimation du trafic aujourd'hui

Modélisation statistique de la
dynamique du trafic

1. Probabilité de distribution des temps de parcours sur les arcs du réseau
2. **Dynamique du trafic à l'échelle du réseau**



**Estimation/
prédiction**

Données

Mesures fournissant des informations
sur l'état du système et/ou les
paramètres du modèle.

Modélisation de la dynamique sur le réseau

Etat de congestion: longueur de la queue:

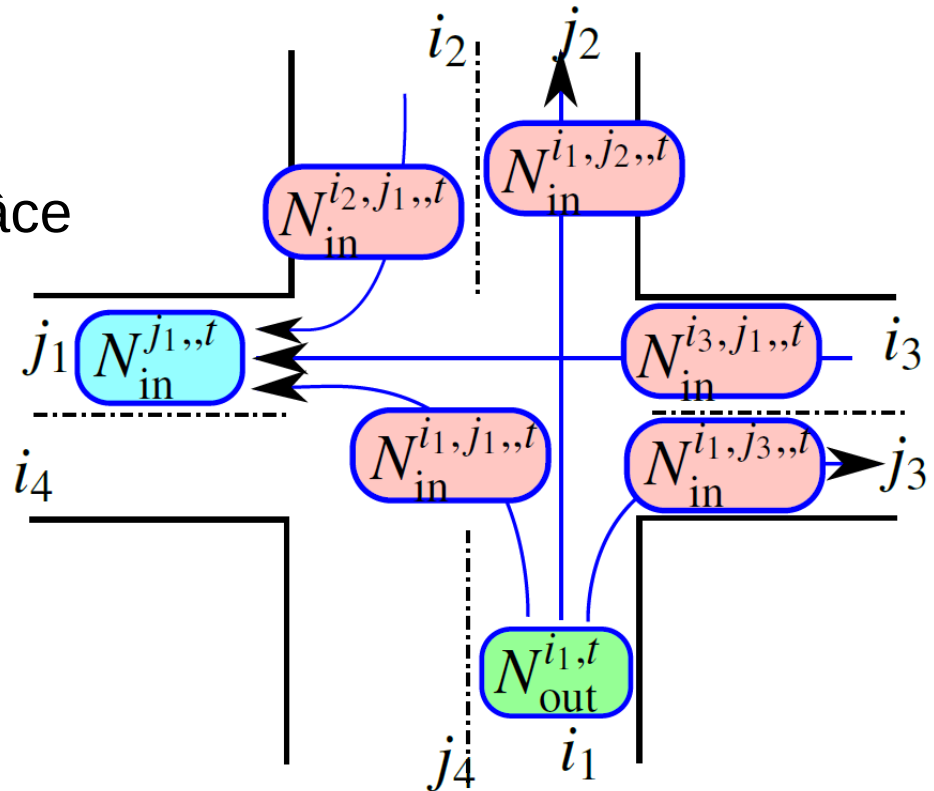
- Régime non saturé: $l_{\max}$
- Régime congestionné: $l_r + l_{\max}^s$

Etat constant pendant les intervalles, évolution entre les intervalles

→ Etat constant par morceaux

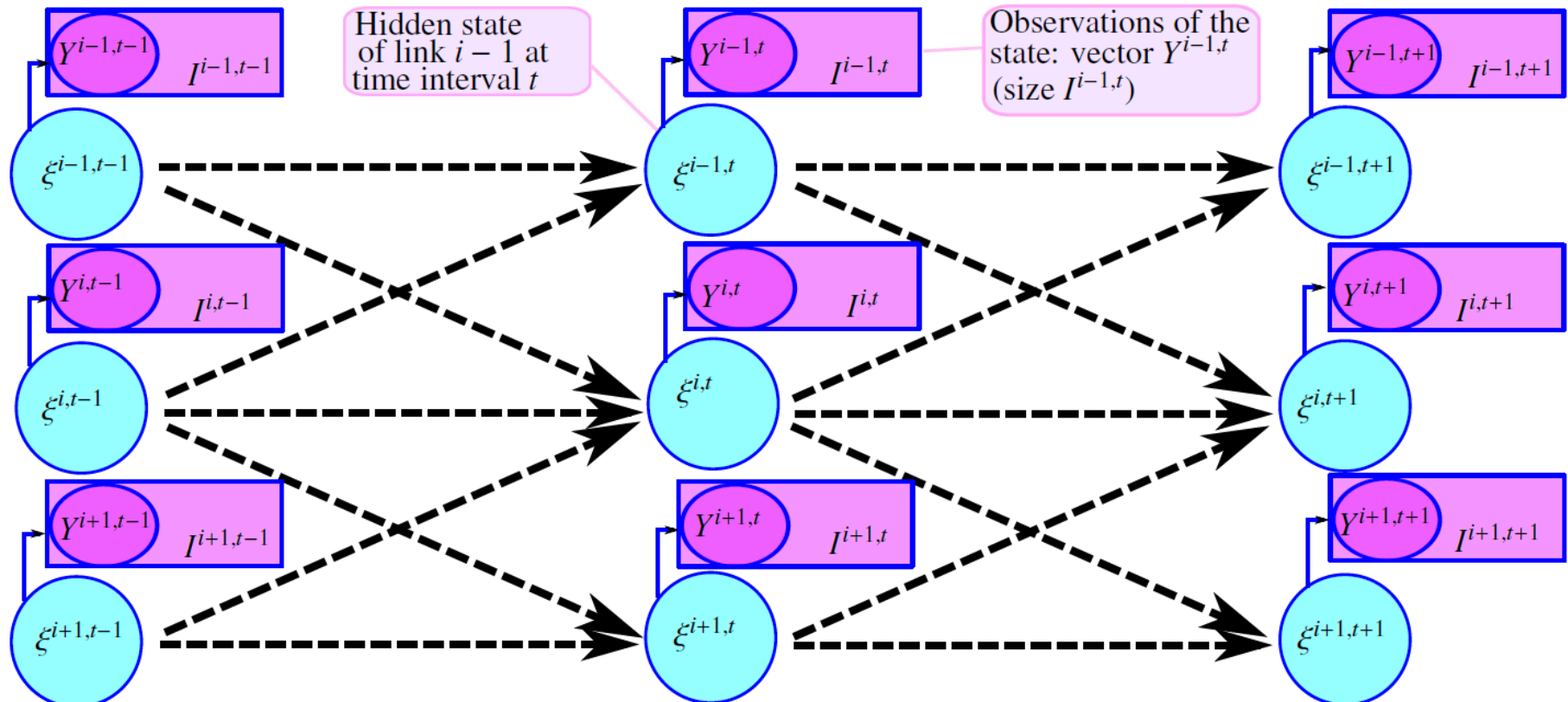
L'évolution de l'état dépend des arcs voisins et est modélisée grâce à la conservation des véhicules à chaque intersection

- Probabilité d'aller à l'arc j , venant de l'arc i : $\nu^{i,j}$
- Taux de création (source)
- Taux de départ (puit)



Réseau dynamique Bayésien

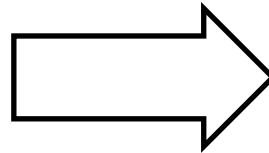
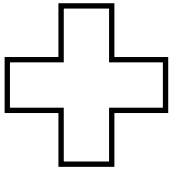
- **Etats de congestion**: longueur de la queue (non observés directement, cachés):
- **Temps de parcours** (observés au travers des données GPS)
 - dépendent des états de congestion sur le chemin
 - dépendent des paramètres des arcs du chemin ($R, C, I_{\max}^s, \varphi^p$)



L'estimation du trafic aujourd'hui

Modélisation mathématique
représentant l'évolution du
système

Modèle



**Estimation/
prédiction**

Données

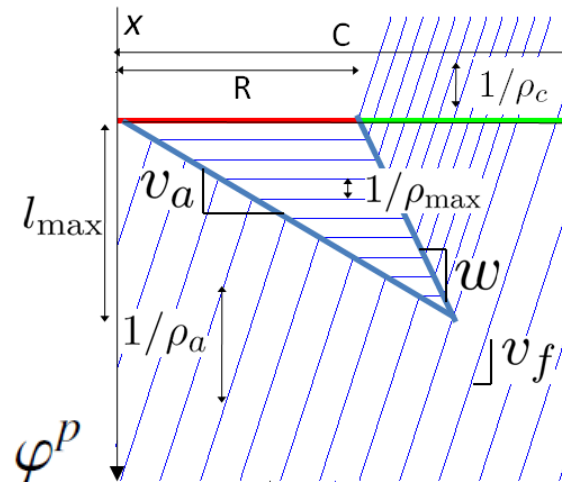
Mesures fournissant des informations
sur l'état du système et/ou les
paramètres du modèle.

Apprentissage des paramètres de la dynamique

- La dynamique est paramétrée par

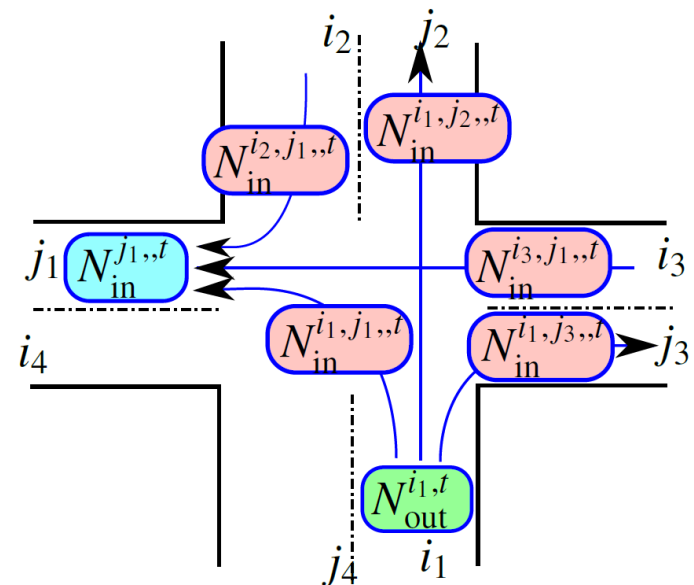
- **Paramètres de chaque arc:**

- Temps de rouge R ,
- Temps de cycle C ,
- Queue à saturation $l_{\max}^s$
- Distributions du pas (inverse de la vitesse) φ^p



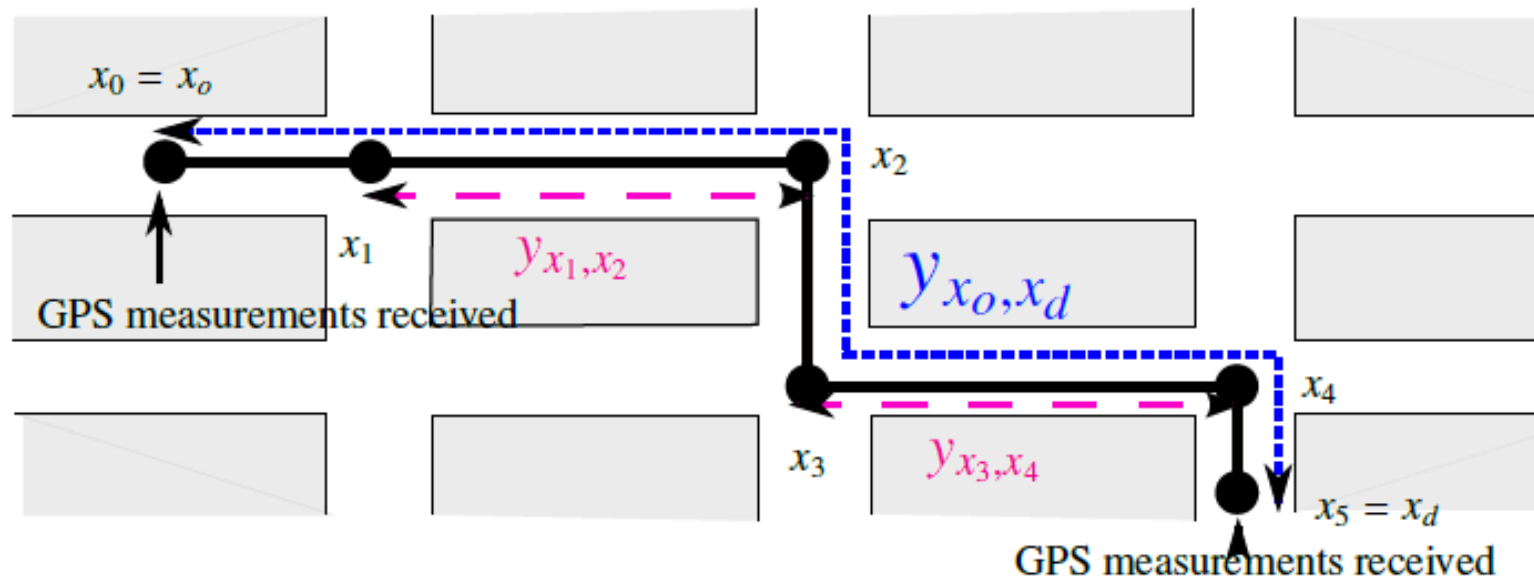
- **Paramètres des intersections:**

- Probabilité d'aller sur l'arc j provenant de l'arc i : $\gamma^{i,j}$
- Taux de création (source)
- Taux de départ (puits)



Décomposition de temps de parcours

- Le véhicule peut traverser plusieurs arcs entre deux positions successives
- Décomposition du temps de parcours total en temps de parcours par arc [Hellinga, 2008]
 - Proportionnellement à la distance parcourue?
 - En maximisant la vraisemblance de la décomposition?



Décomposition de temps de parcours

- Décomposition « optimale » **minimize** : $\sum_{m=0}^M -\ln(g^{i_m}(y_{x_m, x_{m+1}}))$
s.t. : $y_{x_o, x_d} = \sum_{m=0}^M y_{x_m, x_{m+1}}$
- Utilisation de la structure des pdf de temps de parcours (mélange de distributions log-concaves) pour trouver l'optimum global

$$\begin{aligned} \text{minimize : } & \sum_{m=0}^M \sum_{k=0}^{K_{i_m}} -\beta_{i_m, k} \ln(g^{i_m, k}(y_{x_m, x_{m+1}})) \\ & (y_{x_m, x_{m+1}})_m \\ & (\beta_{i_m, k})_{m, k} \\ \text{s.t. : } & y_{x_o, x_d} = \sum_{m=0}^M y_{x_m, x_{m+1}} \quad \beta_{i_m, k} \in \{0, 1\}, \quad \sum_{k=1}^{K_{i_m}} \beta_{i_m, k} = 1. \end{aligned}$$

$\beta_{i_m, k} = 1$ si le véhicule a un délai de type k sur l'arc i_m .



Apprentissage de la dynamique

- Maximisation de la vraisemblance des observations par rapport aux paramètres de la dynamique.



Apprentissage de la dynamique

- Maximisation de la vraisemblance des observations par rapport aux paramètres de la dynamique.
- **MAIS**: la vraisemblance des observations est une fonction difficile à maximiser:
 - Logarithme d'une probabilité marginale (somme sur les états de congestion)

Apprentissage de la dynamique

- Maximisation de la vraisemblance des observations par rapport aux paramètres de la dynamique.
- **MAIS**: la vraisemblance des observations est une fonction difficile à maximiser:
 - Logarithme d'une probabilité marginale (somme sur les états de congestion)
- **PAR AILLEURS**:
 - **Etant donné**: les paramètres de la dynamique, **on peut**: estimer les états cachés (longueur de la queue de chaque arc à chaque intervalle) grâce aux observations GPS
 - **Etant donné**: les états cachés, **on peut**: calculer les paramètres de la dynamique maximisant la vraisemblance des observations GPS





Apprentissage de la dynamique: **algorithme EM**

- *Preuve mathématique* de la maximisation itérative de la fonction de vraisemblance grâce à l'algorithme EM

**Initialization
des
paramètres**



Apprentissage de la dynamique: **algorithme EM**

- *Justification mathématique* de la maximisation itérative de la fonction de vraisemblance:

Initialization
des
paramètres



Calcul des probabilités des
états de congestions **étant
donné** les paramètres de la
dynamique

E step:
espérance



Apprentissage de la dynamique: **algorithme EM**

- *Justification mathématique* de la maximisation itérative de la fonction de vraisemblance:

Initialization
des
paramètres

Calcul des probabilités des
états de congestions **étant
donné** les paramètres de la
dynamique

E step:
espérance

M step:
maximisation

Maximisation par rapport
aux paramètres **étant
donné** les probabilités de
chaque état de congestion





Apprentissage de la dynamique: **algorithme EM**

- *Justification mathématique* de la maximisation itérative de la fonction de vraisemblance:

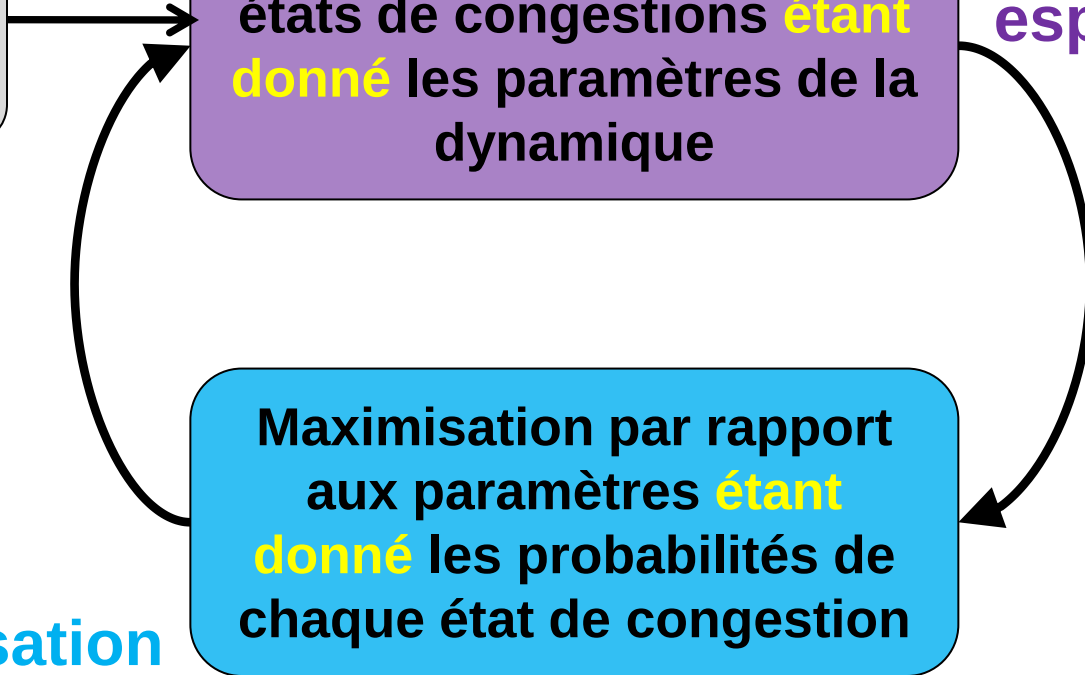
Initialization
des
paramètres

Calcul des probabilités des
états de congestions **étant
donné** les paramètres de la
dynamique

E step:
espérance

Maximisation par rapport
aux paramètres **étant
donné** les probabilités de
chaque état de congestion

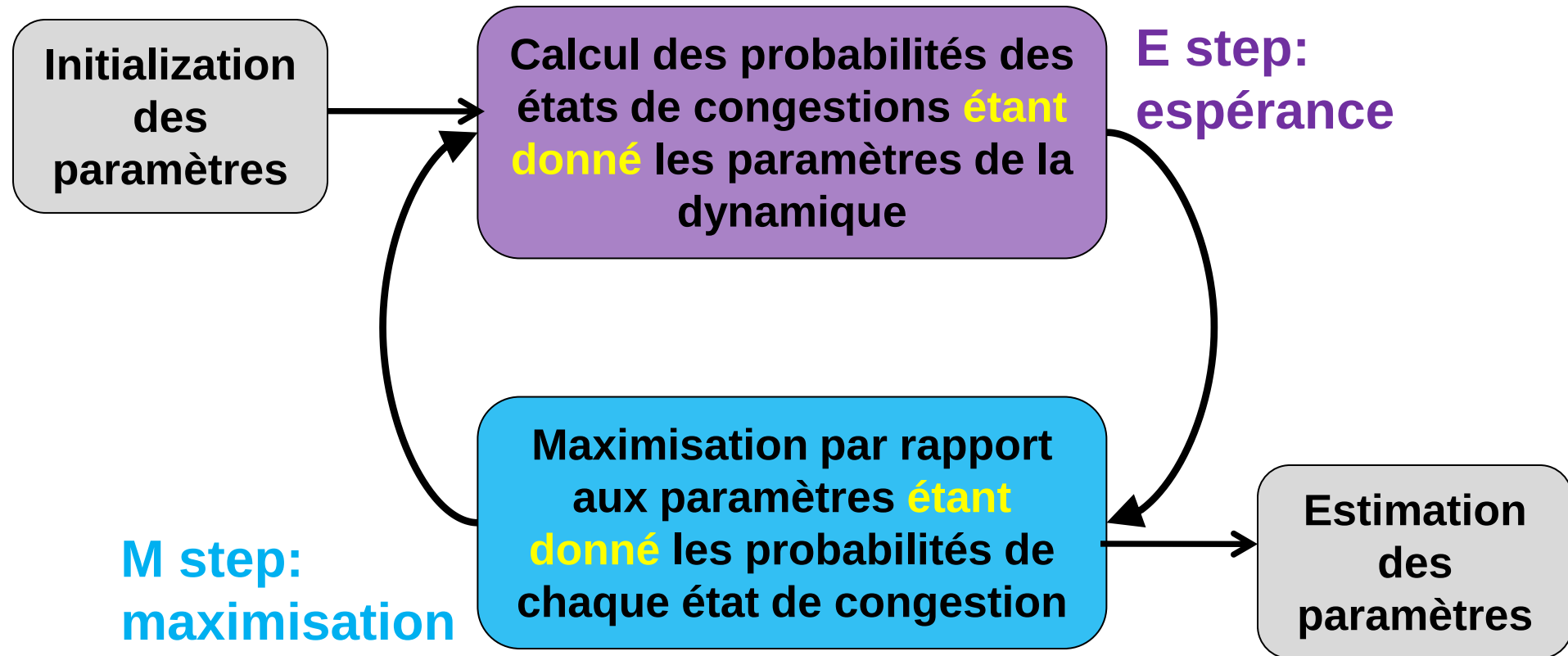
M step:
maximisation





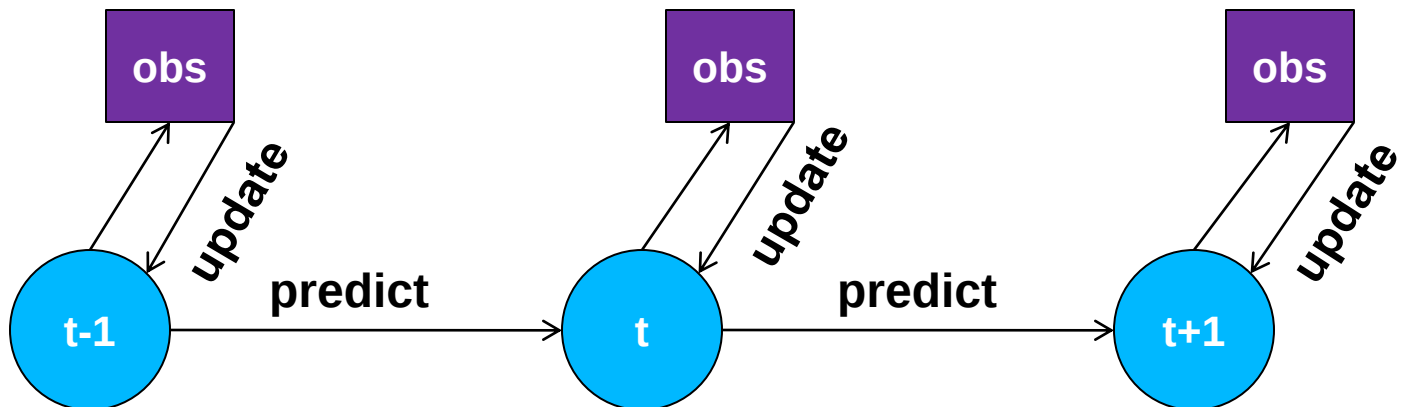
Apprentissage de la dynamique: **algorithme EM**

- *Justification mathématique* de la maximisation itérative de la fonction de vraisemblance:



Estimation/prédiction en temps réel

- Estimation de la probabilité jointe des états de congestion à *l'intervalle t* , étant donné les paramètres de la dynamique et les observations disponibles jusqu'à l'intervalle t : **filtrage/inférence**
- **2 Etapes:**
 - **Prédiction** à t étant donné l'état à $t-1$ et les paramètres de transition
 - **Mise à jour** de l'état à t étant donné la prédiction et les observations disponibles à t





Estimation/prédiction en temps réel

- **MAIS...**



Estimation/prédiction en temps réel

- **MAIS...**
- La densité jointe des états cachés (états de congestion) a une représentation qui croît ***exponentiellement*** avec le nombre d'arcs (malgré la structure du réseau Bayésien)

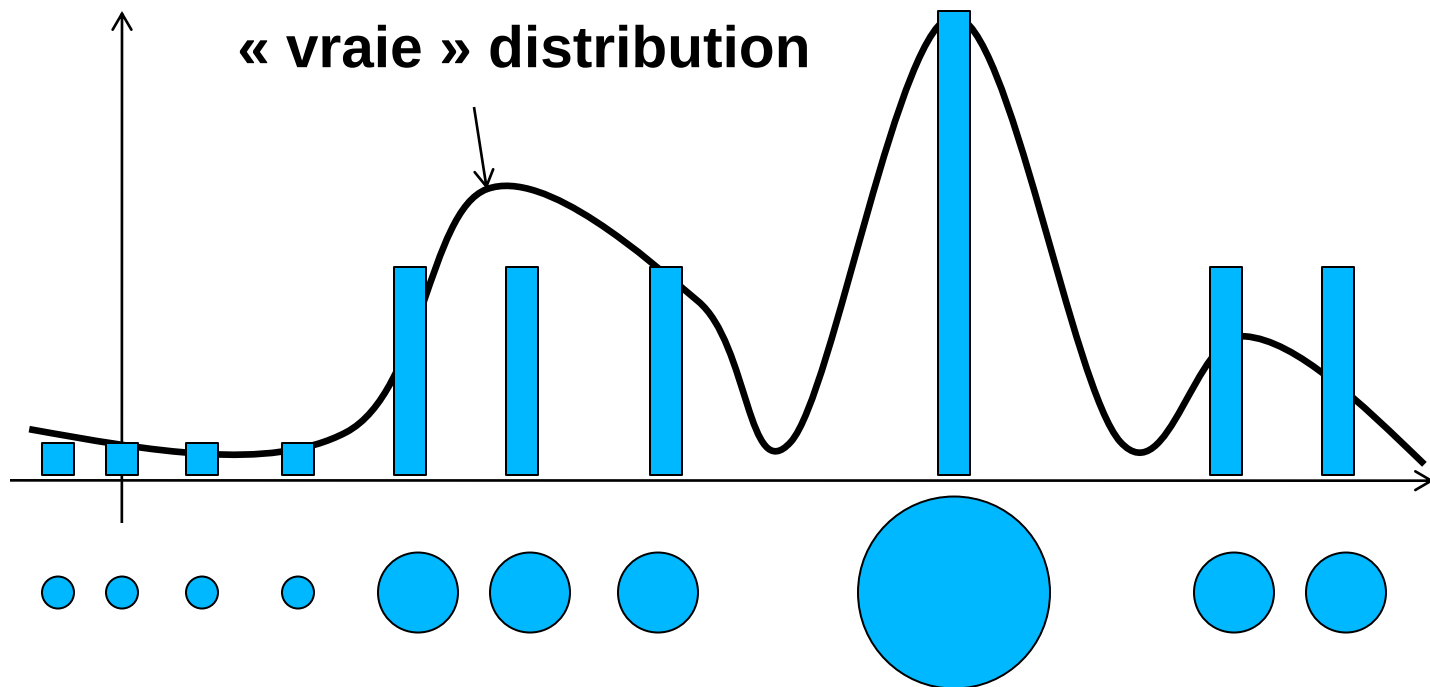


Estimation/prédiction en temps réel

- **MAIS...**
- La densité jointe des états cachés (états de congestion) a une représentation qui croît ***exponentiellement*** avec le nombre d'arcs (malgré la structure du réseau Bayésien)
- ➔ Nécessité de recourir à une approximation pour résoudre le problème d'inférence
 - Belief propagation (exemple pour le trafic: Furtlehner, ITSC 07)
 - Approximate inference [Boyen Koller, 1999]
 - Filtres à particules

L'intuition d'un filtre à particules

- Approximation d'une distribution « compliquée » grâce à des particules
- Chaque particule a un « poids » qui représente son importance



Approximation grâce aux particules



Dans le cas du trafic

- Une **particule** représente une « **instanciation** » de l'état **de congestion** de l'ensemble du réseau
 - Le **poids** de la particule représente la **probabilité** de cette instanciation, étant donné les observations de temps de parcours
- ➔ Les particules évoluent dans le réseau Bayésien et approximent les deux étapes du filtrage
- Prédiction
 - Mise à jour



L'algorithme de filtre à particule pour le trafic

- Echantillonner P particules aléatoirement, représentant l'état de congestion à l'instant initial
- A chaque intervalle de temps

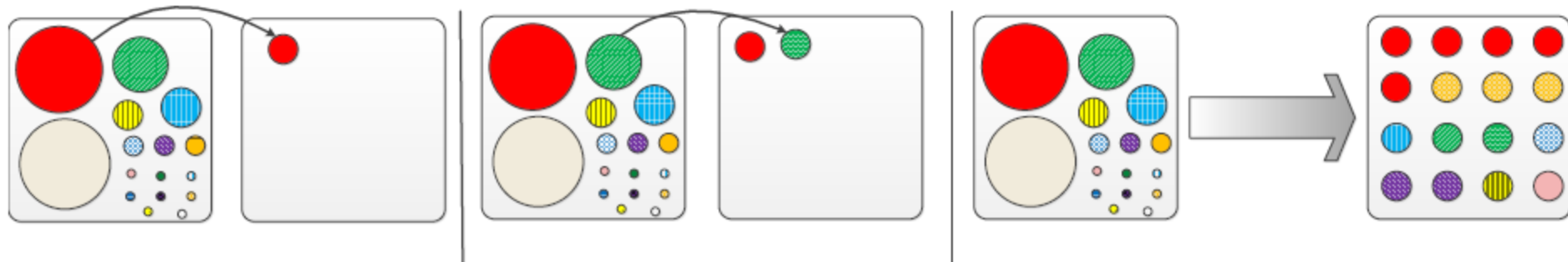
Mise à jour

- Multiplier le poids de la particule par la probabilité des observations étant donné l'instanciation de l'état de congestion de la particule
- Normaliser les poids des particules afin qu'ils somment à 1.
- Ré-échantillonner les particules afin d'éviter la dégénérescence
- Transition vers l'intervalle suivant

Prédiction

Ré-échantillonnage pour éviter la dégénérescence

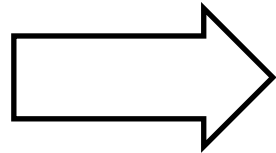
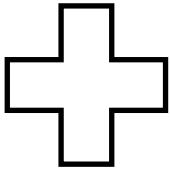
- **But:** éviter que de nombreuses particules aient un poids proche de zéro
- Inutile de garder ces particules qui n'apportent pas/peu d'information sur la distribution
- Échantillonner (avec remplacement) P particules. Chaque particule est choisie avec une probabilité égale à son poids (poids normalisés)



En résumé

Modélisation mathématique
représentant l'évolution du
système

Modèle

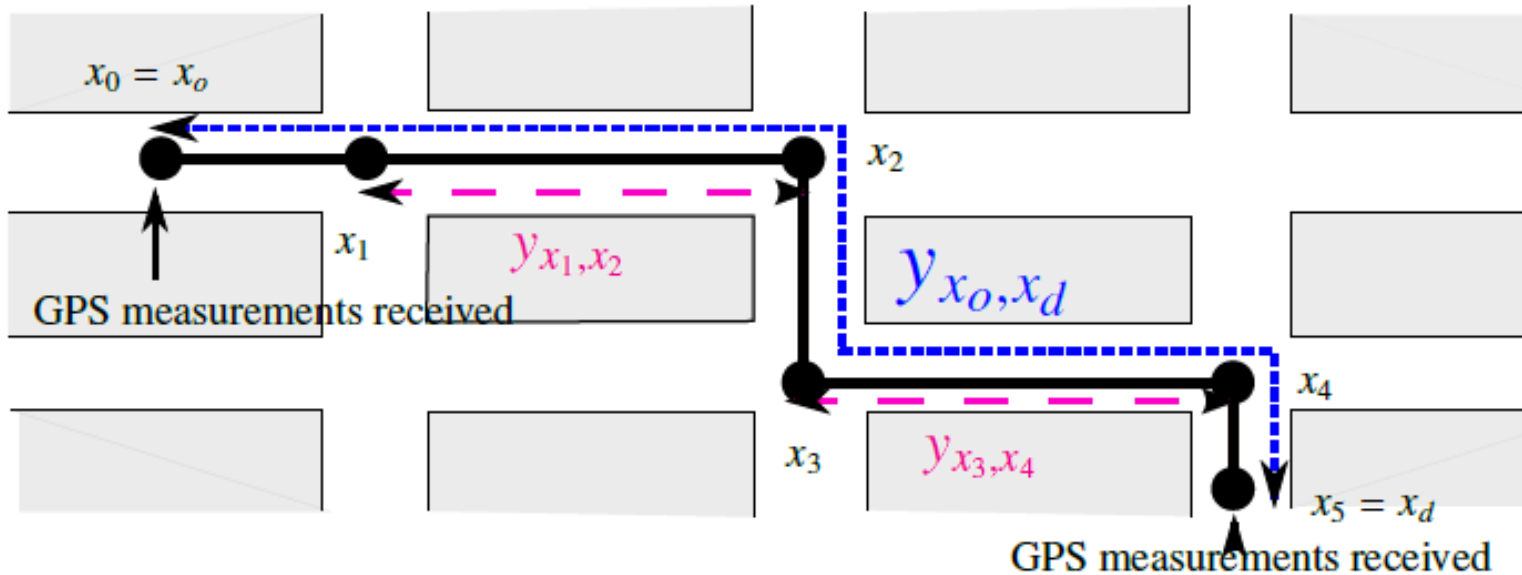


**Estimation/
prédiction**

Données

Mesures fournissant des informations
sur l'état du système et/ou les
paramètres du modèle.

En résumé



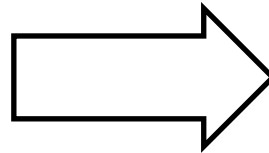
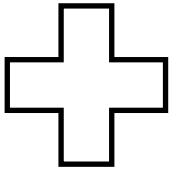
Données

Mesures fournissant des informations sur l'état du système et/ou les paramètres du modèle.

En résumé

Modélisation mathématique
représentant l'évolution du
système

Modèle

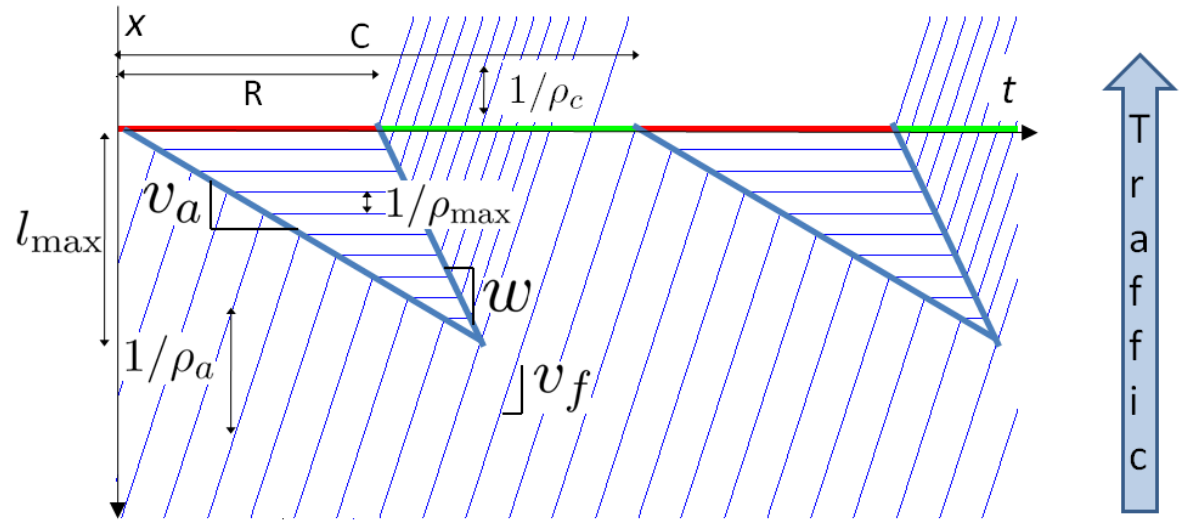
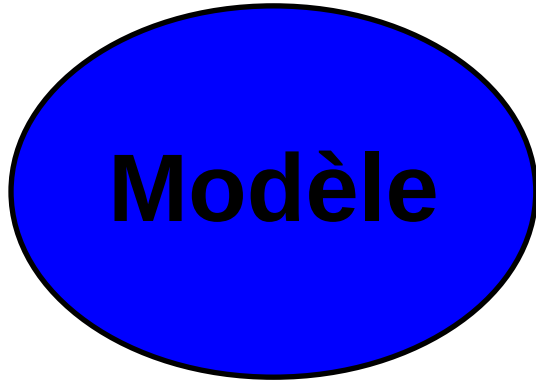


**Estimation/
prédiction**

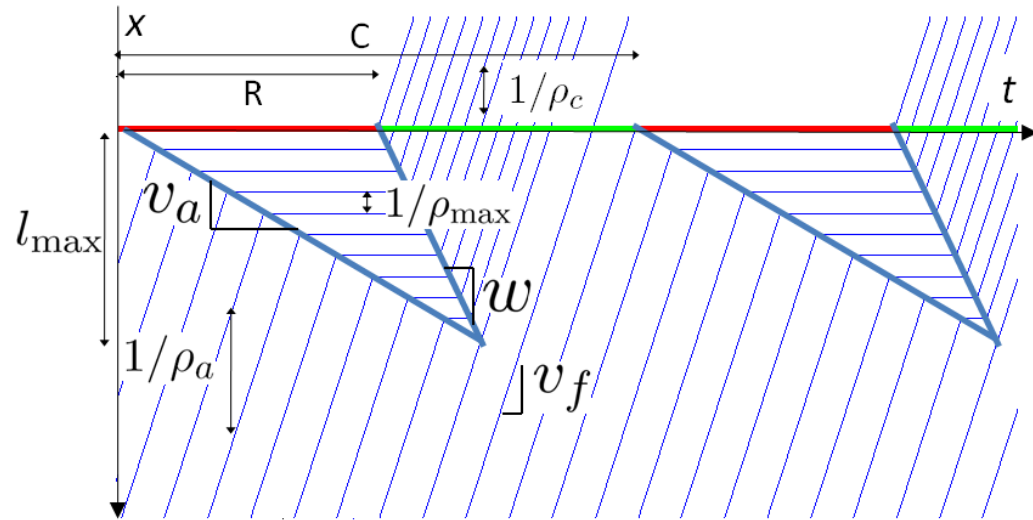
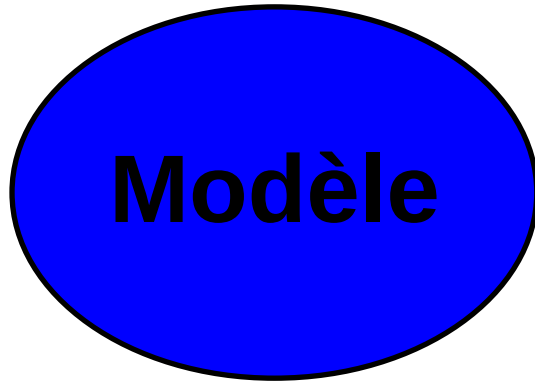
Données

Mesures fournissant des informations
sur l'état du système et/ou les
paramètres du modèle.

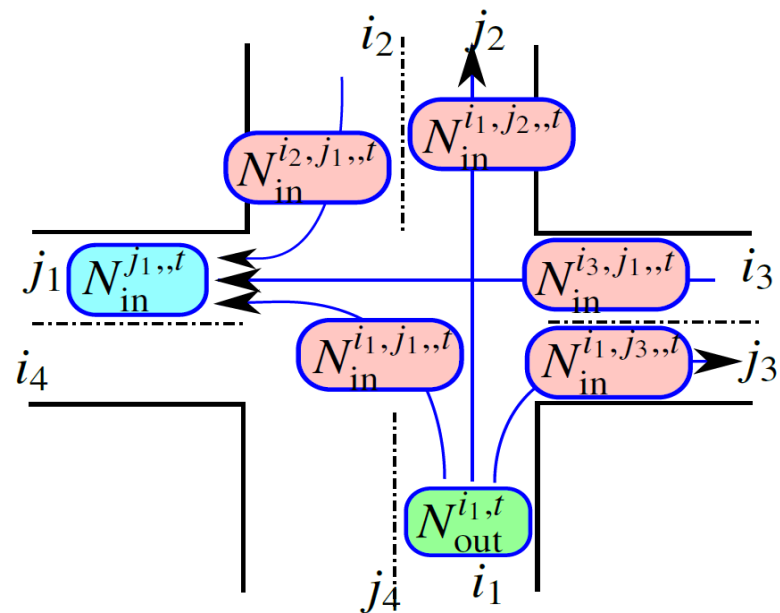
En résumé



En résumé

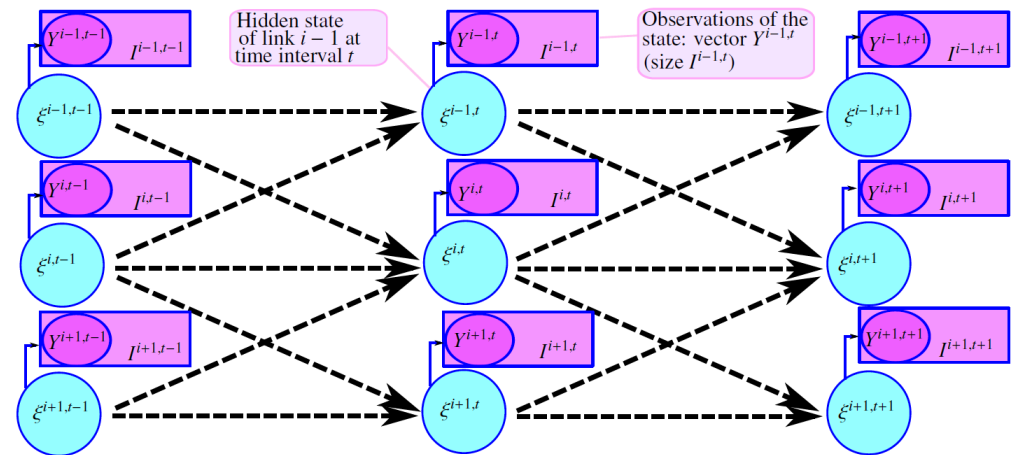
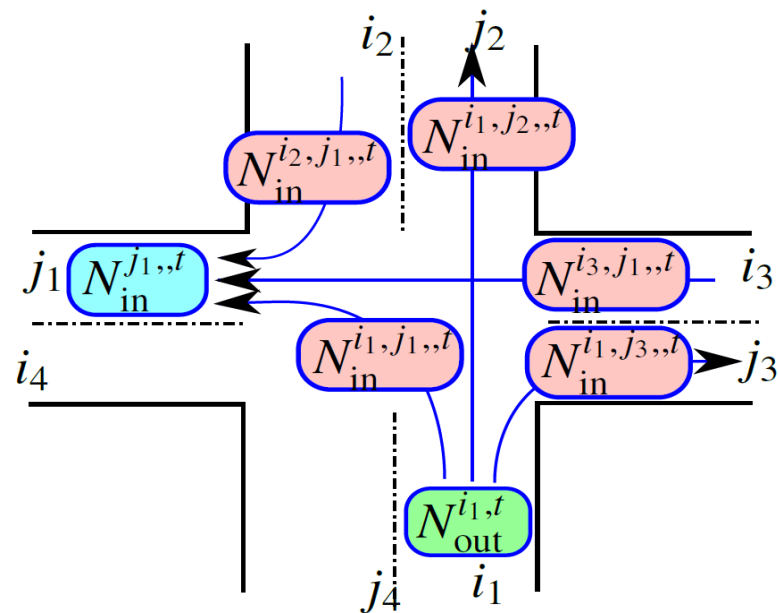
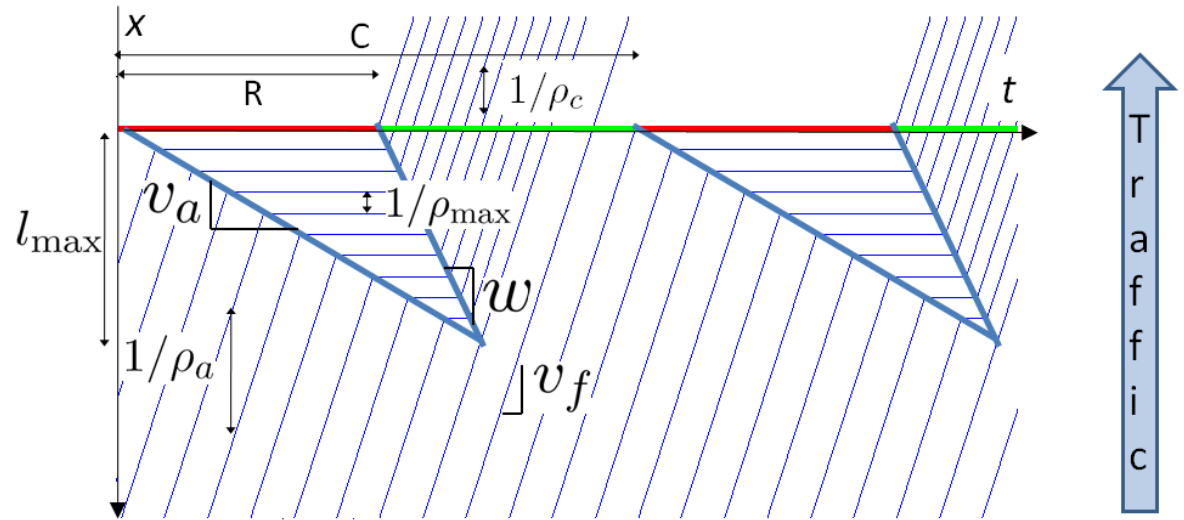


T
r
a
f
f
i
c



En résumé

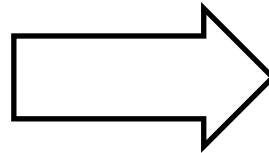
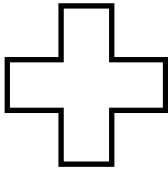
Modèle



En résumé

Modélisation mathématique
représentant l'évolution du
système

Modèle

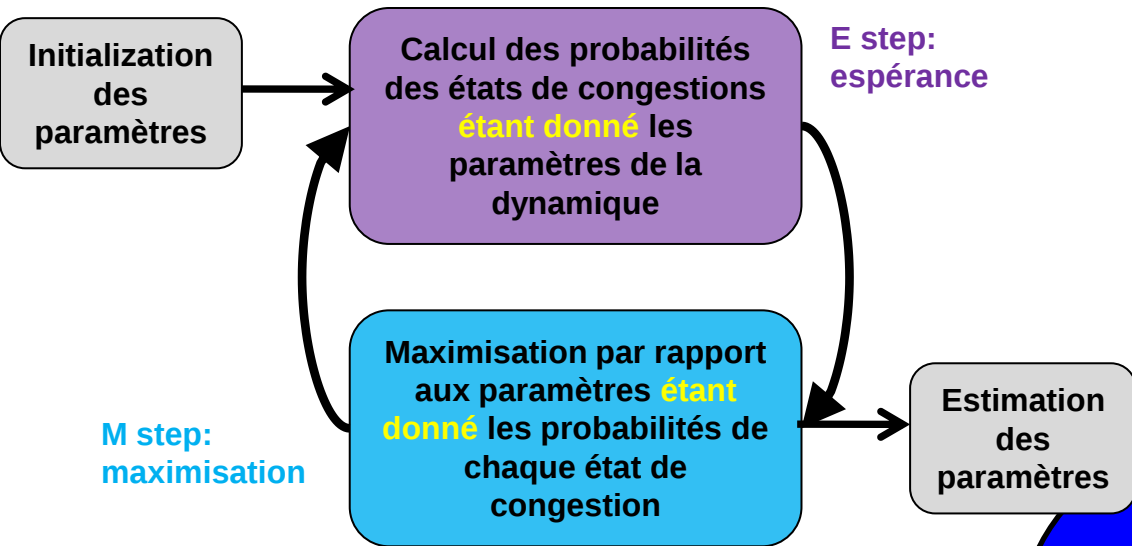


**Estimation/
prédiction**

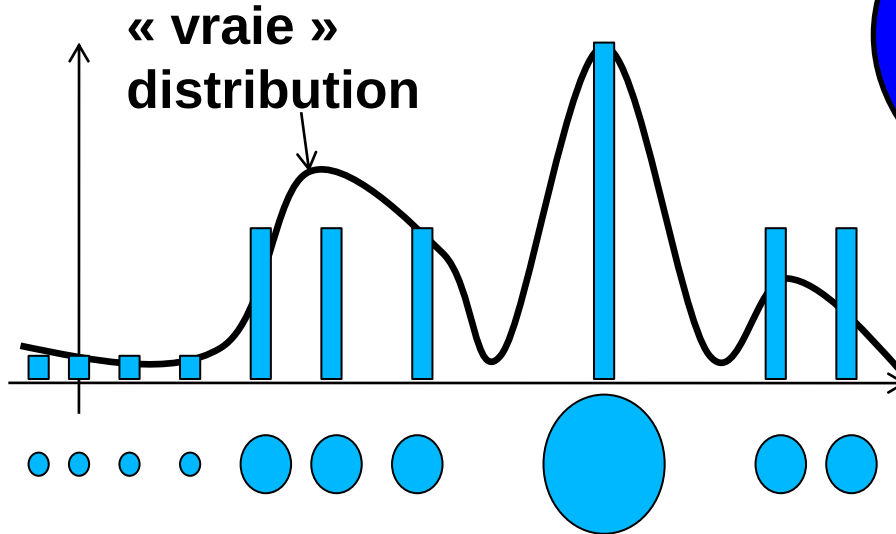
Données

Mesures fournissant des informations
sur l'état du système et/ou les
paramètres du modèle.

En résumé



**Estimation/
prédiction**

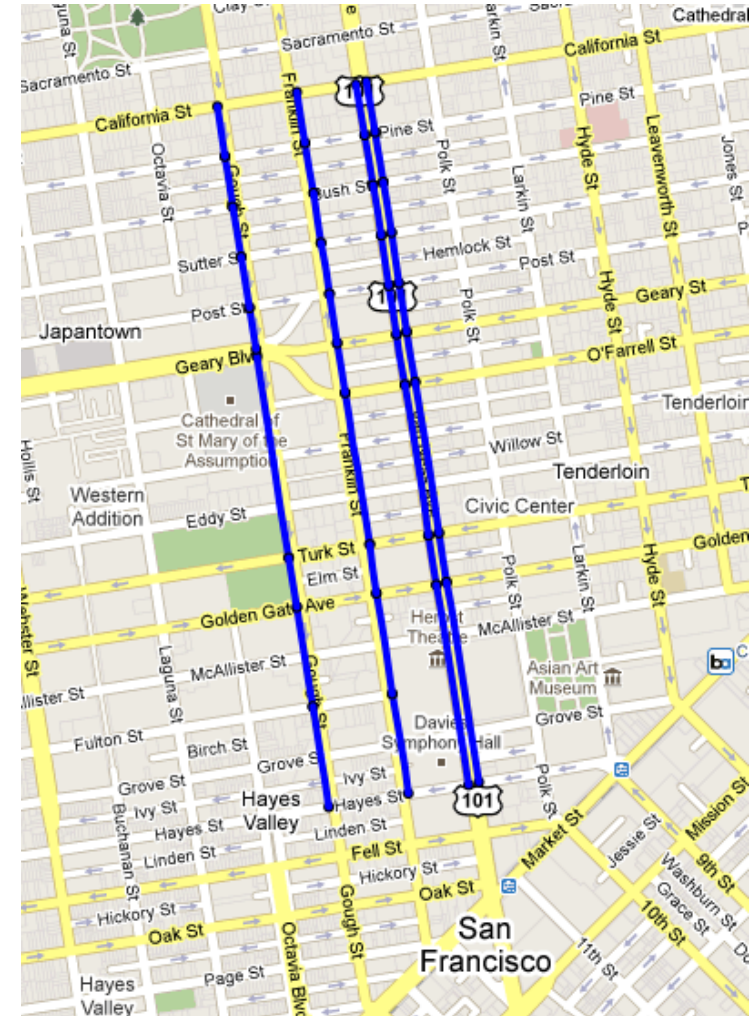


Approximation grâce aux particules

Estimation en réseau urbain (San Francisco)

Mardi 29 juin - jeudi 1er juillet 2010

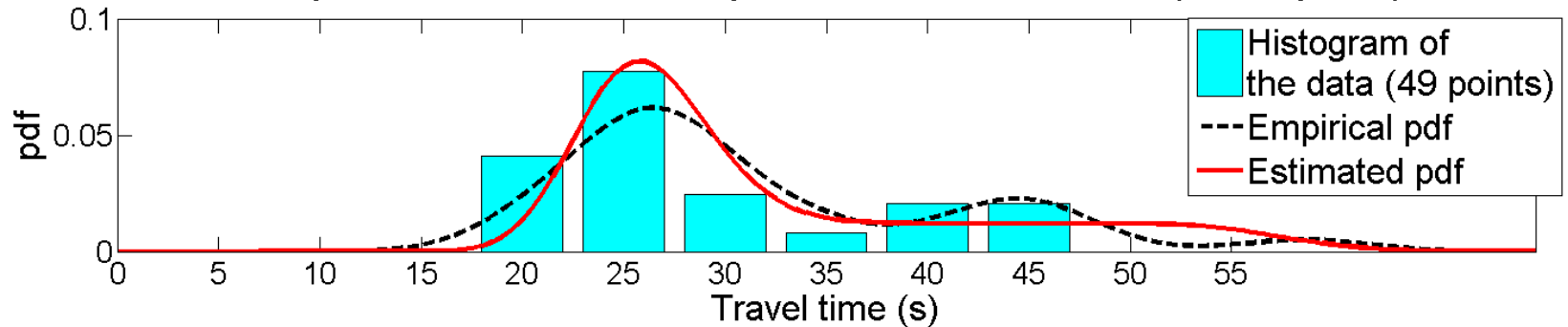
- Test avec 20 véhicules
- 3 heures de tests par jour



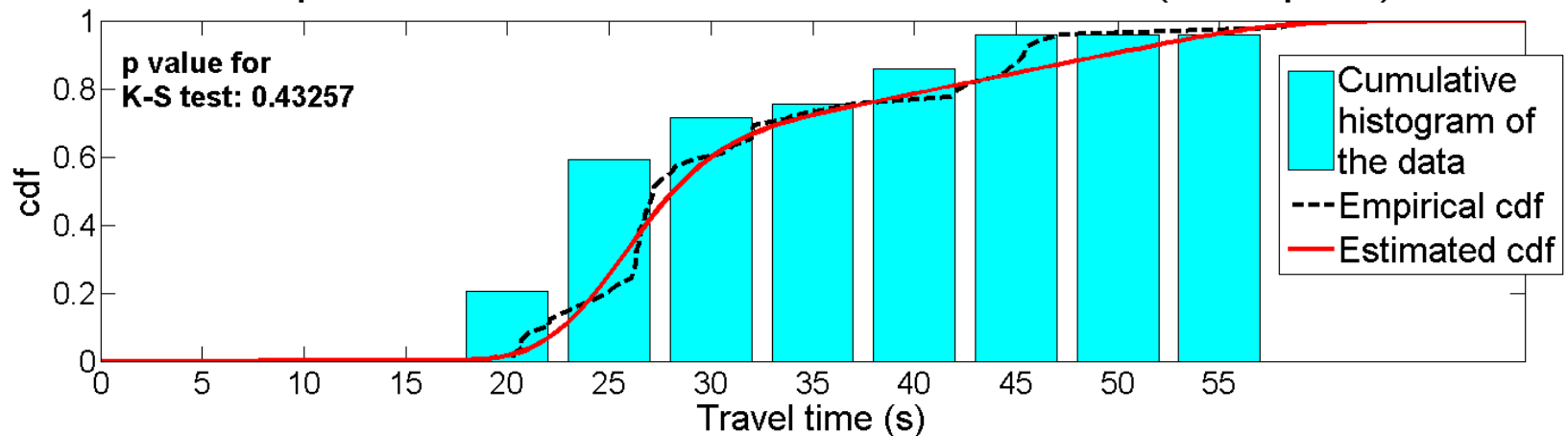
Validation des distributions de temps de parcours

Comparaison des distributions de temps de parcours
(estimations vs. données reçues pendant test)

Empirical and estimated pdf of the travel times (Example 1)



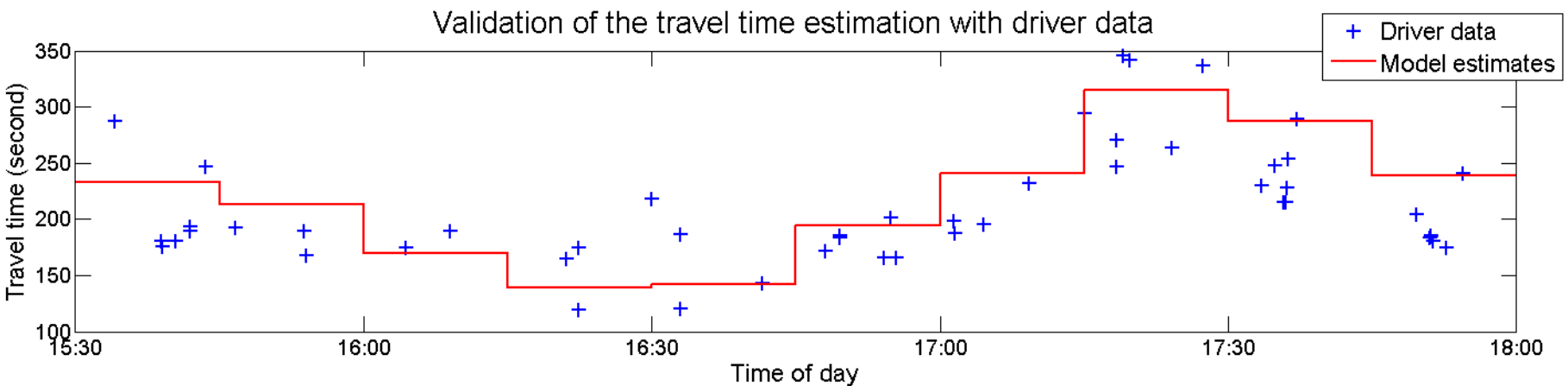
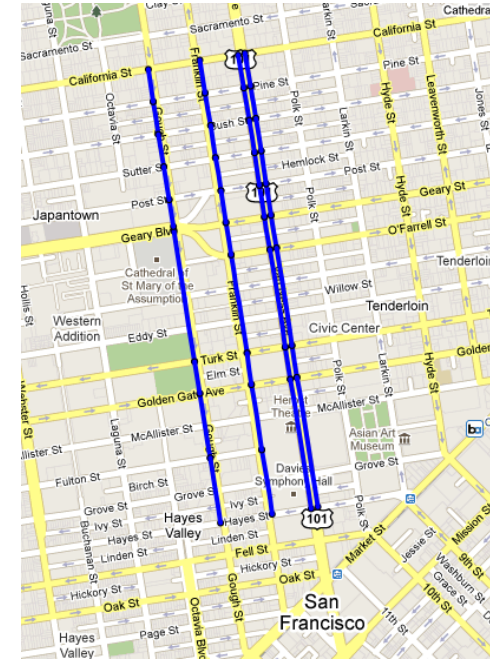
Empirical and estimated cdf of the travel times (Example 1)



Estimation en réseau urbain (San Francisco)

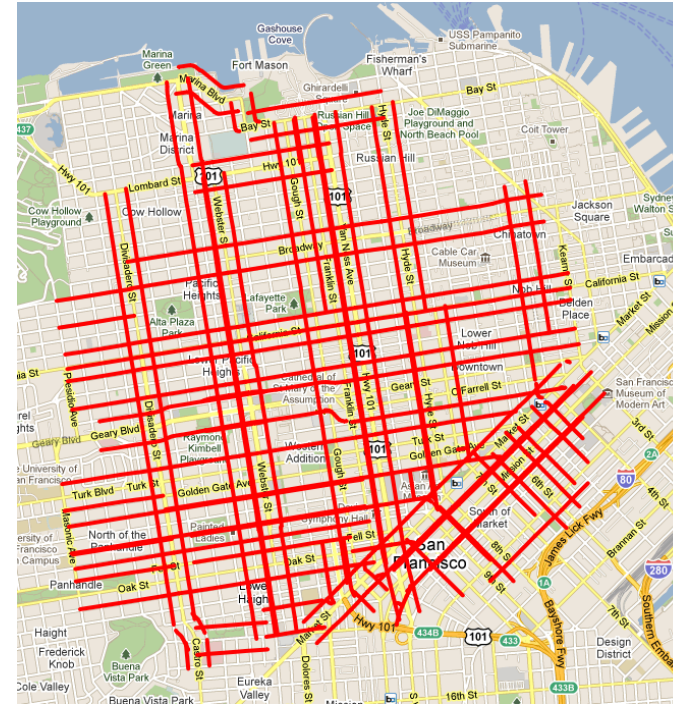
Estimation des temps de parcours sur l'ensemble de la route

- Comparaison avec les conducteurs
- Illustration de la variabilité des temps de parcours en réseau urbain!!!



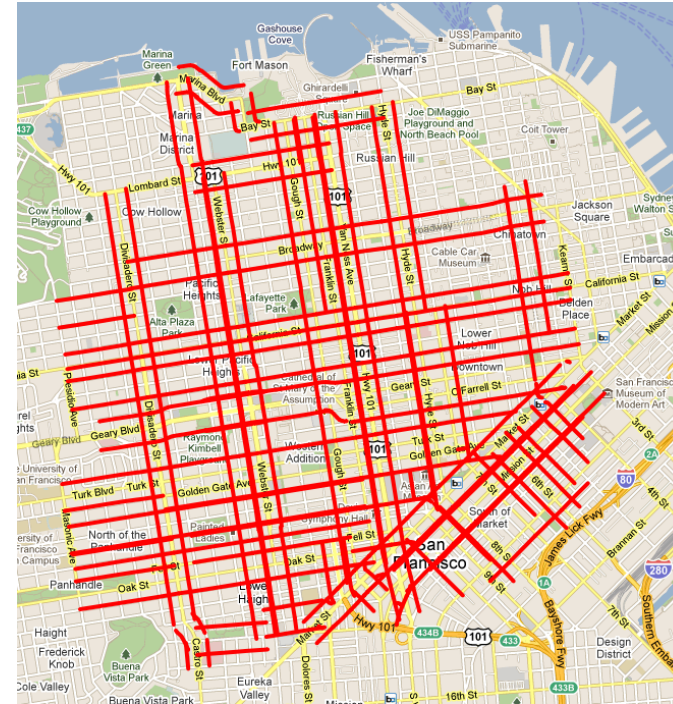
Validation à plus grande échelle

- Apprentissage de la dynamique
- Validation des résultats grâce à des données non utilisées pour l'apprentissage (« validation croisée »)
- Comparaison avec un modèle de type « série temporelle » utilisant la moyenne des observations.



Validation à plus grande échelle

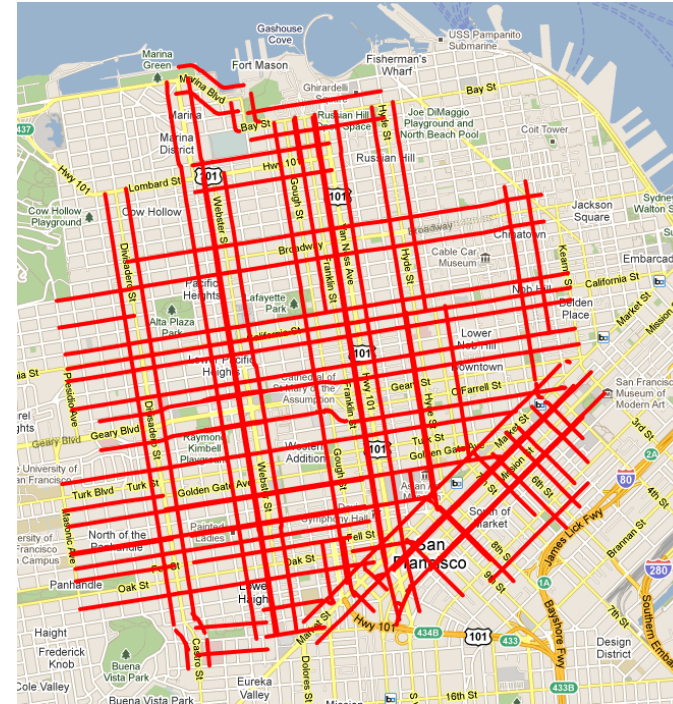
- Apprentissage de la dynamique
- Validation des résultats grâce à des données non utilisées pour l'apprentissage (« validation croisée »)
- Comparaison avec un modèle de type « série temporelle » utilisant la moyenne des observations.



	RMSE	MAE
Traffic Model	25.41	20.23
Baseline Model	31.56	25.69
Improvement (%)	16.32	17.34

Validation à plus grande échelle

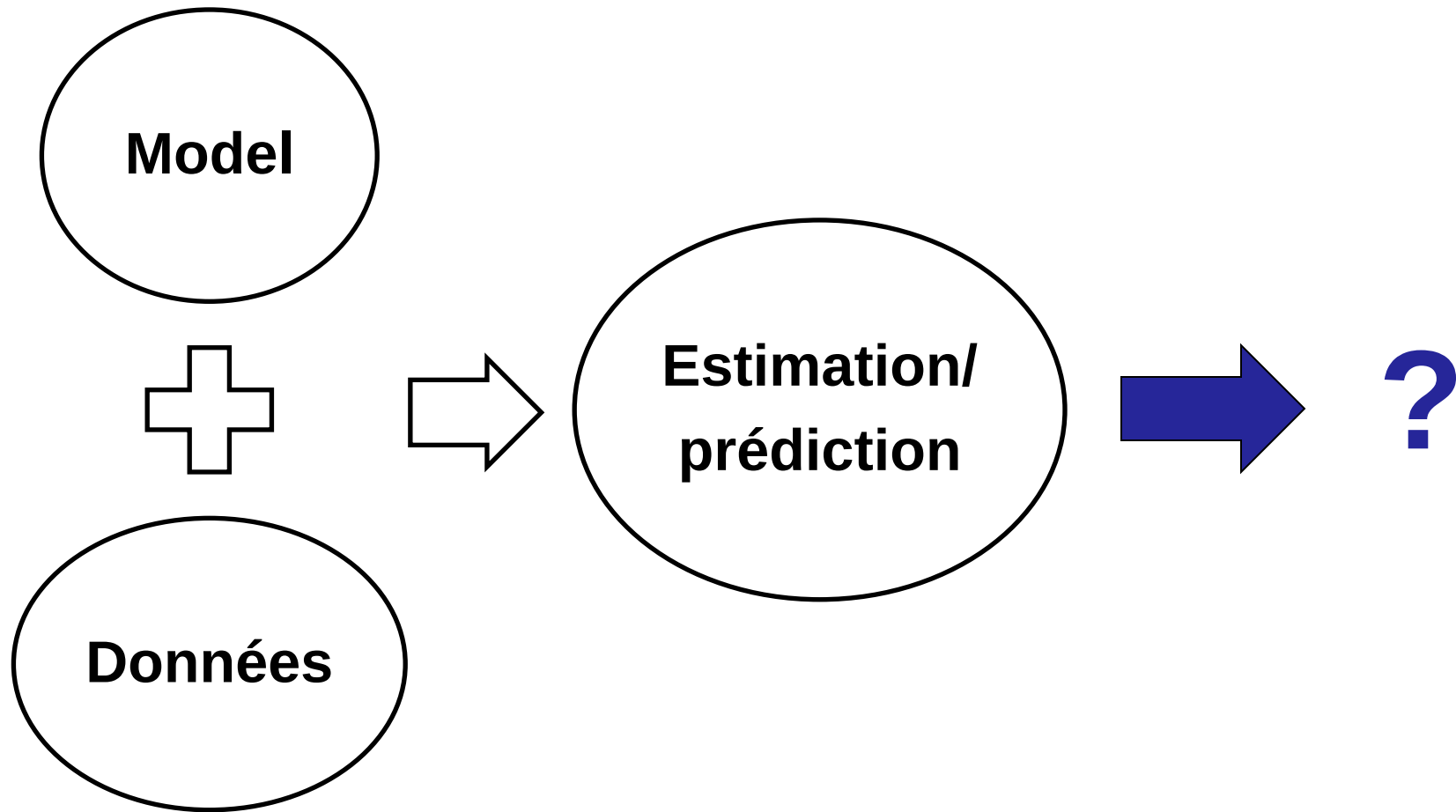
- Apprentissage de la dynamique
- Validation des résultats grâce à des données non utilisées pour l'apprentissage (« validation croisée »)
- Comparaison avec un modèle de type « série temporelle » utilisant la moyenne des observations.



- ➔ Amélioration des erreurs l1 et l2
- ➔ Capacités de prédictions
- ➔ Estimation de la distribution (et non de la moyenne uniquement)
- ➔ Estimation des paramètres du trafic

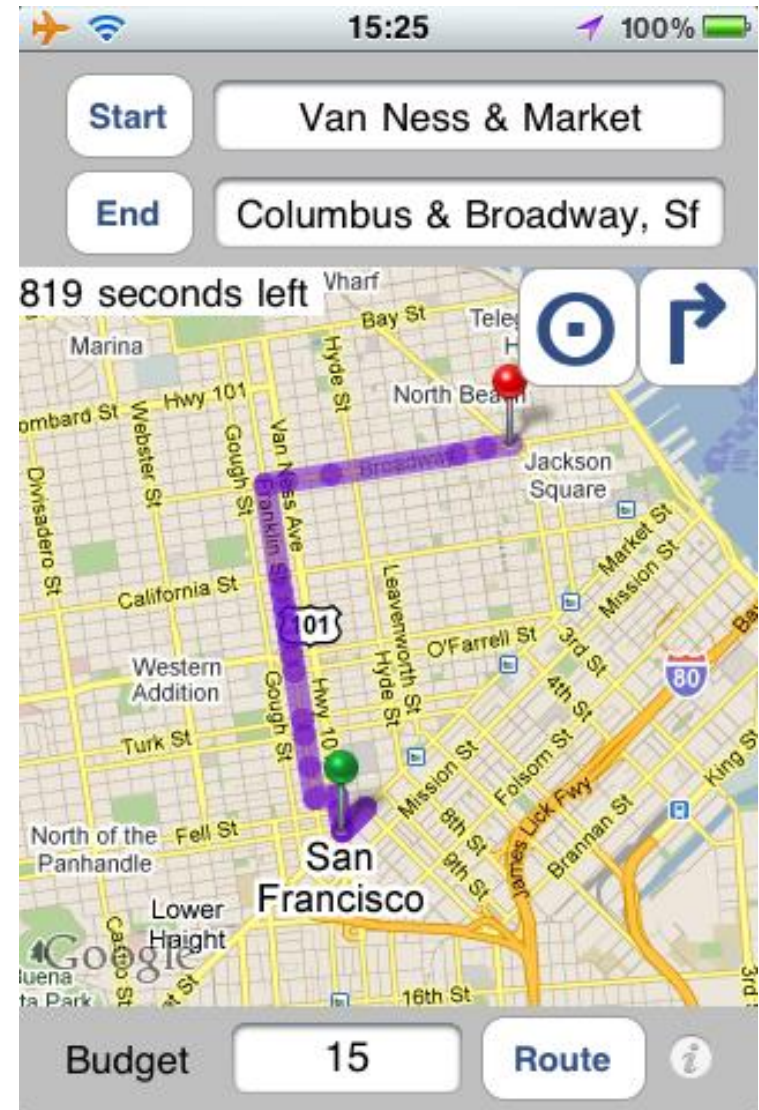
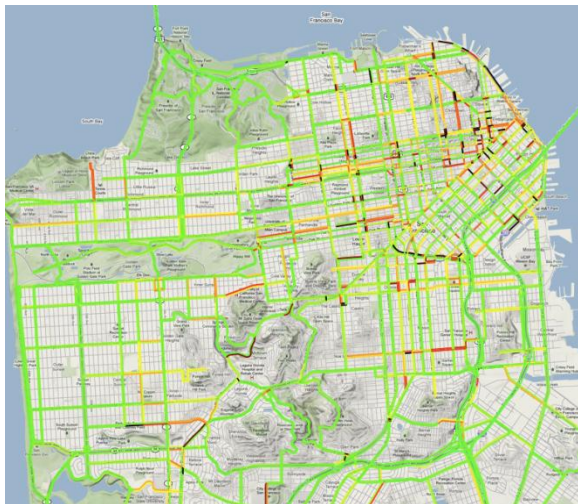
[Hofleitner, Herring, Bayen, Transportation Research B, soumis 2011]

Et après l'estimation du trafic?



Choix d'itinéraires

- Stochastic On Time Arrival (SOTA)
- Disponible pour iPhone:
 - Estimations en temps réel des conditions de circulation sur plus de 2600 arcs du réseau
- Protocole de communication
 - Itinéraire optimal envoyé au téléphone
 - Mise à jour si les conditions de circulation évoluent

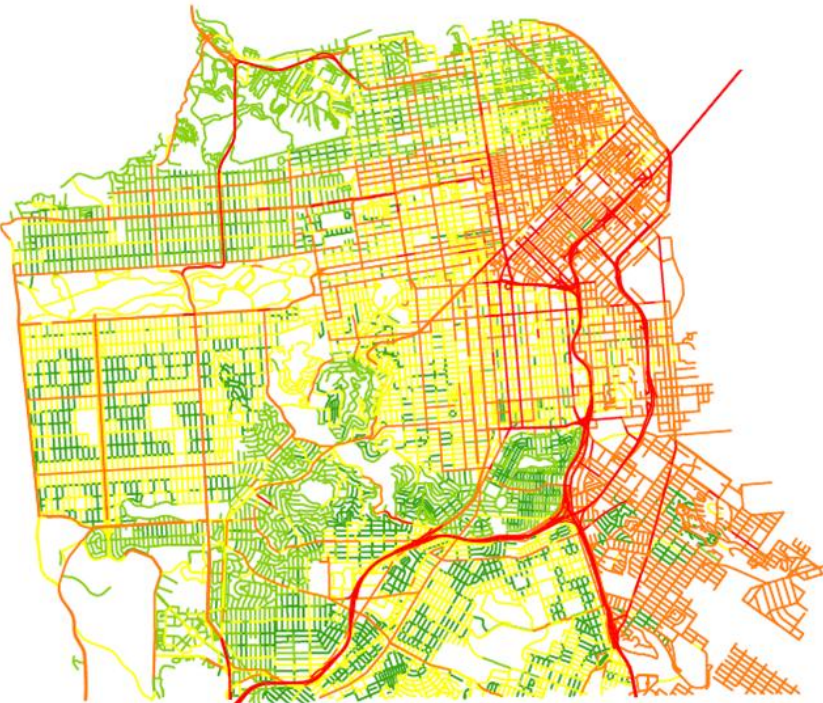


Mobile Millennium: au-delà du trafic

“e-Wellness”

- Niveaux de bruits induits par le trafic: estimation en temps réel. Les études actuelles utilisent des données historiques de nombre véhicules moyen

Aujourd'hui: carte statique



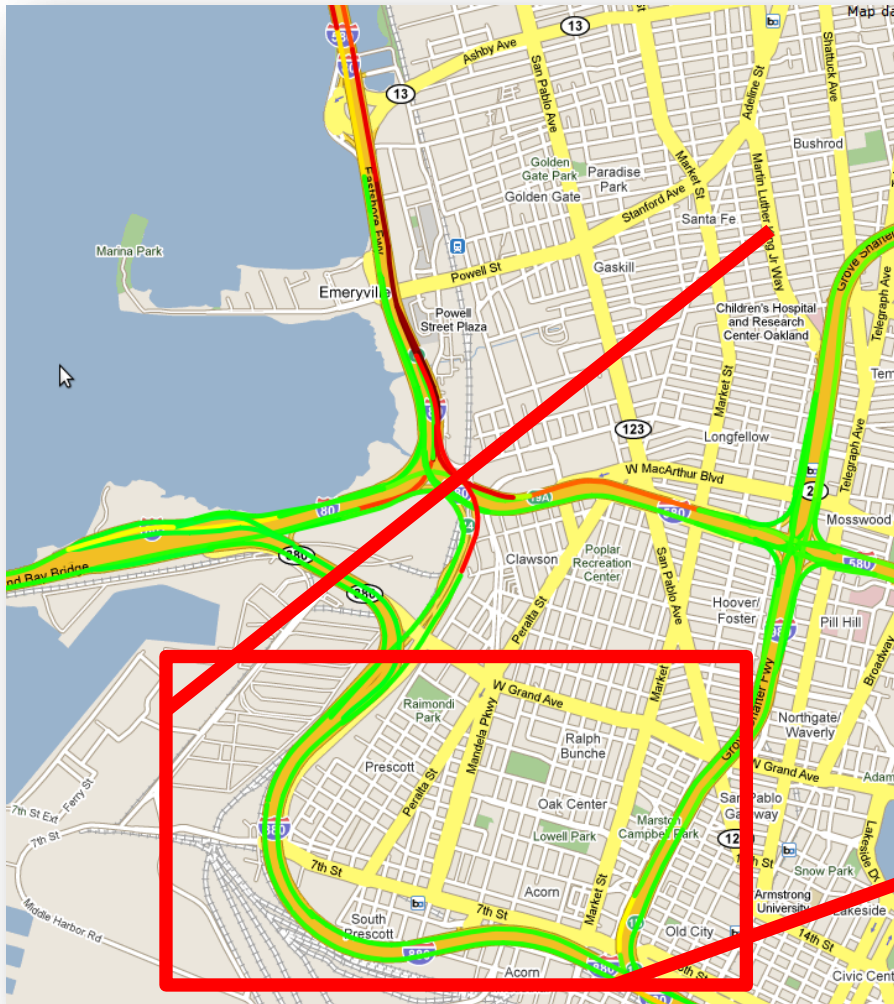
Demain: Estimation en temps réel



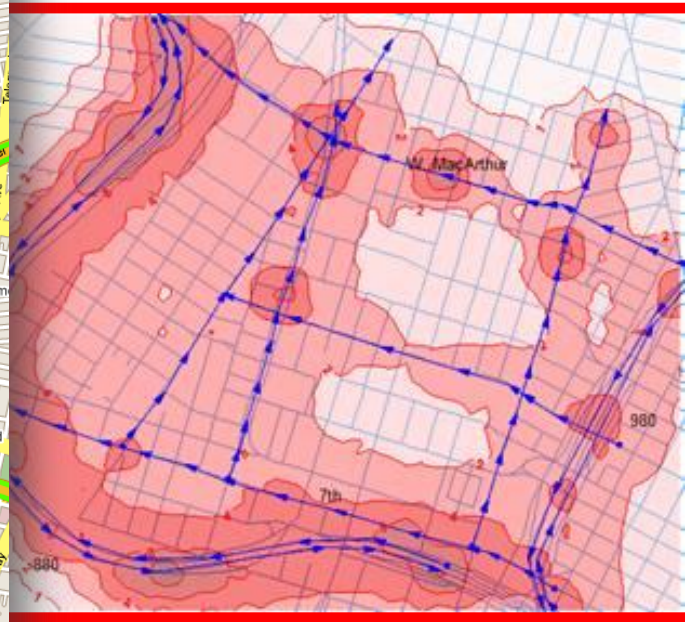
Mobile Millennium: au-delà du trafic

“e-Wellness”

- Estimation des émissions induites du trafic.



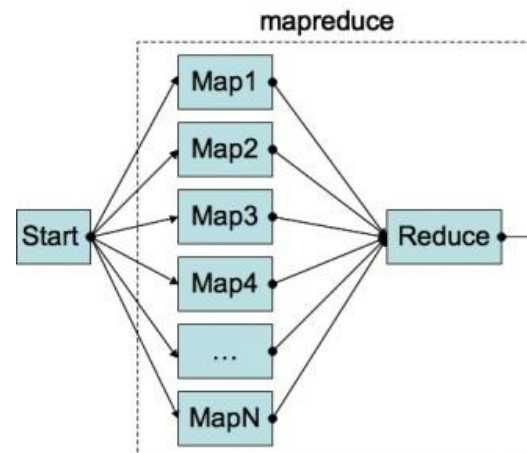
Carte de pollution



Mobile Millennium “on the cloud”

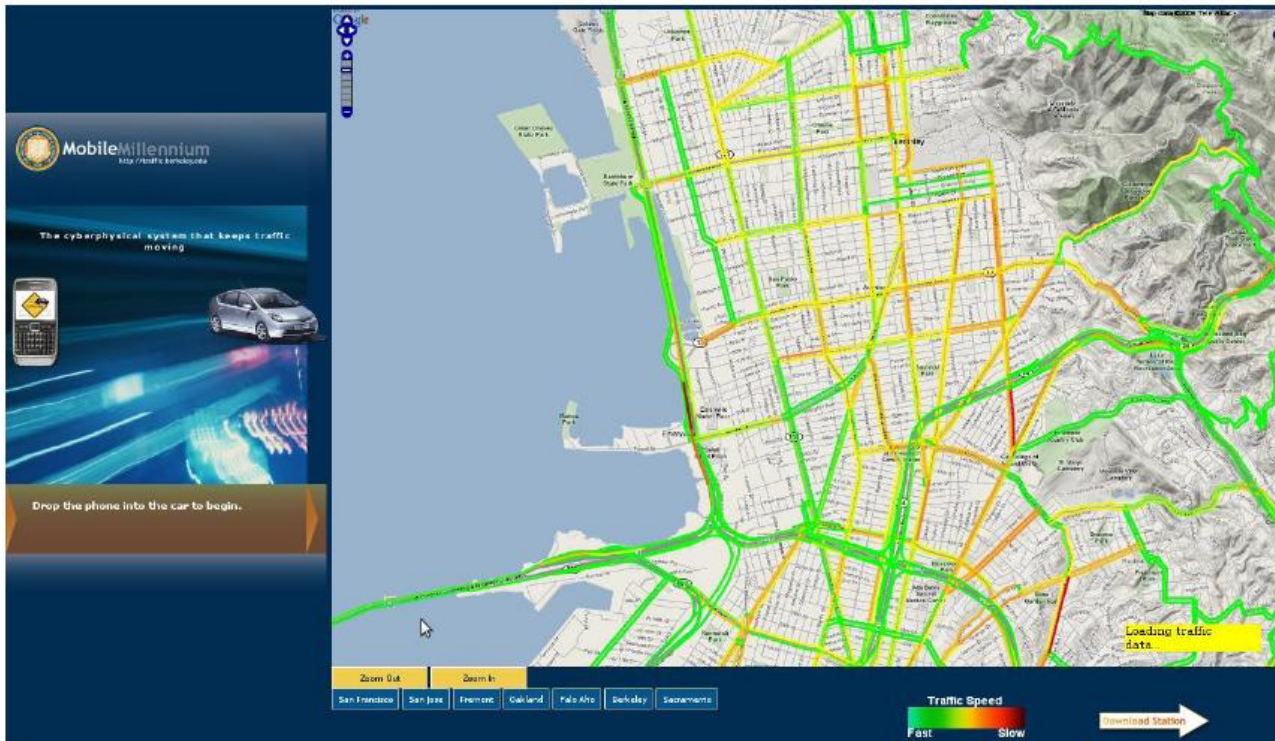
Nouvelle possibilité offerte pour l'estimation à grande échelle

- Algorithmes parallélisables
- Calculs effectués sur un réseau d'ordinateurs
- Différentes possibilités
 - Amazon Map Reduce
 - Mesos, Spark
 - IBM cluster
 - But: 500 noeuds sur le cluster NERSC Magellan



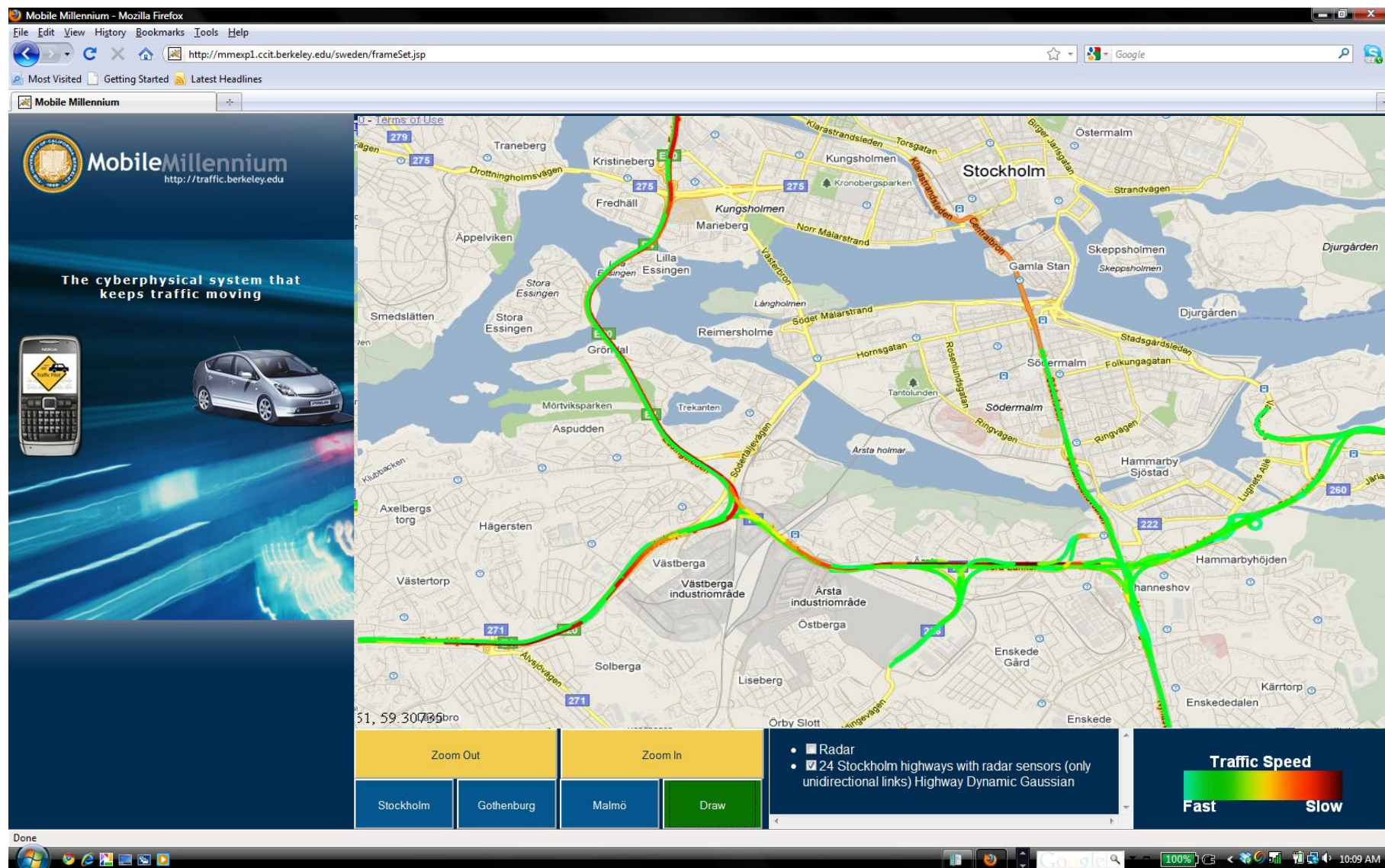
En résumé: *Mobile Millennium* aujourd'hui

- A l'origine, **5000 téléchargements** de l'application Nokia
- Le système collecte environ **60 millions de données** par jour à partir d'une douzaine de sources (téléphones portables, taxis, flotes de véhicules, détecteurs magnétiques, etc.)
- Estimation en **temps réel** (recherche en cours pour la prédiction) sur réseau autoroutier et urbain, choix d'itinéraires, etc.
- Peut intégrer toute source de données (mobile ou statique)





Millennium à l'international: Millennium Stockholm





Remerciements

- Etudiants et post-doc
 - Ryan Herring,
 - Timothée Hunter,
 - Samitha Samaranayake, Anthony Patire, Saurabh Amin, Sébastien Blandin
 - Dan Work et Christian Claudel (maintenant professeurs...)
- Professeurs et chercheurs me conseillant sur mes travaux
 - Alex Bayen, Jean-Patrick Lebacque et Habib Haj-Salem
 - Pieter Abbeel,
 - Laurent El Ghaoui,
- L'équipe du CCIT qui développe le logiciel, l'accès aux données, les outils de visualisation

Modèle statistique du trafic urbain sur un réseau: application à l'estimation à partir de données GPS



Projet Mobile Millennium
<http://traffic.berkeley.edu>

Aude Hofleitner
Thèse Grettia/UC Berkeley
<http://eecs.berkeley.edu/~aude>