



INSTITUT NATIONAL DE RECHERCHE
SUR LES TRANSPORTS ET LEUR SÉCURITÉ



ÉCOLE NATIONALE DES
TRAVAUX PUBLICS DE L'ÉTAT

N.-E.EL Faouzi, J.B Lesort

FUSION DE DONNEES POUR L'ESTIMATION DU TEMPS DE PARCOURS

Réflexions théoriques et méthodologies

Rapport LICIT N° 0101
Novembre 2001

Ce travail a été réalisé dans le cadre de l'action PREDIT-DRAST N° 98MT127 sur la fusion de données et s'inscrit plus spécifiquement dans la démarche de réflexion sur la notion du temps de parcours et son estimation à partir de plusieurs sources de données.

Ce rapport constitue la seconde partie de cette action et porte sur les difficultés théoriques que présente la mesure et l'estimation des temps de parcours, puis sur l'évaluation de l'intérêt de fusionner des mesures d'origines diverses pour améliorer l'estimation des temps de parcours.

L'exposé sur les concepts et les méthodes de fusion de données n'ont pas été développé dans ce rapport dans la mesure où ces aspects ont fait l'objet d'un rapport spécifique (cf. El Faouzi, 2000), joint à ce présent rapport.

Les auteurs tiennent à remercier C. BUISSON pour sa contribution à la première partie de ce rapport ainsi que pour ses relectures des premières versions de ce document.

Une partie des idées contenues dans cette étude est également issue d'un travail réalisé par le en 1999 par le LICIT pour le compte du Conseil Général du département du Val de Marne, et notamment de nombreuses discussions avec M. A. COSTEL.

TABLE DES MATIÈRES

INTRODUCTION.....	7
PARTIE 1 - RÉFLEXIONS THÉORIQUES SUR LE TEMPS DE PARCOURS DÉFINITIONS ET CONSÉQUENCES OPÉRATIONNELLES.....	8
1. NOTION DU TEMPS DE PARCOURS.....	8
2. VITESSE MOYENNE D'UN FLOT DE TRAFIC.....	8
2.1. VARIABLES CARACTÉRISTIQUES DE L'ÉCOULEMENT DU TRAFIC	8
2.1.1. <i>Notion de vitesse moyenne temporelle</i>	9
2.1.2. <i>Notion de vitesse moyenne spatiale</i>	10
2.1.3. <i>Relations entre les différentes formes de vitesse moyenne</i>	10
2.1.4. <i>Concentration et taux d'occupation</i>	13
3. TEMPS DE PARCOURS : APPROCHE THÉORIQUE.....	15
3.1. DÉFINITIONS DE BASE.....	15
3.1.1. <i>Temps de parcours individuels d'itinéraire et de tronçon</i>	15
3.1.2. <i>Temps de parcours moyen d'itinéraire et de tronçons</i>	16
3.2. TEMPS DE PARCOURS ET VITESSE	16
3.2.1. <i>Temps de parcours individuel et vitesse individuelle</i>	16
3.2.2. <i>Temps de parcours moyen et vitesse moyenne</i>	17
3.3. DIFFÉRENTS TEMPS DE PARCOURS.....	18
3.3.1. <i>Temps de parcours expérimenté</i>	18
3.3.2. <i>Temps de parcours instantané</i>	19
3.3.3. <i>Temps de parcours prédictif</i>	21
3.4. SYNTHÈSE	22
4. MESURE ET ESTIMATION DU TEMPS DE PARCOURS.....	22
4.1. PRINCIPES GÉNÉRAUX, NATURE DES TEMPS DE PARCOURS MESURÉS.....	22
4.2. DISTRIBUTIONS DES TEMPS DE PARCOURS	22
4.3. MESURE DIRECTE DU TEMPS DE PARCOURS	24
4.4. ESTIMATION DU TEMPS DE PARCOURS.....	25
4.4.1. <i>Approche statistique</i>	25
4.4.2. <i>Approches physiques et phénoménologiques</i>	26
4.4.3. <i>Approches mixtes</i>	27
4.5. CONCLUSIONS	28
5. DIMENSIONNEMENT D'UN SYSTÈME DE MESURE DU TEMPS DE PARCOURS.....	29
5.1. INTRODUCTION.....	29
5.2. MESURES DIRECTES.....	29
5.2.1. <i>Échantillonnage</i>	30
5.2.2. <i>Taille de l'échantillon</i>	30
5.2.3. <i>Types de flottes</i>	32
5.3. SYSTÈMES D'ESTIMATION	33
5.3.1. <i>Méthodes de stocks</i>	34
5.3.2. <i>Méthodes fondées sur l'estimation ponctuelle de la vitesse</i>	34
5.3.3. <i>Méthodes fondées sur la modélisation de l'écoulement du trafic</i>	34
5.3.4. <i>Méthodes mixtes : le BRP</i>	35
6. CONCLUSION.....	35

ARTIE II - FUSION DE DONNÉES POUR L'ESTIMATION DES TEMPS DE PARCOURS	37
1. INTRODUCTION	37
2. FUSION ET TEMPS DE PARCOURS : ETAT DE L'ART	37
2.1. FUSION STATISTIQUE D'ESTIMATEURS ET D'ALGORITHMES	39
2.2. FUSION CRÉDIBILISTE DES TEMPS DE PARCOURS	41
2.2.1. <i>Recalage et homogénéisation des informations</i>	42
2.2.2. <i>Modélisation des connaissances</i>	43
2.2.3. <i>Évaluation des performances</i>	44
3. CONCLUSIONS	47
PARTIE III - ÉVALUATION DES SYSTÈMES DE MESURE DES TEMPS DE PARCOURS	51
1. INTRODUCTION	51
2. EXIGENCES ATTACHÉES À LA MESURE OU À L'ESTIMATION DES TEMPS DE PARCOURS .52	
2.2. COUVERTURE SPATIALE - CLASSIFICATION DES ARCS	52
2.3. FINESSE SPATIALE	53
2.4. PRÉCISION	53
2.5. FIDÉLITÉ.....	55
2.6. COUVERTURE TEMPORELLE.....	55
2.7. FRAÎCHEUR	56
2.8. DISPONIBILITÉ.....	56
2.9. ÉVOLUTIVITÉ	57
2.10. COHÉRENCE	57
2.11. CONDITIONS DE MISE EN ŒUVRE	58
2.12. COÛTS.....	58
3. INDICATEURS, CRITÈRES ET MOYENS D'ÉVALUATION	58
3.1. COUVERTURE SPATIALE	59
3.2. FINESSE SPATIALE	59
3.3. COUVERTURE TEMPORELLE.....	59
3.4. FRAÎCHEUR.....	60
3.5. DISPONIBILITÉ.....	60
3.6. COHÉRENCE	60
3.7. MISE EN ŒUVRE	61
3.8. ÉVOLUTIVITÉ.....	61
3.9. COÛT.....	61
3.10. PRÉCISION	62
3.11. FIDÉLITÉ.....	62
4. METHODOLOGIES D'ÉVALUATION	65
4.1. PRINCIPALES TECHNIQUES DE VALIDATION A POSTERIORI	65
4.1.1. <i>Validation par resubstitution</i>	66
4.1.2. <i>Validation par échantillon test</i>	66
4.1.3. <i>Validation croisée</i>	67
4.1.4. <i>Autres méthodes de rééchantillonnage</i>	67
5. CONCLUSION	68
PARTIE IV - CONCLUSIONS GÉNÉRALES	69
RÉFÉRENCES	71

TABLE DES ILLUSTRATIONS

Figure I-1 : Distribution théorique des temps de parcours : cas log-normale.....	23
Figure I-2 : Distribution théorique des temps de parcours : cas normale.....	24
Figure II-1 : Recueil multiforme des données de temps de parcours.....	38
Figure II-2 : Calcul des masses de croyances et principe de la fusion.....	44
Tableau II-1 : Performances de la méthodologie de fusion avec concordances strictes.....	45
Tableau II-2 : Performances de la méthodologie de fusion avec concordances larges.....	46
Tableau II-3 : Tableau récapitulatif des exigences, des indicateurs et de leur mode de mesure.....	63

INTRODUCTION

Les enjeux d'une meilleure connaissance des temps de parcours, que ce soit en milieu urbain ou interurbain, sont multiples. Pour les maîtres d'ouvrage et les exploitants, le temps de parcours constitue tout d'abord un indicateur synthétique, immédiatement compréhensible et facile à manipuler de la qualité d'écoulement du trafic sur un réseau. Cet indicateur trouve son utilité hors ligne, dans le suivi des performances d'un réseau ou du système qui le contrôle, mais aussi en ligne comme indicateur d'aide à la prise de décisions tactiques. Communiqué en temps réel aux usagers, le temps de parcours constitue une aide au choix d'itinéraire, mais aussi un élément de confort des déplacements. Influant sur le choix d'itinéraire, il devient également un outil de régulation des flux au service des exploitants.

Tout ceci explique le fort intérêt manifesté par les exploitants pour une meilleure connaissance des temps de parcours, intérêt confirmé par la première partie de l'étude dont ce rapport constitue les deuxièmes et troisièmes volets. Malheureusement, cette demande n'est aujourd'hui satisfaite que très partiellement : les systèmes et méthodes de mesure ou d'estimation des temps de parcours disponibles aujourd'hui se heurtent à des problèmes de différentes natures, liés à la fois à la nature même de la grandeur à mesurer mais aussi aux limites technologiques des moyens de mesure.

C'est pourquoi cette étude a cherché tout d'abord à prendre la mesure des difficultés théoriques que présente la mesure et l'estimation des temps de parcours, puis à faire un point sur les méthodes utilisées. Elle a ensuite, et c'était son objet principal, évalué l'intérêt de fusionner des mesures d'origines diverses pour améliorer l'estimation des temps de parcours. La dernière partie est enfin consacrée à une question souvent négligée, à savoir celle de l'évaluation des méthodes d'estimation de temps de parcours : outre les problèmes généraux liés à toute évaluation (problème de la référence, indicateurs de performance...), ce type d'évaluation présente en effet un certain nombre de difficultés spécifiques. Sans avoir l'ambition de constituer un guide méthodologique, cette étude pointe les difficultés essentielles et propose quelques pistes pour les résoudre.

PARTIE I

RÉFLEXIONS THÉORIQUES SUR LE TEMPS DE PARCOURS *Définitions et conséquences opérationnelles*

1. NOTION DU TEMPS DE PARCOURS

La notion de temps de parcours paraît intuitivement évidente : le temps de parcours d'un usager sur une route ou un tronçon de route se définit sans ambiguïté comme le temps écoulé entre son entrée sur la route ou le tronçon et sa sortie. Pourtant, lorsqu'on cesse de considérer le temps de parcours individuel d'un usager pour s'intéresser à une notion plus globale de temps de parcours caractéristique d'une route ou d'une fraction de route, des difficultés de définition apparaissent immédiatement. On s'aperçoit alors qu'il est possible de définir d'un grand nombre de façons le temps de parcours, chaque définition ayant des caractéristiques propres et impliquant des modes de mesure ou d'estimation différents. Ceci vient en fait des différentes façons possibles d'agréger des temps de parcours individuels pour calculer des moyennes, mais aussi des difficultés comparables qu'on rencontre dans la définition de la vitesse d'un flot de trafic. C'est la raison pour laquelle ce chapitre examinera tout d'abord, ce qui est plus simple, la caractérisation de la vitesse d'un flot de trafic, avant d'examiner plus avant les définitions possibles du temps de parcours et les liens entre celles-ci et la vitesse.

2. VITESSE MOYENNE D'UN FLOT DE TRAFIC

2.1. VARIABLES CARACTERISTIQUES DE L'ÉCOULEMENT DU TRAFIC

L'observation de l'écoulement du trafic peut se concevoir de deux points de vue :

- ❑ soit on observe en un point le trafic qui s'écoule pendant une période de temps donnée (cas de l'observateur placé au bord de la route) ;
- ❑ soit on observe, à un instant donné, la répartition des véhicules sur une portion de route (cas d'une photographie aérienne par exemple).

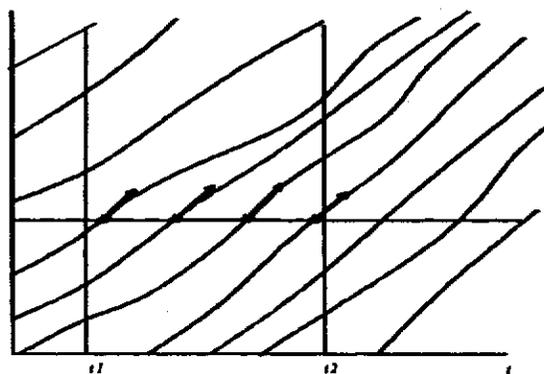
A ces deux points de vue d'observation correspondent deux types de variables différentes pour caractériser l'écoulement du trafic : débit et vitesse moyenne temporelle pour les variables temporelles, concentration et vitesse moyenne spatiale pour les variables spatiales. Si la définition du débit (nombre moyen de véhicules passés en un point par unité de temps) et de la concentration (nombre de véhicules présents sur une section de chaussée à un instant donné) sont relativement sans mystère, ces deux définitions différentes de la vitesse moyenne d'un flot demandent à être examinées d'un peu plus près.

2.1.1. Notion de vitesse moyenne temporelle

On suppose qu'un observateur est placé au bord d'une route et regarde passer les véhicules. Entre les instants t_1 et t_2 il observe ainsi $\eta(x; t_1 \rightarrow t_2)$ véhicules qui ont des vitesses v_1, v_2, \dots, v_n au moment de leur passage au point x :

$$\bar{V}_T(x; t_1 \rightarrow t_2) = \frac{1}{\eta(x; t_1 \rightarrow t_2)} \sum_{i=1}^n v_i, \quad (1)$$

Comme l'illustre la figure suivante, qui représente les trajectoires des véhicules dans un diagramme espace x temps.



L'indice T indique qu'il s'agit d'une moyenne calculée dans le temps (c'est la moyenne des vitesses des véhicules passés en x entre t_1 et t_2). On définira alors la vitesse moyenne temporelle instantanée $V_T(x; t)$ par :

$$V_T(x; t) = \bar{V}_T(x; t - \Delta t/2 \rightarrow t + \Delta t/2) \quad (2)$$

Cette vitesse moyenne à l'instant t ne correspond pas nécessairement à la vitesse d'un véhicule éventuel passant en x à l'instant t .

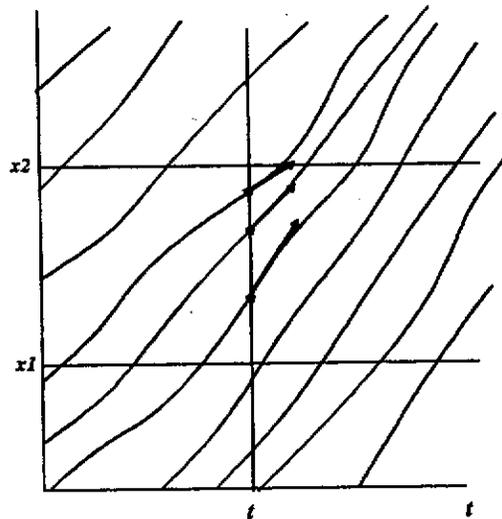
Si l'on cherche à définir une vitesse moyenne instantanée à l'instant t , on sera amené, compte tenu du caractère discret du trafic, à définir une période de référence Δt sur laquelle sera calculé la moyenne.

2.1.2. Notion de vitesse moyenne spatiale

Cette notion est duale de celle de vitesse moyenne temporelle, mais elle en est tout à fait distincte (le chapitre suivant analysera plus précisément les relations existant entre ces différentes notions de vitesse moyenne).

Si l'on appelle v_1, v_2, \dots, v_n la vitesse à l'instant t des n véhicules présents entre les points x_1 et x_2 , la vitesse moyenne spatiale $\bar{v}_E(x_1 \rightarrow x_2; t)$ se définit, comme l'illustre la figure suivante, par :

$$\bar{v}_E(x_1 \rightarrow x_2; t) = \frac{1}{n(x_1 \rightarrow x_2; t)} \sum_{j=1}^n v_j \quad (3)$$



Avec les mêmes hypothèses que précédemment, on peut définir une vitesse moyenne spatiale au point x :

$$v_E = \bar{v}_E(x - \Delta x/2 \rightarrow x + \Delta x/2; t) \quad (4)$$

On voit ainsi qu'il existe deux définitions possibles (spatiale et temporelle), de la vitesse moyenne au point x à l'instant t . Il est donc nécessaire d'étudier plus précisément les rapports entre ces définitions.

2.1.3. Relations entre les différentes formes de vitesse moyenne

2.1.3.1. Cas isovélocique

Ce cas simple est celui d'un écoulement uniforme où tous les véhicules ont une vitesse identique v . Dans ces conditions, la vitesse moyenne U du flot est de façon triviale égale à v . Soit dans ces conditions une portion de chaussée Δx . Elle contient à tout instant un nombre de véhicules $n = K\Delta x$ où K est la concentration.

La vitesse du flot, qui est également celle de tous les véhicules, étant égale à U , il faut un temps $\Delta t = \Delta x/U$ pour que ces n véhicules quittent la zone considérée. Le débit sortant de cette zone est donc égal à :

$$Q = \frac{n}{\Delta t} = K \Delta x \cdot \frac{U}{\Delta x} = KU \quad (5)$$

2.3.1.2. Cas général

Dans le cas général, on utilise la propriété démontrée précédemment comme définition. On définira ainsi la vitesse du flot au point x et à l'instant t , $U(x;t)$:

$$U(x;t) = \frac{Q(x;t)}{K(x;t)} \quad (6)$$

Ceci conduit ainsi à une troisième définition possible de la vitesse du flot.

2.3.1.3. Propriété et Formule de Wardrop

On montre que *la vitesse du flot définie précédemment est égale à la vitesse moyenne spatiale*. Cette propriété, qui nous permet de réduire le nombre de définitions de la vitesse moyenne, est démontrée par ailleurs (cf. Lesort, 1998). On parlera désormais de vitesse du flot pour désigner la vitesse moyenne spatiale.

On montre de même que la vitesse moyenne temporelle n'est pas dans le cas général, égale à la vitesse du flot. Il existe cependant plusieurs relations entre ces deux formes de vitesse moyenne. Une de ces relations, connue sous le nom de formule de Wardrop peut être énoncée de la manière suivante :

$$V_T = U + \frac{\sigma_E^2}{U} \quad (7)$$

où σ_E^2 désigne la variance spatiale de la vitesse.

Ceci montre bien que la vitesse moyenne temporelle est non seulement différente, mais toujours supérieure à la vitesse du flot¹, si ce n'est dans le cas isovélocique où $\sigma_E^2 = 0$ et où les deux valeurs sont confondues. Physiquement, ceci s'explique par le fait que, à concentration égale, le débit est proportionnel à la vitesse. Les véhicules rapides sont donc relativement plus nombreux dans le temps que dans l'espace.

2.3.1.4. Propriété

La vitesse du flot est égale à la moyenne harmonique des vitesses mesurées en un point

¹ Ceci est bien évidemment conforme au résultat général portant sur l'inégalité entre les moyennes arithmétique (\bar{X}) et harmonique (H), inégalité qui s'exprime ainsi : $H \leq \bar{X}$.

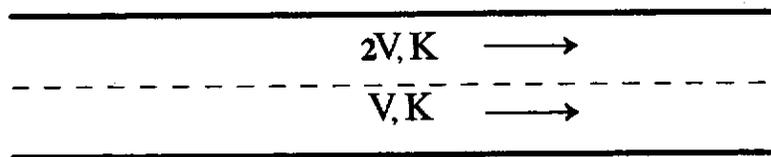
Si η véhicules passent au point x en Δt ceci s'écrit :

$$\frac{1}{U} = \frac{1}{\eta} \sum_{i=1}^{\eta} \frac{1}{v_i} \quad (8)$$

Cette propriété est importante, car elle donne un moyen opératoire d'estimer la vitesse du flot à partir d'un instrument de mesure ponctuel (radar ou boucle). Dans la pratique, la moyenne arithmétique est bien souvent utilisée par simple ignorance de cette propriété, ce qui peut conduire à des estimations de densité erronées. L'exemple suivant permettra de saisir la différence entre ces différentes notions.

2.3.1.5. Exemple : cas de l'autoroute à deux voies

Soit une autoroute à deux voies en situation stationnaire, sur chacune de laquelle roulent deux flux isovélociques, l'un de vitesse V et de concentration K (sur la voie de droite), l'autre de vitesse $2V$ et de concentration K (sur la voie de gauche).



Le débit sur la voie de droite est $Q_1 = KV$, alors que sur la voie de gauche il vaut $Q_2 = 2KV$.

On cherche à calculer l'ensemble des variables qui caractérisent le flux global. La concentration totale est $k = 2K$. Le débit total passant en un point x à l'instant t peut s'écrire : $Q = Q_1 + Q_2 = KV + 2KV = 3KV$.

La vitesse du flot est donc :

$$U = \frac{Q}{K} = \frac{3KV}{2K} = \frac{3}{2}V \quad (9)$$

Calculons maintenant la moyenne arithmétique des vitesses mesurées sur une portion de voie de longueur Δx à un instant t quelconque. Sur la voie de droite, le nombre de véhicules sur cette portion vaut $N_1 = K\Delta x$, leur vitesse est de V ; alors que sur la voie de gauche, $N_2 = K\Delta x = N_1$ et $V_2 = 2V$. La vitesse moyenne spatiale se calcule donc par :

$$\begin{aligned} \bar{V}_E(x \rightarrow x + \Delta x; t) &= \frac{1}{N_1 + N_2} (N_1 V_1 + N_2 V_2) \\ &= \frac{1}{2K\Delta x} K\Delta x (V + 2V) \\ &= \frac{3}{2}V \end{aligned} \quad (10)$$

La moyenne arithmétique des vitesses mesurées pendant une période Δt en un point x peut être calculée si on sait que pendant ce laps de temps passent $M_1 = Q_1 \Delta t = KV \Delta t$ véhicules sur la voie de droite et $M_2 = Q_2 \Delta t = 2KV \Delta t$ véhicules sur la voie de gauche :

$$\begin{aligned} \bar{V}_T(x; t \rightarrow t + \Delta t) &= \frac{1}{M_1 + M_2} (M_1 V_1 + M_2 V_2) \\ &= \frac{1}{(KV + 2KV) \Delta t} (KV \cdot V + 2KV \cdot 2V) \Delta t \\ &= V \frac{KV + 4KV}{KV + 2KV} \\ &= \frac{5}{3} V. \end{aligned} \quad (11)$$

Sur cet exemple simple on voit bien la différence entre la vitesse moyenne spatiale (qui est bien égale à la vitesse de flot) et vitesse moyenne temporelle qui lui est supérieure. Ceci est dû au fait qu'en un point passent dans le même laps de temps plus de véhicules rapides que de véhicules lents, alors que si on regarde une portion de route à un instant, si la concentration est égale, il y aura autant de véhicules rapides que de véhicules lents.

L'inverse de la moyenne harmonique des vitesses mesurées en un point pendant un laps de temps vaut :

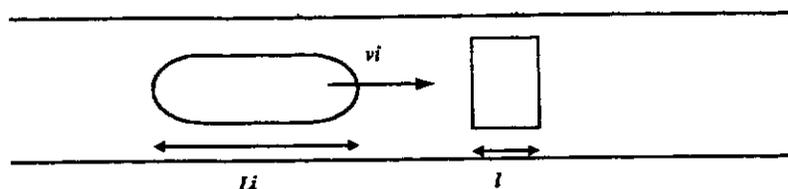
$$\begin{aligned} \frac{1}{\omega} &= \frac{1}{M_1 + M_2} \left(\frac{M_1}{V} + \frac{M_2}{2V} \right) \\ &= \frac{1}{(KV + 2KV)} (K + K) \\ &= \frac{2}{3V} \end{aligned} \quad (12)$$

D'où, comme démontré ci-dessus :

$$\omega = \frac{3}{2} V = \bar{V}_E = U. \quad (13)$$

2.1.4. Concentration et taux d'occupation

La définition de la vitesse d'un flot de trafic comme rapport du débit sur la concentration fournit une méthode de mesure possible de cette vitesse. Malheureusement, la technique de mesure du trafic par boucle électromagnétique, qui est de loin la plus employée aujourd'hui, donne accès à la mesure du débit, mais ni à celle de la concentration ni en général à celle de la vitesse (possible par utilisation de doubles boucles). Elle fournit en revanche un indicateur intéressant : le taux d'occupation, qui dans certaines conditions peut être utilisé comme estimateur de la concentration.



Soit une boucle magnétique de longueur l , sur laquelle on observe les passages de η véhicules pendant un temps Δt . Chaque véhicule i possède une longueur L_i et une vitesse V_i . La boucle est ainsi excitée par le passage du véhicule i pendant un temps T_i :

$$T_i = \frac{L_i + l}{V_i} \quad (14)$$

On définit le taux d'occupation τ pendant la période Δt par :

$$\tau = \frac{1}{\Delta t} \sum_{i=1}^{\eta} T_i \quad (15)$$

C'est donc la proportion du temps pendant laquelle la boucle a été occupée. On peut ainsi écrire :

$$\tau = \frac{1}{\Delta t} \sum_{i=1}^{\eta} \frac{L_i + l}{V_i} \quad (16)$$

En supposant que la longueur des véhicules soit une constante L , il vient :

$$\begin{aligned} \tau &= \frac{L + l}{\Delta t} \sum_{i=1}^{\eta} \frac{1}{V_i} \\ &= (L + l)Q \times \frac{1}{Q\Delta t} \sum_{i=1}^{\eta} \frac{1}{V_i} \\ &= (L + l) \times \frac{Q}{U}, \quad \text{en application du théorème précédent, d'où :} \\ &= (L + l)K. \end{aligned} \quad (17)$$

On peut ainsi avoir, à partir d'une mesure ponctuelle, une estimation de la concentration. L'hypothèse de longueur constante des véhicules est à peu près valide dans un environnement urbain où les poids lourds sont constamment peu nombreux. Cette hypothèse est cependant relativement forte et doit inciter à utiliser le taux d'occupation avec prudence.

3. TEMPS DE PARCOURS : APPROCHE THÉORIQUE

3.1. DEFINITIONS DE BASE

La notion de temps de parcours n'est clairement définissable de façon univoque que d'un point de vue individuel : le temps de parcours d'un véhicule entre un point A et un point B est trivialement le temps qu'il met à aller de l'un à l'autre. Dans des conditions parfaitement stationnaires, uniformes et où tous les véhicules vont à la même vitesse, on peut de façon toute aussi évidente définir le temps de parcours du flux comme celui de chacun des véhicules qui le composent. C'est lorsque ces conditions ne sont pas réalisées que la notion de temps de parcours moyen ou de temps de parcours associé à un itinéraire cesse à la fois d'être intuitive et de répondre à une définition unique.

Trois éléments peuvent en effet induire des variations dans le temps de parcours des véhicules :

- Le premier est la dispersion des comportements des conducteurs et des caractéristiques des véhicules ;
- Le deuxième est la variation des conditions de trafic au cours du temps ;
- Le troisième est la variation de ces conditions dans l'espace.

Ainsi, dans l'exemple ci-dessus d'une autoroute à deux voies, il y a dispersion des comportements, les usagers ayant deux vitesses distinctes (V et $2V$) ; sans qu'il y ait pour autant variation dans le temps (la situation est stationnaire), ni dans l'espace (la situation est uniforme).

Ces trois types de variations ont des influences différentes suivant le type de temps de parcours que l'on calcule. La variation du temps de parcours au cours de la journée est liée à l'évolution des conditions de trafic (qu'elle reflète) mais peu aussi être forte alors que les conditions sont sensiblement identiques (que l'on songe par exemple au temps de parcours sur un tronçon fluide terminé par un feu).

Sur cet exemple du tronçon terminé par un feu, on voit aussi que la vitesse n'est pas uniforme sur le tronçon.

3.1.1. Temps de parcours individuels d'itinéraire et de tronçon.

Le temps de parcours individuel d'itinéraire est, comme son nom l'indique, le temps mis par un individu pour parcourir un itinéraire. Si on partage l'itinéraire i en m tronçons consécutifs disjoints, le temps de parcours T_i^u de l'usager u peut être décomposé comme la somme des temps de parcours T_a^u sur ces différents tronçons :

$$T_i^u = \sum_{a=1}^m T_a^u \quad (18)$$

Ceci est vrai, pour un usager individuel, que les conditions soient ou non stationnaires dans le temps et uniformes dans l'espace.

3.1.2. Temps de parcours moyen d'itinéraire et de tronçons

Les temps de parcours moyens d'itinéraires et de tronçons peuvent évidemment être calculés à partir des temps de parcours individuels. Si pendant une période de temps Δt il est sorti N_i véhicules de l'itinéraire i , Le temps de parcours moyen \bar{T}_i au cours de cette période est égal à :

$$\bar{T}_i = \frac{1}{N_i} \sum_{u=1}^{N_i} T_i^u \quad (19)$$

De même, si N_a véhicules sont sortis du tronçon a de l'itinéraire i pendant le temps Δt , le temps de parcours moyen \bar{T}_a sur le tronçon a s'écrit :

$$\bar{T}_a = \frac{1}{N_a} \sum_{u=1}^{N_a} T_a^u \quad (20)$$

En revanche, et c'est de là que viennent toutes les difficultés de manipulation de la notion de temps de parcours, il n'y a dans le cas général pas d'égalité entre le temps de parcours moyen de l'itinéraire et la somme des temps de parcours moyens des tronçons qui le composent :

$$\bar{T}_i \neq \sum_{a=1}^m \bar{T}_a \quad (21)$$

On voit ainsi que, si les temps de parcours individuels sont additifs, les temps de parcours moyens ne le sont pas dans le cas général. Ceci vient du fait que ces temps n'ont pas été calculés sur les mêmes véhicules. L'égalité n'est vérifiée que si les conditions de trafic sont stationnaires et s'il n'y a pas de dispersion des vitesses individuelles.

3.2. TEMPS DE PARCOURS ET VITESSE

3.2.1. Temps de parcours individuel et vitesse individuelle.

Le temps de parcours individuel du véhicule u sur le tronçon a est relié à la vitesse instantanée du véhicule par la relation suivante :

$$T_a^u = \int_{x=0}^{L_a} \frac{dx}{v_u(x)} \quad (22)$$

où L_a est la longueur du tronçon a et $v_u(x)$ est la vitesse de l'utilisateur u au point x du tronçon. On voit déjà que cette relation n'a de sens que si le véhicule u ne s'arrête pas au cours de son trajet : dans le cas contraire, il faudrait ajouter les temps d'arrêt. On pourra de même écrire sur un itinéraire i composé de m tronçons :

$$T_i^u = \int_{x=0}^{L_i} \frac{dx}{v_u(x)} = \sum_{a=1}^m T_a^u = \sum_{a=1}^m \int_{x=0}^{L_a} \frac{dx}{v_a^u(x)} \quad (23)$$

où $v_a^u(x)$ représente la vitesse du véhicule u à l'abscisse x du tronçon a . On retrouve l'additivité du temps de parcours individuel, qui est liée au caractère linéaire de l'intégrale.

3.2.2. Temps de parcours moyen et vitesse moyenne

En utilisant les résultats ci-dessus, on peut écrire le temps de parcours moyen du tronçon a pendant la période Δt où N_a véhicules sont sortis du tronçon :

$$\bar{T}_a = \frac{1}{N_a} \sum_{u=1}^{N_a} T_a^u = \frac{1}{N_a} \sum_{u=1}^{N_a} \int_{x=0}^{L_a} \frac{dx}{v_a^u(x)} \quad (24)$$

On peut par ailleurs définir une vitesse moyenne pratiquée sur le tronçon par :

$$\bar{V}_a = \frac{1}{\bar{T}_a} \quad (25)$$

On voit que cette moyenne correspond à une agrégation à la fois dans l'espace (le tronçon) et dans le temps (l'ensemble des véhicules sortis du tronçon pendant le temps Δt). On peut alors rechercher les rapports existants entre cette vitesse moyenne et la vitesse du flot U .

Le temps de parcours moyen peut s'écrire :

$$\bar{T}_a = \int_{x=0}^{L_a} \left(\frac{1}{N_a} \sum_{u=1}^{N_a} \frac{1}{v_a^u(x)} \right) dx \quad (26)$$

En conditions stationnaires, et dans ces conditions seulement, où la vitesse au point x ne dépend pas du temps, l'expression entre parenthèse, qui est l'inverse de la moyenne harmonique de vitesses mesurées en un point, est égale à l'inverse de la vitesse $U(x)$ du flot au point x . On peut alors écrire :

$$\bar{T}_a = \int_{x=0}^{L_a} \frac{dx}{U(x)} \quad (27)$$

En conditions non stationnaires les véhicules qui passent au point x pendant le temps Δt n'ont aucune raison d'avoir la même distribution de vitesse que les N_a véhicules considérés pour le calcul de \bar{T}_a . Cette expression n'est donc pas valide.

Si les conditions sont stationnaires, la vitesse moyenne \bar{V}_a peut s'écrire :

$$\bar{V}_a = \frac{1}{\bar{T}_a} = \frac{1}{\int_{x=0}^{L_a} \frac{dx}{U(x)}} \quad (28)$$

On voit que cette vitesse \bar{V}_a est alors une moyenne spatiale *harmonique* de la vitesse du flot. C'est seulement en conditions uniformes, c'est à dire si U ne dépend pas de x , que la vitesse moyenne \bar{V}_a est égale à la vitesse du flot.

On voit ainsi que la notion de temps de parcours, facile à manipuler lorsqu'elle fait référence à un individu, devient extrêmement complexe lorsqu'on recherche le temps de parcours d'un flot. C'est la raison pour laquelle on est amené à envisager d'autres définitions du temps de parcours moyen qui soient plus faciles à manipuler et surtout plus faciles à relier à des notions de vitesse.

3.3. DIFFERENTS TEMPS DE PARCOURS

On peut distinguer trois types de temps de parcours, qui chacun reproduisent un aspect du trafic différent et correspondent à différents types d'estimation.

3.3.1. Temps de parcours expérimenté

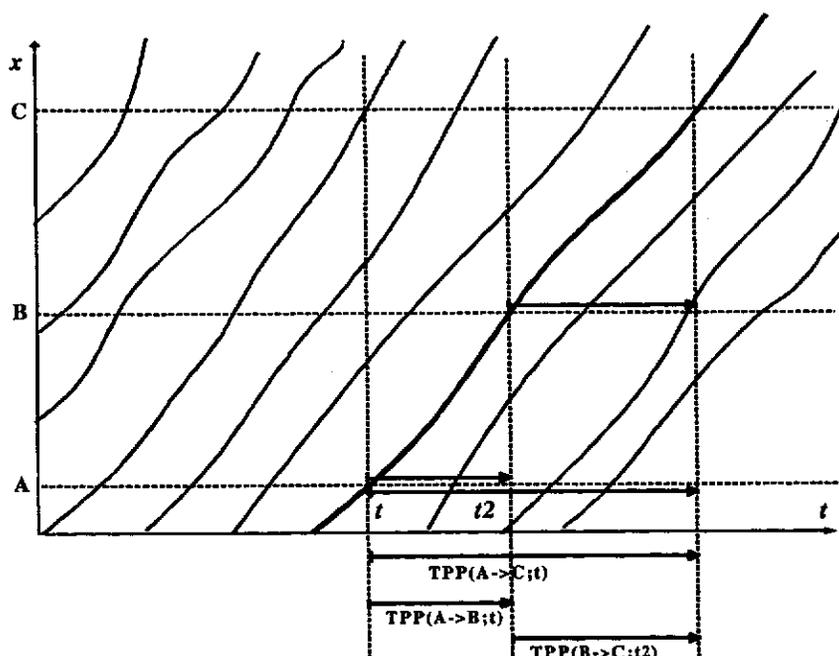
C'est le temps mis par un véhicule pour parcourir un tronçon ou un itinéraire. Il est disponible une fois que le véhicule est sorti et peut donc difficilement être utilisé en temps réel sur des itinéraires longs.

Le temps de parcours expérimenté $TPE(t)$ à l'instant t est donc celui du dernier véhicule sorti du tronçon avant l'instant t , c'est à dire entré sur le tronçon avant l'instant $t - TPE(t)$.

Il est important de noter que les temps de parcours expérimentés ne sont pas additifs c'est à dire que le temps de parcours de A à C mesuré à l'instant t n'est pas nécessairement égal à la somme du temps de parcours de A à B à l'instant t et de B à C à l'instant t . En toute rigueur, il s'écrit :

$$TPE_{A \rightarrow C}(t) = TPE_{B \rightarrow C}(t) + TPE_{A \rightarrow B}(t - TPE_{B \rightarrow C}(t)) \quad (29)$$

Ceci peut être illustré par la figure suivante :



Ce type de temps de parcours est typiquement obtenu à partir des véhicules traceurs. On peut également utiliser un temps de parcours expérimenté moyen, qui est celui évoqué au paragraphe précédent.

3.3.2. Temps de parcours instantané

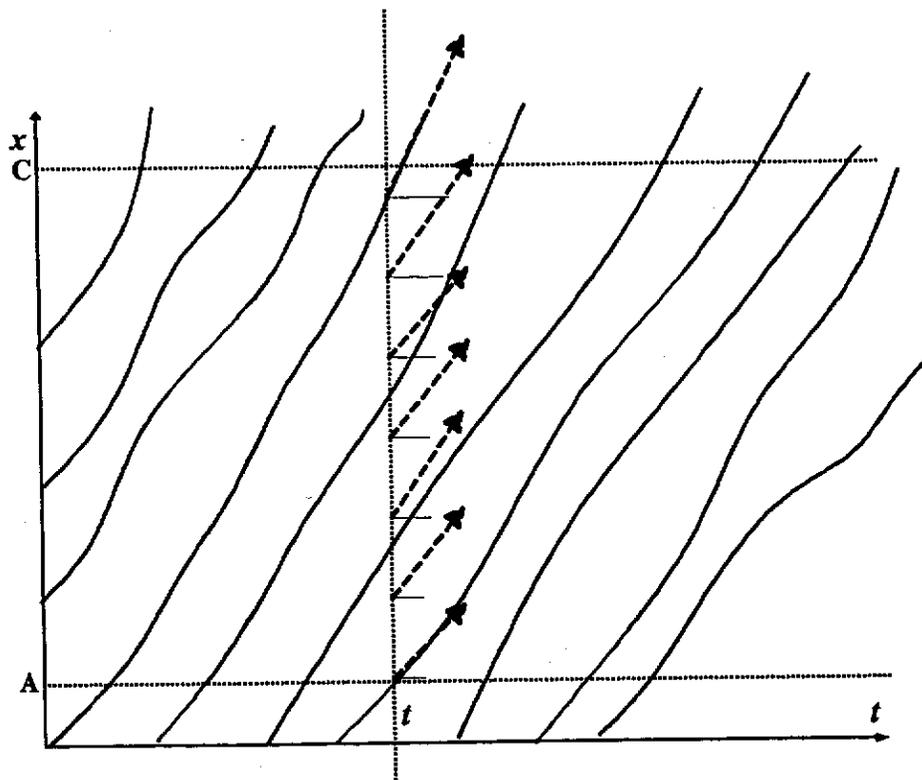
Cette définition dérive directement des considérations faites plus haut sur les relations entre temps de parcours et vitesse moyenne. Elle consiste en effet à utiliser la vitesse du flot, et à calculer un temps de parcours moyen comme si les conditions étaient stationnaires :

$$TPI_{A \rightarrow C}(t) = \int_{x=A}^C \frac{dx}{U(x;t)} \quad (30)$$

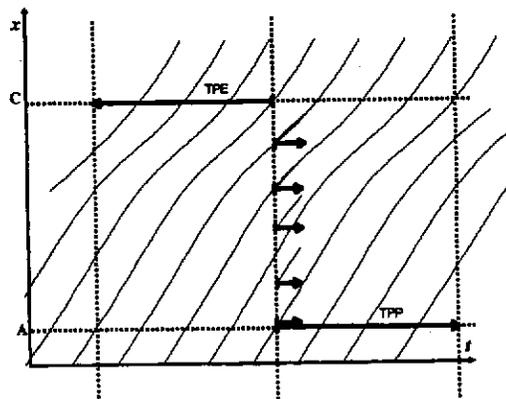
Par nature, le temps de parcours instantané est additif, puisque :

$$\forall B \in [A, C], \text{ on a : } TPI_{A \rightarrow C}(t) = \int_{x=A}^B \frac{dx}{U(x;t)} + \int_{x=B}^C \frac{dx}{U(x;t)} = TPI_{A \rightarrow B}(t) + TPI_{B \rightarrow C}(t). \quad (31)$$

C'est ce que l'on peut voir sur la figure suivante (ou les traits horizontaux figurent les temps de parcours infinitésimaux pour chaque portion de longueur élémentaire).



En situation stationnaire, lorsqu'il n'y a pas de dispersion des comportements individuels et qu'il y ait ou non uniformité le long du tronçon, les deux temps de parcours présentés ci-dessus sont égaux. La figure suivante en fournit une illustration.



Par contre, en dehors des cas de situation stationnaire, ce temps de parcours ne correspond pas à un temps de parcours individuel, puisque les vitesses utilisées étant celles prévalant en tous les points du tronçon à un instant, aucun véhicule n'expérimentera la vitesse $v(x;t)$ lors de son passage en un point x différent de celui qu'il occupe à l'instant t .

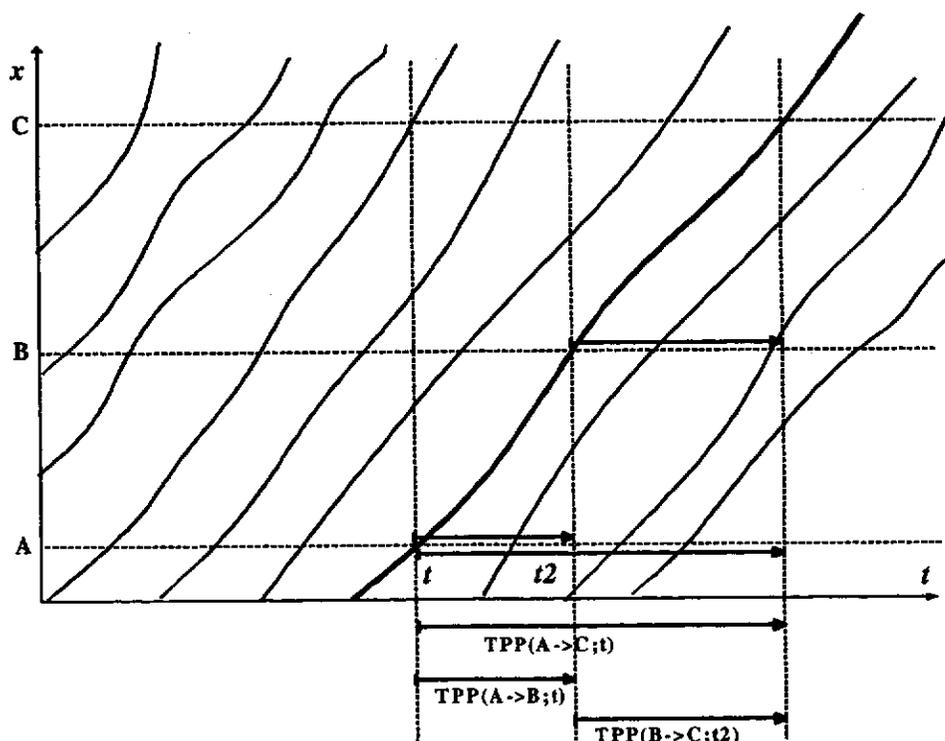
Cette notion de temps de parcours instantané peut être extrapolée à des portions de voies de longueur non infinitésimale, par exemple des tronçons. On peut donc avoir une mesure du temps de parcours instantané sur un itinéraire à partir de la somme des temps de parcours individuels à tout instant. C'est cette mesure qui correspond *grosso modo* à la sommation de temps de parcours obtenus à partir de vitesses moyennes mesurées par des capteurs à boucle. On voit bien que cette définition du temps de parcours, qui utilise les caractéristiques locales du trafic, sera d'une utilisation intéressante lorsqu'on utilisera des capteurs au sol pour la mesure.

3.3.3. Temps de parcours prédictif

Le temps de parcours prédictif est le fruit d'un calcul qui, à partir des conditions courantes de circulation, évalue le temps de parcours qui sera mis par le véhicule en train d'entrer sur le tronçon pour traverser celui-ci, si les conditions de trafic restent les mêmes pendant la durée du trajet. Une formule duale de la précédente montre que ce temps de parcours n'est pas non plus additif, comme le montre la figure suivante :

$$TPP_{A \rightarrow C}(t) = TPE_{A \rightarrow B}(t) + TPP_{B \rightarrow C}(t + TPP_{A \rightarrow B}(t)). \quad (32)$$

C'est ce type d'estimation que l'on obtient en faisant une simulation (à partir des conditions de trafic actuelles) pour calculer le temps de parcours d'un véhicule entrant sur le tronçon à l'instant t .



Le temps de parcours prédictif ne constitue à vrai dire pas une catégorie en soi, puisqu'il est nécessairement constitué à partir de temps de parcours expérimentés ou de temps de parcours instantanés, voire d'une combinaison des deux : dès qu'une valeur mesurée ou estimée de temps de parcours est utilisée dans un contexte opérationnel, c'est qu'on lui prête un certain caractère

prédictif, en supposant par exemple une certaine stabilité des conditions de trafic courantes. En ce sens, toute valeur de temps de parcours utilisée pour l'information ou le guidage des usagers constitue un temps de parcours prédictif.

3.4. SYNTHÈSE

La notion de temps de parcours, intuitivement évidente lorsqu'elle concerne un individu, devient complexe à manipuler lorsqu'il s'agit d'un flot. Suivant l'utilisation qui en est faite et les méthodes utilisées pour le mesurer, des définitions diverses du temps de parcours seront ainsi utilisées. Il faut en pratique être très attentif à la définition utilisée pour éviter les erreurs, notamment lorsqu'il est nécessaire de combiner des informations (calcul de temps de parcours sur un itinéraire à partir de temps de parcours tronçons, calculs de valeurs moyenne, fusion d'estimateurs multiples).

4. MESURE ET ESTIMATION DU TEMPS DE PARCOURS

4.1. PRINCIPES GÉNÉRAUX, NATURE DES TEMPS DE PARCOURS MESURÉS

Qu'il s'agisse de mesure ou d'estimation, le problème est similaire : il s'agit, à partir d'un ensemble de mesures, d'avoir une valeur unique capable de caractériser le temps de parcours moyen d'un flot sur un chaînon ou une portion d'itinéraire. Il existe ainsi deux grandes classes de techniques :

- Soit on mesure le temps de parcours effectif d'un certain nombre de véhicules (voire de tous) et on en déduit une valeur moyenne. On travaille alors à partir de temps de parcours *expérimentés*, au sens défini plus haut ;
- Soit on utilise des mesures globales sur le flot pour en déduire directement une valeur moyenne de temps de parcours. Cette valeur est alors en général un temps de parcours *instantané*, à moins qu'un modèle d'écoulement du trafic soit utilisé pour déterminer un temps de parcours *prédictif*.

Il est bien entendu possible de combiner les deux types de techniques, ce qui peut permettre d'améliorer les performances des méthodes d'estimation. On se trouve alors devant un problème de fusion de données relativement délicat, compte tenu de la nature différente des temps de parcours mesurés par les deux types de techniques. Ce problème de fusion de données sera abordé dans la seconde partie de ce rapport.

4.2. DISTRIBUTIONS DES TEMPS DE PARCOURS

La connaissance des distributions des temps de parcours est nécessaire pour la significativité de la moyenne des temps de parcours et pour la détermination de l'incertitude associée à cette moyenne dans le cas où elle aurait un sens.

Les variations du temps de parcours sont de diverses natures et résultent des modes associés à la circulation automobile et à la dimension spatiale et temporelle du trafic. Ainsi,

- les variations temporelles peuvent provenir de changements de flux du trafic routier, du mélange des types de véhicules et des groupes de conducteurs ; de l'éclairage, des conditions météorologiques, et des éventuels incidents survenus dans la circulation.
- les variations spatiales peuvent être transversales (selon la voie, la direction,...) et/ou longitudinales, dues à des différences de flux du trafic, à la conception du réseau routier, et/ou la régulation du trafic.
- Enfin, les variations modales peuvent provenir de différences de comportement des conducteurs combinées avec des différences de performances de leur véhicule et des conditions de trafic rencontrées.

Hormis une étude portant sur la variabilité des temps de parcours des voies rapides urbaines de la région d'Île-de-France (SIER, 2000), peu de travaux ont réellement porté sur l'analyse des distributions des temps de parcours et les rares études publiées sur ce sujet font état et exclusivement de distributions unimodales faute de savoir traiter le cas multimodal.

C'est ainsi que, par exemple, Richardson et al., (1978) ont observé que les temps de parcours ont souvent des distributions asymétriques et leptokurtiques (*i.e.* distributions à queues lourdes).

Ce leptokurtisme des distributions du temps de parcours traduit essentiellement l'existence de fortes valeurs des temps de parcours, propriété observée empiriquement sur les données. Ce résultat a été partiellement confirmé par des études empiriques menée par Thakuria et Sen (1992) et par Mohammadi (1997), dans lesquelles une modélisation log-normale² a été utilisée (figure I-1).

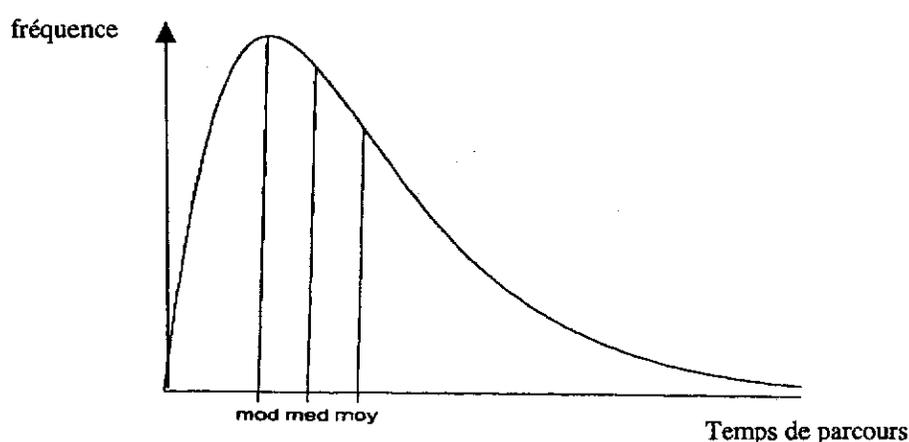


Figure I-1 : Distribution théorique des temps de parcours : cas log-normale.

Sur des bases tout aussi empiriques, d'autres auteurs, ont retenu la distribution normale (cf. figure I-2.) comme distribution des temps de parcours mais avec des variances relativement fortes traduisant ainsi l'aspect plat de la distribution. Ces fortes variances des valeurs observées,

² Distribution en cloche asymétrique à gauche, caractérisée par le fait que le mode soit inférieur à la médiane et celle-ci à la moyenne, *i.e.* : $\text{mod} \leq \text{med} \leq \text{moy}$. Rappelons d'une variable aléatoire suit une loi log-normale si son logarithme suit une loi normale.

traduisant ainsi l'aspect plat de la distribution. Ces fortes variances des valeurs observées, conjuguées à la présence parfois de mode significatif pour des valeurs importantes des temps de parcours mettent en cause la significativité même d'un temps de parcours moyen (cf. partie I).

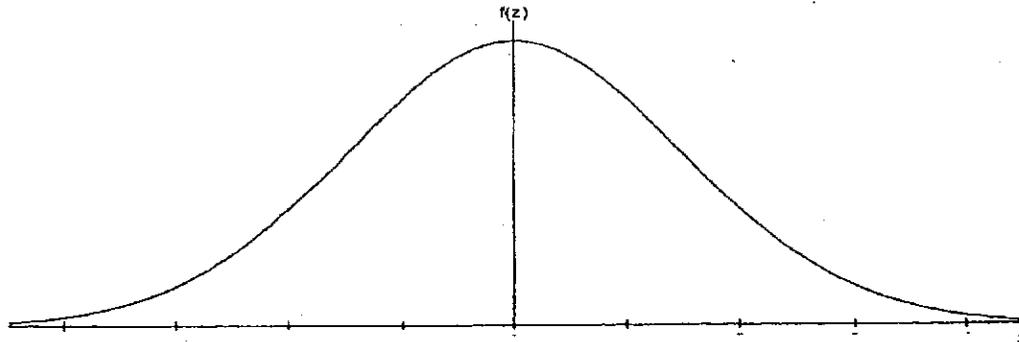


Figure I-2 : Distribution théorique des temps de parcours : cas log-normale.

Notons que l'existence de nombreuses valeurs aberrantes de mesure rendent souvent difficile l'assimilation d'une distribution empirique (histogramme) à une distribution théorique.

Dans le cas où la distribution des temps de parcours fait apparaître clairement plus de deux modes, la démarche classiquement utilisée consisterait à l'ajuster par une combinaison convexe de plusieurs distributions normales, voire log-normales. Une fois les composantes de cette combinaison identifiées, il conviendrait alors de découper les temps de parcours en classes de telle sorte que la distribution des temps de parcours dans chacune des classes corresponde à une des distributions composantes unimodales (cf. Habouzit, 2000). Ceci est peu utilisé dans la pratique.

4.3. MESURE DIRECTE DU TEMPS DE PARCOURS

Il existe principalement deux familles de techniques de recueil du temps de parcours. La première repose sur l'utilisation des moyens de mesures liés à l'infrastructure (capteurs au sol). Il s'agit principalement d'observateurs et/ou de détecteurs de mobiles localisés physiquement le long de routes étudiées. C'est le cas par exemple des capteurs de trafic auxquels sont associés des modules de couplage (reconnaissance de caractères, de paquets de véhicules, de signatures), d'enquêtes minéralogiques réalisées aussi bien manuellement qu'à l'aide de capteurs vidéo via des modules d'identification automatique des véhicules.

La seconde famille repose quant à elle sur l'emploi de détecteurs embarqués à bord de véhicules (GPS, téléphone mobile, système d'information et de guidage embarqué). On retrouve dans cette famille les techniques classiques dites de *véhicules test*. Ces véhicules peuvent être soit actifs *i.e.* instrumentés et ayant pour consigne la collecte du temps de parcours (véhicules flottants) soit passifs *i.e.* véhicules déjà dans le flot de trafic pour d'autres raisons et disposant de moyens de communication et de mesure de leurs temps de parcours (véhicules traceurs) comme par exemple, le système GPS, la poursuite des téléphones mobiles,...

4.4. ESTIMATION DU TEMPS DE PARCOURS

Le coût d'acquisition des données, conjugué aux handicaps inhérents à chaque technique de recueil de données³ font que la mesure directe d'un temps de parcours représentatif et de bonne fiabilité reste limitée à des applications très spécifiques ou à des études d'évaluation de systèmes de gestion de trafic. Par ailleurs, ces techniques fournissent des temps de parcours *a posteriori* (i.e. des temps de parcours réalisés) ce qui exclut les applications temps réel ou en prévision. On s'est alors orienté vers des techniques d'estimation du temps de parcours utilisant certaines données disponibles, permettant une fois le modèle validé, la prévision du temps de parcours.

Deux directions de recherches coexistent dans ce cadre. La première repose sur l'utilisation de techniques statistiques qui engendrent des modèles dits *de fonctionnement* : modèles qui cherchent seulement à reproduire une relation constatée entre des entrées (les causes) et des sorties (les effets) d'un système. L'autre direction fait appel à la physique du phénomène de trafic (théorie du trafic) pour expliciter lois physiques⁴ reliant le temps de parcours (ou la vitesse) à certaines variables caractérisant le flot du trafic. Ce type de modèle peut être qualifié de modèle de *connaissances* car il tente de reproduire le plus fidèlement possible les mécanismes élémentaires qui gouvernent le fonctionnement détaillé du phénomène étudié (ici le trafic).

4.4.1. Approche statistique

Cette première approche traite l'estimation du temps de parcours comme un problème de régression ou d'auto-régression utilisant des caractéristiques du trafic réputées pertinentes (principalement le débit et le taux d'occupation) pour la reconstitution du temps de parcours. Hormis l'hypothèse de causalité entre le temps de parcours et les caractéristiques retenues, aucune loi de comportement n'est imposée et les paramètres du modèle sont estimés à partir des données. Les principaux résultats obtenus reposent sur des modèles linéaires simples : techniques de séries temporelles, régression linéaires (cf. Lam et Rothery, 1970 ; Abours, 1985 ; Dailey, 1993 ; Sisiopiku, 1994).

D'autres méthodes ont été aussi utilisées dans ce contexte. On peut citer des techniques d'estimation non linéaires (Cohen et Nouvelière, 1997) ou encore les modèles dits dynamiques : approches stochastiques et filtre de Kalman (Dailey, 1997 ; Petty et al., 1994). Cette dernière classe permet d'intégrer les données au fur et à mesure de leurs disponibilités avec mise à jour des paramètres du modèle, et s'apparente à un problème de fusion de données dans lequel les informations sont disponibles successivement et non simultanément.

3 Les mesures du temps de parcours issues de l'une ou de l'autre famille souffrent d'un ensemble d'inconvénients identifiés dont le principal est ce que l'on peut appeler le dilemme d'exhaustivité v.s. précision : une technique peut se révéler très satisfaisante en terme de précision des données fournies mais peut manquer de représentativité (spatiale ou temporelle) et inversement.

4 Les lois et les principes fondamentaux de la physique du trafic, ne suffisent pas à écrire un système d'équations comportant autant d'équations que d'inconnues. On introduit alors d'autres lois dites phénoménologiques ou de comportement (c'est le cas par exemple du diagramme fondamental) en vue de fermer le système d'équations.

Ces techniques donnent généralement des résultats corrects (de l'ordre de 12 % d'erreur mesurée par le critère quadratique moyen relatif⁵ pour les plus performantes) dans le cas de situations de trafic stationnaire et de réseaux moyennement chargés. Cependant, elles manquent de robustesse vis-à-vis de la qualité des données d'entrée (valeurs aberrantes, pannes de capteurs,...) et sont très sensibles aux violations des hypothèses qui ont gouverné à leur mise en œuvre (liaisons imposées entre variables, modélisation globale, ignorance de la propagation du trafic,...). Pour une revue des différentes méthodes d'estimation du temps de parcours en milieu urbain et de leurs performances, on renvoie le lecteur intéressé à Sisiopiku et Rouphail (1994a).

Par ailleurs, cette famille de modèles souffre d'un inconvénient majeur, intrinsèque à la modélisation statistique, dans laquelle le modèle est décrit par une représentation mathématique plaquée sur des données expérimentales. Les paramètres ne sont pas *réifiables* (absence de sens physique des paramètres du modèle). Leur portée est alors locale car intimement liée au contexte de modélisation, limitant ainsi la transférabilité et l'universalité d'un système d'estimation statistique du temps de parcours.

L'utilisation d'un tel système sur un site autre que celui sur lequel il a été développé et évalué passe par la collecte de nombreuses données à la fois des variables du trafic et du temps de parcours. Ces données ne peuvent s'obtenir que via des campagnes de mesures relativement coûteuses.

4.4.2. Approches physiques et phénoménologiques

Ce type d'approche consiste à expliciter un ensemble de lois physiques de comportement du trafic et d'en dériver des modèles *a priori*, reliant le temps de parcours (en général un temps de parcours instantané) aux variables de trafic mesurées sur le terrain. C'est de loin cette approche qui a suscité le plus de travaux différents, car elle fait appel à la physique du phénomène lui-même et devrait donc permettre la constitution de modèles plus facilement interprétables et transférables qu'une approche purement statistique. La plupart des systèmes opérationnels aujourd'hui en service sont fondés sur ce type de modèle.

Les modèles les plus rustiques de cette classe reposent uniquement sur la conservation des véhicules, sans autres considérations méthodologiques. On les appelle modèles de stock. Ils consistent à mesurer les flux entrant et sortant de la section où le temps de parcours est recherché, afin d'estimer le nombre de véhicules présents à l'intérieur de cette section (le stock). En divisant ce stock par le débit mesuré en sortie (que l'on suppose donc constant dans le futur proche), on obtient une estimation d'un temps de parcours, qui en l'occurrence n'est pas un temps de parcours instantané mais une prédiction du temps de parcours réel du dernier véhicule entré.

⁵ L'erreur quadratique moyenne relative (EQMR) est définie par : $\|TP_{référence} - TP_{estimé}\|_2^2 / \|TP_{estimé}\|_2^2$ avec :

$$\|x\|_2^2 = \sum_{i=1}^n (x_i)^2$$

Cette méthode, peu utilisée telle quelle, présente deux inconvénients :

- Fondée sur une hypothèse de constance du débit en sortie, elle présente une certaine pertinence en congestion, où le débit sortant est conditionné par l'offre en aval, mais beaucoup moins en fluide où les variations de débit correspondent aux variations de la demande ;
- Totalement en boucle ouverte, l'estimation du stock subit des dérives importantes si notamment la mesure des débits présente un biais systématique (ce qui est le cas le plus fréquent) ; un fréquent recalage manuel du stock est nécessaire.

Le deuxième type de méthodes est fondé sur une estimation (ou parfois une mesure directe) de la vitesse du flot en un certain nombre de points. Ce type de méthode est essentiellement utilisé sur voies rapides urbaines, où la densité de points de mesure est suffisante et la circulation assez homogène. La méthode d'estimation consiste en effet à estimer la vitesse ponctuellement à partir de stations de mesure donnant le débit et le taux d'occupation, la vitesse étant calculée comme le rapport des deux divisé par la longueur électrique moyenne des véhicules. Cette valeur de vitesse est considérée comme représentative de la vitesse du flot sur toute la section entourant la station de mesure, soit une longueur typique de 500 mètres. Le temps de parcours est alors calculé en divisant la longueur de la section par la vitesse estimée. Le temps de parcours global sur un ensemble de section est calculé par sommation des estimations locales : c'est un temps de parcours instantané.

Les performances de ce type de méthode sont suffisantes pour qu'elles soient utilisées opérationnellement, notamment sur les réseaux de voies rapides d'Île-de-France et sur le réseau CORALY.

Une troisième classe de méthode est fondée directement sur la modélisation de l'écoulement du trafic. Les variables d'entrée sont donc les débits entrant dans le réseau, mais à l'équation de conservation seule utilisée par les méthodes de stocks on ajoute la prise en compte de la propagation du trafic, soit sous forme microscopique soit sous forme macroscopique (Sanwal *et al.*, 1996 ; Lebacque, 1996). Ce type de modèles, a priori le plus riche en informations, pose en revanche des problèmes liés à la précision des modèles utilisés, pas toujours assez fins pour permettre une estimation solide des temps de parcours, ainsi que du paramétrage de ces modèles.

Une dernière classe de méthodes utilise la théorie des files d'attente, en particulier pour estimer le retard subi aux feux de circulation en milieu urbain (Kimber et Dailey, 1986 ; Weymann *et al.*, 1994). Le principal inconvénient de ces méthodes, fondées essentiellement sur le modèle de retard de Webster (Webster, 1958), est leur difficulté à prendre en compte les phénomènes de congestion.

4.4.3. Approches mixtes

Certaines méthodes d'estimation des temps de parcours combinent une forme de modélisation phénoménologique avec des techniques statistiques.

C'est ainsi par exemple que Hall et Persaud (1989), à partir de mesures ponctuelles de la vitesse, utilisent des méthodes de filtrage pour interpoler ces vitesses sur les sections séparant les stations de mesure.

C'est ainsi également que la méthode dite BRP (Bonvallet et Robin-Prévallée, 1987) est fondée, après (Abours, 1985), sur l'existence d'une relation linéaire entre le temps de parcours sur un chaînon urbain et le taux d'occupation mesuré sur une station de mesure bien placée. Le fondement du modèle est donc purement statistique, mais la valeur de ses coefficients est calculée à partir de considérations sur le trafic : le modèle est calibré en considérant que le temps de parcours est égal à une valeur déterminée (temps de parcours libre) lorsque le taux d'occupation tend vers zéro, et qu'en congestion totale au contraire ce temps est égal au nombre maximal de véhicules contenus dans le chaînon divisé par le débit mesuré en sortie. Cette méthode est aujourd'hui utilisée pour estimer les temps de parcours en ville, en particulier sur le réseau urbain de Paris.

4.5. CONCLUSIONS

A travers cette revue synthétique des méthodes de recueil et d'estimation du temps de parcours, on mesure l'importance que revêt la problématique du temps de parcours, dont le renouveau grandissant est principalement dû à la profusion de systèmes d'information et de guidage des usagers des infrastructures routières ainsi qu'à l'émergence d'un nouveau marché d'information routière principalement en temps réel.

L'utilisation des systèmes de localisation des véhicules (GPS, positions des téléphones mobiles,...) a relancé l'utilisation des véhicules traceurs pour le recueil en temps réel du temps de parcours, jadis cantonnés dans des applications de suivi et de calibrage des outils d'estimation et de simulation du trafic hors-ligne, faisant apparaître des systèmes de recueil multiformes très prometteurs.

Côté estimation du temps de parcours, malgré la profusion de modèle d'estimation et de leurs sophistications, des études empiriques (cf. El Faouzi et Lesort, 1994, Sisiopiku, 1994) ont montré qu'il est difficile d'obtenir un modèle global d'estimation du temps de parcours à partir des données de trafic, couvrant différentes situations de demande. Il a été ainsi montré qu'en milieu urbain en particulier, le temps de parcours est indépendant, tout au moins linéairement, du débit et du taux d'occupation (deux principales variables explicatives traditionnellement utilisées dans ce contexte) dans le cas de situations de faible demande. Par ailleurs, au-delà d'un certain seuil de taux d'occupation, l'estimation du temps de parcours à partir des données de trafic devient quasi-impossible. Cette impossibilité opérationnelle est essentiellement due aux remontées des files d'attente provoquant l'arrêt de véhicules sur les capteurs.

Seules les estimations du temps de parcours à partir du débit et du taux d'occupation correspondant à des valeurs du taux d'occupation non-extrêmes (supérieures à 20% et inférieures à 80%) sont réellement pertinentes. Ceci couvre bien entendu une large gamme de conditions de trafic, mais pose par exemple des problèmes de localisation des points de mesure sur les réseaux urbains : trop près des feux, ils ne fourniront de l'information qu'en condition de trafic très fluide, trop loin ils ne présenteront d'intérêt qu'en saturation complète (cf. Lesort, 1991 ; Lesort et Mongeot, 1994).

5. DIMENSIONNEMENT D'UN SYSTÈME DE MESURE DU TEMPS DE PARCOURS

5.1. INTRODUCTION

La question du dimensionnement d'un système de mesure ou d'estimation des temps de parcours se pose dans la réalité moins souvent qu'on ne pourrait le croire. Bien des systèmes se sont en effet construits et se construisent encore sur la base d'une infrastructure existante de recueil de donnée, destinée à l'origine à la régulation du trafic (cas par exemple du réseau de détection utilisé dans Paris intra-muros) ou à la surveillance (cas du boulevard périphérique parisien). Dans ces conditions, la question posée n'est pas celle du dimensionnement optimal du système, mais celle de l'utilisation optimale d'un système existant et des performances qu'il rend possibles. Ce n'est que relativement récemment que, à l'occasion de la définition de systèmes de recueil destinés avant tout à la mesure des temps de parcours, la question du dimensionnement s'est trouvée véritablement posée, et ce de deux points de vue :

- ❑ Quel est le dimensionnement optimal pour arriver aux meilleures performances possibles d'un système de mesure ou d'estimation
- ❑ Quel est le dimensionnement minimal permettant d'obtenir des performances significatives

Il est clair que la réponse à ces questions implique d'avoir une idée précise de ce qu'on appelle les performances du système, ce qui n'est pas une question triviale. Ce point sera développé dans la troisième partie de l'étude. Il est possible en revanche de définir, pour chaque type de système, ce que sont les facteurs dimensionnants.

5.2. MESURES DIRECTES

Les techniques de recueil des temps de parcours fondées sur l'utilisation des véhicules traceurs⁶ permettent, lorsqu'elles sont bien dimensionnées, des mesures précises des temps de parcours. La qualité des données obtenues dépend de certaines limites intrinsèques à la technologie de localisation des mobiles utilisée et d'un certain nombre de paramètres de mise en œuvre. Ainsi, sur le plan technologique, on peut citer les problèmes liés à la localisation précise des mobiles et principalement le biais introduit par le GPS et les zones d'ombres en milieu urbain en particulier. Les paramètres de mise en œuvre concernent davantage la composition d'un échantillon statistique de véhicules traceurs. Pour ce faire, certaines recommandations, principalement de nature statistique doivent présider au choix de cette composition. A titre d'exemple, on peut citer, le type de véhicules et des conducteurs qui doivent entrer dans cette composition, les qualités statistiques que doit posséder un tel échantillon pour une mesure des temps de parcours avec un degré spécifié de précision.

⁶ Véhicules déjà dans le flot de trafic pour d'autres raisons et disposant de moyens de communication et de mesure de leurs temps de parcours.

5.2.1. Échantillonnage

Dans la pratique du recueil du temps de parcours, l'échantillonnage⁷ se résume le plus souvent à la spécification de la taille de l'échantillon, i.e. nombre d'observations, et du choix des périodes temporelles de mesures. En effet, dans la plupart des applications de mesure directe du temps de parcours, on a rarement le choix des entités expérimentales (ici les véhicules). Le choix des entités utilisées est principalement dicté par les équipements nécessaires à la mesure du temps de parcours (badges, systèmes embarqués, téléphonie mobile,...). Pour des raisons de coût, ces systèmes ne concernent qu'un pourcentage relativement faible du flux de trafic global, de sorte que le plan de sondage est quasiment imposé.

Pour garantir une précision acceptable du sondage, la théorie statistique impose que l'échantillon retenu soit *représentatif* de la population⁸ totale, faute de quoi, il peut exister un écart entre le résultat de sondage et celui que l'on aurait obtenu par une étude *exhaustive* de la population.

Dans le cas des véhicules traceurs et contrairement aux techniques traditionnelles de mesure du temps de parcours comme celles des enquêtes minéralogiques, véhicules flottants,... le nombre de mesures n'est pas paramétrable dans la mesure où il est fonction de nombre de véhicules instrumentés présents sur un tronçon à un instant donné. Ce nombre lui-même dépend du type de réseau, des profils du trafic, du moment de la journée, Cependant, il est possible de déterminer un nombre de mesures nécessaires pour garantir un niveau de précision fixé à l'avance, au delà duquel les exigences de précisions ne sont plus vérifiées. Le calcul de ce nombre de mesures fait l'objet du paragraphe suivant.

5.2.2. Taille de l'échantillon

L'approche que nous allons présenter ici consiste à déterminer le nombre minimal d'observations des temps de parcours obtenus par des véhicules traceurs nécessaires pour approcher avec une précision donnée le temps de parcours du flot de trafic dont ces véhicules sont issus.

Le calcul de cette taille minimale repose sur l'idée intuitive que la moyenne d'un échantillon particulier ne correspond pas exactement à la moyenne de la population que nous désirons connaître. Toutefois, la valeur calculée sur un échantillon peut nous donner une idée approximative de la valeur de la population. Dans le contexte des temps de parcours, on cherche à quantifier la proximité de la moyenne des temps de parcours (ou de façon équivalente la vitesse moyenne) des véhicules traceurs (moyenne sur un échantillon de véhicules, notée TP_{traceurs}) avec le temps de parcours moyen (ou encore avec la vitesse du flot) obtenu à partir de l'ensemble des véhicules composant le flot de trafic (moyenne de la population des véhicules, notée $TP_{\text{référence}}$). On définit classiquement ce critère de proximité par la probabilité que les écarts absolus ($TP_{\text{référence}} - TP_{\text{traceurs}}$) ne dépassent pas une valeur seuil donnée $\epsilon_a(\alpha)$ (l'erreur maximale autorisée associée à un niveau de précision α , 5% le plus souvent) :

$$P(|TP_{\text{référence}}(t) - TP_{\text{traceurs}}(t)| \leq \epsilon_a(\alpha)) = 1 - \alpha \quad (33)$$

⁷ L'échantillonnage consiste à choisir parmi les éléments de la population (ici le flot de trafic sur un axe par exemple) une sous-population (véhicules) pour laquelle nous obtiendrons des observations.

⁸ Le terme population a une signification particulière en statistique ; c'est l'ensemble d'entités expérimentales à étudier, à partir duquel est extrait un échantillon.

Ce critère, appelé intervalle de confiance au niveau de confiance $(1-\alpha)$, représente la façon algébrique de dire qu'il y a 95% de chance ($\alpha = 5\%$) que le temps de parcours de référence $TP_{référence}(t)$ tombe entre $TP_{traceurs}(t) - \epsilon_a(\alpha)$ et $TP_{traceurs}(t) + \epsilon_a(\alpha)$. Sur la base de ce critère il est maintenant possible de déterminer le nombre d'observations nécessaire à l'estimation du temps de parcours.

Si $u(\Delta x, t)$ désigne la vitesse du flot à l'instant t sur un segment de route Δx , en considérant n véhicules traceurs sur ce segment de route, dont on connaît les vitesses individuelles (v_1, v_2, \dots, v_n) à l'instant t , on peut approximer $u(\Delta x, t)$ par la moyenne arithmétique des vitesses des véhicules traceurs, notée v_E et définie par :

$$v_E = \frac{1}{n} \sum_{i=1}^n v_i \quad (34)$$

Chaque mesure v_i est entachée d'une incertitude δ_i , donc d'après son expression ci-dessus, une erreur δ_i affectant un résultat v_i entraîne sur v_E une erreur égale à δ_i/n .

En supposant que les $(\delta_i)_{i=1, \dots, n}$ de moyenne nulle et variance commune $\sigma^2 = \text{var}(\delta_i)$, les erreurs $(\delta_i/n)_{i=1, \dots, n}$ affectant v_E sont des variables aléatoires indépendantes de moyenne nulle et de variance $\text{var}(\delta_i/n) = \sigma^2/n^2$.

La somme de ces n erreurs est l'erreur aléatoire totale sur v_E et la variance de la distribution d'échantillonnage des moyennes⁹ est donc donnée par :

$$\text{Var}(v_E) = \sigma_{v_E}^2 = \frac{\sigma^2}{n} \quad (35)$$

On voit donc que lorsque n (nombre de véhicules traceurs) s'accroît, la distribution d'échantillonnage de v_E est de plus en plus concentré sur son objectif $u(\Delta x, t)$ et qu'à la limite elle tend vers zéro¹⁰ !.

En supposant σ^2 connue, et en réarrangeant cette équation, on définit la taille de l'échantillon par :

$$n = \left(\frac{\sigma}{\sigma_{v_E}} \right)^2 \quad (36)$$

L'intervalle de confiance de la moyenne, montre que le seuil $\epsilon_a(\alpha)$ figurant dans le critère est donnée par :

⁹ Ce qui suppose l'indépendance statistique des (v_i) , hypothèse vérifiée si l'échantillonnage est aléatoire. Dans le cas contraire, on doit tenir compte des inter-corrélations entre les vitesses des véhicules constituant l'échantillon (cf. Sen et al., 1997).

¹⁰ Il est plus que probable que dans le cadre des véhicules traceurs, le fait d'augmenter le nombre de véhicules ne fera pas tendre la variance du temps de parcours moyen vers zéro mais bien au contraire. Ceci provient de la violation de l'indépendance mentionnée ci-dessus.

$$\varepsilon_a(\alpha) = u_{\alpha/2} \sigma_{V_E} \quad (37)$$

où $u_{\alpha/2}$ désigne la marge d'erreur et se lit dans la table de Gauss-Laplace¹¹ en fonction de la probabilité attribuée au paramètre α .

On pourra alors écrire :

$$n = \left(\frac{\sigma}{\sigma_{V_E}} \right)^2 = \left(\frac{u_{\alpha/2} \sigma}{u_{\alpha/2} \sigma_{V_E}} \right)^2 = \left(\frac{u_{\alpha/2} \sigma}{\varepsilon_a(\alpha)} \right)^2 \quad (38)$$

Comme σ est généralement inconnue, elle est estimée à partir de l'échantillon étudié. Notons s cette estimation empirique, et on a alors la relation suivante :

$$n = \left(\frac{u_{\alpha/2} s}{\varepsilon_a(\alpha)} \right)^2 \quad (39)$$

avec :

n : la taille de l'échantillon requise

s : l'écart - type de l'échantillon

$\varepsilon_a(\alpha)$: précision du sondage, représentant la borne supérieure de l'erreur autorisée par le cahier des charges

$u_{\alpha/2}$: la valeur de l'intervalle de confiance choisie en utilisant la distribution normale ($n > 30$).

On voit alors que le nombre d'observations dépend de la précision requise pour un tel système de mesure.

5.2.3. Types de flottes

La constitution d'un échantillon statistiquement significatif pose le problème de la composition de ce dernier et donc la méthode de sondage à utiliser. Les deux principaux points suivants doivent être examinés préalablement à toute mise en œuvre :

- types de véhicules utilisés : voitures particulières, taxis, bus, autres.
- type de conducteurs : cette question est intimement liée à la première puisqu'elle est une des conséquences du choix du type de véhicule.

La question peut aussi se poser en terme de couverture spatiale du réseau puisque certains types de véhicules utilisent des sous-réseaux bien identifiés du réseau global (par exemple le réseau des

¹¹ Le recours à la loi normale se justifie ici par le fait que si n est suffisamment grand ($n > 30$ suffit), la distribution d'échantillonnage des moyennes est approximativement *normale*, quelle que soit la distribution de la population, pourvue que l'espérance et l'écart-type de cette dernière soient finies.

transports en commun, taxi,...) qui se limitent aux principaux axes routiers et centres de dessertes (aéroports, gares..., ...), là où justement il est possible de disposer d'une information sur l'état de trafic par d'autres moyens de mesure (capteurs par exemple)... Cette dernière composante peut être perçue aussi bien comme une conséquence d'un choix de type de véhicules, que comme critère de choix si l'on impose la couverture de certains axes du réseau routier.

Enfin, la représentativité en termes de conditions de circulation doit être un critère de sélection de la flotte à utiliser. Ainsi, dans le cas d'un échantillon exclusivement constitué d'une flotte de taxis ou de bus, l'estimation du temps de parcours obtenue sera vraisemblablement biaisée. En effet, ce type de flotte tend selon les cas à sous-estimer (dans les cas des taxis) ou à sur-estimer (cas des bus) les temps de parcours réels. A cause de l'utilisation des sites propres, des comportements de conduite des conducteurs, et la dynamique particulière de ce type de flotte ne permettent pas d'approcher aussi finement que possible la vitesse du flot de trafic à partir des vitesses obtenues à partir d'un échantillon de cette flotte.

Notons que selon que l'on s'intéresse à un réseau urbain ou interurbain les flottes potentiellement utilisables ne sont pas les mêmes. C'est ainsi qu'en urbain, on a le choix entre les véhicules de transports en commun (taxis et bus essentiellement) et les voitures particulières, alors qu'en interurbain, le choix se pose entre les poids lourds et les voitures particulières.

5.3. SYSTEMES D'ESTIMATION

La question du dimensionnement d'un système d'estimation des temps de parcours à partir de la mesure des caractéristiques du trafic est relativement peu traitée dans la littérature. Ceci tient tout d'abord à une difficulté conceptuelle : si la mesure directe d'un certain nombre de valeurs de temps de parcours permet de raisonner en termes de représentativité d'un échantillon, la qualité d'une estimation ne peut se mesurer qu'*a posteriori* de façon expérimentale.

Une étude de dimensionnement doit donc reposer sur une expérimentation nécessairement surabondante en termes d'équipements, ce qui est rarement réalisé en pratique. La majorité des travaux menés sur l'estimation des temps de parcours s'est en effet appuyée sur une infrastructure de capteurs donnée, mise en œuvre dans le cadre d'autres préoccupations (mesure des flux) et qu'il s'agissait d'utiliser au mieux. La question du dimensionnement ne s'est donc pas posée. Dans les quelques cas où une infrastructure de mesure a été réalisée spécifiquement à une échelle relativement importante pour la mesure des temps de parcours, l'étude de dimensionnement s'est fondée sur des considérations empiriques et surtout économiques compte tenu du coût d'installation et de raccordement d'une station de mesure.

À défaut d'une véritable étude de dimensionnement, il est en revanche possible de qualifier la sensibilité d'une méthode d'estimation à ses paramètres, ce qui donne une idée de l'importance d'un calibrage fin de ces paramètres. Un certain nombre de considérations empiriques tiennent également en partie lieu d'étude de dimensionnement : pour évaluer la densité nécessaire de stations de mesure de vitesse, il faut avoir une idée de l'homogénéité des vitesses sur les sections concernées.

Enfin, même dans le cas où l'infrastructure de mesure est dédiée à l'estimation des temps de parcours, les considérations économiques sont bien souvent prépondérantes dans son dimensionnement.

5.3.1. Méthodes de stocks

Les méthodes de stock sont fondées sur l'utilisation d'un unique couple de stations de mesures, situées en entrée et sortie de la zone considérée. Par ailleurs, l'erreur relative faite sur l'estimation du temps de parcours dépend uniquement de l'erreur relative faite sur la mesure du stock (erreur sur la différence entre débits entrants et sortant) ainsi que de l'erreur sur la projection du débit sortant. Le système est ainsi peu sensible à sa taille, si ce n'est qu'une erreur absolue sur le stock aura d'autant moins d'influence que celui-ci sera important, c'est à dire la distance entre stations élevée (à condition bien sûr qu'il n'y ait pas de fuite entre les stations).

5.3.2. Méthodes fondées sur l'estimation ponctuelle de la vitesse

La question fondamentale dans le dimensionnement de ce type de système est bien entendu celle de la densité des points de mesure. La réponse couramment apportée est de placer une boucle tout les 500 m, ce qui est par exemple fait sur le boulevard périphérique parisien (à noter que dans ce dernier cas la distance de 500 m est également l'interdistance moyenne entre échangeurs, ce qui fournit une justification à la densité choisie).

Si cette interdistance fournit des résultats convenables, ce qui a été montré expérimentalement, la question de savoir ce qui serait envisageable avec une moindre densité de points de mesure n'a pas de réponse simple.

Il n'existe tout d'abord pas véritablement de référence à laquelle comparer l'estimation : c'est en effet un temps de parcours instantané qui est estimé, et s'il est clair qu'une diminution de la densité des points de mesure dégrade l'estimation de ce temps par rapport à ce que donnerait une connaissance de la vitesse en tout point, la relation est moins claire avec la performance finale, c'est à dire la comparaison entre la valeur estimée et le temps de parcours réel des véhicules entrés au cours de la période d'estimation. Il semble empiriquement clair que la condition de base à respecter est celle de l'homogénéité des sections, à la fois du point de vue de l'infrastructure (nombre de voies, absence d'échangeurs au milieu des sections...) et du point de vue du trafic (présence fréquente de bouchons). Il est également probable que, plus que la densité de points de mesure, le nombre de points utilisés pour le calcul soit déterminant pour la robustesse du système vis à vis des perturbations locales ; ceci veut dire que moins la densité sera élevée, plus les itinéraires renseignés devront être longs. A contrario, il est clair que plus le temps de parcours estimé devient grand devant le temps d'évolution des conditions de trafic, plus l'utilisation d'un temps de parcours instantané devient sujette à caution, les valeurs maximales raisonnables pouvant aller de 15 à 30 ou 45 minutes selon la taille de l'agglomération concernée.

5.3.3. Méthodes fondées sur la modélisation de l'écoulement du trafic

Le dimensionnement des mesures nécessaires au fonctionnement de ce type de méthode renvoie directement à la question du calibrage, et plus précisément du recalibrage en ligne des modèles d'écoulement, question encore totalement du domaine de la recherche et pour laquelle peu de résultats sont disponibles. En théorie, un modèle de ce type devrait se contenter de la connaissance des flux entrants, des possibilités de sortie et des restrictions éventuelles, l'ensemble constituant les conditions aux limites du modèle. Compte tenu du caractère relativement grossier des modèles

d'écoulement du trafic, ceci n'est en pratique pas suffisant, et il faut pouvoir utiliser des mesures intermédiaires (débits, vitesse...) pour recalibrer le modèle. Malheureusement, ce recalibrage introduit nécessairement des discontinuités dans le fonctionnement du modèle d'écoulement rendant le résultat d'ensemble peu satisfaisant. Il est clair en revanche qu'au prix de la résolution de ces difficultés ce type de modèle, qui utilise au maximum l'information, devrait permettre des estimations relativement économes en moyens de mesure.

5.3.4. Méthodes mixtes : le BRP

L'indicateur BRP, au contraire de la plupart des méthodes d'estimation des temps de parcours, a fait l'objet d'un certain nombre d'études de dimensionnement. Des méthodes d'implantation des stations de mesure ont également été implantées (cf. Lesort et Mongeot, 1994).

Les exigences du BRP en termes de mesure sont directement liées à la nature des temps de parcours sur un réseau urbain régulé. Dans ces conditions en effet il est possible de considérer que les vitesses de roulage sont quasi constantes, et que la variance du temps de parcours dépend uniquement du temps passé en file d'attente aux feux. L'implantation des points de mesure doit ainsi avoir pour objectif de prendre en compte les variations de longueur des files d'attente. Une première approche aboutit ainsi à considérer qu'une station de mesure doit être implantée sur chaque chaînon aboutissant à un carrefour à feux : c'est ainsi que fonctionne le premier système de ce type, développé et installé par la Ville de Paris. Il faut noter au passage que les exigences ne sont pas les mêmes pour la mesure du débit, qui peut se faire en un point quelconque du chaînon, et du taux d'occupation qui doit être faite aussi près que possible de l'arrière de la file d'attente pour obtenir une dynamique maximum (trop loin du feu le taux d'occupation sera toujours très faible, trop près toujours trop élevé). Ceci peut d'ailleurs conduire, là où les files d'attente sont très variables, à implanter plusieurs stations de mesure. À l'inverse, certains chaînons terminés par des feux ne supportent quasiment pas de file d'attente, lorsque le feu qui les termine est par exemple pris dans une onde verte. Dans ces conditions l'implantation d'un point de mesure est inutile. Enfin, une congestion exceptionnelle étant toujours possible, une densité minimale de stations de mesure est nécessaire pour y faire face, même aux endroits où les files d'attente sont rares.

La méthode d'implantation montre au passage la fragilité du système lorsque les stations de mesure sont peu abondantes : un simple réglage des feux, modifiant la structure des files d'attente, peut dégrader considérablement les performances de l'estimation. De même, en milieu plutôt périurbain, où la vitesse des véhicules est variable, où il y a peu de feux, de nombreux carrefours sans feux et giratoires, l'implantation est difficile et les performances de la méthode décevantes.

6. CONCLUSION

Ce premier tour d'horizon des différentes notions de temps de parcours et des moyens des les mesurer et de les estimer a mis en évidence la complexité de la question, mais aussi la diversité des sources possibles d'information utilisables pour cette mesure. L'idée vient ainsi naturellement de tenter d'utiliser l'ensemble de l'information disponible, et de combiner les différentes sources pour obtenir une estimation unique meilleure que les estimations partielles fournies par chacune des sources. Les techniques de fusion de données sont précisément destinées à permettre cette

combinaison, en prenant en compte les natures différentes d'information fournies, avec les incohérences voire les contradictions qui peuvent exister entre elles.

PARTIE II.

FUSION DE DONNÉES POUR L'ESTIMATION DES TEMPS DE PARCOURS

1. INTRODUCTION

Les études présentées dans le chapitre précédent ont bien mis en évidence les limites des approches de la mesure des temps de parcours fondées sur une seule source de données.

Devant cette limitation et avec la multiplication des moyens de mesure disponibles (caméras, localisation grâce au GPS ou aux téléphones cellulaires,...), on recourt donc de plus en plus à d'autres sources de données visant à compléter l'information fournie par les moyens de mesure classiques, et par conséquent, à améliorer la qualité de l'estimation du temps de parcours. Le problème de l'estimation du temps de parcours devient alors un problème typique de *fusion de données* (cf. figure II-1.).

2. FUSION ET TEMPS DE PARCOURS : ETAT DE L'ART

Le recours à la fusion de données dans l'appréhension de l'état d'un système est justifié par la capacité des systèmes multi-capteurs à compenser les limites d'un capteur par et à exploiter l'effet de redondance et de complémentarité des informations capteurs.

Une étude sur les concepts de la fusion de données et un état de l'art des méthodes principales de fusion a fait l'objet d'un rapport (cf. El Faouzi, 2000). Nous ne développerons pas ces aspects dans cette partie et nous invitons le lecteur intéressé à consulter ce rapport pour de plus amples explications.

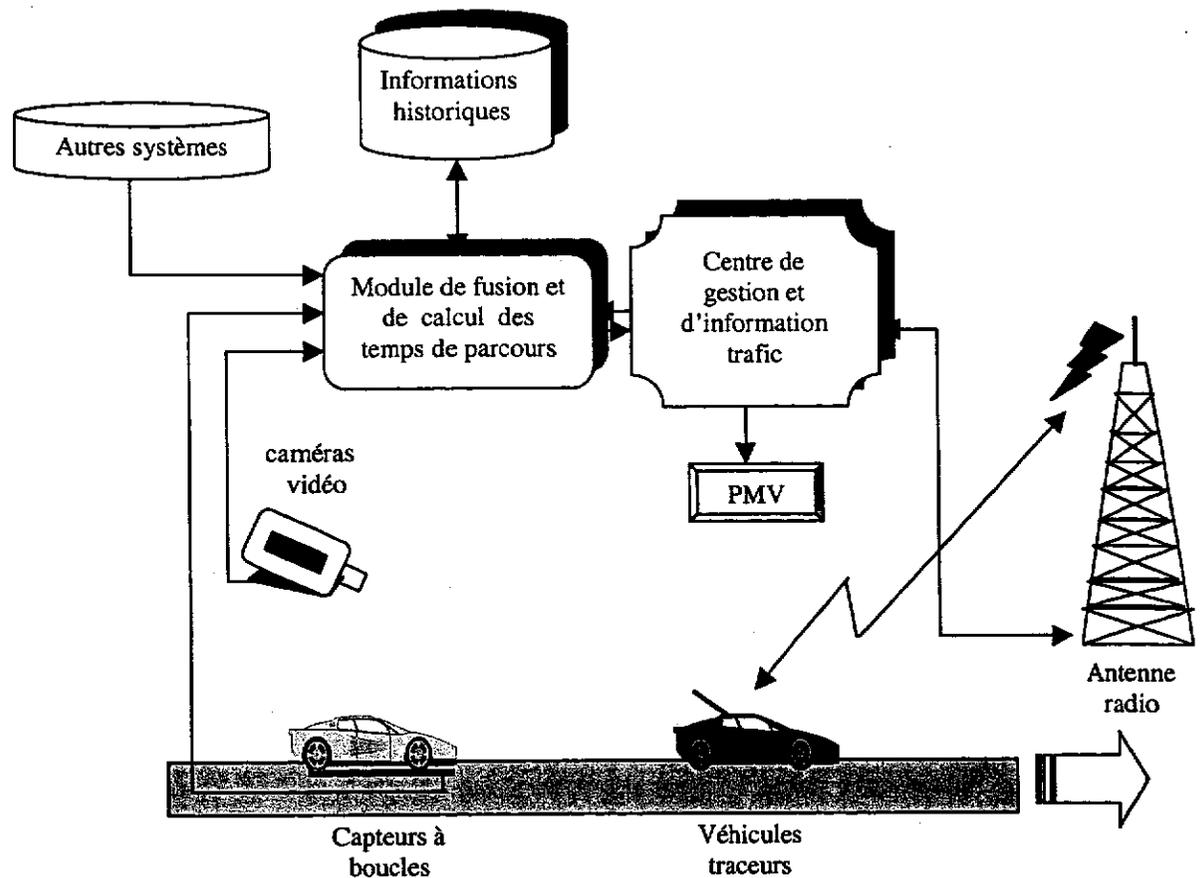


Figure II-1 : Recueil multiforme des données de temps de parcours

S'agissant de la fusion de données pour l'estimation du temps de parcours, elle a déjà été abordée sous deux angles différents :

- la fusion statistique d'estimateurs et d'algorithmes du temps de parcours
- et la fusion crédibiliste des temps de parcours.

La première approche permet de combiner des estimateurs individuels du temps de parcours, calculés à partir de chacune de sources d'informations disponibles. Cette façon classique de procéder peut être vue comme une fusion de données multi-sources préalablement traitées (données de haut niveau). La difficulté se résume alors à l'élaboration d'un système de pondération, associé aux estimateurs disponibles, synthétisant les qualités (précision, fiabilité,...) de chaque estimateur.

La seconde approche, moins classique, travaille directement sur les informations brutes (données de bas niveau) cherchant à exploiter au mieux la complémentarité et la redondance des sources d'information disponibles, en faisant appel à des formalismes adéquats.

2.1. FUSION STATISTIQUE D'ESTIMATEURS ET D'ALGORITHMES

L'idée centrale consiste à combiner différents estimateurs d'une même quantité (résultant d'une ou de plusieurs sources de données) pour améliorer la qualité de l'estimation du caractère étudié. Autrement dit, on se place dans le cadre de plusieurs estimateurs d'un même variable (ici le temps de parcours), et l'on cherche à construire un nouvel estimateur unique et à fort pouvoir synthétique, obtenu par combinaison des estimateurs disponibles.

Les procédures de fusion les plus utilisées sont du type *moyenne pondérée* dans lesquelles on attribue à chaque estimateur un poids qui reflète sa fiabilité. Ces opérateurs se justifient si l'on considère les différentes sources comme une seule source aléatoire dont les informations à agréger sont des réalisations

Pour le besoin de l'exposé, on suppose que les données et informations sur les temps de parcours proviennent de plusieurs sources indépendantes $i = 1, \dots, k$. Ces sources peuvent être basées selon les cas, sur des techniques de mesures directes du temps de parcours (véhicules test, véhicules traceurs, enquêtes minéralogiques, ...) ou sur des techniques d'estimation du temps de parcours. On notera φ_i l'estimateur du temps de parcours issu de la $i^{\text{ème}}$ source, avec $i = 1, \dots, k$.

L'estimateur synthétique recherché s'écrit alors comme une combinaison linéaire et convexe des k estimateurs $\varphi_i, i = 1, \dots, k$ disponibles :

$$\hat{\pi}_\omega = \sum_{i=1}^k \omega_i \varphi_i \quad \text{avec} \quad \sum_{i=1}^k \omega_i = 1 \quad (40)$$

Le problème de fusion se réduit à la recherche d'un système de pondération (ω_i) garantissant une *meilleure* fiabilité de $\hat{\pi}_\omega$.

Plusieurs techniques d'estimation ont été étudiées pour le choix des pondérations dans l'expression précédente. A titre d'exemple, on peut citer la technique dite *de vote* et la technique de régression emboîtée faisant appel à l'emploi de la validation croisée (cf. El Faouzi, 1997). A cela, on peut ajouter d'autres approches qui se fondent sur l'emploi d'heuristiques dont l'intérêt pratique se heurte à l'absence de justification théorique permettant d'évaluer leur degré de légitimité.

Si on suppose que les estimateurs, $\varphi_i, i = 1, \dots, k$ sont sans biais¹², une propriété désirable de l'estimateur issu de la fusion est l'efficacité, *i.e.* $\hat{\pi}_\omega$ de variance minimale, ce qui garantit que la distribution statistique de $\hat{\pi}_\omega$ est fortement concentrée sur sa cible (le temps de parcours réel).

Il s'en suit alors que si l'on recherche un estimateur efficace du temps de parcours, exprimé sous la forme (40), le problème de recherche de pondérations optimales se réduit à un problème d'optimisation où l'on chercherait à trouver les poids qui minimisent la variance de $\hat{\pi}_\omega$.

¹² Rappelons qu'un estimateur est dit sans biais lorsque la moyenne (arithmétique) de ses valeurs est égale au paramètre cible inconnu.

$$\begin{cases} \min_{\omega_i} \text{var}(\hat{\pi}_\omega) \\ \text{avec} \quad \sum_{i=1}^k \omega_i = 1 \end{cases}$$

Dans le cas de l'indépendance des estimateurs $\varphi_i, i=1, \dots, k$ et de variances connues $\sigma_i^2, i=1, \dots, k$, on montre aisément que la variance de $\hat{\pi}_\omega$, donnée par :

$$\text{var}(\hat{\pi}_\omega) = \sum_{i=1}^k \omega_i^2 \sigma_i^2 \quad \text{avec} \quad \sum_{i=1}^k \omega_i = 1 \quad (41)$$

admet pour valeur minimale $1 / \sum_{i=1}^k \sigma_i^{-2}$. Cette valeur est atteinte pour $\omega_i^* = \frac{1}{\sigma_i^2}, i=1, \dots, k$, ce qui donne l'expression suivante :

$$\hat{\pi}_* = \frac{\sum_{i=1}^k \sigma_i^{-2} \varphi_i}{\sum_{i=1}^k \sigma_i^{-2}} \quad (42)$$

Si l'on dispose de deux estimateurs ($k = 2$), l'expression précédente devient :

$$\hat{\pi}_* = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \varphi_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \varphi_2 \quad (43)$$

Cette équation montre que la règle de fusion statistique affecte un poids à chaque estimateur et que ce poids est d'autant plus grand que la fiabilité de l'estimateur (quantifiée ici par l'inverse de la variance de ce dernier) est grande.

C'est cette dernière équation qui est à la base d'une architecture de fusion d'algorithmes d'estimation du temps de parcours, développé et testé dans le cadre du projet américain ADVANCE (cf. Tarko et Rouphail, 1993).

Notons que la fusion statistique dans le cas simple d'estimateurs indépendants et de variances connues, s'apparente à une fusion probabiliste reposant sur l'emploi du théorème de Bayes (cf. El Faouzi, 2000).

La mise en œuvre de cette famille de techniques sur des données réelles (cf. El Faouzi, 1997 ; 1999) a montré sa propension à améliorer la qualité de l'estimateur résultant, en réduisant sa variance et que la réduction de la variance d'estimation est d'autant plus forte que les sources considérées sont complémentaires.

Signalons enfin, qu'au travers de plusieurs simulations, la variance de l'estimateur résultant est toujours inférieure à celle du meilleur (au sens de l'erreur de prévision) estimateur. Le lecteur intéressé par cette approche pourra trouver dans El Faouzi, 1997 et 1999 d'autres schémas d'estimation des poids (ω_i) dans le cas plus général d'estimateurs quelconques et de variances inconnues.

Pour conclure, notons que cette première approche de fusion pour l'estimation du temps de parcours présente plusieurs avantages parmi lesquels on peut citer le fait que, d'une part, elle repose sur des théories statistiques et probabilistes bien établies garantissant des propriétés d'optimalité de l'estimateur résultant et que d'autre part, elle rend possible la prise en compte de la qualité des estimateurs à fusionner au travers de leurs variances. Cependant, elle ne permet pas la prise en compte de la qualité des données utilisées pour élaborer l'estimation et nécessite par ailleurs des échantillons de même taille.

Ce dernier point se traduit dans la pratique par le sacrifice une part de données issues de l'échantillon de grande taille pour pouvoir caler celle-ci sur la taille minimale.

Afin de pallier ces limitations, un autre cadre méthodologique a été étudié. Ce cadre repose sur l'utilisation de la théorie des crédibilités, théorie qui permet la gestion à la fois de l'incertitude, comme la théorie des probabilités permet de le faire, mais aussi les imprécisions inhérentes aux données manipulées.

2.2. FUSION CREDIBILISTE DES TEMPS DE PARCOURS

La théorie des crédibilités est basée sur les travaux de Dempster (1967, 1968), a été formalisée par Shafer en 1976, et présentée comme une généralisation de la théorie de Bayes au traitement de l'ignorance. Une première étude a été réalisée sur la mise en œuvre de cette théorie pour l'estimation du temps de parcours (El Faouzi et Simon, 2000).

La principale difficulté de la fusion de données dans ce cadre est de pouvoir modéliser et évaluer la connaissance apportée par chaque source, afin de tirer le meilleur parti des informations disponibles. Cette opération consiste à exprimer l'imperfection de chaque information en fonction de la source dont elle est issue. En fusionnant des données de sources hétérogènes, nous souhaitons obtenir une information *améliorée* sur l'état d'un système, c'est-à-dire moins imparfaite que chacune des informations issues des sources multiples. Cette notion d'imperfection s'exprime sous la forme de l'incomplétude des données (données non observables, données manquantes...) et de deux notions antagonistes, à savoir l'incertitude et l'imprécision¹³, puisque l'on ne peut en privilégier une qu'au détriment de l'autre (cf. El Faouzi, 2000).

La théorie des crédibilités utilisée dans ce cadre ne propose pas de «méthode» pour modéliser ces connaissances. Or cette étape s'avère déterminante au bon déroulement du processus de fusion et conditionne les performances du système de fusion. Nous avons donc été amenés à proposer un cadre méthodologique général de modélisation des informations issues de chaque source de données. Il s'agit plus précisément d'associer une mesure de crédibilité qualifiant le degré de vraisemblance d'une valeur ou de classe de valeurs du temps de parcours compte tenu des conditions de circulation. Cette étape de modélisation sera détaillée ci-dessous.

Préalablement à ce travail, et pour des raisons de mise en œuvre pratique de la technique des crédibilités, nous avons été amenés à découper en classes les valeurs de temps de parcours. La définition des hypothèses est fondée sur le découpage en r classes $\{\tau_1, \tau_2, \dots, \tau_r\}$ des temps de parcours de référence. Les bornes des classes $\{\tau_1, \tau_2, \dots, \tau_r\}$ sont définies à partir des quantiles des temps de

¹³ Une donnée incertaine fournit une information dont on n'est pas sûr qu'elle soit vraie, alors qu'une donnée imprécise fournit une information vague. Ainsi, une donnée incertaine peut être précise et inversement.

parcours de référence. Ainsi, chaque classe contient le même nombre d'observations, mais l'amplitude de celles-ci varie d'une classe à l'autre.

Dans le souci de tester les performances de la fusion en fonction du nombre de classes, nous avons fait varier ce nombre en considérant respectivement quatre, six, huit et dix classes, c'est-à-dire en augmentant la résolution de la partition et de même le niveau de précision de l'estimation.

Pour plus de détails sur les concepts de fusion et la théorie des crédibilités, le lecteur intéressé pourra se reporter à El Faouzi, 2000.

2.2.1. Recalage et homogénéisation des informations

Selon le degré d'hétérogénéité des données à fusionner, on peut être amené à rechercher un espace de représentation commun, dans lequel les différentes informations pertinentes disponibles renseignent sur une même entité (principe de similarité et de cohérence). Un premier traitement s'attache donc à transformer certaines informations initiales en informations équivalentes dans l'espace commun où s'effectuera la fusion (opération de recalage). Cette phase consiste donc à opérer un homomorphisme des données à la fois dans le temps et dans l'espace.

Dans le cadre de cette étude, deux sources ont été utilisées. Une première source consiste en un recueil exhaustif des données de trafic sur l'itinéraire étudié. La seconde source, quant à elle, repose sur une mesure directe du temps de parcours en utilisant des véhicules tests.

Afin de satisfaire le principe de cohérence énoncé ci-dessus, la conversion des données de trafic issues de la première source en temps de parcours est alors nécessaire. Pour ce faire, nous avons utilisé l'indicateur BRP (cf. § 4.4.3., partie I.) comme algorithme de conversion des données de trafic (principalement le débit et taux d'occupation) mesurées sur l'itinéraire considéré. Cependant, à l'issue de cette opération, le problème de recalage temporel des informations délivrées par les deux sources n'est pas complètement résolu. En effet, le temps de parcours estimé par l'indicateur BRP est un temps de parcours *instantané*, alors que les temps de parcours fournis par les véhicules tests et par l'enquête minéralogique sont des temps *effectivement réalisés* par des véhicules pour parcourir l'itinéraire considéré. Se pose alors la question d'homogénéisation des définitions de ces deux temps de parcours.

Comme il est toujours possible de ramener un temps de parcours réalisé à un temps de parcours instantané en calculant des moyennes temporelles sur les temps de parcours réalisés pendant des intervalles de temps d'amplitude commune Δt , l'étape de recalage est réalisée de la façon suivante : à chaque estimation de temps de parcours par l'algorithme de conversion, correspondant à une période temporelle donnée (multiple de la période d'agrégation des données de trafic), on lui fait correspondre la valeur¹⁴ du temps de parcours réalisé par un véhicule test entré sur l'itinéraire durant cette période. Le temps de parcours de référence pendant cette période est calculé comme la moyenne des temps de parcours de référence réalisés par des véhicules entrés sur l'itinéraire.

¹⁴ Il s'agit, à proprement parler, d'une moyenne sur un nombre d'observations. Mais, vu la faiblesse de la taille de l'échantillon des temps de parcours des véhicules tests, ce nombre est rarement supérieur à l'unité et il arrive même que, pour certaines périodes, aucun véhicule test ne soit entré sur l'itinéraire et que donc la mesure du temps de parcours soit inexistante.

2.2.2. Modélisation des connaissances

Concernant la modélisation des connaissances fournies par les sources multiples, la théorie des crédibilités représente la connaissance fournie par chacune des sources par un jeu de masse de croyances. Ces jeux de masses associés à chaque classe de valeurs du temps de parcours sont ensuite combinés, à l'aide d'un opérateur propre à cette théorie appelé règle de sommation orthogonale. A l'issue de cette opération de combinaison de masses de croyances, on obtient des masses fusionnées synthétisant la connaissance apportée par les différentes sources.

La modélisation de la connaissance par l'intermédiaire de jeux de masses doit être élaborée sur la base des connaissances *a priori* disponibles sur chaque source. L'objectif de cette étape du processus de fusion est en effet d'exprimer la confiance que l'on accorde à une information fournie par une source et de la quantifier sous forme de masse de croyance.

L'approche adoptée dans ce travail est celle de l'apprentissage statistique. Dans notre cas, la confrontation des classes fournies par chaque source de données aux classes de références obtenues à partir des temps de parcours de référence permet de quantifier les performances de chacune des sources. On obtient ainsi, pour chaque source, un tableau de reclassement, appelé aussi *matrice de confusion*, dont les éléments diagonaux sont les pourcentages de concordances entre les classes reconstituées par la source d'information c_{est} et les classes de référence $c_{réel}$. En d'autres termes, ceci représente le nombre de fois où la classe de référence et la classe reconstituée par la source coïncident.

Les éléments hors diagonale donnent quant à eux les pourcentages de discordances entre c_{est} et $c_{réel}$. La trace de cette matrice (somme des éléments diagonaux), rapportée au nombre total d'observations, fournit le taux de bien classés (TBC), critère utilisé pour la qualification des performances de la fusion.

Les éléments de la matrice de confusion sont ensuite normalisés de telle sorte que, pour chaque colonne, la somme des éléments soit égale à l'unité. C'est la matrice de confusion ainsi normalisée qui définit les masses de croyances.

De cette connaissance des masses de croyances de chacune des sources et via le mécanisme de fusion propre à la théorie de l'évidence, on calcule par fusion les masses résultantes, que l'on affecte aux classes $\{\tau_1, \tau_2 \dots \tau_r\}$. Ainsi, comme l'illustre la figure 3, si la source 1 (resp. la source 2) fournit comme hypothèse H_i (resp. H_j), pour chaque source, on affecte aux classes de $\{\tau_1, \tau_2 \dots \tau_r\}$ les masses contenues dans la colonne i (resp. la colonne j) puis, par la règle de combinaison orthogonale, on détermine les masses fusionnées que l'on associe aux classes $\{\tau_1, \tau_2 \dots \tau_r\}$.

Le critère de décision qui s'impose dans ce cas consiste à retenir comme classe probable celle dont la masse de croyance fusionnée est *maximale*.

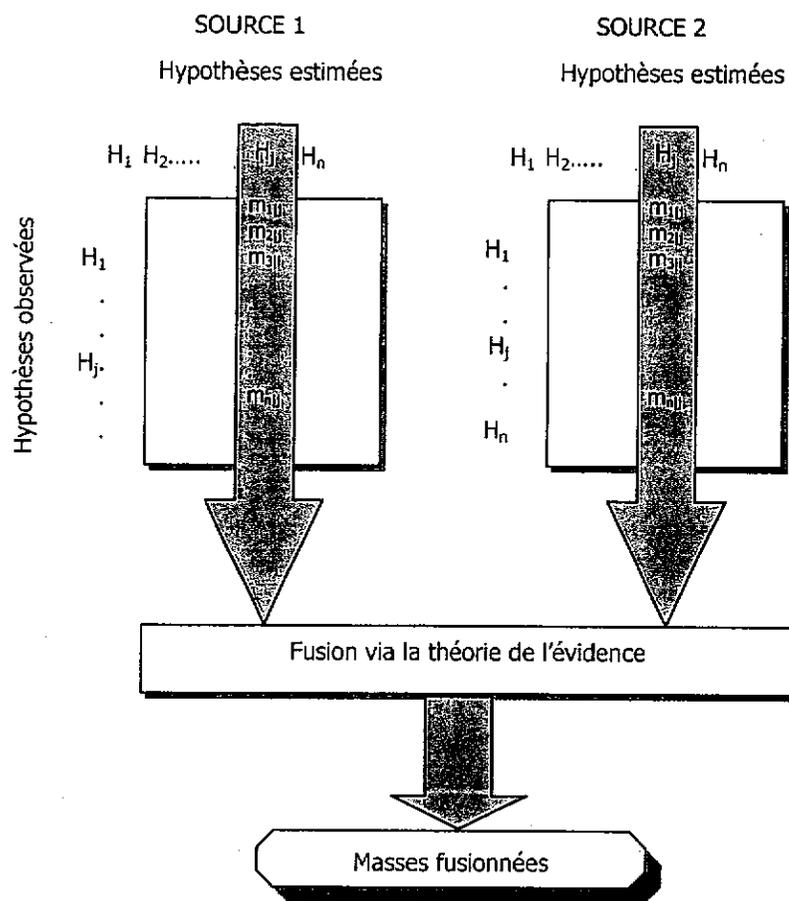


Figure II-2 : Calcul des masses de croyances et principe de la fusion

2.2.3. Évaluation des performances

La fusion crédibiliste a été utilisée sur des données opérationnelles afin d'en estimer les performances. Les données utilisées dans cette étude ont été recueillies à l'issue de la campagne de mesures réalisée à la ZELT, Toulouse en 1994 (cf. Olivero, 1994). L'enquête de mesures a consisté en un recueil de données issues de plusieurs sources :

- ❑ **Source 1** : Données macroscopiques de trafic fournies par le système ZELT sur un itinéraire urbain ;
- ❑ **Source 2** : Données sur la cinématique des quatre véhicules équipés de capteurs embarqués ont été recueillies. Un relevé préalable précis des distances sur l'itinéraire a permis en outre de disposer des temps de parcours entre les divers repères utilisés.
- ❑ **Source 3** : Données sur les temps de parcours recueillies sur le principe d'enquête minéralogique entre les entrées-sorties de l'itinéraire étudié.

Les deux premières sources, que l'on notera «boucles» et «traceurs» respectivement, ont été utilisées pour fournir des informations sur le temps de parcours à fusionner, selon la démarche exposée dans le paragraphe § 2.2.1. précédent. Les temps de parcours fournis par la dernière source constituent la référence pour l'évaluation.

Compte tenu des sorties de cette technique (classes de valeurs de temps de parcours), le critère utilisé pour la qualification de ses performances est le taux de bien classés (TBC), critère communément utilisé dans le cas de la classification. Ce critère est défini comme le rapport du nombre de fois où la classe des temps de parcours résultant de la fusion coïncide avec la classe de référence, sur le nombre total d'informations à fusionner.

Pour chacune des sources, ce taux de bien classés (exprimés en pourcentage) a été calculé, et l'apport de la fusion est quantifié par le critère d'écart absolu ΔTBC défini par :

$$\Delta TBC = TBC(\text{fusion}) - \max[TBC(\text{boucles}), TBC(\text{traceurs})]$$

Ces différents taux ont été calculés à partir d'un échantillon de validation (échantillon n'ayant pas fait partie des données d'apprentissage) obtenu par tirage aléatoire. Trois échantillons ont été utilisés correspondant respectivement à 10 %, 20 % et 30 % des données disponibles.

Pour chaque échantillonnage de validation, nous avons appliqué le processus de fusion en faisant varier le nombre de classes (4, 6, 8 et 10 classes). Les résultats de cette évaluation sont synthétisés dans le tableau ci-dessous.

Taille de l'échantillon de validation Base =230	Méthode d'estimation ou de mesure	Taux de bien classés (TBC) en %			
		4 classes	6 classes	8 classes	10 classes
10 %	Boucles	39,3	28,3	21,7	17,9
	Traceurs	37,5	27,7	22,4	16,8
	Fusion	47,8	35,0	35,2	31,5
	ΔTBC	8,5	6,7	12,8	13,6
20 %	Boucles	39,7	28,8	21,4	18,2
	Traceurs	38,3	27,9	22,4	16,6
	Fusion	46,9	35,4	33,6	31,4
	ΔTBC	7,2	6,6	11,2	13,2
30 %	Boucles	39,8	29,0	21,5	18,5
	Traceurs	37,6	26,5	22,0	15,7
	Fusion	47,6	35,1	33,9	31,0
	ΔTBC	7,8	6,1	11,9	12,5

Tableau II -1 : Performances de la méthodologie de fusion avec concordances strictes

Remarquons tout d'abord la faible qualité de reconstitution du temps de parcours (TBC < 40 %) aussi bien par la méthode BRP que par les mesures fournies par un échantillon de véhicules tests. Dans le cas du BRP, ces performances médiocres peuvent s'expliquer par le fait que cette technique, mis à part sa sensibilité aux paramètres de calibrage, repose sur une incohérence de définitions du temps de parcours utilisées dans la procédure d'identification des paramètres

structuraux de cet indicateur¹⁵ (cf. § 4.4.3., partie I.). Cette incohérence induit la faible reconstitution par cet indicateur des classes contenant de fortes valeurs du temps de parcours (classes extrêmes du temps de parcours).

En ce qui concerne le processus de fusion, il apparaît clairement que la méthode de fusion proposée améliore, dans chacun des cas, la qualité de l'estimation des temps de parcours. L'amélioration de la qualité de l'estimation, en termes de taux de bien classés, varie de 6 % à 13,6 % et celle-ci est d'autant plus importante que le nombre de classes augmente ($\Delta TBC > 12\%$), c'est-à-dire quand le niveau de précision exigée de l'estimation augmente, même si globalement, le taux de bien classés diminue. En revanche, compte tenu de la procédure d'échantillonnage utilisée, la variation de la taille de l'échantillon de validation a peu d'influence sur les différents TBC, même si l'on note une décroissance de l'amélioration avec l'augmentation de la taille de l'échantillon test.

Les performances précédentes ont été calculées sur la base d'une concordance exacte ou stricte entre les classes reconstituées et les classes de référence. Si l'on relaxe cette concordance exacte en s'autorisant la confusion entre les classes adjacentes (*i.e.*, une mesure sera considérée comme bien classée si sa classe restituée coïncide avec sa classe de référence ou si elle correspond aux deux classes encadrant cette dernière), on obtient les résultats du tableau 2.

Taille de l'échantillon de validation Base = 230	Méthode d'estimation ou de mesure	Taux de bien classés (TBC) en %			
		4 classes	6 classes	8 classes	10 classes
30 %	Boucles	64,5	55,0	41,6	29,1
	Traceurs	69,3	52,7	43,5	25,9
	Fusion	82,5	65,7	54,7	48,9
	ΔTBC	13,2	10,7	11,2	19,8

Tableau II - 2 : Performances de la méthodologie de fusion avec concordances larges.

On observe alors une nette amélioration des taux de bien classés après fusion, dont les valeurs nominales oscillent entre 50 % et un peu plus de 82 %. On retrouve la même tendance que celle constatée dans le cas précédent, à savoir que l'amélioration augmente avec le nombre de classes. Cette amélioration ($\Delta TBC > 10\%$) trouve son explication dans le fait que la plupart des erreurs de classement proviennent de la confusion des classes adjacentes. Ceci souligne une certaine sensibilité des performances de la méthode de fusion étudiée aux situations peu ou pas discernables. En effet, l'utilisation exclusive des classes atomiques (simples) oblige à décider d'une classe et une seule du temps de parcours, en dernière phase du processus de fusion, même si la réalisation de cette hypothèse est peu sûre (masse faible), ou même si elle n'est pas beaucoup plus sûre que celle d'un autre (masses de deux ou plusieurs hypothèses très proches).

¹⁵ En effet, la procédure d'identification utilise le temps de parcours libre en situation fluide et le temps passé en situation congestionnée (cf. Bonvalet et Robin-Prevalée, 1987).

3. CONCLUSIONS

L'objectif de ce travail était de démontrer l'opportunité d'une estimation multi-sources du temps de parcours, en proposant un cadre méthodologique à l'estimation de temps de parcours en présence de données de sources hétérogènes. Ce problème a été successivement abordé comme un problème de fusion d'estimateurs et comme celui relevant de fusion de données.

Ces résultats de la fusion d'estimateurs montrent une amélioration, bien que modeste, de la qualité de l'estimation. Devant ces résultats encourageants, des raffinements ont été apportés aux solutions de la fusion d'estimateurs. Ainsi, plusieurs schémas de fusion, essentiellement statistiques, ont été développés et ont montré leur propension à améliorer la qualité de l'estimation en réduisant l'erreur de prévision. De plus, sur plusieurs exemples, la robustesse des méthodes proposées a été mise en évidence, en ce sens que, la qualité de l'estimateur résultant de la fusion est toujours supérieure à celle du « meilleur » estimateur individuel (cf. El Faouzi, 1999).

Nous avons choisi d'étudier également l'approche issue de la théorie des crédibilités qui elle, nous permet une fusion de données proprement dite. La théorie des crédibilités, considérée comme une extension de la théorie bayésienne, a été choisie pour sa souplesse et sa finesse de modélisation de la connaissance. Le formalisme développé a été appliqué aux données d'une expérimentation. Il résulte de cette application que les performances de la fusion sont variables, mais que cette dernière améliore clairement la qualité de l'estimation, en ce sens qu'elle détecte mieux les bonnes classes du temps de parcours que la méthode du BRP (fondée sur les données des capteurs à boucle) ou encore les mesures obtenues par les véhicules tests.

Néanmoins, la méthode développée ici n'utilise pas la théorie des crédibilités de façon optimale : nous n'avons en effet affecté des masses qu'à des hypothèses simples, ce qui oblige à décider d'un état précis du trafic (une seule hypothèse) en dernière phase du processus de fusion, même si la réalisation de cet état n'est pas du tout sûre (masse faible), ou même si elle n'est pas beaucoup plus sûre que celle d'un autre état (masses de deux ou plusieurs hypothèses très proches). Certes, en faisant varier la taille des intervalles de temps de parcours constituant les différentes hypothèses, il est possible d'introduire une notion de précision. Mais, comme le montre le tableau II-2., il serait plus intéressant de prendre en compte des unions d'intervalles (hypothèses composées) lors de l'affectation des masses et de la prise de décision finale, afin notamment de pouvoir modéliser le niveau de l'imprécision de l'information issue de la fusion, et surtout de gérer les situations peu ou pas discernables. Cette direction fait l'objet de recherches en cours.

Il convient également de souligner que l'utilisation de la théorie des crédibilités pour traiter un problème de fusion de données exige une bonne connaissance du système étudié, car elle ne propose de méthodologie, ni pour constituer les hypothèses, ni pour modéliser les jeux de masses. Or ces deux étapes s'avèrent déterminantes pour le bon déroulement du processus de fusion. Ceci vient de son caractère très souple qui, paradoxalement, constitue en même temps un grand avantage : tout problème de gestion d'informations issues de sources hétérogènes peut être abordé par la théorie des crédibilités. Or, ce type de problème est aujourd'hui rencontré dans de nombreux domaines, et il serait sans doute intéressant de se pencher davantage sur l'élaboration d'un cadre méthodologique général de modélisation des jeux de masses.

Notons enfin que pour la mise en œuvre opérationnelle d'un système de fusion de sources multiples pour l'estimation du temps de parcours, la démarche consiste en un processus hiérarchique, découpé en trois phases, que l'on peut résumer de la manière suivante :

1. **Sources d'information** : Il s'agit tout d'abord, de recenser toute l'information *a priori* disponible :

- au niveau des mesures du trafic : par exemple, les performances et précisions des moyens de mesures, disponibilités et fiabilités des mesures,....
- au niveau de l'apprentissage : existence de mesures de référence, de matrices de confusion associées à chaque méthode d'estimation ou de mesure, nuées d'exemples,...
- et éventuellement d'autres informations exogènes plus ou moins subjectives telle que l'expertise sur tel ou tel moyen de mesure et/ou d'estimation, du contexte de l'étude,....

Une fois ce recensement effectué, on doit dresser un bilan concernant les imperfections qui touchent les données provenant de chacune des sources et la façon dont la fiabilité de chaque source doit être appréhendée et quantifiée.

2. **Modélisation et traitements** : préalablement à la modélisation proprement dite, deux traitements sont à effectuer sur les données.

- Le premier traitement correspondant au problème de l'*association* des données fournies par les diverses sources, aussi bien dans le temps que dans l'espace. Dans le cas étudié ici (données trafic et véhicules traceurs) il s'agit d'associer à une mesure de temps de parcours expérimenté par un véhicule traceur sur un tronçon, un vecteur de valeurs prises par les variables macroscopiques du trafic quantifiant les conditions de circulation qu'a rencontré ce même véhicule sur le tronçon étudié. A l'issue de cette étape les données fournies par les diverses sources portent sur la même extension géographique et que la correspondance temporelle entre les différentes mesures est effectuée. On parle dans ce cas de recalage temporel et spatial des données.
- Le second traitement s'attache à ramener les données préalablement recalées dans un espace commun. L'objectif étant de réduire l'hétérogénéité des sources par des conversions de types des données. Ainsi, dans le cas de notre application, les données de trafic (débit, taux d'occupation et éventuellement vitesse) ont été converties en données de temps de parcours via un algorithme adéquat (cf. § 2.2.1, partie II.).

Une fois ce travail préalable terminé, vient alors l'étape de modélisation des informations apportées par chacune des sources puis l'étape de fusion. Ces deux étapes fonction du cadre méthodologique choisi. A ce niveau, il existe plusieurs cadres potentiellement utilisables dont le choix se fera en tenant compte de la nature des imperfections dominantes des sources et de la nature des informations *a priori* disponibles. On peut citer les trois principaux cadres de modélisation permettant la gestion de données multi-sources : le cadre probabiliste, le cadre crédibiliste et le cadre possibiliste (cf. El Faouzi, 2000).

L'ensemble de ces traitements et modélisation doit concourir à l'élaboration d'une information exploitant au mieux les complémentarités et les redondances des sources disponibles.

3. **Evaluation** : Cette évaluation doit porter aussi bien sur les traitements que sur le système lui-même. C'est lors de cette étape que l'on cherche généralement à approcher le compromis performances / capacité et des temps de calcul.

Cette dernière phase constitue l'étape terminale importante dans le développement de tout système et doit être perçue non comme une contrainte mais plutôt comme un maillon important du système qualité et comme un outil puissant qui permettra de mieux définir ses futures améliorations.

PARTIE III.

ÉVALUATION DES SYSTÈMES DE MESURE DES TEMPS DE PARCOURS

1. INTRODUCTION

L'évaluation¹⁶ des systèmes de mesure ou d'estimation des temps de parcours est bien souvent limitée à l'évaluation de la précision de la mesure ou de l'estimation. Encore cette précision est-elle souvent définie de façon vague et sujette à diverses interprétations. Il apparaît en fait clairement que la notion d'évaluation est indissociable d'une notion d'exigences, comme la notion de recette est indissociable de celle de cahier des charges. C'est ainsi que l'évaluation d'un système ne peut se faire qu'en référence à un certain nombre d'exigences, dont l'évaluation a précisément pour objectif de mesurer le degré de satisfaction. La précision d'une estimation dans des conditions données constitue bien entendu l'une des exigences fondamentales d'un système de mesure, mais cette exigence formulée isolément de son contexte (en quels points du réseau ?, dans quelles conditions de trafic ? avec quelle disponibilité du système ? etc...) n'a en fait guère de sens. Il est clair *a contrario* que chaque formulation d'une exigence doit s'accompagner d'une méthodologie d'évaluation de la conformité du système à cette exigence, sous peine de rendre celle-ci inopérante.

Il est enfin évident que les possibilités d'évaluation sont différentes selon qu'on s'intéresse à un prototype, grâce auquel on cherche essentiellement à valider un principe ou une technique de mesure ou d'estimation, à une installation pilote de dimension réduite ou à un système opérationnel.

¹⁶ Pour préciser le sens de ce vocable, notons que deux processus concourent principalement à la vérification de la validité du système. Il s'agit de *validation* et d'*évaluation*. Ces deux processus sont souvent confondus car ils ne correspondent pas à des phases très différenciées dans la pratique. La validation désigne le processus de vérification que le système mis en œuvre obtient les performances prévues, souvent basé sur les résultats d'évaluation. Autrement dit, cette étape permet de s'assurer que la méthode ou le système mis au point répond bien aux objectifs assignés. L'évaluation, quant à elle, se réfère au processus de détermination ou de qualification des performances en comparaison avec une situation de référence et incluant habituellement un processus expérimental. Ce sont là, les définitions retenues dans le cadre du programme CONVERGE du 5^{ème} PCRD consacré aux méthodologies d'évaluation des projets télématiques dans le domaine des transports.

Ce chapitre présente donc dans un premier temps un ensemble d'exigences qu'il est possible de formuler à l'égard d'un système de mesure ou d'estimation des temps de parcours. Dans un deuxième temps il donne quelques pistes de méthodologie pour évaluer la conformité d'un système à ces exigences.

2. EXIGENCES ATTACHÉES À LA MESURE OU À L'ESTIMATION DES TEMPS DE PARCOURS

La formulation précise des exigences du maître d'ouvrage pour la définition d'un système de mesure ou d'estimation des temps de parcours est importante de plusieurs points de vue.

- C'est tout d'abord la base du cahier des charges du système, qui permettra sa conception et surtout sa recette. Il est important à cet égard de noter que derrière chaque exigence on doit trouver une procédure de vérification de conformité, sans laquelle l'exigence reste un simple souhait.
- C'est ensuite l'occasion de prendre explicitement les décisions nécessaires de compromis entre des exigences qui pourraient s'avérer contradictoires.
- Ce peut être enfin la base de la négociation préliminaire à une forme quelconque de contractualisation dans la fourniture de données par le maître d'ouvrage à un opérateur de service ou une agence de presse. Les exigences du maître d'ouvrage sont alors celles qui lui permettront de répondre aux exigences de l'opérateur en qualifiant les données qu'il lui fournit.

Il est ainsi possible de définir un ensemble de classes d'exigences, dans lesquelles on pourra sélectionner un ensemble cohérent qui caractériseront un système de mesure ou d'estimation.

2.2. COUVERTURE SPATIALE - CLASSIFICATION DES ARCS

La couverture spatiale du système de mesure des temps de parcours peut s'envisager de plusieurs manières :

- Elle peut se limiter à certains points critiques du réseau tels que grands échangeurs ou carrefours particulièrement chargés ;
- Elle peut concerner un certain nombre de niveaux du réseau dans la mesure où celui-ci est bien hiérarchisé. Le choix du réseau concerné est alors à son tour un élément de renforcement de cette hiérarchisation ;
- Il faut par ailleurs bien distinguer la couverture spatiale du réseau de mesure de l'organisation spatiale d'un système d'information des usagers. On peut par exemple imaginer que, pour les besoins propres du maître d'ouvrage, l'ensemble du réseau dont il est responsable soit couvert, mais qu'une information diffusée ne concerne que les niveaux supérieurs du réseau.

L'exigence de couverture peut enfin ne pas être uniforme, une couverture plus légère pouvant être suffisante sur certaines zones ou axes. Il est alors clair qu'une cohérence entre couverture spatiale et exigences de performances doit être assurée, un tronçon considéré comme couvert étant celui où les autres exigences seront respectées.

Elle se traduit enfin par des exigences de continuité d'itinéraires, évitant ou traitant spécifiquement les discontinuités (par exemple les transitions entre réseau de surface et réseau de voies rapides).

2.3. FINESSE SPATIALE

La définition d'un temps de parcours peut se concevoir à diverses échelles, allant du tronçon élémentaire (entre deux carrefours) à de grands itinéraires traversant l'ensemble du réseau considéré. Quelque soit l'échelle concernée, il est essentiel de réfléchir aux possibilités de combinaison (exigence d'*additivité*) des mesures élémentaires pour constituer des indicateurs plus globaux. Ceci constitue une exigence très forte, dans la mesure où les grands nœuds du réseau sont précisément les lieux où les temps de parcours sont les plus difficiles à mesurer ou estimer. De même, les questions de raccordement entre réseaux de natures différentes (les bretelles d'accès aux voies rapides par exemple) sont particulièrement difficiles à traiter.

En ce qui concerne la finesse spatiale elle-même, il est souvent difficile de se passer d'une connaissance au niveau du chaînon élémentaire. Ceci peut amener, en fonction des techniques de mesure ou d'estimation retenues, à moduler les exigences de précision en fonction de l'échelle géographique. On sait par exemple que l'estimation du temps de parcours à partir des données de boucles de détection n'est utilisable que sur des itinéraires comprenant un nombre suffisant de tronçons, l'incertitude attachée à la mesure sur chaque tronçon étant trop grande mais les erreurs se compensant par agrégation spatiale. On peut ainsi imaginer de formuler des exigences de connaissance des valeurs élémentaires sur chaque tronçon (ceci afin de permettre des agrégations quelconques), mais de ne formuler d'exigences de précision que sur des ensembles plus vastes (par exemple sur toute agrégation de 5 tronçons ou plus).

2.4. PRECISION

Cette exigence est difficile à formuler, car elle implique à la fois la nécessité d'une *référence* de comparaison (précision par rapport à quoi ?) et d'une procédure de mesure de la précision, l'exigence n'ayant de sens que si son respect peut être vérifié.

Compte tenu de ce qui a été dit plus haut sur la significativité d'un temps de parcours moyen, il est nécessaire de classer les arcs du réseau en trois classes:

- Les arcs où toute estimation de temps de parcours semble vouée à l'échec. Ces arcs feront l'objet d'un traitement particulier et ne seront pas concernés par l'estimation des temps de parcours. Ce sont ceux qui comprennent de nombreux carrefours sans feux présentant des mouvements tournant importants, où qui traversent des zones soumises à des perturbations importantes (stationnement sauvage...).

- Des arcs douteux, où l'estimation semble possible mais pour lesquels il n'est pas possible de prendre d'engagement en termes de performances.
- Les arcs normalement utilisables, où les critères de performance devront être respectés.

La classification des arcs dans l'une des deux premières classes doit être justifiée par la présence d'éléments perturbateurs en nombre et en importance suffisants.

Pour permettre l'évaluation des performances, un certain nombre (typiquement trois à cinq) de classes de trafic homogène doivent être déterminées pour chaque arc ou globalement par portion de réseau. Ces classes sont caractérisées soit par leur créneau horaire, soit par des caractéristiques de trafic (débits, fourchette de longueur de files d'attente ou de taux d'occupation sur les capteurs). Le découpage en classes doit permettre de trouver, à l'intérieur de chaque classe, une distribution des temps de parcours relativement concentrée autour de la moyenne. La validation des performances de l'estimation se fera séparément sur chaque classe.

On peut ensuite formuler l'exigence de précision de la façon suivante (les chiffres sont indicatifs et peuvent être quelconques) : sur chaque itinéraire et au sein de chaque classe de trafic, pendant 90% du temps, 90% des temps de parcours mesurés doivent être dans une fourchette de deux écarts-type autour du temps de parcours moyen réel mesuré. Ceci veut dire qu'en excluant au maximum 10% de temps d'indisponibilité ou de fonctionnement incorrect du système, en excluant ensuite 10% de valeurs estimées de temps parcours qui peuvent être aberrantes, chacune des valeurs restantes doit se trouver dans la fourchette.

Dans le cas où il n'est pas possible de formuler de façon aussi précise des exigences quantitatives de précision, on sera amené à formuler des exigences qualitatives, qui renvoient davantage à des notions de fidélité.

On peut faire plusieurs remarques sur cette formulation d'exigence :

- Elle est à la fois plus précise et plus claire que la formulation classique « *mesurer les temps de parcours à 20% près* »
- Elle a le mérite de proportionner l'exigence à la difficulté ainsi qu'à la perception de l'utilisateur puisque plus les temps de parcours seront dispersés, plus l'écart-type sera grand et l'exigence de précision faible.
- Le fait de demander que l'exigence de performance soit tenue 90% du temps constitue une exigence de disponibilité du système plus précise que l'exigence d'une tolérance à un certain pourcentage de stations de mesure en panne. Dans ce dernier cas en effet, selon la répartition des pannes, l'exigence peut être soit évidente soit impossible à tenir. La réponse à une exigence formulée en termes de disponibilité dans le temps se traduit en redondance du système mais aussi en engagement de maintenance.

2.5. FIDELITE

La fidélité est une exigence fondamentale en matière de mesure. Elle exprime le fait que deux mesures identiques correspondent bien à deux situations identiques sur le terrain. Elle est distincte de l'exigence de précision qui caractérise l'identité de la mesure à une mesure de référence. En d'autres termes, au titre de cette exigence, les mesures de temps de parcours peuvent n'être pas correctes en valeur absolue, mais doivent permettre la comparaison entre des situations ou le suivi des évolutions.

Il peut y avoir nécessité d'un compromis entre précision et fidélité, une moindre exigence de précision pouvant permettre de garantir la fidélité de la mesure : une information qualitative peut ainsi être plus fidèle qu'une information quantitative.

2.6. COUVERTURE TEMPORELLE

Les caractéristiques temporelles de l'information renvoient à plusieurs aspects. Il s'agit tout d'abord de définir si l'information doit être disponible tout au long de la journée, de la semaine etc. S'agissant de systèmes de mesure automatisés, cette question est de peu d'intérêt, cette exigence étant peu contraignante.

Il s'agit ensuite de définir le fonctionnement du système dans les différentes configurations de trafic rencontrées dans le temps : heures creuses, pointes récurrentes, événements particuliers prévisibles, situations de crise, etc. Ceci peut amener à moduler les exigences de précision en fonction des situations. C'est ainsi que l'affichage des temps de parcours sur le boulevard périphérique est abandonné en situation trop fluide, car les temps de parcours sont alors trop dispersés (et présentent au demeurant peu d'intérêt), ainsi qu'en situation trop congestionnée, où le temps de parcours estimé devient grand devant l'échelle de temps des variations du trafic (l'exemple extrême est éclairant : Il n'a pas de sens d'afficher un temps de parcours infini lorsque le trafic est bloqué, car c'est alors l'évolution des conditions de trafic qui va faire que le trafic finira par redémarrer).

On peut ainsi aboutir non pas à une simple liste d'exigences, mais bien à un tableau à plusieurs dimensions définissant des exigences différentes pour chaque situation de trafic, chaque type de voie etc. (rien n'empêchant au demeurant une formulation simplifiée uniformisant ces exigences, à condition que toute simplification soit faite en connaissance de cause).

2.7. FRAICHEUR

La fraîcheur définit le temps qui sépare la mesure physique faite sur le terrain de l'utilisation effective de l'information (affichage, utilisation par l'exploitant). Cette exigence est moins simple à formuler qu'il ne pourrait y paraître, car elle renvoie à différentes caractéristiques du système d'acquisition :

- Les temps d'acheminement et de traitement de l'information ne présentent pas souvent, s'agissant là encore de systèmes automatisés, un caractère très contraignant ;
- Le choix d'une période d'agrégation des mesures et d'une périodicité de mise à jour est en revanche délicat car il conditionne à la fois la réactivité et la sensibilité du système, ce qui militerait pour des périodicités courtes, mais aussi sa robustesse et sa résistance aux perturbations, ce qui tend à faire allonger la période d'agrégation. Le compromis souvent trouvé qui consiste à utiliser une agrégation glissante (on met à jour toutes les cinq minutes une information calculée sur les quinze dernières minutes) présente souvent des inconvénients importants (manque de dynamique, masquage de certains phénomènes transitoires et persistance au contraire de certains autres...);
- Lorsque l'information est constituée à partir de mesures d'origines diverses et asynchrone (par exemple des mesures faites au sol et des temps de trajets fournis par des véhicules traceurs), il n'est parfois même plus possible de déterminer à quelle période précise se rapporte l'information calculée. L'exigence de fraîcheur est alors difficile à formuler.

L'exigence de fraîcheur ne se comprend enfin qu'en lien avec la durée de validité de l'information recueillie.

2.8. DISPONIBILITE

L'exigence de disponibilité correspond essentiellement au comportement du système en situation dégradée du point de vue des équipements de mesure. Les équipements de mesure du trafic (les détecteurs à boucle en particulier) et les équipements de transmission sont en effet fragiles, et il importe de définir comment le système doit réagir à leurs pannes.

L'exigence de disponibilité peut se formuler de diverses manières :

- On pourrait imaginer une exigence de disponibilité totale sans dégradation des performances. Ceci conduirait à une forte redondance des équipements de mesure et de transmissions. Il ne semble pas que cette vision des choses soit pertinente dans le contexte de l'estimation des temps de parcours.
- Il est en revanche plus réaliste de prévoir l'existence de modes dégradés bien identifiés permettant de garantir un fonctionnement minimal du système en cas de défaillance partielle. Ceci peut être fondé par exemple sur des valeurs par défaut (éventuellement définies par classes de situation de trafic), ou sur des modalités de reconstitution d'information manquante à partir de l'information résiduelle disponible. L'exigence de disponibilité se traduit alors là encore par une modulation des exigences de précision ou de fidélité en fonction de l'état matériel du système de mesure.

- On peut par ailleurs formuler des exigences de disponibilité du système de mesure lui-même, qui seront répercutées sur l'organisation de la maintenance des équipements.

2.9. ÉVOLUTIVITE

L'exigence d'évolutivité traduit la capacité du système à s'adapter à l'évolution de son environnement. Celle-ci concerne plusieurs points :

- L'évolution du réseau : ouverture de voies nouvelles éventuellement, mais surtout restructuration du réseau, modification de la hiérarchie des voies, déclassement ou reclassement de certaines, modification de la géométrie des carrefours. Le système doit s'adapter à ce type d'évolution par un simple reparamétrage (ainsi évidemment que par l'adaptation du système de mesure sur le terrain) ;
- L'évolution des techniques de mesure et de transmission doit être possible. Ceci implique une structuration suffisante du système en couches séparées, l'évolution des techniques de mesure ne devant affecter que les couches les plus basses du système.
- L'évolution des plates-formes de fonctionnement : il s'agit des exigences qu'on attache classiquement à tout système mettant en jeu des moyens informatiques.
- L'évolution des conditions de trafic doit non seulement être détectée mais également ne pas remettre en cause les performances du système. Ceci n'a rien de facile si l'on considère par exemple que la position optimale des points de mesure pour l'estimation des temps de parcours se situe aux extrémités des files d'attente : une évolution de la longueur de files d'attente dans un sens ou dans l'autre peut ainsi rendre un poste de mesure totalement inopérant ;

Il en va de même pour l'évolution du système de contrôle, qui peut déplacer ou supprimer des files d'attente, changer la nature de points critiques du réseau, et ainsi remettre en cause non seulement le paramétrage du système de mesure des temps de parcours mais aussi la structure même du système de recueil de données.

2.10. COHERENCE

L'utilisateur qui circule en voiture est amené à traverser les zones de compétences de maîtres d'ouvrage multiples. Il est clair que cet aspect des choses lui est indifférent et que son déplacement constitue un tout. À la rigueur fait-il la distinction entre la part du déplacement effectuée sur voies rapides et celle effectuée sur le réseau de surface.

Ceci implique une recherche de cohérence dans les informations qui lui sont fournies tout au long de son déplacement. Ceci ne veut pas dire que les systèmes qui ont élaboré cette information soient identiques, ni même nécessairement qu'ils fournissent la même information. Le premier niveau d'exigence est que les informations fournies par les uns et les autres ne soient pas contradictoires. A un niveau d'exigence plus élevé, on peut rechercher la complémentarité de

l'information, voire sa complétude, la totalité du réseau étant alors couverte sans discontinuité aux frontières de zones de compétences. Il est également nécessaire que les exigences de fidélité, de précision et de disponibilité soient définies de façon cohérente.

2.11. CONDITIONS DE MISE EN ŒUVRE

Les exigences concernant la mise en œuvre du système sont de plusieurs ordres. Il s'agit d'abord d'exigences sur sa réalisation matérielle (génie civil, intervention sur la voirie, réutilisation d'installations existantes etc.). Celles-ci ne sont guère différentes de ce qui doit être formulé pour un système de gestion des feux par exemple. Il s'agit ensuite, et de façon plus spécifique, de l'ingénierie nécessaire à la mise en œuvre du système, et singulièrement la nécessité de réglage, de paramétrage et de calibrage par des enquêtes.

Il faut ainsi se rappeler que le principal avantage mis en avant par les concepteurs de l'indicateur « BRP » était l'absence de nécessité de calibrage, l'ensemble des paramètres de l'indicateur étant mesurés directement sur plan (Ceci a évolué depuis, l'utilisation de l'indicateur pour des besoins d'information ayant entraîné l'ajout de coefficients qui nécessitent une procédure de calibrage).

Il est bien clair qu'un compromis est à trouver entre les exigences de précision et celles portant sur le calibrage (ce qui n'implique au demeurant pas nécessairement qu'un système très lourd à calibrer soit très performant). Ce qui importe également est de bien dimensionner l'effort nécessaire à ces opérations de mise en œuvre, effort qui peut être très important : organisation d'enquêtes, traitement de masses de données importantes, calculs statistiques.

2.12. COUTS

Les exigences de coût interviennent de façon constante pour limiter les autres exigences attachées au système. Elles sont elles aussi de diverses natures, correspondant aux divers postes de coût :

- ❑ Coût d'investissement pour la réalisation et la mise en œuvre du système
- ❑ Coût de maintenance et de fonctionnement
- ❑ Coûts d'évolution du système (en fonction des évolutions évoquées plus haut)

Par ailleurs certains coûts sont purement financiers (investissements), d'autres se traduisent uniquement en moyens humains (exploitation), d'autres sont éventuellement transformables par le biais d'une réalisation directe ou d'une sous-traitance.

3. INDICATEURS, CRITÈRES ET MOYENS D'ÉVALUATION

A chacun des types d'exigences évoqués plus haut concernant les systèmes de mesure du temps de parcours il est possible d'associer un certain nombre d'indicateurs ou de critères permettant

d'évaluer la conformité d'un système. On peut alors définir une méthodologie permettant de quantifier ces indicateurs (certains sont au demeurant de nature plutôt qualitative, mais on pourra alors tenter de définir des classes de conformité).

3.1. COUVERTURE SPATIALE

La couverture spatiale ne peut évidemment se caractériser que sur un système opérationnel, voire une installation pilote, même si dans certains cas une expérimentation peut permettre la mise en évidence de certaines difficultés (difficulté de couverture de certains types de voies, de carrefours...). Elle se caractérise, à contrario, par l'existence d'un certain nombre de « trous » ou de zones où les temps de parcours ne sont pas connus, ou ne répondent pas à d'autres exigences (précision, disponibilité). Ceci montre bien au passage les interdépendances qui existent entre les différents types d'exigences.

Si l'on peut ainsi considérer une couverture spatiale a priori, en considérant comme non couverte les zones où l'on renonce d'ores et déjà à connaître les temps de parcours, un retour est nécessaire en fin d'évaluation pour vérifier l'effectivité de sa disponibilité de l'information sur l'ensemble du réseau.

Il faut noter que la question de la couverture concerne les chaînons du réseau, mais surtout les nœuds, à travers lesquels la continuité des itinéraires est difficile à assurer. Ceci est particulièrement vrai des échangeurs entre réseau de surface et voies rapides, mais aussi de tous les carrefours compliqués (grands giratoires, carrefours avec une trémie centrale...).

La notion de couverture spatiale dépend enfin du type de technique de mesure utilisé. Si elle est relativement sans mystère lorsqu'on utilise des méthodes d'estimation à partir des données de trafic, elle est étroitement liée à la fréquentation du réseau lorsque la mesure est fondée sur des informations fournies par les véhicules. La couverture spatiale dépend alors non seulement de l'infrastructure en place (lecture de badge, balises...) si celle-ci est nécessaire, mais aussi de la présence effective de véhicules sur les éléments de réseau concernés. L'évaluation en devient relativement complexe (un chaînon affecté par un incident, où le débit devient passagèrement nul, peut ainsi disparaître de la zone couverte).

3.2. FINESSE SPATIALE

Là encore, il est possible de distinguer une finesse spatiale a priori (quel est le plus petit élément où le temps de parcours est mesurable ?) d'une finesse a posteriori (quel est le plus petit élément où les exigences, la disponibilité de précision, etc. sont satisfaites). Une expérimentation limitée peut donner des indications intéressantes sur ce point.

3.3. COUVERTURE TEMPORELLE

La question de la couverture temporelle est assez voisine de celle de la couverture spatiale. Elle est distincte de la question de disponibilité car elle ne concerne pas la sensibilité du système à des pannes ou à des dysfonctionnements, mais plutôt la définition de périodes régulières où le système ne fonctionnerait pas ou ne remplirait pas les exigences liées à la qualité de l'information. C'est

ainsi par exemple que certains types de systèmes peuvent ne pas fonctionner en période de trafic très peu dense (la question de savoir si ceci pose un problème ou non est d'une autre nature, et renvoie à la définition de l'exigence « couverture temporelle ».

3.4. FRAICHEUR

L'exigence de fraîcheur, qui indique le délai maximum entre l'occurrence d'un événement (ou d'une nouvelle mesure) sur le terrain et sa traduction dans l'estimation, est une exigence intermédiaire : il est possible de l'ignorer, et de ne formuler qu'une exigence de précision de l'estimation par rapport aux conditions de trafic prévalant durant la période où l'estimation est utilisée. On peut aussi choisir de décomposer le processus, en estimant d'une part la précision de l'estimation par rapport aux conditions de trafic prévalant durant la période où les mesures sont faites, d'autre part le temps séparant cette période de celle où l'estimation est utilisée. Ce dernier inclut la période d'agrégation des mesures ainsi que tous les délais de calcul et de transmission de l'information, qui peuvent être importants sur des systèmes compliqués.

A noter que ni la définition ni l'évaluation de la fraîcheur ne sont évidentes lorsqu'on utilise des méthodes de lissage ou d'agrégation glissante. On peut alors se limiter à évaluer le temps séparant la dernière prise de mesure de la disponibilité de l'estimation, mais la signification de l'indicateur reste relative.

3.5. DISPONIBILITE

La mesure de la disponibilité se fait à l'intérieur de la période de couverture temporelle. L'indicateur correspondant est la mesure des proportions du temps pendant lesquelles les exigences de qualité d'estimation sont satisfaites :

- soit sur l'ensemble du réseau
- soit sur chacun de ses éléments.

A noter qu'une indisponibilité récurrente (par exemple dans certaines conditions de trafic) correspond davantage à une réduction de la couverture temporelle.

3.6. COHERENCE

L'exigence de cohérence est une exigence multiforme, dont la formulation est difficile et l'évaluation encore plus.

Les indicateurs liés à la cohérence peuvent être définis en termes d'écart ou de dispersion

- des performances sur les différentes parties du réseau
- des définitions utilisées
- des finesses spatiales.

3.7. MISE EN ŒUVRE

L'évaluation complète des difficultés et des contraintes de mise en œuvre ne peut se faire que sur un système opérationnel, mais expérimentation et installation pilote peuvent néanmoins être riches d'enseignements sur ce point. Les indicateurs sont nombreux et couvrent plusieurs domaines :

- les problèmes matériels d'installation (génie civil, équipement de véhicules...)
- l'ingénierie de mise en œuvre : paramétrage, mesures préliminaires, calibrage
- les moyens humains à mobiliser
- la maintenance.

3.8. ÉVOLUTIVITE

Il s'agit d'évaluer quels sont les moyens à mettre en œuvre pour faire évoluer le système face aux changements intervenant dans son évolution. Les indicateurs associés peuvent être qualitatifs ou quantifiés en termes de coûts. L'important est d'être capable de répondre à des séries de questions du type « quelles sont les actions nécessaires pour adapter le système à ... ! ». La suite de la question peut viser :

- L'évolution des moyens de mesure (utilisation de capteurs vidéo au lieu de boucles, lecture de plaques minéralogiques au lieu de badges...).
- L'évolution du réseau (ajout d'un chaînon, modification d'un carrefour, création d'un giratoire, d'un couloir de bus...).
- L'évolution du trafic (apparition ou disparition de files d'attente...)
- La modification de la régulation du trafic.
- Le changement de plate-forme de calcul, de modes de transmission.

La réponse à ces questions devra combiner l'examen des spécifications détaillées du système, de ses nécessités de réglage, de paramétrage, etc. Une partie de ces questions est au demeurant commune à tous les systèmes de gestion du trafic.

3.9. COUT

L'évaluation du coût d'un système d'estimation des temps de parcours ne présente guère de caractère spécifique par rapport à celle du coût d'un système quelconque. Les coûts les plus difficiles à évaluer sont certainement les coûts de mise en œuvre, d'évolution et de maintenance. Ce qui confirme l'intérêt d'une évaluation précise de ces différents points.

3.10. PRECISION

L'exigence de précision a été commentée relativement en détail dans le chapitre précédent. La principale difficulté, déjà évoquée, vient de la constitution de valeurs de référence auxquelles comparer l'estimation faite par le système évalué. Si l'indicateur (erreur quadratique moyenne par exemple, ou distribution des écarts) est sans grande difficulté, il est clair que la définition de cette référence doit être cohérente avec la définition utilisée pour le temps de parcours. C'est ainsi par exemple que comparer le temps de parcours instantané estimé sur un itinéraire à l'instant t avec le temps de parcours réel d'un véhicule entrant dans cet itinéraire au même instant t , qui paraîtrait une méthode d'évaluation assez naturelle, revient en fait à combiner de façon inséparable l'évaluation :

- de la précision d'estimation des temps de parcours
- de la dispersion des temps de parcours réels
- de la stabilité des conditions de trafic au cours d'une période équivalente au temps de parcours lui-même.

La précision d'une méthode de mesure n'est véritablement mesurable que dans un cadre expérimental, compte tenu des moyens de mesure et d'enquête à mettre en œuvre.

3.11. FIDELITE

La fidélité s'évalue simultanément à la précision, mais son évaluation peut nécessiter la mise en œuvre d'échantillons de mesure de plus grande taille. Là encore, l'existence d'une référence est nécessaire, dans les mêmes conditions que pour l'évaluation de la précision. En revanche, l'évaluation porte non pas sur les valeurs relatives de la référence et de l'estimation fournie par le système, mais sur le caractère bijectif de la relation référence \leftrightarrow estimation. On voit ainsi qu'un système peu précis peut être extrêmement fidèle. La non conformité d'un système aux exigences de fidélité peut d'ailleurs conduire à réviser à la baisse les exigences de précision.

Exigence	Indicateurs	Condition d'évaluation	Méthode de mesure
Couverture spatiale	trous dans le réseau renseigné	- système opérationnel - expérimentation	- a priori sur les spécifications du système - a posteriori par la vérification de conformité aux exigences de qualité d'information
Finesse spatiale	- éléments minimaux pour lesquels l'estimation est utilisable	- expérimentation	- a priori et a posteriori comme ci-dessus
Couverture temporelle	- périodes (journalières, hebdomadaires...) où l'estimation est disponible - prise en compte de conditions de trafic diverses	- système opérationnel	- a priori et a posteriori - périodes temporelles/ou conditions de trafic
Fraîcheur	- temps entre l'événement et sa traduction dans l'estimation	- système opérationnel - installation pilote	- calcul à partir des spécifications - mesure sur le système
Disponibilité	- % temps où les exigences de performance sont remplies	- toutes	- suivi des dysfonctionnements
Cohérence	- écarts ou dispersion de définition ou de performance	- système opérationnel + dispositif expérimental d'évaluation	- analyse des spécifications détaillées - échantillonnage
Mise en œuvre	- nombreux indicateurs	- installation pilote	- analyse des spécifications détaillées - points à vérifier lors d'une recette
Evolutivité	- coûts - moyens à mettre en œuvre - couches du système affectées	- système opérationnel	- analyse des spécifications - vérification dans le cadre d'une recette
Coûts	- les coûts eux-mêmes	- système opérationnel	- estimation poste par poste
Précision	- écart à la référence	- expérimentation	- constitution d'une référence solide - mesure des écarts
Fidélité	- relation estimation ↔ référence	- expérimentation	- définition de classes de fonctionnement - vérification de la correspondance des classements obtenus par la situation de référence et par l'estimation (taux de bien classés)

Tableau II - 3 : Tableau récapitulatif des exigences, des indicateurs et de leur mode de mesure.

4. METHODOLOGIES D'ÉVALUATION

L'objectif de cette partie est de regrouper les techniques classiquement utilisés pour l'évaluation de la conformité d'un système aux exigences formulées dans un cahier des charges. Les exigences dont il s'agit ici sont celles relatives à la précision et à la fidélité d'un système d'estimation du temps de parcours, dont l'un des indicateurs associé est la mesure de proximité de la sortie du système étudié à une référence.

Lorsque la sortie est une valeur, cette mesure de proximité s'exprime souvent comme une norme de l'écart à la référence (erreur absolue, erreur relative,...). Cependant dans le cas où la sortie est une classe de valeurs, l'un des critères communément utilisé est le taux de bien classés (TBC). Dans les deux cas, la disponibilité d'une référence est un préalable à la mise en œuvre des techniques développées ci-après. Aussi, on se placera dans ce qui va suivre, dans le cadre de l'existence de données de référence, issue par exemple d'une expérimentation.

4.1. PRINCIPALES TECHNIQUES DE VALIDATION A POSTERIORI

Lorsque le modèle d'estimation du temps de parcours a été construit sous des hypothèses paramétriques explicites (cas de modèles statistiques principalement), il est théoriquement possible, mais pas toujours aisé, de calculer directement le pouvoir prédictif de ce modèle. La complexité des calculs, conjuguée au manque de robustesse des résultats obtenus ont favorisé l'usage de procédure d'évaluation s'affranchissant des hypothèses de nature probabiliste. On parle dans ce dernier cas de procédures non paramétriques.

Dans ce qui va suivre, nous présentons les trois principales méthodes d'évaluation *a posteriori* du pouvoir prédictif d'un système d'estimation du temps de parcours. Comme il a été dit précédemment, on supposera qu'un échantillon de mesures de temps de parcours de référence est disponible, permettant la confrontation des résultats du modèle retenu avec la 'réalité' appréhendée par cet échantillon de mesures de référence.

La constitution de cette 'réalité' revêt une importance capitale à la fois pour la mise en œuvre du système d'estimation des temps de parcours et pour la quantification des performances d'un tel système. En effet, dans la pratique, les mesures des temps de parcours sont utilisées à la fois comme des données d'entrée pour la construction du modèle d'estimation du temps de parcours (ceci est davantage vrai pour les modèles basés sur un apprentissage statistique) et de référence pour la validation de ce dernier.

Dans ce qui va suivre, nous allons passer en revue les trois principales méthodes potentiellement utilisables pour l'estimation du pouvoir prédictif d'un modèle d'estimation et de calcul du temps de parcours. Notons que ces méthodes n'ont pas été développées spécifiquement dans le cas de l'estimation du temps de parcours mais l'ont été pour l'estimation du pouvoir discriminant d'une règle de classement.

4.1.1. Validation par resubstitution

Cette technique de validation consiste à estimer les performances d'un modèle de calcul des temps de parcours en réinjectant les mêmes données (échantillon d'apprentissage) ayant été utilisé pour sa construction (cf. Celeux et Turlot, 1991).

Cette démarche est davantage conseillée pour tester la validité au niveau de l'état du modèle : on effectue un retour sur les hypothèses qui ont été imposées au comportement du phénomène modélisé (ici la propagation du trafic). Cette validation permet de vérifier si les estimations obtenues sont conformes aux hypothèses de calcul¹⁷.

Quand cette technique est appliquée à la validité de prédiction, l'estimation du pouvoir prédictif qui en résulte, appelé pouvoir prédictif *apparent*, a tendance à surestimer le 'vrai' pouvoir prédictif du modèle. En effet, le modèle a été façonné de telle sorte qu'il colle au mieux au jeu de données d'apprentissage et par conséquent, les écarts entre les mesures appartenant à l'échantillon d'apprentissage et les estimations issues du modèle sont les plus faibles. Cette surestimation est d'autant plus grande que la taille de l'échantillon d'apprentissage est petite. Par conséquent, le pouvoir prédictif apparent ne constitue pas une estimation acceptable du 'vrai' pouvoir prédictif d'un modèle. Aussi, fait-on appel à d'autres méthodes de validation, fondées sur des procédures de rééchantillonnage.

4.1.2. Validation par échantillon test

Cette méthode consiste à scinder les données disponibles en deux parties : une partie qui sera utilisée pour bâtir le modèle (échantillon d'apprentissage) et l'autre partie (échantillon test), indépendante de l'échantillon d'apprentissage servira pour mettre à l'épreuve le modèle et de ce fait l'estimation du pouvoir prédictif du modèle. L'estimation du pouvoir prédictif d'un modèle par cette procédure élimine le biais du pouvoir prédictif apparent puisqu'une observation n'est jamais utilisée simultanément pour construire et mettre à l'épreuve le modèle.

Cette technique, par son principe même, exige que le nombre d'observations soit important pour pouvoir fractionner les données en un échantillon d'apprentissage et un échantillon test. Il est à noter par ailleurs, que la précision de l'estimation du pouvoir prédictif croît avec la taille de l'échantillon test.

Bien que cette technique améliore la méthode de resubstitution, elle souffre de deux handicaps qui font qu'elle est de moins en moins préconisée. Le premier handicap est lié à l'incapacité de cette technique à fournir une évaluation de la variance de l'estimation du pouvoir prédictif. Le second est inhérent au principe même de cette méthode, à savoir le fait d'amputer les données disponibles d'une fraction pour constituer l'échantillon test peut conduire à un modèle moins robuste. Aussi, est-il préconisé, après avoir évalué sa qualité à l'aide d'un échantillon test, de construire un modèle définitif sur toutes les données disponibles.

¹⁷ D'une façon générale, la plupart des méthodes s'appuient sur l'hypothèse d'absence de dérèglement systématique dans les estimations. De ce fait, la moyenne (ou encore l'espérance mathématique) de l'erreur d'estimation doit être proche de zéro. De plus, on admet que ces erreurs ne sont pas auto corrélées. Dans ce cas, les tests mis en œuvre reposent sur les critères de blanchiment de l'erreur (bruit blanc) et d'absence de corrélation entre les observations et l'erreur.

4.1.3. Validation croisée

La procédure de validation croisée consiste à retirer de l'échantillon d'apprentissage φ (de taille n) une observation (y_i, x_i) ; $i=1, \dots, n$, puis à construire le modèle sur la base de l'échantillon d'apprentissage privé de l'observation i (φ_{-i}) et enfin calculer l'écart entre la réponse du modèle avec comme entrée x_i et la valeur de référence donnée par y_i . Cet écart représente l'erreur de prévision du modèle.

La norme de ces différents écarts fournit un estimateur sans biais (absence de dérèglement systématique) du 'vrai' pouvoir prédictif du modèle car une observation n'est jamais utilisée simultanément pour construire et mettre à l'épreuve un modèle.

Une généralisation de la validation croisée consiste à découper l'échantillon d'apprentissage de taille n en v parties $\varphi_1, \dots, \varphi_v$ de tailles égales à n_1, \dots, n_v , puis à construire le modèle sur la base de l'échantillon d'apprentissage amputé de la partie n_j , $j=1, \dots, v$ et à calculer les erreurs de prévision pour les observations de φ_j , ($1 \leq j \leq v$).

4.1.4. Autres méthodes de rééchantillonnage

Bien que la méthode de validation croisée décrite dans le précédent paragraphe constitue la référence en matière d'estimation du pouvoir prédictif d'un modèle, il existe cependant d'autres schémas de rééchantillonnage utilisables pour estimer l'erreur de prévision d'un modèle.

La plupart de ces schémas ont été développés pour résoudre des problèmes ne relevant pas de l'évaluation ou de la validation de modèles. On peut citer à titre d'exemple, la procédure dite de *jackknife* initialement proposée par Tukey (Tukey, 1958) pour réduire le biais d'un estimateur. Il est apparenté à la validation croisée dans la mesure où il retire tour à tour chaque observation de l'échantillon d'apprentissage. Cependant, son évaluation du pouvoir prédictif est différente de celle de la validation croisée (cf. Efron 1982, McLachlan 1992).

On peut aussi signaler l'utilisation d'une autre technique connue sous le nom de *bootstrap*. Cette technique de rééchantillonnage qui, dans son principe fondamental substitue à la distribution de probabilité inconnue F dont est issu l'échantillon d'apprentissage la distribution empirique F_n , affecte une masse $1/n$ à chaque réalisation contenue dans l'échantillon d'apprentissage et 0 à tout vecteur non présent dans cet échantillon. Ainsi on obtient un échantillon de taille n dit échantillon bootstrap selon la distribution empirique F_n par n tirages aléatoires avec remise parmi les n observations. On peut ainsi construire un grand nombre d'échantillons bootstrap qui vont permettre d'estimer le pouvoir prédictif du modèle selon la procédure de validation par échantillon test (cf. Efron 1983).

Notons enfin que si l'on parvient, au travers de cette batterie de techniques, à détecter des défauts graves dans la mise en œuvre du système d'estimation du temps de parcours, on sera conduit à mettre en question tout ou partie des choix faits jusqu'alors. Ainsi, si le modèle d'estimation du temps de parcours ne respecte pas les exigences de précision initiales, il faut soit tolérer une identité de comportement moins bonne (à caractérisation constante) et réduire les objectifs, soit enrichir la caractérisation et reprendre le processus de modélisation.

5. CONCLUSION

Sous l'éclairage du paragraphe précédent, la pratique de l'échantillon test est à privilégier dès lors que l'on dispose de données en nombre important. En effet, on montre que la précision de l'estimateur du pouvoir prédictif d'un système prévisionnel, et donc *a fortiori* pour un système d'estimation du temps de parcours, dépend de la taille de l'échantillon test : Cette estimation est bien évidemment sans biais, mais sa variance est importante si la taille de l'échantillon test est faible.

Cependant dans un grand nombre d'applications, il est difficile de disposer de données importantes répondant strictement au protocole d'échantillonnage choisi. La stratégie préconisée est alors d'utiliser l'information de manière exhaustive pour l'élaboration du modèle d'estimation du temps de parcours, et par des techniques de rééchantillonnage on corrige l'optimisme de pouvoir prédictif apparent. Cette correction est le plus souvent effectuée au moyen de la validation croisée ou encore de la technique du bootstrap. Bien que cette dernière technique soit plus lourde dans sa mise en œuvre, elle est cependant beaucoup plus naturelle et peut être mieux adaptée au protocole expérimental.

PARTIE IV.

CONCLUSIONS GÉNÉRALES

L'objectif de cette étude était de vérifier la faisabilité technique d'une estimation multi-sources du temps de parcours, en présence de données de sources hétérogènes. Elle déborde au demeurant ce cadre strict puisqu'elle inclut également une réflexion élargie sur les besoins des exploitants, des opérateurs de systèmes d'information et des usagers en matière de temps de parcours (cf. El Faouzi et Lesort, 2000), ainsi qu'une réflexion sur la nature de cet indicateur et sur l'évaluation des systèmes de mesure et d'estimation.

Dans le cadre de cette étude, deux sources ont été considérées : des capteurs classiques de trafic constitués d'une boucle électromagnétique, qui permettent de mesurer le débit et le taux d'occupation et d'en déduire une estimation de temps de parcours via un algorithme de conversion, et un échantillon réduit de véhicules, dits véhicules traceurs, qui recueillent les temps de parcours qu'ils ont réalisés. Abordant ce problème comme un problème typique de fusion de données, nous avons choisi d'en étudier une approche par la théorie des crédibilités pour sa finesse de modélisation de la connaissance et sa grande souplesse. Dans ce cadre, nous avons développé une méthode générale de fusion de données de temps de parcours basée sur la théorie des crédibilités, que nous avons appliquée aux données d'une expérimentation.

Il résulte de cette étude que des gains non négligeables peuvent être attendus d'un système multi-sources d'estimation du temps de parcours soigneusement conçu, par rapport à l'utilisation d'une source de données unique quelle qu'elle soit. Ces résultats encourageants ne doivent pas occulter les difficultés aussi bien techniques qu'opérationnelles qui n'ont pas encore reçues des réponses complètement satisfaisantes. Aussi, ce travail ne peut donc être considéré que comme une étude préliminaire destinée à donner un cadre d'ensemble à l'utilisation d'informations multiples pour l'estimation des temps de parcours. Elle devrait déboucher sur un projet plus ambitieux associant les exploitants, les opérateurs de systèmes d'information et leurs conseils, et permettant de mettre en place des expérimentations en vraie grandeur.

RÉFÉRENCES

- ABOURS S. – *Estimation de temps de parcours sur un axe urbain à partir de taux d'occupation*, Rapport I.R.T., N° 76, 1985.
- AHMED S.A., COOK A.R. – Analysis of freeway traffic time series data by using Box-Jenkins techniques, *Transportation Research Record* 722, pp. 1-9, 1979.
- AKAHAME H., KOSHI M. – Short-term prediction of inflow volumes of urban freeways., Proceedings of the Second International Conference on Road Traffic Control, April, pp. 35-38., 1986.
- BONVALET F., ROBIN-PREVALLEE Y. – Mise au point d'un indicateur permanent des conditions de circulation en Île-de-France., *Transport Environnement Circulation*, n° 84-85, Septembre-décembre 1987.
- BUISSON Ch., J.-B. LESORT – *Méthodes de mesure automatisée des temps de parcours en milieu urbain : Application au réseau Val de Marne*. Rapport de convention LICIT-Conseil Général de Val de Marne, décembre 1999.
- CELEUX G., TURLLOT J.-C. – Estimation de la qualité d'une règle discriminante. In : «Analyse discriminante sur variables continues». G. Celeux (Editeur), *Collection Didactique INRIA*, pp. 37-49., 1991.
- CHAPOUILLE P. – *Planification et Analyse des expériences*, Masson & Cie, 1973.
- COHEN S., NOUVELIERE C. – Temps de parcours sur artères urbaines régulées : une approche neuronale, *Transport Environnement Circulation*, N° 150, pp. 33-37, 1998.
- COUTON F. – *Application des mélanges de lois probabilité à la reconnaissance de régimes de trafic routier*, Rapport de stage, sous la direction de M. Danech-Pajouh, 1995.
- DAILY D.J. – Travel time estimation using cross-correlation techniques, *Transportation Research*, 27B(2), pp. 97-107, 1993.
- DAILEY D.J. – Travel time estimates using a series of single loop volume and occupancy measurements, *Transportation Research Board*, 76th Annual Meeting, Washington, 1997.
- DAVID Y. ET AL. – *Data Plus : étude prospective dans le domaine du recueil de données de trafic*, Rapport d'étude CERTU-PREDIT, 1999.
- DEMPSTER A.P. – Upper and Lower Probabilities Induced by Multivalued Mapping. *Annals of Mathematical Statistics*, Vol. 38, pp. 325-339, 1967.
- DEMPSTER A. P. – A Generalization of Bayesian Inference, *Journal of the Royal Statistical Society*, 30B, pp. 205-247, 1968.

- EL FAOUZI N.-E., LESORT J.-B. – *Estimation du temps de parcours en milieu urbain*. Rapport de convention N° 9402, INRETS-LICIT, 1994.
- EL FAOUZI N.-E., LESORT J.-B. – *Fusion de données pour l'estimation du temps de parcours. Partie I. Analyse des pratiques et des besoins des exploitants et des opérateurs*. Rapport de convention N° 2001, INRETS-LICIT, Juillet 2000.
- EL FAOUZI N.-E. – *Fusion linéaire d'estimateurs multiples : Méthodes de pondération et de rééchantillonnage*, Rapport technique N° 9701, Licit, INRETS-ENTPE., mars 1997.
- EL FAOUZI N.-E. – *Combining Predictive Schemes in Short-Term Traffic Forecasting.*, *Proceedings of the 14th I.S.T.T.*, A. Ceder (Eds.), pp. 471-487, Jerusalem, 20-23 July, 1999.
- EL FAOUZI N.-E. – *Fusion de données : Concepts et méthodes*. Rapport de Synthèse N° 2002, INRETS-LICIT, décembre 2000.
- EL FAOUZI N.-E., C. SIMON – *Fusion de données pour l'estimation des temps de parcours via la théorie de l'évidence*. *Recherche Transports Sécurité* N° 68, Juillet-Septembre, 2000.
- EFRON B. – *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia, SIAM Monograph, 1982.
- EFRON B. – *Estimating the error rate of a prediction rule : Improvement on cross-validation*. *J.A.S.A.*, **87**, pp. 316-331., 1983.
- HABOUZIT R. – *Identification statistique des régimes de trafic pour l'estimation du temps de parcours*. Projet libre 2^{ème} année ISTIL, sous la direction de N.-E. El Faouzi, 2000.
- HALL F., PERSAUD B.N. – *Evaluation of Speed Estimates made with Single-Detector Data from Freeway Traffic Management Systems*, *Transportation Research Record*, N° 1 457, pp. 149-157, 1989.
- KIMBER R.M., DAILY P.N. – *Time-dependent queuing at Road Junctions: Observation and Prediction*, *Transportation Research*, **20B(3)**, pp. 187-203, 1986.
- KOPPELMAN F., SETHI V. , J. IVAN. – *Calibration of Data Fusion Algorithm Parameters with Simulated Data*, Advance Working Report, TRF-ID-152, 1994.
- LAM T, ROTHERY R.W. – *The Spectral Analysis of Speed Fluctuations on Freeway*, *Transportation Science*, **4**, N° 3, p. 293-309, 1970.
- LEBACQUE J.-P. – *Instantaneous Travel Times for Macroscopic Traffic Flow Models for Networks in the Context of Dynamic Assignment.*, Rapport CERMICS, N° 59-96, 1996.
- LESORT J.-B. – *Étude d'indicateurs de circulation en milieu urbain*, Rapport INRETS, N° 142, juin 1991
- LESORT J.-B., H. MONGEOT – *Mise en œuvre de l'indicateur de vitesse moyenne BRP dans le cadre du projet SITER*, Rapport d'expertise, octobre 1994.
- LESORT J.-B. – *Éléments de la théorie du trafic.*, ENTPE-Mastère I.E.D, 1998.

- LI J., THAKURIAK P., SEN A., ROUPHAIL N. – *Short-term Travel Time Prediction*, Advance Project Report, TRF-TT-01, August 21, 1992.
- McLACHLAN G. J. – Assessing the Performance of an Allocation Rule. *Computer and Mathematics with Applications*, 12(A), pp. 261-272., 1986.
- McLACHLAN G. J. – *Discriminant Analysis and Statistical Pattern Recognition*. New York, Wiley, 1992.
- MOHAMMADI R. – Journey Time Variability in the London Area: 1. Journey time distribution, *Traffic Engineering and Control*, 38(5), pp. 250-257, May 1997.
- OLIVERO P. – *Expérimentation PROGEN CED9 – PROMETHEUS : Réalisation d'une campagne de mesures sur la ZELT*. Rapport de convention, ZELT-10, 1994.
- PETTY K. F., BICKEL P., JIANG J., OSTLAND M., RICE J., RITOV Y., SCHOENBERG F. – *Accurate Estimation of Travel Times from Single-loop Detectors*, Working paper, University of California, Berkeley, 1994.
- SANWAL K., PETTY K., WALRAND J., FAWAZ Y. – An Extended Macroscopic Model For Traffic Flow, *Transportation Research*, 30B(1), pp. 1-9, 1996.
- SEN A., P. THAKURIAH, X.-Q. ZHU, A. KARR. – Frequency of Probe Reports and Variance of Travel Time Estimates, *J.T.E.*, pp. 290-297, 1997.
- SHAFER G. – *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, N. J., 1976.
- STHI V., N. BHANDARI, F. KOPELMAN, J. SCHOFER – Arterial Incident Detection Using Fixed Detector and Probe Vehicle Data., *Transp. Research*, 2(C)., pp. 99-112., 1995.
- SHKABLO S., BHAT C., KOPPELMAN F., LI J., THAKURIAH P., SEN A., ROUPHAIL N. – *Short-term Travel Time Prediction: Review of Literature and Methods*. Advance Project Report, TRF-TT-05, Urban Transportation Center, University of Illinois, Chicago.
- SIMON C. – *Estimation du temps de parcours à partir de données de sources hétérogènes par la théorie de l'évidence*, Rapport de stage de DESS, sous la direction de N.-E. El Faouzi, 1999.
- SIER. – *Variabilité des temps de parcours sur le réseau de voies rapide d'Île-de-France –Année 1999.*, D.R.E.I.F., Novembre 2000.
- SISIOPIKU V. – *Travel Time Estimation from Loop Detector Data for Advanced Travellers Information Systems Applications*, Ph.D. Thesis, Chicago, Illinois, 1994.
- SISIOPIKU V., ROUPHAIL N. – Toward the use of Detector output for Arterial Link Travel Time Estimation: A literature Review, *Transportation Research Record*, N° 1 457, pp. 158-165, 1994a.
- SISIOPIKU V., ROUPHAIL N. – Analysis of Correlation between Arterial travel Time and Detector data from Simulation and Field Studies., *Transportation Research Record*, N° 1 457, pp. 166-173, 1994b.

- SRINIVASAN K. AND P. JOVANIS. – Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network. *Transportation Research Record*, N° 1537, pp.15-22. Année 1996.
- TARKO A., ROUPHAIL N. – *Travel Time Data Fusion in ADVANCE*, Pacific-Rim TransTech Conference: «A Ride Into the Future», July 25-28, 1993, Washington, 1993.
- THAKURIAH P., A. SEN. – *Data Fusion for Travel Time Prediction: A statement of Requirements*. Advance Working Paper, N° TRF-TT-07, 1993.
- TUKEY J. – Bias and Confidence in Not Quite Large Samples. *Ann. Math. Statist.*, **29**, pp. 164, 1958.
- WEBSTER F. V. – *Traffic Signal Settings*, Road Research Technical Paper, No. 39, Road Research Laboratory, London, U. K., 1958
- WEYMANN J., FARGES J-L., HENY J-J. – Dynamic Route Guidance with Queue and Flow Dependent Travel Time, *Transportation Research*, Part C., pp 165-83, 1994.