RECONSTITUTION DE MATRICES ORIGINE-DESTINATION DU TRAFIC ROUTIER INTERURBAIN PAR LES COMPTAGES ET LES SIMULATIONS DE CIRCULATION

Yu-Sen CHEN
Chercheur-thèse
Institut National de Recherche sur les Transports et leur Sécurité
Paris - France

INTRODUCTION

Une matrice Origine-Destination de trafic révèle la répartition spatiale des déplacements. Elle évalue les mouvements du trafic, qualitatifs et quantitatifs, entre les différentes zones. C'est pour cette raison que la matrice O-D joue un rôle important dans les études de trafic et l'évaluation économique du réseau. En particulier, en ce qui concerne:

- la prévision de trafic,
- la planification et l'exploitation du réseau,
- le choix d'investissements,

la connaissance de la matrice O-D est impérative.

Il existe des méthodes d'estimation de ces matrices, pratiques et théoriques :

- celles dites pratiques : "Collage de papillons", "Relevé de numéros minéralogiques", "Photographies aériennes", et "Enquêtes à domicile",
- et celles dites théoriques : "Modèles gravitaires", "Approche par l'interaction", "Approche par l'entropie", "Modèles probabilistes", et "Régression linéaire".

Les méthodes pratiques sont coûteuses tant en main d'oeuvre qu'en traitement des données. Très souvent, les données sont incomplètes. C'est-à-dire qu'un grand nombre de paires O-D n'est pas observé. De plus, les données recouvertes partiellement, peuvent conduire à des informations O-D contradictoires. Le renouvellement et l'exploitation de ces matrices posent toujours des difficultés.

Pour avoir une matrice complète et consistante, les techniques d'estimation des matrices O-D sont utilisées. Bien que les hypothèses et les termes d'optimisation puissent différer, l'objectif pour atteindre la matrice de base (ou de référence) reste le même : trouver le "fitting optimal" entre les estimations et les données disponibles.

D'une part, les méthodes pratiques sont coûteuses. Leur validité est limitée dans le temps et l'espace, ce qui ne permet pas de mettre à jour des matrices O-D. D'autre part, les méthodes théoriques existantes ne donnent aucune garantie sur les résultats. C'est dans ce contexte qu'il faut développer un nouvel outil, qui permet à la fois de reconstituer la matrice O-D et de contrôler les résultats.

Les comptages sont les plus favorables pour ce rôle. Ils ne sont pas coûteux et sont aujourd'hui les moyens les plus utilisés (quasi-standardisés) pour l'acquisition des données routières.

1. ANALYSE DES APPROCHES D'ESTIMATION D'UNE MATRICE O-D

1.1. Equations fondamentales

Les équations fondamentales entre une matrice O-D et les comptages sont les suivantes:

$$R_{kl} = \sum_{ij}^{N} [P^{kl}_{ij} * T_{ij}] \quad , \quad kl = 1,2,...,M \qquad ...(1)$$

où: R_{kl} = trafic de comptages sur le tronçon kl,
P^{kl}_{ij} = fraction de trafic de i à j en passant par kl,
T_{ij} = trafic de i à j, i, j =1,2,...,N,
N = nombre de centroïdes,
M = nombre de compteurs.

1.2. Données nécessaires

Dans l'équation (1), $\{R_{kl}\}$ (nombre: M) sont connus, mais $\{P^{kl}_{ij}\}$ (nombre: M*N*N) et $\{T_{ij}\}$ (nombre: N*N) sont inconnus. Dans la plupart des cas, M est inférieur (même fortement inférieur) à N*N. Donc on ne peut pas trouver la matrice $\{T_{ij}\}$ directement.

En plus, $\{P^{kl}_{ij}\}$ est liée étroitement à $\{T_{ij}\}$. En fait, il est clair que $\{P^{kl}_{ij}\}$ est la fonction de $\{T_{ij}\}$ et de graphe du réseau. Pour avoir une bonne $\{T_{ij}\}$, il faut aussi résoudre le problème de $\{P^{kl}_{ij}\}$.

1.3. Méthodes existantes pour résoudre le problème

Une analyse des méthodes existantes a été effectuée dans un rapport de stage [1]! D'après les premières conclusions de ce rapport, la méthode de l'entropie et celle du maximum de vraisemblance sont apparues comme les plus utilisées.

<u>1.3.1. La méthode de l'entropie</u>

L'introduction de l'entropie dans l'analyse du système de transport a été faite en premier par A.G.WILSON en 1973 [2]. Il a associé à la distribution T_{ij} le plus grand nombre de façons $S(W(T_{ij}))$ par lesquelles les individus peuvent produire la totalité de la distribution.

$$W(T_{ij}) = (T!)/(\prod_{ij} T_{ij}!)$$

dont $T = \sum_{i} O_i = \sum_{j} D_j$ = Nombre total de trajets

En prenant T comme constante, WILSON utilise l'approximation de Stirling pour obtenir la fonction objectif approximative de ce type d'entropie. Donc on peut écrire :

$$\text{minimise} \sum_{j=1}^{m} \sum^{n} \ln (T_{ij} * \ln T_{ij} - T_{ij}) \qquad ...(2)$$

C'est dans ce contexte que L.G. WILLUMSEN a introduit en 1978 le modèle de Maximisation d'Entropie (M.E.) pour l'estimation de la matrice O-D à partir de comptages. En même temps, Van ZUYLEN [3] a conçu le modèle de Minimisation d'Information (M.I.) en incorporant l'information supplémentaire (une matrice a priori). Ces deux modèles sont équivalents [4].

L'intérêt de l'incorporation d'une matrice O-D a priori est qu'elle peut aboutir à une estimation de matrice plus réaliste. Donc la matrice O-D a priori joue un rôle important pour l'estimation de la matrice O-D actuelle [3].

Quant aux comptages, il faut noter, cependant, que chacun ajoute une contrainte pour l'estimation, contrainte qui se traduit par un multiplicateur Lagrangien supplémentaire, et par conséquent, par un coefficient de plus dans le modèle.

Formulations du modèle de M.I. : [4]

$$\text{minimise} \sum_{i}^{n} \sum_{j}^{n} [T^{\dagger}_{ij} * \ln (T^{\dagger}_{ij}/T_{ij})] \qquad ...(3)$$

$$\text{Soumis à :} \sum_{i}^{n} \sum_{j}^{n} [T^{\dagger}_{ij} * P^{kl}_{ij}] = R_{kl} \qquad \forall \quad kl \qquad ...(1)$$

où : $T^{\dagger}_{ij}$ = nombre de trajets de i à j (à estimer),.
T_{ij} = matrice O-D a priori,
si $T_{ij} = 1$, Equation (3) est équivalente à la formulation de M.E.
R_{kl} = Contrainte kl (ex. : comptages),
P^{kl}_{ij} = fraction des trajets ij soumis à la contrainte R_{kl} ,
(ex. : nombre de trajets qui utilise la relation kl)

La minimisation de l'information (3) soumise à la contrainte (1) donne une estimation pour T_{ij}. Cette solution peut être dérivée pour minimiser le Lagrangien (si les conditions de Kuhn-Tucher sont remplies) :

$$T^{\dagger}_{ij} = T_{ij} * X_0 * \prod_{kl} X_{kl} \qquad \forall \quad ij \qquad ...(4)$$

$$\text{où: } \ln(X_0) = -1, \quad \ln(X_{kl}) = \mu_{kl} * \sum_{i} \sum_{j} P^{kl}_{ij} \qquad ...(5)$$

$$\text{et } \sum_{i}^{n} \sum_{j}^{n} [T^{\dagger}_{ij} * P^{kl}_{ij}] = R_{kl} \qquad \forall \quad kl \qquad ...(1)$$

Ces trois équations (4),(5) et (1) permettent de calculer $\{T_{ij}\}$ à l'aide d'une méthode d'itération étant données $\{T^{\dagger}_{ij}\}$ et $\{P^{kl}_{ij}\}$. Comme tous les processus d'itération, les résultats sont obtenus quand le standard de convergence est satisfait.

Ces démarches ont été largement utilisées dans ce genre de méthode (voir la bibliographie de [1]), à noter que les $\{P^{kl}_{ij}\}$ ont été toujours supposées connues.

Avant d'entrer dans les détails de cette méthode d'Entropie, on voit déjà un premier problème. Ce problème est dû au fait que les $\{P^{kl}_{ij}\}$ ont été supposées connues et restent inchangées pendant les processus d'itération. Dans le paragraphe **1.2.**, on a indiqué que les $\{P^{kl}_{ij}\}$ sont fonction de $\{T_{ij}\}$ et du graphe du réseau et sont normalement le produit de l'étape d'affectation. En pratique, les $\{P^{kl}_{ij}\}$ ne sont pas toujours faciles à obtenir. En l'absence d'observations directes, un algorithme devait être utilisé. Donc cette supposition de la connaissance de $\{P^{kl}_{ij}\}$ est loin d'être réaliste et ne donne aucune garantie sur les résultats.

Quant aux modèles, il y a aussi quelques remarques à faire:

a). l'équation (3) n'est pas définie dans le cas où $T_{ij} = 0$. Donc toutes les relations de T_{ij} qui ne sont pas observées ne peuvent pas être utilisées dans le processus de l'estimation. Cela peut devenir un problème sérieux si une matrice observée a priori contenant, pour la plupart, la valeur zéro, est utilisée ,

b). si l'ensemble des contraintes dans l'équation (1) est inconsistant, alors, la solution produite par le modèle n'aura pas de sens ,

c). ces comptages sont utilisés pour ajuster les trafics des seules paires O-D observées. Toutes les autres paires O-D resteront inchangées.

Enfin rappelons que pour faire tourner ce modèle, trois sources principales de données y sont nécessaires (I) les comptages, qui doivent être consistants et bien distribués dans le réseau ; (II) la matrice O-D a priori, la meilleure possible ; (III) des matrices de fractions des trajets qui sont normalement le produit de l'étape d'affectation.

Ces modèles d'entropie semblent convenir spécialement pour actualiser une ancienne matrice O-D [Van ZUYLEN, 3].

<u>1.3.2. La méthode du maximum de vraisemblance</u>

Cette méthode consiste à maximiser la vraisemblance entre les données connues et la matrice $\{T_{ij}\}$ à estimer.

$$L(T, V \mid T^{\dagger}) = L(T \mid T^{\dagger}) * L(V \mid T^{\dagger})$$

où: T -- la matrice a priori, $\{T_{ij}\}$,
V -- les comptages, $\{V_{ij}\}$,
$T^{\dagger}$ -- la matrice à estimer, $\{T^{\dagger}_{ij}\}$.

ou: $$T^{ml} = \ln(L(T, V \mid T^{\dagger})) = \max_{T^{\dagger} \in S} \{\ln L(T \mid T^{\dagger}) + \ln L(V \mid T^{\dagger})\} \quad ...(6)$$

où: S -- l'ensemble de $\{T^{\dagger}_{ij}\}$ possible.

Le calcul de T^{ml} dépend de la distribution des probabilités de T et V. On ne considère que l'échantillon constitué ici par hasard.

Le développement de ce modèle de vraisemblance conduit à des équations identiques à (4),(5) et (1), et la résolution de ces équations a besoin aussi d'une matrice a priori $\{T_{ij}\}$ et des matrices d'affectation $\{P^{kl}_{ij}\}$.

Les mêmes problèmes existent dans ce modèle comme dans l'entropie. Les $\{P^{kl}{}_{ij}\}$ et $\{Tij\}$ sont supposées connues et restent inchangées pendant le calcul.

2. DEVELOPPEMENT D'UNE NOUVELLE METHODE "REMODE"

Les paragraphes précédents ont montré les équations fondamentales et les méthodes existantes pour résoudre le problème. Dans ce paragraphe, nous allons analyser tous ces processus et tenter de proposer les démarches nécessaires et faisables pour résoudre les questions que soulèvent ces méthodes afin d'établir une nouvelle méthode **REMODE** (REconstitution d'une Matrice Origine-DEstination).

2.1. Synthèse du problème

Le problème posé est de reconstituer la matrice O-D par les comptages, et il est très clairement présenté par l'équation (1):

$$R_{kl} = \sum_{ij}^{n} [P^{kl}{}_{ij} * T_{ij}] \quad , \quad kl=1,2,...,M \qquad ...(1)$$

La résolution de ce problème conduit à deux méthodes: méthode de l'entropie et méthode de vraisemblance. Ces méthodes ont demandé toutes deux une matrice a priori et des matrices de fraction de trajets $\{P^{kl}{}_{ij}\}$.

Ces deux conditions prédéfinies sont trop simplifiées et ne présentent aucune garantie de bons résultats, parce que les $\{P^{kl}{}_{ij}\}$ sont la fonction des $\{T\dagger_{ij}\}$ inconnues et le produit de l'étape d'affectation. De plus, la matrice a priori $\{Tij\}$ n'est pas facile à obtenir non plus. Toute erreur dans la matrice $\{Tij\}$ peut conduire à des résultats erronés. Il a été établi que "la précision des résultats d'estimations est extrêmement sensible à la matrice a priori" (NGUYEN, 1982).

2.2. Fonction objectif

La synthèse précédente montre qu'une bonne résolution du problème sollicite les modélisations des matrices $\{Tij\}$ et $\{P^{kl}{}_{ij}\}$.

La fonction objectif du modèle **REMODE** consiste à minimiser la différence entre le comptage et la valeur reconstituée sur tous les compteurs.

Elle peut donc être exprimée comme ci-dessous:

$$\text{Min } Z = \text{Min } \left(\sum_{kl}^{m} | R^{kl}{}_{e} - R^{kl}{}_{m} | \right) \qquad ...(7)$$

$$\text{soumis à:} \quad \sum_{i}^{n} \sum_{j}^{n} [T_{ij} * P^{kl}{}_{ij}] = R^{kl}{}_{e} \qquad \forall \quad kl \qquad ...(8)$$

où: R^{kl}_e = trafic estimé sur le compteur kl,
R^{kl}_m = trafic mesuré sur le compteur kl,
P^{kl}_{ij} = pourcentage de T_{ij} passant par le compteur kl.

Il est évident que la fonction objectif dépend non seulement de T_{ij} , mais aussi de P^{kl}_{ij}. C'est à partir de ce point essentiel qu'on déduit que les P^{kl}_{ij} doivent être des variables comme les T_{ij} pendant les processus de résolution.

Notons que T_{ij} n'est plus une valeur prédéterminée d'une matrice O-D a priori, mais sera initialisé par un sous-modèle (voir 2.3.3. ci-dessous) fondé sur les caractéristiques socio-économiques. La fonction objectif a donc non seulement un sens mathématique mais aussi celui de l'économie.

Le développement de la fonction objectif est reporté au paragraphe **2.4.** après l'introduction de l'initialisation du trafic.

2.3. Modélisation des processus

La résolution de P^{kl}_{ij} demande une simulation de l'écoulement de circulation. Il faut donc développer une série de méthodes: recherche des plus courts chemins, désagrégation des déplacements, etc. La prédétermination de P^{kl}_{ij} par d'autres méthodes exogènes conduit à des résultats statiques et ne satisfait pas la fonction objectif (7).

En fait, la relation entre T_{ij} et les comptages n'est pas directe. Elle est décrite par l'intermédiaire de différentes étapes: composition du trafic, établissement du réseau, écoulement du trafic, etc. Une bonne reconstitution sans connaissance de ces étapes n'aura pas de sens.

2.3.1. Désagrégation des déplacements

La résolution de P^{kl}_{ij} demande des affectations fines des déplacements. Dans ce cas, les déplacements trop agrégés ne conduiront pas à de bons résultats.

Nous avons désagrégé les déplacements selon les motifs et les valeurs du temps de chaque motif. Grâce à la connaissance des partages entre les différents motifs, les désagrégations par motif ne posent pas de problèmes particuliers.

La désagrégation des déplacements pour un motif donné est un peu plus complexe. Nous faisons l'hypothèse que les valeurs du temps des voyageurs pour chaque motif suivent une loi log-normale. Nous obtenons ainsi les différentes tranches des déplacements selon la finesse souhaitée.

Nous avons donc un certain nombre de segments des déplacements. En simulant de l'écoulement de circulation, nous ne connaissons pas la priorité d'un segment particulier à se déplacer. Tous les segments de déplacements seront donc réorganisés selon une loi aléatoire.

2.3.2. Etablissement du réseau

Un réseau routier est nécessaire pour la modélisation. Ce fichier du réseau devait comporter le numéro de la zone étudiée, le type (ou éventuellement la largeur, la rampe et la capacité), les coordonnées et les intitulés de ses deux extrémités (noeuds), la longueur de chaque route.

L'échelle du réseau dépendra du niveau d'étude. Pour une étude du trafic national, le réseau sera composé de routes nationales et autoroutes, et quelques routes départementales si nécessaire.

Dans le cas où plusieurs routes relient deux noeuds, le programme crée un noeud ou un itinéraire fictif pour éliminer cette ambigüité.

2.3.3. Initialisation du trafic

Cette initialisation du trafic par O-D est fondée sur les indicateurs socio-économiques des zones étudiées, avec l'aide d'un modèle de génération-distribution. Le calibrage de ce modèle se fait à partir de certaines O-D connues et plus tard par le programme lui-même pendant le calcul.

Le modèle d'initialisation de matrice $\{T_{ij}\}$ est comme suit:

$$T_{ij} = K * E_i * A_j * F_{ij} * U_{ij} \qquad \text{...(9)}$$

où:
- K -- constante,
- E_i -- émission de zone i,
- A_j -- attraction de zone j,
- F_{ij} -- facteur de séparation entre i et j, fonction du réseau et d'effet de l'offre,
- U_{ij} -- coefficient au niveau d'usager, en fonction du revenu, de la profession (formation), du PIB et de l'interêt touristique des zones i et j.

La matrice ainsi initialisée reflète notre connaissance primaire. Elle a au moins un sens économétrique. En effet, elle sera reconstituée par la fonction objectif.

2.3.4. Calcul des coûts généralisés

Le coût généralisé est un paramètre important, car il est relié à la fois aux usagers et à l'offre. Il est donc fonction de la valeur du temps de l'usager, du type et de la géométrie de la route.

La formule pour ce coût généralisé est donnée comme suit:

$$G(i,l)=v(l)*t(i)+(Cc+Cp+Ce+Cm)*d(i)$$

où:
- $G(i,l)$ = Coût généralisé du tronçon i pour l'usager l,
- $v(l)$ = valeur du temps de l'usager l,
- $t(i)$ = temps de parcours sur tronçon i,
- $d(i)$ = longueur du tronçon i,
- Cc = consommation du carburant sur i,
- Cp = péage sur i,
- Ce = frais d'usure et d'entretien du véhicule,
- Cm = malus d'inconfort.

Le temps de parcours est fondé sur une lois temps-débit, et sera actualisé après chaque chargement du réseau.

2.3.5. Recherche des plus courts chemins

La recherche des plus courts chemins est déduite d'un algorithme de la théorie des graphes. Nous pouvons réaliser cette recherche pour une paire O-D ou pour l'ensemble des O-D selon le cas.

Le critère de recherche peut être le temps de parcours, la longueur ou le coût généralisé. Le plus souvent, nous utilisons le coût généralisé.

Notons que La recherche des plus courts chemins est effectuée pour chaque chargement du réseau.

2.3.6. Simulation de l'écoulement du trafic

Deux hypothèses ont été faites pour assimiler l'écoulement de circulation du trafic:

1. les voyageurs prennent les chemins du moindre coût généralisé,
2. en cas d'encombrements, les voyageurs prennent des itinéraires de délestage.

Ces deux hypothèses sont en harmonie avec les hypothèses faites pour les désagrégations de déplacements. En fait, une désagrégation assez fine de déplacements supporte bien ces hypothèses.

Un autre point important est que l'attribution d'une paire O-D est choisie par le processus aléatoire, ce qui approche l'incertitude des comportements des voyageurs et est tout à fait contraire à la méthode traditionnelle qui attribue les O-D par un simple ordre incrémental d'itération.

2.3.7. Reconstitution du trafic

Les processus de recherche des plus courts chemins et des simulations du trafic se répètent plusieurs fois selon le nombre de segments des déplacements. Pendant ces processus, les coûts généralisés changent en fonction des nouveaux états du trafic.

Après tous les chargements du réseau, la fonction objectif est employée. Elle examine l'écart de chaque compteur, et ensuite déforme la matrice des T_{ij} et p^{kl}_{ij}.

La nouvelle matrice sera jugée par une analyse de corrélation avec les indicateurs socio-économiques des zones étudiées. Une bonne corrélation conduit les processus à la phase suivante et la matrice sert à calibrer les modèles de l'initialisation.

2.4. Réalisation informatique

Ce modèle **REMODE** est donc constitué de processus différents mais cohérents car tous ses sous-modèles ont été écrits de façon continue.

Chaque sous-modèle (modules) est donc un sous-programme. Chaque module est indépendant, mais peut tourner interactivement avec les autres. Cela confère une grande souplesse au modèle.

Une telle tâche ne pourra pas aboutir sans avoir recours aux outils informatiques.

2.4.1. Développement des équations

La fonction objectif et l'initialisation de la matrice permettent de développer l'ensemble des équations mathématiques.

Combinons les équations (7), (8) et (9) et faisons les dérivées de Z sur E_i , A_j , F_{ij} , U_{ij} , K et P^{kl}_{ij} . Nous obtenons une série d'équations. Elles sont assez complexes et ne présentent pas de grand intérêt à 'être reproduites ici.

La résolution de ces équations se fait par itération.

2.4.2. Diagramme des processus

Le diagramme de résolution de toutes ces équations est présenté dans la **Figure 1.**

Notons que les critères d'arrêt sont la précision souhaitée des résultats et le nombre d'itérations maximal du programme. Si ces critères sont satisfaits, le programme s'arrête. Sinon, l'itération continue avec la nouvelle matrice calculée et les nouveaux chemins de parcours.

2.4.3. Résultats provisoires du modèle

Le modèle **REMODE** donne une matrice O-D reconstituée et les trafics O-D décomposés d'un comptage.

Les résultats provisoires montent déjà quelques points intéressants:

1. Le trafic total correspond à peu près au chiffre statistique et reste stable pendant les itérations;
2. Les compteurs retrouvés par la méthode d'affectation sont autour de 85%, dont 95% ont été bien "réconciliés";
3. Les trafics O-D sont interprétables du point de vue socio-économique routier.

3. CONCLUSION

Une nouvelle méthode a été développée pour reconstituer une matrice O-D à partir des comptages. Cette méthode, par rapport aux méthodes existantes, ne demande pas une matrice a priori et des matrices de fraction de trajets. En effet, elles sont modélisées et deviennent une partie intégrante de la méthode développée.

Cette conception rend la méthode beaucoup plus complète et souple. Les modélisations de la matrice O-D a priori et des matrices de fraction permettent de décrire tous les processus de façon continue, ce qui nous offre des résultats plus réalistes.

La modélisation de tous les processus et surtout leur calibrage quasi autonome ouvre une voie intéressante dans les domaines du transport.

Les premiers résultats de l'application sur le réseau français ont montré que la matrice reconstituée est interprétable des points de vue mathématique et économique du réseau.

4. BIBLIOGRAPHIE

1. CHEN Y.S., "Analyse des modèles existants de génération-distribution pour la reconstitution d'une matrice O-D de trafic routier Origine-Destination". Rapport de Stage, INRETS, Décembre 1990.
2. WILSON A.G., "A statistical theory of spatial distribution models". Transp. Res. Vol. 1, pp.253-269, 1967
3. VAN ZUYLEN H.J., "A method to estimate a trip matrix from traffic volume counts". PTRC, Proceedings, pp.305-311, 1978
4. HAMERSLAG Rudi and IMMERS H.Ben, "Estimation of trip matrices: Shortcomings and possibilities for improvement". Transp. Res. Rec. 1203, pp.27-39, 1988
5. SPIESS Heinz, "A maximum likelihood model for estimating Origine-Destination matrices". Transp. Res. -B, Vol. 21B No.5, pp.395-412, 1987
6. NGUYEN S, AMALFI, "Estimating Origine-Destination matrices from observed flows". Italy, October, 1982

7. CASCCETTA Ennio, "A unified framwork for estimating or updating Origine/Destination Matrices from traffic counts". Transp. Res. -B, Vol. 22B No.6, pp.437-455, 1988

[1] Les chiffres entre crochets se rapportent à la bibliographie .

Figure 1 - Diagramme du modèle REMODE