



PREMIER MINISTRE

Commissariat général
à la stratégie
et à la prospective

Département
Développement durable

RAPPORTS
& DOCUMENTS

Juillet 2013

Quelques points à souligner sur les modèles de trafic et leurs usages

Contribution
Marc Gaudry

Tome 2

Rapport

« *L'évaluation socio-économique des investissements publics* »

Groupe de travail
présidé par Émile Quinet

Sommaire

1	Taille et qualité des échantillons	5
2	Faiblesses de <i>design</i> structurantes.....	6
3	Liens avec la localisation des activités et la sécurité civile.....	8
4	La norme européenne SPQR et la notion de pedigree des données et modèles.....	8
5	La centralisation des données publiques	10
6	La décentralisation des modèles	10
7	Références.....	11

Quelques points à souligner sur les modèles de trafic et leurs usages

L'objet de la présente note, présentée le 12 novembre 2012 en séance plénière, est de mettre en lumière certains points utiles à l'évaluation des projets qui dépendent des modèles de trafic. Ces points sont regroupés en six thèmes indiqués au tableau ci-dessous.

Objet et dimensions des thèmes traités dans cette note

	Questions particulières sur les modèles courants	Transparence des outils	Bibliothèques et gouvernance
Établissement des données (<i>data models</i>)	1. Taille et qualité des échantillons	4. La norme européenne SPQR et la notion de pedigree des données et modèles	5. Centralisation des données
Élaboration des modèles (<i>models of data</i>)	2. Faiblesses de design structurantes		6. Mise en concurrence des modèles et listes de vérification
Liens avec le système spatial	3. Liens avec la localisation des activités et la sécurité civile		

Source : Auteur

Il n'est pas question de résumer ici les pratiques variées de la modélisation du trafic dont le début, au moins pour les grands modèles, remonte à CATS (1959-1962). On a connu depuis un développement mondial des grands modèles dont le lien avec la théorie économique n'est pas toujours clair mais suscite des questionnements de plus en plus fréquents et documentés (*e.g.* Gaudry, 2007).

Nous ne reprendrons pas non plus les recommandations déjà élaborées par les rapporteurs, qui semblent sages, mais modifierons peut-être l'accent de certaines d'entre elles.

1 Taille et qualité des échantillons

Dans beaucoup de modèles, les échantillons utilisés semblent faibles.

Voyageurs : donnée urbaines. Considérons les enquêtes globales de transport (EGT). Orfeuil & Massot (2006) en sont satisfaits « pour les déplacements concernant Paris ». Mais se posent en fait à tout analyste plusieurs questions sur la précision et la fiabilité des recueils de données urbains, quand ils existent (l'Ile de France a la chance d'avoir une EGT, mais elle est restée améliorée et elle gagnerait à être plus fréquente), notamment concernant les déplacements effectués à pied, en deux-roues motorisés, en fin de semaine (samedi-dimanche), pour les motifs autres que le travail ou par les non-résidents. Mais la vraie question est en fait celle de la pertinence spatiale : dispose-t-on de données pertinentes pour estimer les flux origine-destination qui sont

à expliquer ? La fiabilité et la précision de ces données est-elle suffisante pour obtenir des estimations correctes ?

Dans d'autres villes, les échantillons sont considérables : à Montréal, il existe une enquête origine-destination de 4 % de la population (160 000 personnes) faite tous les 4 ou 5 ans depuis 1970. À York, municipalité de la banlieue nord de Toronto, l'enquête sur les déplacements est faite pour 500 000 individus pour une conurbation qui a une population plus faible que celle de Paris d'au moins 50 %.

Il faut donc appuyer l'initiative du STIF de former un comité pour examiner la question. On peut penser que les responsables de l'Enquête Nationale Transport 2007 pourraient contribuer de manière utile aux travaux.

Voyageurs : donnée interurbaines. Des améliorations sont en vue pour le rail : une matrice département-département (NUTS-3) serait en préparation. Pour l'avion, les données origine-destination n'existent pas et il est difficile de penser que la situation s'améliorera sans une réglementation adéquate : au Canada, Transports Canada et Statistique Canada échantillonnent 10 % des coupons aériens ; aux États-Unis, le Formulaire 41 « Form 41 » du *Bureau of Transport Statistics* (BTS) impose aux compagnies aériennes de fournir des détails considérables sur leurs activités, y compris par liaison.

La route pose un problème méthodologique important qui est bien connu. Plusieurs pays¹ construisent des matrices origine-destination à partir des comptages sur les segments du réseau, ce qui fournit une matrice bon marché facilement mise à jour ; on lui adjoint souvent des enquêtes sur les routes (*road side interviews* ou RSI) qui ajoutent d'autres informations, y compris sur le motif qui d'est pas décelable à partir de l'observation des seuls comptages.

Ces méthodes sont connues en France, au moins depuis Debaille (1977, 1979) et Danech-Pajouh (1986). Un document de fond (Gaudry, 1999), rédigé à la demande de la Commission Européenne pour encadrer les travaux du projet MYSTIC, fait le point sur les diverses méthodes utilisées depuis 1969.

Il faut donc appuyer les travaux du SETRA qui a débuté l'estimation d'une matrice O-D à partir des comptages routiers.

2 Faiblesses de *design* structurantes

Voyageurs. Le travail n'est plus le premier motif de déplacement : aux États-Unis, seulement 25 % des déplacements urbains sont faits pour ce motif ; en France, seulement 30 % des déplacements aériens courts courriers sont reliés au travail.

Dans ces conditions, le choix de destination devient très important et il faut modéliser la concurrence spatiale (on le fait dans le cadre urbain pour le motif achats). Or, dans les modèles de génération-distribution (sans doubles contraintes), les élasticités

(1) Dont le Canada, qui les utilise depuis 1978.

croisées entre flux sont toutes nulles, car le modèle est compatible avec l'axiome IIA du Luce¹...

Marchandises. Il faut comprendre l'évolution des tonnes et des tonnes-km depuis 1970. La distance moyenne parcourue augmentait avant 2000 et diminue apparemment depuis (excepté sur les voies navigables). Que se passe-t-il ?

Dans les modèles marchandises, l'étape de Distribution devient plus complexe : les structures logistiques sont mal comprises et mal modélisées. Faut-il par exemple une 5^{ème} étape entre la distribution et le choix modal des marchandises, comme chez Ying *et al.* (2004) qui utilisent 7 structures logistiques ? Cette innovation n'est pas sans analogies avec la formulation de tournées dans les modèles voyageurs (Antonin 2, par exemple).

Voyageurs et marchandises. Plusieurs problèmes sont communs à la modélisation des voyageurs et des marchandises. L'un d'entre eux est celui de la remontée dans le modèle de choix modal d'un indicateur adéquat de l'utilité dérivée de l'usage d'itinéraires multiples utilisés dans les réseaux de chaque mode.

Il y a belle lurette qu'on n'utilise plus la valeur moyenne des utilités modales pour expliquer la fréquence des déplacements, mais le *log-sum* de ces utilités. Mais on tarde à faire la même chose pour le choix entre itinéraires par mode. On peut montrer (Gaudry & Quinet, 2012) que l'usage de la moyenne sous-estime toujours l'utilité d'une quantité égale à la mesure d'information de Shannon (égale à moins une fois l'entropie) si le modèle de choix est Logit.

Cette pratique défavorise systématiquement les transports en commun. Parmi les exemples de sous-estimations probables de l'utilité des TC, on peut nommer les travaux sur l'évaluation du projet EOLE et la première vague (d'octobre 2012) des travaux sur le Métro express Grand Paris.

Le lien avec l'utilité. Dans beaucoup de modèles, le lien avec l'utilité n'est pas encore établi. Par exemple, on utilise des modèles de Poisson en accidentologie routière. Ce modèle n'est pas encore relié formellement à l'utilité mais on peut toutefois penser qu'il le sera : le *Logit-repeat* utilisé dans les modèles de tournées (par exemple Antonin 2) pour calculer la probabilité qu'ayant fait un déplacement on en fera un autre, est d'esprit poissonien. De plus, la valorisation du risque d'accident est elle-même exogène au modèle de choix modal : ce risque est traité comme un postiche, mais on peut penser qu'un indicateur de sécurité (par exemple le *logsum* des fréquences d'accident anticipées par catégorie de gravité) sera éventuellement intégré au modèle de choix modal, intégration qui fournira directement des valeurs de l'insécurité.

Cette absence de cohérence dans l'expression de l'utilité se rencontre aussi dans le traitement interne des modèles, par exemple dans le cas où un sous modèle utilise une fonction d'utilité, et un autre une autre.

¹ Un modèle est compatible avec l'axiome « Independence from Irrelevant Alternatives » de Luce (1959) si les rapports de probabilité de choix entre deux possibilités ne dépendent que de leurs caractéristiques propres et aucunement de celles des autres.

3 Liens avec la localisation des activités et la sécurité civile

Le modèle de trafic ne fonctionne pas dans le vide. Par exemple, le cœur de tout modèle LUTI est un modèle de trafic dont dépendront des effets sur la localisation des activités. Ce modèle devrait aussi servir à mesurer la résilience du réseau et à calculer des indices de redondance, si importants en cas de catastrophe.

Il ne faut pas penser que la régionalisation de la gestion des routes, par exemple, empêche l'État de conserver une surveillance de la redondance : on peut imaginer la désignation d'un Réseau inter-municipal (RIM) dont les caractéristiques garantissent un certain bien commun définissable par des mesures (multimodales) de redondance et d'accessibilité.

4 La norme européenne SPQR et la notion de pedigree des données et modèles

Le projet européen SPOTLIGHTS a défini une structure de documentation adaptée aux données et modèles, structure qui permet de fournir un pedigree à ces derniers¹ en adoptant pour les données (sources et dérivées) et pour les modèles (leurs contenus et statistiques dérivées) un quatrain d'étapes documentaires centrées sur la Saisie (S), la formulation de Postulats (P), la Quantification (Q) et la production de Résultats (R), étapes indiquées au tableau ci-après reproduit de Gaudry *et al.* (2002) comme les deux tableaux subséquents.

The SPQR quartet of dimensions applied to information objects

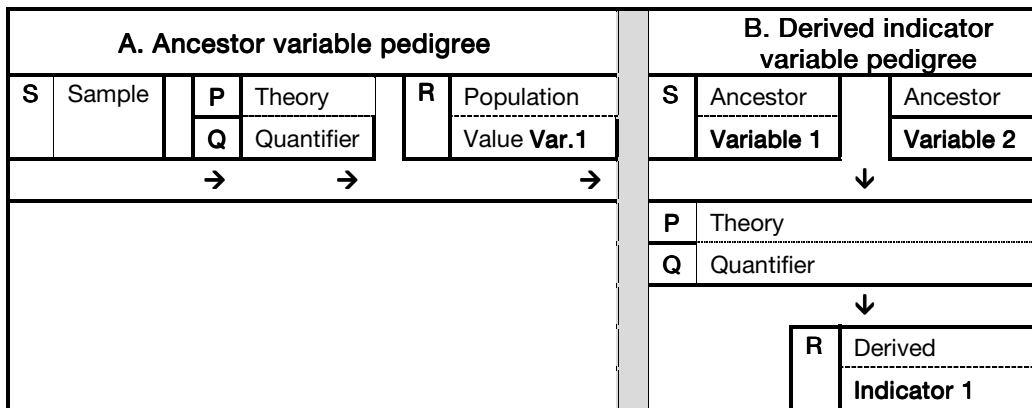
<i>Quartet</i>	<i>Step description</i>	<i>Idea of the step or dimension</i>
S	1. Input data sample	A <i>datum</i> is used; regression sample; (« <i>saisie</i> » de données)
P	2. Propose hypothesis	A vision (<i>theoria</i>) is formulated
Q	3. Quantify relationships	A functional relationship is calculated
R	4. Report output results	A set of raw results is obtained

Source : Gaudry *et al.* (2002)

Pour les variables (data models): la séquence est appliquée soit à une variable ancêtre, soit à une variable dérivée, tel qu'indiqué au tableau ci-dessous. Le sens des opérations successives d'élaboration d'une ou de plusieurs variables ancêtres, et ensuite de leur combinaison par des opérations diverses, est intuitivement clair. Il n'y a pas de données neutres : elles résultent toujours d'un processus de construction qu'il y a lieu de documenter.

(1) En France, le Centre Quetelet a aussi développé des normes pour les bases de données.

Pedigree elements for ancestor and derived variable indicators



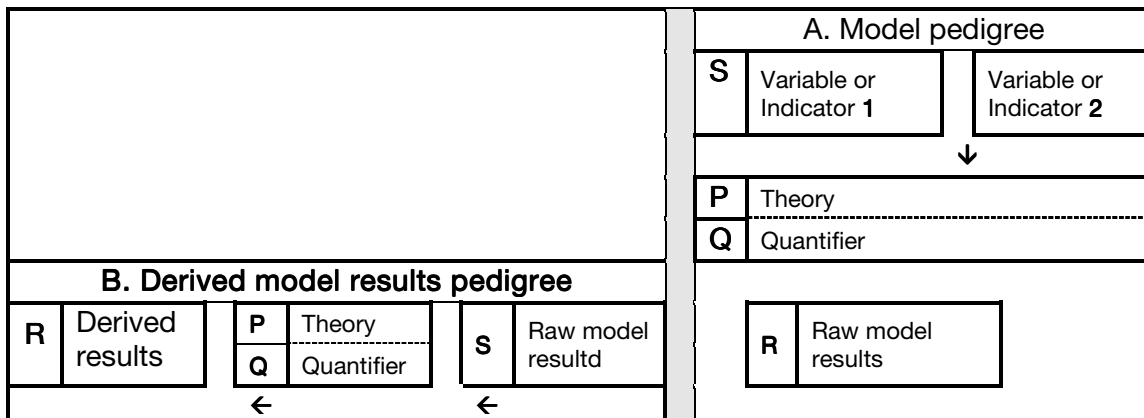
Source : Gaudry et al. (2002)

Pour les modèles (models of data) : la même séquence documentaire est applicable soit à l'estimation d'un modèle fondé sur l'usage de données, modèle qui produit ses résultats en propre, soit à la production de diverses statistiques à partir des résultats du modèle.

Dans un modèle de régression Logit, par exemple, on estimera à l'étape A les coefficients et autres paramètres des variables des fonctions d'utilité par une certaine méthode Q (e.g. la maximisation de la Vraisemblance) implantée dans un algorithme particulier, ce qui produira des résultats R (les paramètres estimés). À l'étape B, ces estimations pourront servir comme entrées S au calcul d'élasticités des probabilités de choix R obtenues selon certaines hypothèses P et procédures Q (par exemple d'évaluation à la moyenne de l'échantillon) appliquées aux sorties du modèle d'estimation.

Les élasticités sont donc des constructions conditionnelles documentables : elles ne sont généralement pas des sorties du modèle lui-même mais des résultats dérivés. Dans un modèle de régression classique, on s'intéressera par exemple fréquemment à diverses constructions complexes dérivables du modèle, comme l'élasticité de divers moments de la variable dépendante.

Pedigree elements for models and derived model results



Source : Gaudry et al. (2002)

L'idée générale de la norme SPQR est de fournir, dans un cadre conceptuel uniforme, une information systématique minimale qui assure aux données et modèles publics la transparence des procédures qui ont conduit leur élaboration et à leurs résultats. Il s'agit de favoriser une « bonne pratique » dans un environnement où les produits financés sur deniers publics ne peuvent et ne doivent pas constituer des « boîtes noires ».

5 La centralisation des données publiques

Comment ne pas avoir de grandes bases publiques, documentées et accessibles à tous ? Est-il normal que les étudiants de Nanterre travaillent sur les données du BART de San Francisco et pas sur celles de Paris ? Pourquoi le modèle Antonin ne serait-il pas ouvert, documenté en long et en large¹, modifiable et opposable, et ses données aussi accessibles que celles du BART ?

6 La décentralisation des modèles

Une autre bonne pratique consiste à mettre les modèles en concurrence : c'est ce qu'a fait VIA RAIL Canada pour les études de grande vitesse ferroviaire en 1989 (en mettant 3 modèles en concurrence) et ce que cette entreprise publique fait en 2012-2013 pour une étude grande vitesse du marché Montréal-Toronto (avec deux modèles qui doivent utiliser la même base de données).

La Commission qui a étudié le projet de grande vitesse ferroviaire en Floride en 1998 a aussi mis en concurrence 3 modèles.

(1) On ne peut pas dire que STIF (2004) remplisse ces conditions.

7 Références

CATS, Chicago Area Transportation Study, (1959-1962). *Final report in three parts*. Study conducted under the sponsorship of [the] State of Illinois, Department of Public Works and Buildings, County of Cook, Board of Commissioners of Cook County. CONTENTS: v. 1, *Survey findings*, December, 1959; v. 2, *Data projections*, July, 1960; *Transportation plan*, April, 1962. Chicago.

Danech-Pajouh, M. (1986). *Estimation des matrices origine-destination par les comptages et la théorie de l'information*. Rapport INRETS n° 7, 23 p., juin.

Debaille, S. (1977). *Un modèle de reconstitution de matrices origine-destination en milieu urbain*. Rapport I.R.T. N° 27, Institut de Recherche des Transports, Décembre 1977.

Debaille, S. (1979). *Reconstitution de matrices origine-destination en milieu urbain : le modèle NEMROD*. Bulletin Recherche-Transports 26, 13-23, Institut de Recherche des Transports, Avril 1979.

Gaudry, M. (1999). *SPQR: the Four Approaches to Origin-Destination Matrix Estimation : Some Considerations for the MYSTIC Research Consortium*. Publication CRT-99-35, Centre de recherche sur les transports (CRT), Université de Montréal, Working Paper N° 2000-10, Bureau d'économie théorique et appliquée (BETA), Université Louis Pasteur, 23 p., Septembre.

Gaudry, M. (2007). Structure de la modélisation du trafic et théorie économique, Ch. 1 dans Maurice, J. & Y. Crozet, (dir.) *Le calcul économique dans le processus de choix collectif des investissements de transport*, 6-97, Economica.

Gaudry, M., Quinet, E. (2012). Shannon's measure of information, path averages and the origins of random utility models in transport itinerary or mode choice analysis. Publication AJD-142, Agora Jules Dupuit, Université de Montréal, 32 p., April.

Gaudry, M., Nielsen, O., Tsamboulas, D., Walker, W., Willumsen, L. (2002). *Dcode: the SPQR pedigree form requirement proposal for European Transportation policy Information System (ETIS) recognised data and models*. Publication AJD-9, Agora Jules Dupuit, Université de Montréal, 29 p, May.

Luce, R.D. (1959). *Individual Choice Behavior: A Theoretical Analysis*. John Wiley & Sons, Inc., New York, ou Dover Publications, Inc., Minelola, New York, 2005.

Orfeuil, J.-P., Massot, M.-H. (2006). *Regards sur la plaquette « Bilans des déplacements » à Paris*. Rapport déposé auprès du Groupe de travail sur les transports et les déplacements du Conseil scientifique de la Mairie de Paris et du Conseil scientifique de la région Île-de-France, 28 p., septembre.

STIF (2004). *Prévisions de trafic régional sur l'Île de France : Fonctionnalités des modèles- Méthodologie, Cas du STIF avec le modèle Antonin*, 16 p.

Ying, Y., Williams, I., Shahkarami, M. (2004). *Integrated regional economic and freight logistics modelling: Results from a model for Trans-Pennine corridor*.