



HAL
open science

Analyse et mesure de l'incertitude dans un modèle de simulation. Les principes, une méthode et l'exemple de l'affectation bicritère du trafic

Fabien Leurent

► **To cite this version:**

Fabien Leurent. Analyse et mesure de l'incertitude dans un modèle de simulation. Les principes, une méthode et l'exemple de l'affectation bicritère du trafic. Génie logiciel [cs.SE]. Ecole Nationale des Ponts et Chaussées, 1997. Français. NNT: . tel-00529468

HAL Id: tel-00529468

<https://pastel.hal.science/tel-00529468>

Submitted on 25 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Fabien LEURENT

X

Thèse de Doctorat de l'Ecole Nationale des Ponts et Chaussées

SPÉCIALITÉ TRANSPORT

Analyse et mesure de l'incertitude dans un modèle de simulation

**Les principes, une méthode et l'exemple de
l'affectation bicritère du trafic**

Soutenue le 29 janvier 1997 devant le jury composé de

Pierre MERLIN, Professeur de l'ENPC, Président du jury

Gabriel DUPUY, Professeur de l'ENPC, Directeur de thèse

Michael FLORIAN, Professeur de l'Université de Montréal, Rapporteur

Olivier MORELLET, Directeur de Recherche de l'INRETS, Rapporteur

Philippe TOINT, Professeur de l'Université de Namur, Examineur



à André Broto et Roger Marche

en souvenir de vos enseignements, votre curiosité intellectuelle, vos intuitions, qui sont à l'origine de ce travail. En remerciement pour vos conseils et votre bienveillance.

Je suis redevable à bien d'autres personnes encore d'idées, de conseils et de soutien. A Gabriel Dupuy, mon Directeur de thèse, qui sans se contenter des aspects formels et algorithmiques m'a engagé à traiter les aspects conceptuels et statistiques. A Olivier Morellet, qui a suivi continûment mon travail, pour des discussions fructueuses, parfois vives mais toujours amicales. A Michael Florian et Philippe Toint pour leurs encouragements et leur participation au jury, ainsi qu'à Pierre Merlin qui l'a présidé fermement.

Mon travail s'est nourri d'échanges avec des praticiens, échanges tant d'idées que de données : je salue ici Jean Delons (et sa société COFIROUTE), Pierre Baillet et Gérard Louah (CETE du Sud-Ouest et de l'Ouest respectivement), Olivier Nalin (SETRA).

Enfin il m'est agréable de remercier mes parents, mes frère et soeurs, et Nathalie ma femme, pour leurs encouragements et leurs incitations à la clarté !

Analyse et mesure de l'incertitude dans un modèle de simulation

Problématique : un modèle est-il proche de la réalité ?

Un modèle est une représentation d'un système. Dans cette représentation figurent des éléments, entre lesquels s'exercent des interactions. Le modèle sert à analyser (fonction de connaissance) et simuler les interactions entre éléments. La simulation peut être qualitative ou quantitative : il s'agit de déduire les conséquences "rationnelles" d'hypothèses sur les variables d'état du système, ou sur des variables de contrôle ou de perturbation.

Comment juger, ou mieux mesurer, si le modèle est proche de la réalité ? L'audit technique d'un modèle consiste à détecter, identifier, et si possible réduire, les erreurs et les incertitudes attachées au modèle.

Objectif : un audit pour porter un jugement technique

J'ai élaboré une méthode systématique d'audit d'un modèle, afin de donner ou refuser des garanties techniques. Un tel audit nécessite la transparence tant des hypothèses internes au modèle, que des hypothèses externes liées à une application particulière.

Méthode : entreprendre l'audit dans autant de dimensions que de types d'erreur dans un modèle

L'erreur dans un modèle est de nature plurielle : nous la décomposons en (i) erreur de *conception* (quels mécanismes explicatifs constituent la *composition conceptuelle*, quelles approximations dans la définition des variables ?), (ii) erreur *formelle* (la synthèse en langage mathématique des mécanismes explicatifs dans une *formule caractéristique*, doit être conforme au contenu conceptuel et cohérente), (iii) erreur *algorithmique* (le dispositif de résolution fournit-il vraiment une solution ? Avec quelle précision ?), (iv) enfin une incertitude de type *économétrique* qui englobe l'erreur d'estimation et l'erreur exogène a priori sur les inputs. L'incertitude de type économétrique peut être quantifiée, après propagation de l'erreur exogène à travers le modèle de simulation.

Tab. 1. Dimensions de l'audit d'un modèle.

Catégorie d'aspect	Explicitation	Disciplines impliquées	Types d'erreurs
Composition conceptuelle	= ensemble d'éléments et de relations entre éléments	Analyse systémique	Erreur de conception
Formule caractéristique	= ensemble de conditions caractéristiques (équations...)	Mathématiques appliquées	Erreur de formulation, par incohérence ou faute logique
Procédure de résolution	Produire une solution par un dispositif ad hoc	Physique, algorithmique, informatique	Erreur de résolution, dont erreur de convergence
Alimentation économétrique	Spécification des fonctions de dépendance et des inputs	Econométrie	Erreur économétrique, dont erreur d'estimation et erreur exogène

Les quatre types d'erreur relèvent chacun d'une discipline scientifique appropriée, d'où la nécessité d'entreprendre l'audit dans autant de dimensions : conceptuelle, formelle, algorithmique, économétrique.

Produit

S'appuyant sur notre analyse théorique de l'erreur dans un modèle, nous avons développé une méthode pluridimensionnelle d'audit (tableau 2).

Tab. 2. Une méthode pluridimensionnelle d'audit.

Dimension	Teneur de l'audit	Discipline	Résultat
Conceptuelle	Explicitation des hypothèses Les discuter et les comparer à l'état de l'art	Analyse systémique	Appréciation de la portée explicative Relativisation du modèle
Formelle	Vérifier la formalisation du contenu conceptuel et l'obtention de la formule caractéristique (FC) Etudier l'existence et l'unicité d'une solution à la FC	Logique mathématique Recherche opérationnelle	Preuve que la FC est conforme au contenu conceptuel Cohérence de la FC donc existence d'une solution Résultats d'unicité
Algorithmique	Vérifier que l'on résout effectivement la FC Vérifier la fiabilité du dispositif de résolution	Physique Informatique Algorithmique	Test rigoureux de convergence pour l'algorithme Capacité de l'algo. à annuler l'erreur de calcul
Econométrique	Quantifier l'erreur exogène a priori (EEA) sur les inputs et les paramètres Propager l'EEA via la FC	Econométrie Recherche opérationnelle	Appréciation de l'EEA propagée Procédure d'estimation de l'EEA

Les tentatives antérieures se cantonnaient aux aspects algorithmiques ou économétriques, avec parfois une allusion au "jugement a priori" pour évaluer le contenu conceptuel. Nous avons démontré le bien fondé d'un audit selon les quatre dimensions, en traitant l'exemple de l'affectation du trafic. Car un tel modèle représente un équilibre entre offre et demande, équilibre dont la formulation et la résolution nécessitent validation.

Perspectives

Après audit d'un modèle, on peut : apprécier la portée explicative, éliminer l'erreur formelle, annuler l'erreur de résolution, quantifier l'erreur exogène propagée donc évaluer le risque statistique.

L'audit se prépare utilement par une capitalisation des concepts, des formulations, des algorithmes et des méthodes d'estimation. Une telle capitalisation produit de grandes économies d'échelle pour l'audit de modèles particuliers.

MOTS-CLEFS

Inéquation variationnelle ; Affectation ; Audit ; Bicritère ; Econométrie ; Erreur ; Modèle ; Simulation ; Recherche opérationnelle ; Transport

Avant-propos

Parfois la modélisation est un exercice périlleux et nécessite une certaine témérité. La lecture également ! Ô lecteur, te voilà prévenu. Le sujet de ma thèse est passablement complexe, je l'ai traité de la façon qui m'a semblé la plus appropriée. Si les formules mathématiques te rebutent, garde-toi de la troisième partie... les deux premières parties et le bilan expliquent suffisamment ma position du problème et mes réponses.

J'ai traité principalement la question de comment prendre des garanties vis-à-vis d'un modèle de simulation. Ma réponse comprend deux volets, relatifs l'un aux aspects techniques et l'autre aux aspects "sémantiques", à la signification de la connaissance incluse dans le modèle. Cette connaissance est-elle scientifique ? Je l'ai supposé a priori, sans m'intéresser au statut philosophique de la connaissance et de la science. Encore que mon modèle de l'erreur en quatre parties : conception, formulation, résolution, exogène, s'apparente fortement au modèle aristotélicien de la cause, décomposée en cause finale, cause formelle, cause efficiente et cause matérielle (Cf. pour une maison : la fonction d'habitation, le plan d'architecte, le maçon et les matériaux de construction). La parenté tient à l'essence causale d'un modèle !

Quant aux aspects techniques d'un modèle, qu'ils relèvent des mathématiques, de la physique, de l'algorithmique ou de l'économétrie, j'en traite autant mais pas plus que nécessaire à la question de l'audit ou à l'exemple de l'affectation. J'emploie pour l'essentiel des éléments exposés dans des références auxquelles je renvoie. Les éléments originaux de la thèse comprennent le modèle d'un modèle et le modèle de l'incertitude dans un modèle, mes contributions techniques à l'affectation sont déjà publiées ailleurs.

Introduction.....	1
Splendeur et misères des modèles.....	1
Traquer l'erreur pour la mesurer et si possible la réduire.....	2
Une méthode à la croisée de plusieurs disciplines.....	2
Organisation de la thèse	3
Partie I - Contexte et prémices	5
1. L'exemple de l'affectation.....	7
1.1 Introduction aux modèles d'affectation	8
1.1.1 Définition provisoire	8
1.1.2 Mise en perspective.....	8
1.1.3 Fonctions et aptitudes : selon la composition sémantique.....	11
1.2 Variations sur un problème de choix d'itinéraire	13
1.2.1 Comment agir sur la congestion ?.....	13
1.2.2 Mieux représenter le comportement de la demande.....	15
1.2.3 Bilan	17
1.3 Codage, calage, trucages, dérapages	17
1.3.1 Codage et calage : de nécessaires arbitrages	17
1.3.2 Quelques trucages	19
1.3.3 Divers dérapages.....	19
1.3.4 Bilan provisoire.....	20
1.4 Les logiciels, instruments des modèles d'affectation	20
1.4.1 Contenu et forme des logiciels d'affectation.....	20
1.4.2 Le marché des logiciels d'affectation	22
1.4.3 Critères pour choisir un logiciel d'affectation	23
1.5 Qu'y a-t-il à auditer ?.....	24
2. Pourquoi auditer ?	25
2.1 Pour ou contre un modèle ? Selon les garanties données !	26
2.1.1 Pour définir et qualifier un modèle	26
2.1.2 Arguments en faveur d'une modélisation	27
2.1.3 Arguments opposés aux modèles.....	28
2.1.4 Synthèse : des garanties sont nécessaires.....	30
2.2 L'audit pour prendre des garanties.....	31
2.2.1 Quelles garanties ?.....	31
2.2.2 Pourquoi auditer l'affectation du trafic ?.....	32
2.2.3 Contribution de la thèse.....	33
2.3 Aspects sociaux et instrumentaux	34
2.3.1 Plan social et institutionnel	34
2.3.2 Plan instrumental.....	35

3.	Antécédents et ressources.....	41
3.1	Le discours économétrique classique	42
3.1.1	<i>Types d'erreur dans la théorie statistique standard.....</i>	<i>42</i>
3.1.2	<i>Quels traitements de l'erreur ?</i>	<i>43</i>
3.1.3	<i>Règles de sélection des modèles</i>	<i>44</i>
3.1.4	<i>Bilan</i>	<i>45</i>
3.2	Questions de spécification.....	46
3.2.1	<i>Le codage, face cachée de la modélisation ?.....</i>	<i>46</i>
3.2.2	<i>Les mécanismes explicatifs</i>	<i>48</i>
3.3	Questions d'estimation	50
3.3.1	<i>Objets et principes méthodologiques de l'estimation</i>	<i>50</i>
3.3.2	<i>Etat de la pratique</i>	<i>51</i>
3.3.3	<i>Du côté de la recherche</i>	<i>52</i>
3.4	Autres contributions	54
3.4.1	<i>Une inconnue : l'erreur de calcul.....</i>	<i>54</i>
3.4.2	<i>Sur la propagation de l'erreur exogène.....</i>	<i>55</i>
3.5	Bilan provisoire	57
Partie II - Théorie et méthode		59
4.	Analyse théorique de l'erreur.....	61
4.1	Plusieurs sources alimentent l'erreur	62
4.1.1	<i>Recensons les sources d'erreur.....</i>	<i>62</i>
4.1.2	<i>L'exemple du modèle d'affectation bicritère</i>	<i>63</i>
4.1.3	<i>Les erreurs se cumulent !.....</i>	<i>64</i>
4.1.4	<i>Discussion</i>	<i>65</i>
4.2	Pour mesurer et traiter l'erreur, décliner l'audit	67
4.2.1	<i>Comment mesurer l'erreur ?</i>	<i>68</i>
4.2.2	<i>Comment traiter l'erreur ?</i>	<i>69</i>
4.2.3	<i>Bilan : explorer, assécher, endiguer, canaliser</i>	<i>71</i>
4.3	Propager l'erreur via le modèle.....	71
4.3.1	<i>Définition du biais et de la marge d'incertitude.....</i>	<i>71</i>
4.3.2	<i>Comment le modèle propage le biais et l'incertitude.....</i>	<i>72</i>
4.3.3	<i>Le cas d'un modèle bicritère simple.....</i>	<i>74</i>
4.3.4	<i>Application numérique.....</i>	<i>76</i>
4.4	Sophistiquer le mécanisme du modèle.....	76
4.4.1	<i>Sophistiquer pour accroître la portée explicative.....</i>	<i>77</i>
4.4.2	<i>Fondement statistique : réduire la variabilité non-expliquée.....</i>	<i>77</i>
4.4.3	<i>L'exemple du réseau à deux arcs (suite).....</i>	<i>78</i>
4.4.4	<i>Quel proportionnement de l'objectif aux contraintes ?</i>	<i>79</i>
5.	Une méthode systématique d'audit.....	81
5.1	Discours de l'audit conceptuel.....	82

5.1.1	<i>L'audit conceptuel et sa justification</i>	82
5.1.2	<i>Comment mesurer et traiter l'erreur de conception ?</i>	83
5.1.3	<i>Plan d'application de la méthode d'exploration systématique</i>	85
5.2	Discours de l'audit formel	86
5.2.1	<i>L'audit formel et sa justification</i>	86
5.2.2	<i>Comment mesurer et traiter l'erreur de formulation ?</i>	89
5.2.3	<i>Pour préparer l'audit formel dans un domaine spécifique</i>	90
5.3	Discours de l'audit algorithmique	91
5.3.1	<i>Une analyse de l'erreur de calcul</i>	91
5.3.2	<i>L'audit algorithmique et sa justification</i>	92
5.3.3	<i>Fautes d'ordre matériel ou logiciel</i>	93
5.3.4	<i>Mesurer et traiter les imprécisions numériques</i>	94
5.3.5	<i>Mesurer et traiter l'erreur de convergence</i>	95
5.3.6	<i>Synthèse</i>	98
5.4	Discours de l'audit économétrique	99
5.4.1	<i>Définir et qualifier l'alimentation économétrique</i>	99
5.4.2	<i>Définir et justifier l'audit économétrique</i>	101
5.4.3	<i>Teneur de l'audit économétrique</i>	103
5.4.4	<i>Pour préparer l'audit économétrique dans un domaine spécifique</i>	105
5.5	Quelles articulations entre les dimensions ?	106
5.5.1	<i>Processus naturels d'audit, dans le développement et l'application d'un modèle</i>	106
5.5.2	<i>Les interactions entre dimensions conduisent à des arbitrages</i>	107
Partie III - Application à l'affectation bicritère		109
6. Plan d'audit de l'affectation bicritère		111
6.1	Caractéristiques de l'affectation	112
6.1.1	<i>Toutes les sources d'erreur sont présentes dans l'affectation</i>	112
6.1.2	<i>Deux complexités inhérentes</i>	113
6.2	Un investissement lourd, à rentabiliser	114
6.2.1	<i>Couvrir toutes les dimensions</i>	115
6.2.2	<i>Comment rentabiliser l'investissement ?</i>	116
6.3	Détail du mode opératoire	118
6.3.1	<i>Un audit dans chaque dimension</i>	118
6.3.2	<i>Une entreprise à plusieurs niveaux</i>	119
6.3.3	<i>Plan de la suite de la partie</i>	119
7. L'audit sémantique : un diagnostic		121
7.1	Composants conceptuels pour l'affectation	122
7.1.1	<i>Hypothèses physiques</i>	122
7.1.2	<i>Notion de coût généralisé</i>	123
7.1.3	<i>Du côté de l'offre</i>	124
7.1.4	<i>Du côté de la demande</i>	125

7.1.5	<i>Des échelles de valeur pour une synthèse</i>	126
7.2	Exemple du modèle d'affectation bicritère	127
7.2.1	<i>Hypothèses physiques</i>	127
7.2.2	<i>Hypothèses sur le coût généralisé</i>	127
7.2.3	<i>Représentation de l'offre</i>	128
7.2.4	<i>Représentation de la demande</i>	129
7.2.5	<i>Exploration systématique</i>	130
7.3	Pertinence et proportionnement	134
7.3.1	<i>Choisir la composition conceptuelle</i>	134
7.3.2	<i>Choisir le format économétrique</i>	135
8.	L'audit formel : une vérification	137
8.1	Ressources formelles pour l'affectation statique	138
8.1.1	<i>Formalisation</i>	138
8.1.2	<i>Formulation en problème type de recherche opérationnelle</i>	141
8.1.3	<i>Questionnement</i>	143
8.2	Formule caractéristique en dimension finie	145
8.2.1	<i>Formule bicritère pour le réseau à deux arcs</i>	145
8.2.2	<i>Extensions sémantiques "en profondeur"</i>	150
8.2.3	<i>Extensions sémantiques "en largeur"</i>	157
8.2.4	<i>Propriétés remarquables des formules caractéristiques</i>	168
8.3	Formule caractéristique en dimension infinie	175
8.3.1	<i>Illustration</i>	175
8.3.2	<i>Formalisation</i>	178
8.3.3	<i>Formules caractéristiques</i>	180
8.4	Commentaires	182
8.4.1	<i>Enseignements de l'expérience d'audit formel</i>	182
8.4.2	<i>La formulation, charnière entre la conception et la quantification</i>	183
9.	L'audit algorithmique : une vérification	185
9.1	Ressources pour l'affectation statique du trafic routier	186
9.1.1	<i>L'erreur de convergence</i>	186
9.1.2	<i>Etat de l'art</i>	187
9.1.3	<i>Inventaire de critères pour mesurer la convergence</i>	189
9.1.4	<i>Inventaire d'algorithmes convergents</i>	196
9.1.5	<i>Récapitulation des critères et des algorithmes, par modèle d'affectation</i>	199
9.2	Application au modèle bicritère	201
9.2.1	<i>Algorithmes et critères</i>	201
9.2.2	<i>Applications numériques</i>	205
9.2.3	<i>Éléments de bilan</i>	212

10. L'audit économétrique : une quantification	213
10.1 Ressources économétriques pour l'affectation.....	214
10.1.1 Qualification économétrique du domaine.....	214
10.1.2 Modèles d'estimation, selon l'objet.....	215
10.1.3 La méthode de Tobin et Friesz pour propager l'erreur exogène à travers le modèle d'affectation.....	220
10.2 Ordres de grandeur pour l'incertitude exogène	223
10.2.1 Fonctions de temps de parcours.....	223
10.2.2 Flots origine-destination.....	227
10.2.3 Valeur du temps	228
10.3 Propager l'erreur exogène via le modèle bicritère	238
10.3.1 Le cas du réseau à deux arcs.....	238
10.3.2 Le cas général.....	248
10.3.3 Un traitement numérique dans le cas général.....	253
10.4 Application à une vraie étude de trafic	254
10.4.1 Le site et le modèle de trafic.....	254
10.4.2 L'erreur exogène et sa propagation.....	255
11. Bilan et perspectives	257
11.1 Récapitulation du sujet.....	258
11.1.1 Niveau des principes et de la méthode.....	258
11.1.2 Préparer l'audit dans un domaine, c'est capitaliser !.....	261
11.1.3 Etat de développement du modèle bicritère.....	262
11.2 L'audit en pratique	264
11.2.1 Vers un modèle certifié ?.....	264
11.2.2 Fonctions de la capitalisation.....	264
11.3 Pistes de recherche.....	265
11.3.1 Du côté de la méthode d'audit.....	265
11.3.2 Capitaliser l'affectation dynamique et aux modes non automobiles	266
En conclusion.....	269
Contribution de la thèse	269
<i>Des concepts et une méthode pour décrire et auditer un modèle.....</i>	<i>269</i>
<i>Dans le domaine de l'affectation.....</i>	<i>269</i>
Modélisation et audit.....	270
<i>Quelles garanties techniques ?.....</i>	<i>270</i>
<i>Du bon usage de l'audit</i>	<i>271</i>
<i>Modéliser, avec ou sans garantie.....</i>	<i>271</i>
Références	273
English summary.....	281
Glossaire des sigles.....	297
Liste des figures et des tableaux.....	299

Introduction

SPLendeur ET MISÈRES DES MODÈLES

Lorsque Gaspard Monge (1746-1818) résolut analytiquement le problème du transport au moindre coût de matériaux via deux ponts sur un fleuve, fondant ainsi la recherche opérationnelle (1776 ; rapporté par Etnier 1987), il n'imaginait probablement pas la fortune future des modèles de simulation.

Un modèle de simulation sert à étudier un objet ou un système dans des conditions spéciales parce que spécifiées, contrôlées, par l'analyste. Comment étudier ou contrôler le monde des choses ? Seules celles qui existent peuvent être empiriquement appréhendées, leur observation est souvent très difficile sinon impraticable, peu d'entre elles finalement se prêtent à des actions commandées. Le modèle repousse certaines limites pratiques de la connaissance et de l'action : il articule le monde des choses et le monde des idées pour appuyer l'imagination sur les choses connues et réciproquement imaginer les choses inconnues.

Selon le domaine d'application, bien des formes sont possibles. Prenons-en pour exemples les équations de la physique, les formules des réactions chimiques, les plans des architectes, les simulations numériques des conditions météorologiques ou des réactions d'un corps à des chocs, les cultures biologiques, la maquette d'avion employée en soufflerie, les topiques en psychologie, les théories sociales et économiques... la liste est pléthorique ! En une génération l'avènement de l'ordinateur, véritable cheval de la pensée, a ouvert des possibilités prodigieuses à la simulation numérique, désormais susceptible d'emploi dans tous les domaines techniques. Ses applications concrètes sont multiples, en particulier dans l'industrie et l'économie.

Cependant l'emploi d'un modèle donne seulement une présomption de qualité à la simulation, non une preuve. Combien de prévisions se sont révélées fausses !

La confusion entre présomption et preuve est volontairement entretenue dans certains débats sociaux ou institutionnels, où des acteurs arguent des résultats d'un modèle (parfois purement inventés, ou d'autres fois contrefaits !) pour soutenir leur discours et impressionner leurs interlocuteurs. D'où par réaction une mise en cause de l'emploi des modèles, appelons-la *critique du sociologue*. Dans certains domaines, en particulier la planification des transports urbains en France, cette critique a considérablement freiné

l'effort de modélisation.

Sur un plan purement technique, la qualité du modèle est non seulement un enjeu scientifique, mais parfois aussi un enjeu économique important lorsque les résultats déterminent de lourds investissements. Nous appelons *critique du financier* la critique technique de la qualité du modèle, afin de marquer le poids des enjeux économiques liés à cette question. En particulier, la critique du financier est à l'origine de notre propre intérêt pour la modélisation, éveillé lors d'un stage dans une société privée concessionnaire d'autoroutes à péage (Cf. §6.2.1.1).

TRAQUER L'ERREUR POUR LA MESURER ET SI POSSIBLE LA RÉDUIRE

Notre thèse a pour finalité de poser et résoudre le problème de l'erreur dans un modèle. Nous reprenons à notre compte la critique du financier et nous la développons de façon constructive. Le problème de l'erreur s'analyse au travers de plusieurs questions :

- quel est l'objet ou le système modélisé ? Quels en sont les traits essentiels, sont-ils ou non représentés dans le modèle ? Le modèle ajoute-t-il des caractères qui n'appartiennent pas à l'objet étudié ?
- le modèle est-il cohérent, exempt de contradiction interne ?
- si le modèle livre des résultats quantitatifs, quelle incertitude leur est attachée ?

Cette liste met en évidence la nature plurielle de l'erreur, qui peut être d'origine sémantique (lien entre objet et modèle) ou syntaxique (cohérence interne) ou économétrique (incertitude sur les résultats quantitatifs). Chaque nature d'erreur relève d'une ou plusieurs disciplines scientifiques particulières : analyse systémique pour l'erreur sémantique, logique mathématique pour l'erreur syntaxique, statistique pour l'erreur économétrique. Aussi notre thèse se situe-t-elle à la croisée de ces disciplines. Elle les articule au service d'un but principal, l'identification de l'erreur et sa mesure, et d'un but accessoire, réduire l'erreur tant que possible.

Nous appelons *audit technique d'un modèle* l'entreprise d'identifier et de mesurer l'erreur. Son produit attendu est une connaissance et une évaluation des qualités et défauts du modèle. A priori, l'identification et la mesure de l'erreur répondent de façon satisfaisante à la critique du financier puisqu'elles lui permettent d'évaluer le risque, donc de le couvrir ou le rémunérer en conséquence.

Ainsi averti de l'amplitude possible de l'erreur, le financier désire probablement la réduire : d'où le but accessoire d'indiquer des moyens pour réduire l'incertitude.

UNE MÉTHODE À LA CROISÉE DE PLUSIEURS DISCIPLINES

Nous avons signalé la nature plurielle de l'erreur et la diversité des disciplines scientifiques impliquées. Notre méthode d'audit est donc pluridisciplinaire par nature. D'avance, nous prions le lecteur d'excuser les imprécisions par rapport aux discours spécialisés de chaque discipline. Nous avons notamment donné à certains termes un sens unique tout au long de la thèse, alors qu'ils changent de sens selon la discipline de modélisation. En particulier, en économie un modèle est réputé "comportemental" s'il reproduit la logique de comportement des acteurs microéconomiques, tandis qu'en

traitement du signal le terme équivalent est "phénoménologique", l'adjectif "comportemental" qualifiant une reproduction purement numérique et non causale !

Un examen des diverses disciplines, en particulier la recherche opérationnelle et l'économétrie, nous a amené à postuler une forme unique pour tout modèle, indépendamment des aspects de présentation ou de support d'application. Cette forme générale comprend quatre éléments : la *composition sémantique*, la *formule caractéristique*, le *dispositif de résolution*, l'*alimentation économétrique*.

Aussi avons-nous mené notre analyse de l'erreur de la façon la plus générale qu'il nous a été possible. La méthode d'audit que nous proposons est vraisemblablement moins générale, elle repose sur des hypothèses particulières quant au dispositif de résolution ; cependant nous pressentons que tout dispositif de résolution (maquette, simulation analogique, simulation numérique...) est lié à l'objet modélisé par un système de formules dont le modélisateur fait un usage explicite ou implicite. Cela rend possible d'adapter le discours que nous avons concentré sur la simulation numérique.

Pour démontrer la viabilité de notre méthode d'audit, nous l'avons appliquée à un modèle d'affectation bicritère du trafic : un tel modèle représente les effets de mesures d'offre (planification, exploitation) sur l'usage d'un réseau de transport par les auteurs des déplacements. On l'emploie en particulier pour simuler le trafic et la recette de projets d'infrastructure à péage (route, train ou avion...). Le modèle considéré a pour trait essentiel la différenciation des déplacements selon les arbitrages auxquels procèdent leurs auteurs entre le temps et le prix des chemins possibles.

Le rattachement de notre thèse à l'économie des transports se justifie par les raisons que voici. D'abord les motivations économiques de la critique du financier. Ensuite la dimension économétrique de l'audit d'un modèle. Enfin l'emploi pratique du modèle traité en exemple, au service de la connaissance et de la simulation de problèmes de transport. Cela étant, la thèse se réfère principalement à des théories et méthodes des mathématiques appliquées : logique mathématique et recherche opérationnelle dans la dimension formelle de l'audit, algorithmique dans la dimension algorithmique, économétrie et recherche opérationnelle encore dans la dimension économétrique. Nous avons tenté d'éviter tout jargon ou toute sophistication inutile ; cependant le traitement particulier de l'affectation bicritère a nécessité des développements probablement indigestes pour un lecteur peu familier des mathématiques.

ORGANISATION DE LA THÈSE

La thèse est articulée en trois parties, respectivement intitulées *Prémices et ressources*, *Théorie et méthode*, *Application à l'affectation bicritère*.

La première partie, **Prémices et ressources**, justifie notre projet d'audit technique. D'abord un premier chapitre développe *l'exemple de l'affectation*, afin tant de présenter les modèles d'affectation que de mettre en évidence les problèmes liés à leur application. Le deuxième chapitre, *Pourquoi auditer ?*, fait une synthèse des arguments pour ou contre les modèles et précise notre propre position, à savoir qu'il faut connaître et circonscrire l'erreur dans un modèle de simulation afin d'en délimiter la portée théorique et quantitative. Le troisième chapitre, *Antécédents et ressources*, examine les traitements de l'erreur proposés par la théorie statistique ou en matière d'analyse des déplacements ; nous concluons que si tous les modélisateurs paraissent conscients de l'erreur

économétrique, seuls certains suspectent l'erreur sémantique et l'erreur de résolution, tandis que l'erreur de formulation reste généralement ignorée.

La deuxième partie **Théorie et méthode**, apporte une analyse théorique de l'erreur et une méthode pratique d'audit. Dans le quatrième chapitre, *Analyse théorique de l'erreur*, nous distinguons entre les différentes natures d'erreur : conception, formulation, résolution, économétrique. Nous expliquons comment ces erreurs se combinent dans le modèle et nous relevons certaines influences mutuelles : en particulier les erreurs sémantique et économétrique sont liées, et nous montrons sur un exemple élémentaire d'affectation qu'enrichir le contenu sémantique permet de réduire l'erreur économétrique. Le cinquième chapitre, *Une méthode systématique d'audit*, suit la distinction entre les différentes natures d'erreur pour prescrire des mesures et des traitements appropriés à chaque nature : l'audit technique se décline selon quatre dimensions respectivement conceptuelle, formelle, algorithmique et économétrique.

La troisième partie **applique la méthode proposée à un modèle particulier** d'affectation du trafic. Le sixième chapitre, *Plan d'audit de l'affectation bicritère*, explique en quoi ce modèle particulier est représentatif vis-à-vis de chaque nature d'erreur ; nous y fixons un plan d'attaque commun aux quatre dimensions d'audit : capitaliser le domaine de modélisation avant d'évaluer le modèle particulier. En effet l'audit technique nécessite un investissement scientifique lourd mais susceptible d'importantes économies d'échelle et d'envergure car les éléments relatifs à un modèle particulier se transposent aisément à des modèles proches. Les quatre chapitres suivants capitalisent l'affectation du trafic et audient l'affectation bicritère selon les dimensions respectivement conceptuelle, formelle, algorithmique et économétrique.

L'audit *sémantique* (chapitre 7) explicite la composition conceptuelle du modèle, afin de la comparer à la connaissance théorique du système étudié et d'apprécier sa portée. Pour chaque composant sémantique, nous recommandons d'établir une échelle de valeur afin d'évaluer la position du modèle.

L'audit *formelle* (chapitre 8) vérifie la correction de la formule caractéristique. Il contrôle l'équivalence entre la formule caractéristique et la composition conceptuelle (correction sémantique) et la justesse logique (correction syntaxique). Il examine l'existence et l'unicité des solutions à la formule caractéristique.

L'audit *algorithmique* (chapitre 9) vérifie la production d'une solution à la formule caractéristique. La production passe par un *dispositif de résolution*, appellation générique qui recouvre tant des calculs ou des simulations numériques, que des expériences sur maquette ou support analogique. Dans le cas d'un traitement numérique, il faut mesurer ou majorer l'écart entre le résultat du calcul et une solution vraie à la formule caractéristique, par un critère de convergence théoriquement rigoureux.

L'audit *économétrique* (chapitre 10) explicite l'erreur exogène relative aux inputs ainsi qu'aux paramètres des fonctions de dépendance, pour la propager à travers la formule caractéristique : ainsi le modèle devient transparent vis-à-vis des hypothèses économétriques.

Enfin le dernier chapitre tire un **bilan** et ouvre des **perspectives** de recherche nouvelle.

Première partie

***CONTEXTE ET
PRÉMICES***

Chapitre 1

L'exemple de l'affectation

Ce premier chapitre présente les modèles d'affectation et les incertitudes qui entachent leur application.

Il se compose de cinq sections. La première présente l'affectation et explique en quoi un modèle de répartition du trafic entre itinéraires concurrents participe à l'étude de la demande de transport. Dans la deuxième section, nous montrons par une série d'exemples didactiques la capacité des modèles d'affectation à simuler des actions de planification et d'exploitation des réseaux.

Encore faut-il adapter le modèle au problème à traiter : la troisième section évoque les difficultés rencontrées en pratique lors de l'application d'un modèle. Cette application est opérée à l'aide d'un logiciel, instrument au service de divers modèles : la quatrième section décrit brièvement les logiciels d'affectation.

Enfin la cinquième section délimite l'audit d'un modèle.

1.1 INTRODUCTION AUX MODÈLES D'AFFECTION

Nous définissons brièvement l'affectation (§1.1.1). Puis nous la plaçons dans la perspective de l'étude de la demande de transport (§1.1.2). Alors nous montrons comment la composition du modèle détermine ses aptitudes (§1.1.3).

1.1.1 Définition provisoire

1.1.1.1 Mettre en scène et simuler le choix d'itinéraire

Imaginons deux lieux mis en relation par deux chemins. Pour passer d'un lieu à l'autre, un déplacement peut emprunter un chemin ou l'autre : un choix d'itinéraire existe s'il n'y a pas de contrainte dictant le passage par un chemin donné. Le résultat du choix peut varier, selon qui l'effectue : le trafic total entre les deux lieux se divise, se répartit, entre les deux chemins. Quelle règle préside au choix ? En pratique, l'auteur du déplacement tend à suivre le chemin le plus commode : celui qui est le plus rapide, le mieux signalé, le plus économique, le plus sûr, le moins fatigant. Si un chemin domine les autres pour tous ces critères, le choix est évident. Sinon l'auteur du déplacement arbitre entre les qualités respectives des chemins qu'il connaît, il fait un compromis.

Un modèle d'affectation représente et simule le choix d'itinéraire sur un réseau de transport, avec de multiples lieux de départ et d'arrivée et potentiellement de multiples chemins entre chaque lieu de départ et chaque lieu d'arrivée. Représenter, pour mettre en scène les chemins possibles et décrire leurs qualités respectives. Simuler, pour répartir le trafic entre itinéraires concurrents selon des règles de comportement.

1.1.1.2 Vocabulaire particulier à l'affectation

La représentation de base est la suivante. Chaque déplacement, selon ses lieux de départ et d'arrivée, fait partie d'un flot entre une zone¹ d'origine et une zone de destination : on parle de *flot origine-destination* (O-D). On regroupe dans la *demande* la spécification des zones d'origine et des zones de destination, la description des comportements et la matrice des flots O-D. La schématisation du territoire d'étude en *zones de demande* est rapprochée d'une *schématisation du réseau* de transport sous la forme de noeuds (carrefours, jonctions) et d'arcs (routes). Chaque chemin sur le réseau est une suite d'arcs parcourus l'un après l'autre depuis un noeud d'origine jusqu'à un noeud de destination. Les chemins sont identifiés selon un principe à préciser. S'il y a plusieurs chemins pour desservir une relation origine-destination donnée, on recourt à une *règle de répartition du trafic entre itinéraires concurrents* pour simuler la concurrence qu'ils se livrent. On regroupe dans l'*offre* la schématisation du réseau, la dépendance des temps de parcours envers le trafic, la donnée des itinéraires.

1.1.2 Mise en perspective

1.1.2.1 Le schéma classique pour étudier la demande de transport

Le schéma classique des études de transport comprend quatre parties qui synthétisent

¹ensemble de lieux ponctuels proches les uns des autres

normativement la demande de transport servie par un réseau. Les quatre parties sont, dans l'ordre déductif originel :

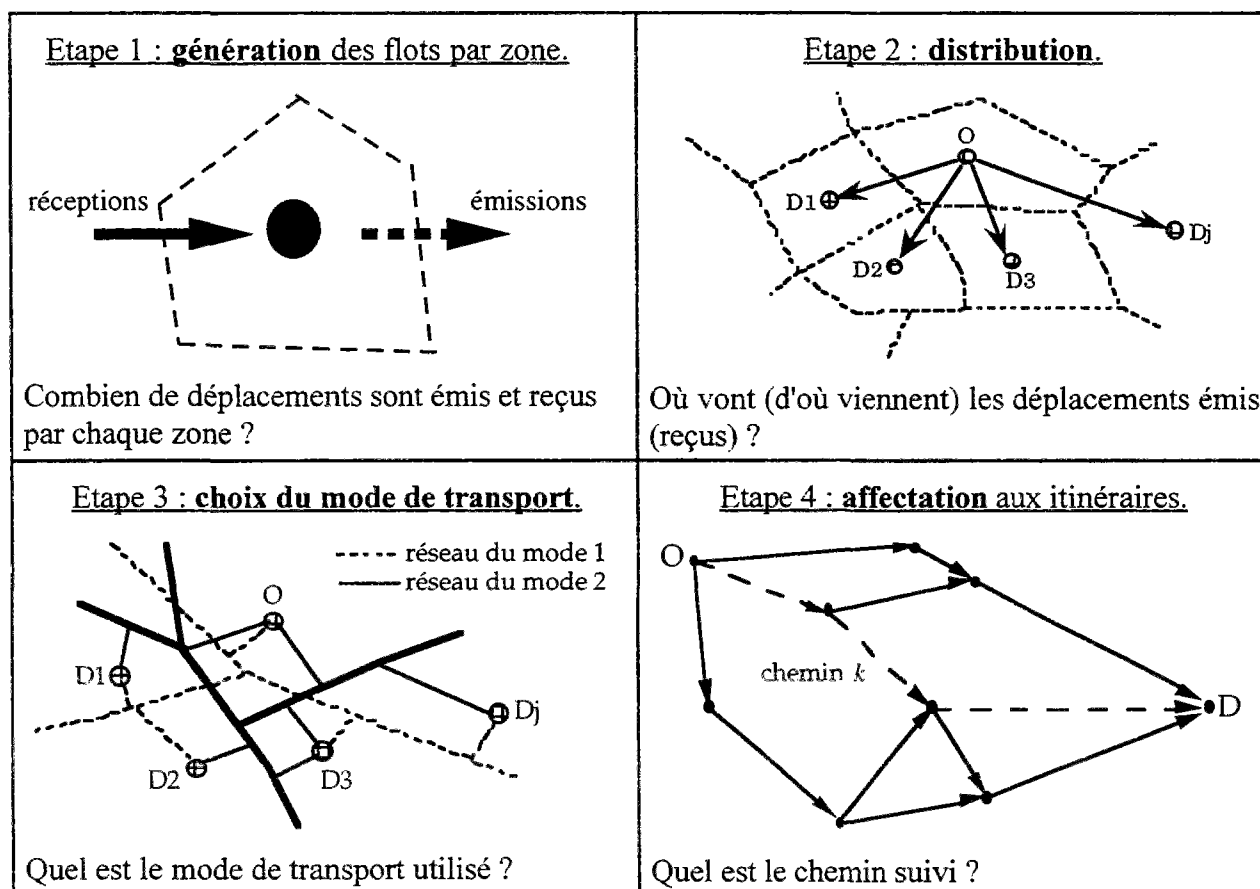
- la *génération*, pour répondre à la question "y a-t-il déplacement ?". Considérant la localisation des activités socio-économiques (résidences, emplois, écoles, commerces, loisirs...), des hypothèses de mobilité servent à quantifier les déplacements émis et reçus par chaque zone d'activité.

- la *distribution*, en réponse à "déplacement entre quels lieux ?". On définit la seconde extrémité géographique de ces émissions et réceptions, de manière à déterminer le volume de déplacement entre chaque lieu d'origine et chaque lieu de destination.

- le *choix modal*, en réponse à "quel mode est utilisé ?". Plusieurs modes s'offrent éventuellement à l'auteur du déplacement. Par exemple, un voyageur dans Paris peut choisir entre l'automobile, le taxi, le bus, le métro, le RER etc. sans oublier la marche et les deux roues. Autre exemple, un chargeur ou un transporteur de marchandises peut choisir pour un envoi à longue distance entre la route, le chemin de fer, l'avion ou le bateau.

- l'*affectation*, en réponse à "quel chemin est emprunté ?". Se fixant les flots de déplacements utilisant un mode donné, le modèle d'affectation rattache chaque déplacement à un chemin sur le réseau de ce mode.

Fig. 1a. Principes du schéma à quatre étapes.



Ainsi le schéma vise à répondre à quatre questions sur les déplacements (figure 1a). Dans la pratique on distingue souvent une étape par question, mais un tel découpage n'a rien d'impératif et il existe des modèles "simultanés" qui traitent plusieurs étapes dans le

même temps.

1.1.2.2 Six paradigmes relatifs à la demande de transport

Le terme de *transport*, décomposable en trans- et -port, comprend à la fois une action, donc un sujet et un objet en relation sous certaines modalités, et une notion de traversée, impliquant un lieu de départ, un lieu d'arrivée ainsi qu'un espace intermédiaire. Le terme de *déplacement* se décompose de façon semblable en dé- et -placement : un sujet change de place un objet, l'accent portant là davantage sur les lieux de départ et d'arrivée que sur l'espace (la distance) qui les sépare.

Par demande de transport, on désigne les besoins de transport manifestés par des voyageurs ou des chargeurs et transporteurs. La représentation théorique de la demande de transport repose sur six paradigmes : économétrique, spatial, économique, comportemental, d'activités, d'évolution temporelle.

i) Paradigme *économétrique* : les mouvements constatés sur une route peuvent être dénombrés, donc le trafic local est une entité mesurable, une variable économétrique. Ce paradigme sous-tend toute l'ingénierie du trafic, formalisée à partir des années 1930.

ii) Paradigme *géographique, spatial* : le trafic local est un mélange de déplacements, dont chacun a une origine et une destination particulières. L'explicitation des origines et des destinations (découpage du territoire en zones de demande) et la schématisation des routes par un graphe (sous forme de sommets et d'arcs) permettent de mener des analyses spatiales : elles constituent le fondement des modèles spatialisés de demande de transport, employés depuis 1950 environ.

iii) Paradigme *économique* : l'auteur d'un déplacement effectue des choix qui sont autant de décisions microéconomiques. Parmi ces choix, citons la génération, la distribution, le mode, l'itinéraire, l'horaire... ou encore la fréquence de déplacement, la motorisation, et jusqu'à la localisation. Le schéma classique est donc une combinaison particulière de décisions ; bien d'autres combinaisons sont possibles. Ce paradigme est d'emploi courant depuis environ 1950¹.

iv) Paradigme *comportemental* : segmenter pour appréhender la diversité des comportements. L'hypothèse d'un comportement moyen n'est souvent qu'un pis-aller. Longtemps retenue dans les modèles par commodité, on tend désormais à lui substituer une approche "par segments", où l'on désagrège la demande en catégories au comportement particulier. En pratique, cela concerne surtout les étapes de génération et de choix modal : par exemple, le modèle bicritère prix-temps suppose une segmentation selon la valeur du temps. Ce paradigme se développe depuis les années 1960.

v) Le paradigme *d'activités* considère le déplacement comme une demande *dérivée* d'une demande primaire pour mener une activité en un lieu donné. Il s'agit d'explicitier certaines modalités du déplacement : son but (l'activité qui le motive), son aspect (caractère contraint ou facultatif, régulier ou non, fréquent ou non...) et plus généralement des

¹Les paradigmes économique et spatial ont été employés de façon combinée par Alan Voorhees vers la fin des années 1940, quand il a conçu le schéma à quatre étapes afin d'analyser des statistiques d'accidents pour le compte d'une compagnie d'assurance. Mentionnons aussi le rôle de Marvin Manheim dans l'identification de la pluralité des combinaisons possibles entre étapes de modélisation, et la conception d'un modèle de demande de transport comme une structure élaborée à partir de composants.

interactions de natures diverses (enchaînement entre activités, relations interpersonnelles...). Ce paradigme est manifeste depuis les années 1970.

vi) Paradigme *dynamique, d'évolution temporelle* de l'offre et de la demande : par modification spontanée de la demande (ex. effets de génération), ou par une modification de l'offre, ou par interaction entre l'offre et la demande.

Notre décomposition en paradigmes ne prétend évidemment pas retracer l'analyse théorique de la demande de transport : en particulier les paradigmes spatial et d'activités sont des évidences de longue date pour les économistes. La décomposition donne des repères pour apprécier le développement et la portée des modèles de demande ; elle aide à qualifier les approches et à relativiser les innovations. Ainsi les modèles de choix discrets, prétendus révolutionnaires au plan théorique, sont en fait une combinaison particulière des paradigmes économique et comportemental ; leur innovation tient à la méthode d'estimation, elle est donc d'ordre technique.

1.1.2.3 Qualification de l'affectation

Un modèle d'affectation met en oeuvre de façon standard les paradigmes économétrique (avec la notion de flot de déplacement), spatial (les relations origine-destination et l'inscription sur un réseau), économique (choix d'un itinéraire). On trouve des instances particulières où intervient également le paradigme comportemental (ex. segmentation selon la valeur du temps), tandis que le paradigme des activités est généralement absent de l'affectation stricto sensu.

Remarquons que tout modèle de déplacement se représente comme un modèle d'affectation à des "super-itinéraires", en définissant un "super-itinéraire" de façon appropriée, par exemple comme la combinaison d'une origine, d'une destination, d'un mode, d'un horaire et d'un chemin modal. Chaque élément de la combinaison prend la forme d'un arc, réel ou fictif, dans un super-réseau approprié (Sheffi et Daganzo, 1978).

1.1.3 Fonctions et aptitudes : selon la composition sémantique

1.1.3.1 Fonctions

On emploie un modèle d'affectation pour simuler le choix d'itinéraire sur un réseau de transport, de façon isolée ou au service de fonctions qui se situent en amont dans le schéma classique : choix modal, distribution etc.

La planification et l'exploitation des réseaux constituent le terrain d'application privilégié : planification d'abord, pour tester l'intérêt de nouvelles infrastructures ; gestion et exploitation ensuite, pour prévoir l'effet de mesures de tarification, de plans de circulation, de partage de la capacité (régulation aux carrefours...). Le modèle simule les réactions de la demande à des actions du côté de l'offre.

1.1.3.2 Des aptitudes qui dépendent de la composition sémantique

Pour remplir effectivement les missions qu'on lui assigne, un modèle d'affectation doit posséder des aptitudes particulières :

- représenter les chemins et leurs caractéristiques,
- comparer les chemins et calculer la répartition du trafic,
- selon la répartition calculée, charger le trafic sur les chemins, donc sur le réseau.

Souvent il convient de compléter ces aptitudes "minimales" afin d'améliorer le réalisme de la simulation :

- lorsque le trafic est suffisamment important pour modifier les conditions de déplacement, en particulier si la congestion accroît sensiblement les temps de parcours, alors il faut prendre en compte cet effet. On peut relier le temps de parcours d'un arc au flot écoulé par cet arc, obtenant alors un modèle à *temps variables*.
- si les conditions de déplacement influent sur le volume de trafic, tout comme le prix d'un produit influe sur le volume de demande, on peut relier le volume de trafic au niveau de service, obtenant un modèle à *demande élastique*.
- si les comportements des auteurs de déplacement sont hétérogènes, il est bon de distinguer plusieurs classes de demande. Par exemple sur un réseau routier, des voitures et des poids lourds n'ont pas les mêmes tenues routières. Autre exemple, certains déplacements sont plus pressés que d'autres ; leurs auteurs acceptent plus volontiers de payer davantage pour économiser du temps. On qualifie de *multiclasse* un modèle qui distingue plusieurs classes.

Nous appelons *mécanisme explicatif* tout composant de modélisation qui confère au modèle une aptitude, une signification. Par exemple, la dépendance du temps de parcours envers le trafic, ou celle du volume de demande envers le niveau de service.

La composition du modèle en mécanismes explicatifs détermine ses aptitudes, partant les fonctions qu'il peut remplir avec pertinence (tableau 1A).

Tab. 1A. Résultats attendus et composants qui leur correspondent.

Résultat de la simulation	Composants de modélisation
Trafic moyen d'une route, hors induction	- zonage et matrice O-D des flots - fonctions de temps de parcours - réseau - paramètres comportementaux - module de sélection des itinéraires - règle de répartition
Prise en compte de l'induction	- fonction de demande
Représentation "a priori" des flots de déplacements	- données socio-économiques par zone - matrice O-D des niveaux de service - modèle de génération - modèle de distribution spatiale - comptages avec identification du trafic local
Evolution temporelle du trafic	- séries chronologiques de données socio-économiques - modèle de projection du trafic
Répartition temporelle du trafic entre périodes types	- plusieurs matrices O-D de flots, une par type de période
Répartition dynamique du trafic	- capacité dynamique dans l'affectation
Evaluation économique	- hypothèses économiques exogènes - module de calcul économique

La référence (Leurent, 1995d) passe en revue des composants sémantiques pour

l'affectation. Mentionnons, du côté de l'offre, des fonctions de temps de parcours pour représenter la congestion, des contraintes explicites de capacité, un type de gestion (optimum individuel ou collectif). Citons, du côté de la demande, une segmentation en plusieurs classes, l'élasticité du volume de déplacement au niveau de service, une minimisation d'un temps généralisé déterministe ou aléatoire, ou encore, à l'intérieur d'une classe, la différenciation selon un attribut particulier (ex. la valeur du temps dans le modèle bicritère prix-temps).

1.2 VARIATIONS SUR UN PROBLÈME DE CHOIX D'ITINÉRAIRE

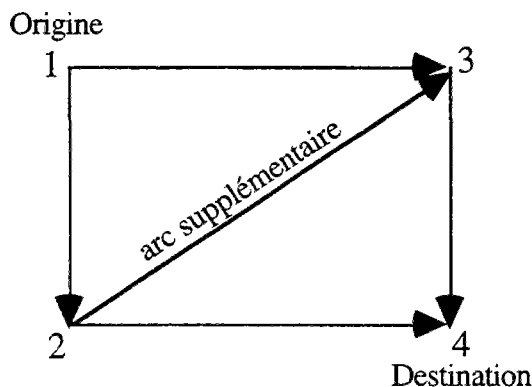
Afin de démontrer la capacité des modèles d'affectation à représenter les actions de planification et d'exploitation d'un réseau, nous avons mené une série d'applications numériques à partir d'un cas d'école¹.

Il s'agit d'une unique relation origine-destination, desservie par un réseau de 5 arcs combinables en trois itinéraires : $\underline{1} = \{(1-2); (2-4)\}$, $\underline{2} = \{(1-3); (3-4)\}$ et $\underline{3} = \{(1-2); (2-3); (3-4)\}$ (figure 1b).

A chaque arc a est associée une fonction de temps de parcours (Cf. loi débit-vitesse), qui relie le temps de traversée T_a au débit local x_a selon la formule

$T_a = t_a^0(1 + 0.15(\frac{x_a}{C_a})^4)$ où t_a^0 est le temps de parcours à débit nul et C_a la capacité pratique.

Fig. 1b. Données pour le cas d'école.



Arc a	t_a^0 (h)	C_a (véh/h)
(1-2)	0,0400	1861
(1-3)	0,1850	2356
(2-3)	0,0154	1233
(2-4)	0,1850	2356
(3-4)	0,0400	1861

1.2.1 Comment agir sur la congestion ?

1.2.1.1 Optimum individuel vs. optimum collectif

Pour marquer la différence entre optimum collectif et optimum individuel², Lam illustre le paradoxe dit de Braess, selon lequel la mise à disposition d'un service supplémentaire peut détériorer l'ensemble des services : c'est un effet de congestion. Pour notre exemple,

¹conçu par Lam (1989) afin d'opposer l'optimum individuel et l'optimum collectif.

²L'optimum collectif, à demande fixée, est défini comme la minimisation du temps total passé en déplacement. L'optimum individuel, dit de Wardrop en transport, correspond à un comportement égoïste des conducteurs, supposés chacun minimiser son coût généralisé : c'est un équilibre de Cournot-Nash.

il faut distinguer deux époques : avant et après la mise en service de l'arc diagonal (2-3).

Avant la mise en service, il n'y a que deux chemins, équivalents quant au temps de parcours. La demande, fixée provisoirement à $q = 6\,000$ véh/h, s'affecte équitablement aux deux itinéraires, pour un temps de parcours individuel de 0.3384 h et un temps horaire collectif de 2 030 h. La répartition est identique pour l'optimum individuel et pour l'optimum collectif.

Après la mise en service, il y a trois chemins, et les deux optimums ne coïncident plus. En l'occurrence, à l'optimum individuel la demande se répartit à raison de 2 000 véh/h par chemin, tandis qu'à l'optimum collectif le troisième itinéraire ne reçoit plus que 800 véh/h contre 2 600 véh/h affectés à chacun des deux autres. Le temps horaire total s'établit à 2 204 h pour l'optimum individuel, contre 1 915 h pour l'optimum collectif : l'ajout de l'infrastructure supplémentaire, bénéfique si l'optimum collectif préside à la répartition aux itinéraires, se révèle nuisible si les conducteurs sont libres d'agir chacun dans son intérêt propre immédiat.

Notons que ce résultat est propre aux données numériques choisies : dans la réalité, le paradoxe de Braess est loin d'être le cas général, heureusement !

1.2.1.2 Limiter la capacité pour amener à l'optimum collectif

Une contrainte de capacité se manifeste sur le terrain de façon physique. Elle peut tenir au dimensionnement géométrique de l'infrastructure (nombre de voies...), mais aussi à un choix d'exploitation : faire un sacrifice local en vue de l'efficacité d'ensemble, tel est le sens du partage de la capacité qu'imposent les signaux de régulation du trafic (feux aux carrefours etc).

Poursuivant notre exemple, il suffit de limiter le débit de l'arc diagonal à celui prévu par l'optimum collectif, i.e.. 800 véh/h, de manière que l'optimum individuel coïncide à nouveau avec l'optimum collectif. Dans ces conditions, on calcule qu'une pénalité temporelle de 0.10 h est associée à la traversée du maillon diagonal.

Comment inciter les conducteurs à se comporter de la façon la meilleure pour la collectivité ? Par exemple, en plaçant à l'issue de l'arc diagonal un signal de trafic qui inflige un retard moyen de 0.10 h.

1.2.1.3 Un péage peut avoir le même effet tout en dégageant une ressource économique

Nous avons donné l'exemple de deux tactiques pour gérer la congestion : ne rien faire, en laissant s'établir sans contrôle l'optimum individuel (§1.2.1.1), ou réguler par les temps de parcours (§1.2.1.2). La tarification constitue une troisième tactique, économiquement préférable aux deux précédentes : il s'agit de remplacer la pénalité temporelle par un péage, susceptible du même effet sur le trafic, afin de convertir les retards infligés volontairement en dépenses monétaires qui elles sont recyclables puisqu'elles apparaissent en recette pour d'autres acteurs économiques.

Supposons que la valeur du temps, i.e. la somme que le conducteur accepte de payer pour gagner une unité de temps, soit de 50 F/h pour chacun des véhicules. Alors il suffit d'imposer sur l'arc diagonal un péage $P = \text{Valeur_du_Temps} * \text{Pénalité_temporelle} = 5.175$ F/véh, pour aboutir au même partage entre itinéraires que par la restriction volontaire de capacité, mais en dégageant une recette horaire totale de $x * P = 4\,140$ F,

sans modifier l'état des usagers puisque ces derniers font la même dépense¹. La tarification dégage une recette financière qui est une ressource pour la collectivité, contrairement aux deux autres tactiques.

1.2.2 Mieux représenter le comportement de la demande

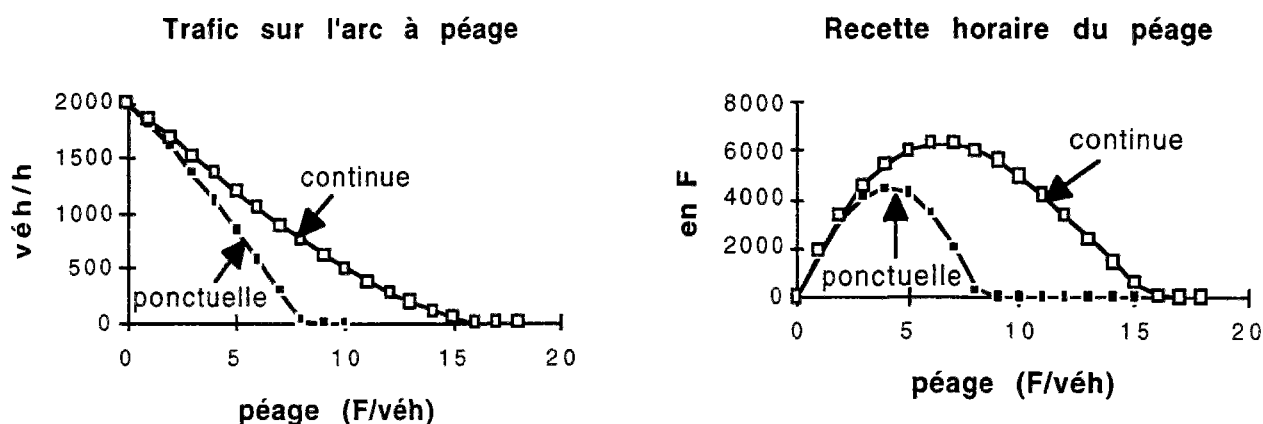
1.2.2.1 Attention aux options de modélisation !

Toujours au sujet des effets du péage, on peut par souci de réalisme distinguer plusieurs valeurs du temps parmi les conducteurs. Si l'on modélise cette dispersion, alors il faut procéder par tâtonnements pour déterminer la valeur convenable du péage, hormis dans les cas élémentaires tel notre exemple.

Admettons que les valeurs du temps suivent une loi triangulaire de moyenne 50 F/h, distribuée entre 0 et 100 F/h. Il suffit de changer le niveau de péage (mis à 7.68 F/véh), pour obtenir le résultat désiré, avec une recette horaire de 6 144 F.

Entre les deux applications numériques précédentes, la valeur du temps moyenne est la même, fixée à 50 F/h. Cependant les prévisions de trafic et de recette diffèrent sensiblement (figure 1c) !

Fig. 1c. Trafics et recettes, en fonction du péage.



La courbe repérée par "continue" désigne le modèle à valeurs du temps différenciées.

1.2.2.2 Considérer l'élasticité de la demande

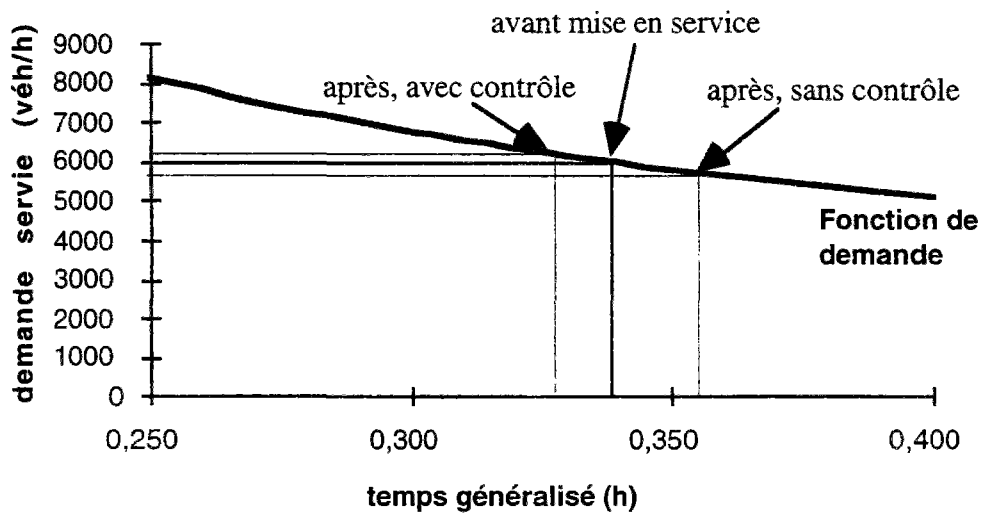
Un modèle qui établit l'équilibre entre offre d'acheminement et demande de déplacement permet de calculer comment la demande réagit de façon quantitative à une certaine modification de l'offre.

Illustrons cela en reprenant notre cas d'école. On suppose que le volume de déplacements dépend du temps de parcours, par une fonction de demande d'élasticité constante -1.0. Si l'on part d'une demande de $q = 6\,000$ véh/h pour les conditions de circulation d'avant la mise en service de l'arc diagonal (i.e.. un temps de trajet individuel de 0.3384 h), alors la mise en service sans contrôle, dégradant l'ensemble du système,

¹en frais d'ordre monétaire dans le cas de la gestion par la tarification, en frais d'ordre temporel dans le cas de la gestion par les temps

réduit la demande servie (de 4%) ; tandis que limiter à 800 véh/h le débit de l'arc diagonal permet au contraire de satisfaire davantage de clients (+ 0.7%).

Fig. 1d. Déplacements de l'équilibre offre-demande.

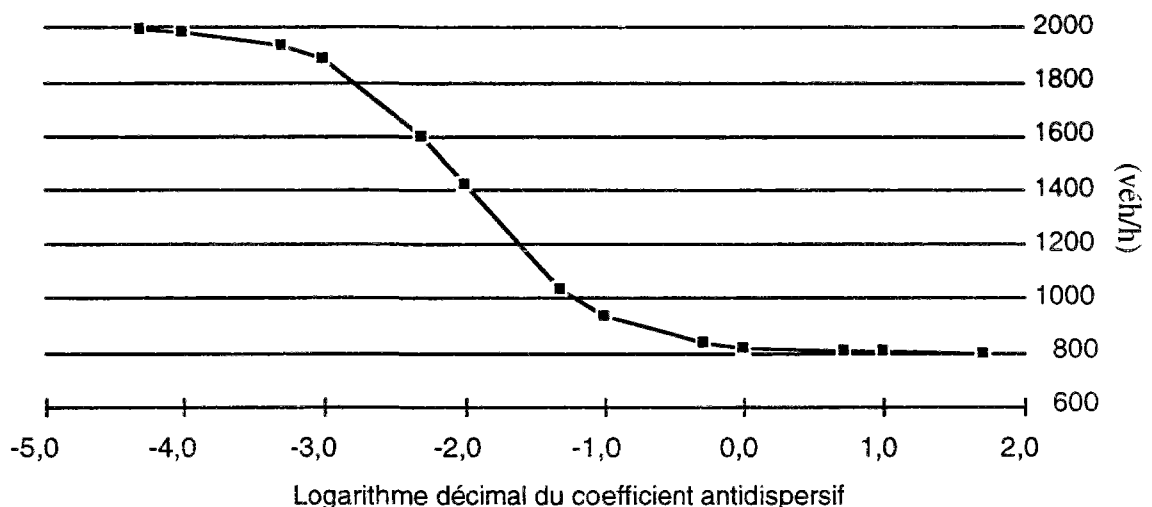


1.2.2.3 Quantifier la précision des perceptions

On qualifie de *stochastique* un modèle d'affectation dans lequel un conducteur attache à un itinéraire un temps généralisé aléatoire : aléatoire au point de vue de l'analyste, qui veut reproduire de manière probabiliste une variabilité qu'il ne désire pas expliquer de façon déterministe.

Dans le modèle logit linéaire, on ajoute un aléa au temps généralisé de chaque itinéraire. La répartition entre itinéraires dépend alors de l'amplitude de l'aléa par rapport à la valeur moyenne du temps généralisé. Pour mesurer le rapport, on emploie souvent un "coefficient antidispersif" θ proportionnel à l'inverse de l'écart-type de l'aléa : plus θ est fort, moins l'aléa influe sur la répartition entre itinéraires.

Fig. 1e. Flot sur l'arc diagonal en fonction du coefficient antidispersif.



Illustrons cela en considérant, pour plusieurs valeurs du coefficient θ , la situation après mise en service de l'arc diagonal et avec une limite de capacité. On voit que la répartition

logit, presque indifférente aux caractéristiques des chemins à θ faible, se rapproche progressivement de la répartition déterministe (figure 1e).

On peut dans ce modèle représenter et mesurer la qualité de l'information mis à disposition des conducteurs. Un coefficient antidispersif θ fort indique une bonne information, tandis qu'un θ faible traduit un manque d'information.

1.2.3 Bilan

Retenons de ces expériences :

- la disposition de composants de modélisation propres à intégrer les considérations théoriques. En particulier nous avons conçu la série de micro-applications comme une justification didactique de la tarification de la circulation.
- l'importance d'une bonne connaissance empirique des paramètres déterminants dans une affectation : valeur du temps en moyenne et en dispersion, coefficient de dispersion aléatoire dans le modèle logit, coefficient d'élasticité du volume de demande au coût généralisé...
- la possibilité pratique d'obtenir des résultats sensiblement différents à partir de données d'application relativement proches, selon les composants de modélisation choisis. Toute tentative d'interprétation devrait donc être accompagnée d'une claire explication du mécanisme d'affectation retenu.

1.3 CODAGE, CALAGE, TRUCAGES, DÉRAPAGES

Une étude de trafic donne lieu à des décisions tant dans la composition du modèle que dans son alimentation en données. Ces décisions comportent une part d'arbitraire, souvent faute de théorisation de la question posée. Le chargé d'étude dispose donc d'une importante marge de manoeuvre dont il tire parti sur commande ou selon ses propres connaissances ou goûts. Nous allons successivement explorer des lieux de choix importants (§1.3.1), exposer des possibilités de trucage (§1.3.2), expliquer en quoi l'application du modèle peut se révéler incontrôlable (§1.3.3), et dresser un bilan provisoire de la contrôlabilité d'une étude (§1.3.4).

1.3.1 Codage et calage : de nécessaires arbitrages

Une étude de trafic nécessite diverses schématisations, discrétisations, approximations :

- de nature phénoménologique, dans la spécification des mécanismes explicatifs. La section précédente a donné des exemples en la matière.
- de nature géographique : la circonscription du territoire d'étude et son découpage en zones, la délimitation temporelle, les modes et les trafics considérés, la schématisation du réseau concret par un ensemble d'arcs et la spécification de certains attributs des arcs (ex. longueur et capacité des routes) etc.
- de nature économétrique, pour la spécification numérique des inputs qui varient selon le scénario : flots origine-destination, paramètres comportementaux qui interviennent dans la répartition du trafic entre itinéraires concurrents (ex. la valeur du temps), certains attributs des arcs (ex. la fonction de temps de parcours d'un arc).

On appelle *codage* l'ensemble des schématisations de nature géographique.

On appelle *calage* l'ajustement numérique des inputs afin de reproduire des observations. Le codage et le calage sont tous deux relatifs à une étude particulière. Ils sont très interdépendants : si l'on calcule des trafics locaux qui excèdent largement les capacités codées, probablement celles-ci sont insuffisantes ou certaines routes "concrètes" ont été omises à tort du codage, ou la matrice O-D est trop forte ; autre raison possible, si les trafics locaux calculés excèdent considérablement les observations, un découpage en zones trop grossier qu'il convient alors de détailler.

1.3.1.1 Arbitrages liés au codage

Les opérations de codage comportent intrinsèquement des arbitrages. Citons :

- l'extension du territoire d'étude. Cela conditionne les effets "visibles" à l'aide du modèle. Par exemple, pour simuler un projet de nouvelle route, un territoire trop réduit ignore le trafic à longue distance qui passait auparavant hors du territoire mais pourrait emprunter le projet.
- le découpage du territoire d'étude en zones. Cela conditionne la qualité de représentation des trajets terminaux.
- la schématisation du réseau : évidemment il faut représenter les routes principales, mais jusqu'à quel niveau de routes secondaires convient-il de descendre ? De cet arbitrage dépendent les itinéraires considérés dans le modèle : correspondent-ils vraiment à ceux disponibles dans la réalité ?

Le cahier des charges de l'étude devrait prescrire des règles en matière de codage. Ce n'est pas le cas général en France. Comme il n'existe pas non plus de normes établies dans la profession, le chargé d'étude dispose d'une grande latitude. Signalons l'existence de règles pratiques en Angleterre : elles sont rassemblées dans plusieurs manuels spécialisés, dont l'un pour l'étude de trafic (le Traffic Appraisal Manual), un autre pour l'évaluation socio-économique des projets (le COst Benefit Analysis).

1.3.1.2 Arbitrages liés au calage

Le calage donne également lieu à des arbitrages :

- les données disponibles pour produire une matrice origine-destination de flots de déplacements sont fragmentaires. Souvent on produit une matrice initiale en appliquant un modèle de génération-distribution dont les inputs sont des données démographiques et socio-économiques au niveau de la zone, puis on ajuste la matrice en s'appuyant sur des mesures locales de débit. De nombreuses méthodes d'ajustement existent ; certains chargés d'étude rectifient "au flair" les effectifs de certaines cases dans la matrice origine-destination...
- souvent on fixe les paramètres du comportement (ex. la valeur du temps) selon une norme ou une tradition, plutôt qu'en fonction d'observations empiriques. Lorsqu'il y a plusieurs valeurs possibles, le chargé d'étude doit choisir.

Les rapports d'étude passent sous silence bon nombre d'opérations de calage. Cela compromet ipso facto les analyses de sensibilité, puisque sont dissimulés des facteurs de doute qui nécessiteraient une étude spécifique !

1.3.2 Quelques trucages

Les arbitrages présentés sont nécessaires. Cependant ils offrent la possibilité de diriger a priori les résultats de l'étude, de les orienter. Parmi les "trucs" rencontrés, voici quelques techniques qui ont pour effet commun d'adapter, ou plus souvent de détourner, le principe de mise en concurrence des itinéraires :

- exclure certains itinéraires de certaines variantes, bien qu'ils en fassent partie !
- affecter tout le trafic d'une relation origine-destination (O-D) sur un seul itinéraire fixé de façon arbitraire.
- diviser artificiellement une relation O-D en plusieurs, avec pour chacune un itinéraire forcé et une certaine part du trafic total. Cela aide au calage... mais modifie implicitement la structure de choix dans le modèle de demande car au lieu de confronter directement les nouveaux itinéraires aux anciens, on étudie pour chaque ancien itinéraire le détournement vers les nouveaux itinéraires.
- modifier artificiellement les caractéristiques (prix, temps) d'un itinéraire en changeant le prix ou en ajoutant à l'itinéraire une arête fictive au temps arbitrairement fixé...

Tous ces "trucs" ne sont pas nécessairement de mauvais aloi : chacun d'eux peut se justifier dans des conditions particulières. Mais l'accumulation de manipulations devient vite incontrôlable, sauf en vue d'un résultat imposé a priori... La responsabilité revient à l'exécutant de l'étude et davantage encore au commanditaire.

Parmi les conséquences des manipulations, mentionnons celles sur l'évaluation des projets d'investissement de transport. La rentabilité socio-économique d'un projet repose essentiellement sur les gains de temps des usagers. Si ces gains sont calculés de façon artificielle (répartition entre itinéraires fixée par l'exécutant de l'étude, temps qui comprennent une part fictive importante), quelle crédibilité accorder à l'étude ?

1.3.3 Divers dérapages

Qui dit trucage sous-entend l'existence d'un manipulateur qui tente de contrôler à sa guise la situation, en l'occurrence l'application du modèle. Un tel contrôle se révèle souvent illusoire, soit parce que l'empilement des manipulations devient inextricable, soit faute d'autorité sur le logiciel d'étude, indispensable instrument du modèle.

L'application rigoureuse du modèle d'affectation nécessite un calcul exact de l'équilibre entre offre et demande de transport. Or certains logiciels calculent une solution qui ne satisfait pas forcément les conditions d'équilibre. Alors le calcul n'a plus de sens !

Une autre possibilité de défaillance réside dans la préparation des données. Le mauvais renseignement d'un seul nombre, par étourderie, est un risque constant. Donnons un exemple lié au logiciel ARIANE (DR, 1986). L'utilisateur code manuellement chaque itinéraire comme une série d'arêtes repérées par des indices. Dans une étude moyenne avec 20 zones de demande, donc 190 relations O-D, à raison de trois itinéraires par relation et de 15 arcs par itinéraire, il lui faut coder environ 8 000 indices ! Si l'un d'eux est faux, le trafic de l'itinéraire est affecté sur l'arc désigné par l'indice codé, non sur l'arc convenable ! Comment prévenir de tels risques ? Il conviendrait d'incorporer au logiciel des fonctions de détection automatique : certains tests sont bien connus (Cf. Thomas 1991) mais ne figurent malheureusement pas dans tous les logiciels.

1.3.4 Bilan provisoire

Concluons qu'une étude de trafic donne lieu à de nombreuses opérations élémentaires tant pour le codage, le calage ou le calcul proprement dit de la répartition entre itinéraires. Sans prétendre juger la pratique, nous nous interrogeons sur la possibilité de contrôler les applications. Le contrôle et la validation d'une étude nécessitent un examen des opérations élémentaires, ou plutôt de leurs résultats, afin de détecter les éventuels truccages et d'identifier les traitements douteux.

Un tel examen est-il possible ? Une condition préalable évidente tient à la coopération entre le contrôleur qui en serait chargé et l'exécutant de l'étude, détenteur des données d'application et des divers jeux d'hypothèses. Si cette condition est remplie¹, il reste à exploiter l'information fournie. Cela requiert un logiciel spécialisé, tout aussi nécessaire pour contrôler précisément l'étude que pour l'exécuter. Car en pratique le volume d'information est tel que seul l'outil informatique permet un examen systématique approfondi.

Ainsi nous ramenons provisoirement la question de la contrôlabilité d'une étude à la question de la possibilité informatique d'un contrôle.

1.4 LES LOGICIELS, INSTRUMENTS DES MODÈLES D'AFFECTION

Avant de répondre à cette question, nous décrirons brièvement les logiciels d'affectation qui désormais sont les instruments irremplaçables des modèles dans les études de trafic (§1.4.1). Puis nous caractériserons le marché des logiciels d'affectation (§1.4.2) et nous montrerons que, parmi les critères de choix d'un logiciel pour une étude, les fonctions de contrôle ne paraissent pas l'élément le plus déterminant (§1.4.3).

1.4.1 Contenu et forme des logiciels d'affectation

Ne confondons pas modèle et logiciel. Le modèle correspond à un référent théorique, tandis que le logiciel offre des fonctions pour instancier et traiter un modèle particulier, spécifié parmi plusieurs possibilités. On peut comparer un logiciel à un récipient dont le contenu varie en fonction de l'étude et comprend un modèle (choisi parmi plusieurs), une méthode de calcul, ainsi qu'un format de données². Cela dit, certains logiciels ne permettent d'instancier qu'un seul et unique modèle.

Plutôt que passer en revue des logiciels particuliers (Cf. Leurent 1996d), nous proposons une grille d'analyse pour qualifier le contenu et la forme des logiciels selon successivement le spectre des composants phénoménologiques (§1.4.1.1), la méthode de calcul (§1.4.1.2), l'ergonomie pour l'utilisateur (§1.4.1.3), enfin la disposition d'utilitaires spécialisés (§1.4.1.4).

Les rubriques choisies traitent des fonctions des logiciels et de leur forme telle que

¹Il semblerait plus logique que le commanditaire soit entièrement propriétaire de toutes les données d'application ; mais ce n'est pas le cas général des études commandées en France !

²certaines fonctions ou paramètres sont spécifiés à l'avance dans le logiciel, ou ne peuvent être choisis par l'utilisateur que parmi quelques options. Notamment, les fonctions de temps de parcours

perceptible par l'utilisateur.

1.4.1.1 Le spectre des composants sémantiques

Nous entendons par là les composants de modélisation que le logiciel met à la disposition de l'utilisateur, particulièrement pour :

- représenter l'offre (spécification des itinéraires, description des arcs, temps de parcours, régime stationnaire ou transitoire de circulation etc.).
- représenter la demande (hypothèse d'information et de rationalité économique, règle(s) de répartition du trafic entre itinéraires concurrents, élasticité de la demande etc.).

La plupart de ces composants figurent dans tout logiciel d'affectation basé sur les principes présentés en section 1.1. S'y ajoutent, selon le cas, diverses possibilités pour traiter les autres modèles de demande que sont la génération, la distribution et le choix modal.

Au sujet des domaines d'application : milieu urbain ou interurbain, trafic de voyageurs et/ou de marchandises, tel et/ou tel mode de transport, certains logiciels se consacrent exclusivement à un seul créneau (ex. ARIANE) tandis que d'autres sont (prétendument) applicables à tout sujet (ex. TRIPS).

1.4.1.2 La méthode de calcul

L'affectation met en équilibre l'offre et la demande d'acheminement par un réseau modal. La résolution quantitative de ce problème nécessite d'abord une formulation mathématique qui est le but du calcul, puis un algorithme qui est la méthode du calcul. Les logiciels d'affectation diffusés commercialement reposent en général sur une méthode de Frank-Wolfe ou une variante (Cf. chapitre 9) ; il convient de vérifier la convergence d'une telle méthode si le modèle traité sophistique la représentation de l'offre ou de la demande par rapport au modèle basique de Beckmann (1956).

1.4.1.3 Ergonomie pour l'utilisateur

A cet aspect correspondent deux lots de fonctions :

- la gestion des fichiers de données. Une étude teste souvent plusieurs scénarios dont chacun nécessite des affectations à différentes époques. Chaque affectation mobilise plusieurs sous-ensembles de données, relatifs à l'offre ou à la demande, ainsi que des fichiers de paramètres généraux.
- l'interface interactive-graphique, moyen moderne et désormais classique de gérer les données spatialisées et de les représenter (ex. carte des capacités du réseau, ou des vitesses moyennes, ou des zones d'émission et réception des déplacements).

Notons que l'accent mis à une époque sur la richesse fonctionnelle dans la gestion et la représentation des données, est moins fort à présent qu'un logiciel d'affectation s'emploie en coopération avec d'autres logiciels plus spécialisés en matière de gestion de données ou de cartographie. L'utilisateur est désormais surtout sensible à la possibilité d'échanger des données entre logiciels.

1.4.1.4 Utilitaires spécialisés

Là encore deux lots de fonctions :

- pour aider au codage, diverses procédures de dépistage d'erreurs éventuelles (ex. tests de connexité, comparaison entre distance à vol d'oiseau et longueur codée des arcs).
- des fonctions de calage (ex. ajustement de matrices O-D sur la base de débits observés).

Ces fonctions sont essentielles pour le contrôle d'une application.

1.4.2 Le marché des logiciels d'affectation

Le logiciel d'affectation, facteur de production des études de trafic, fait lui-même l'objet d'un marché économique. Après une longue période d'autarcie de chaque grand pays industrialisé (bloc nord-américain ; Allemagne ; Royaume-Uni ; France), chaque marché national s'est progressivement ouvert à la concurrence internationale. On peut classer les logiciels en trois catégories : ceux qui participent à la compétition internationale (§1.4.2.1), ceux restreints à une niche nationale (§1.4.2.2), enfin les produits hors commerce (§1.4.2.3).

1.4.2.1 Les produits phares du marché international

Ce sont des boîtes à outils qui offrent à la fois un très large spectre de composants phénoménologiques, une gestion et une présentation de données standardisées (avec un module spécialisé pour chaque type de données), et des utilitaires spécialisés qui sont l'argument principal du discours commercial.

Citons les logiciels EMME/2 (canadien), TRIPS, SATURN et CONTRAM (anglais), MIN UTP et TRANPLAN (américains) : ils sont produits de façon industrielle par des équipes puissantes d'informaticiens et d'analystes-transport.

1.4.2.2 Stratégies de niche

Par le poids de l'habitude ou la force de la loi, certains logiciels détiennent une position privilégiée sur des créneaux étroits. Ainsi en France, pour les études de projet routier en milieu interurbain, une directive du ministère des transports impose le logiciel ARIANE. Autre exemple, l'emploi traditionnel dans les agglomérations françaises du logiciel DAVIS pour les études de trafic routier urbain : position initialement acquise grâce à la mise à disposition gratuite auprès des services techniques des collectivités publiques. Citons encore POLYDROM présent sur le marché suisse et QRS II (américain) qui s'est diffusé via les universités.

Si certains de ces logiciels sont proches du standard international, d'autres sont très en deçà et ne subsistent que faute d'information des exécutants et des commanditaires. La plupart incorporent un composant phénoménologique spécifique qui les différencie des produits internationaux et soutient leur stratégie de niche (ex. affectation bicritère pour DAVIS, affectation logit logarithmique pour ARIANE, fonctions de demande directe dans POLYDROM). Tous se ressemblent par leur mode de production résolument artisanal, entièrement dépendant d'un seul individu crucialement responsable du développement et très peu substituable.

1.4.2.3 Produits hors commerce et sur mesure

Pour mémoire, citons des logiciels développés en vue d'une mission particulière.

Relèvent de cette catégorie :

- des outils de recherche pour le développement d'un modèle. Ainsi le logiciel d'application du modèle MATISSE (Morellet, 1990, 1996), et le logiciel MARS (Leurent, 1995d) destiné à tester divers composants sémantiques et divers algorithmes.
- des logiciels de planification taillés sur mesure pour des exploitants de réseaux de transport (ex. DREIF, RATP).

1.4.3 Critères pour choisir un logiciel d'affectation

Une enquête auprès des praticiens français des études de trafic (Leurent, 1996d) a mis en évidence deux populations :

- des praticiens en situation contrainte par des dispositions administratives (cas d'ARIANE) ou parce que le choix de méthode relève du niveau hiérarchique supérieur (cas des bureaux d'étude privés). Ces praticiens sont captifs d'un logiciel donné qu'il ne leur appartient pas de remplacer par un autre.
- des praticiens en situation de choix, souvent détenteurs de la responsabilité technique des études de trafic dans des bureaux d'étude privés. Ils emploient souvent plusieurs logiciels de façon complémentaire ou alternative, en fonction de leurs composants phénoménologiques respectifs.

Pour un projet d'équipement, voici les critères probables de choix d'un gestionnaire avisé :

- *largeur* du spectre de composants sémantiques, adéquation aux problèmes à traiter,
- *fiabilité du produit* (notamment absence de défaut dans l'algorithme ; le nombre de licences vendues est un indice de fiabilité mais pas une garantie),
- *sécurité d'approvisionnement* (à cet égard une production industrielle est préférable à une production artisanale),
- "*contrôlabilité*", avec la disposition à la fois de fonctions spéciales et d'une documentation suffisante (celle d'EMME/2 est exemplaire),
- *coût total*, incluant non seulement le prix d'achat ⁽¹⁾ mais encore le coût des défaillances, du manque de qualité.

La contrôlabilité n'est qu'un critère parmi d'autres pour l'exécution de l'étude de trafic. Mais dans des missions d'expertise ou de contre-expertise, elle devient le critère principal, essentiel. Il appartient aux acteurs chargés de ces missions d'exprimer leur demande auprès des producteurs de logiciels.

¹Précisons que le matériel informatique requis est généralement le micro-ordinateur compatible IBM-PC, qu'une licence d'un logiciel coûte quelques dizaines de milliers de francs

1.5 QU'Y A-T-IL À AUDITER ?

Récapitulons les principaux éléments établis :

- la disposition de modèles pour simuler la demande de transport et son affectation aux itinéraires (section 1.1).
- la portée des simulations dépend des composants phénoménologiques incorporés au modèle (section 1.2).
- les données d'application et leur mise en forme dans le modèle contribuent à déterminer le résultat, mais offrent des possibilités multiples et dangereuses de manipulation (section 1.3).
- les modèles de trafic sont mis en oeuvre à travers des logiciels ; l'application est-elle conforme, ou donne-t-elle lieu à des adaptations spécifiques ? Cela concerne notamment le but et la méthode du calcul du résultat. Pour en juger, il faut que le logiciel offre des fonctions de contrôle (section 1.4).

Un contrôle du modèle peut donc porter sur la composition sémantique, sur l'alimentation en données, sur la formulation mathématique et sur la procédure de calcul du résultat.

Au terme de "contrôle", ambigu puisqu'il pourrait marquer la soumission à des intérêts particuliers, nous préférons celui d'"audit" pour signifier un contrôle subordonné au seul intérêt de la Connaissance.

Les logiciels et les études nécessitent un audit tout autant que les modèles. Cependant l'audit des modèles paraît prioritaire car il conditionne celui des logiciels, puis celui des études. Dans les limites d'une thèse, nous concentrons notre attention sur deux questions :

- comment auditer un modèle ?
- auditer un modèle d'affectation bicritère du trafic.

La thèse définit le fondement de l'audit d'un modèle, et propose une méthode applicable à tout modèle. Le caractère pratique de la méthode est démontré par le traitement d'un modèle particulier, représentatif des principales difficultés techniques de modélisation.

Chapitre 2

Pourquoi auditer ?

Un modèle comprend plusieurs catégories d'aspects : composition conceptuelle, formule caractéristique, dispositif technique de résolution et alimentation économétrique.

L'audit technique est applicable à tout modèle, quels qu'en soient les composants. Ce deuxième chapitre évalue le besoin, la nature et les limites d'une telle entreprise.

Il se compose de trois sections.

D'abord nous opérons une synthèse des arguments de natures philosophique, sociologique ou pragmatique, favorables ou opposés aux modèles : nous adoptons une position médiane où un modèle est justifié *sous réserve de garanties techniques* (section 2.1).

Comment prendre de telles garanties ? C'est l'objet de l'*audit technique*. Nous précisons sa nature et nous annonçons la contribution de la thèse en ce sens (section 2.2).

Hors du plan technique se posent des questions délicates d'ordre social, institutionnel ou instrumental. Nous les évoquons brièvement et nous suggérons des pistes de réflexion (section 2.3).

2.1 POUR OU CONTRE UN MODÈLE ? SELON LES GARANTIES DONNÉES !

Quel crédit accorder aux résultats d'un modèle quantitatif ? Si les résultats sont douteux, à quoi sert le modèle ? Si le modèle est inutile, pourquoi l'employer ? Afin de développer à ces questions et d'apporter des éléments de réponse, opérons une synthèse des arguments pour ou contre l'usage d'un modèle.

2.1.1 Pour définir et qualifier un modèle

Le dictionnaire (Hachette, 1990) propose pour un modèle les définitions que voici :

- i) « schéma théorique visant à rendre compte d'un processus, des relations existant entre divers éléments du système ».
- ii) « ensemble d'équations et de relations servant à représenter et à étudier un système complexe ».
- iii) « ensemble de caractéristiques que possède en commun une famille de systèmes homomorphes ».

Résumons : un modèle est relatif à un système dont il opère une abstraction (iii, i), qu'il sert à représenter (i, ii) et à analyser ou simuler (ii). Parfois s'ajoute une fonction de prospection, lorsque le modèle incorpore une règle d'évaluation de projet et qu'il sert à définir, tester et classer des variantes.

Pour faciliter l'exposé des arguments pour ou contre la modélisation, convenons d'appeler :

- *modèle conceptuel*, ou *sémantique*, ou *phénoménologique*, le schéma théorique ; à ce stade il y a une identification d'éléments, des mises en relation des éléments, des définitions de systèmes et de sous-systèmes (donc d'environnements), une qualification des relations (causalité, simultanéité d'occurrence etc). Nous appelons *mécanisme explicatif* une relation élémentaire, et *contenu sémantique*, ou *composition sémantique*, l'ensemble des mécanismes explicatifs du modèle conceptuel. Par exemple, en matière d'affectation, identifier un déplacement à la consommation d'un bien économique "acheminement entre une origine et une destination" ; identifier un auteur de déplacement comme un acteur économique qui demande le bien Déplacement ; définir le choix d'itinéraire comme la recherche d'un plus court chemin, donc comme la recherche d'un bien de prix minimum ; supposer que les demandeurs connaissent, donc peuvent choisir, les chemins ; définir le prix du bien Déplacement comme un coût généralisé en précisant les attributs (frais monétaires, temporels...) qui y participent...

- *formule caractéristique*, la synthèse en langage mathématique des mécanismes explicatifs, sous forme d'un problème type de mathématiques appliquées. Le *modèle formel* désigne l'union du modèle sémantique et de la formule caractéristique. Pour obtenir la formule caractéristique d'un modèle, on peut définir une notation mathématique pour chaque élément, puis exprimer chaque mécanisme explicatif comme une relation élémentaire reliant par une fonction un élément (variable endogène de la relation) à d'autres éléments (variables exogènes de la relation), enfin assembler les relations élémentaires dans une *formule synthétique* qu'il reste à rendre caractéristique en la rattachant à un cadre mathématique de référence.

- *modèle résoluble* le modèle formel uni à un dispositif de résolution qui détermine les valeurs des variables endogènes en fonction des valeurs des variables exogènes. Par exemple, un modèle résoluble d'affectation peut comprendre un modèle d'affectation formel et un algorithme de résolution.
- *modèle alimenté* un modèle résoluble joint à des données d'application à un cas particulier donné.

2.1.2 Arguments en faveur d'une modélisation¹

2.1.2.1 Toute représentation abstraite contient un modèle sémantique

Un modèle sémantique rassemble des idées, des intuitions, relatives à un objet ou à un système². L'identification des composants et la mise en évidence des relations fournissent les ingrédients nécessaires à une représentation abstraite.

Une telle représentation peut remplir un rôle *didactique*, en livrant une image simplifiée d'un système éventuellement complexe. Elle permet aussi le *débat* et la *concertation* entre acteurs d'un processus de décision.

Elle sous-tend non seulement toute méthode quantitative visant à reproduire ou prédire un état empirique, mais encore toute analyse *prospective*, qui s'appuie sur une identification des principales relations pour imaginer des évolutions possibles. De la qualité du modèle conceptuel dépend celle de la prospective : pour la rigueur dans l'analyse, la fécondité dans l'imagination, ainsi que le réalisme dans le dessin des cheminements potentiels.

2.1.2.2 Les modèles formel et calculable sont des étapes nécessaires à la quantification

Le modèle formel remplit une fonction didactique plus achevée que le modèle sémantique : comme il procède d'une série d'arbitrages et de décisions, il est le vecteur d'une expertise. La formulation mathématique sert à une mise en cohérence qui prépare la quantification ; par ailleurs elle constitue un exercice d'aguerrissement intellectuel.

Le modèle résoluble précède nécessairement l'application quantitative : une formule sans dispositif de résolution demeure lettre morte.

2.1.2.3 Le modèle alimenté pour éclairer la décision

Le modèle alimenté remplit des fonctions de représentation et de simulation, le cas échéant de prospection. Les retombées pratiques sont évidentes³ : si le système étudié a (réellement ou virtuellement) une existence empirique influencée par les décisions de

¹mentionnons en aparté un argument esthétique, lié à l'attrait d'une architecture satisfaisante pour la Raison

²le système n'a pas nécessairement une existence empirique au moment de la modélisation conceptuelle

³dans la plupart des cas, elles motivent le développement du modèle, du stade conceptuel au stade quantitatif

certains acteurs¹ ou par les évolutions de certains facteurs², alors le modèle alimenté peut non seulement représenter une situation de référence mais encore simuler les effets de possibles décisions ou évolutions. Il peut ainsi servir à éclairer une décision.

Il peut aussi soutenir la compréhension qualitative et la prospective. Car, pour un sujet complexe tel que la demande de transport, l'analyse qualitative ne peut à elle seule hiérarchiser les facteurs déterminants, il faut une quantification pour apprécier leurs rôles respectifs.

2.1.3 Arguments opposés aux modèles

On recense des arguments de natures philosophique, sociologique ou encore technique.

2.1.3.1 Oppositions philosophiques : le modèle inutile ou impuissant

Quelle est l'*utilité finale* d'une prévision ? Si l'évolution du système est parfaitement prévisible, autrement dit déterminée à l'avance, à quoi bon la prévoir puisqu'elle surviendra nécessairement ? Cet argument dénie bien sûr tout libre arbitre aux acteurs. Il repose sur un axiome de prédestination³.

Une approche quantitative n'est-elle pas *fondamentalement impuissante*, impropre à prendre en compte tant la complexité du système étudié⁴, que son caractère évolutif⁵, ou que son instabilité⁶ ?

Ces deux oppositions philosophiques sont susceptibles de réfutations d'une même nature. Par des arguments empiriques, en montrant la justesse de certaines prévisions. Au plan utilitariste, économique : si certains acteurs acceptent de payer pour une modélisation, alors celle-ci est économiquement justifiée, d'autant mieux que son prix est plus élevé.

A l'argument sur l'impuissance fondamentale d'une approche quantitative, nous répondons encore que le bon artisan et le bon outil ne se substituent pas l'un à l'autre ; une analyse qualitative⁷ sera heureusement soutenue, prolongée par un modèle quantitatif capable de projeter systématiquement les principes explicatifs mis en jeu.

Notons enfin que l'interrogation sur l'utilité finale peut être retournée : si tout est déterminé à l'avance, cela vaut pour l'emploi du modèle, pour sa critique, pour la réfutation de la critique etc.

¹par exemple la décision de construire une route, dans le cas de l'affectation

²par exemple l'évolution du contexte démographique ou socio-économique

³Curien et Gensollen (1989) appellent cet argument le "sophisme paresseux"

⁴à la fois complexité d'abondance, en raison du nombre d'acteurs, et complexité de sens, car les relations sont imbriquées les unes dans les autres

⁵au niveau d'ensemble, par exemple la progression de la technologie, comme au niveau individuel, car les acteurs réagissent, s'adaptent et changent

⁶qu'appelle-t-on le hasard et où intervient-il ? Quelle part d'arbitraire joue sur tel ou tel plan ?

⁷le terme d'analyse marque un souci de précision et de rigueur ; il ne s'applique pas à certains verbiages

2.1.3.2 Oppositions sociologiques : le modèle facile à dévoyer ou nocif

Les oppositions de type sociologique tiennent essentiellement aux circonstances et conditions d'emploi d'un modèle.

Application naïve ou brutale. On trouve des mises en garde contre un emploi trop général ou trop rapide des modèles alimentés, fruit d'un enthousiasme immodéré pour appliquer à tous types de situations des moulinettes à chiffres.

Emploi "machiavélique" des prévisions quantitatives. Lorsqu'un modèle de simulation quantitative peut éclairer une décision (par exemple en matière de planification), si l'un des acteurs du processus commande l'étude, il est tenté d'utiliser le modèle à l'appui de sa stratégie propre (notion d'étude-plaidoyer). Plusieurs degrés existent : la dissimulation (omettre exprès certains aspects qui engageraient les autres acteurs à réfléchir), le travestissement (interpréter certains résultats de façon scabreuse), la falsification (obtention de résultats sur mesure grâce à un paramétrage ad hoc du modèle). Comment éviter de tels écueils ?

Répondons à ces deux arguments. Pourquoi faire un procès d'intention à ceux qui utilisent les prévisions quantitatives, même en les subordonnant à leurs propres objectifs ? Il appartient à chaque acteur de pousser à fond sa propre stratégie ; peut-être une "main invisible" vient-elle arbitrer l'ensemble au mieux...

Signalons que l'argument de l'emploi machiavélique exerce à terme des effets pernicieux : il détourne des modèles les praticiens et rétrécit progressivement le milieu technique, jusqu'à priver les modèles conceptuels ou formels de tout support empirique et de tout caractère expérimental.

Autre argument de nature sociologique, l'*éventuelle nocivité* d'un modèle qui éluciderait trop nettement les enjeux d'une décision : certains soutiennent qu'un processus de décision, pour son bon déroulement, nécessite une certaine ambiguïté, favorable à un "compromis dans le malentendu" qui agirait comme l'huile dans les rouages de la société. Nous répondons comme précédemment : il appartient aux acteurs du processus d'en juger. Nous croyons que plus souvent l'éclaircissement des enjeux à la lumière d'un modèle favorise la progression des débats.

2.1.3.3 Arguments pragmatiques

Enfin les techniciens des modèles présentent eux-mêmes des arguments qui induisent à limiter l'emploi de la modélisation :

- l'inutilité pratique d'une modélisation qui sert seulement à accompagner une décision déjà prise, non à éclairer une décision à prendre.
- la difficulté de formalisation et d'emploi d'un modèle conceptuel complexe, ce qui incite à se contenter de représentations plus simples.
- des contraintes économiques, en ressources humaines ou financières, qui forcent à réaliser l'étude de façon très grossière.
- l'incertitude quant aux données d'application, faute de mesures suffisantes.

Ces arguments sont recevables s'ils concluent des efforts sérieux pour améliorer le modèle ou simplement pour en faire un bon usage. Sinon, ils témoignent d'une abdication technique dissimulée sous un masque de sociologie.

2.1.4 Synthèse : des garanties sont nécessaires

2.1.4.1 Les arguments positifs sont recevables à condition d'une garantie technique

Les arguments en faveur de la modélisation tiennent à des possibilités d'emploi au service de buts déterminés : compréhension théorique, pédagogie, prospective, aide lucide à la décision. Ils sont donc conditionnels à la qualité technique du modèle. Peut-on prendre des garanties sur cette qualité technique, si oui de quelle manière ?

2.1.4.2 Contre les oppositions philosophiques, définir sa propre position philosophique

Nous avons vu que les oppositions philosophiques, liées à l'inutilité ou à l'impuissance d'un modèle, peuvent être réfutées par des arguments empirique ou utilitariste. Une réfutation de l'impuissance n'est jamais globale, elle reste locale à une application particulière, et même à une interprétation particulière de cette application : mais un reproche d'impuissance ne traduit lui-même qu'un doute global a priori. Pour nous, nous refusons de poser l'axiome d'une impuissance a priori de la modélisation ; chaque modèle peut tenter de faire ses preuves dans chaque application. Il reste à définir quelles preuves, ce qui nous ramène à la question des garanties techniques.

2.1.4.3 Réponses aux oppositions sociologiques

Les oppositions sociologiques appellent quant à elles des garanties concernant les circonstances d'emploi, les hypothèses d'application et les modalités d'interprétation. Au sujet des *circonstances d'emploi*, pour prévenir une application naïve ou brutale, il convient d'examiner le *proportionnement* du modèle au système à étudier, de prendre une garantie technique sur le modèle sémantique. Pour les *hypothèses d'application*, notamment les valeurs à donner aux variables exogènes, il convient d'adopter les hypothèses les plus crédibles et de *quantifier l'incertitude* qui s'y attache : en propageant cette incertitude via le modèle, on peut expliciter l'incertitude liée au résultat, ce qui donne une garantie technique quant au modèle alimenté¹.

Les *modalités d'interprétation* dépendent dans l'absolu du contenu du modèle, mais dans la pratique un commentateur agit en fonction de ses intérêts propres et de contraintes particulières ; l'imposition de *règles relatives à la présentation* des hypothèses et des résultats peut prévenir d'éventuels détournements. De telles règles procèdent d'arguments techniques mais relèvent de l'administration des processus de décision : en matière d'affectation, le ministère anglais des transports impose des règles de méthode et de présentation² pour toute étude de trafic routier en milieu interurbain (TAM, 1994).

Au total, les oppositions de nature sociologique suggèrent des précautions dans la conception et l'emploi des modèles, plutôt qu'elles n'en condamnent l'usage.

¹cela sous deux conditions : savoir quantifier l'incertitude relative aux données d'application, puis pouvoir la propager via le modèle (analyse de sensibilité)

²production de trois rapports : validation du modèle en situation de référence ; prévision en situation de projet ; évaluation des coûts et des avantages

2.1.4.4 Traitement des oppositions pragmatiques

L'argument de l'inutilité pratique des études concerne seulement la fonction d'aide à la décision, cela seulement dans des cas particuliers. Les autres fonctions d'un modèle, en particulier la fonction de connaissance, ne sont pas affectées.

Quant à l'éventuelle limitation du modèle calculable par rapport à la représentation conceptuelle, elle appelle autant à développer l'effort de modélisation qu'à refuser les applications auxquelles le modèle n'est pas proportionné.

La limitation des moyens et le manque de mesures sont des arguments proches du doute relatif aux données d'application ; nous leur répondons pareillement, en prescrivant de quantifier l'incertitude correspondante et de la propager via le modèle.

2.1.4.5 Bilan d'ensemble : nécessité de garanties techniques

Nous retenons la nécessité de prendre des garanties techniques quant au modèle :

- pour vérifier son utilité à des fins de compréhension théorique, de pédagogie, de prospective, d'aide à la décision.
- pour prouver ou infirmer une aptitude empirique dans le cas d'une application quantitative particulière.
- pour assurer le proportionnement au système étudié et aux questions posées.
- pour prendre en compte explicitement l'incertitude liée aux hypothèses d'application.

Pour donner ces garanties, il faut *auditer* le modèle, de la composition conceptuelle à l'alimentation économétrique. Auditer le modèle, au sens de détecter les éventuelles erreurs, d'en prendre la mesure et si possible de les réduire. Un audit favorable démontre l'utilité du modèle à la connaissance du système étudié. Nous postulons positivement la possibilité et l'intérêt d'une telle connaissance.

2.2 L'AUDIT POUR PRENDRE DES GARANTIES

Cette thèse vise à contribuer de façon technique à l'audit d'un modèle. L'audit technique constitue en lui-même un vaste domaine de recherche, qu'il convient de baliser (§2.2.1). Nous justifions le choix de l'affectation du trafic comme terrain expérimental (§2.2.2), puis nous expliquons la contribution de la thèse (§2.2.3).

2.2.1 Quelles garanties ?

Pour démontrer (définir ?) le bien fondé d'un modèle, il faut prendre des garanties techniques, en effectuant plusieurs contrôles.

Un premier contrôle porte sur la similitude du modèle conceptuel à la représentation abstraite de la réalité : nous l'appelons *audit conceptuel* ou *sémantique*.

Un deuxième contrôle porte à la fois sur la consistance logique, autrement dit la cohérence, et sur la résolubilité du modèle formel, afin de s'assurer que les dépendances suffisent à déterminer les variables endogènes. Nous l'appelons *audit formel*.

Un troisième contrôle porte sur le dispositif technique de simulation et sa capacité effective à produire le résultat attendu. Nous l'appelons *audit algorithmique*.

Un quatrième et dernier contrôle porte sur la validité empirique du modèle, sur sa capacité à reproduire des observations, sur la fiabilité de sa simulation. Nous l'appelons *audit économétrique*.

Ces quatre étapes de contrôle constituent l'*audit technique* du modèle, par définition. Nous en avons précisé les enjeux dans la section précédente.

Nous nous plaçons délibérément sur un plan purement technique, faisant abstraction des considérations d'ordre social ou institutionnel liées à des circonstances particulières d'emploi de modèles dans des processus de décision. La prochaine section évoque brièvement ces aspects.

2.2.2 Pourquoi auditer l'affectation du trafic ?

2.2.2.1 Choix d'un champ d'application restreint...

De fait la modélisation dans son ensemble manque d'audit technique : trop rares sont les études confrontant systématiquement les résultats d'une simulation à des réalisations (Cf. chapitre 3). Ce manque contribue certainement à la diversité des méthodes employées en pratique, car l'utilisateur potentiel n'a guère de lignes directrices à suivre.

2.2.2.2 ... mais porteur d'enjeux pratiques considérables

L'affectation du trafic n'est qu'une partie de la modélisation des déplacements, mais une partie importante puisque c'est elle qui réalise la rencontre entre offre et demande d'acheminement, qui produit les flots sur le réseau, qui calcule les niveaux de service. Cela justifie donc de la choisir comme point de départ.

Notre intérêt pour l'affectation trouve une autre justification dans les enjeux pratiques liés aux modèles qui l'opèrent : la planification des réseaux de transport entraîne des dépenses considérables, tandis que leur exploitation conditionne la mobilité des personnes et des biens.

2.2.2.3 ... et représentatif des difficultés techniques

Enfin, si un modèle d'affectation n'a qu'un contenu phénoménologique limité, il présente des caractères qui nécessitent un traitement technique spécial :

- la formulation implicite des dépendances entre variables endogènes et exogènes, caractéristique des modèles d'équilibre, requiert des outils mathématiques et statistiques perfectionnés (par exemple pour la propagation de l'erreur exogène).
- la solution est généralement obtenue par l'emploi d'un algorithme itératif, ce qui amène la question de la convergence d'un tel processus.
- en raison du caractère spatial, les applications concrètes sont souvent porteuses d'une complexité d'abondance (nombreuses routes schématisées dans le réseau, nombreuses relations origine-destination) ; elles constituent alors des problèmes mathématiques de grande dimension.

Ces caractères soulèvent les principales difficultés mathématiques rencontrées dans la modélisation, du trafic ou autre.

2.2.3 Contribution de la thèse

2.2.3.1 Contenu

La thèse traite de l'audit technique d'un modèle. Elle apporte une analyse de l'erreur et de l'incertitude dans un modèle, ainsi qu'une méthode d'audit technique selon quatre dimensions : audit conceptuel, audit formel, audit algorithmique, audit économétrique.

Elle fait une démonstration sur l'exemple d'un modèle d'affectation bicritère du trafic : toute application de la méthode est particulière, même si la méthode a un caractère général. Afin de préparer l'application à d'autres modèles, la méthode proposée procède d'abord à une capitalisation du domaine, cela pour chaque dimension.

L'audit technique ainsi défini et traité consiste à rendre transparents tous les aspects du modèle : autant le mécanisme de simulation que les inputs. Cette transparence sert à délimiter la portée explicative du modèle, autant qu'à attacher aux résultats l'incertitude qui provient des inputs.

2.2.3.2 Hypothèses de notre entreprise

Prévenons ici du caractère relatif de l'entreprise, qui repose sur des certitudes préalables. Nos principales hypothèses sont :

- la récurrence de certains phénomènes, la reproduction des résultats d'une expérience dans la mesure où l'on reproduit les conditions expérimentales.
- la possibilité d'une connaissance scientifique des phénomènes reproductibles ; connaissance bâtie par la description, l'analyse et l'intuition des phénomènes, ainsi que la déduction.
- la possibilité d'incorporer la connaissance scientifique des phénomènes dans des modèles, formulables à la fois en langage naturel et en langage mathématique basé sur la théorie classique des ensembles.

2.2.3.3 Méthode

Les différentes dimensions de notre audit relèvent de plusieurs disciplines : analyse systémique pour l'audit conceptuel, logique et recherche opérationnelle pour l'audit formel, algorithmique (entre autres) pour l'audit algorithmique, économétrie et recherche opérationnelle pour l'audit économétrique. Donc l'audit technique nécessite d'articuler des concepts relatifs à des disciplines qui toutes sont foisonnantes...

L'auteur prie le lecteur de consentir certains efforts de compréhension : entre disciplines différentes il arrive qu'un même concept s'exprime par des termes différents, ou qu'un terme commun corresponde à des concepts différents. Nous avons forgé notre propre vocabulaire à partir de termes des diverses disciplines, en retenant les acceptions communes ou en imposant un vocable particulier.

2.3 ASPECTS SOCIAUX ET INSTRUMENTAUX

Les chapitres suivants font délibérément abstraction de considérations d'ordre social ou institutionnel. De même ils ne traitent pas des aspects instrumentaux relatifs à l'application des modèles. La présente section examine rapidement ces sujets, sans les traiter avec la profondeur qu'ils méritent.

2.3.1 Plan social et institutionnel

Par sa nature, son développement technologique, son organisation et ses croyances, la société influe sur le modèle et son usage. Voici quelques indices relatifs à l'affectation :

- l'hypothèse que l'auteur du déplacement procède lui-même au choix d'itinéraire est inhérente à une société libérale.
- les études de trafic actuelles mobilisent des ressources informatiques importantes, inimaginables deux générations plus tôt.

Réciproquement le modèle influe sur la société :

- par ses usages didactiques, il transmet et diffuse des éléments de compréhension.
- par les conséquences pratiques de ses applications.

Ces conséquences pratiques méritent un examen.

2.3.1.1 Le modèle pour éclairer ou pour appuyer

Un acteur qui intervient dans un processus de décision a deux raisons fondamentales pour employer un modèle : éclairer un problème, soutenir une argumentation¹. Ces deux fonctions se combinent de multiples manières selon l'objectif de l'acteur :

- éclairer le mieux possible et partager les résultats, sans plaider pour un projet donné.
- soutenir un projet par la "caution scientifique" du modèle, employé avec le souci d'un éclairage de qualité.
- réclamer la "caution scientifique" mais sans souci de qualité d'éclairage, avec jusqu'à l'emploi forcé du modèle, ou la présentation de résultats qui n'ont aucun rapport avec le modèle...
- éclairer le mieux possible et garder les résultats pour soi.
- idem, tout en soutenant un projet donné par une application forcée.

Il ne nous appartient pas de juger ces manières d'employer les modèles. Notre projet se borne à l'appréciation technique.

¹rappelons la féconde analogie entre modèle et lampadaire : la vocation naturelle du lampadaire est *d'éclairer*, mais "l'ivrogne" s'en sert pour *se soutenir, s'appuyer*

2.3.1.2 L'exemple de la planification des transports

Le terme d'*étude-plaidoyer* a été forgé pour désigner certaines études produites pour justifier des projets de transport, de manière outrageusement partielle. Toute indignation à cet égard serait naïve : rien d'étonnant qu'un acteur pousse à fond son projet en faisant feu de tout bois !

Plus profondément, certains ont dénoncé l'emploi des modèles de trafic au service d'une idéologie, par exemple celle du "tout automobile" au sujet des modèles de trafic urbain (Dupuy, 1975). En France, la critique a porté ; désormais la Puissance Publique est plus circonspecte vis-à-vis de ces modèles. Depuis lors, les connaissances et les technologies ont progressé, tandis que leur prix baissait. L'époque où seul l'Etat disposait des ressources nécessaires à la modélisation est révolue, d'autres acteurs se sont pourvus de modèles : des collectivités locales, des exploitants de transport, des banquiers, et jusqu'à des bureaux d'étude propriétaires de leurs modèles.

Dans ces conditions le jeu d'acteurs devient plus équitable : la pseudo caution scientifique n'est plus déterminante puisque partagée. Parfois le rapport des forces s'est renversé entre la Puissance Publique et les autres acteurs : certains acteurs, moins contraints, ont mis au point des modèles plus performants dont ils conservent jalousement les résultats, ce qui leur confère une position de force !

2.3.1.3 Une grille de lecture

Pour apprécier le rôle potentiel ou réel d'un modèle dans un processus de décision particulier, nous suggérons les questions que voici :

- qui sont les acteurs ? Quels sont leurs objectifs et contraintes respectifs ?
- Parmi les acteurs, qui dispose d'un modèle ?
- qui en fait état ?
- qui partage les résultats de son modèle ?
- lorsqu'il y a partage, est-il libre ou contraint ? De quelle manière ?
- les résultats partagés sont-ils expliqués par un rapport détaillé, transparent ? Leur producteur autorise-t-il l'accès aux données d'application ?

Ces questions ouvrent la voie d'un audit "socio-institutionnelle" des modèles.

2.3.2 Plan instrumental

Les caractères instrumentaux de l'audit technique d'un modèle recouvrent à la fois :

- les nécessaires "instruments" : ressources humaines, financières et matérielles, ainsi que leur organisation.
- les usages techniques d'un tel audit, lui-même conçu comme un instrument à des fins pratiques : contrôle d'un logiciel d'application¹ du modèle, contrôle d'une étude produite avec le modèle.

¹par abus de langage, nous appelons *logiciel* le dispositif de résolution d'un modèle, même dans les cas où il n'est pas de nature informatique. Ce choix est expliqué au chapitre 5.

Concernant les ressources nécessaires à l'audit d'un modèle, leur volume dépend probablement de la complexité du modèle : auditer tout un système de modèles exige plus de moyens qu'auditer un seul des modèles dans ce système ! La structure des ressources doit quant à elle respecter la pluralité des disciplines : il faut mobiliser plusieurs compétences scientifiques, donc probablement plusieurs agents. Du côté des outils, nous avons discuté au chapitre précédent des fonctions souhaitables dans un logiciel d'affectation afin de contrôler les données.

Sans nous étendre davantage sur la question des ressources, nous allons approfondir celle des usages techniques. La notion de qualité sera notre point de départ (§.1) pour définir un contrôle de qualité (§.2). Celui-ci requiert une transparence qui s'exprime par des vecteurs privilégiés (§.3). Une répartition rationnelle des rôles entre concepteurs de modèles, développeurs de logiciels et chargés d'étude pourrait faciliter la production d'études "auditables" (§.4), ainsi qu'un contrôle décentralisé (§.5).

2.3.2.1 Qualité du modèle, du dispositif de résolution, de l'étude

Les instances de normalisation (ISO 8402 puis ISO 9000) définissent la qualité comme "l'ensemble des propriétés et caractéristiques d'un produit ou service qui lui confère l'aptitude à satisfaire des besoins exprimés ou implicites". Elles distinguent entre d'une part la qualité de définition, "usage qui devient qualité requise", demandée par le maître d'ouvrage au maître d'oeuvre, et d'autre part la qualité de réalisation, "moyen de parvenir à la qualité requise", demandée par le maître d'oeuvre à l'entrepreneur.

Nous avons défini la qualité technique d'un modèle, enjeu de son audit technique, comme l'ensemble de quatre propriétés :

- portée explicative (ou conceptuelle, ou sémantique),
- correction formelle,
- résolubilité,
- transparence économétrique.

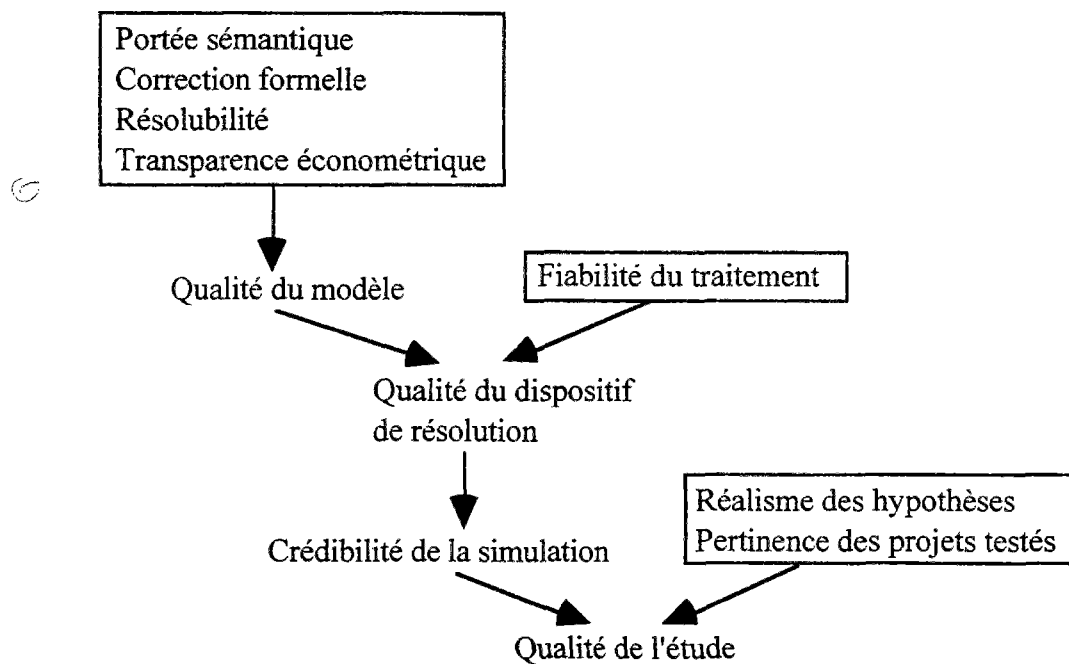
La qualité du dispositif de résolution (du logiciel, suivant notre abus de langage) recouvre la transparence du modèle traité et la fiabilité du traitement.

La transparence du modèle traité est implicite dans les procédures de certification logicielle, qui se concentrent sur la fiabilité du traitement. Celle-ci ne vaut cependant que sous la condition expresse de la transparence du modèle : à quoi bon certifier un logiciel aux principes incompris ?

La qualité de l'étude recouvre quant à elle :

- le réalisme des hypothèses d'environnement,
- la pertinence des projets traités, leur efficacité,
- la crédibilité de la simulation, qui nécessite à la fois la fiabilité du dispositif de résolution et la qualité du modèle.

Fig. 2a. Qualité d'une étude par simulation.



2.3.2.2 Objet et nature d'un contrôle de qualité

Compte tenu des distinctions entre modèle, dispositif de résolution et étude, nous préconisons un contrôle de qualité adapté à chaque objet. Notre thèse prescrit la nature d'un contrôle de qualité pour un modèle :

- vérifier la richesse et la pertinence de la composition sémantique,
- vérifier la correction sémantique et syntaxique de la formule caractéristique,
- contrôler la résolubilité,
- pour l'alimentation économétrique, apprécier et propager l'erreur exogène.

Suggérons des éléments pour contrôler la qualité d'un logiciel :

- élucider les modèles traitables, en fonction des options ouvertes à l'utilisateur. Une documentation technique devrait remplir cette mission, ce n'est cependant pas toujours le cas !
- pour chaque modèle traitable, contrôler la conformité des résultats du logiciel à des résultats attendus, grâce à des jeux de données de test.

Le contrôle de qualité d'une étude engage encore :

- une inspection de la méthode (Cf. modèle et dispositif de résolution),
- un jugement critique des hypothèses d'environnement et des projets testés.
- un commentaire des résultats, afin éventuellement de mettre en lumière les aspects "omis" par le commanditaire...

2.3.2.3 Les vecteurs privilégiés de la transparence

Les éléments du contrôle de qualité partagent une même condition nécessaire, la transparence de l'objet à contrôler. Pour assurer la transparence, il ne suffit pas d'assurer

le libre accès d'un contrôleur aux objets à contrôler : ce serait sous-estimer grandement leur complexité.

La transparence se prépare.

Conformément au vieil adage "verba volent, scripta manent"¹, ses vecteurs privilégiés sont des documents adaptés aux objets concernés : rapport de recherche pour un modèle, document technique pour un logiciel, rapport de présentation pour une étude. Sans préparatifs en ce sens, la qualité tout comme son contrôle deviennent aléatoires !

2.3.2.4 Vers une répartition rationnelle des rôles ?

Comment organiser rationnellement à la fois la production de modèles, de logiciels, d'études, et leur contrôle de qualité ? Nous préconisons une division du travail dans la production, qui s'accorde avec une séparation des pouvoirs dans le contrôle de qualité.

La division du travail consisterait à séparer la production par fonction : recherche pour la production de modèles, fonction de développement logiciel, fonction de bureau d'étude.

En matière d'affectation, les bureaux d'étude qui développent de façon industrielle des logiciels séparent la cellule de développement de la cellule d'utilisation (Cf. l'organisation de MVA par rapport au logiciel TRIPS, celle d'INRO par rapport à EMME/2).

Les confusions entre rôles nuisent non seulement au contrôle de qualité, mais encore :

- au modèle lorsqu'il est confondu avec un logiciel, donc vraisemblablement insuffisamment décrit et explicité.
- au logiciel lorsqu'il est confondu avec un modèle mal décrit et pas certifié : comment décider du contenu du logiciel ?
- le chercheur qui se fait développeur de logiciel se consacre à la diffusion commerciale au détriment de l'information scientifique.
- le développeur qui se fait chercheur ne se trouve pas dans les meilleures conditions pour faire valoir ses idées originales.

2.3.2.5 Pour un contrôle mutuel et réciproque

Une claire séparation des rôles entre recherche, développement logiciel et étude permet de décentraliser le contrôle de qualité, de le répartir entre intervenants de façon qu'ils l'exercent sur eux-mêmes ou l'un sur l'autre (figure 2b).

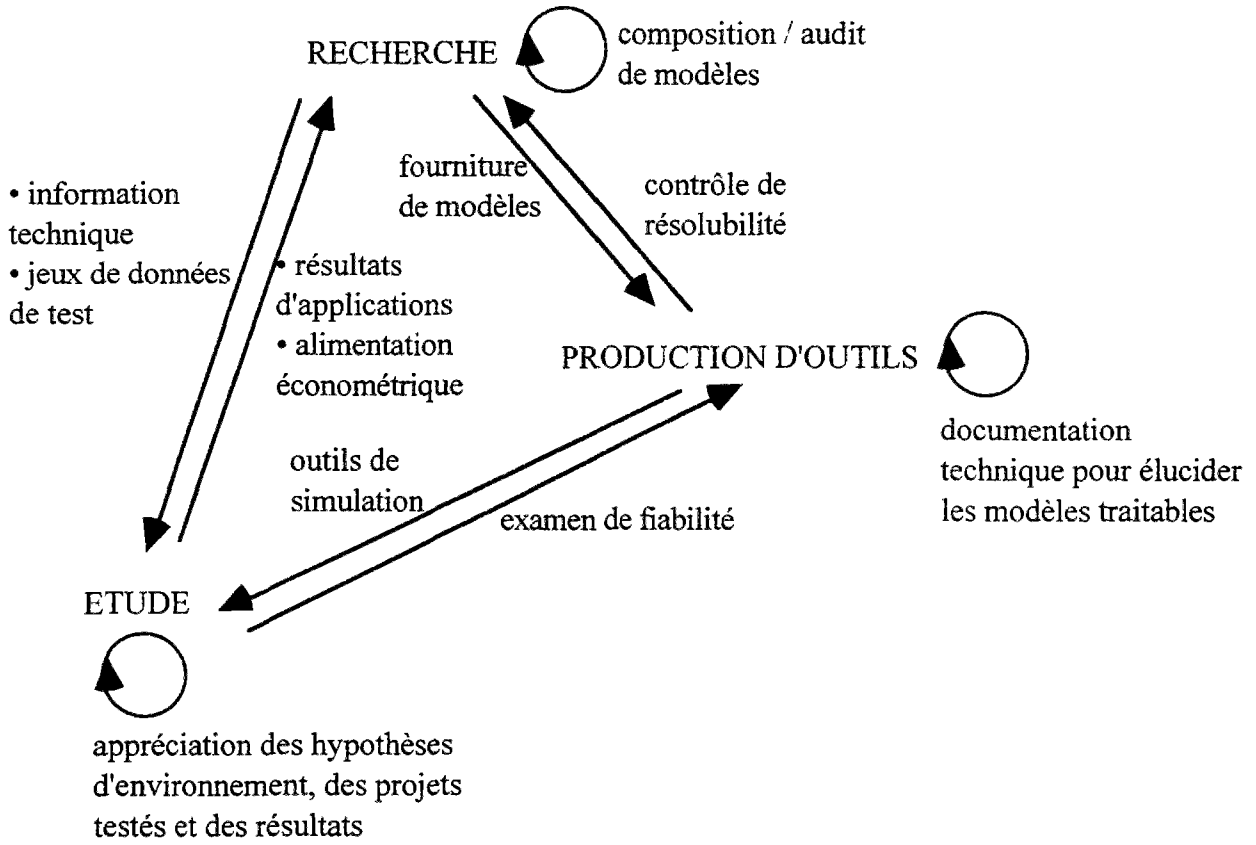
Nous avons commenté ailleurs les possibles implications de la Puissance Publique dans l'encadrement méthodologique des études et la production d'outils (Leurent, 1996d). Une autre contribution possible serait de subventionner un organisme auditeur ou certificateur, chargé :

- d'une mission légère d'accompagnement, de formalisation et de mémoire du processus décentralisé.

¹les paroles s'envolent, les écrits demeurent

- ou d'une mission plus lourde qui ajouterait un pilotage du processus, la responsabilité d'une formation, des actions à long terme (capitalisation méthodologique) ou ponctuelles (ex. un banc d'essai de différents logiciels).

Fig. 2b. Un processus décentralisé d'audit.



Chapitre 3

Antécédents et ressources

Notre ambition d'audit technique naît d'un double constat : insuffisance des pratiques actuelles, et manque de recherches spécifiques, particulièrement en matière d'affectation du trafic. Afin de préciser le constat, ce troisième chapitre dépeint l'état de la pratique et de la recherche.

Le champ de notre investigation s'étend non seulement à l'affectation mais aussi au choix modal, pour deux raisons. Primo les modèles de choix modal sont beaucoup plus proches de l'état de l'art économétrique. Secundo les méthodes de choix modal valent aussi pour l'affectation, en ce qui concerne la répartition du trafic entre itinéraires concurrents et particulièrement la formulation des fonctions de coût généralisé (théorie de l'utilité aléatoire du consommateur).

Nous commençons par un bref rappel du discours économétrique classique, pour relativiser la portée de l'économétrie par rapport au jugement a priori, et dresser une première liste de types d'erreur à considérer de façon particulière : erreur de spécification, erreur de fluctuation d'échantillonnage, erreur de mesure (section 3.1).

Nous suivons approximativement cet ordre pour analyser les modèles de choix de mode ou d'itinéraire, en considérant successivement la spécification (section 3.2) et l'estimation (section 3.3). D'autres contributions sont évoquées ensuite, qui portent les unes sur l'erreur de calcul, les autres sur la propagation de l'erreur via le modèle (section 3.4).

Nous terminons en récapitulant l'existant et en identifiant une source d'erreur jamais invoquée : l'erreur de formulation (section 3.5).

3.1 LE DISCOURS ÉCONOMÉTRIQUE CLASSIQUE

Pour présenter le discours économétrique classique sur l'incertitude et son contrôle, nous allons développer l'exemple du modèle de régression linéaire simple : cela donnera un tableau concret qu'il suffira de compléter par quelques touches pour traiter le cas général¹.

3.1.1 Types d'erreur dans la théorie statistique standard

3.1.1.1 Le modèle de régression linéaire simple

Le modèle de régression linéaire simple consiste en "l'explication" d'une variable endogène (l'effet), par une variable exogène (la cause). Par exemple la quantité d'une récolte à expliquer par la quantité d'engrais. On formule :

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

avec

- i indice d'observation,
- Y_i variable endogène, expliquée,
- X_i variable exogène, explicative,
- α, β coefficients (paramètres) à estimer, appelés respectivement l'ordonnée à l'origine et la pente,
- ε_i terme aléatoire pour intégrer les parties du phénomène inexpliquées par l'analyste.

Les hypothèses probabilistes sur les termes aléatoires ε_i déterminent le modèle. Celui-ci repose sur six hypothèses basiques :

- (H1) le modèle est linéaire en X_i .
- (H2) les valeurs des X_i sont observées sans erreur.
- (H3) $\text{Cov}(X_i; \varepsilon_j) = 0$ entre les erreurs et les variables explicatives.
- (H4) les erreurs ε_i sont centrées : $\text{E}(\varepsilon_i) = 0$.
- (H5) les erreurs ε_i ont la même variance : $\text{E}(\varepsilon_i^2) = \sigma_\varepsilon^2$.
- (H6) les erreurs ε_i vérifient $\text{Cov}(\varepsilon_i; \varepsilon_j) = 0$ si $i \neq j$.

3.1.1.2 Hypothèses basiques ou supplémentaires

Les hypothèses (H1)-(H6) constituent la base du modèle de régression linéaire simple ; elles sont nécessaires pour livrer des estimateurs de α et β au sens des moindres carrés

¹Pour une présentation plus exhaustive, mais moins ciblée, nous renvoyons le lecteur intéressé aux manuels d'économétrie : Wonnacott et Wonnacott (1991), Judge et al. (1988), Bourbonnais (1993), entre autres

ordinaires. Pour aller plus loin, notamment pour effectuer des tests paramétriques à des fins d'inférence, il faut ajouter des hypothèses supplémentaires comme la normalité des erreurs.

3.1.1.3 Interprétation de l'erreur

Dans le modèle de régression linéaire simple, le terme aléatoire ε_i regroupe trois postes :

- erreur de spécification, si la variable explicative ne suffit pas à expliquer toute la variabilité.
- erreur de mesure (d'observation) sur les variables endogènes Y_i .
- erreur de fluctuation d'échantillonnage, car d'un échantillon à l'autre les observations (donc les estimations) varient.

Comme l'hypothèse (H2) est particulière, dans le cas général l'erreur de mesure tient également aux X_i .

Signalons que l'erreur de spécification économétrique n'est pas synonyme d'une erreur sémantique. Certes une erreur sémantique entraîne une erreur de spécification économétrique et probablement, de façon empirique, une insuffisante explication de la variabilité. Mais une variabilité bien expliquée ne signifie pas pour autant l'absence d'erreur sémantique : parfois la variabilité est "expliquée" par de mauvaises raisons (la confusion classique entre causalité et corrélation).

3.1.2 Quels traitements de l'erreur ?

Traiter l'erreur signifie tenter de la réduire, mais encore et de façon plus fondamentale, tenter d'en prendre la mesure et d'en tenir compte dans la prédiction.

3.1.2.1 Agir sur la spécification

Premier traitement possible, changer la spécification du modèle en changeant l'ensemble des variables explicatives (remplacer les X_i par des Z_i d'interprétation différente), ou les spécifications fonctionnelles (passer à une régression non-linéaire etc.). Mais la spécification renvoie aussi à d'autres critères que l'ajustement sur un échantillon : intervient également le *jugement a priori*, à base d'intuition et d'expériences antérieures, qui contraint fortement le dessin du mécanisme explicatif.

3.1.2.2 Agir sur l'erreur de mesure

Pour réduire l'erreur de mesure, il convient d'améliorer la technique de recueil, d'observation, de l'échantillon.

Un autre traitement, en cas d'erreur de mesure importante sur les variables explicatives, consiste à prendre des variables explicatives "instrumentales", qui doivent être fortement corrélées avec les variables explicatives initialement considérées et non corrélées avec les erreurs ε_j .

3.1.2.3 Fluctuation d'échantillonnage

L'erreur de fluctuation d'échantillonnage est justiciable de deux traitements. Primo, réduire a priori cette erreur en accroissant la taille d'échantillon (cela augmente le coût d'enquête), ou encore en choisissant un meilleur estimateur. Le second traitement n'est

pas une réduction mais une prise en compte : il s'agit d'attacher à la prévision l'erreur qui résulte de la fluctuation d'échantillonnage¹.

3.1.2.4 Prendre la mesure de l'erreur de prévision

Pour prendre la mesure de l'erreur tant de spécification que d'échantillonnage, il convient de la propager par le mécanisme du modèle, afin d'attacher l'incertitude propagée à toute simulation. Précisons cela pour le modèle de régression linéaire simple².

Au point X_0 , on veut prédire la valeur moyenne de Y . La meilleure estimation est le point de la droite estimée, $\hat{Y} = \hat{\alpha} + \hat{\beta}X_0$, sur lequel on répercute l'erreur d'estimation des paramètres α et β en attachant une incertitude quantifiée à l'aide de l'écart-type

$s\sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2}}$, où s^2 estime la variance résiduelle (donc σ_ε^2). Pour prédire

maintenant non pas la valeur moyenne de \hat{Y} au point X_0 , mais toute la distribution de

probabilité, il faut considérer l'écart-type $s\sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum(X_i - \bar{X})^2} + 1}$ qui appuie

l'intervalle de prédiction. Le terme supplémentaire sous la racine prend en compte la dispersion autour de la moyenne (rôle potentiel d'un ε_{X_0}).

On constate logiquement que l'incertitude sur une valeur particulière est supérieure à celle sur une valeur moyenne.

3.1.3 Règles de sélection des modèles

Pourquoi, au nom de quoi, préférer un modèle à un autre, pour représenter un phénomène ? Les réponses à cette question procèdent de trois arguments principaux (au delà de la correction syntaxique et de la transparence économétrique, que nous supposons assurées) :

- la portée sémantique, i.e. la proximité des mécanismes explicatifs à la représentation théorique de la réalité.
- l'accord à l'expérience, i.e. la capacité à reproduire des observations.
- la simplicité d'expression.

Ces trois arguments sont de nature et portée différentes. Comparons les à deux critères de qualité et un critère de prix, pour un produit échangé sur un marché économique.

Les approches quantitatives de ce débat se restreignent à opposer les deux derniers critères, sans toucher au premier, dont l'appréciation reste l'affaire du seul jugement a priori. Les approches se rangent en deux classes : la première classe regroupe des comparaisons entre modèles rattachables à un sur-modèle commun, alors que les méthodes de la deuxième classe valent pour tous les cas³.

¹et particulièrement de la variance des estimateurs, propagée par le mécanisme du modèle

²Cf. Wonnacott & Wonnacott (1991) chapitre 12, et l'exemple traité au §10.2.1

³mais si les méthodes de la première classe s'appliquent, elles sont préférables

3.1.3.1 S'il existe un sur-modèle commun

Lorsqu'il existe un sur-modèle commun à deux modèles rivaux, il constitue une référence par rapport à laquelle la qualité de chaque modèle rival se mesure. En effet chaque modèle rival apparaît comme une restriction du sur-modèle, ce qui se traduit par des contraintes sur les valeurs de certains paramètres. Il existe des tests pour juger du bien fondé de telles contraintes ; on peut sélectionner le modèle rival pour lequel la probabilité critique d'acceptation des hypothèses spécifiques est la plus élevée, entre autres principes (Cf. Horowitz 1982).

3.1.3.2 Ratios qualité-prix

On évalue un rapport entre "qualité d'ajustement" et "prix d'obtention" (nombre de paramètres), cela pour chaque modèle, afin de retenir le modèle qui présente le rapport le plus élevé.

Comment calculer le rapport qualité-prix ? Akaike (1973) propose le critère AIC (An Information Criterion) formulé comme la différence entre la log-vraisemblance et le nombre de paramètres, le tout divisé par le nombre d'observations. Il existe bien d'autres formules comparables (Judge et al., 1988 ; Walter et Pronzato, 1994).

Cependant les considérations statistiques invoquées sont d'ordre très général, elles renvoient à l'intuition physique, sans juger du phénomène représenté. On peut encore leur reprocher leur mode de décompte des paramètres : certaines spécifications fonctionnelles, avec quelques paramètres seulement, contiennent en fait une très riche information (Olivier Morellet).

3.1.4 Bilan

Au total, l'économètre dispose de concepts et d'outils pour traiter certains types d'erreur, dans un domaine bien circonscrit qui n'englobe pas toute l'erreur. Il reconnaît l'importance du jugement a priori dans la spécification d'un modèle. Il convient qu'il est vain de chercher à circonscrire le jugement a priori dans la théorie économétrique¹.

Pour notre recherche, nous retenons donc :

- le rôle du jugement a priori,
- que l'économétrie limite son objectif au développement d'une approche qui vise à la cohérence interne et à l'accord avec l'expérience, sans viser une connaissance absolue² du phénomène réel.
- dans ces limites, la présence parmi l'erreur de plusieurs postes distincts.

Pour la suite, nous regroupons l'erreur de mesure et l'erreur de fluctuation d'échantillonnage dans une "erreur d'estimation".

Annonçons dès maintenant que nous allons identifier d'autres sources d'erreur, dont la théorie économétrique classique ne traite pas car elles n'interviennent pas dans ses

¹une branche importante de l'économétrie, l'analyse bayésienne, sert spécifiquement à intégrer de l'information a priori : cela permet d'incorporer une part de jugement a priori relativement aux valeurs des variables, mais pas aux mécanismes explicatifs

²qui serait surnaturelle...

modèles archétypes (comme la régression linéaire) : à savoir l'erreur formelle et l'erreur algorithmique.

3.2 QUESTIONS DE SPÉCIFICATION

La spécification recouvre à la fois le codage (donc des schématisations et des discrétisations) et les mécanismes explicatifs. Dans un problème économétrique standard, le codage se réduit bien souvent à des choix intuitifs, par exemple considérer des séries de données sur une base annuelle plutôt que mensuelle s'il s'agit de faire des prévisions annuelles ; en matière de modélisation du trafic, le codage comprend de nombreux choix moins évidents à justifier, qui soulèvent des questions spécifiques. Aussi nous revenons au codage pour approfondir les considérations de la section 1.3 (§3.2.1), avant d'aborder la spécification des mécanismes explicatifs (§3.2.2).

3.2.1 Le codage, face cachée de la modélisation ?

3.2.1.1 Quelques opérations de codage

Pour simuler le choix modal et l'affectation, il y a lieu de schématiser, coder, l'offre et la demande d'acheminement. Le codage de l'offre comprend :

- l'identification d'un ou plusieurs modes de transport.
- pour chaque mode, l'abstraction du service réel en une représentation de type informatique : identification des sommets (carrefours) et des arcs (routes), rattachement à chaque arc de caractéristiques diverses (longueur, capacité, prix, temps de traversée...). A partir de ces éléments, on construit (habituellement de façon automatique, grâce à un logiciel ad hoc) des itinéraires modaux dont on calcule les attributs de temps, prix, confort, sécurité...

Le codage de la demande comprend :

- le découpage du territoire d'étude en zones d'émission et de réception des déplacements. La constitution des matrices origine-destination relève de phases antérieures au choix modal et à l'affectation, dans le processus de modélisation de la demande de transport.
- pour chaque relation origine-destination, la spécification des comportements, souvent la segmentation en plusieurs groupes aux comportements distincts, dont chacun est caractérisé par ses conditions d'accès aux réseaux modaux, par sa valeur du temps, par l'importance attachée au confort, etc.

3.2.1.2 La pratique courante

Toutes ces opérations relèvent de la spécification du modèle, on pourrait dire qu'il s'agit de la spécification "cachée" du modèle. Leur lourdeur empêche généralement de les reprendre, les améliorer, au cours d'une étude. Ces tâches sont actuellement peu formalisées : les scientifiques de service les délèguent aux techniciens et s'en éloignent vite. Aussi le domaine ne progresse guère au point de vue méthodologique, les seules prescriptions sont d'ordre général.

Le praticien a pour références en matière d'affectation :

- son expérience personnelle, transmise par son formateur et enrichie par la suite.

- des principes très généraux, récapitulés dans les manuels d'enseignement de la planification des transports (Ortuzar et Willumsen, 1990 ; Thomas, 1991).
- des recommandations spécifiques, dans des manuels d'application publiés par les associations d'ingénieurs (ex. TRB 1983) ; qui indiquent par exemple les trafics émis ou reçus selon l'heure de la journée par une grande surface commerciale, en fonction du nombre d'emplois et de la surface marchande.
- des règles impératives, fixées par des documents administratifs. Ainsi, en France, pour les études d'avant-projet routier interurbain, les valeurs des paramètres comportementaux et des taux d'évolution du trafic sont imposées (DR, 1986).

Face à ces possibilités ou contraintes, le praticien conserve toujours beaucoup d'autonomie. Un encadrement rigide masque souvent une foule de petites libertés que s'accorde, comme en contrepartie, l'exécutant de l'étude : codage d'arcs fictifs, scission d'une origine-destination en plusieurs, pour mieux reproduire la situation existante ou se rapprocher des objectifs du commanditaire (Cf. section 1.3).

Un autre mode d'encadrement, sans doute plus efficace, consiste à imposer des *contraintes de résultat* tant sur les données codées que sur les résultats synthétisés par le modèle, et simultanément à fournir au praticien les outils (surtout logiciels) facilitant la *validation préventive* qui permet de satisfaire aux contraintes.

3.2.1.3 La pratique éclairée procède par validation préventive

Comme exemple de procédure pour la validation préventive, voici les recommandations du ministère anglais des transports pour une étude de projet routier interurbain (formulées dans le Traffic Appraisal Manual - Manuel d'Appréciation du Trafic - ou plus cursivement TAM, DOT 1994). Pour un réseau :

- contrôle manuel exhaustif des caractéristiques des arcs (type de route, capacité, vitesse libre...),
- vérification de la connexité par des recherches systématiques de plus court chemin,
- vérification du rapport de la longueur codée d'une route à la distance euclidienne entre les deux noeuds d'extrémité (seuil d'alerte de 1,3),
- temps de parcours : $\pm 10\%$ au seuil de confiance de 95%.

Du côté de la demande, le TAM indique comment attacher des variances individuelles à chaque case d'une matrice origine-destination (O-D) ; à chaque nouvelle mise à disposition d'information (ex. nouvelle enquête origine-destination dans le territoire étudié), les variances des cases concernées sont mises à jour. Les variances O-D doivent être affectées sur le réseau pour produire des variances relatives aux débits des arcs.

Autant de tâches répétitives, vite fastidieuses, qu'il convient d'automatiser tant que possible. Les principaux logiciels commerciaux d'affectation proposent (ou appliquent d'office) de nombreux tests en ce sens : VISEM/VISUM, POLYDROM, DAVIS... Mais seuls certains logiciels anglais, par exemple TRIPS, offrent en standard les opérations de dépiage prescrites par le TAM.

3.2.1.4 Contribution de la recherche : quelques études empiriques pour guider les démarcations statistiques

Par démarcation statistique, nous entendons toute opération visant à distinguer des classes : zonage du territoire, segmentation de la demande et différenciation des comportements, hiérarchisation des arcs du réseau, sélection des itinéraires, etc.

Concernant le zonage du territoire, on trouve des compte-rendus d'expériences diverses (Stover et al. 1972), des plaidoyers pour un traitement ponctuel des origines-destinations (Chapleau, 1993) et quelques investigations théoriques (McFadden et Reid, 1975, Bolduc et al., 1989, Ortuzar et Willumsen, 1990) qui révèlent l'existence d'un biais d'agrégation spatiale.

Le principe de segmenter la demande semble acquis en matière de choix modal mais pas complètement en matière d'affectation. Pour classer les individus, Heggie (1978) propose une typologie selon les dimensions suivantes : sexe, tranche d'âge, catégorie socioprofessionnelle, composition du ménage, position dans le cycle de vie, équipement en moyens de transport et accessibilité aux réseaux. Dans le modèle MATISSE (Morellet, 1990), les voyages sont rattachés à des segments différenciés selon : le motif, la valeur du temps, le moment souhaité, la fréquence, la taille de groupe, la durée de séjour. Bien d'autres propositions existent¹.

Concernant le codage et la hiérarchisation des réseaux, en matière d'affectation un consensus se dégage pour coder, dans l'ordre d'importance décroissante des routes, jusqu'au niveau inférieur au dernier niveau auquel la connaissance des débits affectés présente un intérêt².

Concernant la sélection des itinéraires, on manque singulièrement de références pour l'affectation tant unimodale que multimodale. Il existe quelques indications relatives à la propension des auteurs des déplacements à se rapprocher d'une ligne droite (Tagliacozzo et Pirzio, 1973), aux grands traits des itinéraires choisis (Antonisse et al. 1989, USAP 1992), aux arbitrages entre temps et distance, ou temps et prix (Marche, 1973) ; mais il n'y a pas un guide général. On abandonne généralement au modèle le soin de déterminer un itinéraire d'après un critère de plus court chemin ; pour l'affectation multimodale, il conviendrait de préciser les circonstances et les pratiques de changement de mode au cours d'un déplacement.

3.2.1.5 Bilan : un stade artisanal

Au total, il n'y a pas eu de recherche systématique consacrée à l'élaboration de principes de codage. On dispose actuellement de règles de bon sens, appliquées quand le praticien dispose d'un outil approprié.

3.2.2 Les mécanismes explicatifs

Partie émergée de l'iceberg de la spécification, les mécanismes explicatifs. Il s'agit des "relations causales" incluses dans le modèle, choisies parmi de multiples possibilités. Nous en avons dressé une liste au sujet de l'affectation statique du trafic routier

¹Cf. Jones et al., 1990, pour des références

²concrètement, si seuls comptent les débits sur autoroutes et routes nationales, il est bon de coder également les principales départementales

(Leurent, 1995d), pour l'affectation aux transports collectifs on peut consulter Spiess (1985) et De Cea et al. (1994), et Ortuzar et Willumsen (1990) pour le choix modal.

3.2.2.1 Des mécanismes explicatifs souvent anciens

La conception de la plupart des mécanismes explicatifs est déjà ancienne, souvent son origine est extérieure au domaine des transports : l'idée d'une interaction entre offre et demande remonte au moins au 18ème siècle, la théorie de l'utilité aléatoire du consommateur se précise dans les années 1960 (Lancaster, 1966). Parmi les contributions de l'économie des transports, citons la différenciation de la valeur du temps (Marche, 1973) et la répartition logit à étages (Ben-Akiva, 1973).

3.2.2.2 Côté pratique, on puise dans des boîtes à outils

Concernant l'affectation, le praticien exerce relativement peu de responsabilités : les logiciels courants sont assez rigides, les options y sont souvent restreintes au paramétrage de quelques variables. Toutefois la situation évolue dans le sens d'une plus grande liberté des utilisateurs.

Concernant le choix modal, la liberté est plus grande : l'utilisateur peut généralement segmenter la demande, et spécifier pour chaque segment et chaque mode une fonction d'utilité. Cependant, il doit se contenter (dans la majorité des cas) d'utiliser un modèle logit (multinomial ou à étages).

Au total le praticien reste très soumis aux choix méthodologiques du développeur du logiciel. Dans les lignes ainsi tracées, face à un problème donné, la spécification des mécanismes explicatifs devient une figure imposée : affectation type logit sur un réseau routier faiblement congestionné, différenciation des valeurs du temps pour un problème à péage (d'abord en France, mais de plus en plus couramment aussi dans les pays anglo-saxons), arbitrages modaux selon un modèle logit.

De nombreuses ressources théoriques, telles que l'élasticité de la demande ou l'explicitation des contraintes de capacité dans l'affectation statique, restent encore hors de portée du praticien. D'ailleurs celui-ci, s'il doit résumer les principes de modélisation, se contente souvent d'indiquer le nom du logiciel employé, au lieu de citer les mécanismes explicatifs qu'il a retenus parmi les possibilités offertes par le logiciel.

3.2.2.3 Côté recherche : des propositions à reprendre

Nous comptons **la fourniture et la mise en cohérence de principes explicatifs**, comme des contributions positives à l'audit. En effet, proposer de nouvelles modélisations, ou mettre en cohérence plusieurs mécanismes explicatifs, permet d'élargir la portée de l'audit, à défaut d'en définir la méthode. Parmi les mises en cohérence de principes explicatifs, citons l'intégration simultanée des quatre étapes du schéma classique (Safwat et Magnanti, 1988), le traitement simultané de l'induction, du choix modal et de la différenciation des demandeurs (Morellet et Julien, 1990).

Règles pour choisir entre plusieurs spécifications. Parmi les travaux consacrés au choix circonstancié d'une structure de modèle, citons des comparaisons entre structures alternatives de choix de transport (Williams, 1977), entre diverses spécifications de fonctions d'utilité pour le choix modal (Gaudry et Wills 1978, Bradley et al. 1992). On trouve encore des études systématiques des effets d'hypothèses particulières (Horowitz, 1981 ; Koppelman, 1976).

Ces travaux ont valeur plus de démonstration, d'exemple, que de prescription.

3.2.2.4 Bilan

Il existe donc des contributions qui fournissent ou comparent des mécanismes explicatifs. Les comparaisons, courantes en matière de choix modal, manquent quant à l'affectation. De façon générale, constatons qu'elles n'emportent guère la conviction des praticiens, certainement car soumises à de trop nombreux facteurs de relativisation (non seulement le site d'application, mais encore la qualité du codage, le modèle d'affectation sous-jacent s'il s'agit du choix modal, les observations disponibles...).

3.3 QUESTIONS D'ESTIMATION

Nous rangeons dans les questions d'estimation les aspects liés à l'erreur de mesure et à l'erreur de fluctuation d'échantillonnage. Cependant, nous n'abordons pas ici les problèmes de mesure (dispositifs d'enquête ou de recueil d'information)¹.

Après une position succincte du problème de l'estimation (§3.3.1), nous décrivons successivement l'état de la pratique (§3.3.2) et l'état de la recherche (§3.3.3).

3.3.1 Objets et principes méthodologiques de l'estimation

L'estimation vise les paramètres qui figurent dans la formulation du modèle. Dans un modèle d'affectation, on peut regrouper les paramètres en quatre groupes, selon qu'ils sont liés :

- aux choix discrets entre modes ou entre itinéraires (valeur du temps etc.),
- à la distribution spatiale (matrice origine-destination etc.),
- aux fonctions de demande (élasticité du volume origine-destination à la désutilité du service),
- aux fonctions de temps de parcours (fonctions d'offre).

Cette typologie *d'objets* correspond naturellement aux composants d'un modèle de choix d'itinéraire (le premier groupe concerne également le choix modal et le choix d'horaire) ; elle permet aussi de classer les contributions à l'estimation, qui souvent visent un seul groupe.

La plupart des méthodes d'estimation partagent un principe commun : fixer les valeurs de certains paramètres, de façon à minimiser un certain critère d'écart entre des observations et les résultats du modèle qui leur correspondent. Les méthodes se différencient ensuite par leur objet (les paramètres à fixer), par l'expression et la justification du critère d'écart (pas toujours formalisé, dans la pratique), par les observations utilisées.

¹au sujet des procédures d'enquête, une présentation et des références figurent dans Ortuzar et Willumsen (1990). Les ouvrages de Taylor et Young (1988) et Cohen (1990) tiennent le même rôle pour le recueil d'information sur le trafic routier

3.3.2 Etat de la pratique

3.3.2.1 La pratique sans outil spécialisé : constater plutôt qu'agir...

Contrairement aux contributions théoriques qui fixent sans contrainte les hypothèses nécessaires à la résolution du problème, les méthodes pratiques dépendent crucialement des conditions matérielles de l'application : observations peu nombreuses, fréquemment disparates, ou de qualité douteuse... ainsi que des outils opérationnels.

Souvent, en l'absence de fonctions spécialisées dans le logiciel opérationnel, la partie "ajustement" d'une application se borne au constat des écarts entre observation et reconstitution par le modèle, constat dressé à partir de comparaisons locales (au niveau d'un arc). Certains praticiens modifient les paramètres selon leur intuition, sans garantie quant à l'efficacité finale de leur intervention. Disposant d'observations de débits en un certain nombre de postes, ils illustrent leur "analyse"¹ en calculant une régression d'une variable dépendante "débit simulé" par une variable indépendante "débit observé".

Le plus souvent, les bureaux d'études produisent le diagramme comparant les débits simulés aux débits observés et ils affirment péremptoirement que l'ajustement est satisfaisant (2). Quelques-uns vont plus loin et mesurent la qualité de la régression par des indicateurs statistiques : le coefficient R² de régression (Florian et Nguyen 1976), une erreur quadratique moyenne (le RMSE de Wyley et al., 1981, éventuellement par classes de débit) ou une distance du khi-deux (Cf. Thomas 1991).

3.3.2.2 La pratique éclairée : tirer parti des observations disponibles, grâce à un outil spécialisé

Certains praticiens d'agir sur les paramètres du modèle, afin d'améliorer tant que possible l'ajustement sur les observations disponibles.

Parmi les observations, figurent non seulement des débits, mais encore des temps de parcours³. Intuitivement, cette base d'ajustement est préférable si l'on suppose que l'auteur du déplacement choisit l'itinéraire de coût minimal (donc se réfère à des temps plutôt qu'à des flots).

Pour ajuster les résultats du modèle aux observations, il faut se donner des leviers de commande... et une machine à commander. Les corrections manuelles de paramètres isolés atteignent vite leurs limites ; seule une automatisation, par l'emploi d'un logiciel spécialisé, permet d'aller plus loin. Un logiciel spécialisé offre des avantages déterminants : lisibilité pour un intervenant extérieur, puissance de calcul qui permet de tester de multiples combinaisons. Il en existe pour le choix modal⁴, pour l'estimation de matrices origine-destination⁵, parfois pour plusieurs objets d'estimation (Delons,

¹on ne peut parler d'ajustement, puisqu'il n'y a que constat

²ex. HCG (1993), Janson et al. (1986). Ces derniers auteurs jugent même "surprenante" (sic) la "bonne" reproduction des débits, alors que les temps de parcours par arcs sont eux très médiocrement simulés.

³mesurés par des campagnes de véhicules flottants, entre autres méthodes

⁴logiciels Alogit, CRT, Hiellow

⁵fonctionnalités incluses dans les logiciels TRIPS, SATURN, EMMÉ/2, POLYDROM, OEDIPE

1991¹).

De telles procédures doivent être appliquées à bon escient : il s'agit d'incorporer du bon sens aux données d'application du modèle, non de le contraindre brutalement pour lui faire passer avec succès les tests de validation.

Afin de prévenir d'éventuels excès en la matière, le TAM préconise de laisser de côté certaines mesures de débit et de temps lors de l'ajustement, afin de les utiliser exclusivement pour contrôler l'aptitude reproductive. De manière plus théorique, il est bon que la méthode d'ajustement procède d'un cadre probabiliste clairement défini, auquel s'attache une interprétation physique.

3.3.3 Du côté de la recherche

3.3.3.1 Chaque contribution vise un objet particulier

On a vu qu'en pratique, l'estimation des modèles de trafic dépend à la fois de contraintes institutionnelles (encadrement) et techniques (existence de méthodes économétriquement rigoureuses, disponibilité de ces méthodes dans les logiciels opérationnels). Du côté de la recherche, parmi les travaux relatifs à l'estimation, on trouve aussi bien des contributions ponctuelles (McFadden 1974, Daly 1987) que de larges synthèses (Daganzo, 1979 ; Amemiya, 1981 ; Ben-Akiva et Lerman, 1985) : ces dernières concernent les modèles de choix discret, donc actuellement les principaux modèles de choix modal.

Qu'elle soit ponctuelle ou qu'elle appartienne à une compilation, chaque contribution porte sur un objet donné, pour lequel elle définit (i) un cadre probabiliste, (ii) une méthode théorique d'estimation habituellement choisie entre les moindres carrés ou le maximum de vraisemblance, (iii) ainsi qu'une méthode pratique de calcul. Sa portée dépend à la fois de la richesse du cadre probabiliste (ampleur de l'objet, représentation des interactions entre les variables aléatoires) et de l'applicabilité pratique.

3.3.3.2 Quels cadres probabilistes pour l'estimation ?

En fonction de l'objet d'estimation, nous allons caractériser succinctement la substance des contributions de recherche (voir section 10.2 pour un approfondissement). Nous aborderons successivement les modèles de choix discret pour la formule de répartition entre itinéraires concurrents, les modèles de distribution spatiale pour la répartition du trafic entre les relations origine-destination, puis les fonctions de demande, enfin les fonctions de temps de parcours.

Modèles de choix discret. On représente le choix d'un itinéraire ou d'un mode de transport, comme celui d'une variante d'un certain bien, en l'occurrence l'acheminement entre une zone d'origine et une zone de destination. Chaque variante est choisie avec une fréquence relative qui dépend des caractéristiques des diverses variantes et des individus. Concrètement, on formule une fonction d'utilité pour chaque variante et chaque individu, et l'on suppose que chaque individu choisit la variante qui présente

¹Delons applique une procédure où les variables à estimer sont des niveaux de flots O-D (pour des blocs de la matrice totale) et des paramètres de loi débit-vitesse, le critère d'ajustement est un écart quadratique total qui agrège des écarts en débit et des écarts en temps, la minimisation procède selon une méthode complexe de Box-Jenkins.

pour lui l'utilité maximale. Le plus souvent on formule la fonction d'utilité comme la somme d'un terme déterministe et d'un terme aléatoire.

Le modèle de choix discret le plus répandu est le modèle logit qui correspond à des hypothèses particulières sur les termes aléatoires incorporés dans les fonctions d'utilité. Il existe de très nombreuses contributions à son sujet : exemples d'application, variantes de formulation, test de validité, discussion des hypothèses, mise au point de logiciels d'application. Très banal en matière de choix modal, le modèle logit sert relativement peu à l'affectation, faute de recherche empirique sur les itinéraires réellement suivis dans les déplacements.

Estimation des flots origine-destination. Un modèle de distribution spatiale sert à *produire* des flots origine-destination à partir d'informations fragmentaires : il faut donner une interprétation probabiliste pour estimer les paramètres du processus de production. De nombreuses contributions existent à ce sujet (voir des références dans Ortuzar et Willumsen 1990, ou Bierlaire 1991). On peut les différencier selon les données d'observation (entre enquête origine-destination et comptages), selon l'interprétation probabiliste (théorie de l'échantillonnage ou bayésienne), selon les hypothèses probabilistes sur les flots origine-destination et sur les comptages, selon le modèle d'affectation sous-jacent lorsqu'elles nécessitent des niveaux de service par origine-destination, selon la méthode d'estimation (maximum de vraisemblance ou moindres carrés), enfin selon la portée pratique (dimension des problèmes traitables). On manque non de propositions, mais d'un banc d'essai comparatif pour tester non seulement les combinaisons proposées mais toutes celles envisageables.

Fonctions de demande. Par fonction de demande nous désignons la dépendance d'un flot origine-destination envers les caractères de l'offre sur la relation O-D. Il y a intrinsèquement une dimension temporelle, distincte de la dimension spatiale inhérente à un modèle de distribution. On trouve peu de contributions relatives à l'estimation de fonctions de demande, sinon à l'échelle de l'ensemble d'un territoire ou d'un réseau donc en perdant l'identité des relations O-D : les statistiques de billettique favorisent les modes collectifs (Cf. références dans Goodwin 1992) ; mentionnons aussi des modèles de circulation routière qui simulent l'évolution temporelle d'un agrégat de trafic (souvent dans un cadre national, Cf. Madre 1989 pour la France).

Fonctions de temps de parcours. L'estimation des relations entre temps de parcours et débit fait l'objet de nombreuses contributions de l'ingénierie du trafic, principalement au sujet du mode routier (Cf. Taylor et Young 1988, Cohen 1990, pour des listes de références). On peut classer ces contributions selon qu'elles traitent d'un régime stationnaire ou dynamique et selon qu'elles considèrent les interactions entre les routes ou non.

3.3.3.3 Appréciation d'ensemble

Chacune des contributions de recherche se limite à un objet d'estimation ; très peu font plus qu'exposer une méthode d'estimation et quantifient l'erreur exogène à propager via le modèle de choix de mode ou d'itinéraire (Koppelman 1976 ; Gunn et Bates 1982 ; Bell 1983). Il manque une mise en relation théorique des différents objets par rapport à l'erreur d'estimation dans un modèle d'affectation ; à notre connaissance, les seules tentatives à ce sujet sont empiriques (les analyses de sensibilité des praticiens) ou partielles (procédure anglaise du TAM restreinte aux flots origine-destination) ou agrégées (contribution de Roger Marche à l'échelle d'un réseau).

Au §3.4.2 nous examinons les tentatives d'intégration théorique menées en Angleterre par Ashley (1981) et en France par Roger Marche (1988 à 1991).

3.4 AUTRES CONTRIBUTIONS

Les contributions à l'audit technique concernent essentiellement la spécification et l'estimation. Cependant on trouve aussi des études de l'erreur de calcul (§3.4.1) ainsi que, de façon connexe à l'estimation, des travaux sur la propagation de l'erreur d'estimation (§3.4.2).

3.4.1 Une inconnue : l'erreur de calcul

L'erreur de calcul survient dans les modèles d'équilibre, où les variables endogènes ne s'expriment pas directement en fonction des variables exogènes. Elle concerne donc l'affectation. Elle naît de l'arrêt (nécessairement) prématuré d'un algorithme itératif.

3.4.1.1 Côté pratique : une certaine désinvolture, pas d'interprétation ni d'étude de l'erreur de calcul

Les hommes d'étude de la planification des transports ont semble-t-il pris leur parti du défaut de précision qu'ils subodorent dans les calculs d'affectation. Certains, optimistes, veulent présumer que l'erreur de calcul compense des hypothèses trop fortes dans la modélisation : par exemple, à propos du principe de Wardrop, d'aucuns disent que mieux vaut ne pas pousser au bout l'optimisation des choix d'itinéraire des conducteurs, puisqu'il existe des conducteurs qui n'optimisent pas ! Mais cette position conduit à abdiquer le contrôle du modèle...

Aussi, en matière de condition de convergence, les utilisateurs des modèles opérationnels se bornent à suivre les prescriptions du concepteur de leur logiciel.

Parmi les concepteurs des logiciels, deux attitudes prévalent. D'un côté le clan des "pragmatiques", qui affirment d'emblée qu'aucun algorithme ne peut converger suffisamment vite vers un équilibre, aussi ils adaptent la procédure d'affectation afin qu'elle livre en peu d'itérations un résultat "plausible". C'est la position des utilisateurs de procédés heuristiques (chargements par tranches ou incrémental¹). Leur condition de convergence consiste simplement à atteindre un nombre fixé d'itérations.

De l'autre côté, le clan des fervents adeptes des méthodes mathématiques (surtout l'algorithme de Frank-Wolfe²). Ils suggèrent des critères de convergence basés sur des valeurs caractéristiques du programme mathématique que l'algorithme tente de résoudre.

Au total, ni l'une ni l'autre des deux attitudes ne débouche sur une interprétation statistique du niveau de convergence atteint in fine. Conséquence logique, nous n'avons pas trouvé de document discutant la précision des calculs.

¹voir les anciennes versions de DAVIS, ou des options de MIN UTP et TRANPLAN, ou POLYDROM

²voir d'autres options de MIN UTP et TRANPLAN, ou EMME/2

3.4.1.2 Des recherches fragmentaires

A ce jour, les études systématiques de l'erreur de calcul consistent toutes en des expérimentations numériques sur ordinateur (Horowitz, 1986 ; Rose et al. 1988). Elles comportent :

- une définition de critères de convergence,
- une comparaison de l'évolution de ces critères, en fonction des itérations de l'algorithme d'équilibrage.

Avant d'y revenir plus longuement au chapitre 9, notons dès maintenant que les expérimentations passées considèrent des algorithmes peu précis (sans identification des chemins), qu'elles négligent l'interprétation statistique (notamment quant au biais).

3.4.1.3 Bilan

Au total, de nouvelles études paraissent nécessaires, afin d'approfondir les travaux de recherche antérieurs et de prescrire des règles pour les applications pratiques. Les contributions de recherche se consacrent à la mise au point d'algorithmes, sans remettre en cause les formules mathématiques à résoudre, donc sans conscience d'un risque d'erreur formelle (Cf. chapitre 8).

3.4.2 Sur la propagation de l'erreur exogène

L'erreur exogène se répercute sur les simulations du modèle : le modèle propage l'erreur exogène. Concernant le résultat du modèle, c'est l'amplitude de l'erreur après propagation qui importe directement, non l'amplitude de l'erreur avant propagation.

3.4.2.1 La pratique : quelques analyses de sensibilité

Dans les bonnes études de trafic, on examine la sensibilité des résultats aux hypothèses. Par exemple, dans une étude d'avant-projet d'autoroute à péage, on teste la sensibilité du niveau de recette au niveau moyen des valeurs du temps.

De telles opérations sont très utiles. Il conviendrait de les compléter, en précisant les incertitudes exogènes : quand le résultat dépend de nombreux facteurs, il faut tenir compte des liens entre les facteurs, sous peine de surestimer l'erreur exogène propagée.

3.4.2.2 Des contributions ponctuelles de la recherche

Concernant les modèles de choix discrets, on trouve quelques travaux sur la propagation de l'erreur d'estimation. Koppelman (1976) étudie le biais et la marge d'incertitude dans l'application d'un modèle logit binomial résultant de la procédure de regroupement des individus. Appliquant les formules classiques de propagation des erreurs, il montre que l'erreur d'agrégation est "relativement" indépendante de l'erreur sur les paramètres et les variables en entrée. Horowitz (1981) investigate les effets potentiels de plusieurs hypothèses restrictives du modèle logit multinomial. Il examine tour-à-tour :

- l'hypothèse d'indépendance vis-à-vis des alternatives tierces (concluant que l'erreur reste faible tant que la situation en projet demeure proche de la situation de référence).

- le postulat de généralité des coefficients de la fonction d'utilité¹ : s'il est infirmé, de sérieuses difficultés peuvent surgir.
- l'omission de variables explicatives².
- enfin l'utilisation de mauvaises données de référence pour l'estimation des paramètres.

Horowitz illustre les effets potentiels sur des cas particuliers, sans indiquer d'ordres de grandeur à portée générale.

Concernant la distribution spatiale, Bell (1985) propage à travers un modèle gravitaire l'incertitude sur les marges d'émission et réception (entre autres facteurs).

3.4.2.3 Des travaux anglais sur l'affectation au sein du schéma à quatre étapes concluent à la nécessité d'investigations numériques

Ashley (1981) récapitule une recherche sur l'incertitude dans l'application du schéma à quatre étapes. Evoquant successivement la génération, la distribution, le choix modal et l'affectation, il examine et compare pour chaque étape l'erreur liée à l'estimation, puis l'erreur sur les inputs propagée mécaniquement par le modèle, enfin l'erreur "de prédiction" liée au décalage temporel entre la situation de l'estimation et l'horizon de la simulation.

Pour l'étape de distribution, il recommande de traiter les flots O-D de demande comme des variables aléatoires, dont il regrette de ne pouvoir préciser les covariances.

L'analyse de l'étape d'affectation est moins riche, essentiellement en raison du caractère statistique rudimentaire de la procédure envisagée (le modèle déterministe à temps variables et demande fixe et homogène). Ashley prescrit des simulations numériques (du Monte-Carlo) pour établir des marges d'incertitude "en sortie" à partir de marges d'incertitude "en entrée" sur les flots origine-destination, ainsi que sur les paramètres des fonctions de temps de parcours³ et sur le coefficient d'arbitrage entre temps et distance⁴.

3.4.2.4 L'approche intégratrice de Roger Marche

Roger Marche a développé (1988-1991) une méthodologie d'expertise des prévisions de recette pour une route à péage. Identifiant 4 postes principaux d'erreur, qu'il suppose indépendants les uns des autres :

- le niveau des flots de demande par origine-destination,

¹la non-généralité signifierait que le coefficient pourrait varier selon les individus ; Cf. la notion de dispersion des valeurs du temps

²Wonnacott et Wonnacott, 1991, offrent une présentation particulièrement pédagogique de ce sujet

³pour une fonction de forme $t_a(x_a) = t_a^0 \{1 + \alpha_a (x_a / N_a)^{\beta_a}\}$, il s'agit de considérer que les paramètres t_a^0 , N_a , α_a et β_a sont des variables aléatoires

⁴si l'on remplace la notion de distance par celle de coût monétaire, ce coefficient s'interprète comme l'inverse d'une valeur du temps moyenne

- la moyenne et la dispersion des valeurs du temps,
- les temps de parcours simulés,
- enfin des facteurs divers (taux d'induction...),

Marche additionne les variances sur la recette d'ensemble en provenance de chaque poste, afin d'obtenir une variance globale. Il justifie le niveau de la variance attachée à chaque poste par de multiples présomptions, faisant jouer en définitive un "sentiment d'expert".

Retenons le caractère intégrateur de cette approche, qui prend en compte des facteurs liés aux quatre objets d'estimation.

3.4.2.5 Bilan

Cette revue révèle des oppositions méthodologiques :

- entre praticiens soucieux surtout d'analyse de sensibilité, et chercheurs qui mettent l'analyse de sensibilité au service du calcul de l'erreur propagée.
- entre partisans des expériences numériques (Ashley), et tenants d'une approche économétrique analytique (Marche).
- entre contributions visant un objet particulier, et approches intégratrices.

Ces oppositions sont la conséquence de choix techniques particuliers, non de désaccords quant à l'interprétation statistique. Notre propre position consiste à lever les contraintes techniques pour préserver l'interprétation statistique.

3.5 BILAN PROVISoire

Le tableau 3A récapitule l'état de la pratique et de la recherche, par poste d'erreur dans la modélisation du choix modal et de l'affectation.

Tab. 3A. Synthèse de l'état de la pratique et de la recherche.

Poste d'erreur	Pratique	Recherche
Codage	Peu formalisé Validation préventive d'efficacité variable	Presque muette
Composition sémantique	Choix intuitif, fonction des particularités de l'application	Contributions ponctuelles
Estimation	Oui si fonctions prévues dans le logiciel	Nombreuses contributions ponctuelles
Résolution	Pas de test	Quelques études empiriques
Propagation de l'erreur	Au mieux, tests de sensibilité, propagation de l'incertitude liée aux flots O-D	Quelques contributions ponctuelles Démarches intégratrices
Formelle	- pas de conscience -	Peu de remise en cause

Par rapport à l'économétrie classique, nous constatons :

- l'importance particulière du codage.
- en ce qui concerne l'affectation, la variété des objets d'estimation, partant la difficulté d'élaborer un cadre probabiliste qui les engloberait tous.
- la nécessité d'étudier d'autres postes d'erreur que ceux prévus classiquement, notamment l'erreur de calcul.
- le peu d'attention porté à la propagation de l'erreur, qu'il est pourtant essentiel d'étudier afin d'apprécier l'erreur liée au résultat du modèle¹.

Ces constats montrent la nécessité :

- d'une synthèse et d'une appréciation des concepts déjà élaborés,
- de contributions ponctuelles, pour chaque poste d'erreur,
- d'une trame d'ensemble par rapport à laquelle chaque élément prendrait place et serait mis en relief.

Vis-à-vis de cette liste de tâches, l'objectif de la thèse est d'abord d'élaborer une trame d'ensemble. Puis, pour l'affectation statique du trafic routier, d'habiller la trame en reprenant des contributions ponctuelles antérieures ou en apportant quelques éléments nouveaux, et de faire une synthèse des concepts et outils pour chaque poste d'erreur.

¹à l'intérieur du cadre probabiliste défini initialement

Deuxième partie

THÉORIE ET MÉTHODE

Chapitre 4

Analyse théorique de l'erreur

Dans un modèle, l'erreur découle de plusieurs sources : conception, formulation, calcul, estimation, propagation (section 4.1). Cette décomposition permet de considérer chaque source et de lui appliquer des méthodes spécifiques pour la mesurer et la traiter (section 4.2).

La portée des mesures et traitements dépend de la source d'erreur : il est possible d'assécher l'erreur formelle et l'erreur de calcul, d'endiguer l'erreur de conception et de canaliser l'erreur économétrique.

Nous indiquons comment le mécanisme du modèle propage l'erreur exogène (section 4.3) et nous examinons l'interaction entre erreur conceptuelle et erreur économétrique, en étudiant l'effet d'une sophistication du mécanisme explicatif sur l'interprétation et l'amplitude de l'erreur (section 4.4).

4.1 PLUSIEURS SOURCES ALIMENTENT L'ERREUR

Plusieurs sources alimentent l'erreur d'un modèle. Nous les recensons de façon générale (§4.1.1), puis, à titre d'exemple, de façon particulière au modèle d'affectation bicritère prix-temps (§4.1.2), avant de définir formellement la manière de les cumuler (§4.1.3).

4.1.1 Recensons les sources d'erreur

Remarquons d'abord qu'un recensement précis des sources d'erreur supposerait une connaissance absolue de la réalité, connaissance "surnaturelle" qui pour nous n'a valeur que de référence abstraite, utile à titre didactique. Cette réserve faite, nous recensons cinq classes de sources potentielles d'erreur dans un modèle :

(i) **l'erreur de conception** : dans la schématisation du système réel, dans la définition et le choix des variables explicatives (exogènes) et expliquées (endogènes), dans la mise en relation de ces éléments donc dans l'identification et la formulation des relations "explicatives" qui constituent le *mécanisme*, la *composition conceptuelle* (ou *sémantique*) du modèle.

(ii) dans la formulation qui consiste à synthétiser les mécanismes explicatifs par une formule caractéristique, justiciable d'un problème type de mathématiques. **L'erreur de formulation** procède d'un manque de conformité entre la formule caractéristique et la composition conceptuelle, ou d'un manque de cohérence logique, ou d'une insuffisance de caractérisation. Cette classe comprend les hypothèses mathématiques minimales sur ces fonctions (par exemple, continuité ou croissance).

(iii) **l'erreur de résolution** : dans la résolution de la formule caractéristique, si elle procède par une méthode heuristique ou peu précise. Dans le cas d'une résolution par simulation numérique, la cause peut tenir à des contraintes informatiques : terminaison en un temps fini (troncature d'un algorithme itératif), opérations de précision finie (arrondis).

(iv) **erreur d'estimation** : dans la spécification fonctionnelle (explicitation des fonctions mathématiques) et numérique des paramètres du modèle, autrement dit dans la caractérisation numérique complète des relations de dépendance des variables endogènes entre elles et envers les variables exogènes.

(v) **erreur exogène** : dans l'application à des variables exogènes entachées de biais ou d'incertitude (erreur sur les "entrées", les "inputs").

Les cinq sources d'erreur recensées séparent le modèle de la réalité : elles forment **l'erreur nette** du modèle. Cependant la réalité n'est connue qu'au travers d'observations qui comportent leur propre erreur : cette **erreur de référence** sur le résultat du modèle s'ajoute à l'erreur nette pour constituer **l'erreur brute** du modèle. Pour fixer les idées, considérons un modèle simple de formule :

$$\mathbf{Y} = \mathbf{F}(\mathbf{X})$$

où **Y** est le vecteur des variables endogènes (la réponse, l'effet),
X est le vecteur des variables exogènes (l'argument, la cause),
F est le vecteur des relations de dépendance (le mécanisme).

Le tableau 4A montre le lieu d'occurrence de chaque source d'erreur nette dans le modèle.

Tab. 4A. Classes d'erreur nette dans un modèle.

Niveau	Lieu d'occurrence
Conception	choix de X et Y , idées sur les éléments de F
Formulation	formulation de F (relations de dépendance entre variables explicatives et expliquées, hypothèses mathématiques minimales, synthèse des composants élémentaires de F)
Résolution	erreur dans la résolution de F
Estimation	expression mathématique et numérique de F (erreur d'observation sur les données d'estimation...)
Exogène	erreur sur des valeurs particulières de X

4.1.2 L'exemple du modèle d'affectation bicritère

Le tableau 4B recense les hypothèses sources potentielles d'erreur dans le modèle bicritère prix-temps à demande totale élastique.

Tab. 4B. Hypothèses dans le modèle bicritère.

Classe	Lieu d'occurrence
Conceptuel	<ul style="list-style-type: none"> - divisibilité des demandeurs en flots infinitésimaux - définition des chemins k - comportement optimisant des demandeurs, en fonction uniquement de deux critères de temps et de prix - parfaite information des demandeurs sur les chemins et les temps de parcours, exacte perception subjective des temps de parcours T_a objectifs - dépendance quantitative du niveau de demande envers le temps généralisé - différenciation des demandeurs selon la relation origine-destination et selon la valeur du temps - temps de parcours d'un tronçon croissant en fonction du flot
Formel	<ul style="list-style-type: none"> - $q_{rs} = D_{rs}(\bar{S}_{rs})$ lien agrégé entre le volume total de demande q_{rs} et un temps généralisé moyen \bar{S}_{rs} ; décroissance de la fonction D_{rs} - distribution des valeurs du temps selon une répartition continue H_{rs} - spécification du temps de parcours local T_a en fonction du flot local x_a uniquement (et pas des autres flux, en sens opposé ou croisés aux carrefours), par $T_a = t_a(x_a)$. Croissance de la fonction t_a
Résolution	<ul style="list-style-type: none"> - choix d'un algorithme - nombre d'itérations, exactitude et précision des calculs
Estimation	<ul style="list-style-type: none"> - dans une application, fixer la forme et les paramètres des fonctions t_a, D_{rs}, H_{rs}
Propagation	<ul style="list-style-type: none"> - erreur sur les variables exogènes

Nous nous référons au modèle bicritère prix-temps pour plusieurs raisons. En tant que

modèle d'affectation particulier, tout d'abord. Ensuite parce qu'il sophistique le mécanisme explicatif du modèle de Beckmann, tout en demeurant justiciable d'un traitement analytique aisé ⁽¹⁾. Enfin pour les enjeux financiers liés à son application, dans le but de donner une utilité pratique à la recherche entreprise.

4.1.3 Les erreurs se cumulent !

Les erreurs des cinq types se cumulent. En modifiant la composition conceptuelle (i.e. en changeant la représentation de référence), les importances respectives des types varient : une réduction de l'erreur conceptuelle peut nécessiter un ajout de paramètres donc potentiellement une variation de l'erreur d'estimation, un ajout de variables exogènes, donc potentiellement une variation de l'erreur exogène, un changement des formules et du dispositif de résolution, donc potentiellement une variation des erreurs formelle et algorithmique.

Formalisons ces considérations, en marquant, pour le modèle d'équation $\mathbf{Y} = \mathbf{F}(\mathbf{X})$, l'erreur de chaque type par un symbole propre. Il s'agit de préciser ⁽²⁾ l'expression $\mathbf{Y} = \mathbf{F}(\mathbf{X}) + \varepsilon$ où ε est l'erreur totale du modèle, et \mathbf{Y} et \mathbf{X} des variables observées.

Vient d'abord le risque d'une connaissance inexacte des variables : $\mathbf{Y} = \mathbf{Y}^* + \varepsilon_{\text{réf}}$, où \mathbf{Y}^* est la "vraie" variable dépendante et $\varepsilon_{\text{réf}}$ une erreur de référence. De même $\mathbf{X} = \mathbf{X}^* + \varepsilon_{\text{exo}}$, où \mathbf{X}^* est la "vraie" variable exogène et ε_{exo} une erreur exogène avant propagation par le modèle. Faisons-en provisoirement abstraction.

Soit $\varepsilon_{\text{sém}}$ l'erreur sémantique du modèle "vrai" entre \mathbf{Y}^* et \mathbf{X}^* :

$$\mathbf{Y}^* = \mathbf{F}^*(\mathbf{X}^*) + \varepsilon_{\text{sém}}.$$

où \mathbf{F}^* est le "vrai" mécanisme de dépendance.

S'ajoute l'erreur de formulation ε_{for} qui provient des simplifications et approximations sur \mathbf{F}^* : $\mathbf{F}^{(\text{for})} = \mathbf{F}^* + \varepsilon_{\text{for}}$, donc

$$\mathbf{Y}^* = \mathbf{F}^{(\text{for})}(\mathbf{X}^*) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}}.$$

L'erreur de résolution de $\mathbf{F}^{(\text{for})}$, se solde par $\mathbf{F}^{(\text{sol})} = \mathbf{F}^{(\text{for})} + \varepsilon_{\text{sol}}$, donc

$$\mathbf{Y}^* = \mathbf{F}^{(\text{sol})}(\mathbf{X}^*) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{sol}}.$$

¹Le traitement analytique proposé pour ce modèle présente par surcroît d'intéressants avantages formels sur le modèle de Beckmann, à savoir l'unicité des flots par classe monétaire de chemins à l'équilibre et donc la possibilité d'exprimer ces flots sous la forme de fonctions dérivables des paramètres du modèle. Cela simplifie d'autant l'analyse de sensibilité. Pour une présentation détaillée du modèle de Beckmann, Cf. Leurent 1995d.

²Il s'agit d'illustrer notre propos qualitatif par une formulation. Nous prenons ainsi une petite revanche sur l'incertitude dans un modèle, en proposant un modèle formel de l'incertitude ! Plus sérieusement, ce modèle formel vise seulement le didactisme, non la validité formelle absolue : en particulier nous marquons par un signe "+" le cumul des erreurs, alors que dans certaines circonstances le mode de cumul est plus vraisemblablement multiplicatif

On marque ensuite l'erreur liée à l'estimation, soit $F = F^{(sol)} + \varepsilon_{esti}$:

$$Y^* = F(X^*) + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol} - \varepsilon_{esti}.$$

Enfin s'ajoute l'erreur exogène ε_{exo} sur les inputs X^* :

$$Y^* = F(X - \varepsilon_{exo}) + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol} - \varepsilon_{esti}$$

que l'on approche par

$$Y^* \approx F(X) - \nabla_X F \cdot \varepsilon_{exo} + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol} - \varepsilon_{esti}$$

où l'on identifie $\varepsilon_{nette} \approx -\nabla_X F \cdot \varepsilon_{exo} + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol} - \varepsilon_{esti}$ qui est l'erreur nette totale du modèle. Le dispositif d'observation de la référence à laquelle on compare le résultat, produit l'erreur de référence $\varepsilon_{réf} = Y - Y^*$ qui s'ajoute à l'erreur nette. Donc l'erreur brute totale du modèle est

$$\varepsilon_{brute} \approx -\nabla_X F \cdot \varepsilon_{exo} + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol} - \varepsilon_{esti} + \varepsilon_{réf}.$$

Le tableau 4C récapitule l'empilement des erreurs.

Tab. 4C. Cumul des erreurs.

Types d'erreur cumulés	Formule du modèle
Conception	$Y^* = F^*(X^*) + \varepsilon_{sém}$
Conception & Formulation	$Y^* = F^{(for)}(X^*) + \varepsilon_{sém} - \varepsilon_{for}$
Id. & Résolution	$Y^* = F^{(sol)}(X^*) + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol}$
Id. & Estimation	$Y^* = F(X^*) + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol} - \varepsilon_{esti}$
Id. & Exogène	$Y^* = F(X - \varepsilon_{exo}) + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol} - \varepsilon_{esti}$
Erreur nette + Erreur de référence	$Y = F(X - \varepsilon_{exo}) + \varepsilon_{sém} - \varepsilon_{for} - \varepsilon_{sol} - \varepsilon_{esti} + \varepsilon_{réf}$

4.1.4 Discussion

Nous avons décomposé l'erreur d'un modèle en plusieurs éléments qui en constituent la composition conceptuelle. Cette explicitation est une revanche sur l'incertitude, puisque l'incertitude du modèle devient elle-même l'objet d'un modèle !

La formulation donnée au §4.1.3 remplit essentiellement une fonction analytique et didactique. Les sources d'erreur se distinguent par leur origine et leur nature. Précisons les démarcations afin de prévenir toute confusion.

4.1.4.1 L'erreur de référence se recycle en erreur d'estimation

Lorsque la fonction causale F comporte des paramètres (coefficients scalaires ou formes de fonctions), l'erreur d'estimation ε_{esti} sépare la valeur fixée de leur "vraie" valeur. Comment la valeur fixée est-elle déterminée ? Le plus souvent à partir d'observations préalables du phénomène, par confrontation entre des variables observées $Y_{pré}$ et $X_{pré}$

inexactes. Lors de l'observation préalable, l'erreur de référence $\varepsilon_{\text{pré,réf}}$ et l'erreur exogène $\varepsilon_{\text{pré,exo}}$ se propagent sur le résultat (les paramètres de \mathbf{F}) donc elles contribuent à l'erreur d'estimation $\varepsilon_{\text{esti}}$ dans notre modèle principal.

4.1.4.2 L'erreur d'estimation comme une erreur exogène a priori

A priori chaque paramètre dans \mathbf{F} est une variable exogène ! La démarcation entre variables exogènes \mathbf{X} et fonction causale \mathbf{F} résulte d'un arbitrage commode : elle sépare les variables exogènes au sens strict \mathbf{X} qui sont assez ou très susceptibles de varier d'une simulation à l'autre, des paramètres jugés a priori peu ou pas susceptibles de varier.

S'il s'agit uniquement de détecter et de qualifier l'erreur, la démarcation est artificielle et nous l'ignorons, unissant l'erreur d'estimation et l'erreur exogène dans la catégorie "erreur économétrique". Mais s'il s'agit de quantifier l'erreur économétrique, alors nous distinguons entre les différents inputs, donc entre paramètres et variables exogènes au sens strict.

4.1.4.3 Sur l'erreur de transfert

Le "transfert" d'un modèle de simulation désigne la simulation d'un problème concret à la fois analogue et différent du problème qui a inspiré le modèle. Cela se décline tant pour l'alimentation économétrique, que pour la composition conceptuelle, la formule caractéristique et le dispositif de résolution.

En matière de mathématiques, il n'y a pas de transfert au sens propre : on applique toujours une théorie à un problème ; même si le problème pousse à reformuler la théorie ou à l'amender, la théorie repose uniquement sur des axiomes abstraits ! Cela justifie de transférer immédiatement une formule caractéristique ou un algorithme.

En contrepoint tout transfert d'alimentation économétrique mérite une attention particulière. Des archétypes sont : (i) la simulation d'une époque autre que celle à laquelle se rapporte l'alimentation économétrique (transfert dans le temps), (ii) la simulation d'un système autre que celui auquel elle se rapporte (transfert dans l'espace).

En sciences physiques, on postule qu'un transfert temporel ou spatial n'altère en rien la théorie ni les résultats expérimentaux : c'est le principe de Galilée. Cependant on sait que l'univers se modifie au cours du temps ; que d'un lieu à l'autre du globe terrestre l'intensité de la pesanteur varie¹. Le postulat physique d'absence d'effet d'un transfert est donc avant tout une excellente approximation.

En sciences sociales, on sait que les objets étudiés varient au cours du temps. La notion d'identité apparaît finalement bien fragile ! Il convient donc d'explicitier tout transfert, d'en prévenir autant que possible les effets.

Nous classons l'incertitude relative à un transfert à l'intérieur de l'incertitude exogène a priori. A l'analyste d'apprécier le plus objectivement possible (mais finalement de façon toujours subjective) les hypothèses et les conséquences du transfert, ainsi que l'incertitude qui en résulte.

¹d'où la nécessité d'adapter le paramètre g de gravité en fonction de la latitude

4.1.4.4 Sur le comportement stochastique du système étudié

Est-il possible de connaître avec une absolue précision un phénomène réel ? En physique quantique le principe d'incertitude de Heisenberg limite la précision de la connaissance simultanée, pour un observateur, de la position et la vitesse d'une particule. Dans ces conditions toute connaissance physique apparaît nécessairement probabiliste : i.e. savoir que le système observé peut occuper tel état avec telle probabilité, au lieu de savoir qu'il se trouve dans un état particulier avec une certitude absolue.

Nous classons cette "incertitude intrinsèque" du système modélisé dans l'erreur sémantique du modèle. Elle peut influencer sur l'erreur de référence dont elle est indissociable, donc indirectement sur l'erreur d'estimation.

4.1.4.5 Précisions de vocabulaire

Je ne distingue guère entre erreur et incertitude : la principale nuance entre les deux termes tient au caractère "a posteriori" de l'erreur, par rapport à un constat, alors que l'incertitude conserve un caractère "a priori". L'erreur est plus tangible, concrète, tandis que l'incertitude est plus virtuelle.

La distinction statistique entre biais et incertitude tient à une acception restrictive de l'incertitude, lorsqu'on la suppose nulle en moyenne (i.e. absence de biais). Nous ne suivons cette distinction que pour analyser l'erreur économétrique.

Enfin en théorie du signal on appelle "bruit", ou "bruit aléatoire", toute distorsion du signal que l'on sait à la rigueur quantifier, mais dont on ne connaît pas les causes de façon précise.

4.2 POUR MESURER ET TRAITER L'ERREUR, DÉCLINER L'AUDIT

En cohérence avec la décomposition de l'erreur d'un modèle, nous déclinons l'audit selon chaque source d'erreur :

- audit conceptuel, ou sémantique, pour l'erreur de conception.
- audit formel, pour l'erreur de formulation.
- audit algorithmique, pour l'erreur de résolution.
- audit économétrique, pour l'erreur d'estimation et l'erreur exogène propagée.

Chacun de ces audits mobilise des méthodes provenant de plusieurs disciplines :

- pour l'audit conceptuel, l'analyse systémique, le sens physique.
- pour l'audit formel, la logique mathématique et la recherche opérationnelle.
- pour l'audit algorithmique, la physique, l'informatique, l'algorithmique et l'expérimentation (numérique ou autre).
- pour l'audit économétrique, l'économétrie, le recueil de données et la physique (pour certains dispositifs de mesure), la recherche opérationnelle (pour l'analyse de sensibilité et donc la propagation de l'erreur exogène).

Constatons la variété des disciplines impliquées : il en résulte une complexité

d'abondance, qui s'ajoute aux complexités de sens inhérentes à certaines des disciplines.

Nous disons source d'erreur plutôt que poste d'erreur, afin de développer une analogie : assimiler l'erreur à un fleuve, alimenté par autant d'affluents que de sources d'erreur, et dont l'embouchure correspond au résultat du modèle. Alors réduire l'erreur correspond à assécher le fleuve, tenir compte de l'erreur à endiguer le fleuve, mesurer l'erreur à explorer le fleuve.

Aussi examinons comment prendre la mesure de l'erreur (§4.2.1), avant d'étudier comment la réduire (§4.2.2).

4.2.1 Comment mesurer l'erreur ?

4.2.1.1 Erreur sémantique

On mesure l'erreur de conception soit qualitativement, soit indirectement.

Qualitativement : le jugement a priori, l'analyse systémique servent pour comparer le mécanisme explicatif à la représentation théorique. On apprécie l'écart de façon qualitative, en dressant une liste d'approximations, d'hypothèses simplificatrices.

Indirectement : lorsque des observations expérimentales, donc des réalisations de l'erreur totale ε , laissent déduire la valeur de l'erreur conceptuelle par soustraction des autres sources d'erreur, appréciées elles plus directement.

4.2.1.2 Erreur de formulation

On mesure l'erreur de formulation en évaluant, point par point, la validité logique des chaînons formels qui mènent des hypothèses sur les mécanismes explicatifs à la formule caractéristique de la solution du modèle.

Le résultat de cette mesure est une liste, éventuellement vide, de points d'achoppement ou d'incohérence. Une fois ces points corrigés, il reste à étudier l'existence et l'unicité de la solution à la formule caractéristique.

4.2.1.3 Erreur algorithmique, dans le cas d'une résolution par simulation numérique

Une première action s'apparente à l'audit formel : elle consiste à valider théoriquement la méthode de calcul. Pour cela, il faut examiner s'il y a convergence, si oui vers quel état final, si des applications répétées livrent le même état final, si l'état final correspond à la solution théorique du modèle. Autant de questions au sujet desquelles une garantie théorique est reconfortante... mais ni nécessaire (puisque'une méthode heuristique peut convenir) ni suffisante (il faut encore mesurer numériquement que l'on approche suffisamment près de la solution théorique).

Une seconde action¹ est d'ordre numérique : il s'agit de mesurer quantitativement la distance entre le résultat des calculs et la solution théorique. En pratique, on évalue plutôt un majorant de l'écart que l'écart lui-même. Signalons que la rigueur du critère de convergence importe plus que celle de la méthode de calcul : un test de convergence

¹en plus de la détection et de la correction d'erreurs triviales liées au dispositif de calcul, Cf. section 5.3

théoriquement rigoureux, s'il est numériquement satisfait, démontre la convergence effective d'un algorithme même heuristique.

On attend in fine une mesure quantitative de l'erreur de calcul, sous la forme d'un critère d'écart entre état calculé et solution théorique.

4.2.1.4 Erreur d'estimation et erreur exogène propagée

Rappelons que l'erreur d'estimation recouvre l'erreur d'observation et l'erreur de fluctuation d'échantillonnage. Dans les deux cas, il n'y a pas mesure à proprement parler, mais plutôt estimation d'amplitude, délimitation d'une possible variation.

Cette délimitation permet de prendre des garanties quant à la vraie valeur des paramètres estimés. On entoure prudemment la valeur estimée d'une bande d'incertitude, appelée aussi intervalle "de confiance", auquel la vraie valeur appartient avec un niveau de garantie qui est fonction de la largeur de la bande.

Mesurer l'erreur d'estimation revient donc à prendre en compte explicitement une incertitude, qu'elle tienne à un dispositif d'observation ou à une fluctuation d'échantillonnage. L'économétrie fournit des méthodes en ce sens : elle prescrit notamment d'étudier la variance des estimateurs des paramètres. Pour un modèle donné, il importe de choisir avec soin la méthode économétrique d'estimation, afin de ne pas sous-estimer la bande d'incertitude (donc de bien placer le paramètre), afin de ne pas non plus la surestimer (un modèle trop imprécis est inutile).

La mesure de l'erreur exogène propagée procède en deux étapes. Primo estimer l'erreur exogène a priori (avant toute propagation par le modèle) : en plus de l'erreur d'estimation déjà mentionnée, cela recouvre l'erreur de transfert. Transférer le modèle dans l'espace ou dans le temps nécessite des hypothèses particulières (par exemple sur la croissance du trafic total entre deux horizons temporels). A ces hypothèses est attachée une incertitude à apprécier par délimitation, comme pour l'estimation, mais avec une interprétation différente : liée soit à un modèle extérieur, soit à des croyances subjectives que l'on quantifie intuitivement (sans référence à des observations).

Secundo propager l'erreur exogène a priori jusqu'au résultat du modèle. Il s'agit là d'une procédure déterministe, pour laquelle on dispose de formules mathématiques si l'audit formel a conclu favorablement.

4.2.2 Comment traiter l'erreur ?

Est-il possible de réduire à rien (ou presque) l'erreur d'un modèle ? Pour certaines sources, la formulation et le calcul, oui, un assèchement paraît envisageable. Mais pour d'autres, la conception notamment, un assèchement ne semble pas possible ; le traitement consiste alors à prendre la mesure de l'erreur pour en tenir compte, autrement dit à endiguer le fleuve pour prévenir d'éventuels débordements.

4.2.2.1 Audit conceptuel : explorer l'erreur de conception

Dans la décomposition formelle de l'erreur nette totale, l'erreur de conception est la source la plus difficile à mesurer quantitativement. L'exploration qualitative des écarts entre le mécanisme et la représentation théorique laisse juger a priori son rôle éventuel, mais ne livre pas de repères quantitatifs.

Une mesure indirecte, par soustraction des autres sources d'erreur à une erreur totale

estimée empiriquement, nécessite de contrôler a posteriori de multiples applications : tâche lourde, mais probablement hautement instructive ! Elle permettrait de distinguer les hypothèses cruciales des hypothèses accessoires, de même qu'elle livrerait un montant forfaitaire d'erreur à prendre en compte dans une prévision.

4.2.2.2 Audit formel : tenter d'assécher l'erreur de formulation

L'exploration, la mesure, de l'erreur de formulation, produit une liste de points d'achoppement ou d'insuffisance logique. Par construction mathématique, chaque point particulier peut recevoir une solution, éventuellement au prix d'hypothèses supplémentaires qui risquent de peser sur l'erreur de conception en simplifiant certains des mécanismes explicatifs.

Faut-il assécher l'erreur de formulation, quitte à augmenter l'erreur de conception ? Thèse : oui, car de toute façon comment expliquer, justifier, le résultat d'un modèle mal formulé, incohérent ? Antithèse : un modèle mieux conçu, même mal formulé, pourrait produire un résultat plus proche de la réalité. Synthèse : dans l'optique d'une certification, oui, car il faut connaître précisément les limites du modèle afin d'en garantir le résultat.

4.2.2.3 Audit algorithmique : assécher ou endiguer l'erreur de calcul

La phase théorique de l'audit algorithmique, garantir la rigueur du dispositif de résolution, relève tout comme l'audit formel des mathématiques (et d'autres disciplines si la résolution ne procède pas par simulation numérique). Mais nous avons dit qu'elle n'est pas indispensable : elle apporte une caution logique, mais ne se substitue pas à une nécessaire mesure quantitative d'un critère de convergence.

Le critère de convergence donne le moyen d'endiguer l'erreur de calcul, même de l'assécher quand les coûts de calcul sont modérés. Le critère doit être rigoureux : c'est là qu'apparaît toute l'utilité d'un algorithme théoriquement justifié, car il facilite la définition et l'évaluation d'un tel critère, qui sont moins évidentes avec un algorithme heuristique.

Si l'on se contente d'endiguer l'erreur de calcul, en la ramenant en deçà d'un certain seuil, il reste à l'interpréter en termes statistiques, afin d'en tenir compte au niveau du résultat du modèle.

4.2.2.4 Audit économétrique : endiguer l'erreur d'estimation et canaliser l'erreur exogène propagée

Nous avons vu comment prendre la mesure de l'erreur d'estimation. Pour la réduire, un levier d'action tient à la méthode d'estimation choisie, car, en fonction des hypothèses probabilistes, certains estimateurs sont plus efficaces que d'autres. L'exercice de ce choix suppose une connaissance des principales méthodes existantes ; des résultats généraux d'économétrie (par exemple l'inégalité de Cramer-Rao) indiquent que, pour un jeu donné d'observations, l'efficacité des estimateurs est irréductiblement bornée. Il y a donc une limite à l'assèchement ; d'où la nécessité d'endiguer l'erreur d'estimation qui subsiste.

Pour ce faire, il faut la prendre en compte au niveau du résultat du modèle, en la faisant contribuer à l'erreur propagée. Autrement dit, il faut considérer les paramètres du

modèle comme des variables exogènes auxquelles s'attache une incertitude connue¹.

Pour l'erreur exogène propagée, le traitement procède par une réduction de l'erreur exogène avant propagation, ou par une prise en compte au niveau du résultat du modèle, ce qui est la définition même de la propagation.

Opérer rigoureusement la propagation revient à canaliser l'erreur exogène, à l'empêcher de se répandre hors du lit du fleuve et d'échapper au contrôle du modélisateur.

Une question importante tient à la composition exacte de l'erreur exogène : celle-ci comprend non seulement l'erreur sur les variables exogènes, mais encore l'erreur d'estimation des paramètres du modèle.

4.2.3 Bilan : explorer, assécher, endiguer, canaliser

Au terme de cette analyse abstraite, concluons qu'en théorie² toute l'erreur peut être détectée donc explorée, qualifiée. En pratique, la principale question tient à la quantification de l'erreur de conception, possible seulement de façon indirecte. L'erreur de formulation et l'erreur algorithmique peuvent être réduites, et même asséchées, avec éventuellement des contreparties en termes d'erreur sémantique ou de coût de la résolution. L'erreur algorithmique et l'erreur d'estimation peuvent être endiguées, par une réduction dans la mesure du possible et une prise en compte au niveau du résultat du modèle. Enfin il convient de canaliser l'erreur exogène propagée.

Retenons la possibilité d'arbitrer entre plusieurs sources d'erreur : notamment entre erreur de conception et erreur de formulation. Affirmer le caractère opératoire du modèle peut nécessiter d'en simplifier le mécanisme explicatif³.

4.3 PROPAGER L'ERREUR VIA LE MODÈLE

Pour quantifier l'erreur au niveau du résultat du modèle, il faut propager l'erreur exogène, qu'elle provienne des paramètres estimés ou des données exogènes d'application. Après quelques définitions (§4.3.1), nous indiquons les deux principales formules de propagation de l'erreur (§4.3.2), que nous appliquons à titre d'exemple à l'affectation prix-temps (§4.3.3 et §4.3.4).

4.3.1 Définition du biais et de la marge d'incertitude

Un modèle déterministe prédit une unique valeur, forcément moyenne, pour chacune des variables endogènes. Un modèle stochastique prédit une distribution de probabilité pour chacune des variables endogènes : pour appréhender une telle distribution, deux valeurs caractéristiques sont utiles, la moyenne (moment d'ordre 1, mesure de niveau) et la variance (liée au moment d'ordre 2, mesure de dispersion).

Par erreur de biais, on entend une erreur sur la moyenne. La définition statistique du

¹ou plutôt estimée, à l'aide notamment des variances des estimateurs

²ce qui renvoie à une cohérence de type axiomatique, interne à un cadre d'hypothèses ; non pas à une validité dans l'absolu

³Cf. Paul Valéry : "tout ce qui est simple est faux, tout ce qui est complexe est inutile"

biais est l'écart entre moyenne prédite et moyenne vraie.

La notion de marge d'incertitude renvoie à une mesure de la variance : on déduit de cette mesure un intervalle de confiance pour la valeur moyenne "réelle", intervalle centré sur la valeur moyenne prédite.

Une marge d'incertitude n'a de sens que dans le cadre d'un modèle stochastique (probabiliste). Pour attribuer une marge d'incertitude au résultat d'un modèle déterministe, il faut "plonger" ce modèle dans un cadre stochastique. Pour ce faire, une approche naturelle consiste à donner une distribution statistique aux principaux paramètres.

Rappelons ici qu'il ne faut pas confondre la moyenne et l'écart-type (racine de la variance) d'une distribution de probabilité. Une confusion pourrait fournir des arguments trop simplistes contre l'utilité de sophistication le modèle, dans un cas où la prédiction moyenne du modèle sophistiqué tomberait dans l'intervalle de confiance du modèle simple. Avant de conclure, il convient encore de s'assurer que l'écart entre les moyennes est faible devant l'écart-type, et mieux vaut alors calculer une probabilité critique que de se contenter d'un intervalle de confiance. De plus, la conclusion ne vaut que pour le cas particulier envisagé, elle affirme une inutilité pratique ponctuelle mais pas une inutilité conceptuelle générale.

4.3.2 Comment le modèle propage le biais et l'incertitude

Rappelons maintenant les deux formules classiques¹ de propagation mécanique des erreurs via le modèle, formules relatives l'une au biais et l'autre à la marge d'incertitude. Ces formules découlent du calcul différentiel : nous les donnons pour le cas simple d'une dépendance explicite directe (Cf. §4.1.1).

Soit donc le modèle $\mathbf{Y} = \mathbf{F}(\mathbf{X})$, où l'on note \mathbf{Y} le vecteur des sorties du modèle (variables endogènes, l'effet), \mathbf{F} la fonction vectorielle décrivant la mécanique du modèle, et \mathbf{X} le vecteur des entrées du modèle (variables exogènes, soit contrôlées auquel cas on parle de commandes, soit imposées par un environnement auquel cas on parle de perturbation).

Notant $\delta\mathbf{Z}$ une petite variation de la variable \mathbf{Z} , alors

$$\delta\mathbf{Y} = \nabla\mathbf{F} \cdot \delta\mathbf{X} \quad (\text{F1})$$

où $\nabla\mathbf{F}$ est la matrice des dérivées partielles de \mathbf{F} par rapport à \mathbf{X} , évaluée au point \mathbf{X}_0 .

Cette relation fondamentale repose sur deux hypothèses essentielles :

- la dérivabilité de \mathbf{F} , donc l'existence du jacobien $\nabla\mathbf{F}$,
- la faible variation des entrées, autrement dit que $\delta\mathbf{X}$ est petit.

(F1) est la formule de propagation des "petits" biais. Elle sert également pour les analyses de sensibilité des résultats (quand on interprète $\delta\mathbf{X}$ comme une variation de la commande) et donc dans les calculs d'élasticité.

¹voir Tukey (1957)

La formule (F2) de propagation des marges d'incertitude découle directement de (F1) : supposant que le vecteur \mathbf{X} suit une distribution de probabilité de densité $h(\mathbf{X})d\mathbf{X}$, centrée¹ sur le point \mathbf{X}_0 , on exprime la variance (notée Var) d'une des variables endogènes² $F_i(\mathbf{X})$ comme³

$$\begin{aligned}\text{Var}(Y_i) &= \mathbf{E}(\{Y_i - Y_{i0}\}^2) = \mathbf{E}(\{\delta Y_i\}^2) : \text{on applique (F1)} \\ &= \mathbf{E}(\{\nabla F_i \cdot \delta \mathbf{X} + o(\delta \mathbf{X})\}^2) \cong \int \|\nabla F_i \cdot (\mathbf{X} - \mathbf{X}_0)\|^2 h(\mathbf{X}) d\mathbf{X} \\ &\cong \{\sum_j g_j^2 \text{Var}(X_j)\} + 2\{\sum_{j < k} g_j g_k \text{Cov}(X_j; X_k)\}\end{aligned}$$

où l'on note g_j la j -ème composante du vecteur⁴ ∇F_i : $g_j = \partial F_i / \partial X_j$

Retenons-en (F2) :

$$\text{Var}(Y_i) \cong \left\{ \sum_j \left(\frac{\partial F_i}{\partial X_j} \right)^2 \text{Var}(X_j) \right\} + 2 \left\{ \sum_{j < k} \left(\frac{\partial F_i}{\partial X_j} \right) \left(\frac{\partial F_i}{\partial X_k} \right) \text{Cov}(X_j; X_k) \right\}, \quad (\text{F2})$$

formule valide à condition que $Y_{i0} = F_i(\mathbf{X}_0)$, sans quoi le second membre majore la variance⁵ au lieu de l'approcher.

Au terme de ces nécessaires rappels, concluons à la simplicité des formules de propagation (F1) et (F2). Notons, pour la propagation de la marge d'incertitude, la décomposition en deux termes distincts : le premier uniquement lié aux variances, le second uniquement lié aux covariances. Lorsque les variables X_i sont deux-à-deux

¹ donc $X_0 = \int X h(X) dX = \mathbf{E}(X)$

² nous étudions maintenant une *seule* variable réelle Y_i et non plus un vecteur

³ \mathbf{E} désigne ici l'opérateur d'espérance mathématique par rapport à la loi de \mathbf{X} , donc la moyenne au sens de la distribution de masse $h(\mathbf{X})d\mathbf{X}$

⁴ en fait $g_j = H_{ij}$ si \mathbf{H} désigne le Jacobien de \mathbf{F} , i.e. la matrice des dérivées premières de \mathbf{F} avec en ligne j le vecteur ∇F_j des dérivées partielles de F_j par rapport aux X_i

⁵ au premier ordre, si la masse de h est concentrée près de \mathbf{X}_0 , on vérifie

$\mathbf{E}(Y_i) = \mathbf{E}(F_i(\mathbf{X})) = \mathbf{E}(F_i(\mathbf{X}_0) + \nabla F_i(\mathbf{X}_0) \cdot \delta \mathbf{X}) = Y_{i0} + \nabla F_i(\mathbf{X}_0) \cdot \mathbf{E}(\delta \mathbf{X}) = Y_{i0}$ par définition, et il n'y a pas de biais. Sinon, la formule (F2) fournit une majoration de la variance de \mathbf{Y} , car, notant alors Y_{i1} la moyenne vraie de Y_i tandis que Y_{i0} est seulement définie par $Y_{i0} = F_i(\mathbf{X}_0)$, on a

$$\begin{aligned}\text{Var}(Y_i) &= \mathbf{E}((Y_i - Y_{i1})^2) \\ &= \mathbf{E}((Y_i - Y_{i0})^2) + (Y_{i0} - Y_{i1})^2 - 2\mathbf{E}(Y_i - Y_{i0}) \cdot (Y_{i0} - Y_{i1}) \\ &= \mathbf{E}((Y_i - Y_{i0})^2) - (Y_{i0} - Y_{i1})^2 \text{ et le second membre de (F2) correspond à } \\ &\mathbf{E}((Y_i - Y_{i0})^2)\end{aligned}$$

indépendantes¹, la covariance de deux distinctes d'entre elles est nulle, ainsi donc que le second terme dans (F2).

4.3.3 Le cas d'un modèle bicritère simple

A titre d'exemple, nous reprenons l'analyse agrégée de Roger Marche (1991) pour quantifier l'incertitude d'une prédiction de recette pour une route à péage. Marche considère une variable endogène, la recette $Y = R$, qui dépend de trois variables exogènes : respectivement $X_1 = q$ le niveau de demande dans la zone d'étude, $X_2 = M$ et $X_3 = \sigma$ deux paramètres de la distribution statistique des valeurs du temps. Postulant l'indépendance entre q , M et σ , il conclut que :

$$\text{Var}(R) = \left(\frac{\partial R}{\partial q}\right)^2 \text{Var}(q) + \left(\frac{\partial R}{\partial M}\right)^2 \text{Var}(M) + \left(\frac{\partial R}{\partial \sigma}\right)^2 \text{Var}(\sigma)$$

ce qu'il quantifie prestement, avant de rajouter un quatrième terme directement sur $\text{Var}(R)$ pour tenir compte de divers autres facteurs (l'imperfection des temps de parcours simulés, calcul de l'induction...).

Dans le modèle prix-temps défini par Marche, les temps de parcours ne dépendent pas du débit, ni le volume de demande du niveau de service (temps généralisé moyen). Nous allons préciser les formules pour analyser la sensibilité et l'erreur dans ce modèle "simple" (relativement...), en étudiant le cas élémentaire d'un réseau à deux arcs. Dans la prochaine sous-section, nous pourrons alors mettre en évidence l'apport d'un enrichissement du modèle.

Soit donc un réseau élémentaire formé de seulement deux arcs parallèles en concurrence entre le noeud de départ et le noeud d'arrivée. Le premier arc schématise le réseau gratuit, tandis que le second représente la route à péage : on les repère respectivement par un indice g pour le premier et p pour le second.

Notons :

- T_g et T_p les temps de parcours moyens sur les deux routes, supposés indépendants du trafic, mais susceptibles de variations gaussiennes de variances respectives $\sigma_{T_g}^2$ et $\sigma_{T_p}^2$.
- q le niveau moyen de demande, supposé indépendant du niveau de service, mais susceptible de variations gaussiennes de variance σ_q^2 .
- M et σ respectivement la valeur du temps moyenne et l'écart-type de la distribution des logarithmes des valeurs du temps, susceptibles de variations gaussiennes de variances respectives σ_M^2 et σ_σ^2 . Soit H la fonction de répartition cumulée de la distribution log-normale des valeurs du temps : notant Φ la fonction de répartition cumulée d'une gaussienne centrée normée, on vérifie

¹autrement dit la distribution de probabilité de la variable produit $\mathbf{X} = (X_i)_i$ est égale au produit des distributions des lois marginales : $h(\mathbf{X}) d\mathbf{X} = \prod_i h_i(X_i) dX_i$

$$H(v) = \Phi\left(\frac{\ln(v/M) + \sigma^2/2}{\sigma}\right).$$

- p le niveau de péage, supposé lui bien fixé.

Alors le flot sur l'arc à péage s'exprime, en supposant bien sûr $T_p < T_g$ et en notant $\tilde{v} = p/(T_g - T_p)$ la valeur du temps de basculement entre le chemin gratuit emprunté par les automobilistes de valeur du temps inférieure à \tilde{v} et le chemin à péage choisi par ceux dont la valeur du temps dépasse \tilde{v} :

$$f_p = q \int_{\tilde{v}}^{+\infty} dH(v) = q(1 - H(\tilde{v}))$$

Cela conduit à une recette de $R = p \cdot q \cdot (1 - H(\tilde{v}))$.

Les dérivées partielles de f_p par rapport aux variables "explicatives" q , M , σ , $\Delta T = T_g - T_p$ valent respectivement :

$$[1] \quad \frac{\partial f_p}{\partial q} = 1 - H(\tilde{v}) = \frac{f_p}{q}$$

$$[2] \quad \frac{\partial f_p}{\partial M} = -q \frac{\partial H}{\partial M} = -q \frac{\partial \Phi\left(\left(\ln(\tilde{v}/M) + \sigma^2/2\right)/\sigma\right)}{\partial M} = \frac{q}{\sigma} \frac{\Phi'}{M}$$

en notant Φ' la valeur de la dérivée de Φ au point $\left(\ln(\tilde{v}/M) + \sigma^2/2\right)/\sigma$

$$[3] \quad \frac{\partial f_p}{\partial \sigma} = q \frac{\ln(\tilde{v}/M) - \sigma^2/2}{\sigma^2} \Phi'$$

$$[4] \quad \frac{\partial f_p}{\partial \Delta T} = q(-1) \frac{\Phi'}{\sigma \tilde{v}} \frac{\partial \tilde{v}}{\partial \Delta T} = \frac{q}{\sigma} \frac{\Phi'}{\Delta T}$$

Donc les élasticités¹ ε_{X_i} de f_p aux variables explicatives valent :

$$[1] \quad \varepsilon_q = 1$$

$$[2] \quad \varepsilon_M = \frac{q}{f_p} \frac{\Phi'}{\sigma}$$

$$[3] \quad \varepsilon_\sigma = \frac{q}{f_p} \frac{\Phi'}{\sigma} \left(\ln(\tilde{v}/M) - \sigma^2/2\right)$$

¹rappelons la définition d'une élasticité : c'est le rapport des variations relatives de la variable expliquée et de la variable explicative, donc à la limite

$$\varepsilon_{X_i} = \frac{\partial \ln(f_p)}{\partial \ln(X_i)} = \frac{\partial f_p}{\partial X_i} \frac{X_i}{f_p}$$

$$[4] \varepsilon_{\Delta T} = \frac{q}{f_p} \frac{\Phi'}{\sigma} = \varepsilon_M \text{ logiquement...}$$

4.3.4 Application numérique

Fixons les valeurs numériques à $p = 15$ F, $q = 3000$ véh/h, $M = 60$ F/h, $\sigma = 0.6$, $\Delta T = 0.2145$ h). On calcule d'abord la valeur du temps de coupure $\tilde{v} = 69.9$ F/h et $H(\tilde{v}) = 0.71$, la proportion de trafic capté $f_p/q = 0.29$ et $\Phi' = 0.342$, donc finalement :

$$[1] \varepsilon_q = 1 \quad \text{et} \quad \frac{\partial f_p}{\partial q} = 0.29$$

$$[2] \varepsilon_M = 1.97 \quad \text{et} \quad \frac{\partial f_p}{\partial M} = 28.5 \text{ véh/F}$$

$$[3] \varepsilon_\sigma = -0.037 \quad \text{et} \quad \frac{\partial f_p}{\partial \sigma} = -77.8 \text{ véh/h}$$

$$[4] \varepsilon_{\Delta T} = 1.97 \quad \text{et} \quad \frac{\partial f_p}{\partial \Delta T} = 7968 \text{ véh/h}^2.$$

Les facteurs les plus sensibles sont ici la valeur du temps moyenne et le différentiel de temps de parcours.

Transposons maintenant les hypothèses de Marche quant aux dispersions relatives des variables explicatives :

$$\text{supposant } \frac{\sigma_q}{q} = \frac{\sigma_\sigma}{\sigma} = \frac{\sigma_M}{M} = 10\% \text{ et } \frac{\sigma_{\Delta T}}{\Delta T} = 15\%$$

$$\text{nous obtenons } \frac{\sigma_{f_p}}{f_p} = \frac{\sigma_R}{R} = 37\%.$$

L'incertitude sur les résultats est au moins double ⁽¹⁾ de l'incertitude sur chacune des variables exogènes !

4.4 SOPHISTIQUER LE MÉCANISME DU MODÈLE

Pour juger de l'intérêt de sophistiquer, d'enrichir, le mécanisme explicatif, il faut examiner son effet sur la variabilité expliquée par le modèle.

Nous explicitons l'objectif (§4.4.1) puis sa justification (§4.4.2). Poursuivant l'exemple du modèle prix-temps, nous montrons l'intérêt d'enrichir le mécanisme explicatif (§4.4.3). Nous terminons en discutant des coûts d'une sophistication (§4.4.4).

Nous comparons deux modèles dont l'un est une restriction de l'autre, selon un critère de

¹ce résultat est bien sûr particulier aux valeurs numériques retenues

variabilité expliquée (Cf. §3.1.3).

4.4.1 Sophistiquer pour accroître la portée explicative

La sophistication d'un modèle vise à étendre son domaine de validité, à améliorer son pouvoir explicatif. Elle consiste en un enrichissement du mécanisme explicatif du modèle.

Par exemple, le modèle d'affectation bicritère à temps variables et demande totale élastique sophistique le modèle bicritère à temps fixes et demande fixe par deux nouveaux caractères mécaniques, la dépendance du temps de parcours d'une arête envers la circulation locale, et la dépendance du volume de demande d'une relation origine-destination envers le temps généralisé moyen. Il sophistique le modèle de Beckmann par un nouveau caractère, la dispersion des valeurs du temps.

Au plan statistique, la sophistication sert à prendre en compte des facteurs auparavant incontrôlés, donc à éviter des biais et à améliorer la relation à l'expérience (relation appréciée par des tests statistiques dont la sophistication accroît la puissance).

4.4.2 Fondement statistique : réduire la variabilité non-expliquée

En enrichissant le mécanisme explicatif du modèle, la sophistication vise à accroître la qualité de la prédiction. Elle doit réduire l'incertitude totale $\varepsilon_{\text{nette}}$ sur le résultat en agissant sur deux de ses composantes :

- l'erreur de conception $\varepsilon_{\text{sém}}$ (diminution appréciée qualitativement par un jugement a priori, et quantitativement en mesurant les coefficients de la propagation i.e. les termes de $\nabla_{\mathbf{X}} \mathbf{F}$).
- l'erreur exogène propagée $\approx -\nabla_{\mathbf{X}} \mathbf{F} \cdot \varepsilon_{\text{exo}}$, dont la diminution s'apprécie quantitativement sous réserve d'un audit formel nécessaire pour formuler et calculer rigoureusement $\nabla_{\mathbf{X}} \mathbf{F}$.

En attendant de disposer de suffisamment d'applications du modèle pour mesurer indirectement l'erreur de conception, étudions dès à présent l'amplitude de l'erreur exogène propagée.

L'incertitude sur les variables exogènes tient au manque de contrôle sur leurs variations. La sophistication tente d'expliquer ces variations et ainsi de réduire l'incertitude à une valeur résiduelle. Dans la théorie statistique, la sophistication se rattache donc à l'analyse de la variance (avec ou sans régression). Rappelons que le principe fondamental de l'analyse de la variance se résume dans l'équation :

$$\begin{aligned} \text{Variabilité totale} &= \text{variabilité expliquée} &+ \text{variabilité résiduelle} \\ \sum_i (Y_i - \bar{Y})^2 &= \sum_i (\hat{Y}_i - \bar{Y})^2 &+ \sum_i (\hat{Y}_i - Y_i)^2 \end{aligned}$$

avec : Y_i la i -ème observation de la variable endogène Y ,

\bar{Y} la moyenne des Y_i ,

\hat{Y}_i la valeur prédite par le modèle.

La valeur statistique du modèle s'apprécie alors en rapportant la variance expliquée à la variance résiduelle, la variance différant de la variabilité par une division par le nombre de sources d'information utilisées (les degrés de liberté : nombre de paramètres pour la partie "expliquée", nombre d'observations moins un moins nombre de paramètres pour la partie "résiduelle").

Revenons à l'erreur exogène propagée. La sophistication remplace le mécanisme \mathbf{F}_1 par le mécanisme \mathbf{F}_2 et l'erreur exogène $\approx -\nabla\mathbf{F}_1 \cdot \varepsilon_{\text{exo},1}$ devient $\approx -\nabla\mathbf{F}_2 \cdot \varepsilon_{\text{exo},2}$. Au passage l'amplitude de ε_{exo} décroît car la sophistication décompose la variabilité $\sigma_{\mathbf{X}^*,1}^2$ de $\varepsilon_{\text{exo},1}$ en

$$\begin{array}{lll} \sigma_{\mathbf{X}^*,1}^2 & = (\sigma_{\mathbf{X}^*,1}^2 - \sigma_{\mathbf{X}^*,2}^2) & + \sigma_{\mathbf{X}^*,2}^2 \\ \text{variabilité résiduelle} & = \text{variabilité nouvellement} & + \text{variabilité résiduelle} \\ \text{initiale} & \text{expliquée} & \text{finale} \end{array}$$

La nouvelle erreur exogène a priori $\varepsilon_{\text{exo},2}$ est un résidu de l'ancienne erreur exogène a priori $\varepsilon_{\text{exo},1}$.

La sophistication réduit effectivement le poste "erreur exogène propagée" si la nouvelle propagation (par $-\nabla\mathbf{F}_2$) ne détériore pas la réduction opérée sur l'erreur exogène ε_{exo} .

4.4.3 L'exemple du réseau à deux arcs (suite)

Poursuivons l'exemple de la section précédente, afin de montrer le rôle potentiel d'une sophistication pour un modèle. Soit \mathbf{F}_1 le mécanisme d'affectation bicritère à demande fixe et temps fixes décrit au §4.3.3, et \mathbf{F}_2 le mécanisme d'affectation bicritère à demande totale élastique et temps variables spécifié par Leurent (1993a) (Cf. section 8.2).

4.4.3.1 Ajout d'un lien entre débit et temps de parcours

Anticipons ici sur les résultats du chapitre 10 pour indiquer les élasticités du résultat f_L du mécanisme \mathbf{F}_2 aux variables explicatives $X = [q; M; \sigma; \Delta T]^t$, en conservant les valeurs numériques du §4.3.4. En fixant la demande pour une comparaison cohérente, on calcule :

$$\begin{array}{ll} [1'] \ \varepsilon_q = 0.94 & \text{et} \quad \frac{\partial f_p}{\partial q} = 0.27 \\ [2'] \ \varepsilon_M = 0.59 & \text{et} \quad \frac{\partial f_p}{\partial M} = 8.51 \text{ véh/F} \\ [3'] \ \varepsilon_\sigma = -0.016 & \text{et} \quad \frac{\partial f_p}{\partial \sigma} = -23.2 \text{ véh/h} \\ [4'] \ \varepsilon_{\Delta T} = 0.59 & \text{et} \quad \frac{\partial f_p}{\partial \Delta T} = 2380 \text{ véh/h}^2. \end{array}$$

On constate pour chaque facteur une diminution de l'élasticité. Cela tient à

l'incorporation de mécanismes d'équilibre dans \mathbf{F}_2 par rapport à \mathbf{F}_1 : désormais une variation d'un facteur exerce non seulement un effet direct, mais encore des effets indirects, réactions compensatrices qui atténuent l'effet direct.

Nous supposons maintenant que σ_q/q , σ_M/M et σ_σ/σ ne varient pas d'une spécification à l'autre, mais que $\sigma_{\Delta T}/\Delta T$ décroît de 15% à 5% puisque dorénavant le nouveau mécanisme explique partiellement l'incertitude initiale sur ΔT .

Nous obtenons alors $\sigma_{f_p}/f_p = \sigma_R/R = 12\%$.

La sophistication a permis de diminuer des deux tiers l'incertitude exogène propagée par le modèle, donc d'affermir considérablement la valeur de la prédiction puisque cette diminution ne s'accompagne d'aucune réduction de la portée explicative, au contraire. Il ne faut pas pour autant conclure que la valeur prédite par \mathbf{F}_2 est plus facile à calculer ! Car l'incertitude propagée intervient autour de la prédiction moyenne, sans relation au calcul du niveau de cette prédiction moyenne.

4.4.3.2 Ajout d'un lien entre niveau de service et volume de demande

En considérant maintenant le mécanisme complet de \mathbf{F}_2 (i.e. en gardant l'élasticité de la demande), on obtiendrait (Cf. chapitre 10), pour une fonction de demande d'élasticité constante $e_0 = -0.6$:

[1"] $\varepsilon_{q_0} = 0.76$	et	$\frac{\partial f_p}{\partial q_0} = 0.22$
[2"] $\varepsilon_M = 0.63$	et	$\frac{\partial f_p}{\partial M} = 9.16 \text{ véh/F}$
[3"] $\varepsilon_\sigma = -0.003$	et	$\frac{\partial f_p}{\partial \sigma} = -4.53 \text{ véh/h}$
[4"] $\varepsilon_{\Delta T} = 0.63$	et	$\frac{\partial f_p}{\partial \Delta T} = 1680 \text{ véh/h}^2$.

Ces valeurs numériques montrent que la sensibilité du volume de demande au temps généralisé moyen influe moins que la dépendance du temps de parcours envers la charge de circulation.

En supposant que σ_M/M et $\sigma_\sigma/\sigma = 10\%$, que $\sigma_{\Delta T}/\Delta T = 5\%$ et que $\sigma_{q_0}/q_0 = 10\%$, nous obtenons $\sigma_{f_p}/f_p = \sigma_R/R = 10\%$, donc une réduction supplémentaire par rapport à la précédente sophistication.

4.4.4 Quel proportionnement de l'objectif aux contraintes ?

Sophistiquer le mécanisme explicatif du modèle en accroît les coûts d'application : tant pour l'acquisition des données d'application (pour estimer davantage de paramètres, il faut étendre le dispositif d'enquête), que pour le traitement effectif (formulation et calcul).

Relativisons ce surcroît de dépenses. Concernant l'extension du dispositif d'enquête, une

solution économique peut être de fixer a priori les valeurs des nouveaux paramètres, car autant les renseigner suivant l'intuition plutôt que conserver les valeurs forfaitaires implicitement supposées dans le modèle avant sophistication.

Concernant le traitement effectif, le changement de formulation peut nécessiter un certain investissement mathématique, à évaluer (pas forcément très important). Il entraîne aussi une variation des calculs : pas nécessairement un alourdissement, car le modèle sophistiqué peut jouir de propriétés mathématiques plus favorables ; par ailleurs la régulière croissance des ressources informatiques incite à ne pas rejeter un alourdissement modéré.

Chapitre 5

Une méthode systématique d'audit

Au chapitre précédent, nous avons défini l'audit technique d'un modèle et nous en avons délimité la portée. Les concepts présentés valent de façon générale. Pour les mettre en oeuvre, ce cinquième chapitre propose une méthode systématique qui repose sur deux grands principes :

- pour déterminer les certitudes relatives au modèle, détecter les incertitudes, en prendre la mesure et si possible les réduire.
- distinguer selon la nature de l'incertitude : conceptuelle, formelle, algorithmique ou économétrique.

Nous suivons cette décomposition pour présenter notre discours de la méthode.

L'audit conceptuel (section 5.1) explicite la composition conceptuelle du modèle, afin de la comparer à la connaissance théorique du système étudié et d'apprécier sa portée. Pour chaque composant conceptuel, nous recommandons d'établir une échelle de valeur afin d'évaluer la position du modèle.

L'audit formel (section 5.2) vérifie la correction de la formule caractéristique (FC). Il contrôle l'équivalence entre la FC et la composition conceptuelle (correction sémantique) et la justesse logique (correction syntaxique). Il étudie l'existence et l'unicité des solutions à la FC.

L'audit algorithmique (section 5.3) vérifie la production d'une solution à la FC. La production est assurée par un *dispositif de résolution*, appellation générique qui recouvre tant des calculs ou des simulations numériques, que des expériences sur maquette ou autre support analogique. Dans le cas d'un traitement numérique, il faut mesurer ou majorer l'écart entre le résultat du calcul et une solution vraie à la FC, par un critère de convergence théoriquement rigoureux.

L'audit économétrique (section 5.4) explicite l'erreur exogène relative aux inputs ainsi qu'aux paramètres des fonctions de dépendance, pour la propager à travers la FC : ainsi le modèle devient transparent vis-à-vis des hypothèses économétriques.

5.1 DISCOURS DE L'AUDIT CONCEPTUEL

5.1.1 L'audit conceptuel et sa justification

5.1.1.1 Définitions

Conception du modèle : la schématisation du système étudié, la définition et le choix des variables explicatives (exogènes) ou expliquées (endogènes), la mise en relation de ces éléments (identification des mécanismes explicatifs).

Bien que la conception s'effectue en langage naturel, donnons la forme $Y = F(X)$ à un modèle dans lequel les variables endogènes Y (la réponse, l'effet) dépendent des variables exogènes X (l'argument, la cause) au travers des relations de dépendance F (le mécanisme). La conception consiste à choisir X et Y , et à décrire F en indiquant l'influence de chaque variable exogène sur chaque variable endogène (nullité ou non de $\partial F_i / \partial X_j$, si non nul quel signe, quelle amplitude et quelle régularité ?).

L'*audit conceptuel* consiste à mesurer, traiter et réduire l'erreur liée à la conception. Il s'agit de juger en quoi les hypothèses du modèle incorporent les phénomènes internes ou connexes au système étudié : quels phénomènes sont considérés ou ignorés ? Pour ceux qui sont considérés, de quelle manière, fine ou grossière ?

L'audit conceptuel délimite la *portée explicative* du modèle ; en complément, il livre des pistes pour étendre la portée, pour améliorer la représentation des phénomènes ou ajouter de nouveaux mécanismes.

5.1.1.2 Exemple du réseau à deux arcs

Reprenons l'exemple du réseau à deux routes entre une origine et une destination (Cf. §4.3.3). Le modèle traite du choix de route par des (conducteurs de) véhicules.

La conception du modèle d'affectation bicritère prix-temps à temps fixes et demande fixe repose sur les éléments suivants :

- variables endogènes : le flot sur chacune des routes.
- variables exogènes : le schéma spatial donc les deux routes, les caractéristiques des routes que sont les temps de parcours et les prix, le flot total et la caractéristique des véhicules en déplacement qu'est la distribution des valeurs du temps.
- mécanismes : la connaissance des routes par les conducteurs, l'évaluation de chaque itinéraire via un coût généralisé incluant le prix et le temps valorisé à l'aide d'une valeur du temps, le choix par chaque conducteur de la route qui minimise le coût généralisé.

Signalons que des relations comme l'égalité du flot total à la somme des flots des deux routes, sont des bilans comptables qui appartiennent au contenu phénoménologique du modèle mais n'ont pas de portée explicative.

Moins faciles à repérer, les mécanismes absents : dans l'exemple du modèle à temps fixes, omettre la dépendance du temps de parcours à l'égard du flot et de la capacité constitue une hypothèse à critiquer.

5.1.1.3 Utilité d'un audit conceptuel : juger de la portée explicative

L'audit conceptuel explicite le contenu phénoménologique¹ du modèle : il précise les variables endogènes, les variables exogènes, les relations de dépendance. Il permet ainsi d'apprécier la portée du modèle : quelles questions peut-on traiter avec pertinence, en fonction des composants conceptuels ?

Il peut servir à plusieurs usages. Un usage *didactique* : pour un modèle de compréhension et d'analyse, expliciter et critiquer les hypothèses constitue un exercice de réflexion tant pour l'apprentissage que pour la préparation d'une modification. Mettre au jour les hypothèses implicites et les critiquer constructivement, fait progresser l'analyse des phénomènes.

Un usage *technique* pour l'utilisateur du modèle, ainsi prévenu de la portée de son outil. Cette portée limite les cas d'application rigoureuse, en dehors desquels toute application est a priori litigieuse.

Un usage *sociologique*, lorsque le modèle sert d'outil d'aide à la décision. Tout acteur du processus de décision désireux de contrôler l'outil trouve dans l'audit conceptuel une délimitation a priori de la portée théorique.

5.1.1.4 Antécédents

L'audit conceptuel ainsi défini apparaît en partie dans la plupart des traités de modélisation : tant en physique qu'en économie ou en statistique, les distinctions entre variables endogènes et variables exogènes, les présentations et justifications individuelles de chacune des relations de dépendance, sont des opérations d'audit sémantique. En physique on distingue encore, parmi les variables exogènes, les commandes fixées par l'utilisateur, des perturbations imposées par l'environnement (Cohen 1994, Walter et Pronzato 1994). En macro-économie on distingue, parmi les relations de dépendance, celles qui correspondent à un mécanisme de production ou de consommation ou d'échange, des simples bilans comptables (Muet 1994).

5.1.2 Comment mesurer et traiter l'erreur de conception ?

5.1.2.1 Possibilités a priori de mesure

Dans la décomposition de l'erreur totale en plusieurs sources (conception, formulation, résolution, économétrie), l'erreur de conception est la source la plus difficile à mesurer quantitativement. L'exploration qualitative des écarts entre le mécanisme et la représentation théorique laisse juger a priori son rôle éventuel, mais ne livre pas de repères quantitatifs.

Donc on mesure l'erreur de conception soit qualitativement, soit indirectement.

Qualitativement : le jugement a priori et l'analyse systémique servent pour comparer un

¹nous préférons ce qualificatif à celui de "comportemental", bien que "comportemental" serve habituellement à qualifier un modèle de demande de transport phénoménologique. En effet "comportemental" prête à confusion par rapport à d'autres terrains scientifiques : en matière de traitement du signal (Walter et Pronzato, 1994), un modèle est qualifié de "comportemental" quand il ne vise qu'à reproduire des résultats, non des lois mécaniques ou sociales !

mécanisme explicatif à la représentation théorique. On apprécie l'écart de façon qualitative, en dressant une liste d'approximations, d'hypothèses simplificatrices.

Indirectement : lorsque des observations expérimentales, donc des réalisations de l'erreur totale ε , laissent déduire la valeur de l'erreur conceptuelle par soustraction des autres sources d'erreur, appréciées elles plus directement. Une mesure indirecte, par soustraction des autres sources d'erreur à une erreur totale estimée empiriquement, nécessite de contrôler a posteriori de multiples applications : tâche lourde, mais probablement hautement instructive ! Elle permettrait de distinguer les hypothèses cruciales des hypothèses accessoires, de même qu'elle livrerait un montant forfaitaire d'erreur à prendre en compte dans une prévision.

5.1.2.2 Une méthode d'exploration systématique pour la mesure qualitative

Nous proposons de systématiser la mesure qualitative de l'erreur de conception. Il s'agit d'abord de recenser les objets élémentaires (Cf. des acteurs) et les relations entre objets (Cf. des comportements) internes ou connexes au système étudié¹. C'est là une analyse systémique, qui conduit à préciser la définition du système et de son environnement. On synthétise cette analyse en explicitant les relations élémentaires de conception, les mécanismes explicatifs, qui expriment, portent, la connaissance théorique au sujet du système étudié : en un mot, en explicitant des composants sémantiques.

Alors, considérant chaque composant sémantique, il s'agit d'inventorier les représentations abstraites qui s'y rapportent, puis de les hiérarchiser en fonction de leur élaboration ou de les juxtaposer si elles sont contradictoires ou irréductibles l'une à l'autre, pour enfin constituer une échelle de valeur. C'est dans de telles échelles que l'on apprécie les hypothèses de représentation propres à un modèle donné.

5.1.2.3 Vers des traitements qualitatifs

Une mesure qualitative systématique de l'erreur de conception ouvre la voie à plusieurs traitements qualitatifs, ponctuels ou fondamentaux.

Des traitements ponctuels : vis-à-vis d'une application particulière du modèle, la mesure livre une connaissance théorique de la portée, qui permet de juger la pertinence d'une hypothèse ou d'une interprétation. Autrement formulé, le bon artisan n'est pas un artisan qui dispose d'un bon outil, mais un artisan conscient des qualités et défauts de son outil et capable d'en tirer un bon parti.

Des traitements fondamentaux : l'identification des limites théoriques de la représentation ouvre la voie à des modifications, par correction ou par élaboration. Il s'agit d'améliorer l'outil pour en repousser les limites. A cette fin, la reconnaissance des notions absentes fournit immédiatement des possibilités d'amélioration.

5.1.2.4 Disciplines impliquées

L'audit conceptuel ainsi défini met en oeuvre l'analyse systémique pour l'identification des variables et de leurs rôles respectifs, et le jugement a priori pour l'établissement des

¹il peut être utile de qualifier les relations selon qu'elles traduisent un comportement physique, ou un fonctionnement technique, ou un comportement économique, ou un bilan de conservation (bilan comptable)

échelles de valeur relatives à chacun des phénomènes représentés.

5.1.3 Procédure d'application de la méthode d'exploration systématique

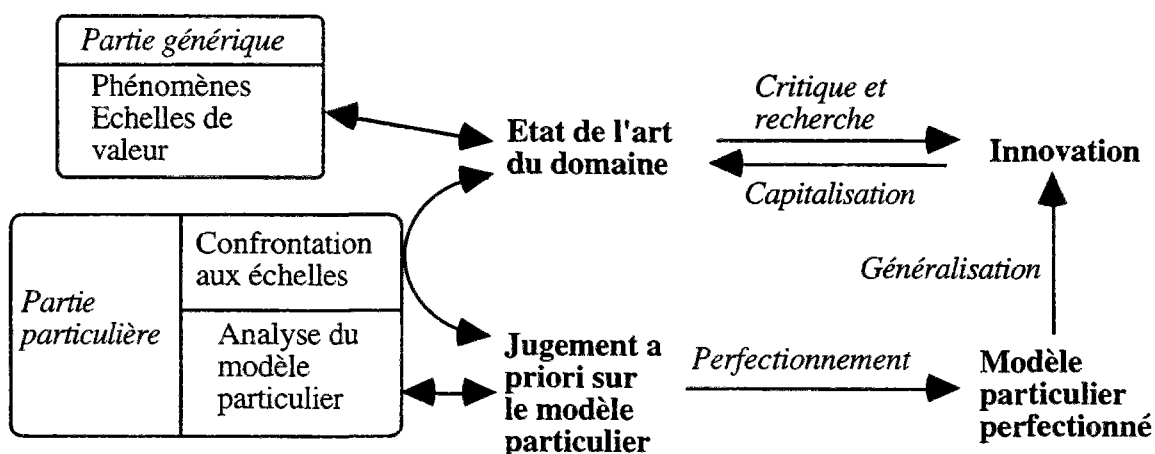
5.1.3.1 Une partie générique et une partie particulière

La méthode d'exploration systématique se compose de deux parties :

- en amont, une partie à portée générique, à savoir l'inventaire des phénomènes connexes à l'objet d'étude et des solutions de représentation qui s'y rapportent, ainsi que la hiérarchisation des solutions en fonction de leur conformité à la représentation théorique. On obtient alors des échelles de valeur.
- en aval, une partie à portée particulière, à savoir la confrontation d'un modèle donné aux échelles de valeur, pour juger a priori de sa pertinence.

Chaque partie a des applications plus larges que le jugement a priori d'un modèle. La partie générique permet d'apprécier l'état de l'art du domaine, d'identifier les insuffisances et de prescrire des pistes de développement. La partie particulière relativise les hypothèses propres au modèle examiné et met en évidence de possibles perfectionnements.

Fig. 5a. La méthode d'exploration systématique.



5.1.3.2 Phasage

Les qualifications d'amont pour la partie générique de l'audit conceptuel et d'aval pour sa partie particulière procèdent d'une exigence logique, notamment pour l'élaboration et l'emploi des échelles de valeur. De fait, la partie générique repose sur l'analyse de divers modèles particuliers, dont elle opère une synthèse, une capitalisation.

5.2 DISCOURS DE L'AUDIT FORMEL

5.2.1 L'audit formel et sa justification

5.2.1.1 Définitions

La *formulation du modèle* comprend une formalisation et une caractérisation. La formalisation énonce en langage mathématique les composants conceptuels dans une *formule synthétique*. La caractérisation transforme, réduit, la formule synthétique en un problème type de mathématique ; elle produit une *formule caractéristique* qui jouit de certaines propriétés d'existence et d'unicité.

Soit $\mathbf{Y} = \mathbf{F}(\mathbf{X})$ un modèle reliant des variables endogènes \mathbf{Y} et exogènes \mathbf{X} au travers d'une dépendance \mathbf{F} . La formalisation consiste à définir précisément \mathbf{F} , à expliciter les hypothèses techniques nécessaires (ex. des hypothèses mathématiques minimales sur les fonctions de dépendance). La caractérisation consiste à placer la formule $\mathbf{Y} = \mathbf{F}(\mathbf{X})$ dans un cadre mathématique de référence (par exemple les systèmes de n équations à n inconnues).

L'*erreur de formulation* procède d'un manque de conformité entre la formule caractéristique et la composition conceptuelle, ou d'un manque de cohérence logique (s'il existe des contradictions internes), ou d'une insuffisance de caractérisation (lorsque les hypothèses ne suffisent pas pour déterminer complètement une solution).

L'*audit formel* consiste à mesurer, traiter et réduire l'erreur de formulation : c'est essentiellement un contrôle logique et mathématique. Il y a naturellement une très forte interaction entre formulation et audit formel : non seulement la formulation livre l'objet de l'audit formel, mais encore, dès qu'il y a recours au langage mathématique, la position des formules s'effectue dans un souci consubstantiel de vérité logique, donc de certitude formelle.

5.2.1.2 Exemple du réseau à deux arcs

Reprenons l'exemple du réseau à deux arcs joignant une origine à une destination (Cf. §4.3.3). Pour un modèle bicritère à demande fixe et temps fixes, la formulation consiste à exprimer les variables endogènes $\mathbf{Y} = (f_g; f_p)$ en fonction des variables exogènes $\mathbf{X} = (q; H; T_g; T_p; P)$.

Vient d'abord une hypothèse technique minimale : $P \geq 0$.

Puis on distingue deux cas, selon les positions relatives de T_g et T_p :

$$(i) (T_g \leq T_p) \text{ et alors } \begin{cases} F_1(\mathbf{X}) = q \\ F_2(\mathbf{X}) = 0 \end{cases}$$

$$(ii) (T_g > T_p) \text{ et alors } \begin{cases} F_1(\mathbf{X}) = q \cdot H(\tilde{v}) \\ F_2(\mathbf{X}) = q \cdot (1 - H(\tilde{v})) \end{cases} \text{ où } \tilde{v} = P / (T_g - T_p).$$

Dans les deux cas, le modèle se formule comme une dépendance directe des variables endogènes envers les variables exogènes, avec $\mathbf{F}(\mathbf{X}) = (F_1(\mathbf{X}); F_2(\mathbf{X}))$. La définition de

F est précise et cohérente. Sur ce cas simple, l'audit formel est trivial //

Pour un modèle bicritère à demande variable et temps variables, la formulation est plus complexe. Les variables exogènes sont $\mathbf{X} = (D; H; t_g; t_p; P)$ où la fonction de demande D relie le volume q au niveau de service \bar{S} via $q = D(\bar{S})$, et où la fonction de temps de parcours t_g (R. t_p) relie le temps local T_g (R. T_p) au flot local f_g (R. f_p) via $T_g = t_g(f_g)$ (R. $T_p = t_p(f_p)$).

Les variables endogènes sont $\mathbf{Y} = (q; f_g; f_p; G; \chi_g; \chi_p; \bar{S}; T_g; T_p)$ avec G fonction de temps généralisé minimum (définie pour une valeur du temps), χ_g fonction indicatrice de l'affectation d'une valeur du temps au chemin gratuit et χ_p fonction indicatrice de l'affectation d'une valeur du temps au chemin à péage.

La dépendance caractéristique F doit prendre en compte :

- la formation des temps de parcours $T_g = t_g(f_g)$ et $T_p = t_p(f_p)$ (5.1) et (5.2)

- le choix du chemin de moindre temps généralisé $G(v) = \min \{T_g; T_p + \frac{P}{v}\}$ (5.3)

et $\forall v \chi_g(v) = 1$ si $T_g \leq T_p + \frac{P}{v}$ ou 0 sinon (5.4)

et $\forall v \chi_p(v) = 1 - \chi_g(v)$ (5.5)

- la formation du niveau de service $\bar{S} = \int G(v) dH(v)$ (5.6)

- la fonction de demande $q = D(\bar{S})$ (5.7)

- l'affectation au chemin de moindre temps généralisé et l'agrégation des valeurs du temps dans les flots :

$$f_g = q \int \chi_g(v) dH(v) \text{ et } f_p = q \int \chi_p(v) dH(v) \quad (5.8) \text{ et } (5.9)$$

Soit au total neuf dépendances élémentaires dont six de dimension 1 et trois de dimension infinie si la répartition H est continue...

Le travail mathématique de formulation consiste à réduire cette complexité, à simplifier la formule caractéristique (ici l'ensemble des relations (5.1) à (5.9)) afin de l'étudier dans un cadre mathématique connu et de pouvoir juger de la cohérence. On peut procéder en deux étapes¹.

La première étape consiste à définir la valeur du temps de coupure \tilde{v} :

$$\begin{cases} t_g(f_g) \leq t_p(f_p) \Rightarrow \tilde{v} = +\infty \\ t_g(f_g) > t_p(f_p) \Rightarrow \tilde{v} = P / (t_g(f_g) - t_p(f_p)) \end{cases} \quad (5.10)$$

Cela permet de court-circuiter les dépendances élémentaires de dimension infinie : (5.3) à (5.5) deviennent

¹solution alternative, rester en dimension infinie, Cf. Leurent 1993b et section 8.3

$$\begin{cases} v \leq \tilde{v} \Rightarrow G(v) = t_g(f_g) ; \chi_g(v) = 1 ; \chi_p(v) = 0 \\ v > \tilde{v} \Rightarrow G(v) = t_p(f_p) + \frac{P}{v} ; \chi_g(v) = 0 ; \chi_p(v) = 1 \end{cases} \quad (5.11)$$

$$(5.4) \text{ devient } \bar{S} = \left\{ \int_0^{\tilde{v}} t_g(f_g) dH(v) \right\} + \left\{ \int_{\tilde{v}}^{+\infty} \left(t_p(f_p) + \frac{P}{v} \right) dH(v) \right\} \quad (5.12)$$

$$(5.8) \text{ et } (5.9) \text{ deviennent } f_g = q \cdot H(\tilde{v}) \text{ et } f_p = q \cdot (1 - H(\tilde{v})). \quad (5.13) \text{ et } (5.14)$$

On peut ainsi transformer \mathbf{Y} en $\tilde{\mathbf{Y}} = (q; f_g; f_p; \bar{S}; \tilde{v})$, tandis que \mathbf{F} se réduit à $\tilde{\mathbf{F}}$ composé de (5.7), (5.10) à (5.14), soit cinq dépendances élémentaires de dimension 1.

La seconde étape consiste à éliminer q , \bar{S} et \tilde{v} . J'ai montré (Cf. section 8.2) que l'on peut transformer $\tilde{\mathbf{Y}}$ en $\bar{\mathbf{Y}} = (f_g; f_p)$ solution de $\min_{\bar{\mathbf{Y}} \geq 0} \mathbf{J}_{\mathbf{X}}(\bar{\mathbf{Y}})$, avec

$$\begin{aligned} \mathbf{J}_{\mathbf{X}}(f_g; f_p) &= \left\{ \int_0^{f_g} t_g(\theta) d\theta \right\} + \left\{ \int_0^{f_p} t_p(\theta) d\theta \right\} \\ &+ (f_g + f_p) P \left(\eta(1) - \eta\left(\frac{f_g}{f_g + f_p}\right) \right) - \int_0^{f_g + f_p} D^{-1}(\theta) d\theta \end{aligned} \quad (5.15)$$

où D^{-1} est la fonction réciproque de D et $\eta(x) = \int_0^x \frac{d\theta}{H^{-1}(\theta)}$ est une primitive de l'inverse de la fonction réciproque de H . Sous des hypothèses de croissance pour t_g , t_p et $-D$, ce programme d'optimisation est convexe, donc il existe une solution, unique si l'une des fonctions croît strictement ou si $P > 0$.

Alors la formule caractéristique n'est pas directe comme dans le cas à temps fixes mais indirecte, via une fonction implicite.

5.2.1.3 Utilité : détecter des erreurs aux effets incontrôlables

En cas d'audit formel défavorable, il faut rejeter ou modifier le modèle. En effet cela signifie que la formule caractéristique

[A] n'est pas conforme aux composants sémantiques : alors le modèle ne remplit pas son objectif technique !

[B] ou comporte des fautes logiques : de telles fautes compromettent le sens des relations entre les variables et risquent d'induire des erreurs aux effets incontrôlables !

[C] ou n'est pas suffisante, c'est-à-dire que les relations de dépendance ne déterminent pas entièrement les variables endogènes.

[A] ou [B] nécessite de corriger le modèle, ou sinon de l'abandonner faute de validité technique. [C] est moins contraignant, car dans certains cas l'on se contente de calculer une solution, sans se soucier de l'éventuelle multiplicité des solutions.

Donc l'audit formel sert à détecter des erreurs qui, en l'absence de correction, retireraient son sens au modèle formel. Il établit qu'un modèle qui l'a subi avec succès jouit d'une formulation rigoureuse et suffisante au plan mathématique.

5.2.1.4 Antécédents

Les formulations mathématiques de modèles constituent autant d'exercices d'audit formel, en ce qui concerne la conformité de la formulation aux mécanismes explicatifs et la vérification de la cohérence logique.

Les traités de logique mathématique sont des manuels pour la partie [B] de l'audit formel (examen de la cohérence logique).

L'identifiabilité (partie [C]) est abondamment traitée en théorie du signal (Cf. Walter et Pronzato 1994) : mais en se bornant au cas d'une structure connue et d'un nombre restreint de paramètres scalaires à estimer, la question étant de savoir si les informations disponibles suffisent pour lever toute ambiguïté. L'identifiabilité ainsi définie relève de l'audit économétrique, non de l'audit formel.

5.2.2 Comment mesurer et traiter l'erreur de formulation ?

5.2.2.1 Mesurer la cohérence logique puis la conformité à la conception

On mesure la cohérence logique en évaluant, point par point, la véracité des hypothèses faites, et la justesse des chaînons formels qui mènent des hypothèses aux formules caractéristiques de la solution du modèle. Il s'agit notamment d'examiner si chaque terme est correctement et suffisamment défini.

Le résultat de cette mesure est une liste, éventuellement vide, de points d'achoppement ou d'incohérence. Il convient de vider cette liste avant de poursuivre.

Lorsque la liste est vide, donc après que la cohérence logique interne de la formule caractéristique a été établie, on peut mesurer sa conformité à la composition conceptuelle en étudiant si la formule caractéristique est compatible avec chacun des composants conceptuels¹. Cette étude doit aboutir favorablement, sous peine de modifier la formule caractéristique.

5.2.2.2 Mesure et traitement vont de pair

Nous avons prescrit de traiter les éventuelles incohérences logiques dès leur détection, et de corriger la formule caractéristique en cas de non conformité aux relations élémentaires de la conception. Le traitement intervient à l'issue de chacune des deux phases de mesure.

Le traitement est de nature mathématique : par construction mathématique, chaque point particulier d'achoppement ou d'incohérence peut recevoir une solution, éventuellement au prix d'hypothèses supplémentaires qui risquent de peser sur l'erreur de conception. S'il y a incompatibilité entre formule caractéristique et relations élémentaires de conception, une solution est de simplifier provisoirement les relations élémentaires (donc la portée explicative) afin d'obtenir la compatibilité, puis de sophistication progressivement la formule caractéristique afin d'obtenir les relations élémentaires telles que posées dans la conception.

¹il suffit de vérifier l'absence d'incompatibilité : de fait la formule caractéristique ne permet pas toujours de retrouver chacun des composants conceptuels car elle résulte d'une réduction mathématique

Mais une telle sophistication progressive n'aboutit pas nécessairement. Faut-il assécher l'erreur de formulation, quitte à augmenter l'erreur de conception ? Thèse : oui, car de toute façon comment expliquer, justifier, le résultat d'un modèle mal formulé, incohérent ? Antithèse : un modèle mieux conçu, même mal formulé, pourrait produire un résultat plus proche de la réalité. Synthèse : s'il s'agit de donner des garanties, oui, car il faut connaître précisément les limites du modèle pour en garantir le résultat.

5.2.2.3 Questions d'existence, unicité et stabilité

La formule caractéristique, après validation de la cohérence logique et de la conformité aux relations élémentaires de la conception, résume en langage formel le contenu sémantique du modèle.

Ce contenu est-il cohérent ? Ou, de façon équivalente, existe-t-il une solution à la formule caractéristique ? Le sens du modèle en dépend crucialement. Etablir l'existence d'une solution à la formule caractéristique permet d'affirmer la consistance logique de la représentation phénoménologique. A contrario, une preuve d'infaisabilité, de non existence, invalide le contenu sémantique et force au remaniement.

Pour étudier l'existence et l'unicité d'une solution, il est bon de plonger la formule caractéristique dans un cadre mathématique connu, où s'appliquent des résultats théoriques généraux. Dans le cas d'un modèle d'affectation, il s'agit de transformer les conditions d'équilibre en un problème type de recherche opérationnelle.

Par stabilité, nous entendons la continuité de la réponse du modèle par rapport à de petites variations des arguments. Une étude à ce sujet permet de circonscrire les éventuels effets d'aléas numériques ou d'incertitudes sur les variables exogènes.

5.2.3 Pour préparer l'audit formel dans un domaine spécifique

Pour auditer formellement un modèle particulier, il est utile de connaître l'état de l'art du domaine, c'est-à-dire les cadres mathématiques de référence pour une formule caractéristique. Contrairement à la méthode d'exploration systématique proposée pour l'audit conceptuel, il n'est pas nécessaire de confronter le modèle à toute une série de possibilités : il suffit de trouver une solution appropriée, de choisir l'outil idoine dans une boîte à outils mathématiques.

Faire un état de l'art du domaine, c'est-à-dire constituer une boîte à outils mathématiques, prépare aussi bien l'audit formel d'un modèle existant, que la formulation d'un nouveau modèle. La boîte à outils est avantageusement documentée par des références à des modèles particuliers, car parfois un modèle particulier emploie des objets originaux (Cf. la notion de valeur du temps de coupure dans le modèle bicritère) : ce sont des ressources de formulation au même titre que les cadres mathématiques de référence.

5.3 DISCOURS DE L'AUDIT ALGORITHMIQUE

Nous considérons le seul cas où le dispositif de résolution prend la forme d'une simulation numérique. Compte tenu du développement de l'informatique, cette forme tend à s'imposer dans des domaines toujours plus nombreux, au détriment (ou en complément) des dispositifs analogiques (ex. maquette). L'erreur de résolution coïncide alors avec l'erreur de calcul.

Celle-ci se décompose en plusieurs postes : défaillance du matériel ou du compilateur, faute d'implantation de l'algorithme, imprécision par arrondi, défaut de convergence (§5.3.1). L'audit algorithmique, vis-à-vis de chaque poste, consiste à détecter, corriger et même prévenir les erreurs (§5.3.2). En cas de défaillance du matériel ou du compilateur, les réponses sortent du cadre de la modélisation, tout comme les fautes d'implantation : cependant il importe de détecter ces erreurs car elles peuvent invalider l'application du modèle (§5.3.3).

Les imprécisions numériques se prêtent à une détection systématique et sont, en l'état actuel des ressources informatiques, aisément traitables (§5.3.4). Enfin la convergence de l'algorithme se mesure par un critère rigoureux (§5.3.5).

5.3.1 Une analyse de l'erreur de calcul

5.3.1.1 Définitions

Algorithme au sens strict. C'est un procédé de calcul itératif qui engendre une suite de nombre réels ou de vecteurs. On formalise un algorithme comme une application \mathbf{W} d'un espace \mathbf{U} dans lui-même, et alors son déroulement à partir d'une valeur initiale $\mathbf{u}_{(0)}$ suit la forme : $\mathbf{u}_{(0)} \in \mathbf{U}$, $\mathbf{u}_{(n)} \mapsto \mathbf{u}_{(n+1)} = \mathbf{W}(\mathbf{u}_{(n)})$, $n = 1, 2, \dots$

Convergence d'un algorithme. On dit qu'un algorithme converge si les suites de valeurs qu'il engendre tendent vers une limite désirée dans l'espace considéré. Pour un modèle formel donné, une limite désirée est évidemment une solution de la formule caractéristique.

J'appelle *dispositif de résolution* d'un modèle, ou *algorithme au sens large*, un dispositif qui produit une solution à la formule caractéristique du modèle formel. Cette définition recouvre tant des calculs ou des simulations numériques, que des expériences sur maquette ou autre support analogique.

La nature du support de simulation est importante en pratique, mais pas déterminante par rapport au problème de l'audit : dans le cas d'une maquette, le lien à l'objet d'étude s'exprime par des formules d'analyse dimensionnelle qui constituent la formule caractéristique du modèle. Le caractère approché de ces formules par rapport au comportement de la maquette constitue l'erreur de résolution, à traiter de façon analogue à l'erreur de calcul du dispositif numérique.

Modèle résoluble. J'appelle modèle résoluble l'assemblage du modèle formel et d'un dispositif de résolution.

5.3.1.2 Décomposition de l'erreur de calcul

Nous définissons l'erreur de calcul comme l'erreur susceptible de survenir dans le modèle algorithmique, sans tenir au modèle formel. Dans le cas d'une résolution numérique, elle est liée à l'algorithme au sens strict ou au support matériel.

Concernant l'algorithme au sens strict, l'erreur de calcul peut tenir au principe s'il y a un caractère heuristique, ou au manque de convergence si l'algorithme est nécessairement itératif, ou enfin au manque de précision dans la représentation des quantités et dans l'exécution des opérations élémentaires.

Au sujet du support matériel, l'erreur de calcul peut tenir à une faute d'implantation (erreur de programmation) ou à une défaillance technique (problème lié au matériel ou au compilateur).

5.3.1.3 Exemple du réseau à deux arcs

Reprenons l'exemple du réseau à deux arcs en concurrence entre une origine et une destination, avec une affectation bicritère à demande variable et temps variables (Cf. §5.2.1.2). Si l'on calcule la solution en appliquant l'algorithme d'égalisation par transvasement (Leurent, 1995c, repris au §9.2.1.3), on risque :

- un défaut de principe si l'on se contente de considérer un seul itinéraire, en omettant de traiter aussi le second : comment l'algorithme identifie-t-il les chemins réputés accessibles ?
- un défaut de convergence, si l'on cesse les transvasements trop précocement.
- un manque de précision numérique : par exemple, si l'on considère une distribution log-normale de la valeur du temps, le calcul de l'impédance d'un chemin nécessite d'évaluer des fonctions qui ne sont pas disponibles dans tous les compilateurs de développement, et qui sont donc programmées de façon particulière dans le logiciel d'application.
- une erreur d'implantation, si l'auteur du logiciel a omis un des termes dans la fonction d'impédance d'un chemin.
- une défaillance technique, si le calcul se fait sur un ordinateur dont le processeur est défectueux¹.

Sans compter d'éventuelles erreurs de l'utilisateur dans le codage des données ou le réglage de l'algorithme !

5.3.2 L'audit algorithmique et sa justification

5.3.2.1 Définir l'audit algorithmique

L'audit algorithmique consiste à mesurer, traiter et réduire l'erreur de calcul. Il se décline selon chacun des postes constitutifs de l'erreur de calcul : pour la plupart d'entre eux, la mesure procède par détection d'une faute, et le traitement par correction de cette faute. Alors la réduction est complète. Dans le cas du manque de convergence d'un algorithme

¹Cf. l'erreur détectée en 1994 par IBM dans le processeur Pentium fourni en première génération par INTEL

itératif, la mesure s'effectue à l'aide de critères quantitatifs, le traitement nécessite de continuer les itérations jusqu'à réduire l'erreur en deçà d'un certain seuil.

5.3.2.2 Utilité : détecter et prendre en compte des erreurs qui risquent d'invalider la quantification

Comment juger a priori des effets d'une erreur de calcul ? Nous l'ignorons ! Ce n'est probablement pas possible. L'ignorance à ce sujet pousse à *dépister* et à *éradiquer* l'erreur de calcul, afin de lever la menace qu'elle constitue à l'égard des résultats du modèle alimenté. L'audit algorithmique tente de détecter et d'éliminer des erreurs aux effets a priori incontrôlables, pour garantir la validité technique du modèle résoluble. Il s'agit surtout d'une action préventive.

5.3.2.3 Lieu

Dans l'erreur de calcul, la partie liée au manque de convergence d'un algorithme itératif est la plus apparente pour l'utilisateur final du modèle alimenté, et la seule qu'il peut contrôler (dans les cas favorables). Au contraire les autres postes engagent la responsabilité de l'auteur de l'algorithme ou de l'auteur du logiciel d'application, ou même des fournisseurs du matériel et du compilateur.

5.3.2.4 Antécédents

Au chapitre 3, nous avons cité deux études empiriques de l'erreur de convergence, en matière d'affectation du trafic. Plus nombreux sont les travaux consacrés à la mise au point d'algorithmes et de critères de convergence rigoureux (Cf. Patriksson 1994 pour un large inventaire d'algorithmes). Au sujet des fautes d'implantation, Easa (1985) corrige un algorithme de recherche des plus courts chemins qui traitait mal certains types de mouvements tournants.

De façon moins spécialisée, les manuels d'apprentissage de la programmation contiennent des règles pour prévenir et corriger les fautes d'implantation logicielle d'un algorithme ; il existe plusieurs méthodes de certification logicielle pour examiner systématiquement si un algorithme remplit correctement sa mission (Hoare 1973).

Enfin les traités d'algorithmique donnent des objets et des méthodes pour mettre au point des algorithmes rigoureux (ex. Knuth 1973, Aho et al. 1974, Minoux et Bartnik 1986, Fenton et Hill 1993).

5.3.3 Fautes d'ordre matériel ou logiciel

5.3.3.1 Cas des défaillances de matériel ou de compilateur

La correction de telles défaillances incombe au constructeur de la machine ou à l'auteur du compilateur : elle échappe à l'auteur du logiciel d'application du modèle. Celui-ci dispose cependant de moyens de détection et de prévention.

Détection. Disposer de jeux de données standards, pour lesquels on connaît à l'avance le résultat d'une application valide, permet de tester une plate-forme matérielle nouvelle ou un nouveau compilateur. La question se pose de façon relativement fréquente, compte tenu de la rapide évolution des ressources informatiques.

Prévention. Si les tests ne livrent pas le résultat escompté, la prudence commande de

signaler le problème au fournisseur et de continuer provisoirement avec les moyens antérieurs.

5.3.3.2 Cas des fautes d'implantation

Les fautes d'implantation sont des erreurs dans le code source du logiciel d'application du modèle ; elles surviennent lorsque l'auteur du logiciel réalise un produit qui ne respecte pas les spécifications fixées. Il s'agit donc d'erreurs par rapport à une tâche assignée a priori, et non d'erreurs dans la tâche assignée (que nous englobons dans l'erreur de convergence, Cf. §5.3.5).

Exemples. Citons quelques possibles fautes d'implantation, pour un modèle d'affectation du trafic :

- ne pas considérer tous les objets définis dans une application, ex. dans la lecture d'un fichier définissant les éléments du réseau, omettre certains des arcs de façon systématique mais difficilement décelable à l'utilisation.
- mal traiter des configurations particulières, des cas limites qui nécessitent un autre traitement que le cas général.

Correction. Le seul traitement admissible, pour une faute d'implantation, est l'éradication totale par correction. Aussi l'auteur du logiciel doit-il *détecter* la faute, afin d'intervenir localement.

Détection. Le compilateur détecte automatiquement certaines fautes d'implantation grossières, qui ne respectent pas la syntaxe obligée du langage de programmation. Pour détecter les autres fautes d'implantation, l'auteur du logiciel "débogue" son produit en l'appliquant à des jeux de données de test et, en cas d'écart au résultat attendu, en suivant étape par étape la réalisation du processus afin de localiser l'erreur dans le code. L'utilisateur final peut également soumettre le logiciel à des tests de dépistage.

Prévention. Il existe des règles de l'art en matière de programmation :

- structuration en modules, chargés chacun d'une fonction particulière, afin de "décomposer pour comprendre".
- lisibilité du code : à obtenir en choisissant des variables informatiques dont la signification est conforme aux éléments de spécification, en donnant des noms explicites aux variables et aux procédures, en incluant des commentaires détaillés.

Appliquer consciencieusement ces règles permet d'éviter de nombreuses fautes d'implantation, et de détecter plus facilement celles qui se glissent toutefois.

5.3.4 Mesurer et traiter les imprécisions numériques

Certaines erreurs de calcul procèdent d'un défaut de précision numérique : elles peuvent se manifester tant dans une opération élémentaire (ex. somme de deux nombres) qu'au niveau d'ensemble de l'algorithme (convergence). Par imprécision, nous entendons défaut de précision numérique dans une opération élémentaire, liée à un arrondi ou à une définition approximative de fonction.

5.3.4.1 Exemples

Une quantité est traitée par l'ordinateur de façon plus ou moins précise, en fonction de

la variable informatique qui la représente, et plus exactement du type de la variable. Dans les langages de programmation courants, plusieurs types sont à la disposition du programmeur : nombre entier sur 1, 2, 4 ou 8 octets, nombre réel sur 4, 6, 8 ou 12 octets etc. Le nombre d'octets propre au type limite la précision de la représentation numérique interne au programme. La tradition du développement informatique prône la parcimonie dans l'emploi des moyens : pour économiser l'espace mémoire et le temps de calcul, ne pas recourir systématiquement à la précision maximale, mais se contenter du strict nécessaire. Comment évaluer le strict nécessaire ? Parfois on arbitre en faveur d'une précision réduite qui permet avec un même matériel de traiter des applications de plus grande taille ; cela n'amène pas forcément de difficulté. Cependant il faudrait justifier rigoureusement le choix et étudier a priori les conséquences possibles.

Donnons maintenant un exemple concret de définition approximative de fonction : le logiciel MARS (Leurent 1995d) est programmé en langage Pascal, qui n'offre pas de façon standard la fonction de répartition d'une loi gaussienne centrée normée. Or cette fonction sert constamment pour traiter, dans le modèle bicritère, une distribution log-normale des valeurs du temps. Donc il a fallu programmer une approximation de la fonction : nous avons choisi $\Phi(x) = 1 - \phi(x)(b_1t + b_2t^2 + b_3t^3 + b_4t^4 + b_5t^5)$ (1) (d'après Saporta 1990), avec une précision meilleure que 10^{-7} : cela nous a paru suffisant mais limite ipso facto la précision des résultats.

5.3.4.2 Prévention, détection et correction

La prévention des imprécisions consiste à employer là où c'est possible (c'est-à-dire pas trop coûteux) une précision plus que minimale, avec par exemple deux chiffres significatifs de plus que le dernier utile en pratique pour obtenir des résultats jugés suffisamment précis.

La détection s'opère à partir d'applications de référence dans lesquelles la précision est un point sensible, ou à la lecture du code source en comparant la précision requise pour chaque variable au type choisi, et en étudiant la précision des fonctions élémentaires programmées par l'auteur du logiciel d'application.

La correction procède par changement de type des variables codées jusqu'alors avec une précision insuffisante, et par amélioration de la précision des fonctions élémentaires programmées directement.

5.3.5 Mesurer et traiter l'erreur de convergence

Par erreur de convergence, nous entendons l'incertitude qui s'attache au résultat d'un algorithme en raison d'un problème de convergence, parce que l'algorithme ne converge pas ou parce qu'il converge mais de façon insatisfaisante ou insuffisante.

L'absence de convergence peut tenir à un défaut théorique, lorsque l'algorithme est heuristique. L'insuffisance de convergence, dans le cas d'un algorithme justifié

¹la formule vaut pour $x \geq 0$; pour les x négatifs, il faut utiliser la symétrie de la fonction de répartition par rapport au point $(0 ; 1/2)$. On a noté $\phi(x)$ la fonction de densité de la distribution gaussienne centrée normée, $t = 1 / (1 + 0.2316419x)$, $b_1 = 0.319381530$, $b_2 = -0.356563782$, $b_3 = 1.781477937$, $b_4 = -1.821255978$, $b_5 = 1.330274429$

théoriquement, tient à l'arrêt d'une procédure itérative à l'issue d'un nombre limité d'itérations, autrement dit à la troncature de l'algorithme.

Vient d'abord une question primordiale : l'algorithme converge-t-il vers un état final, ou tout au moins finit-il par fournir des états qui vérifient certaines propriétés ?

En cas de réponse négative, l'audit algorithmique tourne court : l'algorithme n'a pas de sens vis-à-vis du modèle formel. En cas de réponse positive, il faut répondre à de nouvelles questions au sujet de l'état final (ou des états qui vérifient certaines propriétés) :

- vérifie-t-il la formule caractéristique ?
- avec quelle précision est-il caractérisé ?
- est-il unique, ou dépend-il de l'état initial ?
- est-il stable, par rapport à de petites variations de l'état initial ?

La mesure de l'erreur de convergence est in fine d'ordre numérique, mais elle nécessite un examen théorique des questions posées : l'examen théorique doit établir un critère de convergence rigoureux, dont le calcul effectif fournisse une mesure numérique de l'erreur de convergence.

Le traitement de l'erreur de convergence dépend des résultats de la mesure. Si le critère de convergence indique une incertitude faible, résiduelle, on peut la prendre en compte de façon économétrique (ou l'ignorer si elle est vraiment faible). Si l'incertitude mesurée est élevée, il y a lieu de poursuivre les calculs afin de préciser l'état final.

Au total, la mesure et le traitement s'articulent autour de la notion de critère de convergence. La prévention consiste à déterminer un critère rigoureux et à s'assurer que l'exécution de l'algorithme tend effectivement à le réduire. La détection consiste à vérifier l'existence d'un critère rigoureux, tâche d'ordre théorique, et à examiner la valeur prise par le critère dans l'application numérique de l'algorithme. La correction consiste à réduire la valeur prise par le critère, en prolongeant l'exécution de l'algorithme, ou à modifier l'algorithme s'il ne permet pas une réduction suffisante.

5.3.5.1 Déterminer un critère de convergence rigoureux

Par critère de convergence rigoureux, nous entendons une fonction mathématique continue à valeurs réelles positives ou nulles, appliquée aux variables endogènes, dont l'annulation soit équivalente au fait que les variables endogènes vérifient la formule caractéristique.

Un critère dont l'annulation marque l'obtention par l'algorithme d'un état final, mais n'implique pas que l'état final vérifie la formule caractéristique, n'est pas rigoureux.

Lorsque la formule caractéristique du modèle est $\mathbf{F}(\mathbf{X}; \mathbf{Y}) = \mathbf{0}$ avec \mathbf{X} vecteur des variables exogènes, \mathbf{Y} vecteur des variables endogènes, \mathbf{F} vecteur des fonctions de dépendance, un critère rigoureux naturel est $C_{\mathbf{X}}(\mathbf{Y}) = \|\mathbf{F}(\mathbf{X}; \mathbf{Y})\|^2$.

5.3.5.2 S'assurer que l'algorithme tend à réduire la valeur prise par le critère

Dans une application de l'algorithme, la décroissance vers zéro d'un critère de convergence rigoureux démontre la convergence effective : il s'agit d'une garantie a posteriori, obtenue après l'exécution de l'algorithme. Comment prendre une garantie a

priori, antérieure à l'exécution de l'algorithme ? Réponse : en démontrant la convergence théorique de l'algorithme vers un état final qui annule le critère rigoureux ; une information complémentaire utile est la vitesse de convergence de l'algorithme.

Donnons un vocabulaire pour formaliser l'étude théorique de la convergence et de la vitesse de convergence, en plagiant Culioli (1994). Rappelons qu'un algorithme se formalise comme une application \mathbf{W} d'un espace \mathbf{U} dans lui-même, et que son déroulement à partir d'une valeur initiale $\mathbf{u}_{(0)}$ suit la forme :

$$\mathbf{u}_{(0)} \in \mathbf{U}, \quad \mathbf{u}_{(n)} \mapsto \mathbf{u}_{(n+1)} = \mathbf{W}(\mathbf{u}_{(n)}), \quad n = 1, 2, \dots$$

A \mathbf{W} , on associe l'ensemble de ses points fixes $\mathbf{W}^\infty(\mathbf{U}) = \{\mathbf{u} \in \mathbf{U} ; \mathbf{W}(\mathbf{u}) = \mathbf{u}\}$.

Un point fixe $\bar{\mathbf{u}}$ est dit *attractif* s'il admet un voisinage $\mathbf{V}(\bar{\mathbf{u}}) \subset \mathbf{U}$ tel que

$$\forall \mathbf{u} \in \mathbf{V}(\bar{\mathbf{u}}), \quad \limsup_{n \rightarrow \infty} \mathbf{W}^n(\mathbf{u}) = \bar{\mathbf{u}}.$$

Un sous-ensemble $\bar{\mathbf{U}} \subset \mathbf{U}$ est dit *attractif* si

$$\forall \bar{\mathbf{u}} \in \bar{\mathbf{U}}, \quad \exists \mathbf{V}(\bar{\mathbf{u}}), \quad \forall \mathbf{u} \in \mathbf{V}(\bar{\mathbf{u}}), \quad \lim_{n \rightarrow \infty} \mathbf{W}^n(\mathbf{u}) \in \bar{\mathbf{U}}.$$

Soit \mathbf{U}^* l'ensemble des solutions de la formule caractéristique du modèle : \mathbf{W} doit être tel que $\mathbf{W}^\infty(\mathbf{U})$ soit proche de \mathbf{U}^* et attractif. On appelle *bassin d'attraction* d'un point fixe $\bar{\mathbf{u}}$ l'ensemble des états initiaux \mathbf{u}_0 qui conduisent à une suite $\mathbf{W}^n(\mathbf{u}_0)$ qui converge vers $\bar{\mathbf{u}}$. L'algorithme converge si les suites de valeurs qu'il engendre tendent vers une limite désirée dans l'espace considéré, c'est-à-dire si ses points fixes attractifs satisfont la formule caractéristique (i.e. $\in \mathbf{U}^*$) et si leurs bassins d'attraction recouvrent l'ensemble des états initiaux.

Pour un algorithme, un critère de convergence rigoureux est donc une mesure de la distance entre un état courant \mathbf{u} et l'ensemble des solutions \mathbf{U}^* . La *vitesse de convergence* d'un algorithme concerne la décroissance vers zéro de la distance entre les valeurs engendrées et leur limite : cette notion n'a d'intérêt que si la limite appartient à l'ensemble des solutions. On mesure la vitesse de convergence¹ en comparant la distance entre \mathbf{u}_{n+1} et $\bar{\mathbf{u}} = \mathbf{W}^\infty(\mathbf{u}_0)$, à la distance entre \mathbf{u}_n et $\bar{\mathbf{u}}$. La vitesse s'étudie donc a priori, et suppose une connaissance de l'état limite $\bar{\mathbf{u}} = \mathbf{W}^\infty(\mathbf{u}_0)$.

¹Prenons l'exemple de \mathfrak{R}^N avec la norme euclidienne $\|\cdot\|$:

- si $\limsup_{n \rightarrow \infty} \frac{\|\mathbf{u}_{n+1} - \bar{\mathbf{u}}\|}{\|\mathbf{u}_n - \bar{\mathbf{u}}\|} = \alpha < 1$ la convergence est dite linéaire, de taux α .

- si $\lim_{n \rightarrow \infty} \frac{\|\mathbf{u}_{n+1} - \bar{\mathbf{u}}\|}{\|\mathbf{u}_n - \bar{\mathbf{u}}\|} = 0$ on dit que la convergence est superlinéaire.

- si $\exists \gamma > 1$, $\limsup_{n \rightarrow \infty} \frac{\|\mathbf{u}_{n+1} - \bar{\mathbf{u}}\|}{\|\mathbf{u}_n - \bar{\mathbf{u}}\|^\gamma} = 0$ alors la convergence est superlinéaire d'ordre γ .

5.3.5.3 Examiner, interpréter et prendre en compte la valeur du critère

Dans l'application d'un algorithme intrinsèquement itératif, il est bien improbable que l'on obtienne la solution exacte à l'issue d'un nombre fini d'itérations. Aussi l'utilisateur doit-il forcer la terminaison, en arrêtant la procédure dès que l'état courant satisfait à une certaine condition de convergence, exprimée comme l'abaissement du critère de convergence en dessous d'un certain seuil.

Comment fixer le seuil de convergence ? Il doit correspondre à une erreur "tolérable", suffisamment réduite pour être considérée comme nulle ou pour être propagée par une procédure d'analyse de sensibilité (ce qui permet une prise en compte économétrique). Pour déterminer l'erreur tolérable, il faut interpréter le critère par rapport aux variables endogènes considérées dans l'analyse des résultats. Si le critère n'est pas très intuitif, il peut être bon de lui adjoindre un deuxième critère qui soit intuitif et qui serve non à terminer l'algorithme, mais à titre de "témoin".

5.3.6 Synthèse

Le tableau 5A récapitule les quatre postes d'erreur de calcul et les traitements associés : il fait la part des responsabilités, réparties principalement entre l'auteur et l'utilisateur du logiciel.

Tab. 5A. Récapitulation des quatre postes d'erreur de calcul.

Poste	Outil opératoire	Tâche	Responsable
Défaillance du matériel ou du compilateur	Jeux de données de test	Détection Prévention, correction	Auteur du logiciel Fournisseur
Faute d'implantation	Jeux de données de test	Détection Correction	Auteur/utilisateur du logiciel Auteur du logiciel
Imprécision numérique		Détection, Prévention, Correction	Auteur du logiciel
Convergence	Critère rigoureux de convergence	Conception, implantation Appréciation, interprétation	Auteur du logiciel Utilisateur du logiciel

Constatons pour conclure la diversité des postes d'erreur de calcul : cela peut donner lieu à confusion dans l'application du modèle. Nous avons corrigé ailleurs le modèle logit linéaire (Cf. Leurent, 1994b, 1995b, 1997b et §8.1.1) : la formule caractéristique de Fisk, bien que correcte, a été généralement traitée sans rigueur dans la définition des chemins considérés. Cette erreur est restée longtemps inaperçue, car l'algorithme employé était dépourvu d'un critère de convergence rigoureux : la convergence apparente d'un critère non rigoureux masquait l'erreur de formulation.

Autre exemple, relatif à un modèle d'affectation bicritère : si les chemins traités sont définis de façon approchée, alors la fonction objectif comme la borne inférieure sont calculées de façon non rigoureuse. Cela peut donner lieu à des situations où un critère de convergence, formulé rigoureusement comme la différence entre la valeur de la fonction objectif et la borne inférieure, devient négatif sans pour autant signifier la convergence.

Nous connaissons un tel logiciel, dont l'auteur interprète ces situations comme des imprécisions numériques, au lieu de remettre en cause la formulation et l'implantation !

5.4 DISCOURS DE L'AUDIT ÉCONOMÉTRIQUE

Appliquer un modèle résoluble à un cas concret soulève les questions propres de l'alimentation économétrique : quelles valeurs donner aux paramètres, quelles formes attribuer aux fonctions ? Les réponses procèdent d'investigations spécifiques ou d'expériences antérieures transposées, sinon de l'imagination pure. Le modèle résoluble muni d'une alimentation économétrique constitue le modèle alimenté, paré à la simulation quantitative. L'audit économétrique consiste à apprécier l'alimentation économétrique, à détecter et à prendre en compte l'erreur économétrique.

Nous présentons des éléments pour traiter d'alimentation économétrique (§5.4.1) et d'audit économétrique (§5.4.2) : en particulier la notion de modèle d'estimation, distinct du modèle de simulation. Puis nous expliquons en quoi consiste l'audit économétrique (§5.4.3). Nous terminons en proposant une méthode pour préparer l'audit économétrique d'un modèle appartenant à un domaine spécifique (§5.4.4).

5.4.1 Définir et qualifier l'alimentation économétrique

5.4.1.1 Définitions

La *statistique* a pour objet l'étude d'ensembles ("populations") dont les éléments ("individus") présentent à la fois des ressemblances et des dissemblances, en opérant des qualifications (démarcations entre sous-ensembles) et des quantifications (descriptions numériques). Elle comprend une partie *descriptive*, pour résumer les principaux *caractères* d'une population connue, et une partie *inférentielle*, pour établir une connaissance de l'ensemble d'après l'observation de quelques éléments.

La *théorie des probabilités* sert d'outil à la statistique, afin notamment de construire des modèles probabilistes pour l'inférence. *L'économétrie* est une branche de la statistique dédiée à l'étude des variables et des relations économiques : en raison de la difficulté des expériences en la matière, l'aspect expérimental s'y estompe quelque peu au profit de l'aspect inférentiel et prédictif.

L'*alimentation économétrique* d'un modèle comprend les valeurs numériques des paramètres et de façon plus large les formes des fonctions. Elle englobe aussi les choix concrets de codage, ainsi que le renseignement numérique des hypothèses de simulation en situation virtuelle.

Il faut distinguer entre le *modèle de simulation alimenté*, qui incorpore et prolonge le modèle résoluble, et le (ou les) *modèle(s) d'estimation* par lesquels passe l'alimentation économétrique du modèle de simulation. Un modèle d'estimation est un dispositif d'acquisition et de traitement d'information, afin d'alimenter un modèle de simulation.

Par exemple, en matière d'affectation : le modèle de simulation est le modèle d'affectation, tandis qu'un modèle d'estimation est une représentation particulière de la valeur du temps (ou des flots origine-destination), rattachable à des observations.

Plus généralement, formulons un modèle de simulation par une relation $Y = F(\Theta; X)$ où Y est la réponse (la conséquence), X l'input (la cause), F le mécanisme explicatif (le lien

de cause à effet) et Θ le paramètre. Signalons d'abord la proximité entre X et Θ qui sont tous deux des inputs. Signalons encore la proximité entre Θ et F car F incorpore des hypothèses mathématiques analogues, quoiqu'*implicites*, à celles faites pour spécifier un paramètre. En simulation, on fixe F et Θ , et pour une valeur de l'input X on produit la réponse Y . Autrement dit, Y est inconnue a priori, mais déterminée par la donnée de F , Θ et X . En estimation, on fixe F et des couples input-résultat $(X_i; Y_i)$ dénommés "observations" : Θ est inconnu a priori mais déterminé par une règle d'estimation à partir de F et des $(X_i; Y_i)$. Une règle classique d'estimation consiste à donner à Θ la valeur qui minimise un critère d'écart entre les valeurs observées Y_i et les valeurs simulées $F(\Theta; X_i)$. Ainsi apparaît une relation de dualité entre le modèle de simulation et le modèle d'estimation.

L'adjonction d'une alimentation économétrique à un modèle résoluble produit un *modèle alimenté*, paré à la simulation quantitative.

5.4.1.2 Place de l'alimentation économétrique dans le modèle de simulation

L'alimentation économétrique parachève le modèle résoluble qu'elle transforme en outil quantitatif. Elle peut (et devrait de façon systématique) donner lieu à une confrontation entre modèle de simulation et observations de la réalité, celles-ci menées au travers d'un ou plusieurs modèle(s) d'estimation. Ce caractère empirique est spécifique à l'alimentation économétrique, alors que tant la composition conceptuelle que la formule caractéristique et l'algorithme gardent un caractère théorique.

Cependant la portée pratique de l'alimentation économétrique (donc du modèle alimenté) reste conditionnelle à la portée sémantique, à la cohérence de la formule caractéristique et à l'effectivité du dispositif de résolution.

La distinction entre composition conceptuelle et alimentation économétrique sert à relativiser la place de l'économétrie dans un modèle de simulation : en théorie économétrique, la conception relève du "jugement a priori". Un modèle économétrique est un modèle d'estimation, que l'on peut ou non rattacher à un modèle de simulation ; cependant un modèle de simulation ne se réduit pas nécessairement aux modèles d'estimation qui lui sont rattachés.

5.4.1.3 Erreur économétrique

L'erreur économétrique englobe l'*erreur d'estimation* attachée aux paramètres incorporés aux relations, et l'*erreur exogène* a priori attachée aux variables exogènes. La frontière entre paramètres et variables exogènes est mouvante, selon que l'on traite tel ou tel paramètre comme une variable exogène ou non. Dans une application, on considère qu'une variable exogène à portée globale est un paramètre, contrairement à une variable exogène à portée locale.

Pour notre propos, assimilons provisoirement les paramètres à des variables exogènes¹.

¹la distinction entre erreur d'estimation et erreur exogène tient surtout à une tradition statistique, selon laquelle un paramètre est nécessairement estimé d'après des observations donc connu indirectement, tandis qu'une variable exogène est réputée "observable" donc directement connaissable. Cette tradition correspond au traitement conventionnel d'un modèle de régression linéaire, mais convient moins bien au traitement d'un modèle de simulation.

L'erreur exogène se décompose en erreur d'observation (de mesure) et en erreur de fluctuation d'échantillonnage (Cf. section 3.1), ainsi qu'en "erreur de supputation" survenant lorsque l'on renseigne la valeur d'une variable sans référence à des observations, notamment dans la simulation d'une situation virtuelle.

En passant par le modèle de simulation, l'erreur exogène a priori se transforme en une erreur exogène propagée, attachée au résultat de la simulation.

5.4.1.4 Modèle de simulation et modèle(s) d'estimation

Ainsi l'alimentation économétrique procède par adjonction au modèle de simulation (MS) d'un ou plusieurs modèles d'estimation (ME_i). Les inputs d'un modèle d'estimation ME_i comprennent des inputs et des outputs particuliers du MS ainsi que des observations du système modélisé. Un ME_i produit des outputs, en particulier des paramètres du MS.

Chaque modèle d'estimation ME_i a une *forme* particulière, parmi une gamme de possibilités. A un extrême de la gamme on trouve le modèle probabiliste entièrement spécifié, à l'autre extrême de simples hypothèses a priori, notamment pour une simulation en situation virtuelle. Entre ces deux extrêmes se situent la reprise de résultats livrés par un modèle extérieur, ou la transposition (le transfert) d'expériences analogues.

L'appréciation de la forme n'est pas nécessairement simple : même un modèle probabiliste entièrement spécifié incorpore des hypothèses a priori, par exemple le choix d'une régression linéaire plutôt que non linéaire, ou, dans le cas d'une régression non linéaire, la spécification mathématique de la fonction de dépendance.

5.4.1.5 Exemple du réseau à deux arcs

Reprenons l'exemple du réseau à deux arcs concurrents joignant une origine à une destination. Le MS est le modèle d'affectation du flot origine-destination aux deux itinéraires. Dans le cas d'une affectation bicritère à demande variable et temps variables, le MS a pour inputs la fonction de demande, la distribution de la valeur du temps, les fonctions de temps de parcours ; ses outputs sont le niveau de service composite, le volume origine-destination, les flots et les temps de parcours d'équilibre (Cf. section 7.2).

On peut rattacher trois ME_i à ce MS : pour l'estimation respectivement des valeurs du temps, du flot origine-destination et des fonctions de temps de parcours. Considérons un modèle d'estimation ME de la distribution des valeurs du temps, par une enquête aux itinéraires : les inputs de ce ME sont des flots observés, une simulation de répartition et des variables liées au site d'enquête ; ses outputs sont les paramètres de la distribution des valeurs du temps.

5.4.2 Définir et justifier l'audit économétrique

5.4.2.1 Définition et produit

L'erreur économétrique s'attache au modèle de simulation au travers des modèles d'estimation ; elle se décompose en erreur d'estimation et en erreur exogène.

L'audit économétrique consiste à détecter l'erreur économétrique dans le modèle de

simulation, à en prendre la mesure et, si possible, à la réduire. Il apprécie l'aptitude prédictive du modèle, c'est-à-dire la capacité du modèle à reproduire des observations différentes de celles ayant servi à estimer les paramètres. Il attache des bandes d'incertitude (intervalles de confiance) aux résultats de la simulation.

Il donne des éléments de réponse aux questions que voici :

- que vaut le mécanisme explicatif du modèle ?
- Quelles sont les sources potentielles d'erreur dans une application ?
- Quelle incertitude confèrent-elles aux simulations ?
- Comment en tenir compte dans l'interprétation des résultats ?
- Comment réduire les biais et les incertitudes ?
- Comment estimer au mieux un modèle ?
- Qu'apporte une sophistication du modèle ?

A titre d'exemple, nous avons ébauché l'audit économétrique du modèle d'affectation à deux arcs (Cf. §4.4.3), en établissant des bandes d'incertitude autour des variables endogènes déterminées dans la simulation.

5.4.2.2 Interaction avec l'alimentation économétrique et la conception

L'audit économétrique prolonge et approfondit l'alimentation économétrique du modèle de simulation. Un désaccord entre observations et simulations peut résulter d'une alimentation défailante. Si l'on a audité l'alimentation, par exemple en fixant "au mieux" les paramètres, la persistance d'un désaccord "important" met en cause la conception du modèle, ses caractères sémantiques, ou tout au moins son applicabilité au cas d'étude.

L'analyse de la variance constitue un exemple d'interaction entre audit et alimentation économétriques et composition conceptuelle : on conçoit un modèle de simulation, on l'alimente grâce à un modèle d'estimation qui lui est conforme. Pour l'auditer économétriquement, on répartit la variabilité observée entre une partie dite expliquée qui est attribuable aux inputs et une partie résiduelle, et, en comparant les importances relatives des deux parties, on apprécie la crédibilité du modèle de simulation.

5.4.2.3 Limites : l'audit économétrique demeure relatif et subjectif

Plusieurs raisons militent pour relativiser l'audit économétrique. Tout d'abord sa validité technique est conditionnelle à celle du modèle résoluble, donc aux audits conceptuel, formel et algorithmique. Ensuite l'erreur économétrique n'est jamais identifiée dans l'absolu : vouloir reconnaître et traiter tous les facteurs d'incertitude est une entreprise illusoire, tout comme prétendre aller au fond des choses. Enfin, même pour un facteur d'incertitude dûment identifié, les hypothèses particulières aux situations virtuelles induisent une erreur de supposition.

5.4.2.4 Portée et enjeux

L'audit économétrique remplit des fonctions essentielles par rapport au modèle alimenté : il perfectionne et accomplit l'alimentation économétrique dont dépendent les résultats numériques de la simulation, il donne des repères pour apprécier la signification statistique des résultats.

L'enjeu ultime de l'audit économétrique est d'apprécier de façon empirique et quantitative la qualité de la simulation : au lieu de juger sans démonstration précise que le modèle convient ou ne convient pas, l'audit économétrique pose la question sous forme statistique et quantitative. Quand il conclut positivement, il donne une présomption favorable¹ quant à la validité empirique du modèle.

Alors un résultat de simulation ne se présente plus comme une simple valeur numérique, mais comme une valeur numérique assortie d'une bande d'incertitude, qui laisse apprécier la crédibilité. L'audit économétrique indique comment interpréter cette bande d'incertitude en fonction des hypothèses du modèle : ainsi le modèle devient transparent pour l'utilisateur du résultat, qui obtient en accompagnement du résultat moyen l'incertitude provenant des hypothèses.

Enfin, entre deux modèles de simulation aux principes différents mais aux mêmes objets d'estimation, l'audit économétrique peut indiquer lequel représente le mieux les phénomènes traités (Cf. section 4.4).

5.4.2.5 Antécédents

De façon générale, articles et traités de statistique préparent à l'audit économétrique en exposant des modèles probabilistes où l'erreur économétrique apparaît explicitement. Il faut prendre garde au fait que de tels modèles sont, par construction, des modèles d'estimation (donc aussi de simulation, mais avec une portée restreinte). Cette focalisation sur l'estimation est reprise dans les traités de modélisation qui abordent l'économétrie, même s'ils sont dédiés à une science physique particulière (ex. Walter et Pronzato 1994 pour le traitement du signal) ou à l'ingénierie du trafic (ex. Taylor et Young 1988).

Le traitement d'un modèle d'estimation est donc une opération classique. En matière de modélisation du trafic, il existe de multiples références au sujet de l'estimation tant des flots origine-destination que des paramètres comportementaux comme la valeur du temps. On trouve peu de références concernant le traitement d'un modèle de simulation, encore moins lorsque la formule caractéristique est complexe. Nous avons cité au §3.4.2 deux travaux consacrés à l'affectation, que nous évaluons au §10.1.1.

Cette absence de références s'explique probablement par la multi-disciplinarité intrinsèquement nécessaire au traitement d'un modèle de simulation : il faut recourir non seulement à la statistique, mais encore à la branche des mathématiques appliquées particulière au domaine dont relève le modèle de simulation, par exemple l'optimisation numérique ou l'automatique.

5.4.3 Teneur de l'audit économétrique

5.4.3.1 Schéma d'ensemble

Pour auditer l'alimentation économétrique d'un modèle de simulation, qu'elle soit formalisée ou non, nous préconisons d'identifier les objets d'estimation et, pour chacun d'eux, d'explicitier le modèle d'estimation. L'explicitation du modèle d'estimation sert à détecter l'erreur économétrique, erreur d'estimation ou erreur exogène a priori dont en

¹dès qu'on lui applique un doute philosophique, toute science ne donne au mieux que des présomptions favorables

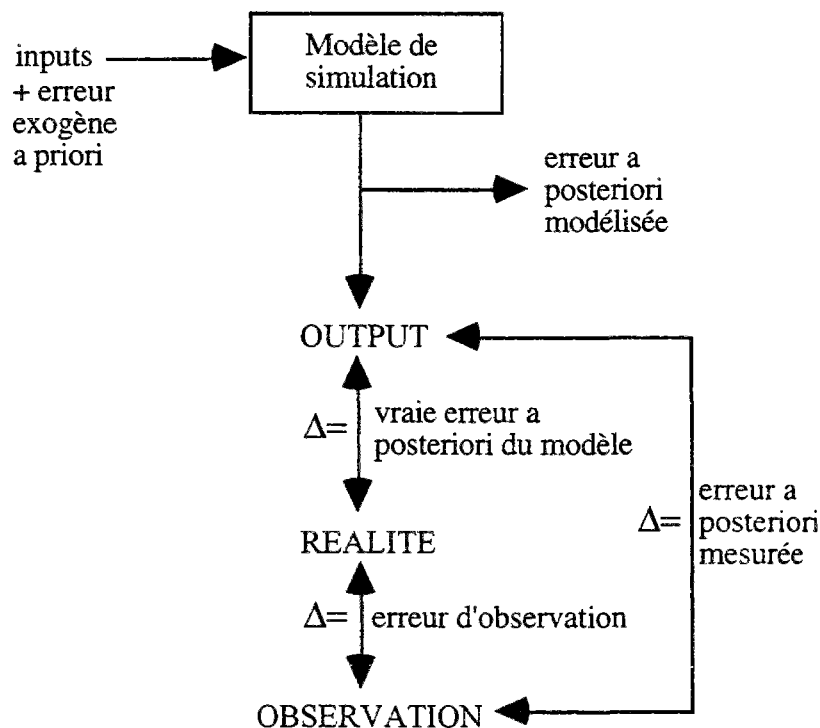
particulier l'erreur de supputation. Cette erreur économétrique, entièrement assimilée à une erreur exogène a priori relative aux inputs du modèle de simulation, se propage via le modèle de simulation et produit alors une erreur exogène propagée qui est une erreur a posteriori attendue.

Que représente l'erreur a posteriori attendue (modélisée) par rapport à l'erreur a posteriori réalisée ? Deux conditions sont nécessaires à une comparaison : d'abord la réalisation d'une erreur a priori, autrement dit le phénomène simulé doit se manifester dans la réalité, puis l'observation de la réalisation. A une erreur d'observation près, l'erreur a posteriori constatée représente l'erreur a posteriori réalisée.

Appliquer le modèle de simulation à une situation de validation, qui correspond à une réalisation pour laquelle on dispose d'observations, permet de calculer l'erreur a posteriori constatée, que l'on peut alors comparer à l'erreur a posteriori attendue. Une proximité entre les deux erreurs a posteriori n'est qu'une présomption de validité économétrique de la simulation ; mais à l'inverse un éloignement dénie la validité et met en cause le modèle de simulation dans son ensemble, conception, formulation, résolution et alimentation économétrique.

Dans la simulation d'une situation virtuelle, l'erreur a posteriori attendue, vis-à-vis de l'erreur a posteriori réalisable, constitue le seul indice.

Fig. 5b. Schéma de principe de l'audit économétrique.



a priori = avant application du modèle de simulation
a posteriori = après application

5.4.3.2 Reconnaître les objets d'estimation

L'audit conceptuel inventorie les sujets et relations élémentaires qui interviennent dans le modèle de simulation. Chaque variable exogène est un objet d'estimation potentiel, ou plutôt chaque type de variable exogène est un objet d'estimation potentiel, car par

généralisation on applique un modèle d'estimation pour un type de variable.

L'identification des objets d'estimation permet de détecter les erreurs exogènes possibles, en marquant leurs lieux d'occurrence potentielle.

Dans la pratique, on conserve moins d'objets d'estimation qu'il n'y a de variables exogènes : dans l'exemple de l'affectation, un arc ne constitue pas un objet d'estimation, au contraire de sa fonction de temps de parcours. Autrement dit, on suppose (consciemment ou non) qu'une erreur non explicitée est négligeable. La rigueur commanderait certes de mener une expérience pour tester une telle hypothèse.

5.4.3.3 Pour un objet, expliciter le modèle d'estimation et l'erreur économétrique

Nous avons évoqué la variété formelle des modèles d'estimation, selon que leur contenu probabiliste est plus ou moins explicite, et plus ou moins rigoureux. En fonction de la forme, l'erreur économétrique liée à l'objet d'estimation est plus ou moins facile à caractériser. L'erreur économétrique contient une erreur de supputation sur les valeurs attribuées aux variables exogènes, particulièrement dans la simulation d'une situation virtuelle : si ces valeurs résultent d'un modèle de simulation extérieur, alors peut-être connaît-on a priori la distribution statistique de l'erreur de supputation. L'explicitation d'un modèle probabiliste permet de préciser l'erreur économétrique.

Finalement, l'erreur économétrique renvoie à des croyances subjectives : hypothèses probabilistes a priori, ou confiance accordée à un modèle extérieur, ou supputations relatives à des variables inobservables. L'explicitation et la mise en forme probabiliste confèrent à ces croyances subjectives une transparence et une rigueur qui en facilitent le partage et leur donnent une certaine objectivité.

5.4.3.4 Intégrer les modèles d'estimation et propager l'erreur exogène

L'audit économétrique a intrinsèquement un caractère systématique : les objets d'estimation doivent recouvrir toutes les variables dont on suppose que la contribution à l'erreur économétrique n'est pas négligeable. Cela peut nécessiter de disposer de plusieurs modèles d'estimation, consacrés chacun à un objet particulier. La juxtaposition des erreurs économétriques attachées aux modèles d'estimation produit l'erreur exogène a priori, à propager par le modèle de simulation pour obtenir l'erreur exogène propagée, désignée encore comme l'erreur a posteriori attendue.

5.4.3.5 Comparaison des modèles

Dans la section 4.4, nous avons expliqué comment comparer deux modèles dont l'un sophistique l'autre, en confrontant les erreurs exogènes propagées.

5.4.4 Pour préparer l'audit économétrique dans un domaine spécifique

L'audit économétrique d'un modèle de simulation particulier apporte des éléments qui sont facilement *transposables* au domaine spécifique dont relève le modèle.

En sens inverse, la connaissance des ressources d'audit économétrique relatives à un domaine spécifique prépare et facilite le traitement d'un modèle particulier. Ces ressources comprennent une liste d'objets d'estimation, puis pour chaque objet des

modèles d'estimation possibles, enfin des méthodes de propagation liées aux formules caractéristiques du domaine.

5.5 QUELLES ARTICULATIONS ENTRE LES DIMENSIONS ?

Pour appliquer l'audit dans chacune de ses dimensions, un ordre naturel consiste à suivre le développement du modèle (§5.5.1). Il représente une possibilité parmi d'autres, une première approche que les relations entre les dimensions d'audit peuvent conduire à adapter et réviser selon le cas traité (§5.5.2).

5.5.1 Processus naturels d'audit, dans le développement et l'application d'un modèle

A quel moment convient-il d'auditer un modèle ? En pratique, la demande de garanties survient en général lors d'une application du modèle alimenté. Cependant, les audits conceptuel, formel et algorithmique ne dépendent pas d'une application particulière, donc ils peuvent être accomplis en laboratoire.

5.5.1.1 Auditer le modèle résoluble en suivant l'ordre de l'élaboration

Dans l'élaboration d'un modèle résoluble, à la conception succède la formulation, puis la mise en place d'un dispositif de résolution. L'audit peut reprendre cet ordonnancement, d'autant qu'en tout état de cause l'audit algorithmique n'a de sens que si l'audit formel aboutit favorablement.

5.5.1.2 Dans une application

Si le modèle résoluble peut être audité en laboratoire, en revanche chaque application, donc chaque modèle alimenté, appelle un audit économétrique particulier. Un travail en laboratoire peut y contribuer en fournissant des outils : une procédure de propagation de l'erreur exogène, ou une méthode d'estimation pour certains paramètres. Cependant il faut alimenter le modèle alimenté propre à l'application en données relatives au système étudié. Cela englobe aussi bien le codage que les variables exogènes.

Là encore, nous affirmons que l'audit économétrique n'a de sens que si l'audit du modèle résoluble a abouti favorablement.

5.5.1.3 Existence de rétroactions

L'ordonnancement de l'audit selon un phasage conceptuel, formel, algorithmique, économétrique, paraît donc assez naturel. Constatons cependant l'existence de rétroactions entre les différentes dimensions. Par exemple les paramètres sont déterminés lors de l'audit économétrique ; or leurs valeurs numériques conditionnent le problème mathématique à résoudre, donc l'audit algorithmique... Autre exemple, entre dimensions conceptuelle et économétrique dans un modèle d'affectation : supposons que l'on estime une valeur du temps à partir d'une enquête aux itinéraires et de temps de parcours moyens mesurés (Cf. §10.2.3) ; si l'audit économétrique établit une dispersion des temps de parcours autour de leur valeur moyenne, alors sans doute faut-il incorporer cette dispersion au modèle, donc reconsidérer la composition conceptuelle.

L'ordonnancement conceptuel-formel-algorithmique-économétrique représente donc une première approche, qui doit mettre en évidence les rétroactions spécifiques au modèle

considéré. La nécessité comme le déroulement des éventuelles approches ultérieures dépendent des résultats du premier examen.

5.5.2 Les interactions entre dimensions conduisent à des arbitrages

5.5.2.1 L'évidente interaction entre les dimensions formelle et algorithmique

L'audit algorithmique s'opère à partir d'une certaine formulation mathématique ; sa portée propre dépend donc de la validité de la formulation, autrement dit de l'audit formel.

5.5.2.2 Entre les dimensions conceptuelle et économétrique

Le codage fournit l'exemple d'une interaction entre audits conceptuel et économétrique. Pensons pour l'affectation au découpage du territoire étudié en zones de demande : le niveau de finesse relève d'un choix de conception, avec des implications au niveau formel (par exemple la distribution des valeurs du temps dépend-elle de la relation origine-destination ?) ; cependant il se traduit au niveau de l'audit économétrique par une incertitude relative aux données origine-destination. Cette incertitude dépend elle-même de la taille des zones de demande. Donc l'audit des choix de conception et l'audit économétrique sont dans une certaine mesure substituables l'un à l'autre.

5.5.2.3 Interaction entre dimensions formelle et économétrique

Nous avons déjà affirmé la vacuité d'auditer économétriquement un modèle dont l'audit formel se révèle négatif. De façon proche, l'audit formel contraint l'audit économétrique en imposant des propriétés à satisfaire : par exemple la monotonie d'une fonction contraint les valeurs des paramètres de cette fonction.

5.5.2.4 Nécessité d'arbitrer entre les dimensions

Les relations entre les dimensions d'audit nécessitent d'arbitrer entre raffinements conceptuels, propriétés formelles, résolubilité et connaissance économétrique. Certains raffinements conceptuels font perdre certaines propriétés formelles ; par exemple considérer des interactions non triviales entre flots de différents arcs implique que les fonctions locales de temps de parcours sont irréductibles aux dérivées partielles d'une fonction de potentiel : alors le modèle d'affectation n'est plus solution d'un programme convexe de minimisation, et l'unicité n'est plus assurée.

Autre exemple d'arbitrage, entre résolubilité et connaissance économétrique : la propagation directe de l'erreur exogène via le mécanisme du modèle, utile pour quantifier l'incertitude économétrique attachée au résultat, suppose beaucoup de précision et de détail dans la résolution du modèle. Pour l'affectation cela implique de mobiliser un algorithme puissant, qui en général présente une forte complexité informatique.

Au total, le modèle alimenté repose sur une série d'arbitrages, de compromis. Cela explique des résistances récurrentes à l'innovation : si une innovation améliore l'un des deux aspects antagonistes mis en équilibre par un arbitrage communément admis, alors la détérioration de l'autre aspect fournit un argument contre l'innovation.

Troisième partie

***APPLICATION À
L'AFFECTATION
BICRITÈRE***

Chapitre 6

Plan d'audit de l'affectation bicritère

Après avoir défini des concepts et une méthode pour auditer un modèle, la troisième partie de thèse traite un exemple relatif à l'affectation du trafic. Chacun des quatre chapitres suivants opère l'audit dans une de ses dimensions : conceptuelle, formelle, algorithmique, économétrique. Le présent chapitre balise le terrain à cet effet.

D'abord nous montrons que l'affectation est un exemple représentatif des diverses natures d'erreur, en raison notamment de complexités mathématiques (section 6.1).

Puis nous montrons que l'audit doit s'exercer dans toutes les dimensions, en raison de l'imbrication des différentes natures d'erreur ; aussi l'investissement à consentir couvre plusieurs disciplines. Pour le rentabiliser, nous proposons de bénéficier d'économies d'envergure en l'effectuant pour un domaine, donc pour un ensemble de modèles particuliers (section 6.2).

Enfin nous précisons le mode opératoire des quatre chapitres suivants : une distinction entre les dimensions et un traitement sur plusieurs plans, ceux respectivement du domaine de l'affectation, du modèle bicritère et des applications particulières du modèle bicritère (section 6.3).

6.1 CARACTÉRISTIQUES DE L'AFFECTATION

Par rapport au modèle de régression linéaire (Cf. section 3.1) dont l'audit s'étend de façon non triviale dans les dimensions conceptuelle et économétrique seulement, l'affectation du trafic nécessite une attention particulière à l'égard des dimensions formelle et algorithmique (§6.1.1). Cela provient d'une complexité d'ordre mathématique qui dans les applications concrètes s'accompagne d'une lourdeur informatique liée à la dimension des problèmes mathématiques à résoudre (§6.1.2).

6.1.1 Toutes les sources d'erreur sont présentes dans l'affectation

Sur l'exemple d'un modèle d'affectation bicritère prix-temps (Cf. §4.1.2), nous avons déjà identifié des éléments relatifs à chacune des quatre natures d'erreur. Reprenons l'examen de façon plus générale.

6.1.1.1 Erreur de conception

Le codage et la composition conceptuelle déterminent l'erreur de conception du modèle d'affectation. Le codage en tant qu'ensemble d'approximations, de schématisations, dans la description physique du cas d'application. La composition conceptuelle en tant qu'ensemble de mécanismes explicatifs, plus ou moins pertinents par rapport à la représentation théorique des déplacements.

A titre d'exemple d'erreur de conception, mentionnons la spécification des fonctions de temps de parcours, supposées le plus souvent dépendre uniquement du débit local et pas d'autres flots (ex. ceux rencontrés au débouché de l'arc).

6.1.1.2 Erreur de formulation

La traduction des mécanismes explicatifs en langage formel mathématique permet d'examiner la cohérence logique (absence de contradiction) et le caractère opératoire : détermination du modèle (i.e. suffisamment de relations par rapport aux variables endogènes considérées), existence et unicité d'une solution. Pour auditer le caractère opératoire, il convient de rattacher les formules caractéristiques du modèle à une théorie mathématique, par exemple l'optimisation convexe, qui fournit des résultats généraux d'existence et d'unicité. Pour l'affectation, les principales théories mathématiques sont l'optimisation convexe, les problèmes de point fixe et les inéquations variationnelles.

Lorsque la composition conceptuelle incorpore des réactions mutuelles entre offre et demande de déplacement, il devient vite impossible d'exprimer directement la solution du modèle en fonction des variables explicatives ; il faut se contenter de résoudre numériquement la formule caractéristique.

6.1.1.3 Erreur de calcul

La résolution numérique d'une formule caractéristique d'affectation nécessite un algorithme itératif, dès qu'il y a des dépendances mutuelles entre certaines des variables endogènes (niveau de service et volume de demande, ou encore flots respectifs de deux itinéraires concurrents sur une même relation origine-destination).

La convergence d'un algorithme itératif n'est parfaite qu'exceptionnellement : dans une

application courante, on arrête prématurément les calculs (*troncature*). D'où évidemment la présence d'une erreur de calcul.

6.1.1.4 Erreur d'estimation et erreur exogène propagée

En matière d'affectation, nous avons identifié quatre objets d'estimation : paramètres du comportement des demandeurs, distribution spatiale des flots, fonctions de demande, fonctions d'offre. Chaque objet renvoie à une erreur d'estimation.

Dès qu'il y a transfert dans l'espace ou le temps, une erreur exogène survient, qu'il y a lieu de propager à travers le mécanisme du modèle en même temps que l'erreur d'estimation. Lorsque le modèle d'affectation incorpore un mécanisme d'équilibre (comme une dépendance mutuelle entre offre et demande de déplacement), la solution ne s'exprime pas analytiquement en fonction des variables exogènes, et les formules de propagation deviennent assez complexes (davantage que celles indiquées au §4.3.2). Leur obtention pose des exigences techniques particulières : elle nécessite en particulier d'explicitier les chemins, chose à laquelle se refusent la plupart des logiciels d'affectation actuels.

6.1.2 Deux complexités inhérentes

6.1.2.1 Une complexité d'ordre mathématique, liée à une caractérisation indirecte

Dans l'affectation, les variables endogènes (= expliquées) sont déterminées par un mécanisme d'équilibre, analogue à l'équilibre économique entre offre et demande. On les caractérise donc de façon indirecte, comme les solutions d'un système d'équations, non de façon directe comme le résultat d'une fonction connue appliquée à des arguments connus.

Formulons cela avec des notations mathématiques. Soit \mathbf{Y} le vecteur des variables endogènes, \mathbf{X} le vecteur des variables exogènes et \mathbf{F} la fonction vectorielle de dépendance. Une caractérisation directe est de forme $\mathbf{Y} = \mathbf{F}(\mathbf{X})$, tandis qu'une caractérisation indirecte est de forme $\mathbf{F}(\mathbf{X}; \mathbf{Y}) = \mathbf{0}$.

Une caractérisation soulève plusieurs questions :

- existe-t-il une solution ?
- une solution est-elle unique ?
- l'éventuelle solution est-elle stable (i.e. varie-t-elle peu si les variables exogènes \mathbf{X} varient faiblement) ?

Questions auxquelles on répond facilement dans le cas d'une caractérisation directe :

- l'existence et l'unicité sont immédiates.
- pour établir la stabilité, il suffit que la fonction \mathbf{F} jouisse de certaines propriétés ; en particulier la continuité suffit.

Dans le cas d'une caractérisation indirecte :

- pour assurer l'existence, il suffit que la fonction \mathbf{F} jouisse de certaines propriétés. Par exemple, si $\mathbf{F}(\mathbf{X}; \mathbf{Y}) = \mathbf{0}$ revient à un problème de point fixe de la même dimension que \mathbf{Y} et sans contradiction interne, la continuité de \mathbf{F} suffit.

- l'unicité nécessite des conditions beaucoup plus fortes. Par exemple que Y soit solution d'un programme convexe paramétré par X .
- pour établir a priori la stabilité, il faut également des conditions fortes.

Tout cela relève de l'audit formel, pour une étude a priori de l'existence, l'unicité et la stabilité. Pour le calcul pratique d'une solution, il faut disposer d'un algorithme non trivial, qui nécessite un audit algorithmique.

6.1.2.2 Une lourdeur informatique liée à la dimension du problème mathématique

Deuxième spécificité importante de l'affectation, la lourdeur des calculs. Le nombre d'opérations élémentaires dans la procédure de calcul varie selon le détail de la modélisation : détail dans le comportement des demandeurs (nombre de classes \bar{m}), dans le zonage du territoire (nombre de zones \bar{r}), dans le maillage (nombre d'arcs \bar{a}), dans la description de l'état final (nombre de chemins \bar{k} pour chaque triplet classe-origine-destination), détail enfin dans la précision des calculs (nombre d'itérations \bar{n}).

Grossièrement, la complexité informatique est proportionnelle au produit de ces facteurs, avec une puissance 2 pour le nombre de zones (si l'on traite individuellement chaque relation origine-destination) et une puissance γ légèrement supérieure à 1 pour le nombre d'arcs (rôle dans la recherche des plus courts chemins) :

$$C = K \bar{n} \bar{m} \bar{r}^2 \bar{a}^\gamma \bar{k}$$

Dans l'exemple d'une application concrète à un réseau comptant $\bar{a} = 5000$ arcs, à une demande spatialisée selon $\bar{r} = 500$ zones et segmentée selon $\bar{m} = 3$ classes, en supposant environ $\bar{k} = 5$ chemins par relation origine-destination et $\bar{n} = 30$ itérations, en adoptant $\gamma = 1$, on évalue :

$$C / K = 5 \cdot 10^{11} \text{ opérations}$$

ce qui n'est pas encore négligeable en l'état actuel des moyens informatiques accessibles à un bureau d'études.

Signalons que ce devis vaut pour des algorithmes qui tirent parti de la décomposition du problème par rapport aux relations origines-destinations et aux classes de demande (Cf. Leurent 1995d), faute de quoi il faudrait l'alourdir très considérablement.

6.2 UN INVESTISSEMENT LOURD, À RENTABILISER

Les deux spécificités évoquées induisent des relations particulières entre les dimensions d'audit : nous en avons fait une expérience personnelle (§6.2.1.1). Plus généralement les interactions entre dimensions incitent à mener l'audit dans toutes les dimensions (§6.2.1.2). L'investissement à consentir est lourd ! Pour l'amortir, le rentabiliser, nous proposons de le faire profiter à plusieurs modèles, afin de bénéficier d'économies d'envergure (§6.2.2).

6.2.1 Couvrir toutes les dimensions

6.2.1.1 Récit anecdotique

Evaluer un projet de voie urbaine rapide à péage. Mon propre intérêt pour l'affectation remonte à un stage effectué dans l'entreprise de bâtiment et travaux publics GTM, alors porteuse du concept LASER. Ce concept consiste en un réseau souterrain de voies rapides urbaines à péage, capable à la fois de contourner, irriguer et soulager une agglomération dense dont la voirie routière de surface subit une forte congestion. La première mouture de LASER concernait le sous-sol parisien et ses promoteurs proposaient un financement d'origine privé.

Un automobiliste pourrait ainsi choisir entre un chemin rapide mais tarifé, ou une route lente mais gratuite. Pour l'opérateur privé, le produit du péage sert à amortir les frais de construction. La rentabilité socio-économique d'une telle infrastructure tient à l'utilité pour ses utilisateurs comme pour les non-utilisateurs qui circulent sur une voirie délestée, ainsi qu'à l'effet d'entraînement de l'investissement.

Des méthodes de simulation incertaines. Pour l'investisseur privé, la rentabilité financière dépend crucialement des recettes, d'où un impératif : prévoir le trafic avec un maximum de fiabilité. Le standard français d'alors (1987) en matière d'affectation du trafic routier en milieu urbain était constitué par le logiciel DAVIS développé à l'INRETS, dans une version avec une classe de demande, une valeur du temps unique, des temps de parcours fonction du trafic mais un chargement incrémental du trafic (par tranches) sans garantie quant à un équilibre entre itinéraires concurrents simultanément chargés. GTM préféra recourir au logiciel d'affectation sur réseau routier interurbain de sa filiale SEEE, avec des chemins codés manuellement, des temps de parcours fixes, mais aussi une distribution continue de la valeur du temps et une segmentation selon le motif de déplacement. Ce choix relevait donc d'un audit conceptuel.

L'INRETS représenté par Roger Marche approuva ce choix, tandis que François Barbier Saint-Hilaire perfectionnait DAVIS afin de représenter plusieurs motifs, une distribution continue de la valeur du temps et un algorithme pour équilibrer les temps entre itinéraires concurrents. Depuis 1991, les différents projets français de type LASER ont été étudiés à l'aide des nouvelles versions de DAVIS, sauf à Marseille où EMME/2 a servi et sert encore à étudier le tunnel Prado-Carénage.

Cependant aucun audit formel ni algorithmique n'était appliqué à DAVIS. Dans son expertise du projet de bouclage à l'ouest de la deuxième rocade routière parisienne (A86), Roger Marche engagea un audit économétrique des recettes simulées, à un niveau agrégé qui éludait les aspects géographiques du trafic (Cf. §3.4.2 et §4.3.3). Cet audit économétrique reposait sur des données numériques obtenues à l'aide de DAVIS, donc il restait conditionnel à une résolution non auditée.

Une reconnaissance progressive des diverses dimensions d'audit. Surpris parfois des résultats obtenus en appliquant DAVIS, j'ai commencé par auditer formellement l'affectation bicritère prix-temps avec une distribution continue des valeurs du temps. J'ai proposé deux voies alternatives pour représenter l'élasticité du volume de demande au niveau de service, ainsi qu'un traitement de contraintes de capacité explicites : ce sont des contributions à la composition conceptuelle (synthèse dans Leurent, 1996a).

Cependant l'algorithme que je proposai d'abord (Leurent 1993a), une méthode des moyennes successives sans explicitation des chemins, nécessitait un fort nombre

d'itérations, sans garantie pratique quant à la précision du résultat obtenu (mais toutefois une garantie théorique a priori). Donc l'audit algorithmique était insuffisant. J'y ai remédié en explicitant les chemins dans la méthode des moyennes successives, puis en étendant l'algorithme d'égalisation par transvasement (Leurent, 1995a, c).

Ces perfectionnements techniques m'ont conduit à développer l'audit économétrique de l'affectation à valeur du temps continûment distribuée : d'abord en précisant la méthodologie (Leurent, 1994c) puis par des contributions ponctuelles, en particulier une méthode d'estimation par le maximum de vraisemblance pour la distribution des valeurs du temps (Leurent, 1995e).

6.2.1.2 Bilan : l'audit économétrique est subordonné aux audits formel et algorithmique

J'ai relaté mon propre itinéraire de recherche afin d'affirmer la *subordination* de l'audit économétrique aux audits formel et algorithmique. Il me semble que si l'on ne sait pas garantir le calcul d'une solution au modèle, alors il est vain de considérer le résultat du calcul comme le pivot d'une analyse économétrique.

Concluons à la stratification et l'enchaînement des tâches : un examen de l'erreur totale nécessite une étude dans toutes les dimensions. Le travail requis est multidisciplinaire, à base notamment d'analyse systémique, de recherche opérationnelle, d'algorithmique et d'économétrie. Les compétences nécessaires ont rarement été rassemblées par le passé¹.

Plus précisément, les étapes formelle et algorithmique ont constitué un véritable verrou technique, domaine réservé des spécialistes de recherche opérationnelle, évité des économètres. Conséquence logique, les modèles sont formulés et les algorithmes spécifiés en fonction de contraintes locales, d'objectifs particuliers (notamment d'économie des ressources informatiques), peu compatibles avec une approche globale de l'audit technique. Le choix, encore très largement suivi, de ne pas expliciter les chemins, empêche toute analyse de sensibilité et donc toute propagation d'erreur (2) ! Il obvie d'une autre manière à l'audit économétrique : comment ajuster une règle de répartition du trafic entre itinéraires concurrents, sans identifier ces itinéraires ?

6.2.2 Comment rentabiliser l'investissement ?

Pour auditer un modèle dans toutes les dimensions, l'investissement à consentir est lourd ! Afin de le rentabiliser, nous proposons de capitaliser tout ce qui est généralisable, pour faciliter l'audit d'autres modèles proches.

6.2.2.1 Capitaliser par des boîtes à concepts et des boîtes à outils

Dans un domaine de modélisation comme l'affectation du trafic, les modèles diffèrent les uns des autres de façon variable : dans leur composition conceptuelle, ou dans leur

¹en raison de concentrations géographiques des compétences : essentiellement en Amérique du Nord pour la recherche opérationnelle, et en Angleterre pour l'économétrie. Tandis que les contributions à la conception sont survenues plus ponctuellement, en France et en Allemagne autant qu'aux Etats-Unis et au Canada. Consulter les Actes des conférences ISTTT pour en juger.

²sauf par répétition de simulations. Mais alors il est difficile de faire la part de chaque source d'incertitude

formule caractéristique, ou dans leur algorithme, ou par leurs méthodes d'alimentation économétrique. Mais ils partagent la majeure partie de leurs principes, de leurs composants.

Les composants communs doivent être audités lorsque l'on traite un modèle particulier : il n'est pas nécessaire de réitérer leur audit pour traiter d'autres modèles du même domaine ! D'où l'idée d'engager l'audit sur le plan du domaine autant que sur le plan du modèle.

L'audit d'un domaine n'est autre que sa capitalisation : sélection des composants valables, détection et élimination des composants défectueux. Donc, comme l'audit, la capitalisation s'étend dans les quatre dimensions, conceptuelle, formelle, algorithmique, économétrique.

Dans chaque dimension nous proposons de capitaliser le domaine en inventoriant les composants et en les organisant dans des boîtes à concepts et/ou des boîtes à outils.

6.2.2.2 Pour transposer à d'autres modèles de trafic

Au long des quatre prochains chapitres, nous allons capitaliser et développer des concepts et méthodes techniques afin de les insérer dans une procédure d'audit. Relatives au cas particulier d'un modèle d'affectation bicritère, ils seront facilement transposables à bien d'autres modèles de trafic. Pour l'affectation en tant que domaine, ils baliseront le terrain, ils constitueront une trame où insérer des compléments d'investigation.

Réduction des modèles de choix de déplacement à des modèles d'affectation. Bien que l'affectation à proprement parler ne constitue qu'une étape parmi les quatre du schéma classique (Cf. §1.1.2), il est souvent possible de formuler et traiter un modèle à plusieurs choix de déplacement, comme un modèle d'affectation. Parmi les combinaisons possibles : destination et itinéraire (Murchland 1966), mode et itinéraire (Florian 1977), destination et mode et itinéraire (LeBlanc et al. 1982), génération et distribution et itinéraire (Safwat et Magnanti 1988).

Une technique, pour transformer un modèle avec plusieurs choix de transport en un modèle d'affectation, consiste à ajouter des arcs fictifs aux arcs basiques du réseau de manière à représenter la structure morphologique du modèle avec plusieurs choix. On désigne cette opération comme la constitution d'un super-réseau (Sheffi et Daganzo 1978).

La modélisation sous forme de super-réseau convient si l'on traite une seule classe de demande, ou plusieurs classes qui n'interagissent pas de façon directe. Mais il n'existe pas actuellement de schéma pour traiter ainsi les modèles avec des transferts entre classes de demande (comme dans MATISSE).

Complexités de l'affectation en général par rapport au modèle bicritère examiné. Certains modèles d'affectation présentent des caractères qui ne figurent pas dans le modèle traité au long des quatre prochains chapitres : ainsi le caractère dynamique, les spécificités des réseaux de transport en commun, certaines interactions entre segments de demande.

Par caractère dynamique, nous entendons l'évolution des phénomènes de circulation d'un instant à l'autre. Cela permet de considérer des files d'attente (éventuellement vides), pour représenter la saturation locale, et si nécessaire de propager la congestion de l'aval

vers l'amont.

Parmi les spécificités des réseaux de transport en commun, citons l'attente des véhicules aux stations, la possibilité pour l'auteur du déplacement de choisir entre plusieurs lignes localement concurrentes celle qui est desservie la première... Des artifices techniques (Spiess, 1985 ; Cascetta et Nguyen, 1988) permettent de traduire ces mécanismes comme des éléments du choix d'un hyper-chemin plutôt que d'un chemin. Alors tout se déroule comme dans un modèle d'affectation routière, à de simples nuances près.

Les interactions entre segments de demande constituent une très prometteuse voie de développement pour la modélisation du trafic : les transferts d'effectif entre segments permettent un traitement cohérent des effets d'induction (MATISSE). L'explicitation des interactions sur le réseau entre plusieurs types de véhicules (ex. voitures et poids lourds) reste une question ouverte, encore très peu défrichée.

Le caractère dynamique comme les interactions entre segments de demande nécessitent des actions spécifiques d'audit, qui ne sont pas menées dans notre thèse. Mais les spécificités des réseaux de transport en commun ne soulèvent pas de difficulté particulière.

Au total, par rapport aux modèles de trafic courants, auditer un modèle d'affectation routière statique n'esquive vraiment que le caractère dynamique. Rappelons que celui-ci n'est actuellement traité que dans des applications spéciales ⁽¹⁾ : pour simuler le trafic routier dans des zones urbaines restreintes ou à des barrières de péage.

6.3 DÉTAIL DU MODE OPÉRATOIRE

Prévenons d'abord de notre situation par rapport à l'audit technique : nous trouvant dans un laboratoire de recherche, nous envisageons chacune des dimensions identifiées. Dans une application concrète, un bureau d'étude peut difficilement faire de même : sa responsabilité en la matière nous paraît réduite aux dimensions conceptuelle et économétrique.

6.3.1 Un audit dans chaque dimension

6.3.1.1 Une première approche...

Pour auditer le modèle d'affectation bicritère, notre ambition se limite à mener une première approche, en suivant l'ordonnancement conceptuel-formel-algorithmique-économétrique. Ce cas d'application suffit pour illustrer les aspects concrets de l'audit ; notre traitement conduit, explicitement ou non, à des arbitrages particuliers entre les dimensions.

Nous ne traitons pas toutes les questions liées aux interactions entre les dimensions d'audit ; la thèse laisse en suspens bien d'autres interrogations plus concrètes, par exemple l'incertitude liée aux fonctions de demande. Cependant la thèse aborde chacune des quatre dimensions ; elle indique les principes, les concepts, les méthodes, disponibles ou souhaitables. Ainsi elle effectue un premier défrichage du terrain.

¹encore très rares en France, mais moins en Angleterre ou aux Etats-Unis

6.3.1.2 ... nécessairement pluridisciplinaire

Nous l'avons déjà mentionné, les dimensions d'audit font appel à des méthodes et outils qui relèvent de disciplines variées. Dans le cas de l'affectation du trafic, les enjeux concrets conduisent à rattacher l'ensemble à l'économie des transports : car la modélisation du trafic remplit des fonctions de planification et d'exploitation, de même qu'elle contribue à établir des coûts de transport pour différents acteurs économiques.

6.3.2 Une entreprise à plusieurs niveaux

La méthode d'audit du chapitre 5 vaut pour tout modèle. Nous allons la particulariser progressivement, en envisageant plusieurs niveaux :

- celui du domaine de l'affectation, dont nous opérerons une capitalisation.
- le niveau du modèle bicritère, considéré sans référence à un cas particulier d'application. La tâche a été préparée au chapitre 5, grâce à une référence continue au réseau à deux arcs : ainsi dégagée de toute complexité d'abondance, l'attention s'est concentrée sur les concepts et outils techniques à employer dans l'audit.
- le niveau des applications du modèle bicritère. Nous traiterons un exemple de taille moyenne (en milieu urbain, 140 zones de demande et réseau de 2000 arcs) lors de l'audit algorithmique. Pour l'audit économétrique, nous traiterons un exemple de petite taille relatif à une liaison routière interurbaine (entre Bordeaux et Pau) étudiée originellement par le CETE du Sud-Ouest.

6.3.3 Plan de la suite de la partie

Nous nous limiterons à une première approche, en envisageant successivement chaque dimension et en considérant seulement les interactions entre les dimensions provoquées par l'antériorité naturelle de la conception à la formulation, de la formulation à la résolution, du modèle calculable à l'alimentation économétrique.

Pour chacune des dimensions d'audit, nous proposerons d'abord une capitalisation de concepts et d'outils pour l'affectation stationnaire du trafic routier, puis nous étudierons le modèle bicritère.

6.3.3.1 Dimension conceptuelle (chapitre 7)

Pour auditer la composition conceptuelle du modèle bicritère, nous examinerons un par un les composants conceptuels de l'affectation statique du trafic routier. Nous pourrions ainsi apprécier le modèle bicritère en le comparant aux ressources du domaine.

Quant au codage, nous le déplaçons vers l'audit économétrique, afin de quantifier l'incertitude qui lui est liée et de la prendre en compte au niveau du résultat du modèle.

6.3.3.2 Dimension formelle (chapitre 8)

Pour l'audit formel, c'est-à-dire la validation de la formulation mathématique et l'étude des propriétés d'existence, unicité et stabilité, notre analyse se bornera à la validation de la formulation mathématique et à l'étude de l'existence. Elle sera incomplète pour l'unicité et la stabilité : nous nous limiterons au cas favorable qui présente une unicité. Un traitement systématique de la non unicité reste à concevoir, au moins dans le cas des modèles de déplacement.

6.3.3.3 Dimension algorithmique (chapitre 9)

Notre audit algorithmique suffira par rapport au cas favorable résultant de l'audit formel. Mais il faudrait l'étendre au cas sans unicité : une telle extension est subordonnée à un audit formel préalable du cas sans unicité.

6.3.3.4 Dimension économétrique (chapitre 10)

Nous avons proposé un schéma général pour l'audit économétrique : quantifier l'incertitude exogène a priori, propager cette incertitude via le modèle (Cf. section 5.4). Ce schéma vaut quel que soit le modèle ; il est applicable sous forme de modules distincts, relatifs soit à la quantification de l'incertitude a priori liée aux variables exogènes d'un certain type, soit à la propagation via le modèle.

Pour notre modèle d'affectation bicritère, nous présenterons trois modules :

- quantification de l'incertitude exogène sur les paramètres comportementaux (dont la valeur du temps).
- quantification de l'incertitude exogène sur les flots origine-destination.
- propagation à travers le modèle de l'incertitude liée à toutes les variables exogènes.

Il manque donc deux modules, pour quantifier l'incertitude exogène liée respectivement aux fonctions de demande et aux fonctions de temps de parcours.

Concernant la réduction d'un modèle de déplacement à un modèle d'affectation, il manque encore d'autres modules, liés aux choix de transport que l'on décide de représenter par des arcs fictifs dans un super-réseau.

Chapitre 7

L'audit sémantique : un diagnostic

L'audit sémantique, ou conceptuel, consiste à détecter et mesurer, éventuellement réduire, l'erreur liée à la composition conceptuelle du modèle. Il s'agit d'explicitier systématiquement les hypothèses de représentation des phénomènes, afin d'en apprécier la portée explicative.

La méthode proposée au chapitre 5 comprend une partie générique, pour inventorier les composants sémantiques relatifs à l'objet d'étude et hiérarchiser les solutions de représentation, et une partie particulière, pour analyser un modèle particulier et le situer par rapport aux solutions disponibles.

Nous appliquons la méthode à l'affectation statique du trafic routier, en recensant et classant les composants conceptuels du domaine (section 7.1), puis en analysant à titre de cas particulier le modèle bicritère (section 7.2).

Les hypothèses de représentation conceptuelle confèrent une certaine portée explicative au modèle, donc une certaine pertinence ; le désir de pertinence est bordé par un souci de proportionnement (section 7.3).

7.1 COMPOSANTS CONCEPTUELS POUR L'AFFECTATION

Afin de détecter et mesurer qualitativement l'erreur de conception dans un modèle d'affectation statique du trafic routier, nous allons recenser les phénomènes relatifs à l'objet d'étude. Pour chacun d'eux, nous citerons les principales possibilités de représentation, nous les hiérarchiserons dans une échelle de valeurs sur laquelle on peut situer chaque modèle particulier.

L'affectation est classiquement conçue comme un choix économique, celui du bien "acheminement par un réseau de transport". Chaque itinéraire est un service offert ; chaque auteur de déplacement est un consommateur, un demandeur. Aussi présenterons-nous les principaux composants conceptuels de l'affectation statique dans l'ordre que voici¹ : hypothèses physiques (§7.1.1), coût généralisé (§7.1.2), composants relatifs à l'offre (§7.1.3) ou à la demande (§7.1.4). Enfin nous récapitulerons l'ensemble (§7.1.5).

7.1.1 Hypothèses physiques

Selon quelles dimensions le modèle permet-il d'analyser le système ? Je qualifie de "physiques" les hypothèses relatives aux dimensions spatiales et temporelles du problème, ainsi que certaines hypothèses d'agrégation économétrique.

7.1.1.1 Caractère spatial

Le réseau de transport tout comme les origines et destinations des déplacements s'inscrivent dans l'espace. Prendre en compte le caractère spatial amène à segmenter le marché du bien "acheminement" selon l'origine et la destination.

Pour l'audit d'un modèle particulier, il convient de préciser comment est menée la spatialisation, ce que recouvre la notion de zone d'origine ou de destination. Sont notamment à examiner :

- la couverture du réseau "réel"² par le réseau codé ; les itinéraires "susceptibles d'être suivis dans la réalité" sont-ils représentés ?
- le proportionnement des zones de demande au réseau : les zones de demande doivent être suffisamment fines pour qu'une scission de l'une d'elles n'entraîne que de petites modifications locales quant aux itinéraires choisis sur le réseau.

7.1.1.2 Temporalité

La dimension temporelle est présente, explicitement ou non, à trois endroits dans l'affectation statique.

Primo le caractère stationnaire. Cela suppose une certaine homogénéité des conditions de trafic durant la période considérée, hypothèse à valider dans une application concrète. Nous y revenons plus loin au sujet du fonctionnement de l'offre (Cf. §7.1.3.2).

¹dans la référence Leurent 1995d, le même ordre sert pour introduire progressivement les différentes notions liées à l'affectation statique du trafic routier

²même si en situation hypothétique

Secundo la période de référence de l'affectation : représente-t-on un jour moyen, ou un jour ouvrable moyen, ou une heure moyenne, ou une heure de pointe moyenne, ou autre chose encore ? Autrement dit, à quel ensemble d'éléments la moyenne fait-elle référence ? Dans cet ensemble, quelle dispersion constate-t-on autour de la moyenne (variabilité temporelle) ?

Tertio l'époque d'application. Quels sont les divers horizons temporels considérés, quelles évolutions suppose-t-on entre des dates successives d'application ?

7.1.1.3 Agrégation économétrique

Les hypothèses spatiales donnent lieu à des agrégations dans l'espace, portant sur les lieux d'origine et de destination, ou sur les itinéraires regroupés en faisceaux ou découpés en sections réputées homogènes. Les hypothèses temporelles donnent lieu à des agrégations dans le temps, portant sur la chronologie des déplacements. De même les comportements donnent lieu à des agrégations spécifiques, sous forme de segments (ou classes) de demande.

Les agrégations concernent les déplacements, sauf celles portant sur les itinéraires. Une hypothèse économétrique fondamentale consiste à considérer le déplacement comme unité statistique, et un ensemble de déplacements comme une variable scalaire positive ou nulle.

7.1.2 Notion de coût généralisé

Le coût généralisé d'un itinéraire pour un segment de demande est un concept opératoire fondamental en matière d'affectation du trafic. Le coût généralisé résume en une seule mesure les facteurs qui opposent une résistance au déplacement. Il combine plusieurs coûts de natures diverses, comme les frais monétaires, les temps passés, les distances parcourues, les nombres de rupture de charge, les dangers perçus, les inconforts et pénibilités... En négatif de ces coûts s'inscrivent des avantages, par exemple la qualité de signalisation, l'attrait esthétique ou les potentialités gastronomiques.

Pour un segment de demande donné, l'itinéraire choisi est celui qui présente le coût généralisé minimal. Pour un itinéraire donné, le coût généralisé est très généralement évalué comme la somme des coûts généralisés des arcs qui composent l'itinéraire.

7.1.2.1 Complétude du coût généralisé

Quels facteurs contribuent explicitement au coût généralisé ? Une exigence minimale est d'incorporer le prix et le temps de parcours. Il est bon de considérer aussi la distance, le confort, la signalisation, la fiabilité.

7.1.2.2 Adéquation du coût généralisé

Les attributs des itinéraires, mesurés chacun dans son unité propre, sont agrégés dans le coût généralisé après conversion dans une unité de compte de référence.

Pour un attribut donné, le facteur de conversion est l'équivalent d'une unité propre de l'attribut en unités de compte de référence : ainsi, pour agréger le temps de parcours au prix, on le multiplie par une "valeur du temps" propre au segment de demande.

Les facteurs de conversion fixés représentent-ils effectivement les préférences des auteurs des déplacements ? La réponse à cette question relève de l'audit économétrique,

mais des choix de conception en laissent préjuger : selon que le modèle inclut ou non une segmentation vis-à-vis de tel ou tel facteur de différenciation, il est plus ou moins apte à représenter la diversité des préférences des auteurs des déplacements, diversité dont témoigne par essence l'emploi concurrent de plusieurs itinéraires sur une même relation origine-destination.

7.1.3 Du côté de l'offre

Au sujet de l'offre d'acheminement, nous avons déjà présenté des hypothèses de spatialisation et de description par des attributs attachés aux arcs du réseau.

7.1.3.1 Les itinéraires potentiels

A quels chemins les déplacements ont-ils accès ? Quelle est leur définition théorique a priori ? En général, on considère des chemins sans boucles, ce qui empêche de représenter certains "errements" dûs à une insuffisante connaissance du réseau par les conducteurs¹.

Dans le modèle déterministe homogène (de Beckmann), tous les chemins sans boucle sont réputés accessibles : seuls les plus courts peuvent recevoir une charge, ce qui limite les besoins d'énumération. Mais dans un modèle stochastique susceptible de charger des chemins au coût non minimal, comment restreindre les chemins à considérer (vu que les chemins sans boucle sont en nombre combinatoire) ? Il faut définir une règle de restriction.

7.1.3.2 Représenter la congestion

La congestion routière, encombrement par accumulation de véhicules, se manifeste par une circulation processionnaire ou par bords en section courante du réseau, ainsi que par des files d'attente aux points de singularité de la section (ex. incident en bord de route) ou aux intersections. Elle dégrade les temps de parcours et le confort de conduite. Elle peut bloquer une infrastructure et se propager en amont, jusqu'à bloquer un sous-réseau par remontée de files entre les carrefours.

La congestion s'explique par la relation entre la charge de trafic et la capacité d'écoulement de l'infrastructure. Si la charge est faible relativement à la capacité, la circulation est fluide. Si la charge est suffisamment forte pour approcher la capacité, ou si la capacité est réduite par un aléa, la congestion fait sentir ses effets.

On ne sait pas représenter explicitement (et rigoureusement) une file d'attente dans un modèle statique. On ne sait guère tenir compte des perturbations aléatoires de la capacité d'une route.

On peut cependant supposer que le temps de parcours dépend du trafic local, que l'écoulement local est borné par une capacité limite, que le temps d'un arc dépend aussi des trafics d'arcs voisins (le sens opposé de circulation, ou les arcs croisés en sortie).

¹l'absence de boucle peut signifier soit ne pas passer deux fois par le même sommet, soit ne pas passer deux fois par le même arc unidirectionnel

7.1.4 Du côté de la demande

Du côté de la demande, les principales hypothèses tiennent (1) à l'élasticité du volume à la désutilité du service, (2) à l'information et à la rationalité des conducteurs, (3) à la différenciation des conducteurs. Elles interagissent les unes avec les autres.

7.1.4.1 Elasticité du volume à la désutilité du service

Sur une relation origine-destination, les coûts généralisés des itinéraires concurrents procurent un service moyen d'une certaine qualité. En général, sur un marché économique libre, le volume de la demande dépend du prix et de la qualité de service. On sait représenter cette dépendance par une fonction de demande, qui relie le volume de demande au niveau de service.

De quoi d'autre dépend la fonction de demande d'une relation origine-destination ? Sans doute aussi des concurrences modales, des substitutions entre destinations etc. Il est bon d'explicitier les autres dépendances, ce qui signifie modéliser non seulement le choix d'itinéraire mais encore d'autres choix, de mode, de destination etc.

7.1.4.2 Information et rationalité des conducteurs

Quelle connaissance un conducteur a-t-il des conditions de circulation locales, des itinéraires possibles, des caractères des itinéraires ? Sur la base de quelle information, et avec quelle rationalité, choisit-il son itinéraire ?

Dans un modèle d'affectation, l'information et la rationalité sont mesurées, rationnées par l'analyste :

- lors de l'identification des chemins pour la connaissance "topologique" des itinéraires possibles.
- à travers l'expression du coût généralisé (facteurs de conversion entre attributs de natures différentes comme le prix et le temps, inclusion ou non d'un terme aléatoire) pour la connaissance des caractères des itinéraires et des conditions de circulation locales.
- même moyen technique pour la rationalité : l'analyste peut différencier les facteurs de conversion par classes d'individus afin de spécifier des comportements spécifiques, des rationalités locales. L'inclusion d'un terme aléatoire dans le coût généralisé permet de "réduire" la rationalité de l'auteur du choix de déplacement.

7.1.4.3 Différencier les demandeurs

Différencier les demandeurs, i.e. distinguer plusieurs classes de demande, sert à considérer et analyser spécifiquement une partie de la demande, notamment si cette classe manifeste des caractères ou comportements spécifiques.

La différenciation fondamentale est d'ordre spatiale : le modèle distingue les déplacements selon leur origine et leur destination. Mentionnons aussi la différenciation par type de véhicule, par exemple entre voiture et poids lourd, car les comportements routiers et les motifs de déplacement diffèrent. Ou encore en fonction de la perception du coût généralisé (donc de façon privilégiée par rapport aux facteurs de conversion comme la valeur du temps).

7.1.4.4 Quelle cohérence d'ensemble ?

Nous avons mentionné une interaction entre l'hypothèse d'information-rationalité et l'hypothèse de différenciation. D'autres interactions existent entre la notion de désutilité du service et l'hypothèse de différenciation (une désutilité par segment de demande, à agréger de quelle manière ?), entre la formulation de la désutilité et l'hypothèse d'information-rationalité (si l'on incorpore un aléa au coût généralisé).

Les composants de modélisation doivent manifester une cohérence d'ensemble. Sinon, quel sens garde le modèle, comment lui accorder le moindre crédit ?

7.1.5 Des échelles de valeur pour une synthèse

Au terme de cette revue systématique (complétée dans Leurent 1995d), nous avons recensé divers phénomènes et, pour chacun d'eux, diverses solutions de représentation. Certaines solutions valent mieux que d'autres, les valeurs se situent dans une échelle orientée. Le tableau joint récapitule les phénomènes et les échelles associées.

Tab. 7A. Phénomènes et solutions de représentation.

	Phénomène ou hypothèse	Echelle de valeur
Spatialité	"Réalisme" du réseau codé	mauvais / moyen / bon
	Segmentation du réseau	mauvais / moyen / bon
	Proportionnement des zones au réseau	mauvais / moyen / bon
Tempo- ralité	Hypothèse de stationnarité	indue / admissible / plausible
	Homogénéité par rapport à la moyenne	mauvais / moyen / bon
Offre	Itinéraires candidats au chargement	mal définis / bien définis
	Attributs pour décrire les routes	un seul / temps et prix / davantage
	Représentation de la congestion	temps local fonction du flot local / contraintes explicites de capacité
	Différenciation des types de véhicules	rien / via des coefficients d'équivalence entre classes / temps particularisés par type
Demande	Elasticité	non / oui + si effets entre segments
	Information et rationalité	totale / partielle + si selon segment
	Différenciation	non / uniquement stochastique / uniquement déterministe / déterministe et stochastique

Chaque ligne correspond à un certain mécanisme explicatif : les solutions disponibles pour représenter ce mécanisme, séparées par des caractères "/", sont classées par valeur croissante.

7.2 EXEMPLE DU MODÈLE D'AFFECTATION BICRITÈRE

Engageons à présent l'audit conceptuel d'un modèle d'affectation statique particulier, le modèle bicritère à demande variable, temps variables et contraintes adjointes. Nous allons préciser les hypothèses conceptuelles et les apprécier par rapport aux ressources sémantiques générales du domaine.

7.2.1 Hypothèses physiques

7.2.1.1 Caractère spatial

On découpe le territoire d'une part en zones d'origine des déplacements et d'autre part en zones de destination. On indexe les \bar{r} zones d'origine par l'identifiant r et les \bar{s} zones de destination par l'identifiant s .

On schématise le réseau par \bar{a} arcs unidirectionnels identifiés par a . Un chemin k entre une origine r et une destination s , astreint à passer au plus une fois par chaque arc, est défini par la donnée d'indicateurs $\delta_{rs}^{ak} = 1$ si l'arc a appartient à k ou 0 sinon.

7.2.1.2 Caractère temporel

Entièrement fonction du cas particulier d'application.

7.2.1.3 Agrégation économétrique

On note f_{rs}^k le flot, volume de déplacements, acheminé par l'itinéraire k depuis l'origine r jusqu'à la destination s . On considère sur chaque relation origine-destination r - s le flot total $q_{rs} = \sum_k f_{rs}^k$.

On considère pour chaque arc a le flot local $x_a = \sum_{rsk} \delta_{rs}^{ak} f_{rs}^k$.

7.2.2 Hypothèses sur le coût généralisé

7.2.2.1 Deux critères principaux, le temps et le prix

Le temps généralisé¹ d'un itinéraire k pour un déplacement de valeur du temps v suit l'expression $G_{rs}^k(v) = T_{rs}^k + P_{rs}^k / v$, avec T_{rs}^k le temps de parcours et P_{rs}^k le prix. Il n'y a pas de terme aléatoire².

Ces deux critères de temps et de prix sont qualifiés de principaux car ils constituent les deux termes de l'arbitrage par rapport auquel les demandeurs sont différenciés, en fonction de la valeur du temps. On pourrait considérer d'autres couples de critères, formés à partir du temps (imposé pour représenter les principaux aspects de la congestion) et d'un autre critère que le prix, par exemple la distance.

¹ici on prend pour unité de compte de référence le temps, plutôt que le prix. Cela suffit de factoriser le temps généralisé par la valeur du temps pour obtenir le coût généralisé

²cette hypothèse peut être levée, Cf. Leurent 1996a

7.2.2.2 Quant aux autres attributs

D'autres attributs liés aux routes, tels que distance, confort, fiabilité etc, peuvent être incorporés avec une pondération particulière¹ soit au prix P_{rs}^k , soit au temps T_{rs}^k . De même le temps pur peut lui-même se décomposer en plusieurs éléments : attente à un noeud, parcours en section courante hors congestion, supplément de temps en section courante imputable à la congestion, etc, à pondérer chacune par un coefficient particulier.

7.2.3 Représentation de l'offre

7.2.3.1 Itinéraires candidats au chargement

A priori tout chemin sans boucle peut recevoir un chargement, sauf contrainte particulière fixée par l'analyste (interdiction de certains mouvements tournants, restriction d'accès de certains arcs pour certaines classes de demande).

En pratique seuls les chemins efficaces, i.e. tels qu'il existe une valeur du temps à laquelle ils procurent le temps généralisé minimum de l'origine-destination, peuvent recevoir du flot.

7.2.3.2 Description des arcs par des attributs

Entièrement fonction du cas particulier d'application, avec pour l'analyste la responsabilité d'incorporer un attribut donné soit au prix, soit au temps. On obtient ainsi, pour un arc a , un temps T_a et un prix p_a . Notons qu'il y a parfois lieu de calculer le prix d'un itinéraire autrement qu'en décomposant l'itinéraire en arcs, par exemple dans le cas d'un péage perçu selon un schéma en cordon origine-destination.

7.2.3.3 Représentation de la congestion

La prise en compte "minimale" consiste à relier sur chaque arc a le temps de parcours T_a au trafic local x_a : soit $T_a = t_a(x_a)$ avec t_a une fonction de x_a (a priori croissante).

On peut aussi contraindre l'écoulement local x_a par une capacité limite L_a : soit $x_a \leq L_a$. Cela conduit à introduire une variable auxiliaire (²) w_a , nulle si $x_a < L_a$ et non négative sinon, à ajouter au temps local selon $T_a = t_a(x_a) + w_a$: on l'interprète comme un temps d'attente subi à l'entrée de l'arc a en cas de saturation de la capacité.

Pour sophistication davantage, on peut supposer plus généralement que le temps local dépend des trafics de plusieurs arcs, soit $T_a = t_a(\mathbf{f})$ avec $\mathbf{f} = \{f_{rs}^k\}_{rsk}$, et que des contraintes $z_b(\mathbf{f}) \leq 0$ traduisent des conditions de fonctionnement du réseau (b identifie une contrainte particulière fixée par l'analyste).

¹ constante pour chaque attribut, sauf à représenter plusieurs classes de demande

² multiplicateur de Lagrange associé à la contrainte

7.2.4 Représentation de la demande

7.2.4.1 Elasticité du volume de demande au niveau de service

Compte tenu de l'hypothèse d'un temps généralisé déterministe, le niveau de service sur la relation origine-destination r - s vaut, pour un déplacement de valeur du temps v :

$$S_{rs}(v) = \min_k G_{rs}^k(v).$$

Comment faire dépendre le volume de demande de ce niveau de service ? La solution rigoureuse est de considérer une fonction de demande par valeur du temps : soit

$$q_{rs}(v) = D_{rs}^{[u]}(S_{rs}(v)).$$

Une solution moins rigoureuse économiquement mais plus commode est de considérer un niveau de service en moyenne par rapport aux valeurs du temps : $\bar{S}_{rs} = \int S_{rs}(v) dH_{rs}(v)$, si $H_{rs}(v)$ désigne la fonction de répartition¹ des valeurs du temps sur la relation origine-destination r - s . On peut alors spécifier $q_{rs} = D_{rs}(\bar{S}_{rs})$. Si les $D_{rs}^{[v]}$ sont identiques et égales à une fonction affine $D_{rs}^{[v]}(\xi) = \alpha_{rs}\xi + \beta_{rs}$, alors les deux solutions se confondent.

7.2.4.2 Information et rationalité

On suppose que chaque auteur de déplacement dispose d'une information totale sur les itinéraires accessibles et leurs caractères (prix et temps), et qu'il est parfaitement rationnel (donc choisit un chemin optimal conditionnellement à sa valeur du temps).

7.2.4.3 Différenciation

Nous avons déjà mentionné la différenciation selon la valeur du temps : le fondement du modèle bicritère est de traiter comme une seule classe des déplacements qui présentent une dispersion des valeurs du temps.

Pour différencier les déplacements vis-à-vis de critères supplémentaires, il est possible de distinguer plusieurs classes de déplacements, chacune d'elles avec sa propre distribution de la valeur du temps. Pour cela, il suffit de segmenter la demande selon \bar{u} classes de véhicules, identifiées par u . Alors les itinéraires k dépendent à la fois de la relation origine-destination² et de la classe³, d'où des flots par chemin f_{urs}^k et des indicateurs δ_{urs}^{ak} pour l'appartenance des arcs aux chemins, des temps généralisés par classe et par valeur du temps $G_{urs}^k(v)$. Les flots par relation origine-destination deviennent aussi particuliers à la classe : q_{urs} . Sur les arcs, il faut individualiser pour chaque classe les flots x_{ua} , les temps T_{ua} et les prix p_{ua} .

¹i.e. la proportion des déplacements de l'origine-destination avec une valeur du temps comprise entre a et $b \geq a$ vaut $H_{rs}(b) - H_{rs}(a)$

²d'ailleurs les découpages en zones d'origine et en zones de destination peuvent aussi dépendre de la classe

³notamment s'il y a des contraintes particulières de gabarit

7.2.4.4 Cohérence d'ensemble

Nous avons défini un niveau de service par origine-destination et par valeur du temps comme le minimum des temps généralisés faisables pour cette valeur du temps, donc en cohérence avec l'hypothèse de rationalité économique et d'information totale. De même, les solutions proposées pour la fonction de demande respectent la différenciation selon la valeur du temps.

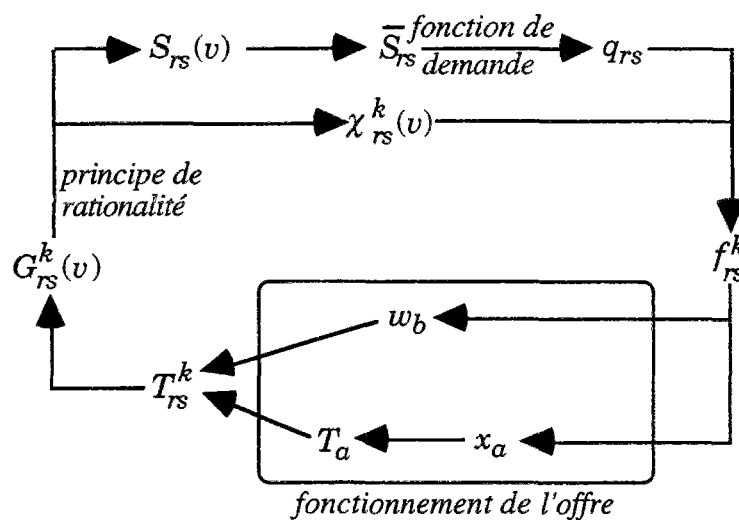
7.2.5 Exploration systématique

Après cette présentation progressive des hypothèses du modèle bicritère, appliquons l'audit conceptuel afin d'apprécier la portée explicative de ce modèle d'affectation particulier.

7.2.5.1 Analyse : variables exogènes ou endogènes, relations de dépendance

Le schéma 7a récapitule les dépendances entre les variables endogènes ou exogènes. Certaines dépendances forment une boucle : cela révèle le mécanisme d'équilibre entre offre et demande.

Fig. 7a. Mécanisme d'ensemble du modèle bicritère.



Les variables exogènes constituent un lot d'analyse physique et comportementale du problème (tableau 7B) et un autre de spécification fonctionnelle (tableau 7C).

Tab. 7B. Variables exogènes d'identification physique et comportementale.

Variable	Type	Sens
a	identifiant	un arc
r	identifiant	une zone d'origine
s	identifiant	une zone de destination
k	identifiant	un chemin (fixé par le modélisateur)
v	réel ≥ 0	une valeur du temps
u	identifiant	une classe
b	identifiant	une contrainte adjointe

Tab. 7C. Variables exogènes de spécification fonctionnelle.

Variable	Type	Sens
δ_{rs}^{ak}	indicateur $\in \{0;1\}$	définition de l'itinéraire à partir des arcs
t_a	fonction réelle	temps de parcours de l'arc a
p_{rs}^k	réel	frais monétaires sur le chemin k
z_b	fonction réelle	contrainte adjointe
D_{rs}	fonction réelle	fonction de demande de l'O-D $r-s$
H_{rs}	fonction réelle	répartition des valeurs du temps de l'O-D $r-s$

Recensons maintenant les variables endogènes (tableau 7D).

Tab. 7D. Recensement des variables endogènes.

Variable	Type	Sens
T_a	réel	frais temporels sur l'arc a
T_{rs}^k	réel	frais temporels sur le chemin k
w_b	réel ≥ 0	pénalité temporelle liée à la contrainte adjointe $z_b(\mathbf{f}) \leq 0$
$G_{rs}^k(v)$	réel	temps généralisé du chemin k pour la VdT v
$S_{rs}(v)$	réel	niveau de service sur l'O-D $r-s$ pour la VdT v
\bar{S}_{rs}	réel	niveau de service en moyenne des VdT sur l'O-D $r-s$
q_{rs}	réel	flot total de l'O-D $r-s$
$\chi_{rs}^k(v)$	fonction indicatrice	emploi du chemin k par les déplacements de VdT v =1 pour un seul chemin k qui procure un temps généralisé minimum à v , et 0 pour les autres chemins
f_{rs}^k	réel	flot sur le chemin k
x_a	réel	flot local sur l'arc a

Recensons enfin les relations de dépendance (tableau 7E).

Signalons que certaines variables, exogènes ou endogènes, sont des fonctions au stade de la conception. C'est au stade économétrique que l'on décide d'une spécification fonctionnelle particulière, combinant des fonctions classiques et des paramètres particuliers qui seuls subsistent comme variables exogènes (l'incertitude relative à la spécification fonctionnelle particulière comprend donc deux postes : l'un d'incertitude sur les paramètres scalaires, l'autre d'incertitude sur l'emploi des fonctions classiques et qui participe à l'incertitude résiduelle).

Tab. 7E. Relations de dépendance.

Sens	Produit (s)	Formule(s)	Entrée(s)	
			EXO.	ENDO.
Formation du temps, <i>congestion</i>	T_{rs}^k w_b	$T_{rs}^k = \{\sum_a \delta_{rs}^{ak} t_a\} + \sum_b w_b \frac{\partial z_b}{\partial f_{rs}^k}$ obtention complexe pour w_b ¹	t_a, z_b	w_b, f_{rs}^k
Formation du temps généralisé <i>coût généralisé</i>	$G_{rs}^k(v)$	$G_{rs}^k(v) = T_{rs}^k + P_{rs}^k / v$	P_{rs}^k, v	T_{rs}^k
Minimisation du temps généralisé <i>rationalité</i>	$S_{rs}(v)$ $\chi_{rs}^k(v)$	$S_{rs}(v) = \min_k G_{rs}^k(v)$ $G_{rs}^k(v) > S_{rs}(v) \Rightarrow \chi_{rs}^k(v) = 0$		$G_{rs}^k(v)$
Niveau de service <i>différenciation</i>	\bar{S}_{rs}	$\bar{S}_{rs} = \int S_{rs}(v) dH_{rs}(v)$	$H_{rs}(v)$	$S_{rs}(v)$
Fonction de demande <i>élasticité</i>	q_{rs}	$q_{rs} = D_{rs}(\bar{S}_{rs})$	D_{rs}	\bar{S}_{rs}
Flot sur chemin (affectation)	f_{rs}^k	$f_{rs}^k = q_{rs} \int \chi_{rs}^k(v) dH_{rs}(v)$	H_{rs}	$q_{rs}, \chi_{rs}^k(v)$
Flot sur arc (agrégation)	x_a	$x_a = \sum_k \delta_{rs}^{ak} f_{rs}^k$	δ_{rs}^{ak}	f_{rs}^k

Nous constatons l'absence de certains composants conceptuels :

- pas de terme aléatoire dans le coût généralisé,
- pas de différenciation comportementale selon d'autres dimensions que la valeur du temps,
- interactions entre destinations par rétroaction de l'ensemble des niveaux de service $\{\bar{S}_{rs}\}_{rs}$ sur l'ensemble des fonctions de demande $\{D_{rs}\}_{rs}$,
- effet d'offre sur la structure de la demande, par rétroaction de la fonction $S_{rs}(v)$ sur la distribution $H_{rs}(v)$,
- dimension temporelle, tant instantanée (temps d'horloge) que par succession des périodes de référence (temps de calendrier) que chronologique (évolution historique).

On peut combler certaines de ces lacunes (Leurent 1993b, 1996a).

7.2.5.2 Appréciation par rapport aux échelles de valeur

Au terme de l'analyse du modèle bicritère, nous pouvons juger a priori sa portée explicative en situant ses hypothèses dans les échelles de valeur établis au §7.1.5.

¹par comparaison entre itinéraires concurrents, et aussi par rapport aux fonctions de demande

Tab. 7F. Audit sémantique du modèle bicritère.

Composant conceptuel		Valeur
Caractères spatial et temporel		Entièrement fonction du cas particulier d'application
Offre	Itinéraires candidats au chargement	bien définis
	Description des routes par des attributs	moyen (temps et prix)
	Représentation de la congestion	médiocre si seulement temps local fonction du flot local ; moyen si contraintes explicites de capacité
	Différenciation des types de véhicules	Selon choix de l'analyste
Demande	Elasticité	Selon choix de l'analyste, au mieux moyen supérieur
	Information et rationalité	médiocre (information totale)
	Différenciation	moyen (uniquement déterministe)

Au total, le modèle bicritère a une portée explicative relative. Selon le composant, l'appréciation varie du bon (si les options convenables sont retenues), au médiocre (dans le cas contraire).

7.2.5.3 Portée du modèle bicritère et limites à l'interprétation

Le modèle bicritère sert à répartir le trafic entre des itinéraires concurrents qui présentent des caractéristiques diverses en termes de prix et de temps. Il convient pour étudier des arbitrages entre routes à péage et routes sans péage, à condition de vérifier l'hypothèse d'information des conducteurs.

Signalons que certains phénomènes inhérents au mode routier (remontée de files d'attente, stationnement) ne sont pas considérés.

Concernant le choix modal, l'emploi de fonctions de demande par origine-destination permet a priori de prendre en compte les autres modes¹. Ce composant peut servir si le modèle s'intègre à une représentation plus élaborée, autrement dit à un système de modèles. L'adaptation du modèle bicritère au choix modal nécessite des précautions : pour comparer des modes de transport individuel, ou des modes de transport collectif, différencier selon la valeur du temps peut suffire ; mais pour comparer des modes individuels à des modes collectifs, il convient probablement d'ajouter d'autres facteurs de différenciation.

¹ainsi d'ailleurs que d'éventuelles interactions entre périodes temporelles

7.3 PERTINENCE ET PROPORTIONNEMENT

L'audit sémantique d'un modèle produit un jugement a priori sur sa portée explicative : les aspects absents de la composition conceptuelle ne peuvent être traités avec pertinence. La question de la pertinence concerne non seulement la composition conceptuelle, mais encore le codage, défini comme la représentation d'objets concrets par des objets abstraits.

Intuitivement, une composition conceptuelle plus élaborée et un codage plus fin assurent une meilleure pertinence : mais jusqu'où aller ? En pratique on s'arrête à un certain stade : ce stade suffit-il ? Autant d'interrogations relatives au proportionnement du modèle aux objectifs d'étude. Le souci de proportionnement complète, borde, la recherche de pertinence. Le proportionnement met l'outil et l'usage en relation de dépendance réciproque : la pertinence requiert à la fois un outil proportionné à l'usage désiré, et un usage proportionné à l'outil employé.

7.3.1 Choisir la composition conceptuelle

Nous avons marqué les limites phénoménologiques du modèle bicritère (Cf. §7.2.5.3). De telles limites de pertinence nécessitent un proportionnement des simulations et des interprétations.

7.3.1.1 L'exemple de la congestion routière

En particulier le caractère statique limite la pertinence de la représentation de la congestion : on n'explicite pas l'éventuelle file d'attente au débouché d'une route, donc on ne calcule pas la durée d'attente en fonction de la longueur de la file et de la capacité impartie, pas plus que l'on ne propage la congestion dans l'espace (en direction des routes en amont).

Cependant l'utilisateur du modèle, s'il connaît les capacités limites d'écoulement, peut leur comparer les flots affectés afin de détecter les saturations et dépassements de capacité. Un usage proportionné à la portée de son outil lui permet de localiser les problèmes de capacité : il lui reste alors à les examiner de façon particulière.

7.3.1.2 Détecter pour brancher vers des représentations plus fines

L'exemple des contraintes de capacité démontre l'utilité de connaître a priori les limites phénoménologiques, afin de leur proportionner l'interprétation des résultats du modèle et de détecter leurs manifestations concrètes. Ces manifestations n'obèrent pas l'utilité du modèle à des fins de connaissance pratique : leur mise en évidence constitue en soi une connaissance utile. A l'utilisateur de juger dans quelle mesure les problèmes soulevés influent sur l'ensemble des résultats, et s'il y a lieu de les étudier de façon particulière (à l'aide d'enquêtes spécifiques et/ou de modèles plus fins).

7.3.1.3 De l'importance de la quantification

Sans mesure, comment évaluer objectivement un phénomène ? Poursuivant l'exemple de la congestion au débouché d'une route, une mesure qualitative révèle l'existence ou non d'une file d'attente, mais on évalue l'importance de l'attente en quantifiant la durée individuelle moyenne d'attente, ou la longueur moyenne de la file etc.

En l'absence d'observations, la connaissance est synthétisée par modèle. Si le modèle reste conceptuel, qualitatif, il est bien sûr possible d'enchaîner des hypothèses : mais l'intérêt pratique est limité par la puissance de représentation abstraite, le raisonnement manque de repères pour supputer la probabilité d'occurrence d'une situation ou juger des contributions respectives de diverses causes à tel effet... Un modèle quantitatif soutient la représentation abstraite, fournit des repères pour guider le raisonnement, permet de hiérarchiser les déterminants d'un phénomène en fonction de leurs contributions respectives.

7.3.2 Choisir le format économétrique

Le codage désigne la schématisation d'objets concrets par des objets abstraits. Ainsi le codage d'une route produit un arc doté de certains attributs. En matière d'affectation du trafic routier, le codage porte sur le réseau de transport et sur l'inscription dans l'espace de la demande.

7.3.2.1 Coder le réseau routier

Distinguons entre la structure topologique, la segmentation en routes et la caractérisation abstraite des routes par des attributs.

La structure topologique du réseau codé doit reproduire celle du réseau "réel" avec un détail suffisant pour que les chemins "effectivement choisis" soient identifiables dans le modèle. Une règle raisonnable de proportionnement est de représenter toutes les routes de niveau supérieur ou égal au niveau de détail juste inférieur à celui souhaité pour l'analyse des résultats.

La segmentation en routes doit respecter la configuration des jonctions "réelles", ainsi que l'homogénéité des routes par rapport à leurs caractéristiques abstraites.

La caractérisation abstraite d'une route par les attributs d'un arc sert à la représentation (notamment cartographique) et à la simulation : il faut coder les attributs qui entrent dans la composition du coût généralisé.

7.3.2.2 Rattacher la demande à des zones

Distinguons entre l'agrégation des lieux, la couverture du territoire et le traitement du trafic local.

L'agrégation des lieux d'extrémité des déplacements en zones de demande doit respecter les conditions du choix d'itinéraire pour un déplacement individuel localisé ponctuellement : les zones doivent être proportionnées au réseau, scinder l'une d'elles ne devrait modifier les choix d'itinéraire que marginalement.

L'extension géographique du territoire d'étude dépend évidemment des objectifs poursuivis ; elle doit être cohérente avec la question du choix d'itinéraire, de façon à représenter les chemins susceptibles d'être choisis. Si la décision de passer par un chemin plutôt qu'un autre intervient en amont sur le réseau, la couverture spatiale doit s'étendre jusqu'au lieu en amont. Le proportionnement intervient ici en relativisant l'importance des trafics avec au moins une extrémité extérieure au territoire d'étude : cette extrémité est assignée non pas à une zone intérieure mais à une zone "du cordon" (i.e. une zone fictive en bordure du territoire circonscrit).

Pour traiter les déplacements dont l'itinéraire est invariable ou qui restent à l'intérieur

d'une zone, on considère qu'ils constituent un trafic local que l'on attache au réseau et non aux zones de demande. Il convient d'expliciter la composition du trafic local, par rapport au contenu des matrices origine-destination de flots.

Chapitre 8

L'audit formel : une vérification

L'audit formel d'un modèle consiste à détecter et expliciter l'erreur de formulation, éventuellement à l'éradiquer. Cette erreur procède d'un manque de conformité entre la composition conceptuelle et la formule caractéristique, ou d'une erreur "syntaxique" dans la formule caractéristique, ou d'une insuffisante détermination de la formule caractéristique. Ainsi la formule caractéristique tient une place centrale ; sa nature intrinsèquement mathématique nous amène à effectuer l'audit formel sur ce terrain.

Le chapitre se compose de quatre parties. D'abord des préparatifs : nous recensons les ressources standards pour formuler un problème d'affectation statique du trafic routier (section 8.1).

Le modèle bicritère ajoute un composant sémantique original (la différenciation continue des valeurs du temps) au modèle standard de Beckmann. Nous donnons des éléments formels pour traiter ce composant, aussi bien en dimension finie (section 8.2) qu'en dimension infinie (section 8.3 ; en effet une distribution continue de la valeur du temps est un objet mathématique de dimension infinie).

Enfin nous tirons quelques enseignements de cette expérience d'audit formel qui met en évidence des arbitrages entre conception et formulation, ainsi qu'entre audit formel et audit algorithmique (section 8.4).

Les parties dépendent peu les unes des autres. L'inventaire de ressources nécessite un bagage mathématique minimal. La formulation en dimension finie est présentée de façon très progressive et, je l'espère, didactique. La formulation en dimension infinie requiert des connaissances mathématiques beaucoup plus étendues.

8.1 RESSOURCES FORMELLES POUR L'AFFECTATION STATIQUE

L'affectation présente une complexité d'ordre mathématique (Cf. §6.1.2.1) : les dépendances circulaires entre variables endogènes font que la solution du modèle se caractérise de façon *implicite*. Cela nécessite des formules caractéristiques adaptées : inéquation variationnelle, problème de point fixe, programme de complémentarité non linéaire, programme extrémal.

Depuis 30 ans, la modélisation des déplacements constitue un terrain privilégié de recherche mathématique sur ces formules caractéristiques. Nous nous contenterons ici de les rappeler (§8.1.2) après avoir présenté les éléments de formalisation (§8.1.1). Puis nous indiquerons les principaux résultats relatifs aux questions d'existence, d'unicité et de stabilité (§8.1.3).

8.1.1 Formalisation

8.1.1.1 Eléments pour le modèle déterministe homogène

Pour formuler les objets et les mécanismes élémentaires de conception, nous avons déjà présenté des notations formelles (Cf. section 7.2). Nous les récapitulons brièvement dans les tableaux joints (tableaux 8A à 8C), à propos du modèle d'affectation déterministe (sans aléa dans le coût généralisé) homogène (un comportement identique pour tous les déplacements).

Tab. 8A. Variables exogènes pour les dimensions physiques et comportementales.

Variable	Type	Sens
a	identifiant	un arc
r	identifiant	une zone d'origine
s	identifiant	une zone de destination
k	identifiant	un chemin (fixé par le modélisateur)
b	identifiant	une contrainte adjointe

Tab. 8Abis. Variables exogènes pour la spécification fonctionnelle.

Variable	Type	Sens
δ_{rs}^{ak}	indicateur	définition de l'itinéraire à partir des arcs
t_a	fonction réelle	temps de parcours de l'arc a
z_b	fonction réelle	contrainte adjointe
D_{rs}	fonction réelle	fonction de demande de l'O-D $r-s$

Tab. 8B. Recensement des variables endogènes.

Variable	Type	Sens
T_a	réel	frais temporels sur l'arc a
w_b	réel ≥ 0	pénalité temporelle liée à la contrainte adjointe $z_b(\mathbf{f}) \leq 0$
G_{rs}^k	réel	temps généralisé ¹ (TG) du chemin k sur l'O-D $r-s$
S_{rs}	réel	niveau de service sur l'O-D $r-s$
q_{rs}	réel	flot total de l'O-D $r-s$
χ_{rs}^k	fonction indicatrice	emploi du chemin k par les déplacements de l'O-D $r-s$: =1 pour un des chemins qui procurent un TG minimum et 0 pour tous les autres
f_{rs}^k	réel	flot sur le chemin k
x_a	réel	flot local sur l'arc a

Tab. 8C. Relations de dépendance.

Sens	Produit (s)	Formule(s)	Entrée(s)	
			EXO.	ENDO.
Formation du temps généralisé <i>congestion et coût généralisé</i>	G_{rs}^k w_b	$G_{rs}^k = \{ \sum_a \delta_{rs}^{ak} t_a \} + \sum_b w_b \frac{\partial z_b}{\partial f_{rs}^k}$ Obtention complexe pour w_b^2	t_a, z_b	w_b, f_{rs}^k
Minimisation du TG et niveau de service <i>rationalité</i>	S_{rs} χ_{rs}^k	$S_{rs} = \min_k G_{rs}^k$ $G_{rs}^k > S_{rs} \Rightarrow \chi_{rs}^k = 0$		G_{rs}^k
Fonction de demande <i>élasticité</i>	q_{rs}	$q_{rs} = D_{rs}(S_{rs})$	D_{rs}	S_{rs}
Flot sur chemin (affectation)	f_{rs}^k	$f_{rs}^k = q_{rs} \chi_{rs}^k$		q_{rs}, χ_{rs}^k
Flot sur arc (agrégation)	x_a	$x_a = \sum_k \delta_{rs}^{ak} f_{rs}^k$	δ_{rs}^{ak}	f_{rs}^k

8.1.1.2 De l'importance de bien définir les termes : un exemple

Chaque notation formelle reflète la conception du modèle, non seulement parce qu'elle renvoie à un objet "concret" (c'est-à-dire supposé tel dans la représentation théorique),

¹dans le cas monocritère, le temps généralisé est simplement le temps pur, ou une somme pondérée à poids invariants de plusieurs attributs : dans ce cas on ne décompose pas la somme

²par comparaison entre itinéraires concurrents, et aussi par rapport aux fonctions de demande

mais encore parce que la précision de sa définition est en rapport direct avec la précision de la conception. Une notation insuffisamment définie témoigne d'une difficulté non résolue, parfois non identifiée, dans la représentation théorique. Cela peut invalider le modèle formel.

Citons en exemple le modèle d'affectation logit à temps variables. Fisk (1980) a formulé ce modèle comme la solution d'un programme d'optimisation convexe, dans lequel la fonction objectif dépend des flux sur les chemins notés f_{rs}^k , avec $r-s$ indice d'origine-destination et k indice de chemin. La conception suppose que tous les chemins identifiés (donc désignés par un indice k) reçoivent une portion non nulle du trafic origine-destination. Fisk a donné une formulation correcte, mais ensuite elle a proposé un algorithme itératif lors duquel l'ensemble des chemins identifiés peut changer de façon incontrôlable : conséquence, l'algorithme ne converge généralement pas ! Le fait est longtemps passé inaperçu (Thomas 1991 finit par le mentionner), faute d'un critère de convergence rigoureux... Dans cet exemple, la formulation a négligé la nécessité mathématique de considérer des objets bien identifiés. Un traitement formel existe : il suffit de fixer l'ensemble des chemins et de traiter constamment tous les chemins identifiés et eux seuls (Leurent 1994b, 1995b).

S'agit-il d'audit formel ou d'audit algorithmique ? Le symptôme du problème apparaît dans la résolution du modèle, mais la cause réside essentiellement dans une omission de la formulation.

8.1.1.3 Notations vectorielles

Des notations vectorielles servent à exprimer les dépendances entre variables de façon concise.

Tab. 8D. Notation vectorielle des variables et fonctions.

Symbole	Signification
\mathbf{x}	$:= [x_a]_a$ vecteur des flux sur les maillons
\mathbf{t}	$:= [t_a]_a$ vecteur des fonctions de temps de parcours des maillons
\mathbf{z}	$:= [z_b]_b$ vecteur des contraintes adjointes
\mathbf{w}	$:= [w_b]_b$ vecteur des multiplicateurs associés aux contraintes adjointes
Δ	$:= [\delta_{rs}^{ak}]_{rsk,a}$ matrice de définition des chemins à partir des maillons
\mathbf{f}	$:= [f_{rs}^k]_{rsk}$ vecteur des flots par chemins
\mathbf{G}	$:= [G_{rs}^k]_{rsk}$ vecteur des temps généralisés sur les chemins
\mathbf{q}	$:= [q_{rs}]_{rs}$ matrice des flots par relation origine-destination
\mathbf{D}	$:= [D_{rs}]_{rs}$ matrice des fonctions de demande
\mathbf{A}	$:= [A_{rsk}^{r's'}]_{rsk,r's'}$ matrice d'expansion de la dimension de \mathbf{q} à celle de \mathbf{f} , $A_{rsk}^{r's'} = 1$ si $\{ (r = r') \text{ et } (s = s') \}$ et 0 sinon.
\mathbf{S}	$:= [S_{rs}]_{rs}$ matrice des niveaux de service entre origines et destinations

Tab. 8E. Forme vectorielle des relations de dépendance.

Formule	Sens
$\mathbf{f} \geq \mathbf{0}$ $\mathbf{S} = \min \mathbf{G}$ et $\mathbf{f} \cdot (\mathbf{G} - \mathbf{AS}) = 0$	comportement rationnel
$\mathbf{q} = \mathbf{D}(\mathbf{S})$	fonction de demande
$\mathbf{G} = \Delta \cdot \mathbf{t}(\mathbf{f}) + \nabla_{\mathbf{f}} \mathbf{z}(\mathbf{f}) \cdot \mathbf{w}$ $\mathbf{z}(\mathbf{f}) \leq \mathbf{0}$ et $\mathbf{w} \geq \mathbf{0}$ et $\mathbf{z}(\mathbf{f}) \cdot \mathbf{w} = 0$	formation des temps généralisés

8.1.2 Formulation en problème type de recherche opérationnelle

Nous avons présenté des termes et des relations élémentaires : comment les intégrer dans une formulation synthétique qui ait une portée mathématique ? Comment obtenir une formule caractéristique qui se prête à étudier l'existence, l'unicité et la stabilité ?

Lorsque les relations élémentaires peuvent être combinées en une formule directe $\mathbf{Y} = \mathbf{F}(\mathbf{X})$, il suffit d'étudier les propriétés mathématiques de la fonction \mathbf{F} : a priori l'existence et l'unicité sont assurées, tandis que la stabilité tient à la continuité de \mathbf{F} .

En l'absence de formule directe, les relations élémentaires suivent la forme $\{\mathbf{F}(\mathbf{X}; \mathbf{Y}) = \mathbf{0}\}$ et $\{\mathbf{G}(\mathbf{X}; \mathbf{Y}) \geq \mathbf{0}\}$, avec \mathbf{F} et \mathbf{G} deux fonctions. On se trouve dans cette situation dès qu'il y a une circularité dans les relations élémentaires, avec une variable endogène en argument d'une fonction élémentaire qui livre une autre variable endogène et réciproquement (revoir la figure 7a).

Alors l'étude n'est pas triviale, même dans un cas avec seulement quelques variables scalaires (Cf. exemple du réseau à deux arcs). Il est bon d'appliquer des résultats mathématiques généraux.

En matière d'affectation du trafic, le cadre mathématique de référence est la théorie de l'optimisation numérique, avec pour principales formulations l'inéquation variationnelle (§.1), le problème de point fixe (§.2), le problème de complémentarité non linéaire (§.3) et le programme extrémal (§.4).

Dans ce qui suit, nous désignons par \mathbf{U} un sous-ensemble non vide, fermé et convexe de \mathcal{R}^N et par \mathbf{V} une application allant de $\mathbf{U} \subseteq \mathcal{R}^N$ dans \mathcal{R}^N .

8.1.2.1 L'inéquation variationnelle

La formule caractéristique d'une inéquation variationnelle est :

$$\text{"Trouver } \mathbf{u}^* \in \mathbf{U} \text{ tel que } \forall \mathbf{u} \in \mathbf{U}, \mathbf{V}(\mathbf{u}^*) \cdot (\mathbf{u} - \mathbf{u}^*) \geq 0\text{"}. \quad (\text{FC1})$$

Pour représenter l'affectation aux itinéraires (Smith, 1979), on prend $\mathbf{u} = \mathbf{f}$ un vecteur de flot par chemins, et $\mathbf{V}(\mathbf{u})$ vecteur des temps généralisés des chemins. L'inéquation variationnelle signifie que si, à partir de l'équilibre, certains déplacements changent d'itinéraire (remplacement de \mathbf{u}^* par \mathbf{u}), alors leur temps total croît (c'est $\mathbf{V}(\mathbf{u}^*) \cdot \mathbf{u}$ qui est au moins aussi grand que $\mathbf{V}(\mathbf{u}^*) \cdot \mathbf{u}^*$). Dans le cas du modèle déterministe

homogène, on peut également prendre pour \mathbf{u} un vecteur de flot par arcs "faisable", i.e. tel qu'il existe un flot par chemins sous-jacent. Dans le cas à demande élastique et contraintes adjointes, on peut prendre $\mathbf{u} = (\mathbf{f}; \mathbf{w})$ et $\mathbf{V}(\mathbf{u}) = (\mathbf{G} - \mathbf{A}\mathbf{S}; -\mathbf{z})$, à condition de vérifier l'existence d'un point satisfaisant aux contraintes adjointes.

Dans le cas d'un réseau à deux arcs F et T, si la demande est fixe alors la variable endogène est $\mathbf{u} = (f_F; f_T) \in \mathbf{U} \equiv \{(f_F; f_T); f_F \geq 0, f_T \geq 0, f_F + f_T = q\}$ où q est le volume de demande. La condition (FC1) équivaut au système

$$\begin{cases} f_F > 0 \Rightarrow V_F = \min\{V_F; V_T\} \\ f_T > 0 \Rightarrow V_T = \min\{V_F; V_T\} \end{cases} \quad (\text{FC1bis})$$

8.1.2.2 Le problème de point fixe

Soit \mathbf{W} une application multivoque de \mathbf{U} dans l'ensemble de ses parties. La formule caractéristique d'un problème de point fixe associé à l'application multivoque \mathbf{W} est :

$$\text{"Trouver } \mathbf{u}^* \in \mathbf{U} \text{ tel que } \mathbf{u}^* \in \mathbf{W}(\mathbf{u}^*)\text{"}. \quad (\text{FC2})$$

Pour le choix d'itinéraire, on peut prendre pour \mathbf{u} un flot par chemins \mathbf{f} , avec $\mathbf{W}(\mathbf{u}) = \overline{\mathbf{D}}(\mathbf{t}(\mathbf{f}))$ un opérateur de demande qui, en fonction des niveaux de service liés à $\mathbf{t}(\mathbf{f})$, calcule les volumes origine-destination ainsi qu'une répartition aux itinéraires selon le principe comportemental du modèle (par exemple en affectant aux plus courts chemins).

On passe d'une inéquation variationnelle à un problème de point fixe en définissant $\mathbf{W}(\mathbf{u}) = \arg \min_{\mathbf{y} \in \mathbf{U}} \mathbf{V}(\mathbf{u}) \cdot (\mathbf{y} - \mathbf{u})$.

8.1.2.3 Le problème de complémentarité non linéaire

La forme standard d'un problème de complémentarité non linéaire est :

$$\text{"Trouver } \mathbf{u}^* \in \mathbf{U} \text{ tel que } \begin{cases} \mathbf{u}^* \geq \mathbf{0} \\ \mathbf{V}(\mathbf{u}^*) \geq \mathbf{0} \\ \mathbf{u}^* \cdot \mathbf{V}(\mathbf{u}^*) = 0 \end{cases} \text{"}. \quad (\text{FC3})$$

Un tel problème se ramène à une inéquation variationnelle de forme (FC1). La réciproque est vraie lorsque $\mathbf{U} = \mathfrak{R}^{+N}$.

8.1.2.4 Programme extrémal ; la transformation de Beckmann

Enfin la formule caractéristique d'un programme de minimisation, pour une fonction objectif J de \mathbf{U} dans \mathfrak{R} , est :

$$\text{"Trouver } \mathbf{u}^* \in \mathbf{U} \text{ solution de } \min_{\mathbf{u} \in \mathbf{U}} J(\mathbf{u})\text{"}. \quad (\text{FC4})$$

Alors \mathbf{u}^* est solution de l'inéquation variationnelle basée sur $\mathbf{V} = \nabla_{\mathbf{u}} J$. En sens réciproque, si \mathbf{V} est intégrable (i.e. si $\forall i, j \frac{\partial V_i}{\partial u_j} = \frac{\partial V_j}{\partial u_i}$), la solution de l'inéquation variationnelle basée sur \mathbf{V} minimise tout potentiel associé J tel que $\nabla_{\mathbf{u}} J = \mathbf{V}$.

Divers programmes extrémaux sont concevables pour caractériser un modèle d'affectation : par exemple, à partir du problème de point fixe, si \mathbf{W} est univoque on

peut former $J(\mathbf{u}) = \|\mathbf{u} - \mathbf{W}(\mathbf{u})\|^2$.

Mais le programme extrémal fondamental est

$$\min_{\mathbf{f}} J(\mathbf{f}) = \sum_{\alpha} \int_0^{x_{\alpha}} t_{\alpha}(\theta) d\theta - \sum_{rs} \int_0^{q_{rs}} D_{rs}^{-1}(\theta) d\theta \quad (\text{FC5})$$

soumis aux contraintes $f_{rs}^k \geq 0$, $x_{\alpha} = \sum_{rsk} \delta_{rs}^{ak} f_{rs}^k$ et $q_{rs} = \sum_k f_{rs}^k$.

Beckmann (1956) l'a formulé en montrant que ses conditions de Kuhn-Tucker sont équivalentes aux mécanismes explicatifs du modèle déterministe homogène à demande élastique. Plusieurs raisons justifient le caractère fondamental du programme de Beckmann :

- des hypothèses mathématiques intuitives sur les t_{α} et D_{rs} (respectivement croissantes et décroissantes) entraînent la convexité, donc l'existence d'une solution et la convexité de l'ensemble des solutions.
- algorithmiquement, on peut le résoudre en manipulant des variables liées uniquement aux arcs et aux relations origine-destination, non aux chemins.
- par analogie économique, il signifie la maximisation d'un surplus (Gartner 1980).
- historiquement, il a inspiré les formulations de modèles d'affectation plus élaborés : logit (Fisk 1980), combinaison d'étapes (Cf. références citées au §6.2.2.2), stochastique (Sheffi et Powell 1982), à attente aux correspondances (Spiess 1985), bicritère (Leurent 1993a).

8.1.2.5 L'ajout de relations élémentaires peut conduire à changer de problème type

Nous avons énoncé les problèmes types de recherche opérationnelle sous forme vectorielle, sans préciser la dimension mathématique de chaque problème. La dimension dépend à la fois du cas particulier d'application (ex. variables liées aux arcs, ou aux relations origine-destination) et de l'élaboration du modèle conceptuel, autrement dit de la richesse des mécanismes explicatifs qui sont synthétisés dans la formule caractéristique. Un modèle en sophistique un autre par ajout de mécanismes explicatifs. Soit l'ajout préserve la formule caractéristique, par exemple l'ajout d'un caractère logit préserve la forme extrémale à la Beckmann, soit il la détruit : par exemple considérer des temps de parcours locaux qui dépendent de façon non symétrique des flots de plusieurs arcs, empêche tout programme extrémal à la Beckmann ; dans ce cas on recourt à une formulation plus générale telle que l'inéquation variationnelle ou le problème de point fixe.

8.1.3 Questionnement

Chaque problème type de recherche opérationnelle se prête à une étude des questions d'existence, d'unicité et de stabilité. Pour chaque problème et chaque question, des hypothèses spécifiques sont requises afin de produire une réponse.

8.1.3.1 Existence

Si $\{V \text{ est continue}\}$ et $\{U \text{ est compact, ou } V \text{ est coercive, ou } V \text{ est fortement monotone}\}$, alors il existe une solution à l'inéquation variationnelle.

Cela s'applique par exemple au modèle déterministe homogène à demande élastique et contraintes adjointes. Il faut d'abord supposer l'existence d'un point faisable qui satisfait les contraintes adjointes. Si \mathbf{G} et \mathbf{z} sont continues, alors \mathbf{V} est continue. La compacité s'obtient en bornant \mathbf{U} ; dans le cas général, il n'y a pas de coercivité ni de forte monotonie de \mathbf{V} .

Si $\{\mathbf{U}$ est compact $\}$ et $\{\mathbf{W}$ est univoque et continue $\}$, alors il existe une solution au problème de point fixe.

Cela s'applique par exemple au modèle logit linéaire à demande fixe : définissant $\mathbf{u} = \{f_{rs}^k\}_{rsk}$ et $\mathbf{W}(\mathbf{u}) = \{q_{rs} \exp(-\theta T_{rs}^k(\mathbf{u})) / [\sum_{\ell} \exp(-\theta T_{rs}^{\ell}(\mathbf{u}))]\}_{rsk}$, alors \mathbf{U} est compact, \mathbf{W} est univoque, et aussi continue dès que les t_a donc les T_{rs}^k sont continues.

Si $\{J$ est continue $\}$ et $\{\mathbf{U}$ est compact, ou J est fortement convexe $\}$, alors il existe une solution au programme extrémal.

Cela s'applique en particulier au programme de Beckmann où sans restriction pratique on peut supposer la compacité de \mathbf{U} , tandis que la continuité de J découle immédiatement du caractère borné des t_a et D_{rs}^{-1} .

8.1.3.2 Unicité

S'il existe une solution à l'inéquation variationnelle et si \mathbf{V} est strictement monotone, alors la solution est unique.

Reprenons l'exemple du modèle déterministe homogène à demande élastique et contraintes adjointes : l'emploi de fonctions de temps de parcours asymétriques compromet en général la monotonie. Les contraintes de capacité $\{x_a \leq C_a\}$ ont un effet identique : si deux arcs successifs a et b de mêmes caractères, sans autre arc confluent en leur sommet commun, sont contraints de la même manière, alors le couple $(w_a ; w_b)$ n'est pas unique en général.

Si $\{\mathbf{U}$ est compact $\}$ et si $\{\mathbf{W}$ est univoque et continue et strictement contractante $\}$, alors il existe une unique solution au problème de point fixe.

S'il existe une solution au programme extrémal et si J est convexe, alors l'ensemble des solutions est convexe. Si de plus J est strictement convexe, la solution est unique.

Pour l'affectation du trafic, la question de l'unicité se décline sous plusieurs formes pas nécessairement équivalentes. Entre autres, dans le cas avec plusieurs classes indicées par u :

- y a-t-il unicité des flots par chemins f_{urs}^k , des temps par chemins T_{urs}^k ?
- y a-t-il unicité des flots par arcs x_{ua} , des temps par arcs t_{ua} ?
- y a-t-il unicité de flots composites par arcs x_a , de temps composites par arcs t_a ?

Il convient de répondre à ces questions, autant pour relativiser les résultats d'un algorithme que pour choisir les objets d'observation et d'estimation dans l'audit économétrique.

8.1.3.3 Stabilité

Fiacco (1978) fournit des résultats de stabilité pour la solution d'un programme convexe. Tobin (1986) les étend à la solution d'une inéquation variationnelle en dimension finie. Nous récapitulerons ces résultats au §10.1.3.

8.2 FORMULE CARACTÉRISTIQUE EN DIMENSION FINIE

J'ai donné plusieurs formules caractéristiques pour l'affectation bicritère : en dimension finie (FCI, ¹) ou en dimension finie (FCF, Leurent 1991b, 1993a, 1996a). La dimension infinie est inhérente à une distribution continue des valeurs du temps : la densité est une fonction continue, donc un objet mathématique de dimension infinie. Ma FCF esquivait la complexité en considérant pour variable endogène un vecteur de flot de dimension finie. Ma FCI a pour variables endogènes un vecteur de flot en dimension infinie. Plus générale que la FCF, elle n'étend cependant la portée sémantique que pour l'élasticité de la demande qu'elle représente de façon désagrégée (une fonction de demande individuelle par valeur du temps), tandis que la FCF traite la demande élastique de façon agrégée.

Dans cette section, nous menons de façon constructive l'audit formel du modèle bicritère à formule caractéristique en dimension finie. Au lieu d'analyser une formule générale qui rassemble les divers composants sémantiques (Cf. Leurent 1996a), nous présentons un composant formel bicritère caractéristique (§8.2.1) puis nous l'étendons, d'abord "en profondeur" par combinaison avec d'autres mécanismes explicatifs (§8.2.2), ensuite "en largeur" en passant du réseau à deux arcs au cas général d'un réseau à plusieurs arcs par chemin, plusieurs chemins par relation O-D et plusieurs O-D (§8.2.3).

8.2.1 Formule bicritère pour le réseau à deux arcs

Considérons provisoirement un réseau à deux arcs concurrents entre une même origine et une même destination. Nous donnons les concepts essentiels du traitement formel (§.1) et nous les illustrons (§.2), puis nous proposons une formule caractéristique adaptée (§.3).

8.2.1.1 Valeurs du temps de coupure offerte et demandée

Nous considérons la concurrence entre une route gratuite F et une route payante T , de temps de parcours respectifs T_F et T_T . Soit P la valeur du péage. Le coût généralisé de déplacement, pour une valeur du temps v , vaut $G'_F(v) = v.T_F$ sur la route gratuite et $G'_T(v) = P + v.T_T$ sur la route à péage.

Pour quelles valeurs du temps v la route gratuite présente-t-elle un coût généralisé minimal ? Il faut $G'_F(v) \leq G'_T(v)$ soit $v.T_F \leq P + v.T_T$ ou encore

$$v.(T_F - T_T) \leq P. \quad (8.1)$$

Comme v est positive, lorsque $T_F - T_T \leq 0$ toute valeur de v vérifie (8.1) : si la route à

¹Cf. Leurent 1991a, 1993b. Un travail antérieur de Dafermos (1981) ne considère pas la demande élastique, contrairement à ma FCI que je reprends dans la section 8.3

péage est plus lente, nul n'a intérêt à la prendre !

Lorsque $T_F - T_T > 0$ alors (8.1) entraîne

$$v \leq P / (T_F - T_T) \equiv v^o. \quad (8.2)$$

La valeur remarquable v^o délimite les valeurs du temps auxquelles la route gratuite procure un coût généralisé minimum, celles comprises entre 0 et v^o , et les valeurs du temps auxquelles la route à péage procure un coût généralisé minimum, celles entre v^o et $+\infty$ car $G'_F(v) \geq G'_T(v) \Leftrightarrow v \geq v^o$. Nous appelons v^o la *valeur du temps de coupure offerte*, puisqu'elle est déterminée par les attributs P , T_F et T_T des itinéraires (donc des services, donc de l'offre).

Supposons que les valeurs du temps se répartissent selon une fonction cumulée $H(v)$, i.e. la proportion de déplacements avec une valeur du temps inférieure à x vaut $H(x)$. Si chaque déplacement s'affecte au chemin qui lui procure un coût généralisé minimum, et si $T_F - T_T > 0$ donc si v^o existe, alors la part de marché "comportementale" de la route gratuite F vaut $H(v^o)$ et celle de la route à péage T vaut $1-H(v^o)$.

Considérons maintenant une répartition quelconque du volume q de déplacements entre les deux routes, soit f_F sur la route gratuite et f_T sur la route à péage. Naturellement, $f_F \geq 0$ et $f_T \geq 0$ et $f_F + f_T = q$. Selon cette affectation, la part de marché de la route gratuite est f_F / q et celle de la route à péage $f_T / q = 1 - f_F / q$.

L'affectation $(f_F; f_T)$, bien que quelconque ⁽¹⁾ et sans hypothèse a priori sur la rationalité des demandeurs, définit intrinsèquement une *valeur du temps de coupure demandée* : sachant que la route à péage est plus chère que la route gratuite, nous pouvons postuler que f_F rassemble les déplacements aux valeurs du temps inférieures à une valeur de coupure demandée v^d , et f_T ceux dont les valeurs du temps dépassent v^d . Par définition, v^d se caractérise par

$$H(v^d) = f_F / q. \quad (8.3)$$

Comme la fonction de répartition cumulée H est croissante au sens large, elle est inversible. Nous pouvons définir sa réciproque, $H^{-1}(\alpha) \equiv \sup\{v ; H(v) < \alpha\}$ que nous prolongeons par continuité en 0 : i.e. $H^{-1}(0) \equiv \lim_{v \rightarrow 0^+} H^{-1}(v)$ aussi égale à $\inf\{v ; dH/dv > 0\}$.

La condition (8.3) entraîne

$$v^d = H^{-1}(f_F / q). \quad (8.3bis)$$

Il y a un équilibre entre l'offre et la demande lorsque chaque déplacement est affecté à un chemin de coût généralisé minimum (pour la valeur du temps de ce déplacement).

¹ sous les seules contraintes $f_F \geq 0$ et $f_T \geq 0$ et $f_F + f_T = q$

L'équilibre correspond ou bien à un "vrai" partage entre itinéraires avec $f_F \in]0; q[$ (cas A), ou bien à un point extrême (cas B) avec soit $f_F = q$ (cas Ba) soit $f_F = 0$ (cas Bb).

Dans le cas A, certaines valeurs du temps prennent le chemin à péage donc le chemin gratuit est le plus lent, donc v^o existe et par définition de l'équilibre $v^o = v^d$, soit encore

$$T_F = T_T + P/H^{-1}(f_F / q). \quad (8.4A)$$

Dans le cas Ba, le chemin à péage n'est le meilleur pour aucune valeur du temps avec une masse non nulle dans la distribution. Soit (i) il n'est jamais le meilleur, i.e. $T_T \geq T_F$ et v^o n'existe pas. Soit (ii) il est le meilleur mais seulement pour des valeurs du temps sans masse, i.e. $T_T < T_F$ mais $v^o \geq \max v^d = H^{-1}(1)$ donc aussi $T_T + P/H^{-1}(1) \geq T_F$. Cette dernière condition résume tout le cas Ba puisqu'elle est moins forte que celle du sous-cas Ba(i), sans en altérer l'hypothèse $f_F = q$. On la réexprime comme

$$\begin{cases} f_F = q \\ T_F \leq T_T + P/H^{-1}(f_F / q) \end{cases} \quad (8.4Ba)$$

Dans le cas Bb, aucune valeur du temps avec une masse ne prend le chemin gratuit. Alors le chemin à péage est plus rapide, v^o existe bien, et du côté de la demande toutes les valeurs du temps avec une masse dépassent v^o , i.e. $v^o \leq \min v^d = H^{-1}(0)$. Donc $P/(T_F - T_T) \leq H^{-1}(0)$ ou encore

$$\begin{cases} f_F = 0 \\ T_F \geq T_T + P/H^{-1}(f_F / q) \end{cases} \quad (8.4Bb)$$

La réunion des conditions (8.4) met en relation les variables exogènes de l'équilibre (T_F, T_T, P, q) et les variables endogènes (f_F, f_T). Par construction, l'équilibre nécessite que l'une ou l'autre des conditions soit vérifiée. En sens réciproque, si l'une d'elles est vérifiée, alors nécessairement on se trouve à l'équilibre. Donc la réunion¹ des conditions (8.4) équivaut à la définition de l'équilibre, elle *caractérise* l'équilibre bicritère entre offre et demande.

Insistons sur la non redondance des deux valeurs du temps de coupure. Celle côté offre n'indique a priori rien quant à l'affectation réelle de la demande. Celle côté demande n'indique a priori rien quant à la situation de l'offre. Seul le principe de rationalité individuelle met les deux valeurs en équilibre, lorsqu'il y a un vrai partage entre les chemins.

8.2.1.2 Illustration

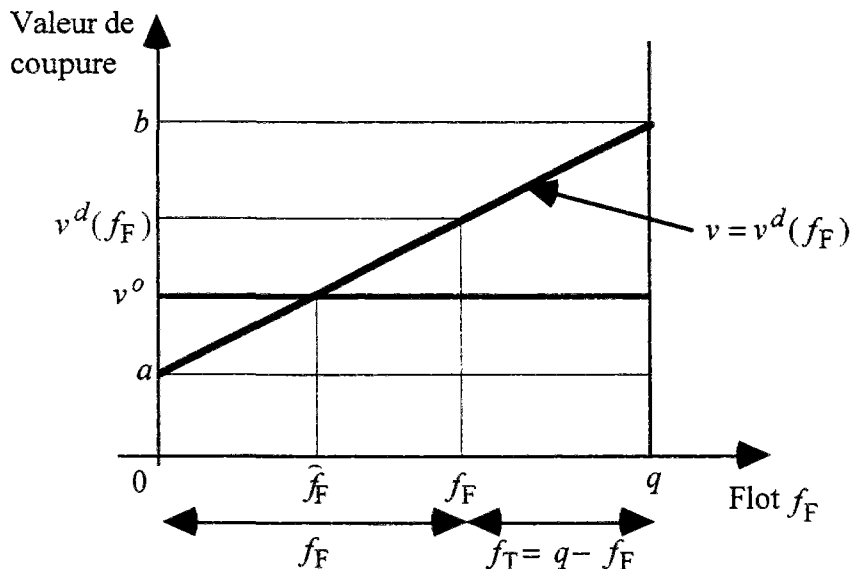
Soit un diagramme avec en abscisse le volume sur le chemin gratuit F, compté entre 0 et q : nous repérons une affectation par un point du segment $[0; q]$, distant de f_F de la droite verticale d'abscisse 0, et de f_T de la droite verticale d'abscisse q . En ordonnée figurent les valeurs de coupure : nous portons la valeur de coupure demandée, v^d

¹précisément, la disjonction logique des sous-systèmes (8.4A), (8.4Ba) et (8.4Bb)

conçue comme une fonction de f_F (et de q , mais celui-ci est fixe).

Pour une situation d'offre déterminée (P , T_F et T_T fixés) dans le cas A, avec $T_F > T_T$, la valeur de coupure offerte vaut $v^o = P / (T_F - T_T)$ donc se représente par une droite horizontale sur le diagramme. Alors l'équilibre entre offre et demande se trouve à la valeur \hat{f}_F du flot sur la route gratuite (figure 8a).

Fig. 8a. Egalité des valeurs de coupure respectivement offerte et demandée.



A titre d'exemple, nous considérons une valeur du temps distribuée uniformément entre deux valeurs a et b (rectangle). Donc $H^{-1}(\alpha) = a + \alpha(b - a)$ pour $\alpha \in [0;1]$, ce qui se représente par une portion de droite entre les points $(0;a)$ et $(q;b)$.

La figure 8a permet aussi d'analyser les situations de déséquilibre : en un point $(f_F; v^d(f_F))$ donné tel que $0 < f_F < q$, si $v^d(f_F)$ est inférieure à v^o cela signifie que les valeurs du temps entre v^d et v^o sont affectées à tort à la route gratuite. Tandis qu'une $v^d(f_F)$ supérieure à v^o implique que les valeurs entre v^o et v^d sont affectées à tort à la route à péage.

Toujours de façon graphique, on représente le cas Ba(i) sans ligne horizontale v^o et le cas Ba(ii) avec une ligne horizontale v^o qui dépasse $v^d(q)$. Le cas Bb se représente par une ligne horizontale v^o au dessous de $v^d(0)$.

8.2.1.3 Formule caractéristique

Comment faire la synthèse des trois situations d'équilibre dans (8.4) ? Elles partagent :

- les hypothèses générales du problème, notamment $f_F \geq 0$, $f_T \geq 0$ et $f_F + f_T = q$.
- l'intervention de T_F et $T_T + P / H^{-1}(f_F / q) \equiv \tilde{T}_T$.
- la mise en comparaison de T_F et \tilde{T}_T .

La ressemblance paraît s'arrêter là puisque la comparaison de T_F avec \tilde{T}_T change selon la situation d'équilibre ! Cependant, soit $\lambda' \equiv \min\{T_F; \tilde{T}_T\}$. Le cas Ba impose $f_T(\tilde{T}_T - T_F) = 0$ si $\tilde{T}_T \geq T_F$, donc $f_T(\tilde{T}_T - \lambda') = 0$; également, comme dans ce cas $T_F = \lambda'$, $f_F(T_F - \lambda') = 0$. De son côté, le cas Bb impose $f_F(T_F - \tilde{T}_T) = 0$ si $T_F \geq \tilde{T}_T$, donc $f_F(T_F - \lambda') = 0$, tandis que $f_T(\tilde{T}_T - \lambda') = f_T \cdot 0 = 0$. Enfin, dans le cas A, $T_F = \tilde{T}_T = \lambda'$ donc $f_F(T_F - \lambda') = f_T(\tilde{T}_T - \lambda') = 0$.

Au total, chaque état d'équilibre nécessite que

$$T_F - \lambda \geq 0 \tag{8.5a}$$

$$\tilde{T}_T - \lambda \geq 0 \tag{8.5b}$$

$$f_F(T_F - \lambda) = 0 \tag{8.5c}$$

$$f_T(\tilde{T}_T - \lambda) = 0, \tag{8.5d}$$

dans lequel nous avons noté λ plutôt que λ' afin de rendre la définition de λ endogène à (8.5), en nous affranchissant désormais de la définition $\lambda' \equiv \min\{T_F; \tilde{T}_T\}$. En effet celle-ci devient superflue : si $f_F > 0$ alors (8.5c) l'entraîne, si $f_T > 0$ alors (8.5d) l'entraîne, tandis que si $f_F = f_T = 0$ alors $q = 0$ donc le problème de répartir q est trivial !

Ainsi l'équilibre bicritère implique (8.5). L'implication réciproque est triviale : à partir de (8.5), on distingue selon la nullité de f_F et/ou f_T . Si $f_F > 0$ et $f_T > 0$ on aboutit à donc à (8.4A), si $f_F > 0$ et $f_T = 0$ à (8.4Ba) ou (8.4A), enfin si $f_F = 0$ et $f_T > 0$ à (8.4Bb) ou (8.4A).

Nous reconnaissons dans (8.5) les prémices d'un programme de complémentarité non linéaire (NCP, Cf. §8.1.2.3) avec pour variables endogènes le vecteur $\mathbf{u} = (f_F; f_T; \lambda)$ et pour application \mathbf{F} qui à \mathbf{u} associe le vecteur

$$\mathbf{F}(\mathbf{u}) = \begin{bmatrix} T_F - \lambda \\ \tilde{T}_T - \lambda \\ F_\lambda \end{bmatrix},$$

cela sans contrainte particulière sur la composante F_λ ! Profitons-en pour placer le problème dans le domaine $(\mathfrak{R}_+)^3$ de préférence à $\mathbf{F}_q \times \mathfrak{R} = \{(f_F; f_T); f_F \geq 0, f_T \geq 0, f_F + f_T = q\} \times \mathfrak{R}$, en imposant $\lambda \geq 0$ et en définissant $F_\lambda(\mathbf{u}) = f_F + f_T - q$: cet artifice prépare l'ajout des composants "temps variables" et surtout "demande élastique".

Une solution du NCP sur $(\mathfrak{R}_+)^3$:

$$\mathbf{u} \geq \mathbf{0}, \mathbf{F}(\mathbf{u}) \geq \mathbf{0} \text{ et } \mathbf{u} \cdot \mathbf{F}(\mathbf{u}) = 0 \tag{8.6}$$

satisfait (8.5), et si $q > 0$ alors (8.5c, d) implique $\lambda = T_F$ ou $\lambda = \tilde{T}_T$, d'où $\lambda > 0$, donc $F_\lambda = 0$ et $f_F + f_T = q$. Comme (8.5) implique l'équilibre bicritère, (8.6) également.

Au total le NCP (8.6) caractérise l'équilibre bicritère. Comme il est défini sur $(\mathcal{R}_+)^3$, il équivaut à une inéquation variationnelle. Formellement, l'inéquation variationnelle est plus simple que le NCP car elle ne nécessite pas d'explicitier λ . Elle a pour variable endogène le vecteur $\mathbf{f} \in \mathbf{F}_q$ et pour application associée

$$\mathbf{V}: \mathbf{f} = (f_F; f_T) \mapsto \mathbf{V}(\mathbf{f}) = (T_F; \tilde{T}_T). \quad (8.7)$$

8.2.2 Extensions sémantiques "en profondeur"

Grâce aux concepts de valeurs de coupure offerte et demandée, nous obtenons une formule caractéristique standard, exprimée en fonction des mêmes variables endogènes (les flots par chemins) que les composants plus traditionnels de l'affectation (temps variables, contraintes de capacité, demande élastique). Comme ceux-ci prennent aussi la forme d'un NCP ou d'une inéquation variationnelle, le composant formel bicritère en dimension finie se laisse aisément combiner avec eux.

8.2.2.1 Cas des temps variables

Supposons maintenant que les temps de parcours des routes dépendent du vecteur des flots $\mathbf{f} = (f_F; f_T)$ selon des fonctions

$$T_F = t_F(\mathbf{f}) \text{ et } T_T = t_T(\mathbf{f}). \quad (8.8a, b).$$

Alors la définition de l'équilibre entre offre et demande change, puisque l'on ajoute le composant formel (8.8) de "fonctionnement technique" de l'offre. Le nouvel équilibre se caractérise par la conjonction de (8.4) et de (8.8), donc encore par la conjonction de (8.6) et (8.8). Par substitution, l'équilibre bicritère à temps variables se caractérise comme la solution du NCP défini sur $(\mathcal{R}_+)^3$ pour l'application \mathbf{F} :

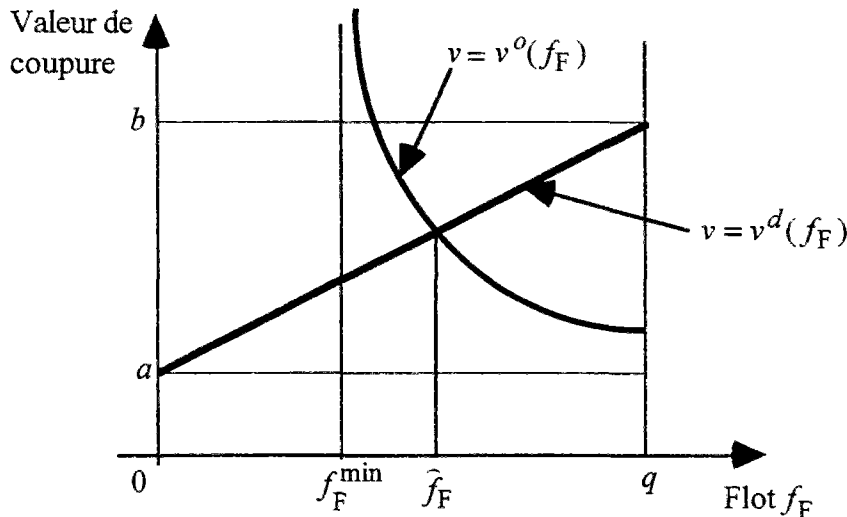
$$\mathbf{u} = \begin{bmatrix} f_F \\ f_T \\ \lambda \end{bmatrix} \mapsto \mathbf{F}(\mathbf{u}) = \begin{bmatrix} t_F(\mathbf{f}) - \lambda \\ t_T(\mathbf{f}) + P/H^{-1}(f_F / q) - \lambda \\ f_F + f_T - q \end{bmatrix}. \quad (8.9)$$

Pour un volume O-D fixé à q , (8.8) revient à des dépendances $T_F = t_F(f_F)$ et $T_T = t_T(q - f_F)$ donc $v^o = P / (t_F(f_F) - t_T(q - f_F))$ est une fonction de f_F . Le cas A des conditions (8.4) se divise maintenant en deux éventualités. Soit (cas Aa) il existe $f_F^{\min} \in]0; q[$ telle que $t_F(f_F^{\min}) = t_T(q - f_F^{\min})$, auquel cas les valeurs de $f_F \in]0; f_F^{\min}]$ ne peuvent relever d'un équilibre de type A, ni d'ailleurs d'un autre type. Soit (cas Ab) une telle valeur f_F^{\min} n'existe pas, ce qui implique par continuité ou bien que $t_F(0) \geq t_T(q)$, ou bien que $t_F(q) \leq t_T(0)$ ce qui relève du cas Ba et non du cas A. Donc le cas Ab correspond uniquement à $t_F(0) \geq t_T(q)$.

De façon graphique, la figure 8b illustre le cas Aa. Elle se distingue de la figure 8a car v^o est désormais une fonction de f_F , et par la présence de la valeur spéciale f_F^{\min} . Dans le cas Ab cette valeur n'existe pas. Dans le cas Ba, la courbe $v = v^o(f_F)$ n'existe pas, ni

non plus la valeur f_F^{\min} . Dans le cas Bb, la courbe $v = v^o(f_F)$ existe mais reste en dessous de $\min v^d(f_F)$, donc il n'y a pas de valeur singulière f_F^{\min} .

Fig. 8b. Egalité des valeurs de coupure, cas des temps variables.



La valeur de coupure offerte $v = v^o$, constante sur la figure 8a, devient une fonction variable $v = v^o(f_F) = P / (t_F(f_F) - t_T(q - f_F))$.

Les situations de déséquilibre s'analysent sur la figure 8b comme sur la figure 8a, à la nuance près qu'en un point de déséquilibre, par exemple $f_F < \tilde{f}_F$, l'affectation des valeurs du temps comprises entre $v^d(f_F)$ et $v^o(f_F)$ peut n'être pas totalement erronée : car la réaffectation à la route gratuite de certaines valeurs du temps supérieures à $v^d(f_F)$, tend à augmenter f_F , donc à accroître $t_F(f_F)$ et réduire $t_T(q - f_F)$, donc à diminuer v^o . Le rapprochement de v^d et v^o par transfert de certaines valeurs du temps, donc de flot, s'effectue par un mouvement conjoint vers l'équilibre.

Toute cette analyse se résume dans le NCP (8.9) analogue de (8.6), donc dans l'inéquation variationnelle (8.10) analogue de (8.7) à la substitution près de (8.8) :

$$V: f \in F_q \mapsto V(f) = (t_F(f); t_T(f) + P/H^{-1}(f_F/q)). \tag{8.10}$$

8.2.2.2 Contraintes de capacité

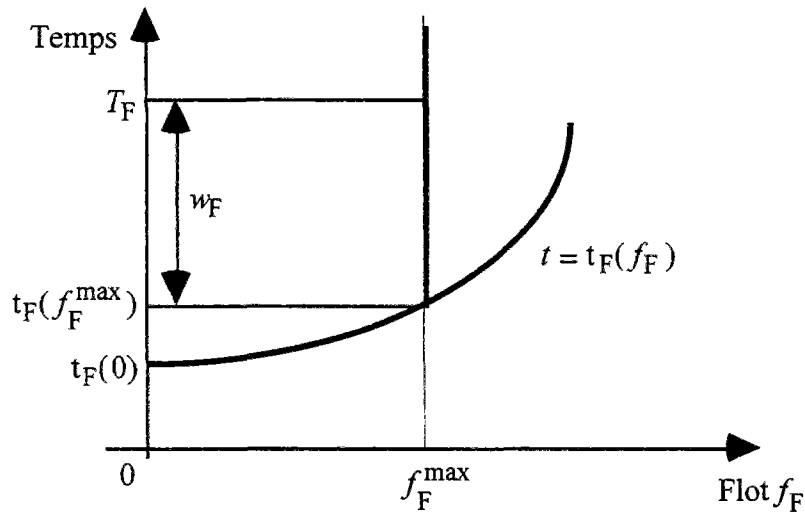
Commençons par expliquer la nature et l'interprétation d'une contrainte de capacité. Une telle contrainte limite le débit d'un service, pour des raisons physiques comme le nombre de files ou la distance minimale entre usagers. Lorsque le débit maximum est atteint, une file d'attente se forme et augmente le temps moyen de service. De quoi dépend la durée d'attente ? A la fois des arrivées des clients et des autres services accessibles aux clients potentiels, qui constituent autant de possibilités de détournement. La durée d'attente apparaît ainsi comme une variable endogène du problème.

Pour notre réseau à deux arcs, soient les contraintes

$$(f_F \leq f_F^{\max}) \text{ et } (f_T \leq f_T^{\max}). \tag{8.11a, b}$$

La contrainte (8.11a) n'intervient que si $f_F^{\max} < q$. Si la fonction de temps de parcours dans (8.8a), indépendamment de q , se formule $t_F(f_F)$, alors le temps total T_F comprend à la fois le temps de service proprement dit t_F et la durée d'attente $w_F \geq 0$ associée à (8.11a), non nulle seulement si $f_F = f_F^{\max}$. La figure 8c récapitule ces considérations.

Fig. 8c. Effets d'une contrainte de capacité.



Si $f_F = f_F^{\max}$, w_F s'interprète comme un retard, une attente subie avant d'accéder à la route F. Il s'agit d'un retard perçu, subjectif, puisqu'il résulte des comportements des clients : il excède probablement la valeur moyenne véritable, objective de l'attente, car les clients ressentent les temps d'attente avec davantage d'intensité que les temps de service.

Nous modifions donc (8.8a, b) en respectivement

$$T_F = t_F(\mathbf{f}) + w_F \text{ avec } w_F \geq 0 \text{ et } w_F(f_F^{\max} - f_F) = 0, \quad (8.12a)$$

$$T_T = t_T(\mathbf{f}) + w_T \text{ avec } w_T \geq 0 \text{ et } w_T(f_T^{\max} - f_T) = 0. \quad (8.12b)$$

L'équilibre entre offre et demande pour le modèle bicritère à temps variables et contraintes de capacité se définit par la conjonction de (8.4) et de (8.12), ou par la conjonction de (8.6) et de (8.12). Or les conditions (8.12) ressemblent elles-mêmes à une partie de NCP pour des variables endogènes w_F et w_T ! Unissons donc le NCP (8.6) et (8.12) : soit \mathbf{F} l'application qui à $\mathbf{u} = (f_F; f_T; \lambda; w_F; w_T) \in (\mathfrak{R}^+)^5$ associe le vecteur

$$\mathbf{F}(\mathbf{u}) = \begin{bmatrix} t_F(\mathbf{f}) + w_F - \lambda \\ t_T(\mathbf{f}) + w_T + P/H^{-1}(f_F/q) - \lambda \\ f_F + f_T - q \\ f_F^{\max} - f_F \\ f_T^{\max} - f_T \end{bmatrix}. \quad (8.13)$$

Par construction, le NCP basé sur (8.13) équivaut à la conjonction de (8.6) et (8.12),

donc il caractérise un équilibre bicritère à temps variables et contraintes de capacité.

Ainsi la spécification exogène de q , H , P , t_F , t_T , f_F^{\max} et f_T^{\max} détermine les variables endogènes f_F , f_T , w_F et w_T à l'équilibre. Seule exception, si $f_F^{\max} + f_T^{\max} < q$ le réseau ne peut écouler le volume de demande donc il n'y a pas d'équilibre entre offre et demande. Alors le NCP basé sur (8.13) n'admet pas de solution.

Etudions maintenant la relation entre un équilibre contraint (solution du NCP basé sur (8.13)) et un équilibre libre, i.e. sans contraintes de capacité (solution du NCP (8.9)). Pour assurer l'existence d'un équilibre contraint, nous imposons $f_F^{\max} + f_T^{\max} > q$. Si la contrainte (8.12a) vérifie $f_F^{\max} < q$ alors l'équilibre "libre" de type Ba avec $\hat{f}_F = q$ est exclu. De même, si la contrainte (8.12b) vérifie $f_T^{\max} < q$ alors un équilibre libre de type Bb avec $\hat{f}_F = 0$ est exclu puisque nécessairement $\hat{f}_F \geq q - f_T^{\max}$. Supposons donc à la fois $f_F^{\max} + f_T^{\max} > q$, $f_F^{\max} < q$ et $f_T^{\max} < q$. Alors il y a un "vrai" partage du volume entre les deux chemins, donc égalité entre $v^o(f_F; w_F; w_T)$ et $v^d(f_F)$ de façon analogue à la situation d'équilibre libre A.

Raisonnons comme si l'on imposait les contraintes de capacité (8.12) de façon très progressive, à partir d'un équilibre libre "antérieur", jusqu'à l'équilibre contraint. Soit l'équilibre libre ne sature aucune des contraintes de capacité : donc il résout le problème contraint, et il se caractérise par

$$\begin{cases} \hat{w}_F = 0 \\ \hat{w}_T = 0 \\ t_T(\hat{\mathbf{f}}) + P/H^{-1}(\hat{f}_F / q) = t_F(\hat{\mathbf{f}}) \end{cases} \quad (8.14A.A)$$

Soit l'équilibre libre comporte "trop de trafic sur la route F", cas Ba ou cas A avec $\hat{f}_F \geq f_F^{\max}$, alors la contrainte de capacité (8.12a) agit sur \hat{f}_F , donc \hat{w}_F est fixé par

$$\begin{cases} \hat{f}_F = f_F^{\max} \\ \hat{w}_T = 0 \\ t_T(\hat{\mathbf{f}}) + P/H^{-1}(\hat{f}_F / q) = t_F(\hat{\mathbf{f}}) + \hat{w}_F \end{cases} \quad (8.14A.Ba)$$

condition caractéristique d'une situation contrainte A.Ba analogue d'un cas libre Ba.

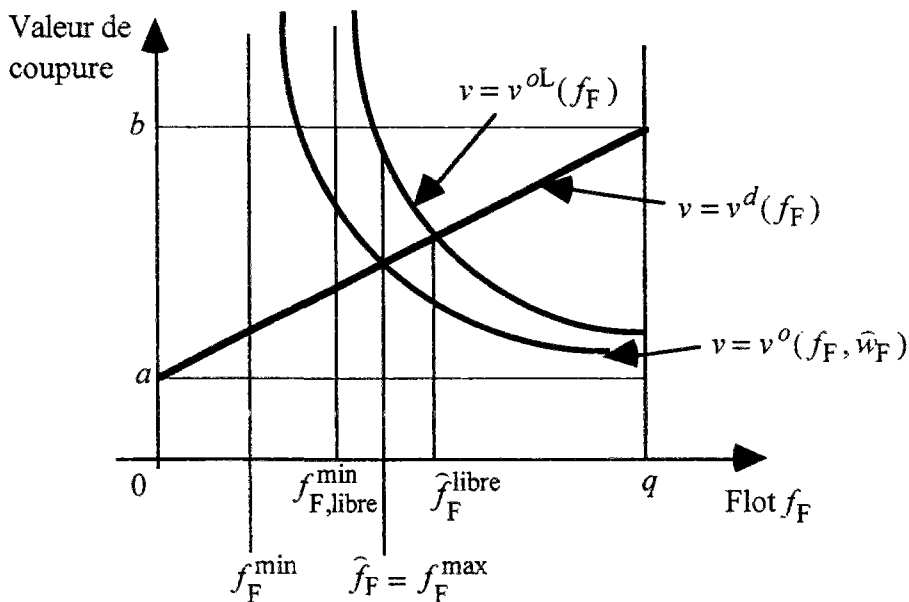
Soit l'équilibre libre comporte "trop de trafic sur la route T", cas Bb ou cas A avec $\hat{f}_T \geq f_T^{\max}$: cette fois la contrainte de capacité agit sur \hat{f}_T , donc \hat{w}_T est fixé par

$$\begin{cases} \hat{f}_F = q - f_T^{\max} \\ \hat{w}_F = 0 \\ t_T(\hat{\mathbf{f}}) + P/H^{-1}(\hat{f}_F / q) + \hat{w}_T = t_F(\hat{\mathbf{f}}) \end{cases} \quad (8.14A.Bb)$$

condition caractéristique d'une situation contrainte A.Bb analogue d'un cas libre Bb.

De façon graphique, un équilibre contraint de type A.A nécessite que la valeur de coupure offerte libre $v^{oL} = P / (t_F(f) - t_T(f))$ intercepte la valeur de coupure demandée $v^d(f_F)$ entre les points limites f_F^{\max} et $q - f_T^{\max}$. Donc, par continuité, comme v^d croît et v^{oL} décroît, que $v^d(f_F^{\max}) \leq v^{oL}(f_F^{\max})$ et $v^d(q - f_T^{\max}) \geq v^{oL}(q - f_T^{\max})$. La dernière relation implique que $v^{oL}(q - f_T^{\max})$ est borné, donc $q - f_T^{\max} > f_F^{\min}$ du modèle libre : si $q > f_F^{\min} + f_T^{\max}$ on ne peut se trouver dans la situation d'équilibre sous contraintes A.A, on se situe nécessairement dans la situation A.Ba de manière à modifier, contraindre la définition de f_F^{\min} .

Fig. 8d. Un équilibre contraint de type A.Ba : décalage de la valeur de coupure offerte.



Dans un équilibre de type A.Ba, la demande sans contrainte pour la route F sature la capacité. La courbe des valeurs de coupure offerte du cas libre est déplacée de façon à satisfaire $v^o(\hat{f}_F, \hat{w}_F) = v^d(\hat{f}_F)$. Dans le cas (i) avec $\exists f_F^{\min} \in]0; q[$ telle que $t_F(f_F^{\min}) = t_T(q - f_F^{\min})$, l'asymptote verticale $f = f_F^{\min}$ se décale vers la gauche.

8.2.2.3 Demande élastique

Un problème contraint tel que le volume de demande q dépasse la somme des capacités f_F^{\max} et f_T^{\max} des chemins, fournit une bonne introduction à l'élasticité de la demande. Si la capacité offerte ne peut satisfaire la demande a priori q , celle-ci ne constitue pas le bon paramètre du problème, il faut considérer la demande servie, connue a posteriori.

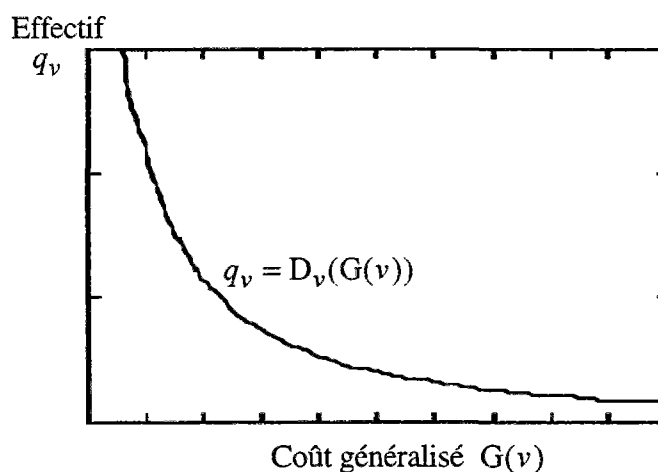
Sur un marché économique, la demande servie dépend non seulement de la capacité offerte qui la borne, mais encore du prix du bien échangé. Si le bien est normal, une baisse de son prix attire davantage de clients (induction) tandis qu'une hausse décourage certains clients (désinduction). Sur le marché du bien "acheminement entre une origine et une destination", un consommateur potentiel d'itinéraire prend en compte dans son comportement "d'achat" non seulement le prix mais encore d'autres aspects (temps,

confort etc.), ce que l'on résume dans le "coût généralisé".

Pour un type v de consommateur, i.e. pour une valeur du temps donnée v dans le modèle bicritère, le volume de demande q_v dépend du temps généralisé $G(v)$ par une fonction particulière, $q_v = D_v(G(v))$ (dont la figure 8e donne un exemple).

Les clients d'un chemin sont indifférents à une modification de l'autre chemin tant que celui-ci ne devient pas plus intéressant pour eux que leur itinéraire actuel ; donc leurs effectifs q_v restent stables. Cependant sur le chemin modifié les clients réagissent probablement à un changement qui les concerne, auquel leurs effectifs q_v s'adaptent. Au total, *la structure de la demande* (sa répartition entre les classes, entre les valeurs du temps) *se déforme*.

Fig. 8e. Une fonction de demande.



Dans notre formulation bicritère en dimension finie, nous négligeons la déformation structurelle de la demande par effet d'offre et nous supposons que le volume total $q = \int q_v dH(v)$ s'adapte au temps généralisé moyen $\bar{G} = \int G(v) dH(v)$ via une fonction de demande D :

$$q = D(\bar{G}). \quad (8.15)$$

Comment formuler \bar{G} en fonction des variables endogènes f_F , f_T et de T_F , T_T et P ? Nous choisissons de l'exprimer comme un *temps généralisé moyen possible*, en considérant pour chaque valeur du temps le meilleur temps généralisé $G(v)$ possible :

$$G(v) \equiv \min \{T_F ; T_T + P/v\}. \quad (8.16)$$

Si $T_F \leq T_T$ alors $G(v) = T_F$ pour toute valeur $v > 0$. Si $T_F > T_T$ alors la valeur de coupure offerte v^0 existe et délimite les valeurs du temps $v \leq v^0$ pour lesquelles $G(v) = T_F$ et les valeurs du temps $v > v^0$ pour lesquelles $G(v) = T_T + P/v$. Dans ce cas $\bar{G} = (\int_0^{v^0} T_F dH) + (\int_{v^0}^{+\infty} (T_T + \frac{P}{v}) dH) = T_F H(v^0) + T_T(1 - H(v^0)) + P \int_{v^0}^{+\infty} \frac{dH(v)}{v}$ dépend encore uniquement des variables d'offre T_F , T_T et P .

Résumons l'élasticité de la demande en fonction des variables endogènes par

$$f_F + f_T = D \left(\int \min \left\{ T_F ; T_T + \frac{P}{v} \right\} dH(v) \right). \quad (8.17)$$

L'équilibre bicritère à demande élastique et temps fixes se définit par la conjonction de (8.4) et de (8.17). L'équilibre bicritère à demande élastique et temps variables et contraintes de capacité se définit quant à lui par la conjonction de (8.14) et de (8.17). Donc il se *caractérise* par la conjonction de (8.17) et du NCP basé sur (8.13), ce que l'on résume par un NCP défini pour les variables endogènes $\mathbf{u} = (f_F; f_T; \lambda; w_F; w_T)$ dans $(\mathcal{R}_+)^5$ en remplaçant dans la composante F_λ le volume q par le membre de droite de (8.17) : soit

$$\mathbf{F}(\mathbf{u}) = \begin{bmatrix} t_F(\mathbf{f}) + w_F - \lambda \\ t_T(\mathbf{f}) + w_T + P/H^{-1}(f_F/q) - \lambda \\ f_F + f_T - D(\bar{G}(\mathbf{f}; w_F; w_T)) \\ f_F^{\max} - f_F \\ f_T^{\max} - f_T \end{bmatrix}. \quad (8.18)$$

Comment déterminer graphiquement la solution d'un tel problème ? Imaginons un diagramme en trois dimensions, qui correspondent respectivement à f_F , f_T et aux valeurs de coupure. Pour des valeurs fixées de w_F et w_T , les valeurs de coupure demandée et offerte sont deux fonctions des deux variables scalaires f_F et f_T : chaque courbe représentative dans le cas de la demande fixe devient une surface représentative dans le cas de la demande variable. De même la fonction de demande définit une autre surface d'équation $f_F + f_T - D(\bar{G}(\mathbf{f}; w_F; w_T)) = 0$. L'intersection des trois surfaces livre l'équilibre bicritère à demande élastique.

8.2.2.4 Autres éléments

Concluons à la possibilité de caractériser un modèle bicritère combiné ou non aux composants "temps variables", "contraintes de capacité", "demande élastique", sous la forme d'un problème mathématique standard, le problème de complémentarité non linéaire (ou par une inéquation variationnelle, de façon équivalente ici).

D'autres combinaisons possibles tiennent à la segmentation discrète selon d'autres critères que la valeur du temps : relation O-D, type de véhicule, valeur de la distance ou du confort etc. J'ai évoqué ailleurs les possibilités existantes (Leurent 1996a).

Les aspects stochastiques et la dépendance des prix en fonction des trafics (ex. parce que les frais de carburant dépendent de la congestion) restent des questions ouvertes, qui n'ont pas encore reçu de formulation en dimension finie.

8.2.3 Extensions sémantiques "en largeur"

Jusqu'à présent, notre audit formel du modèle bicritère conclut à l'existence de formules caractéristiques rigoureuses pour une composition sémantique "profonde", avec plusieurs mécanismes explicatifs donc une complexité de sens. Cependant une composition sémantique "large" suppose de traiter non seulement un réseau élémentaire, mais le cas général d'un réseau avec des itinéraires composés de plusieurs routes et qui peuvent se chevaucher, avec plusieurs itinéraires par relation O-D et plusieurs relations O-D : soit une complexité d'abondance.

Notre audit formel dissocie volontairement la complexité de sens "en profondeur", de la complexité d'abondance "en largeur". Pour le modèle bicritère formulé en dimension finie, la dissociation est naturelle puisqu'elle n'entraîne ni rupture ni redondance. Le seul point délicat réside dans la pluralité des itinéraires par relation origine-destination : elle nécessite un traitement formel original que nous présentons (§.1) et caractérisons (§.2), pour ensuite expliquer comment considérer la demande élastique (§.3), plusieurs relations O-D (§.4), plusieurs arcs par itinéraire et des temps variables (§.5).

8.2.3.1 Classes isoprix, valeurs de coupure offertes et demandées

Pour étudier la concurrence entre deux chemins, nous avons d'abord composé les temps et les prix, puis mis en relation les valeurs du temps de coupure offerte et demandée, enfin nous avons associé à chaque chemin une fonction d'impédance qui caractérise ses qualités pour les consommateurs. Lorsque la concurrence oppose $\bar{k} \geq 2$ chemins, comment les comparer ? Comment définir les valeurs du temps de coupure, entre toute paire de chemins ou entre certains chemins "voisins" ? Existe-t-il une fonction d'impédance caractéristique de l'équilibre ?

De fait les notions utiles pour deux chemins se généralisent pour \bar{k} chemins, sous réserve de concepts intermédiaires plus sophistiqués : classes isoprix, valeurs de coupure offertes et demandées, fonctions d'impédance (tableau 8F).

Tab. 8F. Les principales notions originales de l'équilibre bicritère.

Notation	Sens	Rappel cas à 2 arcs
m	classe isoprix, groupe les chemins de même prix	les deux routes F et T
E_m E'_m	ensemble H-efficace de la classe m agrégat d'ensembles H-efficaces	si $T_F \leq T_T$, $E_F = \mathfrak{R}_+$ et $E_T = \emptyset$, sinon $E_F = [0; v^o]$ et $E_T =]v^o; +\infty]$
v_m^o	valeur de coupure offerte	v^o
v_m^d	valeur de coupure demandée	v^d
\bar{I}_m	fonction d'impédance de la classe m	$\bar{I}_F = T_F$ et $\bar{I}_T = T_T + P / H^{-1}(\frac{I_F}{q})$

Soit T_k le temps d'un chemin k , P_k son prix et $G_k(v) \equiv T_k + P_k / v$ son temps généralisé pour un déplacement de valeur du temps v . Le chemin peut servir aux déplacements s'il

existe des valeurs du temps auxquelles il procure un temps généralisé minimum, i.e. s'il est efficace au sens de la définition suivante :

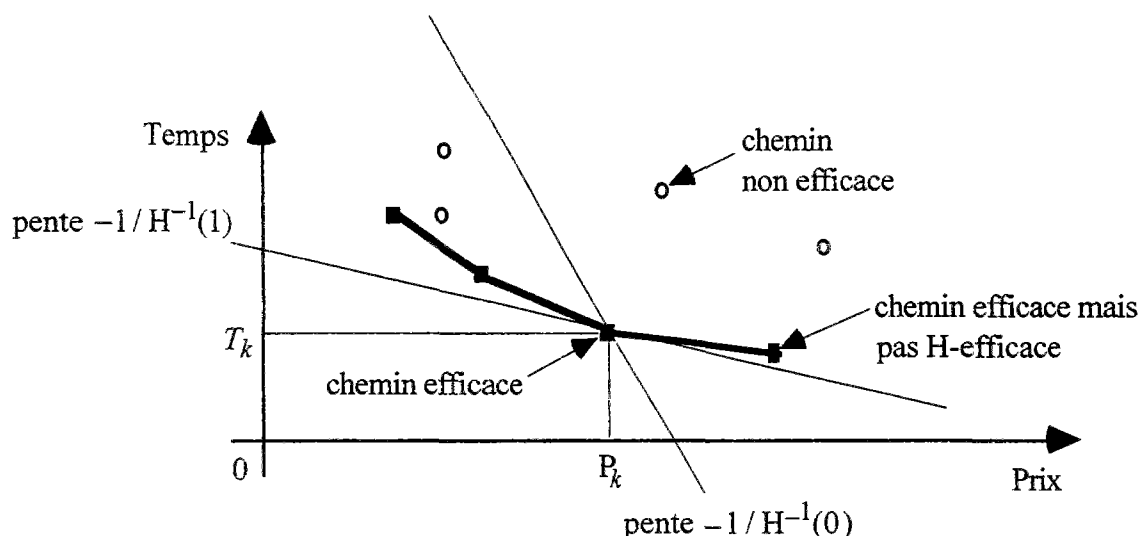
$$\text{Un chemin } k \text{ est efficace si } \exists v > 0 \quad \forall k' \quad G_k(v) \leq G_{k'}(v). \quad (8.19)$$

Cette définition de l'efficacité est générale, trop générale si les valeurs du temps à servir se situent dans un sous-ensemble strict de \mathfrak{R}_+ . Soit $\Omega = [H^{-1}(0); H^{-1}(1)]$ le support de la distribution des valeurs du temps sur la relation origine-destination, et soit $\Omega^* = \Omega \setminus \{0\}$. Restreignons la définition (8.19) :

$$\text{Un chemin } k \text{ est H-efficace si } \exists v \in \Omega^* \quad \forall k' \quad G_k(v) \leq G_{k'}(v). \quad (8.20)$$

A l'équilibre bicritère, le volume de demande q se répartit uniquement entre les chemins H-efficaces, les autres ne reçoivent pas de flot. (8.20) induit un principe de discrimination : si un chemin k domine un autre chemin k' au sens de $T_k \leq T_{k'}$ et $P_k < P_{k'}$, alors k' n'est pas H-efficace. Mais (8.20) est plus général que la règle de domination (figure 8f).

Fig. 8f. Les itinéraires H-efficaces.



Les itinéraires efficaces forment l'enveloppe convexe des points $(P_k; T_k)$ dans un diagramme prix-temps. Entre deux chemins consécutifs sur cette enveloppe, la valeur de coupure offerte v^o vérifie $\Delta P + v^o \Delta T = 0$ donc $v^o = -\Delta P / \Delta T$ est l'opposé de l'inverse de la pente du segment qui relie les deux points. En tirant à partir d'un point efficace les droites de pentes respectives $-1/H^{-1}(1)$ et $-1/H^{-1}(0)$, on met en évidence les points qui ne sont pas H-efficaces (même si certains sont efficaces).

Comment déterminer les itinéraires H-efficaces ? Au plan théorique, la définition (8.20) suffit : mais elle est peu maniable puisqu'elle porte sur chaque valeur du temps, or celles-ci sont en nombre infini ! Aussi définissons-nous d'autres concepts afin de caractériser la H-efficacité.

Définition des classes isoprix. Nous regroupons tous les chemins concurrents de même prix P_k dans une classe isoprix notée m , de prix $\bar{P}_m = P_k$ pour tout chemin k dans m . Nous associons à cette classe le temps

$$\bar{T}_m \equiv \min_{k \in m} T_k. \quad (8.21)$$

Les \bar{m} classes isoprix forment une partition de l'ensemble des chemins concurrents ⁽¹⁾. Nous les rangeons dans l'ordre des prix \bar{P}_m croissants et nous les numérotons par m allant de 1 à \bar{m} : donc $\bar{P}_1 < \dots < \bar{P}_m < \dots < \bar{P}_{\bar{m}}$. Définissons encore le temps généralisé de la classe m pour un déplacement de valeur du temps v : $\bar{G}_m(v) \equiv \bar{T}_m + \bar{P}_m/v$. Soit f_k le flot du chemin k et soit q_m le volume affecté à la classe m , donc $q_m = \sum_{k \in m} f_k$.

La répartition bicritère concerne intrinsèquement les classes isoprix plutôt que les chemins : car à l'intérieur d'une classe isoprix le principe d'optimalité exclut les chemins de temps non minimum, mais il ne régit pas le partage entre deux itinéraires de même prix et de même temps minimum.

Pour caractériser la H-efficacité d'une classe isoprix m , nous définissons son domaine de H-efficacité, l'ensemble H-efficace E_m tel que

$$E_m \equiv \{v \in \Omega^* ; \forall n \bar{G}_m(v) \leq \bar{G}_n(v)\}. \quad (8.22)$$

ainsi que l'ensemble H-efficace "cumulé" E'_m tel que

$$E'_m \equiv \bigcup_{n \leq m} E_n. \quad (8.23)$$

E'_m regroupe les valeurs du temps a priori intéressées par les chemins de prix inférieur ou égal à \bar{P}_m , tandis que $\Omega^* \setminus E'_m$ regroupe les valeurs du temps a priori intéressées par les chemins de prix strictement supérieur à \bar{P}_m .

La valeur du temps de coupure offerte par la classe isoprix m est définie comme

$$v_m^o \equiv \sup E'_m. \quad (8.24)$$

Par convention, si $E'_m = \emptyset$ nous attribuons à v_m^o la valeur $H^{-1}(0)$.

Soit $e^-(m)$ la classe isoprix H-efficace de prix immédiatement inférieur ou égal à \bar{P}_m .

Propriétés des ensembles H-efficaces : on vérifie facilement que

- (i) $\Omega^* = \bigcup_m E_m = \bigcup_m \text{H-eff } E_m$.
- (ii) $E_m = \emptyset$ si et seulement si $\forall v \in \Omega^*, \exists n \neq m, \bar{G}_m(v) > \bar{G}_n(v)$.
- (iii) E_m est soit vide soit convexe ⁽²⁾.

¹pour une relation origine-destination donnée, comme dans tout ce paragraphe et le suivant

²Soient deux points v_0, v_1 dans E_m , donc $v_0 \bar{G}_m(v_0) = \min_n v_0 \bar{G}_n(v_0)$ et $v_1 \bar{G}_m(v_1) = \min_n v_1 \bar{G}_n(v_1)$. Pour tout point $v_\alpha = \alpha v_0 + (1-\alpha)v_1$ de l'intervalle $[v_0; v_1]$ (donc $\alpha \in [0;1]$), on vérifie $\forall n$ que $v_\alpha \bar{G}_m(v_\alpha) = \alpha v_0 \bar{G}_m(v_0) + (1-\alpha)v_1 \bar{G}_m(v_1) \leq \alpha v_0 \bar{G}_n(v_0) + (1-\alpha)v_1 \bar{G}_n(v_1) = v_\alpha \bar{G}_n(v_\alpha)$ donc $v_\alpha \in E_m$

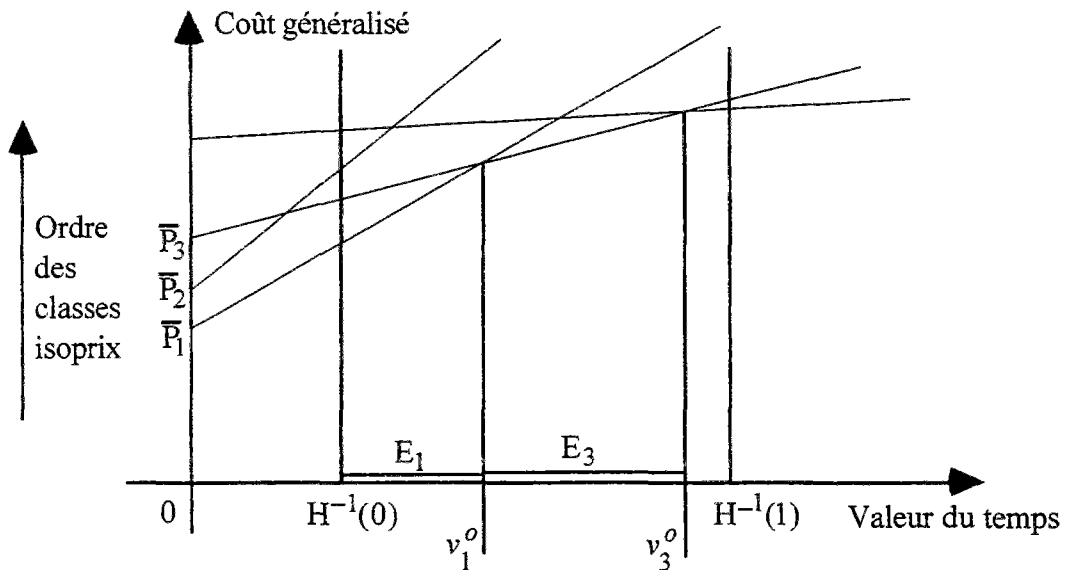
(iv) Si $\bar{P}_m < \bar{P}_n$ et E_m et E_n sont non vides, alors $\sup E_m \leq \inf E_n$ (1) : donc $E_m \cap E_n$ avec $m \neq n$ est soit vide soit réduit à un singleton.

(v) Entre deux classes H-efficaces consécutives $m < n$, $\sup E_m = \inf E_n = \frac{\bar{P}_m - \bar{P}_n}{\bar{T}_n - \bar{T}_m}$.

(vi) Si $E'_m \neq \emptyset$ alors $v_m^o = \sup E_{e^-(m)}$.

(vii) Si $E'_m \neq \emptyset$ alors $\sup (E'_m \setminus E_{e^-(m)}) \leq \inf E_{e^-(m)}$ avec égalité sauf peut-être pour la plus petite classe H-efficace.

Fig. 8g. Les ensembles H-efficaces.



Sur un diagramme des coûts généralisés $\bar{G}'_m(v) = \bar{P}_m + v\bar{T}_m$ en fonction de la valeur du temps, on repère graphiquement les valeurs du temps de coupure offertes, les ensembles H-efficaces E_m et les agrégats E'_m .

La propriété (vi) des ensembles H-efficaces implique que la classe m suffit pour représenter tous les chemins de prix inférieur ou égal à \bar{P}_m dans la concurrence avec les itinéraires de prix strictement supérieur. Elle se complète par la propriété (vii) qui découle également de la propriété (iv).

Toutes ces propriétés géométriques des ensembles E_m et E'_m découlent ultimement de la propriété d'optimalité d'un chemin pour une valeur du temps donnée. Pour une

¹ par convexité, si E_m et E_n sont non vides ce sont des intervalles. Dans ce cas $\forall u \in E_m$ $\bar{G}_m(u) \leq \bar{G}_n(u)$ donc $(\bar{P}_m - \bar{P}_n) / u \leq \bar{T}_n - \bar{T}_m$, tandis que $\forall v \in E_n$ $\bar{G}_m(v) \geq \bar{G}_n(v)$ donc $(\bar{P}_m - \bar{P}_n) / v \geq \bar{T}_n - \bar{T}_m$. En combinant il vient $(\bar{P}_m - \bar{P}_n) / u \leq \bar{T}_n - \bar{T}_m \leq (\bar{P}_m - \bar{P}_n) / v$.

Si $n > m$ i.e. $\bar{P}_n > \bar{P}_m$, cela entraîne $u \leq v$ donc $\sup E_m \leq \frac{\bar{P}_m - \bar{P}_n}{\bar{T}_n - \bar{T}_m} \leq \inf E_n$ ce qui limite l'intersection à au plus un point

distribution H fixe donc un ensemble Ω fixe, elles résument les potentialités de l'offre qui résultent des \bar{k} couples $(T_k; P_k)_k$ via les \bar{m} couples $(\bar{T}_m; \bar{P}_m)_m$.

Définissons maintenant les valeurs de coupure demandées. Soit $\mathbf{q} = (q_m)_m$, vérifiant $q_m \geq 0$ et $\sum_m q_m = q$, une répartition du volume de demande q entre les \bar{m} classes isoprix. Notons $q'_m \equiv \sum_{n \leq m} q_n$ le volume agrégé des m premières classes isoprix. Posons à nouveau le *postulat fondamental* : en l'absence d'autre information, q_m rassemble les valeurs du temps de l'intervalle $[H^{-1}(\frac{q'_{m-1}}{q}); H^{-1}(\frac{q'_m}{q})]$.

Nous définissons la m -ème *valeur de coupure demandée* par

$$v_m^d \equiv H^{-1}(\frac{q'_m}{q}). \quad (8.25)$$

Un équilibre bicritère entre classes isoprix correspond à l'affectation de chaque valeur du temps dans Ω à la classe qui lui procure un coût généralisé minimal. Alors les valeurs du temps d'un ensemble H -efficace E_m ont leurs masses agrégées dans q_m , d'où $q_m \geq q \int_{E_m} dH(v)$. Comme $\sum_m q_m = q$ et $\bigcup_m E_m = \Omega^*$, nécessairement à l'équilibre

$$\forall m, q_m = q \int_{E_m} dH(v) \quad (8.26)$$

qui est la **propriété fondamentale d'un équilibre bicritère entre classes isoprix**.

Cette propriété fondamentale, complétée par le postulat fondamental, *caractérise complètement* un équilibre bicritère ! (8.26) signifie précisément que l'ensemble E_m a la même mesure (par rapport à dH) que la plage de valeurs du temps attribuée à la classe m en vertu du postulat fondamental.

Comme $E'_m = \bigcup_{n \leq m} E_n$ et les E_n sont disjoints (hormis peut-être aux points frontières), (8.26) équivaut à

$$\forall m, q'_m = q \int_{E'_m} dH(v), \quad (8.27)$$

qui à son tour équivaut à

$$\forall m, q'_m = q H(\sup E'_m) = q H(v_m^o). \quad (8.28)$$

Si $q > 0$, en recourant aux définitions (8.25) et (8.24), (8.28) équivaut à

$$\forall m, v_m^d = v_m^o. \quad (8.29)$$

Par rapport au réseau à deux arcs, les définitions sophistiquées du cas général aboutissent à des égalités entre les valeurs de coupure offertes et demandées, qu'il y ait ou non du trafic sur les classes isoprix.

8.2.3.2 Formule caractéristique à temps et volume fixes

Il nous reste à rendre commode, opératoire, la formule caractéristique (8.26) : nous voulons caractériser l'équilibre bicritère sans recourir explicitement aux variables

ensemblistes E_m , et sous la forme d'un problème type de recherche opérationnelle comme nous l'avons fait pour le réseau à deux arcs.

Etablissons donc une fonction d'impédance pour chaque classe isoprix, suivant nos contributions antérieures (Leurent, 1993a, 1996a). Pour cela, soit $e^+(m)$ la classe isoprix H-efficace (i.e. $E_{e^+(m)} \neq \emptyset$) de prix immédiatement strictement supérieur à \bar{P}_m .

Donc $\forall m$, si $e^-(m) > 1$ alors $e^+(e^-(m) - 1) = e^-(m)$.

A l'équilibre, pour toute classe n H-efficace sauf la dernière $\bar{e} = e^-(\bar{m})$, (8.29) jointe à la propriété (v) des ensembles H-efficaces entraîne que $v_n^d = (\bar{P}_{e^+(n)} - \bar{P}_n) / (\bar{T}_n - \bar{T}_{e^+(n)})$, donc $\bar{T}_n + (\bar{P}_n - \bar{P}_{e^+(n)}) / v_n^d = \bar{T}_{e^+(n)}$.

En sommant de proche en proche sur les classes H-efficaces, on établit que

$$\forall n \text{ H-efficace, } \bar{T}_n + \sum_{l \text{ H-eff, } l=n}^{e^-(\bar{e}-1)} (\bar{P}_l - \bar{P}_{e^+(l)}) / v_l^d = \bar{T}_{\bar{e}},$$

valable même pour $n = \bar{e}$ en convenant qu'alors la somme des termes $\Delta P / v_l^d$ est nulle.

Or on peut inclure les classes non H-efficaces dans la sommation, car entre chaque l et $e^+(l)$ on vérifie $\sum_{m=l}^{e^+(l)} (\bar{P}_m - \bar{P}_{m+1}) / v_m^d = \sum_{m=l}^{e^+(l)} (\bar{P}_m - \bar{P}_{m+1}) / v_l^d = (\bar{P}_l - \bar{P}_{e^+(l)}) / v_l^d$. Ainsi, toujours à l'équilibre,

$$\forall n \text{ H-eff, } \bar{T}_n + \sum_{m=n}^{\bar{m}-1} (\bar{P}_m - \bar{P}_{m+1}) / v_m^d = \text{Constante.}$$

D'où l'idée d'associer à chaque classe isoprix n , efficace ou non, la fonction d'impédance

$$\bar{I}_n \equiv \bar{T}_n + \sum_{m=n}^{\bar{m}-1} (\bar{P}_m - \bar{P}_{m+1}) / v_m^d \tag{8.30}$$

qui dépend uniquement des temps \bar{T}_n , des prix \bar{P}_m et du vecteur $(q_m)_m$.

Par construction, à l'équilibre les impédances de deux classes H-efficaces sont égales. Toujours à l'équilibre, on montre que l'impédance d'une classe non H-efficace dépasse l'impédance commune des classes H-efficaces ⁽¹⁾.

¹ Montrons que toute classe non H-efficace j présente une impédance supérieure ou égale à celle d'une classe H-efficace. Si $q'_j > 0$, soit $m = e^-(j)$ la classe efficace telle que $q'_j = q'_m$. $\forall i \in [m; j-1]$, $v_i^d = v_m^d$ donc $\sum_{i=m}^{j-1} \frac{\bar{P}_i - \bar{P}_{i+1}}{v_i^d} = \frac{\bar{P}_m - \bar{P}_j}{v_m^d}$ donc encore $\bar{I}_j - \bar{I}_m = \bar{T}_j - \bar{T}_m + (\bar{P}_m - \bar{P}_j) / v_m^d = \bar{G}_j(v_m^d) - \bar{G}_m(v_m^d)$. Comme m est H-efficace, $v_m^d = v_m^o \in E_m$ donc $\bar{I}_j - \bar{I}_m \geq 0$. Si $q'_j = 0$, donc $v_j^d = H^{-1}(0)$, considérons un nombre ε qui tend vers

Théorème (1) : formule caractéristique pour un équilibre bicritère à demande et temps fixes. La répartition entre classes isoprix $(q_m)_m$ est la trace d'un équilibre bicritère pour les attributs $(\bar{T}_m; \bar{P}_m)_m$ si et seulement si

$$\forall m, q_m > 0 \Rightarrow \bar{I}_m = \min_n \bar{I}_n. \quad (8.31)$$

$H^{-1}(0)$ par valeurs supérieures : donc $\varepsilon > 0$. Soit m la première classe H-efficace ; s'il est suffisamment petit $\varepsilon \in E_m$ donc $(\bar{T}_j - \bar{T}_m)\varepsilon + \bar{P}_j - \bar{P}_m \geq 0$. Or

$$\bar{I}_j - \bar{I}_m = \lim_{\varepsilon \rightarrow H^{-1}(0)^+} \bar{T}_j - \bar{T}_m + \sum_{i=j}^{m-1} \frac{\bar{P}_i - \bar{P}_{i+1}}{\varepsilon} = \lim_{\varepsilon \rightarrow H^{-1}(0)^+} \frac{(\bar{T}_j - \bar{T}_m)\varepsilon + \bar{P}_j - \bar{P}_m}{\varepsilon} \geq 0.$$

¹Nous avons déjà montré le sens direct. En sens réciproque, montrons qu'une répartition qui satisfait (8.31) vérifie $\forall m \in S = \{m; q_m > 0\}$, $O_m \equiv [H^{-1}(\frac{q'_m-1}{q}); H^{-1}(\frac{q'_m}{q})] \subset E_m$ où les E_m découlent des prix \bar{P}_m et des temps \bar{T}_m . Cette propriété [$O_m \subset E_m$] entraîne pour $m \in S$ que $q_m = q \int_{O_m} dH \leq q \int_{E_m} dH$. Or $\Omega^* \subset \bigcup_{m \in S} O_m \subset \bigcup_m E_m = \Omega^*$ donc $q \leq \sum_{m \in S} q_m = q \sum_{m \in S} \int_{O_m} dH \leq q \sum_{m \in S} \int_{E_m} dH \leq q$, i.e. partout égalité donc $\forall m \in S$, $q_m = q \int_{E_m} dH$. De plus, $q = q \sum_m \int_{E_m} dH$ donc $\forall m \notin S$, $q \int_{E_m} dH = 0$ en accord avec $q_m = 0$ par définition de S. Au total, pour toute classe m $q_m = q \int_{E_m} dH$, donc (8.26).

Montrons donc la propriété [$O_m \subset E_m$]. Notons d'abord $\Omega^* \subset \bigcup_{m \in S} O_m$ donc $\forall v \in \Omega^*$, $\exists m_v \in S$, $v \in O_{m_v}$. Pour montrer que v appartient à E_{m_v} , montrons que pour tout n on vérifie $\bar{G}_n(v) \geq \bar{G}_{m_v}(v)$. Soit $n > m_v$: alors

$$\begin{aligned} \bar{G}_n(v) - \bar{G}_{m_v}(v) &= \bar{I}_n - \bar{I}_{m_v} + \left(\sum_{l=m_v}^{n-1} (\bar{P}_l - \bar{P}_{l+1}) / v_l^d \right) + (\bar{P}_n - \bar{P}_{m_v}) / v \\ &\geq \bar{I}_n - \bar{I}_{m_v} + \left(\sum_{l=m_v}^{n-1} (\bar{P}_l - \bar{P}_{l+1}) / v_{m_v}^d \right) + (\bar{P}_n - \bar{P}_{m_v}) / v \text{ car } 0 \leq v_{m_v}^d \leq v_l^d \\ &\geq \bar{I}_n - \bar{I}_{m_v} + (\bar{P}_n - \bar{P}_{m_v})(1/v - 1/v_{m_v}^d) \geq 0 \text{ comme somme de deux termes non négatifs.} \end{aligned}$$

$$\text{Sinon } n < m_v : \bar{G}_n(v) - \bar{G}_{m_v}(v) = \bar{I}_n - \bar{I}_{m_v} - \left(\sum_{l=n}^{m_v-1} (\bar{P}_l - \bar{P}_{l+1}) / v_l^d \right) + (\bar{P}_n - \bar{P}_{m_v}) / v$$

$$\geq \bar{I}_n - \bar{I}_{m_v} - \left(\sum_{l=n}^{m_v-1} (\bar{P}_l - \bar{P}_{l+1}) / v_{m_v-1}^d \right) + (\bar{P}_n - \bar{P}_{m_v}) / v \text{ car } 0 \leq v_l^d \leq v_{m_v-1}^d$$

$$\geq \bar{I}_n - \bar{I}_{m_v} + (\bar{P}_n - \bar{P}_{m_v})(1/v - 1/v_{m_v-1}^d) \geq 0 \text{ comme somme de deux termes non négatifs.}$$

Ainsi $\bar{G}_n(v) - \bar{G}_{m_v}(v) \geq 0 \quad \forall n$, donc $v \in E_{m_v}$ et $O_{m_v} \subset E_{m_v}$.

Théorème (1) : NCP pour un équilibre bicritère entre classes isoprix. Soit $\mathbf{u} = {}^t[\mathbf{q}; \lambda_+; \lambda_-] \in (\mathfrak{R}_+)^{\bar{m}+2}$ et \mathbf{F} l'application qui à \mathbf{u} associe $\mathbf{F}(\mathbf{u})$ dont les \bar{m} premières composantes valent $\bar{I}_m - \lambda_+ + \lambda_-$ et les deux dernières $(\sum_m q_m) - q$ et $q - \sum_m q_m$ respectivement. Alors toute solution $\hat{\mathbf{u}} = {}^t[\hat{\mathbf{q}}; \hat{\lambda}_+; \hat{\lambda}_-]$ du programme de complémentarité non linéaire

$$\mathbf{u} \geq \mathbf{0}, \mathbf{F}(\mathbf{u}) \geq \mathbf{0} \text{ et } \mathbf{u} \cdot \mathbf{F}(\mathbf{u}) = \mathbf{0} \quad (8.33)$$

vérifie que $\hat{\mathbf{q}}$ est un équilibre bicritère. Réciproquement, tout équilibre bicritère $\hat{\mathbf{q}}$ associé à $\hat{\lambda}_+ \equiv \max\{0; \min_m \bar{I}_m(\hat{\mathbf{q}})\}$ et $\hat{\lambda}_- \equiv \max\{0; -\min_m \bar{I}_m(\hat{\mathbf{q}})\}$ résout le NCP (8.33).

Nous avons dédoublé le multiplicateur de Lagrange $\lambda \equiv \lambda_+ - \lambda_-$ en deux réels λ_+ et λ_- non négatifs, pour couvrir les cas où $\min_m \bar{I}_m$ prendrait des valeurs négatives : en effet la définition (8.30) n'implique pas $\min_m \bar{I}_m \geq 0$, contrairement au cas du réseau à deux arcs.

Comment redescendre au niveau des chemins ? Pour un chemin k à l'intérieur d'une classe isoprix $m(k)$, si à l'équilibre $T_k > \bar{T}_{m(k)}$ alors $f_k = 0$. Donc on peut compléter \mathbf{u} en lui ajoutant autant de composantes f_k et μ_m que de chemins et de classes, respectivement, et incorporer à \mathbf{F} des composantes qui correspondent aux nouvelles variables selon $F_{fk} \equiv T_k - \mu_{m(k)}$ et $F_{\mu m} \equiv (\sum_{k \in m} f_k) - q_m$. Alors à l'équilibre si $f_k > 0$ on vérifie $T_k = \mu_{m(k)} = \min_{k' \in m(k)} T_{k'} = \bar{T}_{m(k)}$ par définition de $\bar{T}_{m(k)}$.

Ainsi l'on obtient un NCP pour l'équilibre bicritère en fonction des chemins. Un tel NCP mobilise davantage de composantes qu'il n'est strictement nécessaire puisque les μ_m et les $F_{\mu m}$ n'apportent pas d'information mathématique autre que des définitions ! Nous préférons le simplifier en définissant une fonction d'impédance par chemin

$$I_k \equiv T_k + \sum_{n=m(k)}^{\bar{m}-1} (\bar{P}_n - \bar{P}_{n+1}) / H^{-1}(\frac{q'_n}{q}), \text{ avec } q'_n \equiv \sum_{k; P_k \leq \bar{P}_n} f_k. \quad (8.34)$$

Ainsi l'équilibre bicritère en fonction des chemins résout le NCP défini sur $(\mathfrak{R}_+)^{\bar{k}+2}$, pour l'application \mathbf{F} suivante

$$\mathbf{u} = \begin{bmatrix} \mathbf{f} = (f_k)_{k=1.. \bar{k}} \\ \lambda_+ \\ \lambda_- \end{bmatrix} \mapsto \mathbf{F}(\mathbf{u}) = \begin{bmatrix} (I_k - \lambda_+ + \lambda_-)_{k=1.. \bar{k}} \\ (\sum_k f_k) - q \\ q - \sum_k f_k \end{bmatrix}. \quad (8.35)$$

Si, à l'équilibre, une classe H-efficace m comprend au moins deux chemins de temps minimal, alors la solution de (8.35) n'est pas unique, puisque toute répartition de $q_m > 0$ entre les différents chemins $k \in m$ de temps minimal est une solution.

¹*Preuve.* La réciproque est triviale. En sens direct, le cas $q=0$ est trivial. Si $q > 0$, la solution $\hat{\mathbf{u}}$ du NCP vérifie $\exists m, q_m > 0$, donc $\bar{I}_m = \hat{\lambda}_+ - \hat{\lambda}_- \equiv \hat{\lambda}$ et même $\hat{\lambda} = \min_m \bar{I}_m$ car $\mathbf{F}(\mathbf{u}) \geq \mathbf{0}$. De $\mathbf{u} \cdot \mathbf{F}(\mathbf{u}) = \mathbf{0}$ on tire $\forall m, q_m > 0 \Rightarrow \bar{I}_m = \hat{\lambda} = \min_m \bar{I}_m$ i.e. $\hat{\mathbf{q}}$ vérifie (8.31)

Au total, nous avons établi une formule caractéristique rigoureuse et commode pour la répartition bicritère à temps et demande fixes, analogue à celle donnée pour le réseau à deux arcs (Cf. §8.2.1.3). Les seules variables "vraiment" exogènes dans (8.35) sont la distribution H et les prix P_k qui régissent la définition des q_m . Les variables "apparemment exogènes" q et T_k peuvent devenir endogènes, ce qui permet d'étendre la formule caractéristique à la demande élastique et aux temps variables.

8.2.3.3 Représenter la demande élastique

Pour les mêmes raisons qu'au §8.2.2.3, nous supposons que le volume de demande peut varier en fonction de l'offre via une fonction de demande D :

$$q = D(\bar{G}), \quad (8.36)$$

avec
$$\bar{G} = \int \min_k \{T_k + P_k / v\} dH(v).$$

Comment exprimer le temps généralisé composite \bar{G} sous une forme explicite par rapport aux P_k et aux T_k ? Il vient

$$\begin{aligned} \bar{G} &= \int_{\Omega^*} (\min_k G_k(v)) dH(v) = \int_{\Omega^*} (\min_n \bar{G}_n(v)) dH(v) \\ &= \sum_m \int_{v_{m-1}^d}^{v_m^d} (\min_n \bar{G}_n(v)) dH(v) \text{ puisque } \Omega^* = \bigcup_m [H^{-1}(\frac{q'_{m-1}}{q}); H^{-1}(\frac{q'_m}{q})]. \end{aligned}$$

Comme $\forall v \in [v_{m-1}^d; v_m^d]$, $\min_n \bar{G}_n(v) \leq \bar{G}_m(v)$, nécessairement

$$\bar{G} \leq \sum_m \int_{v_{m-1}^d}^{v_m^d} \bar{G}_m(v) dH(v) \equiv \check{G} \quad (8.37)$$

avec égalité à l'équilibre bicritère, grâce au postulat fondamental. On reformule

$$\check{G} = \sum_m \bar{T}_m \frac{q_m}{q} + \bar{P}_m \left(\eta\left(\frac{q'_m}{q}\right) - \eta\left(\frac{q'_{m-1}}{q}\right) \right) \quad (8.38)$$

où la fonction η , définie sur $[0;1]$, est une primitive de $1/H^{-1}$: par le changement de variable $v = H(\alpha)$ en effet, $\int_{H^{-1}(x)}^{H^{-1}(y)} \frac{dH(v)}{v} = \int_x^y \frac{d\alpha}{H^{-1}(\alpha)}$. Nous notons désormais η' au lieu de $1/H^{-1}$.

Grâce à (8.38), \check{G} apparaît comme une fonction explicite des T_k et des P_k via les classes isoprix m donc les \bar{T}_m , \bar{P}_m , q_m et q'_m . De fait \bar{G} et \check{G} sont des fonctions seulement des parts de marché relatives q_m / q : indépendamment du volume total q donc indépendamment de (8.36), à l'équilibre bicritère $\bar{G} = \check{G}$ donc on peut substituer \check{G} à \bar{G} dans l'équation (8.36) dès qu'on la combine aux conditions (8.35).

Nous pouvons rendre la fonction de demande (8.36) intrinsèque au NCP en modifiant

dans (8.35) ⁽¹⁾ l'application \mathbf{F}

$$\mathbf{u} = \begin{bmatrix} \mathbf{f} = (f_k)_{k=1..\bar{k}} \\ \lambda_+ \\ \lambda_- \end{bmatrix} \mapsto \mathbf{F}(\mathbf{u}) = \begin{bmatrix} (I_k - \lambda_+ + \lambda_-)_{k=1..\bar{k}} \\ (\sum_k f_k) - D(\check{G}(\mathbf{f})) \\ D(\check{G}(\mathbf{f})) - \sum_k f_k \end{bmatrix}. \quad (8.39)$$

Montrons que le NCP basé sur (8.39) admet pour solutions les équilibres bicritères à demande variable et eux seulement. En sens direct, soit $\mathbf{u} = {}^t[\mathbf{f}; \lambda_+; \lambda_-]$ une solution du NCP : la non négativité des deux dernières composantes de \mathbf{F} fait que $\sum_k f_k = D(\check{G})$, tandis que les \bar{k} premières composantes impliquent l'équilibre bicritère donc $\bar{G} = \check{G}$ donc la condition de volume est satisfaite. Le sens réciproque est trivial : comme un équilibre bicritère satisfait (8.36), on lui associe $\hat{\lambda}_+ \geq 0$ et $\hat{\lambda}_- \geq 0$ tels que $\hat{\lambda}_+ - \hat{\lambda}_- = \min_k I_k$ pour obtenir une solution de (8.39).

8.2.3.4 Plusieurs relations origine-destination sur le réseau

A temps fixes, l'équilibre bicritère à demande élastique est intrinsèque à une relation origine-destination : pour traiter plusieurs relations O-D il suffit de traiter autant de NCP de forme (8.39) que de relations, ou de façon rigoureusement équivalente un NCP global dans lequel on individualise chaque relation par un identifiant i . On adapte alors les notations : chaque répartition O-D du flot par chemins devient $\mathbf{f}_i = (f_{ik})_{k=1..\bar{k}_i}$, les prix et les temps des chemins deviennent P_{ik} et T_{ik} , de même que les classes isoprix et les multiplicateurs associés λ_{i+} et λ_{i-} . Enfin la distribution des valeurs du temps peut varier parmi les relations O-D, ce qui donne lieu à des notations individuelles H_i , η_i et η'_i , de même que les temps généralisés $G_{ik}(v)$, $\bar{G}_{im}(v)$, \check{G}_i et que la fonction de demande D_i .

Une telle conjonction d'équilibres particuliers peut évidemment avoir lieu vis-à-vis d'autres critères de segmentation que la relation O-D traitée ici : en particulier le type de véhicule, le motif de déplacement etc. sans problème tant que la segmentation demeure discrète (i.e. avec un nombre fini de segments).

8.2.3.5 Itinéraires composés, temps variables, contraintes adjointes

Jusqu'à présent nous avons considéré des temps d'itinéraire sans dépendance explicite vis-à-vis des flots. Or, sur un réseau de transport concret, les temps varient en fonction du trafic. Comment représenter cette dépendance ? On peut exprimer les temps T_{ik} des chemins k des relations O-D i comme des fonctions explicites

$$T_{ik} = t_{ik}(\mathbf{f}) \quad (8.40)$$

du vecteur de flots par chemins $\mathbf{f} = (f_{ik})_{ik}$: il suffit de substituer $t_{ik}(\mathbf{f})$ à T_{ik} dans le NCP (8.39).

Précisons la dépendance (8.40). Deux chemins composés d'une majorité de tronçons

¹dans I_k on remplace aussi q par $\sum_k f_k$

communs présentent certainement de fortes ressemblances dans leurs fonctions de temps. Aussi exprime-t-on usuellement le temps d'un chemin comme la somme des temps des tronçons a qui le composent, soit

$$T_{ik} = \sum_{a \in k} T_a = \sum_a \delta_a^{ik} T_a \quad (8.41)$$

avec δ_a^{ik} un indicateur égal à 1 si le tronçon a est incident à k ou sinon égal à 0.

Si les déplacements des différentes valeurs du temps et des différentes relations O-D ont tous le même "comportement d'écoulement" sur chaque tronçon, leur influence sur les temps se manifeste uniquement au travers des agrégats "flots par arcs"

$$x_a(\mathbf{f}) \equiv \sum_k \text{passe par } a f_{ik} = \sum_{ik} \delta_a^{ik} f_{ik} \quad (8.42)$$

Considérons le vecteur des flots par arcs $\mathbf{x}(\mathbf{f}) = (x_a(\mathbf{f}))_a$: il est relié à \mathbf{f} par $\mathbf{x} \equiv \Delta \cdot \mathbf{f}$ avec $\Delta \equiv [\delta_a^{ik}]_{a,ik}$ la matrice d'incidence des arcs aux chemins. Sous l'hypothèse de conformité d'écoulement des déplacements, on vérifie que

$$T_a(\mathbf{f}) = t_a(\mathbf{x}(\mathbf{f})) \quad (8.43)$$

ce qui simplifie grandement la spécification des T_a . En pratique, seuls les flots de l'arc a et de certains arcs voisins agissent sur le temps T_a : on se contente souvent de

$$T_a(\mathbf{f}) = t_a(x_a(\mathbf{f})), \quad (8.44)$$

réduction qui donne de "bonnes" propriétés mathématiques au NCP d'affectation.

Cependant des fonctions explicites $t_{ik}(\mathbf{f})$ ou $t_a(x_a)$ ne rendent pas compte de toutes les interactions possibles entre les flots sur le réseau. L'adjonction de contraintes $z_b(\mathbf{f}) \leq 0$ permet d'aller plus loin, afin notamment de représenter des contraintes de capacité (en spécifiant $z_b(\mathbf{f}) \equiv x_a(\mathbf{f}) - x_a^{\max} \leq 0$).

Associons à chaque contrainte adjointe b un multiplicateur w_b tel que

$$z_b(\mathbf{f}) \leq 0, w_b \geq 0 \text{ et } w_b \cdot z_b(\mathbf{f}) = 0 \quad (8.45)$$

et notons $\mathbf{w} = (w_b)_b$. Le temps de parcours d'un chemin k devient

$$T_{ik}(\mathbf{f}; \mathbf{w}) = t_{ik}(\mathbf{f}) + \sum_b w_b \cdot \frac{\partial z_b}{\partial f_{ik}} \quad (8.46)$$

L'équilibre bicritère à demande élastique, temps variables et contraintes adjointes est donc la solution d'un NCP qui injecte (8.46) dans (8.39) et ajoute (8.45). Ce NCP a pour variables endogènes $\mathbf{f} \equiv (f_{ik})_{i,k=1..k_i}$, $\lambda_+ \equiv (\lambda_{i+})_i$, $\lambda_- \equiv (\lambda_{i-})_i$, $\mathbf{w} \equiv (w_b)_b$.

Terminons par deux remarques. Primo, au stade formel il est possible mais pas nécessaire d'agréger par arcs les flots par chemins : on peut parfaitement agréger moins, par exemple avec des flots par classes de véhicule et par arc. Secundo, à propos de la définition des chemins : dans tout modèle d'affectation il convient de la fixer précisément (au travers de la matrice d'incidence Δ des arcs aux chemins). Souvent on retient tous les

chemins sans boucles, auquel cas il faut encore préciser si cela exclut seulement les chemins qui passent au moins deux fois par un arc, ou plus généralement tous les chemins qui passent au moins deux fois par un noeud.

8.2.4 Propriétés remarquables des formules caractéristiques

Nous avons caractérisé l'équilibre bicritère à demande élastique, temps variables et contraintes adjointes par un problème de complémentarité non linéaire (NCP). Afin d'en déduire des réponses aux questions d'existence et d'unicité, nous allons adapter marginalement cette formule caractéristique, en "cachant" les variables endogènes duales λ_{i+} , λ_{i-} et w_b .

Nous commençons par récapituler le NCP et indiquer une inéquation variationnelle équivalente (§8.2.4.1). Puis nous "réduisons" l'inéquation variationnelle en reformulant la condition sur le volume de demande et en restreignant le domaine de définition (§8.2.4.2). Alors nous établissons un programme extrémal équivalent dans des cas particuliers importants (§8.2.4.3). Ensuite nous étudions l'existence qui ne présente pas de spécificité par rapport au modèle monocritère (§8.2.4.4), et l'unicité qui constitue une propriété remarquable du modèle bicritère : sur une relation O-D donnée, le volume total se répartit de façon unique entre les classes isoprix (§8.2.4.5).

8.2.4.1 Synthèse des NCP ; l'inéquation variationnelle équivalente

Théorème : NCP caractéristique de l'affectation bicritère à demande élastique, temps variables et contraintes adjointes. Soit pour variables exogènes (i) les fonctions de répartition H_i des valeurs du temps, (ii) les fonctions de demande D_i , (iii) les prix P_{ik} des chemins, (iv) les fonctions de temps de parcours $t_{ik}(\mathbf{f})$, (v) les contraintes adjointes $z_b(\mathbf{f}) \leq 0$. L'équilibre bicritère à demande élastique, temps variables et contraintes adjointes a pour variables endogènes le vecteur des $\bar{K} \equiv \sum_i \bar{k}_i$ flots par chemins, $\mathbf{f} = (f_{ik})_{ik}$, les deux vecteurs des \bar{i} multiplicateurs associés aux fonctions de demande, $\lambda_+ = (\lambda_{i+})_i$ et $\lambda_- = (\lambda_{i-})_i$, et le vecteur des \bar{b} multiplicateurs associés aux contraintes adjointes, $\mathbf{w} = (w_b)_b$. Il se caractérise comme une solution du NCP associé à l'application \mathbf{F} définie sur l'ensemble $\mathbf{U} \equiv (\mathfrak{R}_+)^{\bar{K}+2\bar{i}+\bar{b}}$:

$$\mathbf{u} = \begin{bmatrix} \mathbf{f} = (f_{ik})_{ik} \\ \lambda_+ = (\lambda_{i+})_i \\ \lambda_- = (\lambda_{i-})_i \\ \mathbf{w} = (w_b)_b \end{bmatrix} \mapsto \mathbf{F}(\mathbf{u}) = \begin{bmatrix} \left(T_{ik} + \left(\sum_{n=m_i(k)}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) \eta'_i \left(\frac{q_{i,n}}{q_i} \right) \right) - \lambda_{i+} + \lambda_{i-} \right)_{ik} \\ \left(\sum_{k \in i} f_{ik} \right) - D_i(\bar{G}_i)_i \\ \left(D_i(\bar{G}_i) - \sum_{k \in i} f_{ik} \right)_i \\ (-z_b(\mathbf{f}))_b \end{bmatrix} \begin{matrix} (8.47a) \\ (8.47b) \\ (8.47c) \\ (8.47d) \end{matrix}$$

soumis aux contraintes de définition

$$T_{ik} \equiv t_{ik}(\mathbf{f}) + \sum_b w_b \cdot \frac{\partial z_b}{\partial f_{ik}} \quad (8.47e)$$

$$q_i(\mathbf{f}) \equiv \sum_{k \in i} f_{ik} \quad (8.47f)$$

$$q'_{i,n}(\mathbf{f}) \equiv \sum_{k \in i; P_{i,k} \leq \bar{P}_{i,n}} f_{ik} \quad (8.47g)$$

$$\check{G}_i \equiv \sum_{n=1}^{\bar{m}_i} \frac{q'_{i,n} - q'_{i,n-1}}{q_i} (\min_{k \in n} T_{i,k}) + \bar{P}_{i,n} \left(\eta_i \left(\frac{q'_{i,n}}{q_i} \right) - \eta_i \left(\frac{q'_{i,n-1}}{q_i} \right) \right). \quad (8.47h)$$

Comme \mathbf{U} est un orthant positif, le NCP (8.47) équivaut à l'inéquation variationnelle suivante :

$$\text{"trouver } \hat{\mathbf{u}} \in \mathbf{U} \text{ tel que } \forall \mathbf{u} \in \mathbf{U}, (\mathbf{u} - \hat{\mathbf{u}}) \cdot \mathbf{F}(\hat{\mathbf{u}}) \geq \mathbf{0}." \quad (8.48)$$

Cette forme est très générale : elle englobe aussi bien le cas de la demande élastique que celui de la demande fixe (en spécifiant une fonction D_i constante). Mais la généralité se paie par l'absence d'une méthode de résolution évidente : comment chercher la solution, en ajustant toutes les variables endogènes à la fois, ou tour-à-tour les variables primales et les variables duales ?

Remarquons déjà que la forme d'inéquation variationnelle se prête à une réduction du nombre de variables λ_j . Il suffit de définir $\mathbf{U}' \equiv (\mathfrak{R}_+)^{\bar{K}} \times \mathfrak{R}^{\bar{I}} \times (\mathfrak{R}_+)^{\bar{b}}$, de remplacer dans (8.47a) $\lambda_{i+} - \lambda_{i-}$ par un multiplicateur λ_i libre de varier dans tout \mathfrak{R} , et de conserver seulement une des deux conditions sur le volume de demande $D_i(\check{G}_i)$, par exemple (8.47b).

8.2.4.2 Une inéquation variationnelle réduite

Bien qu'au stade formel nous ne traitons pas de résolution, nous désirons la préparer en proposant des formules caractéristiques alternatives, qui se prêtent à différentes méthodes de résolution. Montrons donc comment "cacher" les variables duales λ_j et w_b . Pour ces dernières, les contraintes adjointes bornent l'ensemble faisable \mathbf{U} : nous en tenons compte en restreignant celui-ci à $\mathbf{U}'' \equiv \{\mathbf{f} \in (\mathfrak{R}_+)^{\bar{K}} ; \forall b, z_b(\mathbf{f}) \leq 0\}$.

Mais comment cacher les variables λ_j et les conditions (8.47b) ? Il faut conserver l'information relative à la fonction D_i , l'exprimer en fonction des variables endogènes subsidiaires f_{ik} , la rapatrier dans les conditions (8.47a) d'où en revanche λ_i doit disparaître ! L'idée est de passer par le temps généralisé composite \check{G}_i : à l'équilibre λ_i s'en déduit puisque

$$\begin{aligned} \sum_{k \in i} f_{ik} \lambda_i &= \sum_{k \in i} f_{ik} I_{ik}, \text{ on remplace } I_{ik} \text{ par sa définition,} \\ &= \sum_{m \in i} \left(q_{i,m} \bar{T}_{i,m} - \sum_{n=m}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) \eta'_i \left(\frac{q'_{i,n}}{q_i} \right) q_{i,m} \right), \text{ donc d'après (8.38)} \\ q_i \lambda_i &= q_i \left(\check{G}_i - \bar{P}_{i,\bar{m}_i} \eta_i(1) + \bar{P}_{i,1} \eta_i(0) + \sum_{n=1}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) \left(\frac{q'_{i,n}}{q_i} \eta'_i \left(\frac{q'_{i,n}}{q_i} \right) - \eta_i \left(\frac{q'_{i,n}}{q_i} \right) \right) \right). \quad (8.49) \end{aligned}$$

Si la fonction de demande D_i est monotone, nous pouvons reformuler la condition sur le volume de demande (8.36) en appliquant à chaque membre la fonction réciproque D_i^{-1} définie par $D_i^{-1}(q) \equiv \max\{x ; q < D_i(x)\}$. Ainsi (8.36) devient

$$\tilde{G}_i = D_i^{-1}(\sum_{k \in i} f_{ik}) \quad (8.50)$$

d'où à l'équilibre, si $q_i > 0$,

$$\lambda_i = D_i^{-1}(q_i) - \bar{P}_{i, \bar{m}_i} \eta_i(1) + \bar{P}_{i,1} \eta_i(0) + \sum_{n=1}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) \left(\frac{q'_{i,n}}{q_i} \eta'_i \left(\frac{q'_{i,n}}{q_i} \right) - \eta_i \left(\frac{q'_{i,n}}{q_i} \right) \right). \quad (8.51)$$

Nous substituons cela dans (8.47a) pour définir, à l'aide de l'indicateur $\Gamma_{in}^k \equiv \partial q'_{i,n} / \partial f_{ik}$ égal à 1 si $P_{i,k} \leq \bar{P}_{i,n}$ ou égal à 0 sinon,

$$F'_{ik} \equiv T_{ik} + \bar{P}_{i, \bar{m}_i} \eta_i(1) - \bar{P}_{i,1} \eta_i(0) + \sum_{n=1}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) \left(\eta_i \left(\frac{q'_{i,n}}{q_i} \right) + \left(\Gamma_{in}^k - \frac{q'_{i,n}}{q_i} \right) \eta'_i \left(\frac{q'_{i,n}}{q_i} \right) \right) - D_i^{-1}(q_i). \quad (8.52)$$

La fonction F'_{ik} dépend encore de \mathbf{w} , ce que l'on désire éviter. Soit donc encore la fonction F''_{ik} définie comme $F''_{ik}(\mathbf{f}) \equiv$

$$t_{ik}(\mathbf{f}) + \bar{P}_{i, \bar{m}_i} \eta_i(1) - \bar{P}_{i,1} \eta_i(0) + \sum_{n=1}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) \left(\eta_i \left(\frac{q'_{i,n}}{q_i} \right) + \left(\Gamma_{in}^k - \frac{q'_{i,n}}{q_i} \right) \eta'_i \left(\frac{q'_{i,n}}{q_i} \right) \right) - D_i^{-1}(q_i). \quad (8.53)$$

L'inéquation variationnelle réduite s'énonce de la manière suivante.

Théorème : inéquation variationnelle réduite caractéristique. Si les fonctions z_b sont convexes, l'équilibre bicritère à demande élastique, temps variables et contraintes adjointes résout l'inéquation variationnelle définie sur l'ensemble \mathbf{U}'' , pour l'application \mathbf{F}'' qui à $\mathbf{f} = (f_{ik})_{ik}$ associe le vecteur $\mathbf{F}''(\mathbf{f}) = (F''_{ik}(\mathbf{f}))_{ik}$. Réciproquement, soit une solution $\hat{\mathbf{f}}$ de l'inéquation variationnelle réduite : si les ∇z_b pour les contraintes b telles que $z_b(\hat{\mathbf{f}}) = 0$ sont libres, alors il existe des variables duales λ_i réelles et w_b non négatives telles que le vecteur $(\hat{\mathbf{f}}, \lambda, \mathbf{w})$ résout l'inéquation variationnelle (8.48).

La preuve du théorème découle de résultats généraux sur les inéquations variationnelles en dimension finie : voir par exemple les énoncés de Tobin et Friesz (1988) traduits au §10.1.3.1. Leur théorème AS1 implique le sens réciproque ⁽¹⁾, tandis que le sens direct résulte du théorème AS2 ⁽²⁾.

¹sans besoin d'une hypothèse de convexité sur les z_b . On suppose que $\hat{\mathbf{f}}$ résout l'inéquation variationnelle réduite. Avec les notations de Tobin et Friesz, il n'y a que des contraintes inégalités $g_{ik}(\mathbf{f}) \equiv f_{ik} \geq 0$ et $g_b(\mathbf{f}) \equiv -z_b(\mathbf{f}) \geq 0$. Associons-leur respectivement des multiplicateurs π_{ik} et w_b . Le théorème AS1 implique $\forall i, k, F''_{ik} - \pi_{ik} + \sum_b w_b \frac{\partial z_b}{\partial f_{ik}} = 0$ avec $\forall b, w_b \geq 0$ et $w_b z_b = 0$, ainsi que $\forall i, k, \pi_{ik} \geq 0$ et $\pi_{ik} f_{ik} = 0$.

Donc $\forall i, k, F'_{ik} = \pi_{ik} \geq 0$ et $F'_{ik} f_{ik} = 0$, donc $0 = \sum_{k \in i} F'_{ik} f_{ik} = q_i (\tilde{G}_i - D_i^{-1}(q_i))$ car les termes en η'_i s'évanouissent. Donc si $q_i > 0$ alors $\tilde{G}_i = D_i^{-1}(q_i)$: il ne reste qu'à choisir $\lambda_i \equiv \min_{k \in i} I_{ik}$. Quant au cas $q_i = 0$, il est trivial pour l'affectation.

²à un vecteur $(\hat{\mathbf{f}}, \lambda, \mathbf{w})$ qui résout l'inéquation basique (8.48), associons des multiplicateurs

8.2.4.3 Programme extrémal

La formulation extrémale donnée par Beckmann (1956) au modèle monocritère repose sur une analogie économique, expliquée par Gartner (1980) : la fonction objectif ⁽¹⁾ correspond au surplus total des acteurs, offreurs et demandeurs, sur le marché économique "service d'acheminement". Les offreurs sont les arcs a qui produisent des unités x_a au coût marginal t_a . Leur surplus vaut

$$\sum_a \left(x_a t_a - \int_0^{x_a} t_a(\theta) d\theta \right).$$

Quant aux demandeurs, les relations O-D i qui tirent chacune un profit brut $\int_0^{q_i} D_i^{-1}(\theta) d\theta$ mais déboursent $q_i \tilde{G}_i^{\text{mono}}$, leur surplus vaut

$$\sum_i \left(\int_0^{q_i} D_i^{-1}(\theta) d\theta - q_i \tilde{G}_i^{\text{mono}} \right),$$

donc au total

$$-J^{\text{mono}}(\mathbf{f}) \equiv \sum_a \left(x_a t_a - \int_0^{x_a} t_a(\theta) d\theta \right) + \sum_i \left(\int_0^{q_i} D_i^{-1}(\theta) d\theta - q_i \tilde{G}_i^{\text{mono}} \right)$$

ce qui se simplifie car les recettes des offreurs sont les dépenses des demandeurs, donc $\sum_a x_a t_a = \sum_i q_i \tilde{G}_i^{\text{mono}}$ et

$$J^{\text{mono}}(\mathbf{f}) = \sum_a \int_0^{x_a} t_a(\theta) d\theta - \sum_i \int_0^{q_i} D_i^{-1}(\theta) d\theta. \quad (8.54)$$

Pour le modèle bicritère, j'ai développé une formulation extrémale comparable (Leurent 1993a) : dans le surplus monocritère seules les dépenses temporelles apparaissent (quitte à inclure les autres frais subis sur un arc dans le temps de l'arc, mais avec un facteur de conversion unique pour tous les déplacements). Ma formulation extrémale repose sur un "potentiel bicritère" qui distingue les frais temporels des frais monétaires, et reconnaît les différents arbitrages entre temps et argent :

$$-J^{\text{bic}}(\mathbf{f}) \equiv \sum_a \left(x_a t_a - \int_0^{x_a} t_a(\theta) d\theta \right) + \sum_i \left(\int_0^{q_i} D_i^{-1}(\theta) d\theta - q_i \tilde{G}_i^{\text{bic}} \right)$$

$$J^{\text{bic}}(\mathbf{f}) = \sum_a \int_0^{x_a} t_a(\theta) d\theta + \sum_i q_i \left(\sum_{n=1}^{\bar{m}_i} \bar{P}_{i,n} \left(\eta_i \left(\frac{q'_{i,n}}{q_i} \right) - \eta_i \left(\frac{q'_{i,n-1}}{q_i} \right) \right) \right) - \int_0^{q_i} D_i^{-1}(\theta) d\theta. \quad (8.55)$$

Grâce à l'inéquation (8.37), tout point d'équilibre bicritère *minimise* ce potentiel bicritère (Leurent 1993a). Signalons qu'il faut se garder de toute interprétation économique :

$\pi_{ik} \equiv F'_{ik} - \min_{k' \in i} F'_{ik'}$ aux pseudo-contraintes $g_{ik}(\mathbf{f}) \equiv f_{ik} \geq 0$. Par construction $\pi_{ik} \geq 0$, $F'_{ik} - \pi_{ik} \frac{\partial g_{ik}}{\partial f_{ik}} = 0$, tandis que $\pi_{ik} f_{ik} = 0$ résulte des propriétés caractéristiques du NCP associé à (8.48). Donc les hypothèses du théorème AS2 de Tobin et Friesz (1988) sont rassemblées et le vecteur $\hat{\mathbf{f}}$ résout l'inéquation variationnelle réduite.

¹plus exactement son opposé

l'analogie de Gartner a sa propre interprétation, qui n'est pas celle attendue a priori pour un problème à péage justiciable de l'affectation bicritère !

Le lien avec les formules caractéristiques précédentes est le suivant : le gradient du potentiel J^{bic} par rapport aux flots f_{ik} a pour composantes les fonctions d'impédance F''_{ik} qui apparaissent dans l'inéquation variationnelle réduite. Détaillons les calculs du terme proprement bicritère $J_i^{\text{bic}} \equiv q_i \sum_{n=1}^{\bar{m}_i} \bar{P}_{i,n} (\eta_i(\frac{q'_{i,n}}{q_i}) - \eta_i(\frac{q'_{i,n-1}}{q_i}))$.

Formellement, nous dérivons J_i^{bic} par rapport à $q'_{i,n}$ pour n entre 1 et \bar{m}_i , et par rapport à q_i , en oubliant provisoirement les contraintes de définition qui imposent $q'_{i,\bar{m}_i} = q_i$, pour ensuite calculer la dérivée de J_i^{bic} par rapport à f_{ik} grâce à la règle de dérivation d'une fonction composée :

$$\begin{aligned} \frac{\partial}{\partial q'_{i,n}} J_i^{\text{bic}} &= (\bar{P}_{i,n} - \bar{P}_{i,n+1} \delta_{n,\bar{m}_i}) \eta'_i(\frac{q'_{i,n}}{q_i}) \text{ avec } \delta_{n,m} \equiv 1 \text{ si } n = m \text{ ou } 0 \text{ sinon, et} \\ \frac{\partial}{\partial q_i} J_i^{\text{bic}} &= \sum_{n=1}^{\bar{m}_i} \bar{P}_{i,n} (\eta_i(\frac{q'_{i,n}}{q_i}) - \eta_i(\frac{q'_{i,n-1}}{q_i})) + q_i \sum_{n=1}^{\bar{m}_i} \bar{P}_{i,n} (\frac{-q'_{i,n}}{(q_i)^2} \eta'_i(\frac{q'_{i,n}}{q_i}) - \frac{-q'_{i,n-1}}{(q_i)^2} \eta'_i(\frac{q'_{i,n-1}}{q_i})) \\ &= \left(\sum_{n=1}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) (\eta_i(\frac{q'_{i,n}}{q_i}) - \frac{q'_{i,n}}{q_i} \eta'_i(\frac{q'_{i,n-1}}{q_i})) \right) \\ &\quad + \bar{P}_{i,\bar{m}_i} \left(\eta_i(\frac{q'_{i,\bar{m}_i}}{q_i}) - \frac{q'_{i,\bar{m}_i}}{q_i} \eta'_i(\frac{q'_{i,\bar{m}_i}}{q_i}) \right) - \bar{P}_{i,1} \left(\eta_i(\frac{q'_{i,0}}{q_i}) - \frac{q'_{i,0}}{q_i} \eta'_i(\frac{q'_{i,0}}{q_i}) \right) \end{aligned}$$

en convenant que $q'_{i,0} \equiv 0$. Par la règle de dérivation d'une fonction composée,

$$\begin{aligned} \frac{\partial}{\partial f_{ik}} J_i^{\text{bic}} &= \left(\sum_{n=1}^{\bar{m}_i} \bar{P}_{i,n} \frac{\partial J_i^{\text{bic}}}{\partial q'_{i,n}} \frac{\partial q'_{i,n}}{\partial f_{ik}} \right) + \frac{\partial J_i^{\text{bic}}}{\partial q_i} \frac{\partial q_i}{\partial f_{ik}} \text{ avec } \frac{\partial q'_{i,n}}{\partial f_{ik}} = \Gamma_{in}^k \text{ et } \frac{\partial q_i}{\partial f_{ik}} = 1, \text{ donc} \\ \frac{\partial}{\partial f_{ik}} J_i^{\text{bic}} &= \left(\sum_{n=1}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) (\eta_i(\frac{q'_{i,n}}{q_i}) + (\Gamma_{in}^k - \frac{q'_{i,n}}{q_i}) \eta'_i(\frac{q'_{i,n-1}}{q_i})) \right) + \bar{P}_{i,\bar{m}_i} \eta_i(1) - \bar{P}_{i,1} \eta_i(0). \end{aligned}$$

D'où le résultat annoncé et le théorème que voici.

Théorème : formulation extrémale pour l'affectation bicritère. Si les fonctions de temps de parcours t_a sont symétriques, alors l'équilibre bicritère à demande élastique, temps variables et contraintes adjointes se caractérise comme la solution de

$$\min_{\mathbf{f} \geq 0} J^{\text{bic}}(\mathbf{f}) \quad (8.56a)$$

soumis aux contraintes

$$z_b(\mathbf{f}) \leq 0 \quad (8.56b)$$

$$x_a(\mathbf{f}) \equiv \sum_{ik} \delta_a^{ik} f_{ik} \quad (8.56c)$$

$$q_i(\mathbf{f}) \equiv \sum_{k \in i} f_{ik} \quad (8.56d)$$

$$q'_{i,n}(\mathbf{f}) \equiv \sum_{k \in i} \Gamma_{in}^k f_{ik}. \quad (8.56e)$$

Les dérivées secondes de J_i^{bic} par rapport aux f_{ik} servent pour l'analyse de sensibilité (Cf. section 10.3). On les obtient également en dérivant une fonction composée, grâce à

la relation $\frac{\partial^2}{\partial f_{ik'} \partial f_{ik}} J_i^{\text{bic}} = \left(\sum_{n=1}^{\bar{m}_i-1} \frac{\partial^2 J_i^{\text{bic}}}{\partial q_{in} \partial f_{ik}} \Gamma_{in}^k \right) + \frac{\partial^2 J_i^{\text{bic}}}{\partial q_i \partial f_{ik}}$. On calcule

$$\frac{\partial}{\partial q_{i,n}} \frac{\partial}{\partial f_{ik}} J_i^{\text{bic}} = (\bar{P}_{i,n} - \bar{P}_{i,n+1}) \eta_i'' \left(\frac{q_{i,n}}{q_i} \right) \left(q_i \Gamma_{in}^k - q'_{i,n} \right) / (q_i)^2 \text{ pour } n \neq \bar{m}_i \text{ et}$$

$$\frac{\partial}{\partial q_i} \frac{\partial}{\partial f_{ik}} J_i^{\text{bic}} = \sum_{n=1}^{\bar{m}_i-1} (\bar{P}_{i,n} - \bar{P}_{i,n+1}) \eta_i'' \left(\frac{q_{i,n}}{q_i} \right) \frac{q_{i,n}}{(q_i)^3} (-\Gamma_{in}^k q_i - q'_{i,n}), \text{ donc}$$

$$\frac{\partial^2}{\partial f_{ik'} \partial f_{ik}} J_i^{\text{bic}} = \sum_{n=1}^{\bar{m}_i-1} \frac{\bar{P}_{i,n} - \bar{P}_{i,n+1}}{(q_i)^4} \eta_i'' \left(\frac{q_{i,n}}{q_i} \right) (q_i \Gamma_{in}^k - q'_{i,n}) (q_i \Gamma_{in}^{k'} - q'_{i,n}). \quad (8.57)$$

8.2.4.4 L'existence

Pour qu'il existe au moins une solution, il suffit que l'ensemble faisable U'' soit non vide et compact, et que l'application support de l'inéquation variationnelle soit continue. La continuité de F'' est facile à assurer. La compacité de U'' est immédiate si la demande est fixe, ainsi que dans les problèmes à demande élastique avec une fonction de demande bornée (ce qui est toujours le cas en pratique !).

Théorème : existence d'une solution à l'inéquation variationnelle réduite. Nous supposons que l'ensemble faisable U'' n'est pas vide. Si les fonctions t_{ik} , D_i^{-1} , $\partial z_b / \partial f_{ik}$ et H_i sont continues, alors il existe une solution à l'inéquation variationnelle (8.53).

Bien sûr on peut affaiblir l'hypothèse de continuité en semi-continuité. Plus intéressant, le problème d'étendre les formules caractéristiques à des répartitions H_i continues par morceaux seulement, afin de considérer aussi des masses discrètes dans la distribution des valeurs du temps.

8.2.4.5 L'unicité

De façon générale, si le jacobien de l'application F'' associée à l'inéquation variationnelle est monotone, alors l'ensemble des solutions est convexe. C'est le cas notamment si les fonctions de temps de parcours sont croissantes au sens large et les fonctions de demande décroissantes au sens large : puisqu'alors les termes correspondants dans l'impédance sont monotones, donc aussi leur somme, donc encore F'' toute entière puisque le terme proprement bicritère est lui-même monotone.

Montrons la monotonie du terme proprement bicritère dans le cas où la fonction H_i est dérivable : alors la dérivée de η_i' vaut $\eta_i''(x) = -\frac{1}{2} \left\{ H_i'(H_i^{-1}(x)) \times (H_i^{-1}(x))^2 \right\} \leq 0$. En tout point $\tilde{\mathbf{f}}_i$ particulier, $\forall \mathbf{f}_i \equiv (f_{ik})_{k \in i}$,

$$\mathbf{f}_i \cdot \nabla^2 J_i^{\text{bic}} \Big|_{\tilde{\mathbf{f}}_i} \cdot \mathbf{f}_i = \sum_{k,k'} \frac{\partial^2 J_i^{\text{bic}}}{\partial f_{ik'} \partial f_{ik}} \Big|_{\tilde{\mathbf{f}}_i} f_{ik} f_{ik'}$$

$$\begin{aligned}
&= \sum_{k,k'} \left(\sum_{n=1}^{\bar{m}_i-1} \frac{\bar{P}_{i,n}-\bar{P}_{i,n+1}}{(\bar{q}_i)^4} \eta_i'' \left(\frac{\bar{q}'_{i,n}}{\bar{q}_i} \right) (\bar{q}_i \Gamma_{in}^k - \bar{q}'_{i,n}) (\bar{q}_i \Gamma_{in}^{k'} - \bar{q}'_{i,n}) \right) f_{ik} f_{ik'} \\
&= \sum_{n=1}^{\bar{m}_i-1} \frac{\bar{P}_{i,n}-\bar{P}_{i,n+1}}{(\bar{q}_i)^4} \eta_i'' \left(\frac{\bar{q}'_{i,n}}{\bar{q}_i} \right) \sum_{k,k'} f_{ik} f_{ik'} (\bar{q}_i \Gamma_{in}^k - \bar{q}'_{i,n}) (\bar{q}_i \Gamma_{in}^{k'} - \bar{q}'_{i,n}) \\
&= \sum_{n=1}^{\bar{m}_i-1} \frac{\bar{P}_{i,n}-\bar{P}_{i,n+1}}{(\bar{q}_i)^4} \eta_i'' \left(\frac{\bar{q}'_{i,n}}{\bar{q}_i} \right) \left(\sum_k f_{ik} (\bar{q}_i \Gamma_{in}^k - \bar{q}'_{i,n}) \right)^2 \\
&= \sum_{n=1}^{\bar{m}_i-1} \frac{\bar{P}_{i,n}-\bar{P}_{i,n+1}}{(\bar{q}_i)^4} \eta_i'' \left(\frac{\bar{q}'_{i,n}}{\bar{q}_i} \right) (\bar{q}_i \bar{q}'_{i,n} - \bar{q}'_{i,n} \bar{q}_i)^2
\end{aligned}$$

ce qui est positif ou nul puisque $\eta_i'' \leq 0$ et $\bar{P}_{i,n} \leq \bar{P}_{i,n+1}$.

Si $H'_i > 0$ sur l'intérieur du support Ω^* , alors la monotonie du terme proprement bicritère est stricte et les parts de marché des classes isoprix, les $q_{i,n} / q_i$, sont uniques à l'équilibre.

De même si t_a croît strictement alors x_a est unique à l'équilibre, et si D_i décroît strictement alors q_i est unique à l'équilibre.

8.3 FORMULE CARACTÉRISTIQUE EN DIMENSION INFINIE

Je rappelle ici le traitement formel en dimension infinie du modèle bicritère, d'après Leurent (1991b, 1993b). La dimension infinie permet seule de représenter à la fois une distribution continue des valeurs du temps, et une élasticité désagrégée du volume de demande au niveau de service. Cependant elle nécessite des connaissances mathématiques plus étendues que la formulation en dimension finie, et sa résolution se révèle plus ardue.

Nous commençons par illustrer comment la demande réagit de façon désagrégée à une modification de l'offre (§8.3.1), puis nous proposons une formalisation (§8.3.2) et nous résumons les principaux résultats mathématiques (§8.3.3).

8.3.1 Illustration

Pour illustrer la représentation désagrégée de l'élasticité de la demande dans le modèle bicritère, simulons les effets de la mise en service d'un nouveau moyen de transport.

8.3.1.1 Les services en concurrence

Soit un système élémentaire de transport avec trois moyens de déplacement pour les voyageurs potentiels, tous demandeurs d'un trajet entre une même origine et une même destination :

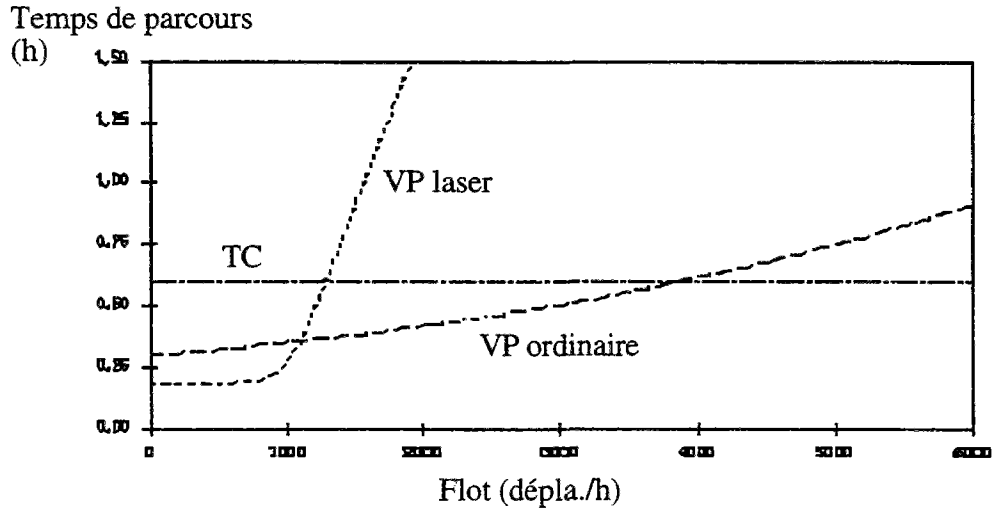
- un mode de transport collectif, peu coûteux mais relativement lent, dont la performance ne dépend pas du flot écoulé.
- le mode automobile ordinaire par un itinéraire qui n'est pas tarifé, ne peut être saturé mais dont les performances décroissent sensiblement en fonction du flot écoulé.
- un mode automobile suivant le concept LASER, c'est-à-dire un itinéraire tarifé de sorte que l'usager en attend de la rapidité.

Nous supposons que la capacité du mode LASER est suffisamment inférieure à celle de la route gratuite, afin que son temps de parcours croisse plus vite en fonction du flot. Pour chaque mode k , le temps de déplacement dépend du trafic du mode selon $t_k(f_k) = t_k^0 (1 + \gamma_k + ([\alpha_k (1 - \frac{f_k}{N_k})]^2 + \beta_k^2)^{1/2} - \alpha_k (1 - \frac{f_k}{N_k}) - \beta_k)$ où t_k^0 est le temps de parcours à trafic nul, N_k une capacité pratique, α_k , β_k et γ_k des paramètres pour prendre en compte la congestion et qui vérifient $\alpha_k > \gamma_k$ et $\beta_k = \frac{\gamma_k(\alpha_k - \gamma_k/2)}{\alpha_k - \gamma_k}$.

Tab. 8G. Caractères des trois modes.

Mode k	Prix P_k (Francs)	Temps libre t_k^0 (heures)	Capacité pratique (dépla/heure)	α_k	γ_k
TC	3,6	0.60	50 000	1.0	0.0
voiture ordinaire	18,4	0.30	5 000	2.5	1.5
route première classe	54,4	0.18	1 000	4.0	0.5

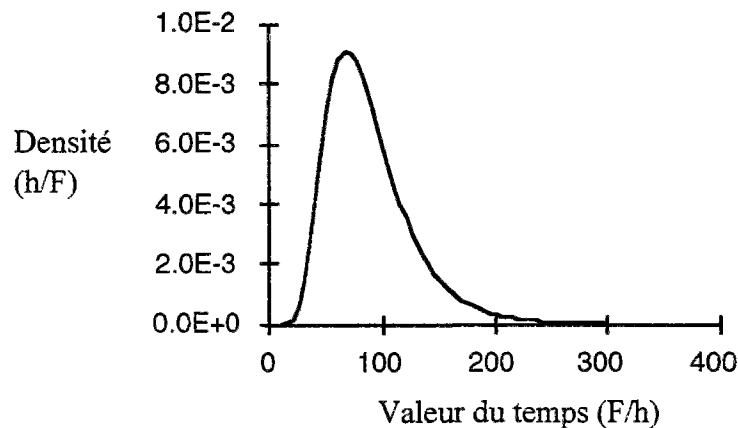
Fig. 8h. Les fonctions de temps de parcours des trois modes.



8.3.1.2 La demande

Nous supposons que les demandeurs sont différenciés selon leur valeur du temps, qui suit une distribution log-normale tronquée $d\mu(v) = \frac{1}{\sigma\sqrt{2\pi v}} \exp(-\frac{1}{2}(\frac{\log(v/\text{med})}{\sigma})^2)dv$ sur $[v_0; v_1]$ et 0 ailleurs. On fixe $v_0 = 1$ F/h, $v_1 = 400$ F/h, $\text{med} = 80$ F/h, $\sigma = 0.6$.

Fig. 8i. Distribution des valeurs du temps.



Pour chaque valeur du temps v et chaque mode k , le temps généralisé comprend le temps proprement dit et les frais monétaires, soit $G_k(v) = t_k(f_k) + P_k / v$.

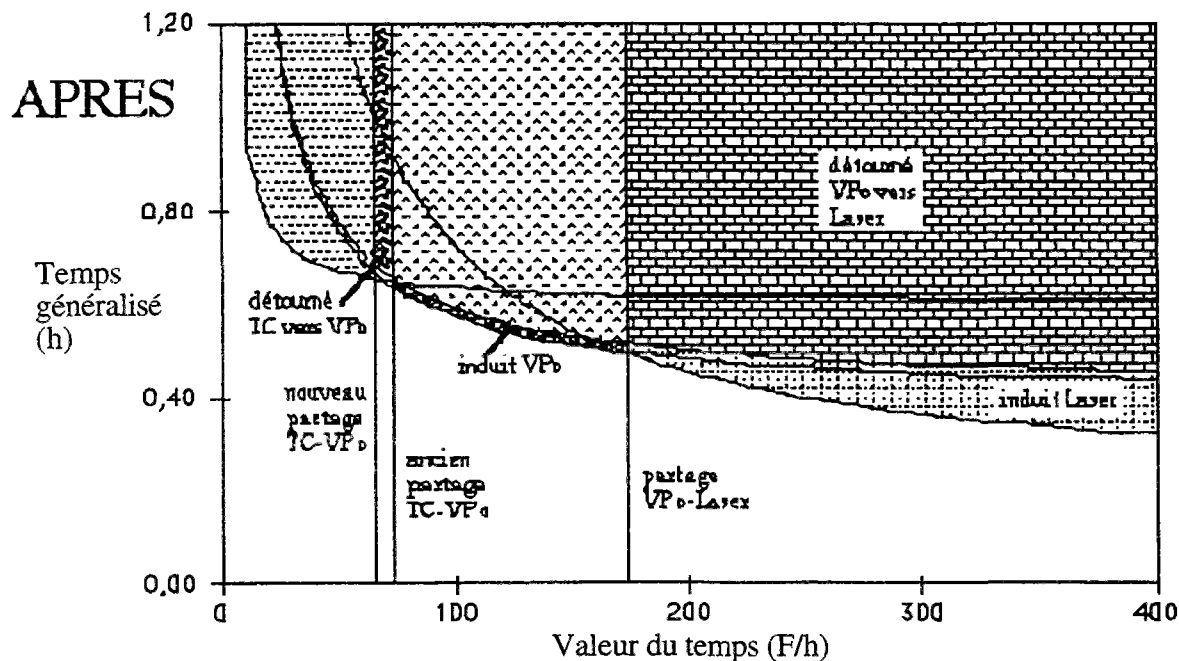
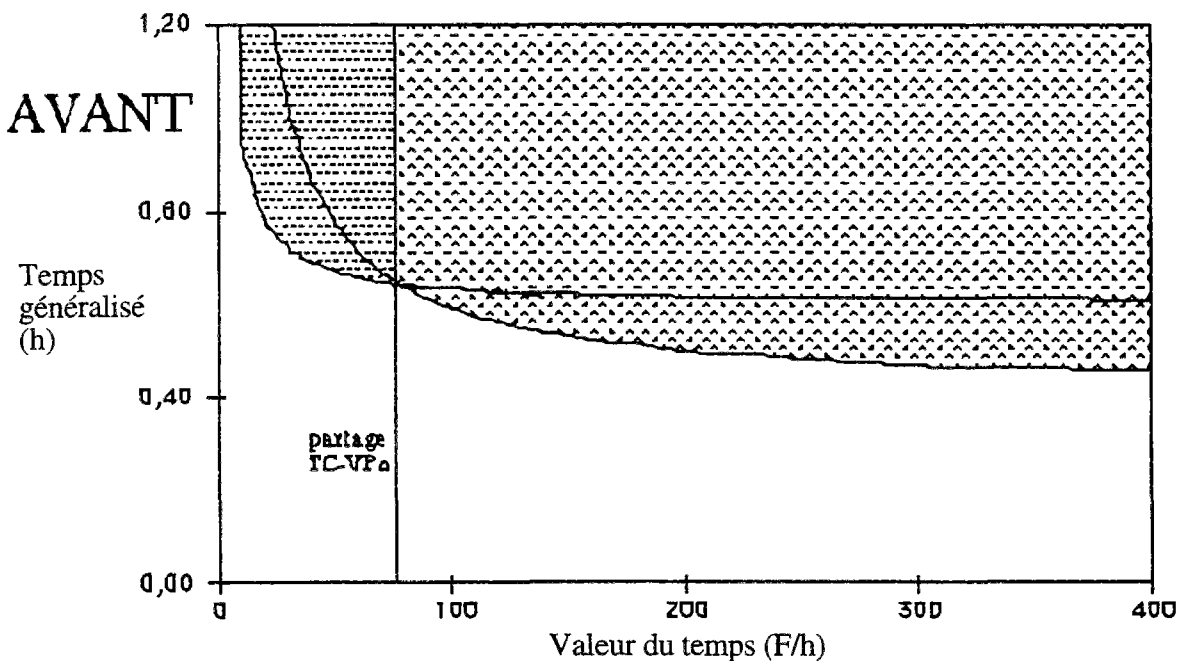
Nous supposons que pour toutes les valeurs du temps, le volume réagit de façon identique au temps généralisé : $D_{rs}^{[v]}(S) = D(S) = D_0[(S/S_0)^{-\alpha} - (S_1/S_0)^{-\alpha}]$ sur $[S_0; S_1]$ et 0 ailleurs, avec pour paramètres $S_0 = 0.15h$, $S_1 = 3.0h$, $\alpha = 0.6$ et $D_0 = 12\ 000$ déplacements par heure (revoir la figure 8e).

8.3.1.3 Simulation des situations avant et après la mise en service

Les deux dates d'application du modèle sont définies comme "avant" et "après" la mise en service du troisième mode. Le tableau 8H présente les résultats de l'application du modèle aux deux dates.

Tab. 8H. Résultats avant et après la mise en service.

Mode	Ensemble	TC	VP ordinaire	VP laser
flot (dépla/h)	3162	1341	1821	-
VC _{inf} (F/h)	1.0	1.0	76.4	-
VC _{sup} (F/h)	400.0	76.4	400.0	-
vitesse (km/h)	24.6	20.0	29.5	-
flot (dépla/h)	3244	1205	1622	417
VC _{inf} (F/h)	1.0	1.0	71.2	173.3
VC _{sup} (F/h)	400.0	71.2	173.3	400.0
vitesse (km/h)	27.1	20.0	30.6	65.0



Après la mise en service du troisième mode, on constate naturellement que plus de déplacements s'effectuent au total, à une vitesse moyenne plus élevée.

8.3.2 Formalisation

8.3.2.1 Éléments standards

Les objets et dépendances élémentaires du modèle déterministe homogène (Cf. §8.1.1) sont tous repris dans le modèle bicritère (Cf. section 7.2) qui ajoute un composant conceptuel sans en retrancher aucun autre. En particulier il reste utile de considérer les flots par chemins f_{rs}^k et les flots par arcs x_a . On suppose que les chemins k forment un ensemble bien identifié, grâce à la donnée de la matrice d'incidence arcs-chemins $\Delta = [\delta_{rs}^{ak}]_{rsk,a}$.

8.3.2.2 Éléments originaux liés aux arbitrages entre prix et temps

Le modèle bicritère ajoute une différenciation des déplacements par rapport à la valeur du temps, facteur d'arbitrage entre prix et temps. On considère deux attributs de coût pour chaque chemin $r-s-k$: un temps pur T_{rs}^k et un prix P_{rs}^k , à combiner de façon particulière pour chaque valeur du temps v . Alors le temps généralisé est une fonction $G_{rs}^k(v) = T_{rs}^k + \tilde{P}_{rs}^k(v)$ où la notation $\tilde{P}_{rs}^k(v)$ généralise la dépendance P_{rs}^k/v . Dans les notations T_{rs}^k , \tilde{P}_{rs}^k et G_{rs}^k , la dépendance envers un vecteur de flot par chemins est sous-entendue.

Du côté de la demande, les flots sont caractérisés par une dimension supplémentaire, la valeur du temps. Comment préciser la distribution statistique de ce caractère, pour une relation origine-destination donnée ? En toute rigueur, chaque segment de valeur du temps a sa propre fonction de demande : pour préciser la distribution statistique de la valeur du temps dans la demande servie, il faut analyser l'équilibre entre offre et demande, et la seule connaissance a priori est relative à la distribution statistique des fonctions de demande par segment, les $D_{rs}^{[v]}(\mathbf{S})$ où \mathbf{S} est un vecteur d'attributs des services offerts. Cette connaissance se formalise par un couple $(\mu_{rs}(\cdot); D_{rs}^{[1]}(\cdot))$ où μ_{rs} est une mesure de masse (1) et $\{D_{rs}^{[v]}(\cdot)\}_v$ est une famille de fonctions de demande (2).

Dans ces conditions, la demande servie associée à un état des services \mathbf{S} vaut $\int_{total} D_{rs}^{[v]}(\mathbf{S}) d\mu_{rs}(v)$; la proportion des déplacements avec une valeur du temps comprise entre x et y vaut $\left(\int_x^y D_{rs}^{[v]}(\mathbf{S}) d\mu_{rs}(v) \right) / \left(\int_{total} D_{rs}^{[v]}(\mathbf{S}) d\mu_{rs}(v) \right)$.

Pour quantifier le flot affecté à l'itinéraire k de la relation origine-destination $r-s$, on emploie une fonction de clientèle $\tilde{f}_{rs}^k(v)$ telle que $\int_x^y \tilde{f}_{rs}^k(v) d\mu_{rs}(v)$ soit l'effectif des

¹définie a priori sur l'ensemble des valeurs du temps, supposé un sous-ensemble de \mathfrak{R}_+

²définies chacune sur un convexe fermé de \mathfrak{R}_+ , à valeurs dans \mathfrak{R}_+

déplacements entre r et s passant par le chemin k et dont la valeur du temps est comprise entre x et y .

Alors l'équilibre entre offre et demande s'établit, de façon désagrégée, pour chaque segment de valeur du temps : $\forall v \sum_k \bar{f}_{rs}^k(v) = D_{rs}^{[v]}(\mathbf{S})$.

Un flot par arc et tranche de valeur du temps s'exprime :

$$x_a|_x^y = \sum_{rsk} \delta_{rs}^{ak} f_{rs}^k|_x^y = \sum_{rs} \int_x^y \left(\sum_k \delta_{rs}^{ak} \bar{f}_{rs}^k(v) \right) d\mu_{rs}(v).$$

Définition di-D1, vecteur de clientèle. Nous appelons vecteur de clientèle et notons $\bar{\mathbf{f}}$, un vecteur fonctionnel à autant de composantes que de chemins, dont chaque composante est une fonction de consommation du chemin r - s - k par les demandeurs de valeur du temps v : $\bar{f}_{rs}^k(v)$. Nous notons $\bar{f}_{rs}^0(v) = \sum_k \bar{f}_{rs}^k(v)$ et $f_{rs}^k = \int \bar{f}_{rs}^k(v) d\mu_{rs}(v)$.

8.3.2.3 Conditions d'équilibre

Définition di-D2, équilibre bicritère. Un équilibre bicritère est un couple constitué d'un vecteur de clientèle $\bar{\mathbf{f}}$ et d'un vecteur \mathbf{w} de multiplicateurs associés à des contraintes adjointes, tel que :

(i) fonctionnement de l'offre :

- $\forall r, s, k \quad T_{rs}^k = \left(\sum_a \delta_{rs}^{ak} t_a(\bar{\mathbf{f}}) \right) + \sum_b w_b \cdot \frac{\partial z_b}{\partial f_{rs}^k}$,
- $\forall b \quad z_b(\bar{\mathbf{f}}) \leq 0$ et $w_b \geq 0$ et $w_b \cdot z_b(\bar{\mathbf{f}}) = 0$.

(ii) formation des coûts généralisés ⁽¹⁾ :

- $\forall r, s, v, k \quad G_{rs}^k(v) = T_{rs}^k + \bar{P}_{rs}^k(v)$.

(iii) comportement rationnel de la demande :

- $\forall r, s, v \quad S_{rs}(v) = \min_k G_{rs}^k(v)$,
- $\forall r, s, v, k \quad G_{rs}^k(v) > S_{rs}(v) \Rightarrow \bar{f}_{rs}^k(v) = 0$.

(iv) équilibre offre-demande :

- $\forall r, s, v \quad \sum_k \bar{f}_{rs}^k(v) = D_{rs}^{[v]}(S_{rs}(v))$.

Définition di-D3, niveau de service différentiel. Soit $\bar{\mathbf{V}}$ l'application qui au couple $\mathbf{u} = (\bar{\mathbf{f}}; \mathbf{w})$ associe le couple $(\bar{\mathbf{G}} - \Gamma \cdot \bar{\mathbf{D}}^{-1}; -\mathbf{z})$ où $\bar{\mathbf{G}} - \Gamma \cdot \bar{\mathbf{D}}^{-1}$ est un vecteur fonctionnel

¹convenons pour la suite de noter $\forall v$ au lieu de $d\mu_{rs}$ -presque partout en v

à autant de composantes $G_{rs}^k(v) - D_{rs}^{[v]-1}(\bar{f}_{rs}^0(v))$ que de chemins, et \mathbf{z} est le vecteur des fonctions liées aux contraintes adjointes.

Prendre le produit scalaire entre un vecteur $(\bar{\mathbf{f}}_0; \mathbf{w}_0)$ et un vecteur $\bar{\mathbf{V}}(\bar{\mathbf{f}}; \mathbf{w})$ revient à sommer deux produits scalaires :

- le premier produit scalaire ⁽¹⁾ concerne $\bar{\mathbf{f}}_0$ et $(\bar{\mathbf{G}} - \Gamma \cdot \bar{\mathbf{D}}^{-1})(\bar{\mathbf{f}}; \mathbf{w})$, il est égal à :

$$\sum_{rsk} \int \bar{f}_{0rs}^k(v) \cdot [G_{rs}^k(v) - D_{rs}^{[v]-1}(\bar{f}_{rs}^0(v))] d\mu_{rs}(v).$$

- le second entre \mathbf{w}_0 et $-\mathbf{z}(\bar{\mathbf{f}})$, égal à $-\mathbf{z}(\bar{\mathbf{f}}) \cdot \mathbf{w}_0 = -\sum_b z_b(\bar{\mathbf{f}}) \cdot w_{0b}$.

Interrompons ici le discours formel pour constater la complexité inhérente à la formalisation et à l'énoncé rigoureux du modèle. Tandis que la conception d'une portion du modèle reste qualitative, la formulation nécessite de préciser l'architecture d'ensemble, elle révèle les questions liées à la cohérence des relations élémentaires.

8.3.3 Formules caractéristiques

Un équilibre bicritère se formule sous des hypothèses très générales comme une inéquation variationnelle. En fortifiant certaines hypothèses, on le reformule comme un programme extrémal. Les théorèmes di-T1 et di-T2 et les corollaires di-C1 et di-C2 sont prouvés dans Leurent (1993b).

8.3.3.1 Inéquation variationnelle

Théorème di-T1, inéquation variationnelle pour le modèle bicritère à demande élastique désagrégée et contraintes adjointes. $(\bar{\mathbf{f}}^*; \mathbf{w}^*) \geq \mathbf{0}$ est un équilibre bicritère si et seulement s'il résout l'inéquation variationnelle

$$\forall (\bar{\mathbf{f}}; \mathbf{w}) \geq \mathbf{0}, \quad \bar{\mathbf{V}}(\bar{\mathbf{f}}^*; \mathbf{w}^*) \cdot (\bar{\mathbf{f}} - \bar{\mathbf{f}}^*; \mathbf{w} - \mathbf{w}^*) \geq 0.$$

Dans cette formulation, les attributs du coût généralisé sont des T_{rs}^k et $\bar{P}_{rs}^k(v)$ astreints à dépendre de $(\bar{\mathbf{f}}^*; \mathbf{w}^*)$ d'une façon à préciser. A priori tant les T_{rs}^k que les $\bar{P}_{rs}^k(v)$ peuvent dépendre des flots f_{rs}^k liés à $\bar{\mathbf{f}}^*$.

Corollaire di-C1, existence. S'il existe un couple $(\bar{\mathbf{f}}; \mathbf{w})$ satisfaisant $\mathbf{z}(\bar{\mathbf{f}}) \leq \mathbf{0}$, et si les T_{rs}^k et $\bar{P}_{rs}^k(v)$ dépendent continûment de $\bar{\mathbf{f}}$ et \mathbf{w} ⁽²⁾, alors il existe au moins une solution à l'inéquation variationnelle de di-T1, donc au moins un équilibre bicritère.

¹ dans un espace L^2 , Cf. Leurent 1993b. Les vecteurs fonctionnels $\bar{\mathbf{f}}$ et $(\bar{\mathbf{G}} - \Gamma \cdot \bar{\mathbf{D}}^{-1})(\bar{\mathbf{f}}; \mathbf{w})$ sont définis dans l'espace de Hilbert des applications $\otimes_{rs} d\mu_{rs}$ mesurables de carré $\otimes_{rs} d\mu_{rs}$ intégrable, allant de $\otimes_{rs} \mathcal{R}_+$ muni de sa tribu borélienne $\otimes_{rs} \mathcal{B}(\mathcal{R}_+)$ dans $\otimes_{rs} \mathcal{R}$ muni de sa tribu borélienne $\otimes_{rs} \mathcal{B}(\mathcal{R})$

²⁺ détail des hypothèses techniques complémentaires, Cf. Leurent 1993b

8.3.3.2 Programme extrémal

Lorsque les fonctions de temps de parcours t_a sont symétriques par rapport aux f_{rs}^k et les fonctions de prix $\tilde{P}_{rs}^k(v)$ sont indépendantes de $\bar{\mathbf{f}}$ et de \mathbf{w} , alors l'inéquation variationnelle du théorème di-T1 s'intègre en un programme extrémal à la Beckmann.

Théorème di-T2, programme extrémal en dimension quelconque. Si chaque t_a dépend uniquement de $x_a(\bar{\mathbf{f}})$ et si les prix $\tilde{P}_{rs}^k(v)$ sont indépendants de $\bar{\mathbf{f}}$ et de \mathbf{w} , alors un couple $(\bar{\mathbf{f}}^*; \mathbf{w}^*) \geq \mathbf{0}$ est un équilibre bicritère si et seulement si $(\bar{\mathbf{f}}^*; \mathbf{w}^*)$ est point selle du lagrangien défini pour $(\bar{\mathbf{f}}; \mathbf{w}) \geq \mathbf{0}$ par :

$$\mathcal{L}(\bar{\mathbf{f}}; \mathbf{w}) \equiv \left(\sum_a \int_0^{x_a(\bar{\mathbf{f}})} t_a(\theta) d\theta \right) - \left(\sum_b z_b(\bar{\mathbf{f}}) \cdot w_b \right) + \sum_{rs} \int \left[\left(\sum_k \tilde{P}_{rs}^k(v) \cdot \bar{f}_{rs}^k(v) \right) - \bar{f}_{rs}^0(v) \cdot D_{rs}^{[v]-1}(\bar{f}_{rs}^0(v)) - \int_{D_{rs}^{[v]-1}(\bar{f}_{rs}^0(v))}^{+\infty} D_{rs}^{[v]}(\theta) d\theta \right] d\mu_{rs}(v).$$

Donc $\bar{\mathbf{f}}^* \geq \mathbf{0}$ participe à un équilibre bicritère si et seulement s'il résout le programme extrémal $\min_{\bar{\mathbf{f}} \geq \mathbf{0}} J_{\text{bic}\infty}(\bar{\mathbf{f}}) \equiv \mathcal{L}(\bar{\mathbf{f}}; \mathbf{0})$ soumis aux contraintes $z_b(\bar{\mathbf{f}}) \leq 0$.

Corollaire di-C2, unicité. Lorsqu'il existe une solution au programme extrémal du théorème di-T2, alors, en un point d'équilibre $\bar{\mathbf{f}}^*$:

- si les fonctions t_a croissent et si les fonctions $D_{rs}^{[v]-1}(\cdot)$ décroissent (presque partout en v), alors la fonction objectif $J_{\text{bic}\infty}$ est convexe donc l'ensemble des solutions est convexe.
- de plus, si la fonction t_a croît strictement, le flot $x_a(\bar{\mathbf{f}}^*)$ est unique. Si les fonctions $D_{rs}^{[v]-1}(\cdot)$ décroissent strictement presque partout en v , alors les $\bar{f}_{rs}^0(v)$ calculés au point $\bar{\mathbf{f}}^*$ sont uniques.

8.4 COMMENTAIRES

Nous avons mené l'audit formel sur les trois plans de la méthodologie (section 5.2), du domaine de l'affectation statique (section 8.1), du modèle bicritère (sections 8.2 et 8.3). Dans le cas du modèle bicritère, nous avons repris les étapes de la formulation.

Nous en tirons des enseignements relatifs à l'audit formel (§8.4.1) et à ses relations avec les autres dimensions d'audit (§8.4.2).

8.4.1 Enseignements de l'expérience d'audit formel

8.4.1.1 Aptitudes du modèle bicritère

Aptitude formelle du modèle bicritère conceptuel. Les relations élémentaires de conception incorporées au modèle bicritère ont été formulées de façon cohérente, avec des propriétés mathématiques qui assurent la résolubilité. Le modèle formel ainsi constitué se prête à la quantification.

Aptitude phénoménologique du modèle bicritère formel, grâce à la possibilité d'une quantification. Une application numérique a illustré l'aptitude du modèle bicritère formel à simuler quantitativement les phénomènes incorporés dans sa conception. Cependant, pour conserver une simplicité relative à la formulation, nous avons sacrifié la représentation désagrégée de l'élasticité de la demande, au profit d'une représentation agrégée qui constitue une approximation sémantique.

8.4.1.2 Enseignements méthodologiques

La capitalisation des ressources formelles du domaine de l'affectation statique du trafic routier *facilite* la formulation du modèle bicritère, tant pour transformer des objets et relations élémentaires de conception en variables et fonctions mathématiques, que pour rattacher la formule caractéristique à un cadre mathématique de référence dans lequel existent des résultats théoriques d'existence et d'unicité d'une solution.

La construction mathématique du modèle bicritère *manifeste une alternance continue entre induction et déduction* : l'induction procède à coups de définitions, tandis que la déduction s'exprime par des théorèmes.

La formulation peut receler des complexités propres. Nous avons présenté et employé divers objets et concepts mathématiques, dont certains témoignent d'une sophistication élevée (par exemple le vecteur de clientèle et son espace d'appartenance). Cette sophistication entraîne une complexité formelle : car pour manipuler rigoureusement des objets sophistiqués, il faut vérifier de multiples propriétés. Concluons que la garantie théorique de validité formelle peut avoir un coût non négligeable. Ce n'est pas toujours le cas, lorsqu'il y a dépendance directe des résultats envers les variables exogènes, le coût est négligeable. Mais pour un modèle d'équilibre, le choix est ouvert entre une garantie théorique coûteuse ou une formulation douteuse. A notre avis, il faut choisir la garantie théorique.

8.4.2 La formulation, charnière entre la conception et la quantification

8.4.2.1 Quel arbitrage entre validité conceptuelle et validité formelle ?

Le choix entre garantie théorique coûteuse et formulation douteuse suppose un choix antérieur, celui d'une composition sémantique sophistiquée plutôt que fruste. Si l'on désire une garantie théorique sans pouvoir consentir un investissement suffisant, alors il faut remettre en question le choix antérieur, donc se contenter d'une représentation phénoménologique plus simple. Par exemple, dans le cas d'un modèle de trafic, renoncer à la dépendance du temps de parcours local envers des flots d'autres arcs.

8.4.2.2 Dualité entre composition et forme du modèle

La conception et la formulation traitent des mêmes objets, sur des plans différents : plan phénoménologique pour la conception, plan mathématique pour la formulation. L'audit formel complète l'audit conceptuel en posant une contrainte de cohérence d'ensemble ; il permet de détecter des inconsistances ou des impossibilités, qui poussent à modifier le modèle conceptuel.

8.4.2.3 Arbitrage entre dimension formelle et dimension informatique de la modélisation

Nous avons présenté une formulation en dimension finie du modèle bicritère à demande élastique, non pas en raison d'une insuffisance mathématique de la formulation en dimension quelconque, mais faute de maniabilité pratique en l'état actuel des techniques¹. La formulation occupe une place intermédiaire entre la conception et la résolution du modèle.

8.4.2.4 La formulation offre un intérêt propre, d'ordre mathématique

Nous avons constaté la complexité inhérente à certaines opérations de la formulation. Cette complexité d'ordre mathématique participe à la complexité d'ensemble de la modélisation. En contrepartie, elle confère un intérêt mathématique au modèle, objet tant d'application que de perfectionnement des connaissances mathématiques générales.

¹un autre argument en faveur de la formulation en dimension finie tient à la moindre complexité mathématique

Chapitre 9

L'audit algorithmique : une vérification

Au chapitre 5, nous avons décomposé l'erreur de résolution par simulation numérique en quatre postes : défaillance de matériel ou de compilateur, ou faute d'implantation du logiciel d'application, ou imprécision par arrondi, ou insuffisante convergence d'un algorithme itératif. Chaque poste est justiciable d'actions de détection, correction et prévention. En particulier il faut un critère rigoureux pour mesurer l'erreur de convergence et en observer l'évolution au long des itérations de l'algorithme.

Nous recensons des ressources pour examiner l'erreur de convergence dans un modèle d'affectation statique du trafic routier : tant des critères de convergence pour la détecter, que des algorithmes pour la réduire (section 9.1).

Dans le cas particulier du modèle bicritère, nous comparons les algorithmes et les critères de convergence, sur une application de taille moyenne (section 9.2).

9.1 RESSOURCES POUR L'AFFECTATION STATIQUE DU TRAFIC ROUTIER

Nous allons indiquer des ressources pour étudier l'erreur de calcul, plus précisément l'erreur de convergence dans les modèles d'affectation. Au sujet des fautes d'implantation, nous renvoyons aux méthodes standards de certification des logiciels, et nous supposons dans la suite qu'il n'en subsiste aucune. Nous supposons également que l'on échappe aux défaillances de matériel ou de compilateur, et que les opérations élémentaires sont effectuées avec une précision suffisante.

Nous commençons par examiner la question de la convergence dans un modèle d'affectation statique du trafic routier (§9.1.1) ; puis nous dressons un état de l'art (§9.1.2). Nous présentons ensuite des critères pour mesurer la convergence (§9.1.3) et enfin des algorithmes convergents qui laissent espérer une résolution effective de la formule caractéristique du modèle (§9.1.4).

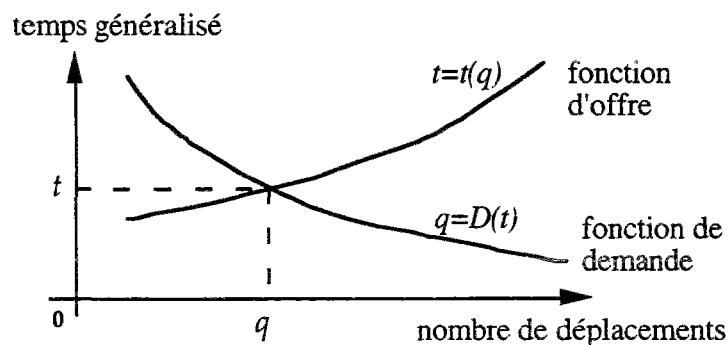
9.1.1 L'erreur de convergence

Dans un modèle d'affectation statique du trafic routier, l'erreur de convergence tient à une insuffisance ou à une absence de convergence de l'algorithme vers une solution de la formule caractéristique. La résolution de la formule caractéristique nécessite des calculs itératifs, car dès que l'on spécifie que les temps de parcours dépendent du trafic, il y a des dépendances circulaires entre les variables endogènes (Cf. §8.1.2).

9.1.1.1 Principe général des algorithmes

Les algorithmes d'équilibrage ont pour principe général d'ajuster progressivement un jeu de variables endogènes, jusqu'à satisfaire la formule caractéristique. Dans le cas de l'affectation, le jeu de variables endogènes comprend tant des variables liées à l'offre (temps de parcours) que des variables liées à la demande (flots par chemins ou par arcs).

Fig. 9a. Equilibre entre offre et demande d'acheminement.

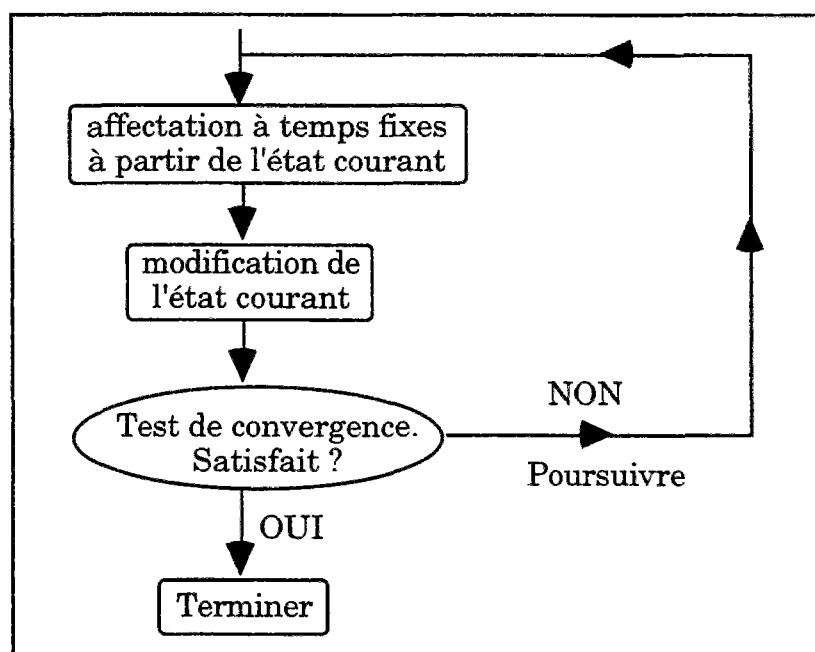


L'étape principale des calculs comprend :

- i) un établissement des prix du marché (application de la fonction d'offre sur la figure 9a).
- ii) une répartition à prix constants (temps fixes) de la demande entre les services concurrents (application de la fonction de demande sur la figure 9a).

L'algorithme consiste à répéter l'étape principale tant qu'une condition de terminaison n'est pas satisfaite (figure 9b).

Fig. 9b. Principe des calculs.



9.1.1.2 Mesurer la convergence

Un critère de convergence rigoureux mesure la distance entre un point courant (valeur courante des variables endogènes) et l'ensemble des solutions de la formule caractéristique. En pratique, on ne dispose pas d'une connaissance numérique préalable de l'ensemble des solutions, mais seulement d'une connaissance théorique liée aux propriétés mathématiques de la formule caractéristique. Donc le critère doit être formulé en fonction seulement du point courant et non de l'ensemble des solutions.

9.1.2 Etat de l'art

9.1.2.1 Positions des praticiens : déterminés par les auteurs des logiciels

Les auteurs des logiciels d'affectation manifestent des attitudes diverses (Cf. §3.4.1.1) :

- certains implantent un algorithme convergent muni d'un critère de convergence rigoureux.
- d'autres, niant l'existence d'algorithmes convergents et efficaces, implantent des algorithmes heuristiques dont ils fixent a priori l'issue, sans référence à un critère de convergence.
- enfin, dans le cas de modèles d'affectation plus sophistiqués que le modèle de Beckmann (surtout logit et probit), certains auteurs implantent un critère de convergence qui jouit d'une interprétation physique mais n'est pas rigoureux (le RMSE évoqué ci-dessous).

Les utilisateurs des logiciels sont captifs des options prises par les auteurs. A notre connaissance, l'erreur de convergence qui subsiste au terme de l'algorithme n'est jamais prise en compte de façon économétrique.

9.1.2.2 Côté recherche, quelques expérimentations

Du côté de la recherche, les travaux publiés sur l'erreur de convergence rapportent des expérimentations numériques sur ordinateur, qui examinent le déroulement de certains algorithmes à la loupe de plusieurs critères de convergence.

Horowitz (1986) effectue plusieurs séries de tests : il compare deux modèles (le modèle de Beckmann et celui de Bruynooghe-Evans qui combine distribution entropique et affectation) et trois algorithmes (l'un par "substitution", la méthode de Frank-Wolfe et une méthode des moyennes successives¹), cela pour deux contextes d'application. Il mesure l'erreur de convergence par un critère d'écart-type entre "flot par arcs affectés en \bar{n} itérations" et "flot par arcs à l'équilibre". Il montre que la méthode de Frank-Wolfe se comporte légèrement mieux que la méthode des moyennes successives². Par ailleurs Horowitz insiste sur la nécessité d'un algorithme "galiléen" (qui appliqué dans des conditions identiques livre des résultats identiques) et sur l'indépendance vis-à-vis du point initial.

Rose et al. (1988) appliquent la méthode de Frank-Wolfe au modèle de Beckmann, pour comparer plusieurs critères de convergence :

- i) d'une part un critère empirique (analogue à celui d'Horowitz) basé sur les écarts entre flots calculés en \bar{n} itérations et flots d'équilibre.
- ii) d'autre part des critères de type algorithmique qui sont des fonctions de valeurs caractéristiques du programme d'optimisation (fonction objectif, borne inférieure sur la valeur optimale), parmi lesquels notamment "l'écart de dualité".

Ces auteurs montrent que les différents critères évoluent dans le même sens (heureusement !) et que l'écart de dualité varie sensiblement au même rythme que le critère empirique.

Retenons de ces deux contributions :

- i) la nécessaire définition de critères de convergence.
- ii) la variété des critères de convergence envisageables.
- iii) la nécessité d'effectuer un certain nombre d'itérations pour réduire suffisamment l'erreur de troncature.
- iv) que les critères proposés sont liés aux arcs du réseau ou au réseau dans son ensemble, mais pas individuellement aux chemins suivis. Cela s'explique sans doute par l'absence d'unicité par rapport aux chemins dans le modèle de Beckmann.

Parmi les autres recherches qui ont contribué à la mesure de l'erreur de convergence, la proposition du critère d'écart de dualité par Murchland (1966) mérite une mention particulière. Citons aussi les travaux qui proposent un algorithme alternatif à la méthode

¹nous dirons encore Monte-Carlo ; ou MSA pour Méthode de Simulation Aléatoire, qui a l'avantage de correspondre à l'acronyme anglo-saxon usuel pour Method of Successive Averages

²mais s'il avait mesuré la convergence atteinte au bout d'un certain temps et non à l'issue de \bar{n} itérations, peut-être la conclusion aurait-elle changé car le calcul du pas de descente dans la méthode de Frank-Wolfe est une opération relativement onéreuse

standard (Frank-Wolfe pour le modèle de Beckmann) et qui comparent les deux méthodes à l'aide de critères de convergence (ex. Schittenhelm 1990, Larsson et Patriksson 1992, Leurent 1995c).

9.1.2.3 Quelques possibilités ignorées

Les études de l'erreur de convergence dans un modèle d'affectation ont ignoré certaines possibilités :

- choisir un algorithme précis capable de réduire à néant l'erreur de convergence.
- traquer l'erreur de convergence dans des modèles plus élaborés que le modèle déterministe homogène de Beckmann.
- choisir des critères de convergence qui utilisent des informations par chemins, et pas seulement liées aux arcs ou aux relations origine-destination.
- étudier de façon théorique la vitesse de convergence.

9.1.2.4 Un projet d'inventaire et d'expérimentation systématiques

Pour systématiser l'étude de l'erreur de convergence dans un modèle d'affectation, nous proposons de mener un inventaire et une comparaison des critères et des algorithmes, cela pour le modèle déterministe homogène mais aussi pour d'autres modèles d'affectation (stochastique, bicritère...).

L'inventaire portera sur les critères de convergence (§9.1.3) et sur les algorithmes (§9.1.4), tant pour le modèle bicritère que pour d'autres modèles d'affectation. La valeur respective des critères et des algorithmes dépend sans doute du modèle d'affectation considéré : après un examen a priori (§9.1.5), nous mènerons une comparaison dans le cas du modèle bicritère (section 9.2).

9.1.3 Inventaire de critères pour mesurer la convergence

Un critère de convergence se formule en fonction des variables endogènes du modèle, évaluées au point courant (plus éventuellement en d'autres points qui lui sont liés). Pour assurer la rigueur, nous imposons que le critère soit continu, qu'il prenne des valeurs positives ou nulles sur l'ensemble de définition des variables endogènes, qu'il s'annule sur l'ensemble des solutions de la formule caractéristique et là seulement.

Donc la spécification du critère est relative à la formule caractéristique, ainsi qu'aux problèmes types de recherche opérationnelle auxquels la formule caractéristique se rattache.

Pour inventorier les principaux critères de convergence en matière d'affectation du trafic, nous les rattacherons aux différents problèmes types rappelés au §8.1.2 : d'abord le programme extrémal (§.1), puis le problème de point fixe (§.2), ensuite l'inéquation variationnelle et le problème de complémentarité non linéaire (§.3). Alors nous examinerons en détail le critère d'écart de dualité qui occupe une place majeure (§.4) : nous en discuterons la calculabilité (§.5), car elle pose parfois des difficultés qui motivent l'emploi de critères plus simples mais moins rigoureux, de type empirique (§.6). Nous terminerons par une évaluation comparative des principaux critères (§.7).

De façon générale, convenons de noter \mathbf{u} le vecteur des variables endogènes qui sont effectivement considérées dans l'algorithme de résolution : \mathbf{u} se compose du vecteur des

flots par chemins \mathbf{f} , ou simplement du vecteur des flots par arcs $\mathbf{x}(\mathbf{f})$, éventuellement de variables duales comme les multiplicateurs \mathbf{w} associés aux contraintes $\mathbf{z}(\mathbf{f}) \leq \mathbf{0}$. Soit U l'ensemble admissible de \mathbf{u} . Soit n l'indice de l'itération courante.

9.1.3.1 Critères pour un programme extrémal

Si \mathbf{u} résout un programme extrémal $\min_{\mathbf{u} \in U} J(\mathbf{u})$ de valeur optimale J^* , alors le critère $C(\mathbf{u}) = J(\mathbf{u}) - J^*$ convient. Dans un programme extrémal à la Beckmann, on ne connaît pas a priori J^* ; souvent on considère un critère non rigoureux $\hat{C}(\mathbf{u}_n) = J(\mathbf{u}_n)$ pour en observer la décroissance et la stabilisation, ou encore la différence $\hat{C}(\mathbf{u}_n) = J(\mathbf{u}_{n-1}) - J(\mathbf{u}_n)$ pour terminer l'algorithme lorsqu'une itération marginale n'apporte qu'un gain très minime ou même une perte. Signalons que comme $J(\mathbf{u}_n) - J^*$ est une série à termes "presque tous" (1) positifs ou nuls, il faut imposer une condition plus contraignante sur $J(\mathbf{u}_{n-1}) - J(\mathbf{u}_n)$ que sur $J(\mathbf{u}_n) - J^*$.

Concluons que pour un programme à la Beckmann, les critères de ce type sont insatisfaisants, soit par défaut de calculabilité, soit par défaut de rigueur.

Dans le cas d'un programme extrémal contraint, le vecteur \mathbf{u} à considérer dépend de l'algorithme choisi. S'il fait intervenir explicitement les multiplicateurs attachés aux contraintes, alors il faut les incorporer au critère, sinon on revient au cas d'un programme extrémal libre. Voir le cas d'un problème de complémentarité non linéaire pour considérer les multiplicateurs.

9.1.3.2 Problème de point fixe

Le problème de point fixe avec une application \mathbf{W} univoque se prête naturellement à l'établissement d'un critère de convergence rigoureux : si le modèle se formule par une équation $\mathbf{u} = \mathbf{W}(\mathbf{u})$, alors la fonction $C(\mathbf{u}) = \|\mathbf{u} - \mathbf{W}(\mathbf{u})\|^2$ constitue un critère de convergence rigoureux, calculable dès que l'on sait évaluer $\mathbf{W}(\mathbf{u})$.

Certains modèles d'affectation vérifient $\mathbf{f} = \mathbf{D}(\mathbf{t}(\mathbf{f}))$ avec \mathbf{D} une fonction de demande et \mathbf{t} une fonction de temps de parcours (d'offre). On peut alors formuler d'autres critères, par exemple $\|\mathbf{f} - \mathbf{D}(\mathbf{t}(\mathbf{f}))\|^2$, ou de façon duale $\|\mathbf{t}(\mathbf{f}) - \mathbf{D}^{-1}(\mathbf{f})\|^2$, ou de façon agrégée $\|\mathbf{x}(\mathbf{f}) - \mathbf{x}(\mathbf{D}(\mathbf{t}(\mathbf{f})))\|^2$ en désignant par $\mathbf{x}(\mathbf{f})$ le vecteur des flux par arcs déduit de \mathbf{f} .

Pour le modèle de Beckmann, on construit trivialement une application \mathbf{t} pourvue des propriétés désirées. *Mais on ne peut associer* à des temps \mathbf{T} un flot unique $\mathbf{D}(\mathbf{T})$ qui soit caractéristique (2).

Pour certains modèles stochastiques, comme le logit linéaire additif (Leurent, 1994b) ou le logit logarithmique (Cf. Leurent 1995d), deux étapes (une seule dans le cas de la demande fixe) suffisent pour construire une telle fonction de demande \mathbf{D} :

¹tous dans le cas du modèle de Beckmann traité avec l'algorithme de Frank-Wolfe

²on peut considérer un ensemble caractéristique de flots associés à un état \mathbf{T} , mais l'absence d'unicité empêche d'aller bien loin, Cf. Patriksson (1994)

- i) dans le cas de la demande variable, on commence par calculer les volumes O-D en appliquant pour chaque O-D $r-s$ une fonction de demande D_{rs} à un niveau de service moyen S_{rs} .
- ii) puis on répartit les flots par O-D entre les différents chemins envisagés selon la règle de répartition du logit.

Cela vaut également pour le modèle bicritère à condition que, pour chaque relation origine-destination, les chemins présentent des prix deux-à-deux distincts.

Dans le cas du logit linéaire additif, l'algorithme dual proposé dans (Leurent 1994c) calcule un gradient $\mathbf{X}(\mathbf{t}^{-1}(\mathbf{T}) - \mathbf{D}(\mathbf{T}))$ dont la norme pourrait servir de critère de convergence.

Retenons le critère $\|\mathbf{f} - \mathbf{D}(\mathbf{t}(\mathbf{f}))\|^2$ que nous appelons "écart de substitution".

9.1.3.3 Inéquation variationnelle et problème de complémentarité non linéaire

Deux solutions s'offrent pour traiter une inéquation variationnelle "Trouver \mathbf{u}^* tel que $\forall \mathbf{u} \in \mathbf{U}, \mathbf{V}(\mathbf{u}^*) \cdot (\mathbf{u} - \mathbf{u}^*) \geq 0$ " : la transformer en un problème de point fixe¹ et considérer des critères de point fixe, ou employer une *fonction d'écart* ("gap function").

Une fonction d'écart associée à l'inéquation variationnelle est une fonction $\psi : \mathbf{U} \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$ telle que $\psi(\mathbf{u}) \geq 0 \forall \mathbf{u} \in \mathbf{U}$ et $\psi(\mathbf{u}) = 0 \Leftrightarrow \mathbf{u} \in \mathbf{U}^*$. Une telle fonction livre évidemment un critère rigoureux, mais pas nécessairement facile à calculer.

On peut considérer la fonction d'écart primal définie par $\psi_P(\mathbf{u}) = \sup_{\mathbf{v} \in \mathbf{U}} \mathbf{V}(\mathbf{u}) \cdot (\mathbf{u} - \mathbf{v})$, justifiée dans le cas où \mathbf{V} est monotone par l'équivalence entre l'inéquation variationnelle et un problème de point selle pour le lagrangien $\mathcal{L}(\mathbf{u}; \mathbf{v}) = \mathbf{V}(\mathbf{u}) \cdot (\mathbf{u} - \mathbf{v})$. Bien d'autres fonctions d'écart ont été proposées (voir Patriksson 1994 pour une liste de quelques possibilités).

Dans le cas d'un problème de complémentarité non linéaire "Trouver \mathbf{f} tel que $\mathbf{f} \geq \mathbf{0}$, $\mathbf{F}(\mathbf{f}) \geq \mathbf{0}$ et $\mathbf{f} \cdot \mathbf{F}(\mathbf{f}) = 0$ ", on se ramène à une inéquation variationnelle ou à un problème de point fixe ; ou, directement, on forme un critère de convergence rigoureux et naturel : $C_{\mathbf{X}}(\mathbf{f}) = \sum_i (\max(-f_i; 0))^2 + \max(-F_i(\mathbf{f}); 0)^2 + |f_i \cdot F_i(\mathbf{f})|$.

9.1.3.4 Le critère d'écart de dualité

La fonction d'écart primal $\psi_P(\mathbf{u}) = \sup_{\mathbf{v} \in \mathbf{U}} \mathbf{V}(\mathbf{u}) \cdot (\mathbf{u} - \mathbf{v})$ se révèle particulièrement utile pour le modèle de Beckmann : en un point \mathbf{u} , le vecteur $\mathbf{V}(\mathbf{u})$ représente les coûts généralisés de déplacement (par chemins ou par arcs, selon la définition de \mathbf{u}) ; donc on obtient un vecteur $\mathbf{v} = \arg \sup_{\mathbf{v} \in \mathbf{U}} \mathbf{V}(\mathbf{u}) \cdot (\mathbf{u} - \mathbf{v})$ en affectant les flots O-D aux plus courts chemins. Soit \mathbf{v}_n un tel vecteur calculé par rapport au point \mathbf{u}_n . Le critère de

¹ Soit \mathbf{U}^* l'ensemble des solutions de cette inéquation variationnelle : on montre que $\mathbf{u}^* \in \mathbf{U}^* \Leftrightarrow \mathbf{u}^* = \mathbf{P}_{\mathbf{U}}^{\mathbf{B}}(\mathbf{u}^* - \mathbf{B}^{-1} \mathbf{V}(\mathbf{u}^*)) = \mathbf{W}(\mathbf{u}^*)$, avec $\mathbf{P}_{\mathbf{U}}^{\mathbf{B}}$ projection sur \mathbf{U} selon la norme déduite de la matrice définie positive \mathbf{B} . Le cas particulier où $\mathbf{B} = \mathbf{I}$ revient à définir $\mathbf{W}_i(\mathbf{u}) = \max\{u_i - V_i(\mathbf{u}), 0\}$.

convergence $\mathbf{V}(\mathbf{u}_n) \cdot (\mathbf{u}_n - \mathbf{v}_n)$ est à la fois rigoureux et calculable ; de plus on l'interprète naturellement comme le temps total perdu du fait que l'affectation \mathbf{u}_n ne s'accorde pas forcément aux plus courts chemins évalués sur la base des temps $\mathbf{V}(\mathbf{u}_n)$.

Plaçons-nous dans le cas d'un modèle d'affectation qui comme celui de Beckmann relève d'un problème type $\min_{\mathbf{u} \in \mathbf{U}} J(\mathbf{u}) = J_1(\mathbf{u}) + J_2(\mathbf{u})$ avec $J_1(\mathbf{u}) = \sum_a \int_0^{x_a(\mathbf{u})} t_a$ et, dans le cas à demande élastique homogène, $J_2(\mathbf{u}) = -\sum_{rs} \int_0^{q_{rs}(\mathbf{u})} D_{rs}^{-1}$ (la définition de J_2 change pour un modèle à intégration d'étapes, ou stochastique, ou bicritère). Alors, si J_1 et J_2 sont convexes, la solution \mathbf{v}_n du programme auxiliaire

$$\min_{\mathbf{v} \in \mathbf{U}} J_1(\mathbf{u}_n) + \nabla J_1(\mathbf{u}_n) \cdot (\mathbf{v} - \mathbf{u}_n) + J_2(\mathbf{v})$$

livre une borne inférieure sur la valeur optimale J^* du programme d'optimisation de base, à savoir $\text{LBE}^{(n)} := J_1(\mathbf{u}_n) + \nabla J_1(\mathbf{u}_n) \cdot (\mathbf{v}_n - \mathbf{u}_n) + J_2(\mathbf{v}_n)$. Un critère possible de convergence est donc

$$\begin{aligned} & J(\mathbf{f}^{(n)}) - \text{LBE}^{(n)} \\ &= J_2(\mathbf{u}_n) - J_2(\mathbf{v}_n) - \nabla J_1(\mathbf{u}_n) \cdot (\mathbf{v}_n - \mathbf{u}_n) \\ &\geq (\nabla J_1(\mathbf{u}_n) + \nabla J_2(\mathbf{v}_n)) \cdot (\mathbf{u}_n - \mathbf{v}_n) \quad \text{puisque } J_2 \text{ est convexe,} \\ &\geq \sum_a t_a^{(n)} \cdot (x_a(\mathbf{u}_n) - x_a(\mathbf{v}_n)) - \sum_{rsk} S_{rs}^{(n)} (f_{rs}^k(\mathbf{u}_n) - f_{rs}^k(\mathbf{v}_n)) \end{aligned}$$

par construction de \mathbf{v}_n , en notant $S_{rs}^{(n)}$ le temps origine-destination minimal entre r et s sur la base des temps par arcs $t_a^{(n)} = t_a(x_a^{(n)})$ et $f_{rs}^k(\mathbf{u})$ le flot du chemin k de l'O-D r - s lié au vecteur \mathbf{u} .

$$\geq \sum_{rsk} (T_{rs}^{k(n)} - S_{rs}^{(n)}) f_{rs}^k(\mathbf{u}_n) \quad \text{car les termes en } \mathbf{v}_n \text{ se compensent mutuellement.}$$

Autrement dit, le critère majore la somme, chemin par chemin, de l'écart entre le temps du chemin $T_{rs}^{k(n)}$ et le temps minimum de l'origine-destination $S_{rs}^{(n)}$, multiplié par le flot du chemin. Rapporté à la somme des flots origine-destination, le critère majore l'écart moyen entre temps supportés et temps minimum : s'il s'annule, alors l'équilibre de Wardrop est atteint ; la réciproque est également vraie.

L'appellation d'écart de dualité provient de ce que la borne inférieure $\text{LBE}^{(n)}$ n'est autre que la valeur de la fonction objectif duale (Cf. Leurent 1995d) au point $\mathbf{T} = \{t_a^{(n)}\}_a$. Donc $J(\mathbf{u}_n) - \text{LBE}^{(n)}$ majore l'écart $J(\mathbf{u}_n) - J^*$ sans qu'il y ait besoin de connaître a priori J^* . Un encadrement plus fin de l'écart $J(\mathbf{u}_n) - J^*$ est fourni par $J(\mathbf{u}_n) - \text{LBE}$, avec LBE initialisée à $-\infty$ et actualisée à chaque itération selon $\text{LBE} := \max\{\text{LBE}; \text{LBE}^{(n)}\}$.

L'écart de dualité de Rose et al. (1988) vaut $(J(\mathbf{u}_n) - \text{LBE}^{(n)}) / \sum_a t_a^{(n)} \cdot x_a(\mathbf{u}_n)$: c'est donc un écart relatif moyen entre temps et temps minimum pour chaque chemin. Une telle moyenne relative favorise sans doute excessivement les chemins courts, déjà sur-représentés puisqu'ils supportent les flots les plus forts.

Les deux autres critères proposés par Rose et al. (1988) équivalent à l'écart de dualité, ou sont moins intéressants que lui. L'indicateur de "relative lower bound gap", c'est-à-dire "l'écart relatif à la borne inférieure", défini comme $1 - \text{LBE}^{(n)} / J(\mathbf{u}_n)$, équivaut à l'écart de dualité au facteur $1/J^*$ près. Quant à l'indicateur $(J(\mathbf{u}_{n-1}) - J(\mathbf{u}_n)) / J(\mathbf{u}_n)$, il équivaut dans le cas de la méthode de Frank-Wolfe à $\alpha^{(n-1)} \nabla J(\mathbf{u}_{n-1}) \cdot (\mathbf{v}_{n-1} - \mathbf{u}_{n-1}) / J(\mathbf{u}_n)$, donc au produit de l'écart de dualité relatif par le pas de descente $\alpha^{(n)}$ qui tend rapidement vers 0. Cet indicateur est donc dégénéré par rapport à l'écart de dualité.

9.1.3.5 Calculabilité de l'écart de dualité

Dans le cas du modèle de Beckmann, le critère d'écart de dualité possède toutes les propriétés désirées : rigueur, calculabilité, interprétation physique. Pour d'autres modèles justiciables d'une formulation à la Beckmann, son calcul requiert davantage d'efforts, dès que la fonction objectif partielle J_2 incorpore des termes exprimés directement en fonction des flots par chemins : c'est d'ailleurs l'évaluation de $J_2(\mathbf{u}_n)$ plus que celle de $J_2(\mathbf{v}_n)$ qui pose problème.

Nous avons montré (Leurent 1994b, Leurent 1995b) que dans le cas où le programme auxiliaire a une solution unique, l'on peut évaluer $J(\mathbf{v}_n)$ plutôt que $J(\mathbf{u}_n)$ afin de former le critère $J(\mathbf{v}_n) - \text{LBE}^{(n)} = J_1(\mathbf{v}_n) - J_1(\mathbf{u}_n) - \nabla J_1(\mathbf{u}_n) \cdot (\mathbf{v}_n - \mathbf{u}_n)$. Cela permet de traiter un modèle logit linéaire avec énumération automatique des chemins, à l'aide uniquement de variables liées aux arcs et aux relations O-D.

Pour le modèle probit ou le modèle bicritère à demande élastique, il est difficile de calculer \mathbf{v}_n , donc le critère d'écart de dualité est peu valable.

9.1.3.6 Le critère empirique lié aux arcs

Lorsqu'aucun des critères rigoureux précédents ne se calcule facilement, on recourt à des pseudo-critères de convergence. Le plus connu est l'écart-type entre flots affectés et flots d'équilibre, calculé arc par arc. Notant \bar{a} le nombre d'arêtes du réseau, $\hat{x}_a^{(n)}$ le flot affecté sur l'arc a à l'issue de l'itération n des calculs, et \bar{x}_a le flot d'équilibre sur cet arc ($\bar{x}_a = \hat{x}_a^{(+\infty)}$ si l'algorithme doit converger vers le vrai équilibre), alors la définition

$$\text{formelle } (1) \text{ est } \text{RMSE}^{(n)} := \sqrt{\frac{1}{\bar{a}-1} \sum_a (\hat{x}_a^{(n)} - \bar{x}_a)^2}.$$

Ce critère est continu, à valeurs positives ou nulles, et il ne s'annule que lorsque les flots par arcs correspondent à la solution du modèle. Mais dans la forme donnée, il nécessite de connaître à l'avance l'état désiré ! On l'utilise en fait sous une forme bâtarde, dans laquelle \bar{x}_a n'est pas le flot d'équilibre mais la moyenne des itérations passées. Il s'agit alors d'un RMSE "relatif" et non "absolu", d'un pseudo-critère pouvant s'annuler en dehors d'un point d'équilibre, et pas nécessairement nul en un point d'équilibre.

¹Nous notons $\text{RMSE}^{(n)}$ conformément à la tradition anglo-saxonne (pour Root Mean Square Error, racine de l'erreur moyenne sur les carrés)

Ce pseudo-critère empirique constitue donc un dernier recours, notamment pour le modèle probit dans lequel, en l'absence de restriction, les chemins sont en nombre combinatoire, sans possibilité pratique, même à temps fixes, de répartir entre eux le trafic de l'origine-destination.

Le RMSE relatif contrebalance son manque de rigueur mathématique par la possibilité d'une interprétation statistique, en terme d'erreur sur la prédiction d'une variable élémentaire telle que le flot écoulé par un arc. Le RMSE mesure un écart-type pour les flots de tous les arcs qui comptent chacun pour un poids identique. Une approximation probabiliste consiste à supposer que les flots calculés $\hat{x}_a^{(n)}$ sont des réalisations de variables aléatoires gaussiennes $\hat{x}_a^{(n)} = \bar{x}_a + \sigma^{(n)} \varepsilon_a$ avec les ε_a des résidus gaussiens centrés normés indépendants.

Alors $\text{RMSE}^{(n)}$ est un estimateur de $\sigma^{(n)}$, dont on déduit un intervalle de confiance à 95% pour $\hat{x}_a^{(n)}$ selon $\hat{x}_a^{(n)} = \bar{x}_a \pm 1.96 \text{RMSE}^{(n)} / \sqrt{a}$. En sens inverse on peut inférer que

$$\bar{x}_a = \hat{x}_a^{(n)} \pm 1.96 \frac{\text{RMSE}_r^{(n)}}{\sqrt{a}}$$

où l'on a remplacé le $\text{RMSE}^{(n)}$ absolu par un $\text{RMSE}_r^{(n)}$ relatif, de façon que le membre de droite contienne uniquement des termes calculables en fonction des calculs des itérations antérieures, sans référence à des variables absolues \bar{x}_a dont la connaissance préalable ôterait tout son sens à une application de l'algorithme.

On voit ainsi que le critère brut du RMSE néglige les disparités entre les arcs, notamment les capacités de transport, tant en capacité ponctuelle d'écoulement qu'en possibilité spatiale d'acheminement à distance, autrement dit en longueur. Mieux vaudrait prédire

$$\bar{x}_a = \hat{x}_a^{(n)} \pm 1.96 N_a \frac{\text{RMSE}_{r,NL}^{(n)}}{\sqrt{a}}$$

avec $\text{RMSE}_{r,NL}^{(n)} := \sqrt{\frac{1}{\sum_a L_a} \sum_a L_a \left(\frac{\hat{x}_a^{(n)} - \bar{x}_a^{(n)}}{N_a} \right)^2}$

en notant N_a la capacité pratique de l'arc a et L_a sa longueur. Nous qualifions ce dernier critère de RMSE relatif normé.

Les critères de type algorithmique, aussi bien de type fonction objectif qu'écart de dualité ou écart de substitution, ne semblent pas se prêter à une interprétation statistique simple, en termes de variables aléatoires.

Rappelons cependant l'interprétation physique naturelle du rapport entre écart de dualité et somme des flots origine-destination : c'est l'écart moyen entre temps d'un chemin et temps minimum de l'origine-destination correspondante.

9.1.3.7 Evaluation comparative

Nous avons passé en revue des critères de signification et de valeur diverses :

- critères liés à une fonction objectif à la Beckmann, sans notion d'écart de dualité.
- critère d'écart de substitution pour un problème de point fixe.
- critère de fonction d'écart pour une inéquation variationnelle.
- critère d'écart de dualité, cas particulier remarquable des fonctions d'écart.
- pseudo-critère empirique.

Le tableau 9A les range et marque notre appréciation par rapport aux exigences que nous avons posées a priori.

Tab. 9A. Evaluation technique des principaux critères de convergence.

TYPE	CRITERE	Cas d'application	EXIGENCES		
			nul \Leftrightarrow solution	facile à calculer	interprétation physique
Empirique	RMSE relatif	Tous	non	oui	moyenne
	RMSE relatif normé	Tous	non	oui	bonne
Fonction objectif à la Beckmann	niveau $J^{(n)} = J(\mathbf{f}^{(n)})$	Seulement si programme extrémal convexe	non	selon modèle	moyenne
	variation $J^{(n-1)} - J^{(n)}$		selon l'algo. ⁽¹⁾	selon modèle	médiocre
Ecart de dualité	$J^{(n)} - LBE^{(n)}$	inéquation variationnelle	oui	selon modèle	bonne
Ecart de substitution	$\ \mathbf{f} - \mathbf{D}(\mathbf{t}(\mathbf{f}))\ ^2$	seulement si \mathbf{D} bien définie	oui	nécessaire d'identifier les chemins	assez bonne

L'évaluation débouche sur un classement des critères par ordre de mérite :

- en tête, l'écart de dualité satisfaisant à tous les égards *là où il est calculable*.
- suivi de l'écart de substitution, limité en termes d'interprétation statistique et aussi en potentialités d'application (valable surtout pour les modèles logit).
- loin derrière les critères basés sur la seule fonction objectif. Notons toutefois que pour les algorithmes chargés de faire décroître la fonction objectif d'itération en itération, ce critère permet un arrêt naturel quand à une étape il n'y a plus décroissance, mais stabilité ou augmentation. Cela signifie que l'on atteint les limites de la précision numérique.
- enfin des RMSE relatifs, critiquables pour leur comportement au voisinage de zéro.

¹oui dans le cas d'une méthode de Frank-Wolfe pour le programme de Beckmann. Non s'il s'agit d'une MSA appliquée au même problème

9.1.4 Inventaire d'algorithmes convergents

Plusieurs références (Sheffi, 1984, Thomas, 1991, Patriksson, 1994, Leurent, 1995d, entre autres) décrivent la méthode des moyennes successives et la méthode de Frank-Wolfe qui s'y apparente : ce sont les deux algorithmes de référence pour l'affectation du trafic, la première car la plus simple et applicable à de très nombreux contextes, la seconde en tant que principale méthode employée pour résoudre avec précision le modèle de Beckmann. Les méthodes courantes de l'optimisation numérique, du premier ordre (de type gradient) ou du second ordre (de type Newton), n'ont été appliquées à l'affectation que dans quelques travaux de recherche : en effet leur mise en oeuvre nécessite le traitement individuel des chemins, chose encore exclue dans les logiciels "du commerce".

Nous nous contentons ici d'indiquer le principe de quelques algorithmes : méthodes des moyennes successives (MSA), de Frank-Wolfe, de décomposition simpliciale (§.1), de génération de colonne (§.2), de décomposition cyclique (en particulier l'égalisation par transvasement, §.3), de l'application contractante pour les problèmes de point fixe (§.4) ; puis une méthode de diagonalisation pour les inéquations variationnelles (§.5) et une méthode de lagrangien augmenté pour traiter les contraintes adjointes (§.6), enfin une approche duale (§.7).

9.1.4.1 Des moyennes successives à la décomposition simpliciale

La méthode des moyennes successives (MSA). On fixe une suite de nombre réels α_n vérifiant $0 \leq \alpha_n \leq 1$, $\sum_n \alpha_n = +\infty$, $\sum_n \alpha_n^2 < +\infty$. Soit \mathbf{u} un vecteur de flots par chemins. Soit C un critère de convergence et ε un seuil de convergence, une erreur tolérable. Pour un modèle d'affectation justiciable d'un programme à la Beckmann $\min_{\mathbf{u} \in U} J(\mathbf{u}) = J_1(\mathbf{u}) + J_2(\mathbf{u})$ avec $J_1(\mathbf{u}) = \sum_{\alpha} \int_0^{x_{\alpha}(\mathbf{u})} t_{\alpha}$ et J_2 fonction convexe dépendant du principe d'affectation, la MSA se compose de trois étapes :

- [0] Fixer un état initial \mathbf{u}_0 , initialiser le compteur d'itérations $n := 0$.
- [1] Déterminer \mathbf{v}_n solution de $\min_{\mathbf{v} \in U} J_1(\mathbf{u}_n) + \nabla J_1(\mathbf{u}_n) \cdot (\mathbf{v} - \mathbf{u}_n) + J_2(\mathbf{v})$.
- [2] Actualiser $\mathbf{u}_{n+1} := \mathbf{u}_n + \alpha_n (\mathbf{v}_n - \mathbf{u}_n)$.
- [3] Si $C(\mathbf{u}_{n+1}) < \varepsilon$ alors terminer, sinon faire $n := n + 1$ et retourner au pas 1.

Autrement dit, la MSA affecte des flots, actualise les temps de parcours, et réoriente une fraction α_n des flots en fonction des nouveaux temps.

La méthode de Frank-Wolfe est une variante de la MSA, dans laquelle les pas de descente α_n ne sont pas fixés a priori mais déterminés en cours de route dans une étape intermédiaire qui s'insère entre [1] et [2] :

- [1bis] Déterminer α_n solution de $\min_{\alpha \in [0;1]} J(\mathbf{u}_n + \alpha(\mathbf{v}_n - \mathbf{u}_n))$.

Lorsque l'on sait évaluer J simplement, cette étape intermédiaire revient à une recherche "en ligne" facile.

Décomposition simpliciale. Au lieu de faire une combinaison convexe entre l'état courant \mathbf{u}_n et un point extrême \mathbf{v}_n comme dans les deux méthodes précédentes, la

décomposition simpliciale calcule le point courant comme la meilleure combinaison convexe de plusieurs points extrêmes obtenus lors des itérations antérieures.

9.1.4.2 Génération de colonnes

On appelle génération de colonne la recherche d'un (plus court) chemin pour une relation origine-destination : en effet une description de chemin forme une colonne de la matrice d'agrégation par maillon des flots sur chemins (la matrice Δ de la section 8.1). En traitant le modèle de Beckmann, on constate que le nombre de chemins effectivement utilisés sur chaque relation origine-destination est faible (rarement plus de 3 ou 4).

Un algorithme de génération de colonne comprend deux types de tâches, d'une part la génération de colonne à proprement parler et d'autre part la répartition des flux par origine-destination entre les chemins, c'est-à-dire entre les colonnes. L'importance relative donnée à chacune des deux tâches varie selon que l'on souhaite réduire les temps de recherche des plus courts chemins ou le temps de calcul d'une meilleure répartition. Le mode de répartition n'est pas nécessairement agrégé : si l'on peut se contenter d'un même jeu de coefficients pour l'ensemble des relations origine-destination, on peut également particulariser un jeu de coefficients pour chaque relation (cas de l'algorithme DSD de Larsson et Patriksson 1992). En optimisation numérique, un algorithme de génération de colonne appartient à la famille des méthodes de décomposition-coordination (Cohen 1994, Culioli 1994).

9.1.4.3 Algorithme de décomposition cyclique et méthode d'égalisation par transvasement

Si l'on décompose le vecteur \mathbf{f} : par exemple $\mathbf{f} = (\mathbf{f}_1, \dots, \mathbf{f}_{r-1}, \mathbf{f}_r, \mathbf{f}_{r+1}, \dots, \mathbf{f}_{\bar{r}})$ en notant $\mathbf{f}_r = (f_{rs}^k)_{sk}$ le vecteur des flots sur les chemins issus d'une même origine r vers toutes les destinations s , alors la fonction objectif $J(\mathbf{f})$ se réécrit $J(\mathbf{f}_1, \dots, \mathbf{f}_{r-1}, \mathbf{f}_r, \mathbf{f}_{r+1}, \dots, \mathbf{f}_{\bar{r}})$. Le choix de \mathbf{f}_r comme variable de pivot n'est pas le seul possible.

Une méthode de *décomposition cyclique* utilise cette forme en résolvant cycliquement un sous-problème pour chaque origine, et en itérant le cycle principal. Il existe deux schémas usuels pour le cycle principal, celui de Gauss-Seidel et celui de Jacobi qui mènent respectivement aux sous-problèmes suivants dans le cycle i pour l'origine r :

$$\min_{\mathbf{f}_r \geq 0} J(\mathbf{f}_1^{(i+1)}, \dots, \mathbf{f}_{r-1}^{(i+1)}, \mathbf{f}_r, \mathbf{f}_{r+1}^{(i)}, \dots, \mathbf{f}_{\bar{r}}^{(i)}) \quad (\text{Gauss-Seidel})$$

$$\min_{\mathbf{f}_r \geq 0} J(\mathbf{f}_1^{(i)}, \dots, \mathbf{f}_{r-1}^{(i)}, \mathbf{f}_r, \mathbf{f}_{r+1}^{(i)}, \dots, \mathbf{f}_{\bar{r}}^{(i)}) \quad (\text{Jacobi}).$$

Le schéma de Gauss-Seidel converge plus rapidement dans une application "en série", tandis que le schéma de Jacobi offre des possibilités évidentes de parallélisation.

La méthode de décomposition cyclique jouit d'une interprétation simple en termes de théorie des jeux : il suffit d'assimiler le problème d'affectation à un jeu, dans lequel les joueurs sont les zones d'origine et luttent pour la capacité du réseau. Ainsi la méthode fait figure d'un processus itératif d'ajustement des stratégies de transport, dans lequel chaque joueur, son tour venu, choisit une stratégie en fonction des décisions antérieures des autres joueurs.

La méthode d'égalisation par transvasement est un schéma de Gauss-Seidel avec pour variable de pivot le vecteur $\mathbf{f}_{rs} = (f_{rs}^k)_k$ des flux sur chemins d'une origine-destination. Pour résoudre le sous-problème d'ordre r à l'itération i , on fixe l'origine r et l'on considère successivement chaque destination s : pour la relation origine-destination r - s , on transvase le flot du chemin chargé le plus long vers le chemin le plus court, jusqu'à égaliser les coûts des chemins ou vider le chemin plus long. Cette méthode, appliquée par Gibert (1968) au modèle de Beckmann, a été étendue aux modèles logit et au modèle bicritère à demande élastique agrégée par Leurent (1995c).

9.1.4.4 Une méthode de point fixe applicable à des inéquations variationnelles

Dafermos (1980) a proposé de résoudre une inéquation variationnelle en lui associant un problème de point fixe équivalent. Supposant la forte monotonie de \mathbf{V} , alors, pour toute matrice définie positive \mathbf{G} , en choisissant un réel positif ρ convenable, la solution \mathbf{u}^* est également l'unique solution du problème de point fixe $\mathbf{W}_\rho(\mathbf{u}) = \mathbf{u}$, où l'application \mathbf{W}_ρ associe à \mathbf{u}_0 l'unique solution $\mathbf{u} = \mathbf{W}_\rho(\mathbf{u}_0)$ de l'inéquation variationnelle auxiliaire $\forall \mathbf{v}, (\mathbf{v} - \mathbf{u}) \cdot (\mathbf{G}(\mathbf{u} - \mathbf{u}_0) + \rho \mathbf{V}(\mathbf{u}_0)) \geq 0$.

Avec un choix judicieux de ρ , toute suite d'itérées $(\mathbf{W}_\rho)^n \mathbf{f}_0$ converge vers la solution unique de l'inéquation variationnelle principale.

9.1.4.5 La méthode de diagonalisation, applicable aux inéquations variationnelles

La méthode de diagonalisation consiste à résoudre une suite de problèmes auxiliaires où seuls les termes diagonaux de \mathbf{V} varient en fonction de \mathbf{u} . Précisément, la fonction auxiliaire $\mathbf{V}^{(n)}$ de l'étape n est définie composante par composante selon :

$$\mathbf{V}_i^{(n)}(\mathbf{u}) = \mathbf{V}_i(u_1^{(n)}, \dots, u_{i-1}^{(n)}, u_i, u_{i+1}^{(n)}, \dots, u_i^{(n)})$$

où $\mathbf{u}^{(n)}$ est la solution de l'étape $n-1$. Quand \mathbf{V} est monotone, Florian et Spiess (1982) montrent la convergence locale de la méthode vers un équilibre local \mathbf{u}^* si $\|\mathbf{I} - \mathbf{D}^{-1/2}(\mathbf{u}^*) \mathbf{B}(\mathbf{u}^*) \mathbf{D}^{-1/2}(\mathbf{u}^*)\|_2 < 1$, avec \mathbf{I} matrice identité, \mathbf{B} matrice jacobienne de \mathbf{V} , \mathbf{D} partie diagonale de \mathbf{B} , et $\|\cdot\|_2$ norme matricielle associée à la norme euclidienne des vecteurs.

9.1.4.6 La méthode d'Uzawa pour un lagrangien augmenté, afin de traiter des contraintes adjointes

Pour traiter un programme convexe contraint $\min_{\mathbf{u} \in \mathbf{U}} J(\mathbf{u})$ soumis à $\mathbf{z}(\mathbf{u}) \leq \mathbf{0}$, on définit un lagrangien augmenté

$$\mathcal{L}_\alpha(\mathbf{u}; \mathbf{w}; \tau) = J(\mathbf{u}) + \frac{1}{2\tau} \sum_b (\max\{0; w_b + \tau \cdot z_b(\mathbf{u})\}^2 - w_b^2),$$

dont les points-selles sont identiques à ceux du lagrangien naturel du programme contraint (Larsson et Patriksson, 1992). Pour calculer un point-selle de \mathcal{L}_α , la méthode d'Uzawa enchaîne plusieurs itérations, dont la i -ième se compose d'un programme de

minimisation non contrainte $\min_{\mathbf{u} \in \mathbf{U}} J^{[i]}(\mathbf{u}) = \mathcal{L}_a(\mathbf{u}; \mathbf{w}^{[i]}; \tau)$, puis d'une actualisation des variables duales \mathbf{w} par $w_b^{[i+1]} := \max\{0; w_b^{[i]} + \rho \cdot z_b(\mathbf{u}^{[i]})\}$.

9.1.4.7 Une méthode de sous-gradient pour le programme dual

Si le modèle d'affectation se formule comme un programme convexe $\min_{\mathbf{u} \in \mathbf{U}} J(\mathbf{u})$ en fonction du vecteur de flot \mathbf{u} , alors le vecteur des temps par arcs $\mathbf{p} = \{p_a\}_a$ est solution d'un programme dual associé $\max_{\mathbf{p} \geq \mathbf{t}(0)} G(\mathbf{p})$ ⁽¹⁾. Un sous-gradient $\nabla G(\mathbf{p})$ de G au point \mathbf{p} a pour coordonnées $g_a = -t_a^{-1}(p_a) + \sum_{rs} \delta_{rs}^{ak_p} D_{rs}(T_{rs}^{k_p}(\mathbf{p}))$ avec k_p un plus court chemins entre r et s sur la base des temps \mathbf{p} . On peut résoudre le programme dual en montant dans la direction $\nabla G(\mathbf{p})$ à partir de \mathbf{p} , avec une amplitude fonction de la valeur courante de la fonction duale, d'une borne supérieure sur sa valeur optimale et de la norme du gradient ⁽²⁾.

9.1.4.8 Remarque

Parmi les algorithmes présentés, certains traitent entièrement et uniquement une affectation sans contrainte : MSA, Frank-Wolfe, décomposition simpliciale, génération de colonne, décomposition cyclique. La méthode duale s'applique également à l'affectation sous contrainte de capacité (Leurent 1994b). Les autres algorithmes, application contractante, diagonalisation, lagrangien augmenté, servent à décomposer un problème plus complexe en une suite d'itérations dont chacune comprend une affectation sans contrainte : cette deuxième classe emploie un algorithme de la première classe en tant que sous-traitant.

9.1.5 Récapitulation des critères et des algorithmes, par modèle d'affectation

9.1.5.1 Jugement a priori

Vis-à-vis des cinq principaux types d'affectation : Beckmann, bicritère, logit linéaire, logit logarithmique, probit, les critères de convergence et les algorithmes que nous avons inventoriés ont une portée variable. Nous portons une appréciation a priori dans le tableau 9B.

Voici la liste des algorithmes considérés pour chaque modèle :

- i) **Beckmann.** Frank-Wolfe (FW), les moyennes successives (MSA), la procédure d'égalisation par transvasement (PET).
- ii) **Bicritère.** Une MSA sans explicitation des chemins (MSA-arc), une MSA avec explicitation progressive des chemins notée MSA-iti, une PET avec explicitation progressive.

¹où $G(\mathbf{p}) := \min_{\mathbf{u} \geq 0, \mathbf{w}} J(\mathbf{u}) + \sum_a w_a \cdot (p_a - t_a(\mathbf{u}))$

²Fukushima 1984 pour le modèle de Beckmann, Leurent 1994b pour le modèle logit linéaire à demande élastique

iii) **Logit linéaire.** Une MSA avec énumération implicite des itinéraires (MSA-arc), une MSA avec identification des chemins notée MSA-iti, la méthode duale notée Polyak du nom de l'inventeur de l'algorithme de sous-gradient, la PET.

iv) **Logit logarithmique.** Une MSA-iti et la PET.

v) **Probit.** Une MS-arc, une MSA SAM et une FW SAM où l'acronyme SAM marque un calcul approché de la répartition entre itinéraires concurrents (Maher et Hughes 1995).

Les algorithmes convergent tous vers une solution de la formule caractéristique du modèle, à l'exception des méthodes pour traiter le probit SAM ou le logit logarithmique.

Tab. 9B. Applicabilité des critères aux algorithmes, selon le modèle.

Modèle	Algorithme	Critère empirique	Fonction objectif à la Beckmann	Variation de fonction objectif	Ecart de dualité	Ecart de substitution
Beckmann	FW	+	+	+	+	
	MSA	+	+		+	
	PET	+	+	+		
Bicritère	MSA-arc	+				
	MSA-iti	+	+		+	
	PET	+	+	+		
Logit linéaire	MSA-arc	+	+		+	+
	Polyak	+	+		+	+
	MSA-iti	+	+	+		+
	PET	+	+	+		+
Logit logarithmique	MSA-iti	+				+
	PET	+				+
Probit	MSA-arc	+				
	MSA SAM	+	+		+	+
	FW SAM	+	+		+	+

9.1.5.2 Commentaires

L'écart de dualité n'est pas marqué pour la méthode de transvasement. Pour le calculer, il suffit de choisir l'endroit dans l'organigramme où faire une répartition à temps fixes.

L'écart de substitution ne s'applique que dans le cas des modèles logit. Lorsque le traitement des chemins est implicite (cela concerne le logit linéaire avec une énumération implicite des chemins par la procédure STOCH3, Cf. Leurent 1994b), l'écart doit être calculé au niveau des arcs et non des chemins.

La fonction objectif à la Beckmann ne sert, pour un modèle pourvu d'une formulation extrémale convexe (ce qui exclut par exemple le logit logarithmique), que si l'algorithme employé permet de l'évaluer. Cela exclut le modèle bicritère à demande élastique résolu sans expliciter les chemins. Lorsqu'un algorithme ne réduit pas nécessairement la fonction objectif à chaque nouvelle itération, on considère non pas la valeur de la fonction objectif au point courant, mais son minimum sur toutes les itérations depuis la première.

La variation de fonction objectif s'applique moins fréquemment que le niveau de la fonction objectif, puisque ce critère n'a pas de sens pour les algorithmes qui ne font pas nécessairement décroître la fonction objectif à chaque nouvelle itération.

9.2 APPLICATION AU MODÈLE BICRITÈRE

Dans le cas particulier du modèle d'affectation bicritère, nous allons décrire les algorithmes et les critères de convergence (§9.2.1), pour les tester sur une étude de trafic de taille moyenne (§9.2.2). Enfin nous tirerons quelques conclusions (§9.2.3).

9.2.1 Algorithmes et critères

Pour calculer un équilibre bicritère à demande élastique agrégée sans contraintes adjointes, nous avons proposé (Leurent 1993a) une méthode des moyennes successives (MSA) qui ne nécessite pas d'explicitement les chemins. On peut également implanter une MSA avec explicitation progressive des chemins (Leurent 1993b). Dial (1996) et Marcotte et Zhu (1994) ont proposé une méthode de Frank-Wolfe pour traiter le cas de la demande fixe sans expliciter les chemins : la différence, par rapport au traitement du modèle de Beckmann (Cf. §9.1.4.1), est de déterminer un point extrémal non par une simple recherche de plus court chemin, mais par une recherche paramétrique de plus court chemin qui est une adaptation de la méthode du simplexe paramétrique. Enfin nous avons adapté la méthode d'égalisation par transvasement qui suppose d'identifier les chemins (Leurent 1995c). Pour prendre en compte les contraintes adjointes, il est possible d'appliquer une méthode de lagrangien augmenté (Leurent 1995d).

Nous décrirons successivement les applications au modèle bicritère *symétrique* (i.e. les temps de parcours par arcs dérivent d'un potentiel) de la MSA avec explicitation des chemins (§.1), de la MSA sans explicitation des chemins (§.2), de la procédure d'égalisation par transvasement (§.3), d'une méthode de lagrangien augmenté (§.4). Nous terminerons par un examen des critères de convergence possibles (§.5).

Nous reprenons les notations de la section 8.3, en ajoutant le symbole C_{rs} pour désigner la collection des chemins de la relation origine-destination $r-s$.

9.2.1.1 La MSA avec explicitation progressive des chemins

Nous définissons une Répartition Bicritère à Temps Fixes (RBTF) comme une répartition du flot origine-destination entre les chemins concurrents selon le principe bicritère, avec un ensemble donné de chemins et des temps de parcours maintenus constants. Une RBTF calcule les parts de marché $(f_{rs}^k/q_{rs})^{(n)}$ des itinéraires connus, puis le volume de l'O-D $r-s$ selon $q_{rs}^{(n)} = D_{rs}(\bar{S}_{rs}^{(n)})$ et les flots par chemins selon $f_{rs}^{k(n)} = q_{rs}^{(n)}(f_{rs}^k/q_{rs})^{(n)}$, qu'enfin elle charge sur le réseau.

Si les prix des chemins P_{rs}^k découlent de prix par arcs p_a via $P_{rs}^k = \sum_a \delta_{rs}^{ak} p_a$, la MSA avec explicitation progressive des chemins se compose des étapes suivantes.

[0] *Initialisation.*

[0.1] Mettre à zéro le compteur d'itérations : $n := 0$. Choisir une suite de nombres réels α_n tels que $0 \leq \alpha_n \leq 1$, $\sum_n \alpha_n = +\infty$, $\sum_n \alpha_n^2 < +\infty$.

[0.2] Sur la base des temps par arcs $t_a(0)$, calculer un plus court chemin $k_{rs}^{(0)}$ par relation O-D $r-s$.

[0.3] Initialiser les collections de chemins : $\forall r, s \quad \mathbf{C}_{rs} := \{k_{rs}^{(0)}\}$.

[0.4] Déduire des flots par chemins $\mathbf{f}^{(0)}$ d'une RBTF basée sur les temps $t_a(0)$.

[1] *Mettre à jour les temps de parcours.*

[1.1] Fixer $t_a^{(n)} := t_a(x_a(\mathbf{f}^{(n)}))$.

[2] *Recherche de direction.*

[2.1] Rechercher des plus courts chemins $k_{rs}^{(n)}$ sur la base des temps généralisés par arcs $t_a^{(n)} + p_a / v^{[r](n)}$, où $v^{[r](n)}$ est une valeur du temps tirée au hasard.

[2.2] $\forall r, s$, inclure dans \mathbf{C}_{rs} le chemin $k_{rs}^{(n)}$ s'il ne s'y trouve déjà.

[2.3] Par une RBTF basée sur les temps $t_a^{(n)}$, calculer des flots par chemins $\mathbf{g}^{(n)}$.

[3] *Mettre à jour le vecteur de flot courant.*

[3.1] Fixer $\mathbf{f}^{(n+1)} := \mathbf{f}^{(n)} + \alpha^{(n)} (\mathbf{g}^{(n)} - \mathbf{f}^{(n)})$.

[4] *Test de convergence.*

[4.1] Appliquer un test de convergence. S'il réussit, terminer, sinon faire $n := n + 1$ et retourner en [1].

9.2.1.2 La MSA sans explicitation des chemins

Pour appliquer la MSA sans expliciter les chemins, on modifie l'algorithme précédent :

- supprimer les pas [0.3] et [2.2] pour ne plus considérer les collections \mathbf{C}_{rs} .
- considérer les flots par arcs $x_a(\mathbf{f}^{(n)})$ et $x_a(\mathbf{g}^{(n)})$, et non les flots par chemins.
- dans le cas de la demande variable, stocker des temps généralisés composites par relation origine-destination $\bar{S}_{rs}^{(n)}$ et $\tilde{S}_{rs}^{(n)}$. Initialiser au pas [0.2] $\bar{S}_{rs}^{(0)} := T_{rs}^{k_0}$. Calculer au pas [2.1] $\tilde{S}_{rs}^{(n)}$ comme le temps généralisé de $k_{rs}^{(n)}$. Enfin, mettre à jour $\bar{S}_{rs}^{(n+1)} := \bar{S}_{rs}^{(n)} + \alpha^{(n)} (\bar{S}_{rs}^{(n)} - \tilde{S}_{rs}^{(n)})$ avant le pas [3.1].

- au pas [3.1], calculer $x_a(\mathbf{g}^{(n)})$ en fixant $q_{rs}(\mathbf{g}^{(n)})$ à $D_{rs}(\overline{S}_{rs}^{(n+1)})$ et en l'affectant à $k_{rs}^{(n)}$. Même changement au pas [0.4] par rapport à $k_{rs}^{(0)}$.

9.2.1.3 La procédure d'égalisation par transvasement

Exposons maintenant comment appliquer la procédure d'égalisation par transvasement (PET) au modèle bicritère. Nous notons $K_{rs}^{k_1 k_2}(\mathbf{f}; \alpha)$ le vecteur $\mathbf{g} = \{g_{rs}^k\}_{rs k}$ qui ne diffère de \mathbf{f} que sur les deux composantes $g_{rs}^{k_1} = f_{rs}^{k_1} - \alpha$ et $g_{rs}^{k_2} = f_{rs}^{k_2} + \alpha$: \mathbf{g} est un \mathbf{f} modifié en transvasant, sur la relation origine-destination r - s , le flot α du chemin k_1 au chemin k_2 . Quand les prix des chemins découlent des prix par arcs p_a via $P_{rs}^k = \sum_a \delta_{rs}^{ak} p_a$, la PET se compose des étapes suivantes.

[0] Initialisation.

- [0.1] Initialiser le compteur d'itérations : $n := 0$.
- [0.2] Initialiser les collections de chemins : par exemple $\forall r, s \quad \mathbf{C}_{rs} := \{k_{rs}^{(0)}\}$ où $k_{rs}^{(0)}$ est un plus court chemin calculé sur la base des temps de parcours $t_a(0)$.
- [0.3] Charger le réseau en affectant le trafic de l'O-D r - s sur le chemin $k_{rs}^{(0)}$.

[1] Progression d'ensemble.

- [1.1] Incrémenter le compteur d'itérations $n := n + 1$. Choisir la première origine r .

[2] Augmenter la collection de chemins.

- [2.1] Depuis le sommet d'origine r , tirer au hasard" (selon la distribution des valeurs du temps) une valeur du temps $v^{[r](n)}$ et chercher les plus courts chemins, sur la base des temps généralisés par arcs $t_a(x_a(\mathbf{f})) + p_a/v^{[r](n)}$.
- [2.2] Pour chaque destination s , si le plus court chemin ainsi calculé entre r et s n'est pas déjà dans \mathbf{C}_{rs} , l'y placer.
- [2.3] Choisir la première destination s .

[3] Equilibrer l'O-D r - s .

- [3.1] Si la collection \mathbf{C}_{rs} est vide ou n'a qu'un élément, aller au Pas 4.
- [3.2] Prendre dans \mathbf{C}_{rs} un chemin chargé de temps maximum, k_1 de charge $f_{rs}^{k_1}$, et un chemin de temps minimum, k_2 .
- [3.3] Si $I_{rs}^{k_1}(\mathbf{f}) \leq I_{rs}^{k_2}(\mathbf{f}) + \varepsilon_T$, alors aller au Pas 4. Sinon, soit $I_{rs}^{k_1}(K_{rs}^{k_1 k_2}(\mathbf{f}; f_{rs}^{k_1})) \geq I_{rs}^{k_2}(K_{rs}^{k_1 k_2}(\mathbf{f}; f_{rs}^{k_1})) + \varepsilon_T$ auquel cas on retient $\bar{\alpha} := f_{rs}^{k_1}$, soit on recherche la solution $\bar{\alpha}$ du problème $I_{rs}^{k_1}(K_{rs}^{k_1 k_2}(\mathbf{f}; \alpha)) = I_{rs}^{k_2}(K_{rs}^{k_1 k_2}(\mathbf{f}; \alpha))$ pour $\alpha \in [0; 1]$.

[3.4] Actualiser $\mathbf{f} := K_{rs}^{k_1 k_2}(\mathbf{f}; \bar{\alpha})$. Retourner au pas [3.2].

[4] *Nouvelle destination et nouvelle origine.*

[4.1] Si s n'est pas la dernière destination, alors considérer la destination suivante et aller au pas [3.1].

[4.2] Si s est la dernière destination, si r n'est pas la dernière origine, alors considérer l'origine suivante et retourner au pas [2.2].

[5] *Test de convergence.*

[5.1] Soit $\mathbf{f}^{(n)}$ le vecteur de flot par chemins obtenu à la fin de la n -ième itération. Si $\mathbf{f}^{(n)}$ satisfait un test de convergence, alors Terminer. Sinon, retourner au pas [1.1].

Dans le cas de la demande élastique, mieux vaut recourir à la formulation de la demande en excès et spécifier $I_{rs}^k = T_{rs}^k + \sum_{i=m_{rs}(k)}^{\bar{m}_{rs}-1} (\bar{P}_{rs}^i - \bar{P}_{rs}^{i+1}) \eta'_{rs}(q'_{rs,i} / q_{rs})$ pour la fonction d'impédance plutôt que la définition standard (Cf. section 8.2). On définit un chemin supplémentaire fictif d'indice $k = 0$, qui est un "complément de flot O-D par rapport à un point de la courbe de demande", avec pour impédance

$$I_{rs}^0 = D_{rs}^{-1}(q_{rs}) - \left\{ \eta_{rs}(1) \bar{P}_{rs}^{\bar{m}_{rs}} + \sum_{i=1}^{\bar{m}_{rs}-1} (\bar{P}_{rs}^i - \bar{P}_{rs}^{i+1}) \left(\eta_{rs} \left(\frac{q'_{rs,i}}{q_{rs}} \right) - \frac{q'_{rs,i}}{q_{rs}} \eta'_{rs} \left(\frac{q'_{rs,i}}{q_{rs}} \right) \right) \right\}.$$

Pour une parfaite rigueur, il faudrait, dans une recherche de plus courts chemins, chercher pour chaque O-D non pas un seul chemin mais un ensemble de chemins tels que pour chaque valeur du temps l'un d'entre eux soit efficace. Pour simplifier l'on peut se contenter de chercher un seul chemin à la fois depuis chaque origine au Pas 2, chemin tel qu'il soit efficace pour une certaine valeur du temps $v^{[r](n)}$ tirée au hasard selon la loi de probabilité propre à l'O-D. Mais alors, dans le test de convergence, il faut imposer par prudence la stabilité de la collection des itinéraires pendant la dernière itération et si possible quelques précédentes.

9.2.1.4 Une méthode de lagrangien augmenté pour traiter les contraintes adjointes

Si l'on adjoint des contraintes de capacité $x_a(\mathbf{f}) - L_a \leq 0$ au programme $\min_{\mathbf{f} \geq 0} J(\mathbf{f})$, le lagrangien augmenté

$$\mathcal{L}_a(\mathbf{f}; \mathbf{w}; \tau) = J(\mathbf{f}) + \frac{1}{2\tau} \sum_a (\max\{0; w_a + \tau(x_a(\mathbf{f}) - L_a)\}^2 - w_a^2)$$

possède les mêmes points-selles que le lagrangien naturel du programme contraint. Pour déterminer un point-selle, la méthode d'Uzawa procède par itérations, la i -ième itération consistant à :

- chercher $\mathbf{f}^{[i]}$ solution de $\min_{\mathbf{f} \geq 0} J^{[i]}(\mathbf{f}) = J_1^{[i]}(\mathbf{f}) + J_2(\mathbf{f})$ avec $J_1^{[i]}(\mathbf{f}) = \sum_a \int_0^{x_a(\mathbf{f})} t_a^{[i]}$ où $t_a^{[i]}(\theta) = t_a(\theta) + \max\{0; w_a^{[i]} + \tau(x_a(\mathbf{f}^{[i-1]}) - L_a)\}$, et $J_2(\mathbf{f}) = \sum_{rs} [q_{rs} \{ \sum_{i=1}^{\bar{m}_{rs}} \bar{P}_{rs}^i (\eta_{rs}(\frac{q'_{rs,i}}{q_{rs}}) - \eta_{rs}(\frac{q'_{rs,i-1}}{q_{rs}})) \} - \int_0^{q_{rs}} D_{rs}^{-1}]$.

- actualiser $w_a^{[i+1]} := \max\{0; w_a^{[i]} + \rho(x_a(\mathbf{f}^{[i]}) - L_a)\}$.
- tester la stabilité des $w_a^{[i+1]}$ (par rapport aux $w_a^{[i]}$), et terminer s'il y a lieu.

Notons que le vecteur $\mathbf{f}^{[i]}$ peut servir de point initial pour la minimisation libre de l'itération suivante.

9.2.1.5 Critères de convergence

Examinons l'applicabilité des critères de convergence présentés au §9.1.3 aux algorithmes indiqués pour traiter le modèle bicritère. Les pseudo-critères RMSE relatifs s'appliquent mais ne sont pas rigoureux.

Le niveau de fonction objectif s'applique pour la MSA avec explicitation des chemins et pour la PET, mais pas pour la MSA sans explicitation (car alors on ne peut évaluer $J_2(\mathbf{f})$). La décroissance de fonction objectif ne vaut que pour la PET, puisque pour une MSA elle n'est pas assurée d'une itération à l'autre. L'écart de dualité ne convient pas, car il faudrait calculer un état auxiliaire de façon exacte et non approchée. L'écart de substitution ne convient pas dès qu'une classe monétaire compte plus d'un chemin.

Au total, la MSA sans explicitation n'est justiciable que d'un RMSE relatif donc non rigoureux, la MSA avec explicitation comme la PET sont justiciables d'un RMSE relatif et du niveau de la fonction objectif. Dans le cas de la demande fixe, la méthode de Frank-Wolfe procure un écart de dualité (Marcotte et Zhu 1994).

Dans les applications numériques qui suivent, nous remédions au manque de critères rigoureux de convergence en appliquant *préalablement* la PET afin d'obtenir une solution de référence. La solution de référence livre à la fois la valeur optimale de la fonction objectif ce qui rend rigoureux le critère de niveau de la fonction objectif ⁽¹⁾, et les flots par arcs à l'équilibre, qui nous servent à calculer des RMSE absolus donc rigoureux.

9.2.2 Applications numériques

Nous avons appliqué trois algorithmes à une situation sans contraintes de capacité. Une application avec contraintes de capacité est rapportée dans Leurent (1996a).

9.2.2.1 Conditions expérimentales

Les données de l'application. La société COFIROUTE nous a aimablement communiqué deux jeux de données codant respectivement un réseau routier urbain à l'ouest de Paris (Cf. figure 9c) et une matrice origine-destination des déplacements à l'heure de pointe du soir.

Le réseau comprend environ 2 000 arêtes unidirectionnelles. Les fonctions de temps de parcours sont de type BPR, paramétrées selon $t_a(x) = t_a^0(1 + 0.15(\frac{x}{C_a})^4)$, avec C_a capacité pratique ⁽²⁾. Il y a quinze arcs à péage (valeurs entre 7 F et 15 F).

¹de fait on se sert de la valeur optimale comme d'une borne inférieure dans le calcul d'un écart de dualité

²cela pour les arcs qui ne sont pas des connecteurs. Pour ces derniers, on suppose que le temps de parcours ne dépend pas du niveau de trafic

Il y a 141 zones de demande, donc près de 20 000 paires O-D. On considère des fonctions de demande d'élasticité constante ε_0 : $q_{rs} = D_{rs}(t) = q_{rs}^0 (t / t_{rs}^0)^{\varepsilon_0}$ dans laquelle $(t_{rs}^0 ; q_{rs}^0)$ est un point de référence particulier à la relation O-D r - s . On fixe la valeur de ε_0 à -0.6. On a retenu comme distribution des valeurs du temps une loi log-normale de médiane 60 F/h, et d'écart-type de loi du logarithme naturel 0.6. A chaque itération, depuis chaque origine, on recherche les chemins de temps généralisé minimal pour une valeur du temps tirée au hasard.

Fig. 9c. Schéma du réseau des premières études A86 ouest.



Le logiciel d'application. Tous les algorithmes testés sont mis en oeuvre à l'aide d'un même logiciel, MARS (Leurent 1995d). Ainsi la comparaison entre algorithmes porte non seulement sur le nombre d'itérations, mais encore et davantage sur le temps total passé, calculé équitablement.

Le matériel informatique. Les calculs sont effectués sur un micro-ordinateur portable compatible IBM-PC, équipé d'un processeur Pentium cadencé à 120 MHz (donc représentatif des outils actuels courants dans un bureau d'études de transport).

9.2.2.2 Détail des tests

Nous avons d'abord appliqué la PET avec au plus trois transvasements par relation O-D dans une itération et, dans chaque transvasement, un seuil de temps variable selon l'itération ($\varepsilon_T = 10^{-5}$ h à la première itération puis selon une suite géométrique de raison

0.85). Nous avons arrêté l'algorithme à l'issue de la 65^{ème} itération ⁽¹⁾, en limite de la précision numérique du logiciel MARS ⁽²⁾.

L'état courant des variables endogènes nous a servi de "*solution de référence*" ⁽³⁾, par abus de langage et faute de pouvoir calculer plus précisément la solution véritable. Celle-ci n'a vraisemblablement pas été atteinte : à partir de la 36^{ème} itération la valeur courante de la fonction objectif n'a plus nécessairement décré d'une itération à la suivante, contrairement au principe de l'algorithme. Dans l'analyse des résultats, nous garderons ce fait en mémoire et nous comparerons les autres méthodes non seulement au résultat "final" de la PET, mais encore au résultat de la 36^{ème} itération.

Nous basant sur la "*solution de référence*" et en particulier la valeur finale J^* de la fonction objectif, nous avons appliqué successivement la PET (65 itérations), puis la MSA-iti qui explicite les chemins (avec un pas de descente à l'étape n égal à $1/(n+1)$ et 200 itérations), enfin la MSA-arc sans explicitation des chemins (mêmes pas de descente mais 400 itérations). Pour chaque algorithme nous avons mesuré les critères de convergence suivants :

- RMSE absolu et RMSE absolu normé,
- RMSE relatif et RMSE relatif normé,
- ainsi que la valeur de la fonction objectif pour la PET et la MSA-iti.

9.2.2.3 Comparaison en nombre d'itérations

Dans l'application de référence de la PET, nous avons constaté une décroissance progressive de la fonction objectif jusqu'à la 36^{ème} itération. Ensuite des incidents de parcours se sont produits de plus en plus fréquemment⁴ : la valeur atteinte à la 36^{ème} itération, $J^{(36)} = 165\,985.60$ véh.h, est très proche du minimum de la série " $J^* = J^{(64)} = 165\,981.14$ véh.h.

La figure 9d représente l'évolution du critère lié à la fonction objectif en fonction du nombre d'itérations n . Elle montre pour la PET comme pour la MSA-iti la valeur non pas de $J^{(n)}$ mais du logarithme naturel $\log(J^{(n)} / J^* - 1)$ qui mesure plus nettement la précision.

En nombre d'itérations, la PET se révèle l'algorithme le plus efficace : en 10 itérations elle réduit davantage J que la MSA-iti en 200 !

Les critères de RMSE confirment la supériorité de la PET : tant le critère relatif avec norme (figure 9e ⁽⁵⁾) que le critère absolu avec norme (figure 9f).

¹pour un seuil $\varepsilon_T = 2.6 \cdot 10^{-10}$ h

²une commande de 70 itérations a engendré un bouclage infini à la 67^{ème}... Dans MARS les nombres réels sont codés sur six octets

³caractérisée en termes de flots par chemins et par arcs, ainsi que de valeur de la fonction objectif

⁴précisément, aux itérations 37, 40, 43, 44, 47, 49, 51, 52, 53, 55, 56, 57, 59, 60, 63, 65

⁵signalons que le RMSE relatif sans norme se comporte de façon tout-à-fait analogue au RMSE relatif avec norme, à un facteur de l'ordre de 2000 près

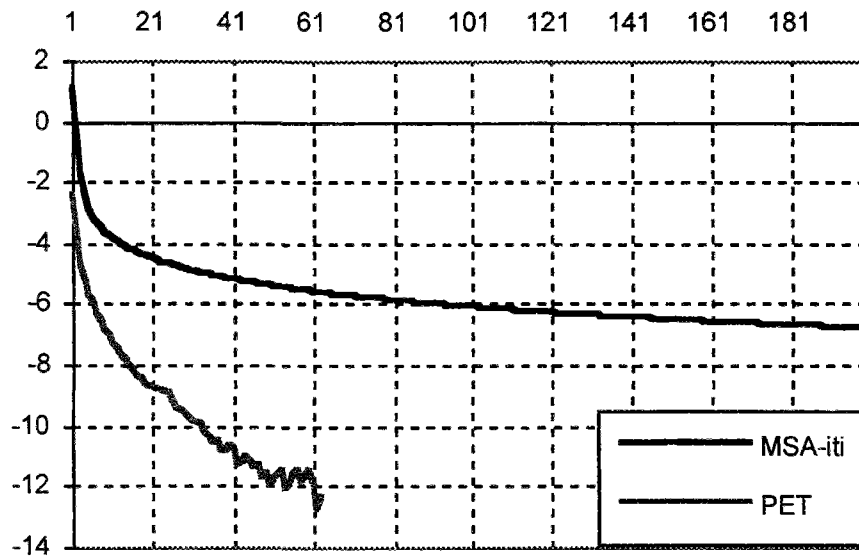
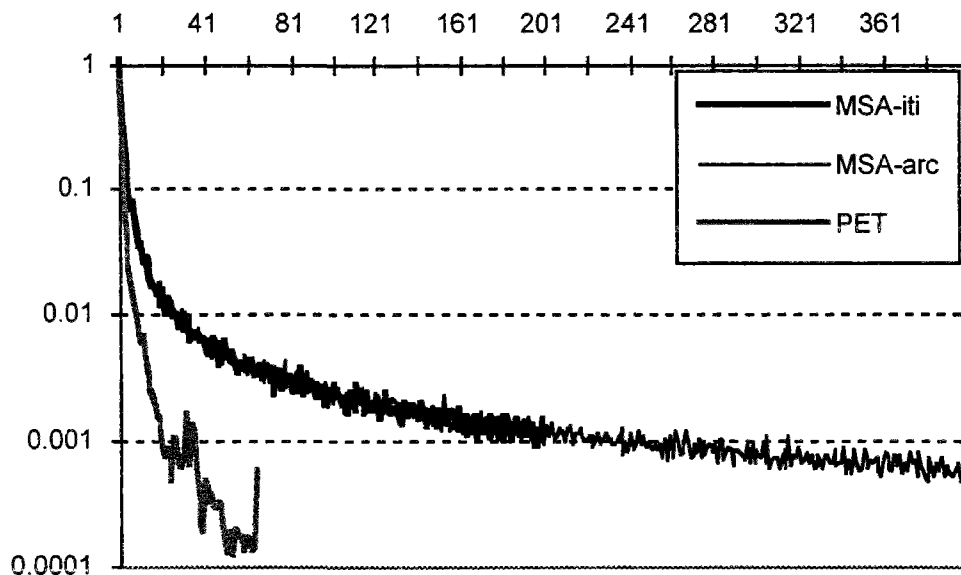
Fig. 9d. Evolution de $\log(J^{(n)} / J^* - 1)$ en fonction du nombre d'itérations n .

Fig. 9e. RMSE relatif normé, en fonction du nombre d'itérations (échelle logarithmique).



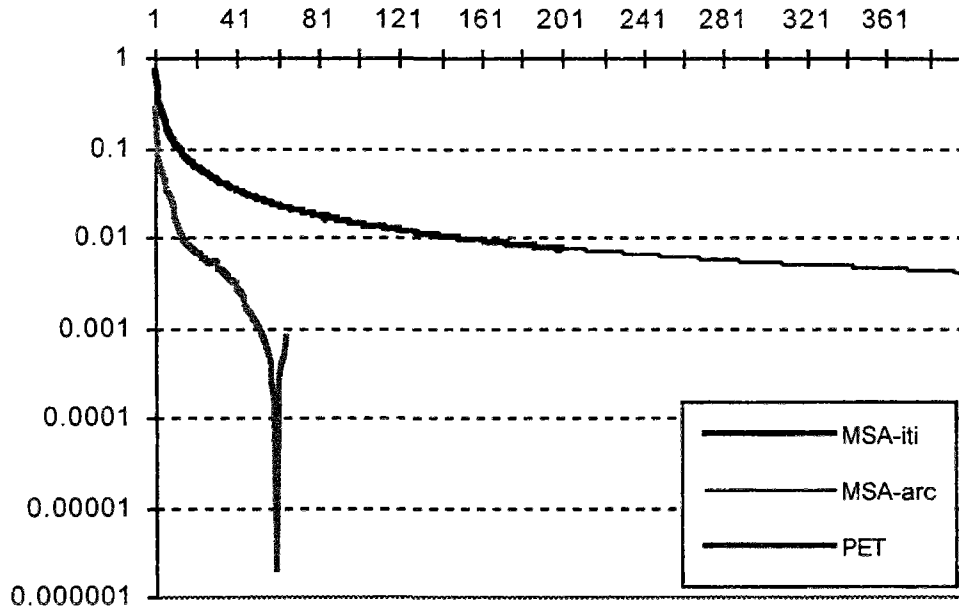
Pour le RMSE absolu avec norme il ne faut pas donner une grande importance au comportement de la PET après la 40ème itération, en raison des imprécisions numériques qui entraînent des incidents de parcours.

Entre les deux méthodes des moyennes successives, la MSA-iti et la MSA-arc, les évolutions des critères RMSE sont presque identiques. Ce constat, surprenant a priori car l'explicitation des chemins fournit une meilleure direction de descente dans la MSA-iti, tient probablement au faible nombre d'arcs à péage dans l'application traitée. En effet, dans les situations sans péage, l'explicitation des chemins n'influe pas sur la convergence de la MSA.

Le nombre limité d'arcs à péage pourrait également expliquer la supériorité apparente de la PET dans le cas bicritère, car dans le cas monocritère la PET est supérieure aux méthodes de combinaison convexe (Frank-Wolfe et MSA, Cf. Schittenhelm 1990 et Leurent 1995c).

Dans de petites applications avec une forte proportion d'arcs à péage (Leurent 1995e, f), la PET et la MSA-iti se sont révélées grossièrement équivalentes et supérieures à la MSA-arc : l'explication tient à la convergence linéaire des méthodes de combinaison convexe avec explicitation des chemins dans les modèles à différenciation de la demande (explicite ou implicite), constatée pour le modèle logit linéaire par Leurent (1995b) et pour le modèle bicritère à demande fixe par Marcotte et Zhu (1996).

Fig. 9f. RMSE absolu normé, en fonction du nombre d'itérations (échelle logarithmique).



Dans l'application traitée ici, la PET présentait une convergence linéaire de taux 0.8 entre les itérations 10 (suffisamment éloignée du démarrage) et 36 (la dernière avant les incidents de parcours). La figure 9g représente la régression linéaire de $Y = \|\mathbf{x}^{(n+1)} - \mathbf{x}^\infty\|$ par $X = \|\mathbf{x}^{(n)} - \mathbf{x}^\infty\|$ où $\mathbf{x}^{(n)}$ est le vecteur courant des débits par arcs à l'itération numéro n , pour les points n entre 10 et 36. Cette régression donne $Y = 0.00097 + 0.7882 X$ avec un coefficient de corrélation $R = 0.984$. La pente correspond au taux de convergence de l'algorithme (Cf. §5.3.5.2).

Le taux de convergence de la MSA-iti, mesuré entre les itérations 50 et 150 est quant à lui très proche de 1, comme le montre la figure 9h qui représente la régression linéaire de $Y = \|\mathbf{x}^{(n+1)} - \mathbf{x}^\infty\|$ par $X = \|\mathbf{x}^{(n)} - \mathbf{x}^\infty\|$ où $\mathbf{x}^{(n)}$ est le vecteur courant des débits par arcs à l'itération numéro n , pour les points n entre 50 et 150. Les résultats exacts sont $Y = 0.0001 + 0.9797 X$ avec un coefficient de corrélation $R = 0.9999$.

On peut probablement améliorer le taux de convergence de la MSA-iti en spécifiant autrement les pas de descente.

Fig. 9g. Taux de convergence de la PET entre les itérations 10 et 36.

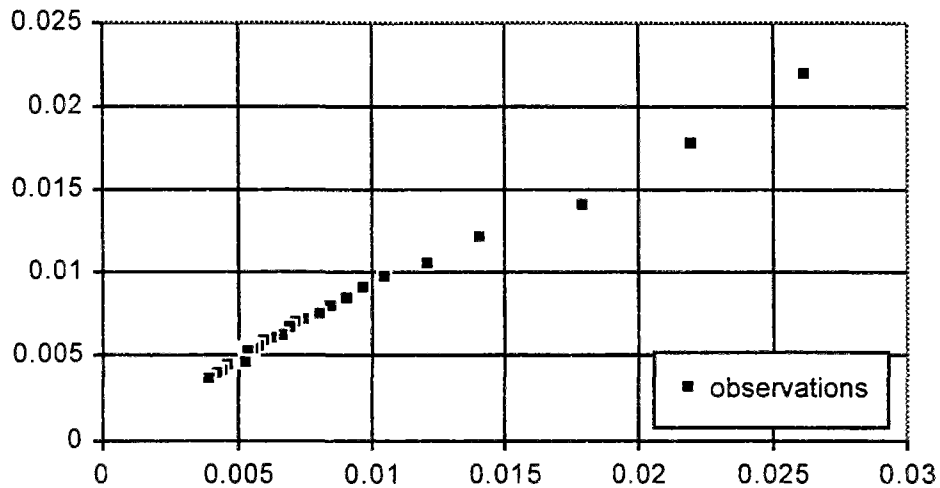
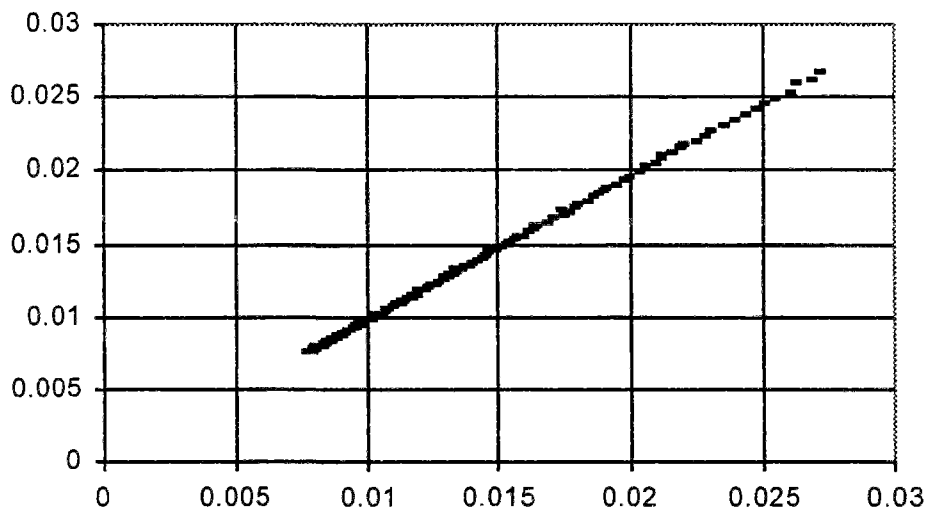


Fig. 9h. Taux de convergence de la MSA-iti.



9.2.2.4 Temps de calcul des algorithmes

En temps de calcul total, la comparaison des algorithmes paraît moins favorable à la PET. Les 65 itérations de celle-ci ont consommé 1 650 minutes, contre 450 minutes pour 200 itérations de la MSA-iti et 127 minutes pour 400 itérations de la MSA-arc ! Ce bilan total inciterait à préférer la MSA-iti à la PET dans une application opérationnelle, s'il ne fallait l'infléchir car les dernières itérations de la PET n'améliorent guère la précision tout en consommant l'essentiel du temps de calcul. Les 36 premières itérations de la PET consomment "seulement" 140 minutes, soit en moyenne 3.9 minutes par itération, à comparer aux 2.2 minutes par itération de la MSA-iti.

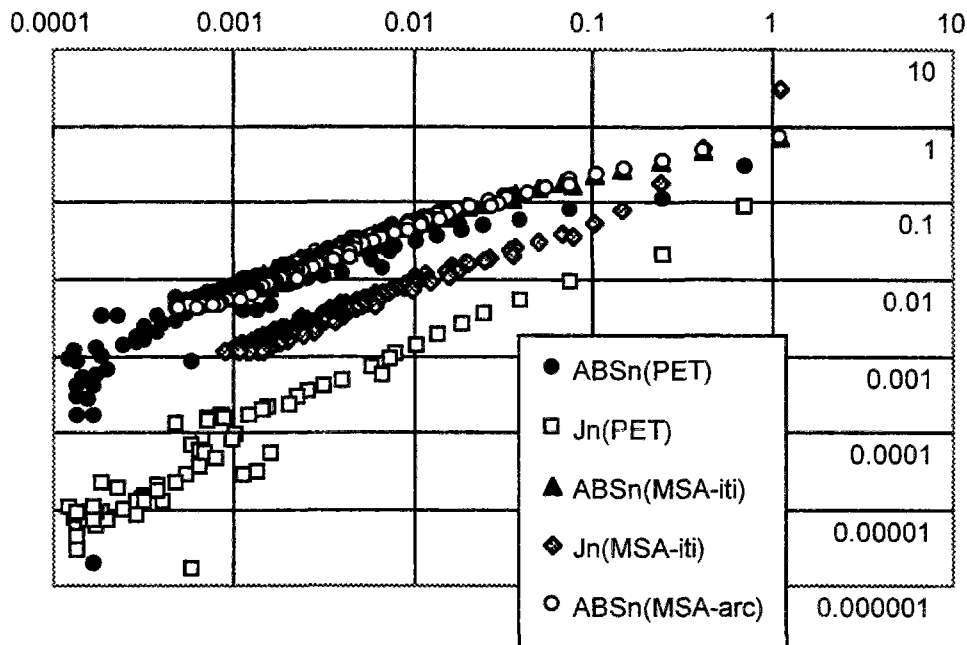
Ainsi notre comparaison des algorithmes conclut à une supériorité de la PET dans une situation typique d'infrastructure routière à péage en milieu urbain, pour calculer une solution précise dans un temps raisonnable. Signalons que les performances tant de la PET que des MSA sont perfectibles, en modifiant le réglage des divers paramètres ; la

détermination de paramètres optimaux relève davantage de l'analyse numérique que de l'audit.

9.2.2.5 Comparaison empirique des critères de convergence

La figure 9i compare les critères de convergence pour les trois algorithmes testés. Pour la PET, nous avons tracé deux séries de points avec autant d'observations que d'itérations : une série avec en abscisse le RMSE relatif avec norme et en ordonnée le RMSE absolu avec norme, l'autre série avec la même abscisse et en ordonnée la valeur $J^{(n)}/J^{*-1}$ liée à la fonction objectif. Même chose pour la MSA-iti, tandis que seule la série avec les deux RMSE normés (relatif et absolu) a été tracée pour la MSA-arc.

Fig. 9i. RMSE relatif normé vs. RMSE absolu normé ou vs. $J^{(n)}/J^{*-1}$, (échelle logarithmique).



En échelle logarithmique nous observons des relations approximativement affines entre les différents critères. Tous décroissent vers zéro ⁽¹⁾, à des allures différentes mais proportionnelles.

Pour les MSA et les premières itérations de la PET, le RMSE absolu décroît 10 fois plus vite que le RMSE relatif ; mais pour la PET en dessous du seuil de 10^{-3} , les deux RMSE se rapprochent.

Pour la MSA avec explicitation des chemins, le critère lié à la fonction objectif et le RMSE relatif évoluent de façon étonnamment comparable (les points de la série sont sur la diagonale principale). Dans le cas de la PET, le critère de fonction objectif décroît légèrement plus vite que le RMSE relatif.

¹vers le point origine (0 ; 0) qui serait dans le coin en bas à gauche de la figure 9i

9.2.3 Eléments de bilan

9.2.3.1 Critères de convergence pour le modèle bicritère

Deux conclusions pratiques s'imposent :

i) le critère du RMSE relatif, bien qu'approximatif, est un bon substitut des critères rigoureux que sont le RMSE absolu et $\{J^{(n)}/J^*-1\}$. Un seuil de 10^{-3} pour le RMSE relatif normé équivaut à un seuil de 10^{-4} pour $\{J^{(n)}/J^*-1\}$ avec la PET, à un seuil de 10^{-3} pour $\{J^{(n)}/J^*-1\}$ avec la MSA-iti et à un seuil de 10^{-2} pour le RMSE absolu normé avec tous les algorithmes testés. Cela remédie de façon pragmatique au manque d'un critère de convergence rigoureux pour nos algorithmes de résolution du modèle bicritère à demande élastique ⁽¹⁾.

ii) mesurer l'erreur numérique sur le débit d'un arc à l'aide du RMSE relatif en lieu et place du RMSE absolu (Cf. §9.1.3.6) conduit à une *sous-estimation* importante de l'erreur, d'un facteur environ 10 dans notre application !

Le pseudo-intervalle de confiance à 95% pour le ratio entre débit et capacité pratique d'un arc, exprimé au §9.1.3.6 comme $\bar{x}_a / N_a = \hat{x}_a^{(n)} / N_a \pm 1.96 \text{RMSE}_{NL}^{(n)} / \sqrt{\bar{a}}$, donne dans la situation traitée $\bar{x}_a / N_a = \hat{x}_a^{(n)} / N_a \pm 4.4 \cdot 10^{-2} \text{RMSE}_{NL}^{(n)}$. Au seuil de $\text{RMSE}_{r,NL}^{(n)} = 10^{-2}$, la largeur du pseudo-intervalle de confiance pour \bar{x}_a / N_a est évaluée à $8.8 \cdot 10^{-4}$ si l'on utilise le RMSE relatif, alors qu'elle vaut $8.8 \cdot 10^{-3}$ si l'on emploie le RMSE absolu ! Pour un arc de capacité pratique $N_a = 5\,000$ véh/h, le pseudo-intervalle sur le débit x_a a pour "vraie" largeur 44 véh/h, faussement évaluée à 4.4 véh/h si l'on ne corrige pas le RMSE relatif.

9.2.3.2 Algorithmes pour traiter le modèle bicritère

Plusieurs algorithmes existent pour résoudre le modèle bicritère à demande élastique. Dans le cas d'une application de taille moyenne, avec un nombre limité d'arcs à péage, la procédure d'égalisation par transvasement permet une résolution précise en un temps raisonnable ; de plus elle produit des résultats par itinéraires.

¹signalons toutefois une possibilité de test rigoureux, pour la MSA-iti comme pour la PET. Pour chaque relation O-D, il s'agit de rechercher les chemins optimaux pour les bornes INF k et SUP k de chaque plage de valeurs du temps affectées à un itinéraire k . Si les chemins optimaux ainsi déterminés procurent à la valeur du temps correspondante le même coût généralisé que l'itinéraire k , alors on a effectivement déterminé la frontière d'efficacité de l'O-D dans le diagramme prix-temps.

Chapitre 10

L'audit économétrique : une quantification

L'erreur économétrique d'un modèle de simulation englobe l'erreur d'estimation liée à la spécification numérique des relations de dépendance, ainsi que l'erreur exogène a priori relative aux variables exogènes. Elle dépend aussi de la validité conceptuelle, formelle et algorithmique du modèle de simulation.

L'audit économétrique proposé au chapitre 5 attache au résultat de la simulation une incertitude a posteriori modélisée, que l'on obtient en propageant via la formule caractéristique l'erreur relative aux variables exogènes. Il s'agit de rendre transparent au plan statistique le modèle de simulation, en explicitant l'incertitude qui résulte des hypothèses d'application.

Pour préparer l'audit économétrique d'un modèle d'affectation, nous dressons une synthèse de diverses ressources (section 10.1). D'une part des modèles d'estimation pour déterminer les variables exogènes relatives aux paramètres du comportement, ou aux flots origine-destination, ou aux fonctions de demande, ou aux fonctions de temps de parcours. D'autre part une procédure pour propager l'erreur relative aux variables exogènes via le modèle d'affectation.

Ensuite, à l'aide de divers jeux d'observations, nous indiquons quelques ordres de grandeur pour l'erreur exogène relative aux fonctions de temps de parcours, aux flots origine-destination et aux valeurs du temps (section 10.2). Pour estimer les paramètres d'une distribution continue de la valeur du temps, nous proposons un modèle de choix discret entre itinéraires et une méthode du maximum de vraisemblance.

Enfin nous élaborons une méthode pour propager l'erreur exogène via le modèle d'affectation bicritère ; le cas général requiert une résolution numérique des formules de propagation, tandis que le réseau à deux arcs se prête à une expression analytique directe de l'erreur exogène a posteriori en fonction de l'erreur exogène a priori (section 10.3).

Au total, les principaux ingrédients d'un audit économétrique de l'affectation bicritère à demande fixe se trouvent réunis. Nous propageons l'erreur exogène dans une étude routière en milieu interurbain, relative à la liaison entre Bordeaux et Pau (section 10.4).

10.1 RESSOURCES ÉCONOMÉTRIQUES POUR L'AFFECTATION

Un examen de l'état économétrique de l'affectation révèle quatre principaux objets d'estimation (§10.1.1). Nous exposons les principes des modèles d'estimation relatifs à ces objets (§10.1.2). Enfin nous présentons des résultats mathématiques puissants qui livrent des formules pour propager l'erreur exogène (§10.1.3).

10.1.1 Qualification économétrique du domaine

De nombreuses contributions ponctuelles fournissent des éléments pour l'audit économétrique d'un modèle d'affectation ; cependant les tentatives d'intégration sont rares et incomplètes.

10.1.1.1 Une reconnaissance progressive des objets d'estimation

Le développement et le perfectionnement des modèles d'affectation ont révélé progressivement des objets d'estimation : fonction de temps de parcours, flots origine-destination, fonctions de demande, paramètres du comportement. Le statut d'objet d'estimation n'a pas été immédiatement reconnu : par exemple les flots origine-destination furent d'abord acceptés comme le produit d'un modèle de distribution spatiale, sans examen économétrique.

Les examens économétriques ont été menés de façon ponctuelle, surtout pour les fonctions de demande et les paramètres du comportement. Un intérêt spécifique s'est manifesté pour les fonctions de temps de parcours au début des années 1960, puis pour les flots origine-destination au début des années 1980.

10.1.1.2 Prise en compte de l'erreur exogène

A notre connaissance, il y a eu deux tentatives concrètes pour prendre en compte l'erreur exogène et la propager à travers le modèle d'affectation : l'une menée en Angleterre par Ashley (1981) qui a produit une procédure incluse au Traffic Appraisal Manual (TAM, DOT 1981-1994), l'autre menée en France par Roger Marche (1988 à 1991) pour étudier des voies rapides urbaines à péage.

La procédure du TAM est entièrement conforme au schéma de principe pour l'audit économétrique d'un modèle de simulation :

- application à une situation de référence pour laquelle on dispose d'observations de débit.
- position d'une erreur exogène a priori relative aux flots origine-destination et propagation de cette erreur à travers le modèle d'affectation.
- comparaison entre erreur a posteriori attendue et erreur a posteriori constatée, en confrontant, pour chaque route avec une observation de débit, d'une part le flot observé et la bande d'incertitude résultant de l'erreur d'observation, d'autre part le flot calculé et la bande d'incertitude liée à l'erreur exogène propagée.

Cependant la procédure du TAM n'est pas présentée avec une justification théorique, mais plutôt comme une recette ; surtout, elle ne considère qu'un objet d'estimation, les flots origine-destination.

Nous avons déjà décrit l'approche économétrique de Roger Marche (Cf. §3.4.2) : elle intègre plusieurs objets d'estimation, à savoir les flots origine-destination, la distribution

des valeurs du temps et les temps de parcours, mais elle aliène la dimension spatiale du problème d'affectation et considère seulement deux arcs ; par ailleurs la propagation est menée à demande fixe et temps fixes, donc pour un modèle bicritère fruste.

Les deux approches sont complémentaires. Notre audit économétrique les unifie et les prolonge.

10.1.1.3 L'état de l'art manifeste des carences explicables

En matière d'alimentation économétrique de l'affectation, l'état de l'art est inégal selon l'objet d'estimation (Cf. §10.1.2) : relativement avancé pour les paramètres du comportement et les flots origine-destination, rudimentaire pour les fonctions de demande et les fonctions de temps de parcours. L'audit économétrique se réduit à un simple constat, hormis dans les études qui obéissent au TAM ou de façon très ponctuelle.

Cet état de l'art tient sans doute aux spécificités économétriques de l'affectation : complexité du traitement technique, variété et parfois difficile mesurabilité des objets d'estimation, sous-détermination de certains problèmes d'estimation donc nécessité de modèles annexes pour combler les lacunes.

La complexité du traitement technique, plus encore peut-être son obscurité pour des économètres peu familiers de l'optimisation numérique, explique le peu de liens entre pratique et recherche, entre d'une part analyses empiriques de sensibilité donc études de propagation, et d'autre part mesure de l'erreur d'estimation donc d'une erreur à propager.

La variété des objets d'estimation (fonctions de coût généralisé des chemins, niveaux des flots O-D, fonctions de demande, lois de temps de parcours) explique quant à elle le caractère fragmentaire des contributions, d'autant plus que le lien entre les incertitudes provenant des différents objets n'est pas établi au niveau des résultats (1). Le recueil d'observations requiert parfois des moyens considérables, particulièrement pour les enquêtes avant et après modification de l'offre de transport qui servent à estimer les fonctions de demande : le coût de revient rend problématique la mesurabilité.

Enfin certains problèmes d'estimation sont sous-déterminés, ils comportent beaucoup plus de variables endogènes qu'il n'y a d'observations disponibles en pratique. En particulier le nombre de relations O-D excède souvent de beaucoup le nombre des déplacements saisis dans une enquête, ou le nombre de postes de comptage sur le réseau. Cette sous-détermination pousse à recourir à des modèles extérieurs au mécanisme d'affectation proprement dit, par exemple à des modèles de distribution en ce qui concerne les flots O-D. La complexité technique en est renforcée lorsque le modèle d'estimation a lui-même pour fonction de synthétiser de l'information à partir d'observations partielles.

10.1.2 Modèles d'estimation, selon l'objet

Présentons maintenant les principaux modèles d'estimation relatifs respectivement aux paramètres du comportement (§.1), aux flots origine-destination (§.2), aux fonctions de

¹sans compter une autre difficulté technique : l'explicitation des chemins nécessaire aux calculs de propagation

demande (§.3), enfin aux fonctions de temps de parcours (§.4). Signalons que ces modèles sont généralement employés de façon indépendante (§.5).

10.1.2.1 Paramètres du comportement : des modèles de choix discret pour simuler la répartition du trafic entre itinéraires concurrents

Le coût généralisé d'un chemin synthétise tout ce qui constitue ou diminue la résistance au déplacement : temps passé, prix etc. (Cf. §7.1.2). On suppose généralement qu'un déplacement s'affecte au chemin qui lui offre un temps généralisé minimum ; donc la répartition du trafic entre itinéraires concurrents dépend de la formule du temps généralisé. Habituellement, la formule est une somme pondérée des divers facteurs tels que le temps pur ou le prix ; les coefficients de pondération (ou paramètres du comportement) sont des taux de change entre le facteur et un facteur de référence, ainsi la valeur du temps entre le temps pur et le prix.

On estime les paramètres du comportement à l'aide d'un modèle de choix discret entre plusieurs variantes d'un même bien ; on suppose que chaque variante est choisie avec une fréquence qui dépend des caractères de toutes les variantes et des individus. Sous certaines conditions, la fréquence d'une variante pour un individu correspond à la probabilité qu'une fonction d'utilité de la variante pour l'individu soit maximale sur l'ensemble des variantes.

Il suffit d'accorder le temps généralisé et la fonction d'utilité pour modéliser la répartition du trafic entre itinéraires concurrents comme un choix discret en théorie de l'utilité aléatoire du consommateur. Par exemple on peut exprimer l'utilité d'un chemin k pour un individu i comme

$$U_i(k)(\omega) = -G_i(k) + \xi_i(k)(\omega)$$

avec $G_i(k)$ le coût généralisé du chemin k pour l'individu i , ω un aléa, $\xi_i(k)(\omega)$ une erreur aléatoire. Alors la fréquence de choix de k par i est :

$$P_i(k) = \Pr(\{U_i(k) \geq U_i(l) \quad \forall l\}) = \Pr(\{U_i(k) \geq \max_{l \neq k} U_i(l)\})$$

Dans le modèle logit multinomial, on stipule que les erreurs aléatoires $\xi_i(k)$ sont indépendantes et identiquement distribuées selon une loi de Weibull de variance $\pi^2 / 6\theta^2$. Alors

$$P_i(k) = \frac{\exp(-\theta G_i(k))}{\sum_{\ell} \exp(-\theta G_i(\ell))}$$

Cette simplicité analytique permet l'estimation de structures complexes de choix (Daly 1987) : aussi le modèle logit est-il de loin le modèle de choix discret le plus largement répandu, sous sa forme multinomiale ou sous la forme de logit à étages (le "nested logit" de Ben-Akiva 1973). Notons que Gaudry et Wills (1978) préconisent d'opérer des transformations de Box-Cox sur les termes aléatoires du logit.

Le modèle probit ouvre une autre possibilité, en supposant que les $\xi_i(k)$ sont des variables aléatoires gaussiennes multivariées : très séduisant en théorie car il peut représenter des interactions entre termes aléatoires et des demandeurs différenciés, le probit se révèle peu maniable en pratique.

Dans tous les cas, la méthode d'estimation privilégiée est celle du maximum de vraisemblance (ex. Amemiya 1981).

Le modèle bicritère prix-temps donne l'exemple d'une différenciation déterministe des demandeurs. Les méthodes d'estimation en vigueur suivent des principes de moindres carrés (Marche 1973, Papon 1988) ; mentionnons un modèle à la fois prix-temps et logit (Bradley et al., 1992).

Pour récapituler, des cadres probabilistes existent ainsi que des méthodes d'estimation. Dans les cas favorables où il y a une méthode pratique de calcul, celle-ci est mise à la disposition des praticiens dans des logiciels commerciaux applicables au logit et à ses dérivés, ou au probit, mais pas au prix-temps.

Les modèles de choix modal ont fait l'objet de très nombreuses applications des modèles de choix discret. Koppelman (1976) indique comment propager une erreur exogène à travers un modèle logit. Horowitz (1982) spécifie des tests pour détecter et mesurer l'erreur économétrique dans un modèle logit. Daganzo (1979) pour le probit et Ben-Akiva et Lerman (1985) pour le logit dressent de larges synthèses. Pour l'affectation, les seules publications sont, au sujet du logit linéaire, Robillard (1974), Fisk (1977) et Ben-Akiva et al. (1984) ; au sujet du logit logarithmique et du prix-temps, Leurent (1995e) que nous reprenons au §10.2.3.

10.1.2.2 Flots O-D : des modèles de distribution spatiale pourvus d'une interprétation probabiliste

Originellement conçus à partir d'analogies avec la physique, les principaux modèles de distribution spatiale¹ (gravitaire, opportunitaire, facteurs de croissance, entropique) ont été progressivement pourvus d'interprétations probabilistes afin de les estimer avec des données de comptages (Robillard 1975 ; Van Zuylen et Willumsen 1980 et de nombreux autres).

En particulier le modèle entropique (Wilson 1967) repose sur l'hypothèse probabiliste que voici : que les flots O-D suivent une distribution multinomiale dont on observe l'état le plus vraisemblable. Alors les flots q_{rs} entre les zones d'origine désignées par r et les zones de destination désignées par s résolvent le problème de maximisation d'entropie $\max W(\{q_{rs}\}_{rs}) = (\sum_{rs} q_{rs})! / \prod_{rs} q_{rs}!$ (2).

En contraignant ce programme selon diverses hypothèses relatives à des sommes par origine ou par destination des flots, ou à des comptages, on retrouve le modèle gravitaire, ou la méthode des facteurs de croissance... Les contraintes représentent des informations fragmentaires, des observations versées au modèle d'estimation des flots O-D.

Il existe de nombreuses contributions relatives à l'estimation des flots O-D. Elles se différencient les unes des autres selon :

¹présentation de nombreuses méthodes par Ortuzar et Willumsen (1990), ou Bierlaire (1991)

²c'est une fonction de vraisemblance. La log-vraisemblance associée s'exprime comme $\log W = -\sum_{rs} q_{rs} \log \frac{q_{rs}}{\sum_{rs} q_{rs}}$ grâce à l'approximation de Stirling, $\log(x!) \approx x(\log(x) - 1)$

- les données d'observation (entre enquête origine-destination et comptages).
- l'interprétation probabiliste (théorie de l'échantillonnage chez la plupart, approche bayésienne pour Maher 1983).
- les hypothèses probabilistes sur les flots O-D¹ et sur les comptages², sur les marges (Bell, 1983).
- le modèle d'affectation sous-jacent (éventuellement dynamique).
- la méthode d'estimation (maximum de vraisemblance pour la plupart quoique tous ne le mentionnent pas explicitement ; moindres carrés généralisés pour Cascetta 1984).

Chacune des contributions citées développe une méthode d'estimation. Celle-ci, en sus de son résultat principal (estimation des valeurs moyennes des paramètres), peut livrer l'incertitude attachée à son résultat. Cette incertitude devient à son tour un input pour le modèle d'affectation, au travers duquel on peut la propager afin d'en évaluer la contribution à l'incertitude exogène sur les résultats de l'affectation. Wyley et al. (1981) montrent comment propager une incertitude relative aux flots O-D à travers l'affectation. Bell (1985) indique comment propager des erreurs relatives aux marges ou aux comptages à travers un modèle d'estimation de type entropique.

En fait, il manque un banc d'essai comparatif des diverses méthodes proposées (3). Un tel banc d'essai pourrait comprendre une discussion des hypothèses probabilistes ainsi que des comparaisons numériques.

10.1.2.3 Fonctions de demande

L'estimation d'une dépendance fonctionnelle entre le volume O-D et le niveau de service (temps généralisé moyen) de l'O-D, a fait l'objet de contributions dont la portée dépend du mode examiné :

- pour les transports collectifs, les données de billétique permettent de nombreuses études d'élasticité de la demande au tarif et à la qualité de service (Cf. synthèse de Goodwin, 1992).
- pour les modes individuels, dont l'automobile, les procédures d'enquête disponibles n'ont pas du tout la même valeur. Au niveau d'un territoire régional ou national, on mesure des élasticités du trafic automobile au prix des carburants, du péage, à la taille du réseau autoroutier (Madre 1989 pour la France entière). Mais à notre connaissance il n'existe pas de mesure empirique au niveau des origines-destinations. Signalons le travail préparatoire de Bonsall (1991) qui met en évidence la nécessité de tenir compte des transferts entre itinéraires, entre modes, entre horaires, ce que confirment des études empiriques (comme Loos et al. 1992).

Rares sont les logiciels d'affectation "du commerce" qui laissent représenter l'élasticité de la demande de façon directe, à l'aide de fonctions de demande propres à un mode et à

¹des variables aléatoires multinomiales pour Wilson 1967 puis bien d'autres, gaussiennes pour Maher 1983 ou poissonniennes indépendantes pour Spiess 1985

²constants ou VA poissonniennes ou gaussiennes

³dont certaines sont mises à disposition des praticiens dans des logiciels du commerce, entre autres EMME/2 et TRIPS

une relation origine-destination. Plus souvent les logiciels offrent une élasticité par report modal (EMME/2, TRIPS) ou spatial (POLYDROM).

10.1.2.4 Fonctions de temps de parcours

Dans un modèle statique, les fonctions de temps de parcours sont des relations entre le temps de parcours d'un arc et le flot écoulé par l'arc pendant la période de référence ; éventuellement on considère les flots d'arcs voisins. L'estimation des relations entre temps de parcours et débit fait l'objet de nombreuses contributions de l'ingénierie du trafic, principalement au sujet du mode routier : consulter le Highway Capacity Manual (HCM 1994), Taylor et Young (1988) et Cohen (1990) pour de nombreuses méthodes et références.

Branston (1976) recense les formules proposées jusqu'alors pour de telles fonctions, entre autres la célèbre fonction du BPR ⁽¹⁾

$$t(x) = t_0 \left(1 + \alpha \left(\frac{x}{C}\right)^\beta\right)$$

où x est le flot de l'arc, t_0 le temps de parcours à flot nul, C la capacité pratique, α et β des paramètres habituellement renseignés à 0.15 et 4.0 respectivement.

D'autres propositions figurent dans Planco (1986), Spiess (1990), Leurent (1991c). Mais plus rares sont les contributions qui expliquent comment estimer ces fonctions. Signalons les travaux de Taylor (1977) pour la fonction de Davidson (1966), ou du SETRA (1984) pour une fonction qui distingue les voitures des poids lourds ; dans les deux cas, l'estimation suit la méthode des moindres carrés.

A notre connaissance, les contributions n'ont pas examiné l'erreur d'estimation des fonctions de temps de parcours en tant d'erreur exogène participant à l'erreur économétrique d'un modèle d'affectation. Autre carence plus remarquable encore : la plupart des modèles routiers agrègent les différents types de véhicules, aussi bien voitures particulières que poids lourds, dans un flot composite où un véhicule d'un type donné est converti en voitures particulières via un coefficient d'équivalence. Cette approximation peut se justifier pour représenter l'influence du trafic sur le temps de parcours subi par un véhicule d'un certain type, mais elle ne devrait pas conduire à traiter tous les véhicules comme des parcelles indiscernables du trafic total ⁽²⁾ !

10.1.2.5 Quelle coordination entre objets d'estimation ?

Souvent les objets d'estimation sont considérés de façon isolée, sans préciser les relations entre les modèles d'estimation. Pourtant une coordination semble nécessaire dans au moins deux circonstances : lors de l'estimation et lors de la propagation de l'erreur exogène.

A notre connaissance les coordinations sont excessivement rares. Au sujet de la propagation de l'erreur exogène, l'approche intégratrice de Marche constitue la seule

¹Bureau of Public Roads

²Une conséquence usuelle est d'assigner aux poids lourds les mêmes vitesses qu'aux voitures. Le cas où cette approximation paraît la moins dicutable est l'affectation statique en milieu urbain et en période de pointe : dans ces conditions, le trafic total comprend peu de poids lourds, et s'il y a congestion les vitesses tendent à l'uniformité.

tentative antérieure. Au sujet de l'estimation, les praticiens traitent un objet puis un autre, avec éventuellement des retours. Ainsi le TAM mentionne plusieurs opérations : ajustement de la matrice, validation du réseau, ajustement du coefficient d'équivalence entre distance et temps de parcours ; sans marquer d'ordre ni de dépendance entre elles. Delons (1991) ajuste simultanément les paramètres du comportement, les fonctions de temps de parcours et la matrice O-D, en minimisant une fonction d'écart globale dans laquelle les observations relatives aux différents objets sont pondérés par expertise : une telle approche simultanée constitue un important progrès, à consolider en explicitant un modèle probabiliste global.

10.1.3 La méthode de Tobin et Friesz pour propager l'erreur exogène à travers le modèle d'affectation

Lorsque la formule caractéristique du modèle est une inéquation variationnelle, des résultats théoriques généraux livrent des formules pour l'analyse de sensibilité (AS) et la propagation de l'incertitude. Récapitulons ces résultats en traduisant une synthèse due à Tobin et Friesz (1988), avant de les adapter à nos notations pour l'affectation bicritère.

10.1.3.1 Récapitulation de résultats généraux pour l'analyse de sensibilité

Soit $\mathbf{F} : \mathcal{R}^n \rightarrow \mathcal{R}^n$ une application continue, $\mathbf{g} : \mathcal{R}^n \rightarrow \mathcal{R}^m$ une application différentiable, et $\mathbf{d} : \mathcal{R}^n \rightarrow \mathcal{R}^p$ une application linéaire affine. Définissons

$$\mathbf{K} = \left\{ \mathbf{x} \in \mathcal{R}^n ; \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \mathbf{d}(\mathbf{x}) = \mathbf{0} \right\}. \quad (\text{AS1})$$

Nous cherchons le vecteur \mathbf{x}^* de \mathbf{K} qui résout :

$$" \forall \mathbf{x} \in \mathbf{K}, \mathbf{F}'(\mathbf{x}^*)^t \cdot (\mathbf{x} - \mathbf{x}^*) \geq \mathbf{0} ". \quad (\text{AS2})$$

La formule (AS2) est un problème d'inéquation variationnelle dont \mathbf{x}^* est une solution.

THÉORÈME AS1. Conditions nécessaires sur la solution. Si le vecteur $\mathbf{x}^* \in \mathbf{K}$ est une solution de l'inéquation variationnelle (AS2) et si les gradients $\nabla g_i(\mathbf{x}^*)$ pour les indices i tels que $g_i(\mathbf{x}^*) = 0$ et $\nabla d_i(\mathbf{x}^*)$ pour tous les indices i entre 1 et p sont linéairement indépendants, alors il existe un vecteur $\boldsymbol{\pi} \in \mathcal{R}^m$, un vecteur $\boldsymbol{\mu} \in \mathcal{R}^p$, tels que

$$\mathbf{F}'(\mathbf{x}^*) - \nabla \mathbf{g}(\mathbf{x}^*)^t \cdot \boldsymbol{\pi} - \nabla \mathbf{d}(\mathbf{x}^*)^t \cdot \boldsymbol{\mu} = \mathbf{0} \quad (\text{AS3a})$$

$$\boldsymbol{\pi}^t \cdot \mathbf{g}(\mathbf{x}^*) = 0 \quad (\text{AS3b})$$

$$\boldsymbol{\pi} \geq \mathbf{0}. \quad (\text{AS3c})$$

THÉORÈME AS2. Conditions suffisantes sur la solution. Si les $g_i(\mathbf{x})$ pour i entre 1 et m sont concaves et si $\mathbf{x}^* \in \mathbf{K}$, $\boldsymbol{\pi} \in \mathcal{R}^m$ et $\boldsymbol{\mu} \in \mathcal{R}^p$ vérifient (AS3a, b, c), alors \mathbf{x}^* est une solution de l'inéquation variationnelle (AS2).

THÉORÈME AS3. Conditions suffisantes pour l'unicité locale d'une solution. Sous les hypothèses du Théorème AS2, si de plus \mathbf{F} est différentiable et satisfait

$$\mathbf{y}^t \cdot \nabla \mathbf{F}(\mathbf{x}^*) \cdot \mathbf{y} > 0 \quad (\text{AS4})$$

quel que soit le vecteur $\mathbf{y} \neq \mathbf{0}$ tel que

$$\nabla g_i(\mathbf{x}^*)^t \cdot \mathbf{y} \geq 0 \text{ pour les indices } i \text{ tels que } g_i(\mathbf{x}^*) = 0 \quad (\text{AS5a})$$

$$\nabla g_i(\mathbf{x}^*)^t \cdot \mathbf{y} = 0 \text{ pour les indices } i \text{ tels que } \pi_i > 0 \quad (\text{AS5b})$$

$$\nabla d_i(\mathbf{x}^*)^t \cdot \mathbf{y} = 0 \text{ pour tous les indices } i \text{ entre } 1 \text{ et } p, \quad (\text{AS5c})$$

alors \mathbf{x}^* est une solution localement unique de l'inéquation variationnelle (AS2).

Soit ε un vecteur de paramètres de perturbation de dimension q , et soit $\mathbf{F} : \mathfrak{R}^n \times \mathfrak{R}^q \rightarrow \mathfrak{R}^n$ une application une fois continûment différentiable en $(\mathbf{x}, \varepsilon)$, $\mathbf{g} : \mathfrak{R}^n \times \mathfrak{R}^q \rightarrow \mathfrak{R}^m$ une application concave en \mathbf{x} et deux fois continûment différentiable en $(\mathbf{x}, \varepsilon)$, et $\mathbf{d} : \mathfrak{R}^n \times \mathfrak{R}^q \rightarrow \mathfrak{R}^p$ une fonction linéaire affine en \mathbf{x} et continûment différentiable en ε . Considérons l'inéquation variationnelle perturbée suivante, notée $\text{IV}(\varepsilon)$: "Trouver $\mathbf{x}^* \in \mathbf{K}(\varepsilon)$ tel que

$$\mathbf{F}(\mathbf{x}^*, \varepsilon)^t \cdot (\mathbf{x} - \mathbf{x}^*) \geq 0 \quad (\text{AS6})$$

sur l'ensemble $\mathbf{K}(\varepsilon) = \left\{ \mathbf{x} \in \mathfrak{R}^n ; \mathbf{g}(\mathbf{x}, \varepsilon) \geq \mathbf{0}, \mathbf{d}(\mathbf{x}, \varepsilon) = \mathbf{0} \right\}$ ".

THÉORÈME AS4. Théorème de la fonction implicite. Supposons les hypothèses du Théorème AS3 satisfaites pour $\text{IV}(\mathbf{0})$ (en remplaçant les $\mathbf{F}(\mathbf{x}^*)$, $\mathbf{g}(\mathbf{x}^*)$, $\mathbf{d}(\mathbf{x}^*)$, π et μ par les $\mathbf{F}(\mathbf{x}^*, \mathbf{0})$, $\mathbf{g}(\mathbf{x}^*, \mathbf{0})$, $\mathbf{d}(\mathbf{x}^*, \mathbf{0})$, π^* et μ^* respectivement) : donc les gradients $\nabla g_i(\mathbf{x}^*, \mathbf{0})$ pour les indices i tels que $g_i(\mathbf{x}^*, \mathbf{0}) = 0$ et $\nabla d_i(\mathbf{x}^*, \mathbf{0})$ pour tous les indices i entre 1 et p sont linéairement indépendants. Supposons de plus que la condition de stricte complémentarité

$$\pi_i^* > 0 \text{ quand } g_i(\mathbf{x}^*, \mathbf{0}) = 0$$

soit satisfaite. Alors π^* et μ^* sont uniques. Sur un voisinage de $\varepsilon = \mathbf{0}$, il existe une fonction continûment différentiable $[\mathbf{x}(\varepsilon)^t, \pi(\varepsilon)^t, \mu(\varepsilon)^t]^t$, dans laquelle $\mathbf{x}(\varepsilon)$ est une solution unique de $\text{IV}(\varepsilon)$ et $\pi(\varepsilon)$, $\mu(\varepsilon)$ sont les multiplicateurs associés qui satisfont les hypothèses du Théorème AS3 pour une solution localement unique, et telle que

$$[\mathbf{x}(\mathbf{0})^t, \pi(\mathbf{0})^t, \mu(\mathbf{0})^t]^t = [\mathbf{x}^{*t}, \pi^{*t}, \mu^{*t}]^t.$$

De plus, sur un voisinage de $\varepsilon = \mathbf{0}$, l'ensemble des contraintes inégalités actives reste constant, et les gradients des contraintes actives sont linéairement indépendants au point $\mathbf{x}(\varepsilon)$.

Pour $\varepsilon = \mathbf{0}$ et $(\mathbf{x}, \pi, \mu) = (\mathbf{x}^*, \pi^*, \mu^*)$, en vertu du Théorème AS1

$$\mathbf{F}(\mathbf{x}, \mathbf{0}) - \sum_{i=1}^m \pi_i \nabla g_i(\mathbf{x}, \mathbf{0})^t - \sum_{i=1}^p \mu_i \nabla d_i(\mathbf{x}, \mathbf{0})^t = \mathbf{0} \quad (\text{AS7a})$$

$$\pi_i g_i(\mathbf{x}, \mathbf{0}) = 0 \text{ pour } i \text{ entre } 1 \text{ et } m \quad (\text{AS7b})$$

$$d_i(\mathbf{x}, \mathbf{0}) = 0 \text{ pour } i \text{ entre } 1 \text{ et } p. \quad (\text{AS7c})$$

Désignons les matrices jacobiniennes des membres de gauche des équations (AS7a, b, c) par rapport à ε comme $\nabla_{\varepsilon} \mathbf{J}$ et par rapport à $\mathbf{y} = [\mathbf{x}^t; \pi^t; \mu^t]^t$ comme $\nabla_{\mathbf{y}} \mathbf{J}$.

COROLLAIRE AS1. Dérivation de la solution de l'inéquation variationnelle IV(0) par rapport à ε . Sous les hypothèses du Théorème AS4, $\nabla_{\mathbf{y}} \mathbf{J}$ est inversible, et les dérivées partielles de $\mathbf{y}^* = (\mathbf{x}^*, \pi^*, \mu^*)$ par rapport à ε sont données par

$$\nabla_{\varepsilon} \mathbf{y}^* = \begin{bmatrix} \nabla_{\varepsilon} \mathbf{x} \\ \nabla_{\varepsilon} \pi \\ \nabla_{\varepsilon} \mu \end{bmatrix} = (\nabla_{\mathbf{y}} \mathbf{J}(\mathbf{y}^* | \mathbf{0}))^{-1} (-\nabla_{\varepsilon} \mathbf{J}(\mathbf{y}^* | \mathbf{0})). \tag{AS8}$$

Les conditions techniques nécessaires à l'obtention de (AS8) comprennent notamment l'unicité locale de la solution. Or cette condition est rarement satisfaite en matière d'affectation. Aussi Tobin et Friesz (1988) ont-ils étendu les résultats de Tobin afin de traiter certains cas sans unicité. Décrivons brièvement leurs principes :

- (i) supposer que l'application \mathbf{F} est monotone, donc que les solutions à l'inéquation variationnelle forment un ensemble convexe.
- (ii) choisir un point extrême non dégénéré dans cet ensemble convexe des solutions et, à ce point, déterminer les dérivées de la solution d'un problème restreint qui vérifie les conditions d'unicité. Dans le problème restreint, seuls les chemins chargés d'un flot strictement positif au point extrême sont considérés.
- (iii) grâce à une condition de stricte complémentarité des écarts, la solution du problème restreint perturbé résout également le problème original perturbé, et ne dépend pas du point extrême particulier sur lequel on s'appuie.

10.1.3.2 Adaptation aux modèles d'affectation

Dans le cas de l'affectation (Cf. section 8.1), plusieurs transpositions sont possibles :

- par exemple :

$$\begin{aligned} \mathbf{x} &\leftarrow \mathbf{f} \\ \mathbf{F}(\mathbf{x}) &\leftarrow \mathbf{V}(\mathbf{f}) = \{V_{rs}^k(\mathbf{f})\}_{rsk} \\ \mathbf{g}(\mathbf{x}) &\leftarrow (\mathbf{f}; -\mathbf{z}(\mathbf{f})) \\ \mathbf{d}(\mathbf{x}) &\leftarrow \{\sum_k f_{rs}^k - \bar{q}_{rs}\}_{rs} \text{ si la demande est fixe} \end{aligned}$$

avec ⁽¹⁾ $V_{rs}^k = \left(\sum_a \delta_{rs}^{ak} t_a \right) - D_{rs}^{-1}$
 $+ \bar{p}_{rs}^{\bar{m}_{rs}} \eta_{rs}(1) + \sum_{i=1}^{\bar{m}_{rs}-1} (\bar{p}_{rs}^i - \bar{p}_{rs}^{i+1}) \left\{ \eta_{rs} \left(\frac{q'_{rs,i}}{q_{rs}} \right) + \left(\Gamma_{rs}^{ik} - \frac{q'_{rs,i}}{q_{rs}} \right) \eta'_{rs} \left(\frac{q'_{rs,i}}{q_{rs}} \right) \right\}$

¹on convient ici que l'indicateur Γ entre un chemin k et une classe monétaire i , tous deux relatifs à la relation origine-destination $r-s$, est défini par $\Gamma_{rs}^{ik} = 1$ si $m_{rs}(k) \leq i$ ou $\Gamma_{rs}^{ik} = 0$ si $m_{rs}(k) > i$, Cf. section 8.2.4

de manière que les variables duales \mathbf{w} associées aux contraintes adjointes apparaissent au sein du vecteur $\boldsymbol{\pi}$,

• ou encore :

$$\begin{array}{l} \mathbf{x} \leftarrow (\mathbf{f}; \mathbf{w}) \\ \mathbf{F}(\mathbf{x}) \leftarrow (\mathbf{I}(\mathbf{f}; \mathbf{w}); -\mathbf{z}(\mathbf{f})) \\ \mathbf{g}(\mathbf{x}) \leftarrow (\mathbf{f}; \mathbf{w}) \end{array}$$

pour traiter les variables primales \mathbf{f} et les variables duales \mathbf{w} sur un même pied.

Notons ici que les possibilités théoriques pour propager l'erreur n'ont pas été mises à profit, hormis dans les illustrations numériques élémentaires des contributions méthodologiques (vecteur \mathbf{x} de dimension 3 dans Tobin et Friesz 1988). L'explication tient à la nécessité de traiter explicitement les chemins, chose exclue dans les algorithmes traditionnels mais permise par les méthodes comme l'égalisation par transvasement développée au chapitre 9.

10.2 ORDRES DE GRANDEUR POUR L'INCERTITUDE EXOGÈNE

Nous avons préconisé de considérer l'erreur exogène a priori relative à chacun des différents objets d'estimation, pour propager le tout via le modèle de simulation et livrer l'erreur a posteriori attendue. Avant de propager l'erreur exogène, il faut la quantifier : chose plus facile dans la simulation d'une situation de référence pour laquelle on dispose d'observations, que dans la simulation d'une situation virtuelle où intervient une erreur de supputation.

On constate une quasi-absence de quantifications de l'erreur exogène a priori. Pour démontrer la possibilité d'une telle quantification, et fournir des ordres de grandeur de référence, examinons donc des observations relatives respectivement aux fonctions de temps de parcours (§10.2.1), aux flots origine-destination (§10.2.2) et aux valeurs du temps (§10.2.3). Pour ces trois objets il existe des modèles d'estimation relativement simples ; concernant la valeur du temps, nous apportons une contribution originale, à savoir une méthode pour estimer une distribution continue grâce à des données individuelles.

10.2.1 Fonctions de temps de parcours

10.2.1.1 Principe

Compte tenu de l'hypothèse de stationnarité dans une affectation statique, nous considérons une dépendance du temps de parcours y envers le flot local x hors effets de forte saturation. Supposons que la dépendance est de forme $y = f_{\alpha, \beta}(x)$ où α et β sont deux paramètres à estimer. Nous voulons déterminer, sur une route particulière, la part de la variabilité des temps de parcours qui est expliquée par les variations du flot : il s'agit de décomposer la variabilité totale en variabilité expliquée et variabilité résiduelle.

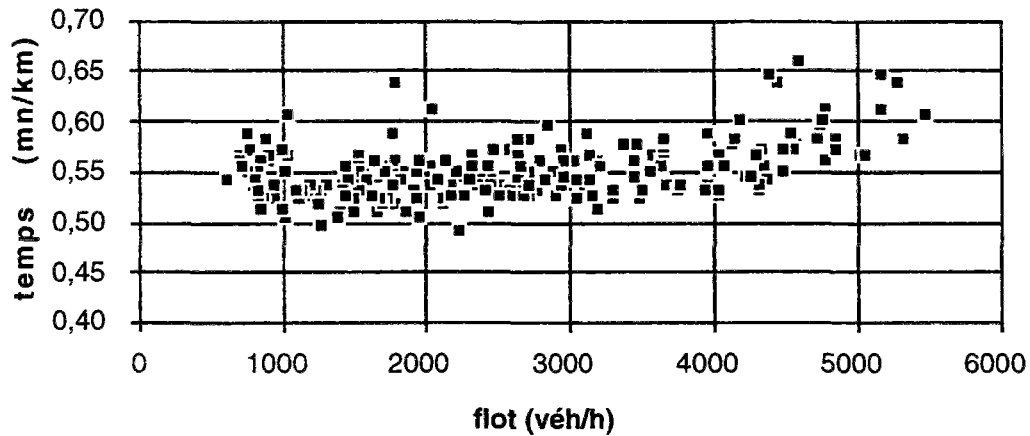
Soient des observations $(x_i; y_i)$ indicées par i et en nombre n ; notons $\bar{y} = \frac{1}{n} \sum_i y_i$ et $\hat{y}_i = f_{\alpha, \beta}(x_i)$. Alors la variabilité totale vaut $VT = \sum_i (y_i - \bar{y})^2$, la variabilité expliquée $VE = \sum_i (\hat{y}_i - \bar{y})^2$ et la variabilité résiduelle $VR = \sum_i (y_i - \hat{y}_i)^2$.

L'objectif¹ de l'estimation est de minimiser la variabilité résiduelle en fonction des paramètres α et β , et, au point $(\hat{\alpha}; \hat{\beta})$ de minimisation, de quantifier l'incertitude attachée à une prédiction $y = f_{\hat{\alpha}, \hat{\beta}}(x)$.

10.2.1.2 Les observations utilisées

On dispose de couples (débit ; vitesse instantanée) mesurés durant toute une journée de 1990 sur un sens de l'autoroute A7 entre Lyon et Valence, avec une circulation sur trois voies. La base temporelle des mesures est de six minutes (donc $n = 240$). On transforme les vitesses instantanées v_i en temps de parcours kilométrique instantané $y_i = 1/v_i$.

Fig. 10a. Le jeu d'observations.



Source : ASF (1990).

10.2.1.3 Estimer une fonction affine par une régression linéaire simple

Donnons une spécification affine à la fonction de temps de parcours : $y = \alpha + \beta x$. Sous les hypothèses du modèle de régression linéaire standard (Cf. section 3.1), on obtient des estimations² $\hat{\beta} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$ pour β et $\hat{\alpha} = \bar{y} - \bar{x}\hat{\beta}$ pour α , en notant

$$\bar{x} = \frac{1}{n} \sum_i x_i. \text{ A.N. : } \hat{\beta} = 1.04 \cdot 10^{-5} \frac{\text{mn}}{\text{km.pcu}} \text{ et } \hat{\alpha} = 0.52 \text{ mn/km.}$$

Considérant que les y_i sont des variables aléatoires en tant que tirages (réalisations) de y , les estimateurs $\hat{\alpha}$ et $\hat{\beta}$ sont eux-mêmes des variables aléatoires. Sous les hypothèses

¹L'objectif ultime de l'estimation est de déterminer les vraies valeurs des paramètres. Sous certaines hypothèses probabilistes, *en moyenne* les vraies valeurs minimisent la variabilité résiduelle

²au sens des moindres carrés ordinaires

classiques relatives aux y_i , $\hat{\beta}$ suit une distribution approximativement normale de moyenne β et d'écart-type $SE_{\hat{\beta}} = \sigma / \sqrt{\sum_i (x_i - \bar{x})^2}$, où σ est l'écart-type de l'aléa résiduel $y - \hat{y} = y - f_{\alpha, \beta}(x)$. On estime σ^2 par la variance résiduelle $s^2 = \left(\sum_i (\hat{y}_i - y_i)^2 \right) / (n - 2)$. A.N. : $s = 2.4310^{-2} \frac{\text{mn}}{\text{km}}$ et $SE_{\hat{\beta}} = 1.2210^{-6} \frac{\text{mn}}{\text{km.pcu}}$.

Alors l'intervalle de confiance au seuil $1-z$ de la pente β est $\beta = \hat{\beta} \pm t_{z/2}^{(n-2)} SE_{\hat{\beta}}$, avec $t^{(n-2)}$ variable de Student à $n-2$ degrés de liberté. A.N. : au seuil de confiance $1-z = 95\%$, $\beta = \hat{\beta} \pm t_{0.025}^{(238)} SE_{\hat{\beta}} = 1.04 \cdot 10^{-5} \pm 2.32 \cdot 10^{-6} \frac{\text{mn}}{\text{km.pcu}}$.

Autrement dit : l'erreur exogène a priori relative au paramètre β suit, au facteur $1/SE_{\hat{\beta}}$ près, une loi de Student à $n-2$ degrés de liberté. Quand le nombre d'observations n est grand, la distribution de l'incertitude $\hat{\beta} - \beta$ est approximativement normale centrée d'écart-type $SE_{\hat{\beta}}$. L'erreur exogène a priori relative à α s'en déduit, via $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$.

En un point x_0 , l'intervalle de confiance au niveau $1-z$ pour la valeur moyenne de y (1)

est $\hat{y}^{+/-} = \hat{\alpha} + \hat{\beta}x_0 \pm t_{z/2}^{(n-2)} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$ et prend en compte uniquement

l'incertitude liée à l'estimation des paramètres. Tandis que l'intervalle de confiance à $1-z$ pour une valeur particulière de y en x_0 incorpore de plus l'incertitude résiduelle (2) et

vaut $\hat{y}^{+/-}(x_0) = \hat{\alpha} + \hat{\beta}x_0 \pm t_{z/2}^{(n-2)} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} + 1}$ (fig. 10b pour l'exemple traité).

On mesure la crédibilité du modèle en comparant la variabilité expliquée à la variabilité résiduelle. Dans le cas de la régression linéaire, de façon équivalente, on évalue la probabilité critique de l'hypothèse "pas de pente", donc pas de lien entre y et x , c'est-à-dire encore ($\beta = 0$), en évaluant la probabilité critique p.c. qui est la probabilité d'observer $\{t^{(n-2)} = \hat{\beta} / SE_{\hat{\beta}}\}$ sous l'hypothèse que $\beta = 0$. On mesure la p.c. comme l'aire de la queue d'une distribution $t^{(n-2)}$ au delà de la valeur $t = \hat{\beta} / SE_{\hat{\beta}}$.

Pour notre jeu d'observations, on observe que $t = 8.49$ donc une probabilité critique inférieure à 1‰. Au niveau de confiance 99.9% on peut rejeter l'hypothèse nulle "pas de pente" et conclure que le temps de parcours augmente effectivement quand le trafic local croît.

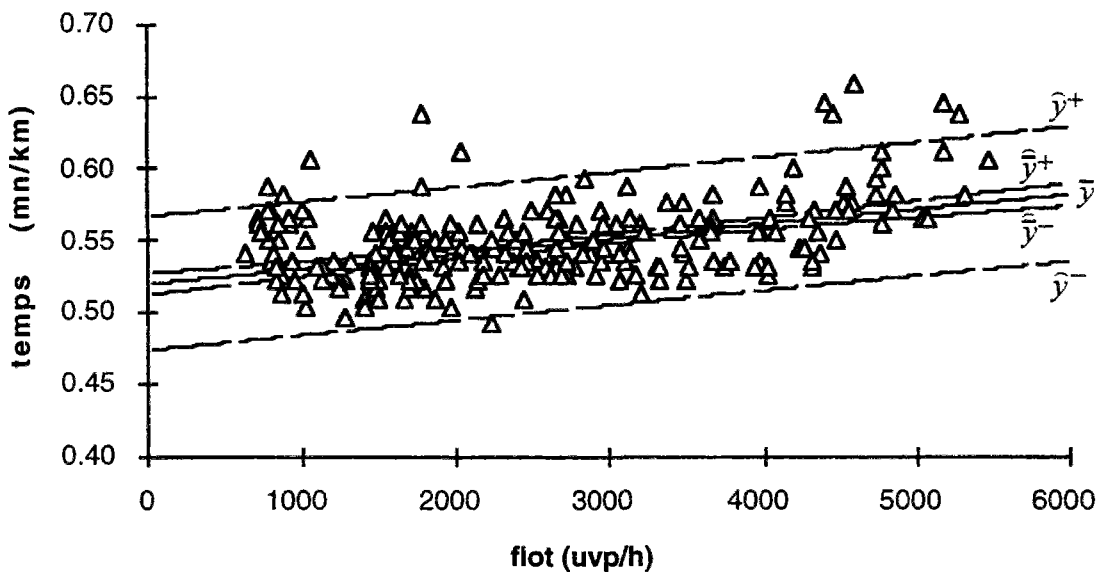
Considérer une fonction de temps de parcours avec une pente $\hat{\beta}$, plutôt qu'une pente nulle donc un temps de parcours constant, permet de réduire la variabilité inexpliquée de la variabilité totale VT à la variabilité résiduelle VR. Dans une affectation à temps fixes,

¹c'est une variable aléatoire via $\hat{\alpha}$ et $\hat{\beta}$

²quand n est grand, l'incertitude résiduelle domine largement.

la variabilité inexpliquée attachée au temps de parcours est VT, tandis que dans une affectation à temps variables elle se réduit à VR. Dans notre exemple, à temps fixes on a au seuil de confiance $1-z = 95\%$ l'intervalle $y = \bar{y} \pm t_{z/2}^{(n-1)} \sigma_{tf} = 0,549 \pm 0,052$ mn/km car $(n-1)\sigma_{tf}^2 = VT$, tandis qu'à temps variables¹ on a $y = \hat{\alpha} + \hat{\beta}x \pm t_{z/2}^{(n-2)} \sigma_{tv} = 0,549 \pm 0,046$ mn/km car $(n-2)\sigma_{tv}^2 = VR$. A temps fixes, $\frac{\sigma_y}{y} \approx 5.0\%$ tandis qu'à temps variables $\frac{\sigma_y}{y} \approx 4.4\%$.

Fig. 10b. Bandes de confiance.



L'intervalle de confiance à 95% pour la valeur *moyenne* de la prédiction est donné, en chaque valeur x du flot, par $[\hat{y}^-(x); \hat{y}^+(x)]$. L'intervalle de confiance à 95% pour la prédiction au point x est donné par $[\bar{y}^-(x); \bar{y}^+(x)]$.

10.2.1.4 Bilan

Nous avons appliqué la méthode statistique classique d'estimation selon les moindres carrés ordinaires à des mesures conjointes de débit et de temps de parcours. Cela nous a permis de calculer l'incertitude exogène relative aux temps de parcours dans deux modèles, avec ou sans dépendance du temps envers le débit.

Nous avons constaté que l'incorporation de la dépendance du temps envers le débit permet de réduire l'incertitude exogène, de façon assez faible mais significative dans l'exemple considéré. Signalons que ces données se rapportent à un site exempt de fortes congestions comme d'interactions avec d'autres arcs, et à une journée sans incident. Tout écart à ces conditions "idéales" nécessiterait de modifier le modèle d'estimation, par exemple en spécifiant une fonction non linéaire plutôt qu'une fonction affine linéaire.

¹en négligeant l'incertitude liée à l'estimation des paramètres

10.2.2 Flots origine-destination

10.2.2.1 Variabilité autour de la situation moyenne

Dans l'affectation, on considère une période de temps durant laquelle les conditions de circulation sont homogènes, stationnaires. Cette période de référence correspond à une situation moyenne, par exemple le jour moyen de l'année en milieu interurbain. Il y a probablement des écarts, des variations autour de cette situation moyenne : dans le cas des flots O-D, leur variabilité se répercute sur les flots des arcs, puis par leur entremise sur les temps de parcours, enfin sur les choix d'itinéraire.

10.2.2.2 Les observations utilisées

Afin d'évaluer l'ordre de grandeur des phénomènes de variabilité des flots O-D, nous avons analysé les flots O-D journaliers de l'année 1994 entre dix postes d'entrée/sortie (gares de péage) de l'autoroute A10 entre Paris et Tours. Ces données nous ont été gracieusement communiquées par la société COFIROUTE.

10.2.2.3 Analyse des données

Nous avons calculé pour chaque relation O-D i le flot journalier moyen \bar{q}_i et son écart-type σ_{q_i} . Entre chaque couple de relations O-D, $i = r-s$ et $j = r'-s'$, nous avons évalué la covariance $\text{Cov}(q_i; q_j)$ des flots journaliers q_i et q_j . Le tableau 10A expose les résultats relatifs à trois relations O-D qui présentent des flots moyens contrastés.

Tab. 10A. Moments de la (vraie) distribution empirique.

Relation O-D i	moyenne \bar{q}_i (voitures / jour)	écart-type σ_{q_i} (voitures / jour)	$\frac{\text{Cov}(q_i; q_j)}{\bar{q}_i \cdot \sum_j \bar{q}_j}$
Paris-Tours	5 727	4 163	0.231
Orléans-Tours	1 311	763	0.202
Arthenay-Orléans	123	47	0.011

Source : COFIROUTE (1995).

Le tableau 10A appelle deux commentaires :

i) la dispersion relative du flot journalier, i.e. le rapport entre son écart-type et sa moyenne, ne décroît pas en fonction du flot moyen, à l'encontre des hypothèses conventionnelles (Wyley et al. 1981). Est-ce parce que celles-ci se rapportent à des échantillons partiels plutôt qu'à l'ensemble de la population ? Réponse au prochain paragraphe !

ii) la dépendance entre O-D différentes, mesurée de façon agrégée par la corrélation entre une O-D et l'ensemble de toutes les O-D, paraît plus importante pour les O-D avec un flot fort. Cela tient peut-être aux pointes saisonnières qui affectent essentiellement le trafic à longue distance, entre les grandes villes (Paris, Orléans, Tours) qui échangent ou font transiter des flots importants.

10.2.2.4 Bilan et transposabilité des résultats

Dans la plupart des cas pratiques, la connaissance exhaustive des flots O-D est

inaccessible. Les enquêtes aux itinéraires interurbains procèdent couramment, en France, par un échantillonnage à deux niveaux : l'un porte sur les jours de l'année (souvent entre 1 et 3 jours d'enquête), l'autre sur les véhicules (taux de sondage de l'ordre de 20%).

Connaissant la vraie distribution pour notre jeu de données, nous pouvons calculer l'incertitude exogène moyenne résultant d'une telle procédure d'enquête. Supposons que l'enquête porte sur N jours et que chaque jour k l'on mesure un flot $q_{i(k)}$ pour chaque relation O-D i . Alors la moyenne d'échantillon est $\hat{q}_i = \{\sum_k q_{i(k)}\} / N$, égale en moyenne à \bar{q}_i , tandis que l'écart-type d'échantillon est $\{\frac{1}{N-1} \sum_k (q_{i(k)} - \hat{q}_i)^2\}^{1/2}$, égal en moyenne à $SE_{q_i}(N) = \sigma_{q_i} / \sqrt{N}$: la précision croît comme la racine carrée du nombre de jours d'enquête. Le tableau 10B présente les écarts-types moyens pour différentes durées d'enquête N .

Les résultats du tableau 10B fournissent une base de comparaison pour une enquête aux itinéraires routiers en milieu interurbain. Ils *infirment* l'hypothèse conventionnelle selon laquelle la dispersion relative décroît en fonction du flot moyen. Signalons toutefois que la matrice O-D considérée est très particulière, en ce qu'elle correspond à des véhicules dont le choix d'itinéraire est déjà effectué puisqu'ils circulent sur l'autoroute plutôt que sur les routes parallèles gratuites.

Tab. 10B. Moments de quelques distributions d'échantillons.

Relation O-D i	$\frac{SE_{q_i}(N=1)}{\bar{q}_i} = \frac{\sigma_{q_i}}{\bar{q}_i}$	$\frac{SE_{q_i}(N=5)}{\bar{q}_i} = \frac{\sigma_{q_i}}{\sqrt{5} \bar{q}_i}$	$\frac{SE_{q_i}(N=20)}{\bar{q}_i} = \frac{\sigma_{q_i}}{\sqrt{20} \bar{q}_i}$
Paris-Tours	0.73	0.33	0.17
Orléans-Tours	0.58	0.26	0.13
Arthenay-Orléans	0.38	0.17	0.09

Source : COFIROUTE (1995).

10.2.3 Valeur du temps

La simulation du choix d'itinéraire repose généralement sur un principe explicatif simple : chaque auteur de déplacement est affecté à l'itinéraire de temps généralisé (y compris les frais de circulation, les inconforts et insécurités) minimum. En complétant ce principe par l'expression du temps généralisé ressenti le long d'un chemin, on obtient une règle de répartition du trafic entre itinéraires concurrents, autrement dit un modèle d'affectation aux itinéraires. Dans un tel modèle, la valeur du temps et le malus d'inconfort sont des *paramètres du comportement* dont les valeurs numériques déterminent les proportions du trafic affectées aux itinéraires concurrents. On peut estimer leurs valeurs numériques à l'aide d'observations de répartition du trafic entre itinéraires concurrents, en estimant un *modèle de choix discret*.

Nous avons déjà rappelé comment traiter un choix entre itinéraires concurrents comme un modèle de choix discret (Cf. §10.1.2.1). Si un tel traitement est chose fréquente pour le choix modal, il n'en va pas de même pour l'affectation : seule la répartition entre itinéraires selon une règle logit linéaire a été ainsi étudiée (Robillard 1974, Fisk 1977, Ben-Akiva et al. 1984). Pour d'autres répartitions comme le logit logarithmique ou le prix-temps, il y avait un manque théorique à combler afin notamment d'exploiter

efficacement les enquêtes aux origines-destinations¹.

Nous développons donc une telle méthode² dans ce qui suit. Après avoir donné des notations pour formaliser un choix entre itinéraires comme un modèle de choix discret (§.1), nous rappellerons les procédures disponibles pour mesurer les attributs des variantes et des auteurs des déplacements (§.2). Ensuite nous présenterons notre méthode d'estimation selon le maximum de vraisemblance (§.3). Nous l'adapterons au cas particulier du modèle prix-temps (§.4) puis nous traiterons deux jeux d'observations (§.5) : une enquête de préférences déclarées (source USAP 1992) et une enquête aux itinéraires classique (source CETE de l'Ouest 1993). Nous terminerons par une discussion des hypothèses et des résultats (§.6).

10.2.3.1 Notations

Pour représenter un choix discret entre itinéraires, soit :

- r la zone d'origine du déplacement.
- s la zone de destination du déplacement.
- k un chemin, de longueur L_k , parcouru en un temps T_k , pour lequel l'auteur du déplacement dépense le prix M_k , et subit l'inconfort I_k ... De façon générique, soit a un indice d'attribut de coût généralisé, et $X_{a,k}$ l'attribut d'indice a dans le coût généralisé attaché au chemin k .
- i un auteur de déplacement entre r et s , de valeur du temps v_i , de valeur du confort c_i , de coefficient d'équivalence entre distance et numéraire d_i . De façon générique, nous notons $r_{i,a}$ le coefficient de l'attribut d'indice a pour l'individu i .
- $\xi_i(k)$ un terme aléatoire attaché au chemin k par l'individu i (en toute rigueur, il faudrait noter $\xi_i(k)(\omega)$, avec ω un aléa).

Alors le coût généralisé déterministe subi sur le chemin k par l'individu i vaut :

$$G_i(k) = M_k + v_i T_k + d_i L_k + c_i I_k = \sum_a r_{i,a} \cdot X_{a,k}$$

Formulons l'utilité aléatoire attachée au chemin k par l'individu i comme :

$$U_i(k) = A_k^i - G_i(k) + \xi_i(k),$$

avec A_k^i une "constante" modale fonction seulement du chemin et de l'individu. La spécification des $\xi_i(k)(\omega)$ détermine la règle de répartition du trafic entre itinéraires concurrents (Cf. §10.1.2.1).

Les paramètres du comportement sont ceux des distributions des coefficients d'équivalence $r_{i,a}$ et des termes aléatoires $\xi_i(k)$, ainsi que les "constantes" A_k^i . Soit $\Theta = \{\theta_m\}_m$ le vecteur des paramètres à estimer.

¹qui sont courantes en France pour le mode routier en milieu interurbain

²élaborée à l'occasion d'une convention pour le compte du SETRA (Leurent 1995e)

On suppose que l'individu i choisit le chemin k qui maximise $U_i(k)$. Pour cet individu, la fréquence d'emprunt du chemin k est modélisée comme la probabilité que l'utilité aléatoire qu'il lui attache dépasse celles des autres itinéraires k' , soit

$$\pi_{ik}(\Theta) = \Pr(\{U_i(k) \geq \max_{k'} U_i(k')\}). \quad (10.0)$$

Pour marquer le résultat des enquêtes aux itinéraires, notons :

- j un groupe (un sous-ensemble d'individus i , défini par des valeurs particulières des attributs, par exemple le motif, ou la valeur du temps).
- f_k le trafic mesuré sur le chemin k .
- f_{jk} le trafic du groupe j mesuré sur le chemin k (donc $f_k = \sum_j f_{jk}$).
- q_j l'effectif du groupe j , donc $q = \sum_j q_j$ est l'effectif total des individus enquêtés¹.

La probabilité théorique d'emploi du chemin k par les membres du groupe j est la moyenne des probabilités individuelles π_{ik} des individus i qui composent ce groupe :

$$\pi_{jk}(\Theta) = \left(\sum_{i \in j} \pi_{ik}(\Theta) \right) / \left(\sum_{i \in j} 1 \right).$$

Les méthodes d'estimation ont pour principe commun de rapprocher les trafics simulés $q_j \pi_{jk}$ des trafics mesurés f_{jk} afin de fixer la valeur de Θ . Pour cela il faut connaître les q_j et les f_{jk} , connaître "tout ce qui n'est pas Θ " dans $\pi_{jk}(\Theta)$, et disposer d'un critère de proximité entre les $q_j \pi_{jk}$ et les f_{jk} .

10.2.3.2 De quelles mesures dispose-t-on ?

Examinons d'abord comment connaître les q_j , les f_{jk} et les paramètres autres que Θ dans les $\pi_{jk}(\Theta)$. En pratique on sépare la tâche en deux lots : d'abord la mesure des caractères intrinsèques des chemins (les $X_{\alpha,k}$), puis l'observation des caractères individuels des individus (choix, critères de regroupement, indices sur les $r_{i,\alpha}$).

Attributs des auteurs des déplacements : enquêtés par interview. De façon courante, une enquête aux itinéraires d'une relation origine-destination procède par interception lors d'un déplacement. Chaque interviewé répond à quelques questions : il indique son motif de déplacement (au départ et à l'arrivée), ses lieux d'origine et destination, plus éventuellement d'autres renseignements qui servent à regrouper les individus : liés au ménage (composition, motorisation, position dans le cycle de vie), à l'individu (âge, sexe, Profession et Catégorie Sociale) ⁽²⁾ et au déplacement lui-même (motif, horaire, lieux de départ et d'arrivée). L'enquêteur consigne également le site d'interception, l'heure de passage et le type de véhicule. Comme l'on observe un seul choix par individu, il n'est pas possible d'estimer les attributs qui lui sont propres : d'où

¹dont la réponse est utilisable

²en sus l'enquêteur peut faire déclarer les attributs du chemin suivi : temps de parcours, prix... mais l'expérience montre que ces évaluations sont bien trop subjectives, avec jusqu'à 200% ou 300% d'erreur !

la nécessité de faire des classes. Le regroupement le plus grossier, prendre la population dans son ensemble, est le seul possible si l'observation porte uniquement sur la relation origine-destination et le chemin suivi.

Une telle enquête permet de révéler les préférences des auteurs des choix, en "expliquant" des choix observés. D'autres techniques d'enquête sont maintenant répandues, basées sur le principe d'une "déclaration des préférences" ⁽¹⁾. Elles reposent sur une mise en situation : l'enquêteur présente au répondant plusieurs scénarios, un scénario désignant une situation de choix entre plusieurs itinéraires décrits de façon abstraite (avec les attributs $X_{a,k}$). Le répondant annonce un choix par scénario : sa déclaration est assimilée à une observation. En variant les scénarios, on obtient une information riche, qui peut suffire pour estimer les attributs de l'individu², mais que l'on emploie le plus souvent avec un regroupement en classes.

Ainsi l'USAP (1992) a réalisé des interviews d'automobilistes, leur demandant de choisir un itinéraire parmi plusieurs différenciés en termes de temps T_k et de prix P_k . Le tableau 10C décrit les options proposées (pour des déplacements à distance d'environ 150 km) et indique pour chacune d'elles l'effectif f_k des répondants qui l'ont choisie.

Tab. 10C. Concurrence entre itinéraires, déplacements à environ 150 km.

k	T_k (h)	P_k (F)	G_k (h) pour une valeur du temps de 60 F/h	f_k , motifs privés	f_k , motifs professionnels
1	3.0	100	4.67	135	13
2	2.5	120	4.50	60	22
3	2.0	150	4.50	135	32
4	1.75	170	4.58	70	24
5	1.25	220	4.92	65	56

Source : USAP (1992), pour le cas de l'heure creuse.

Attributs des chemins : le plus souvent un codage. Concernant les caractères des chemins, la technique usuelle consiste à décrire un itinéraire comme l'enchaînement de routes élémentaires (arcs) sur le réseau modal, puis à coder pour chaque arc des attributs de prix, temps, distance... dont on fait la somme au niveau de l'itinéraire. L'opération paraît naturelle pour des grandeurs additives comme le prix, le temps, la distance ; moins pour l'inconfort dont la quantification est délicate ⁽³⁾. Mentionnons une éventuelle complication, lorsque les attributs $X_{a,k}$ ne sont pas des constantes, mais varient au cours de la période modélisée. Il faut tenir compte de cette variabilité si elle influe sur les

¹enquêtes de Préférences Déclarées, ou en anglais Stated Preferences. Cf. Ortuzar et Willumsen 1990, Ben-Akiva et Lerman 1985, le numéro spécial XXII/1 du Journal of Transport Economics Policy en 1988

²Cf. Judge et al. 1988, Daito et al. 1992, Wynter 1994. La richesse de l'information collectée est relativisée par la virtualité des situations de choix et la subjectivité des perceptions.

³pour un réseau de transport collectif, l'inconfort peut être quantifié comme la probabilité de ne pas disposer d'une place assise, par exemple. Pour le mode routier, il peut être relié entre autres à la fiabilité du temps de parcours, ou à une fluidité de trajet appréhendée par le nombre d'arrêts en cours de route, au jalonnement, à l'agrément...

choix modélisés ⁽¹⁾.

Le codage des attributs des chemins s'appuie sur des informations relevées sur le terrain : distance, péage, temps de parcours mesuré par des campagnes de véhicules (flottants ou suiveurs) ou par photographie aérienne ou par détection automatique (Cf. Taylor et Young 1988, Cohen 1990). Nous supposons dans ce qui suit que les attributs des arcs sont mesurés de façon suffisamment précise pour qu'il n'y ait pas à distinguer les valeurs codées des valeurs réelles ⁽²⁾.

10.2.3.3 Une méthode du maximum de vraisemblance qui tire parti des observations individuelles

Pour opérer l'estimation, on connaît deux méthodes traditionnelles : minimiser une somme pondérée d'écarts³ entre quantités observées et quantités modélisées, ou maximiser la vraisemblance des observations. Cette seconde méthode est l'outil statistique privilégié des modèles de choix discret. Il s'agit de trouver l'estimateur qui maximise la vraisemblance des observations. Pour une valeur donnée des paramètres, on définit la vraisemblance d'une observation comme la probabilité d'occurrence de la valeur observée. La vraisemblance d'un ensemble d'observations indépendantes est égale au produit des vraisemblances individuelles des différentes observations.

L'observation des chemins $k(i)$ respectivement choisis par les T individus i de la population enquêtée, a pour vraisemblance

$$V(\{k(i)\}_i | \Theta) = \prod_{i=1}^T \pi_{k(i)}(\Theta). \quad (10.1)$$

Cette expression montre le caractère individuel de la méthode, puisque chaque individu i contribue au produit d'ensemble par le terme $\pi_{k(i)}(\Theta)$. Si la probabilité $\pi_{k(i)}(\Theta)$ dépend de i seulement au travers du chemin suivi $k(i)$ et des caractères du groupe j auquel appartient i , alors (10.1) se réduit à

$$V(\{k(i)\}_i | \Theta) = V(\{f_{kj}\}_{j,k} | \Theta) = \prod_{j,k} (\pi_{kj}(\Theta))^{f_{kj}}. \quad (10.2)$$

En particulier, si chaque groupe j correspond à une relation origine-destination $r-s$, la vraisemblance de l'ensemble des observations s'exprime

$$V(\{f_{rs,k}\}_{rs,k} | \Theta) = \prod_{rs} \prod_{k \in rs} \{\pi_{rs,k}(\Theta)\}^{f_{rs,k}}, \quad (10.3a)$$

donc la log-vraisemblance vaut

$$L(\Theta) = \log V(\{f_{rs,k}\}_{rs,k} | \Theta) = \sum_{rs} \sum_{k \in rs} f_{rs,k} \log(\pi_{rs,k}(\Theta)). \quad (10.3b)$$

¹notamment en ce qui concerne les temps de parcours ; à titre d'attribut de fiabilité on peut retenir la dispersion de la distribution statistique

²cela n'a rien d'évident a priori...

³généralement des écarts élevés au carré, auquel cas la méthode s'appelle naturellement méthode des moindres carrés. Dans le cas d'une enquête aux itinéraires d'une relation O-D, un écart prendrait typiquement la forme $q_j \pi_{jk}(\Theta) - f_{jk}$

L'estimateur du maximum de vraisemblance (EMV) $\hat{\Theta}$ est la valeur qui maximise la vraisemblance de l'observation, ou de façon équivalente la log-vraisemblance. Pour le calculer, on exprime des conditions du premier ordre sur l'annulation des dérivées partielles au premier ordre du critère $L(\Theta)$. Les dérivées partielles au premier et au second ordre de $L(\Theta)$ découlent des dérivées partielles au premier et au second ordre des probabilités simulées $\pi_{kj}(\Theta)$:

$$\frac{\partial L}{\partial \theta_m} = \sum_{rs} \sum_{k \in rs} f_{rs,k} \frac{\partial \pi_{rs,k}}{\partial \theta_m} \quad (10.4a)$$

$$\frac{\partial^2 L}{\partial \theta_m \partial \theta_n} = \sum_{rs} \sum_{k \in rs} f_{rs,k} \left\{ \frac{\partial^2 \pi_{rs,k}}{\partial \theta_m \partial \theta_n} - \frac{1}{\pi_{rs,k}} \frac{\partial \pi_{rs,k}}{\partial \theta_m} \frac{\partial \pi_{rs,k}}{\partial \theta_n} \right\}. \quad (10.4b)$$

Les dérivées partielles au second ordre $\frac{\partial^2}{\partial \theta_m \partial \theta_n} L(\Theta)$ de la fonction de log-vraisemblance, par rapport aux paramètres du comportement $\Theta = \{\theta_m\}_m$, s'emploient dans la matrice d'information $I(\Theta) = -E \left[\frac{\partial^2}{\partial \theta_m \partial \theta_n} L(\Theta) \right]$. Après inversion, la matrice d'information livre une estimation de la variance des estimateurs : $\sqrt{T}(\hat{\Theta} - \Theta) \xrightarrow{d} N \left(0; \lim_{T \rightarrow \infty} \left[\frac{1}{T} I(\Theta) \right]^{-1} \right)$. Cela sert pour mesurer la qualité de l'estimation statistique (Amemiya, 1981 ; Judge et al. 1988).

Insistons sur la forte dépendance de toutes ces expressions vis-à-vis du critère de regroupement. Même si Θ paraît la seule quantité à estimer en maximisant $L(\Theta)$, la forme de L repose essentiellement sur des hypothèses relatives au regroupement des individus, en l'occurrence ici la relation O-D et le chemin suivi. De façon analogue, on pourrait encore segmenter selon les différents motifs de déplacement.

Résolution numérique. Pour calculer l'EMV Θ , autrement dit pour maximiser numériquement la fonction de log-vraisemblance, on recommande généralement la méthode de Newton ou la méthode de score qui emploient les dérivées au premier et au second ordre de la fonction objectif afin d'annuler les dérivées au premier ordre (Amemiya, 1981). En omettant les indices $r-s$, si l'on se donne les formules des π_k en fonction de Θ , en un point Θ_0 donné on peut évaluer $\pi_k(\Theta_0)$ et les dérivées partielles $\partial \pi_k / \partial \theta_m$ et $\partial^2 \pi_k / \partial \theta_m \partial \theta_n$. Les valeurs de $L(\Theta_0)$ et des dérivées $\partial L / \partial \theta_m$ et $\partial^2 L / \partial \theta_m \partial \theta_n$ s'en déduisent grâce aux équations (10.3) et (10.4).

10.2.3.4 Cas du modèle bicritère

Jusqu'ici nous avons suivi une méthode standard pour formuler un modèle de choix discret, sans spécifier complètement la règle de répartition du trafic entre itinéraires concurrents donc l'expression des $\pi_{rs,k}(\Theta)$. Nous apportons maintenant des éléments originaux, à savoir un traitement d'une distribution continue des valeurs du temps conformément à l'hypothèse bicritère.

Supposons donc que la valeur du temps v suit une distribution continue, avec pour

fonction de répartition $H_{\mu,\sigma}(x) = \Pr\{v \leq x\}$ et pour densité $h_{\mu,\sigma}(x) = \frac{\partial}{\partial x} H_{\mu,\sigma}$. Ces deux fonctions dépendent non seulement de l'input x (une valeur du temps particulière) mais encore de deux paramètres μ et σ à fixer (1).

Le vecteur de paramètres $\Theta = \{\theta_m\}_m$ se compose des paramètres de la distribution des valeurs du temps, notés μ et σ , ainsi que de coefficients r_a pour la pondération des attributs $X_{a,k}$ rattachés à la partie monétaire du coût généralisé.

Nous formulons donc le coût généralisé d'un chemin k pour un déplacement i de valeur du temps v_i selon $G_i(k) = v_i T_k + (\sum_a r_a \cdot X_{a,k})$ où apparaît la décomposition entre partie temporelle et partie monétaire. Selon l'hypothèse bicritère déterministe, l'utilité aléatoire se réduit à l'opposé du coût généralisé, soit $U_i(k) = -v_i T_k - (\sum_a r_a \cdot X_{a,k})$. Alors, sur une relation O-D r - s donnée, la proportion de choisir la classe monétaire efficace k vaut (Cf. section 8.3) :

$$\pi_k(\Theta) = \int_{I_k}^{S_k} h_{\mu,\sigma}(v) dv = H_{\mu,\sigma}(S_k) - H_{\mu,\sigma}(I_k), \tag{10.5}$$

en notant I_k la valeur du temps de coupure avec la classe efficace de prix immédiatement inférieur, et S_k celle avec la classe efficace de prix immédiatement supérieur. Pour simplifier nous confondons ici entre chemin et classe monétaire. Une classe monétaire k non efficace a une probabilité nulle d'être choisie.

Par construction, les probabilités $\pi_k(\mu; \sigma; \{r_a\}_a)$ dépendent de μ et σ uniquement par l'intermédiaire de la fonction $H_{\mu,\sigma}$, et des coefficients r_a uniquement via les S_k et I_k qui sont des valeurs de coupure offertes égales respectivement à $S_k = (P_{k+1} - P_k) / (T_k - T_{k+1})$ et $I_k = (P_k - P_{k-1}) / (T_{k-1} - T_k)$.

Par linéarité de la dérivation, on vérifie que

$$\frac{\partial}{\partial \mu} \pi_k = \frac{\partial}{\partial \mu} H_{\mu,\sigma}(S_k) - \frac{\partial}{\partial \mu} H_{\mu,\sigma}(I_k),$$

et de façon générale pour tout opérateur de dérivation D dans $\{\frac{\partial}{\partial \mu}; \frac{\partial}{\partial \sigma}; \frac{\partial^2}{\partial \mu \partial \sigma}; \frac{\partial^2}{\partial \mu^2}; \frac{\partial^2}{\partial \sigma^2}\}$,

$$D \pi_k = D H_{\mu,\sigma}(S_k) - D H_{\mu,\sigma}(I_k). \tag{10.6a}$$

Pour les dérivées partielles par rapport aux paramètres r_a , remarquons d'abord que

$$\frac{\partial S_k}{\partial r_a} = \frac{X_{a,k+1} - X_{a,k}}{T_k - T_{k+1}} \text{ et } \frac{\partial^2 S_k}{\partial r_a \partial r_b} = 0 = \frac{\partial^2 I_k}{\partial r_a \partial r_b}, \text{ tandis que } \frac{\partial I_k}{\partial r_a} = \frac{X_{a,k} - X_{a,k-1}}{T_{k-1} - T_k}. \text{ Donc}$$

¹comme dans le cas d'une distribution uniforme, ou d'une loi log-normale. L'extension à une distribution caractérisée par plus de deux paramètres est immédiate. On se cantonne ici à deux paramètres car cela suffit pour détailler les interactions entre paramètres qui se manifestent dans la dérivation au second ordre de la log-vraisemblance

$$\frac{\partial \pi_k}{\partial r_a} = \frac{\partial H_{\mu,\sigma}(S_k)}{\partial x} \frac{\partial S_k}{\partial r_a} - \frac{\partial H_{\mu,\sigma}(I_k)}{\partial x} \frac{\partial I_k}{\partial r_a} \tag{10.6b}$$

$$\frac{\partial^2 \pi_m}{\partial r_a \partial r_b} = \frac{\partial^2 H_{\mu,\sigma}(S_m)}{\partial x^2} \frac{\partial S_m}{\partial r_a} \frac{\partial S_m}{\partial r_b} - \frac{\partial^2 H_{\mu,\sigma}(I_m)}{\partial x^2} \frac{\partial I_m}{\partial r_a} \frac{\partial I_m}{\partial r_b} \tag{10.6c}$$

$$\frac{\partial^2 \pi_k}{\partial r_a \partial \mu} = \frac{\partial^2 H_{\mu,\sigma}(S_k)}{\partial \mu \partial x} \frac{\partial S_k}{\partial r_a} - \frac{\partial^2 H_{\mu,\sigma}(I_k)}{\partial \mu \partial x} \frac{\partial I_k}{\partial r_a} \tag{10.6d}$$

$$\frac{\partial^2 \pi_k}{\partial r_a \partial \sigma} = \frac{\partial^2 H_{\mu,\sigma}(S_k)}{\partial \sigma \partial x} \frac{\partial S_k}{\partial r_a} - \frac{\partial^2 H_{\mu,\sigma}(I_k)}{\partial \sigma \partial x} \frac{\partial I_k}{\partial r_a} \tag{10.6e}$$

On en déduit, suivant les indications données au §.3, les dérivées partielles de la log-vraisemblance. J'ai donné ailleurs (Leurent, 1996a) un formulaire pour traiter une distribution log-normale des valeurs du temps.

10.2.3.5 Deux applications numériques

Calculons d'abord des estimateurs EMV pour les paramètres d'une distribution log-normale des valeurs du temps à partir des observations de l'USAP (Cf. tableau 10C). Les résultats obtenus (tableau 10D) sont conformes aux valeurs usuelles, avec des valeurs du temps moyennes $M = \exp(\mu + \sigma^2/2)$ de l'ordre de 60-70 F/h et des écarts-types $\hat{\sigma}$ fort proches de la valeur 0.6 déjà préconisée par Roger Marche (1988). La très basse dispersion relative indique que les résultats sont très significatifs : la qualité des résultats tient vraisemblablement à la forme de l'enquête USAP, conçue spécialement pour mettre en évidence la distribution continue de la valeur du temps.

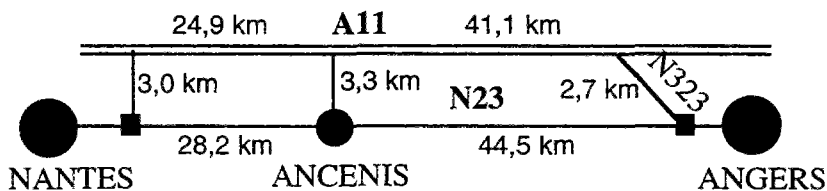
Tab. 10D. Déplacements VL, motif privé, heure creuse, à distance de 150 km.

Hypothèse	$\hat{\mu}$	$\hat{\sigma}$	M (F/h)	$\hat{\sigma}_{\hat{\mu}}$	$\hat{\sigma}_{\hat{\sigma}}$	Cov($\hat{\mu}, \hat{\sigma}$)	L(θ)
EMV à μ fixé	=3.9	0.634	60.4	-	0.030	-	-767.4
EMV à σ fixé	4.06	=0.600	69.4	0.030	-	-	-753.0
EMV	4.06	0.570	68.2	0.028	0.028	1.0e-4	-752.5

Nous avons mené trois estimations : la première en fixant la médiane, la seconde en fixant le paramètre de dispersion, enfin en estimant conjointement les deux paramètres.

Le second jeu d'observations est relatif à la liaison routière entre Nantes et Angers, qui s'effectue essentiellement à l'aide de deux itinéraires principaux (figure 10c) : une autoroute à péage (A11) vs. une route nationale (la RN23).

Fig. 10c. La concurrence A11-N23 entre Nantes et Angers.



Le CETE de l'Ouest (1993) a rassemblé des informations en provenance de plusieurs enquêtes aux itinéraires (surtout l'enquête cordon de 1989 à Nantes et celle de 1990 à Angers). Nous en avons retenu les résultats relatifs à trois paires origine-destination : respectivement Angers-Nantes, Nantes-Ancenis, Angers-Ancenis. Le tableau 10E présente les attributs objectifs des itinéraires en concurrence et les résultats de l'enquête aux itinéraires.

Tab. 10E. Attributs "objectifs" des itinéraires¹ et résultats de l'enquête.

Paire origine-destination		T_k (h)	D_k (km)	P_k (F)	dont péage ²	VL/jour observés
Nantes- Angers	N23	0.967	72.7	65.4	0	1500
	A11	0.617	71.7	86.7	36	3500
Nantes- Ancenis	N23	0.40	28.2	25.4	0	2600
	A11	0.30	31.2	33.9	11	1300
Angers- Ancenis	N23	0.567	44.5	40.1	0	500
	A11	0.417	47.1	50.8	17	450

Source : CETE de l'Ouest 1993.

Les valeurs estimées sont plausibles, et donnent un bon ajustement (tableau 10F).

Tab. 10F. Comparaison entre observation et simulations, liaison Angers-Nantes.

Relation origine-destination		Observation	valeurs standards des paramètres ³	paramètres estimés ⁴
Nantes- Angers	VL/jour par N23	1500	2619	1522
	VL/jour par A11	3500	2381	3477
Nantes- Ancenis	VL/jour par N23	2600	2850	2621
	VL/jour par A11	1300	1050	1279
Angers- Ancenis	VL/jour par N23	500	597	456
	VL/jour par A11	450	353	494
$L(\theta)$		-	-6765	-6199
Valeur du temps moyenne (F/h)		-	= 70	77.1

L'avant-dernière colonne affiche les valeurs simulées avec des paramètres fixés a priori, selon des valeurs conventionnelles. La dernière colonne montre les valeurs estimées.

¹les frais monétaires incorporent le péage et les frais de disposition et d'utilisation du véhicule, ainsi que le malus d'inconfort évalué forfaitairement selon la pratique du ministère français des transports, Cf. DR 1986

²Pour un véhicule léger. En plus du péage, on a intégré au prix le malus d'inconfort sur les sections non autoroutières et les coûts de disposition et circulation du véhicule, calculés avec les valeurs prescrites par la directive

³en fixant $\sigma = 0.6$ et $M = 70$ F/h donc $\mu = 4.07$

⁴avec $\hat{\mu} = 4.284$, $\hat{\sigma} = 0.348$, $\hat{\sigma}_{\hat{\mu}} = 0.0045$, $\hat{\sigma}_{\hat{\sigma}} = 0.01$ et $\text{Cov}(\hat{\mu}, \hat{\sigma}) = -1e-5$

10.2.3.6 Discussion

La méthode d'estimation proposée est simple, pour peu que l'on dispose d'un logiciel d'application convenable ⁽¹⁾. Elle possède une interprétation désagrégée, puisque chaque individu fournit une observation. Elle jouit des propriétés théoriques générales de l'estimateur du maximum de vraisemblance, avec en particulier l'estimation de l'incertitude sur les valeurs estimées.

Nos deux applications numériques démontrent la précision de la méthode : l'incertitude associée aux paramètres estimés est faible. Dans les deux cas nous avons trouvé une valeur du temps moyenne proche des valeurs conventionnelles pour une voiture circulant sur un réseau interurbain.

Mentionnons toutefois deux hypothèses restrictives importantes :

(i) nous avons supposé que tous les itinéraires enquêtés étaient efficaces, c'est-à-dire susceptibles d'offrir un temps généralisé minimum pour au moins une valeur du temps. Dans certaines enquêtes aux itinéraires, en particulier celle entre Bordeaux et Pau évoquée dans la section 10.4, ce n'est pas le cas parce que le prix et le temps moyen ne suffisent pas à expliquer le choix d'itinéraires secondaires. Ce défaut est mineur tant qu'il concerne seulement des itinéraires secondaires.

(ii) l'hypothèse que les attributs des itinéraires sont connus avec une précision parfaite, et qu'ils valent indistinctement pour tous les déplacements de la relation O-D. Cela néglige d'abord les parcours terminaux à l'intérieur des zones d'origine et de destination : assez plausible pour l'interurbain routier mais bien moins pour l'urbain ou l'interurbain non routier. Cela néglige surtout la variabilité intrinsèque de certains attributs, en particulier du temps de parcours. Imaginons un exemple dans lequel les temps varient au cours de la journée : un individu qui se déplace à une heure donnée ne subit pas les mêmes temps que ceux partis lors d'autres périodes ! Dans les deux cas traités ici, cela n'intervient que pour l'enquête entre Nantes et Angers, et là probablement de façon mineure car sur les itinéraires concernés les temps varient en général moins d'une minute (source CETE de l'Ouest 1993).

¹Nous avons développé un tel logiciel pour le compte du SETRA, Cf. Leurent (1995e)

10.3 PROPAGER L'ERREUR EXOGENE VIA LE MODÈLE BICRITÈRE

Propager l'erreur exogène a priori via le modèle de simulation sert à mettre en évidence l'incertitude qui pèse sur le résultat de la simulation en raison de l'incertitude relative aux variables exogènes. Cela permet de justifier le résultat obtenu et d'en quantifier la crédibilité statistique.

Nous avons examiné les méthodes antérieures pour propager l'erreur exogène : tests de sensibilité des praticiens qui donnent des indices mais pas des affirmations statistiques (Cf. section 3.4), affectation sur le réseau d'une incertitude relative aux flots origine-destination (TAM, Cf. §10.1.1.1), approche intégratrice de Roger Marche sans dimension spatiale (Cf. section 4.3 et §10.1.1.2). Ces méthodes sont limitées de par leur principe statistique (tests de sensibilité) ou le champ couvert, puisque le TAM ne prend pas en compte les objets d'estimation autres que les flots O-D, tandis que Marche élimine la dimension spatiale.

Nous formons donc le projet d'une méthode rigoureuse au plan statistique, couvrant les différents objets d'estimation et préservant la dimension spatiale, afin de propager l'erreur exogène via le modèle d'affectation bicritère. Pour cela nous adaptons les ressources théoriques qui assurent la rigueur statistique (Cf. §10.1.3) ; cela nécessite un algorithme d'affectation qui explicite les chemins, tels ceux présentés au chapitre précédent (MSA-iti et PET du §9.2.1).

En ouverture, nous traitons le cas du réseau à deux arcs qui se prête à un traitement entièrement analytique ; nous étudions la contribution potentielle d'un enrichissement conceptuel à la réduction de l'erreur exogène propagée (§10.3.1). Dans le cas général (§10.3.2), les formules de propagation ne se prêtent pas à une expression analytique : nous proposons donc une résolution numérique (§10.3.3).

10.3.1 Le cas du réseau à deux arcs

Considérons un réseau à deux arcs qui sont deux chemins concurrents joignant une unique origine à une unique destination : c'est le cas le plus élémentaire de concurrence entre itinéraires. Nous supposons qu'une des deux routes, notée F, est gratuite, tandis qu'un péage P est perçu sur l'autre route, notée T (en fait les routes pourraient être toutes deux payantes, et P serait alors la différence de prix).

La route à péage n'est compétitive que si son temps de parcours, T_T , est inférieur à celui de la route gratuite, T_F . Sous cette hypothèse, les usagers à faible valeur du temps v choisissent la route gratuite car elle est peu coûteuse même si elle est lente, tandis que les usagers à forte valeur du temps choisissent la route à péage car elle est rapide même si elle est chère. Formellement, nous supposons que chaque usager mesure la (dés)utilité d'un chemin k par rapport à sa propre valeur du temps v selon la formule de coût généralisé $G_k(v) = P_k + v.T_k$. Les usagers tels que $G_F(v) = v.T_F$ est moindre que $G_T(v) = P + v.T_T$ choisissent la route gratuite, tandis que ceux pour lesquels $G_T(v)$ est moindre que $G_F(v)$ choisissent la route à péage.

La valeur du temps \hat{v} qui délimite les deux groupes d'usagers est celle qui rend égaux les

temps généralisés des deux itinéraires :

$$T_F = T_T + P / \hat{v} \quad (10.7)$$

donc
$$\hat{v} = P / (T_F - T_T). \quad (10.8)$$

Notant H la fonction de répartition cumulée de la VOT, la proportion des usagers sur la route gratuite est $\Pr(v \leq \hat{v}) = H(\hat{v})$ tandis que la part de marché de la route à péage vaut $\Pr(v > \hat{v}) = 1 - H(\hat{v})$. Pour un volume O-D de q , le flot de la route gratuite est $f_F = qH(\hat{v})$ tandis que le flot sur la route à péage est

$$f_T = q(1 - H(\hat{v})). \quad (10.9)$$

10.3.1.1 Analyse de sensibilité à temps fixes et demande fixe

Intéressons-nous au flot sur la route à péage, désigné dans cette paragraphe comme le résultat du modèle. Si nous fixons les temps de parcours T_F et T_T ainsi que le volume de demande q , le résultat dépend directement des variables exogènes T_F , T_T , P , q et H (qui est une fonction). Une incertitude exogène δX sur un input X se répercute sur le résultat Y comme une petite variation δY . Selon Tukey (1957), pour un modèle à dépendance directe $Y = F(X)$, on a :

$$\delta Y = \nabla_X F \cdot \delta X, \quad (10.10)$$

qui est la formule de base pour l'analyse de sensibilité et la propagation des (petits) biais. En particulier une variation δX peut correspondre à un échantillon de l'erreur exogène ε_X ; dans ce cas δY est un échantillon de l'erreur exogène propagée ε_Y . En élevant au carré les deux membres de (10.10) et en prenant la moyenne des distributions d'erreur, nous obtenons la formule de base pour la propagation de la variance et de la covariance :

$$E(\varepsilon_Y^2) = \sum_{n,m} \left(\frac{\partial F}{\partial X_n} \frac{\partial F}{\partial X_m} \right) \text{Cov}(\varepsilon_{X_n}; \varepsilon_{X_m}), \quad (10.11)$$

où ε_{X_n} est l'erreur exogène sur la composante X_n de X .

Analyse de sensibilité du réseau à deux arcs. Dans ce cas la formule (10.10) livre, pour chaque composante X_n :

$$\delta f_T = \frac{\partial}{\partial X_n} \left(q(1 - H_\Theta \left(\frac{P}{T_F - T_T} \right)) \right) \cdot \delta X_n, \quad (10.12)$$

où Θ est le vecteur des paramètres de la fonction H . Nous considérons le vecteur d'inputs $X = [T_F; T_T; q; \theta_i]^t$. Les dérivées partielles sont respectivement :

$$\frac{\partial f_T}{\partial T_F} = -\frac{\partial f_T}{\partial T_T} = -q \frac{\partial H}{\partial v}(\hat{v}) \frac{\partial \hat{v}}{\partial T_F} = q \frac{P}{(T_F - T_T)^2} \frac{\partial H}{\partial v}(\hat{v}), \quad (10.13a, 10.13b)$$

$$\partial f_T / \partial q = 1 - H(\hat{v}) = f_T / q, \quad (10.13c)$$

$$\frac{\partial f_T}{\partial \theta_i} = -q \frac{\partial H}{\partial \theta_i}(\hat{v}). \quad (10.13d)$$

Nous considérons la dérivée partielle de H par rapport à v , plutôt que la dérivée totale, parce que $H_{\Theta}(v)$ dépend aussi des paramètres Θ .

Application numérique. Soit $P = 3$ \$, $q = 3\,000$ véh/h, $T_F = 0.431$ h, $T_T = 0.215$ h et une valeur du temps dont le logarithme naturel suit une distribution normale de moyenne $\mu = 2.48$ et d'écart-type $\sigma = 0.6$, ce qui fixe la valeur du temps moyenne à $M = \exp(\mu + \sigma^2/2) = 12$ \$/h. La valeur du temps de coupure est $\hat{v} = 14.0$ \$/h donc $H(\hat{v}) = 0.71$ et la part de marché du péage vaut $f_T/q = 0.29$. Les dérivées partielles de H valent respectivement : $\partial H/\partial v = 0.0082$, $\partial H/\partial \mu = -0.57$ et $\partial H/\partial \sigma = -0.32$ (voir Leurent, 1996e). Alors les dérivées partielles et les élasticités du résultat aux variables endogènes valent :

$$\begin{aligned} \frac{\partial f_T}{\partial T_F} &= -\frac{\partial f_T}{\partial T_T} = 7968 \text{ véh/h}^2 & \text{et} & \quad e_{\Delta T} = \partial \log f_T / \partial \log(T_F - T_T) = 1.97 \\ \frac{\partial f_T}{\partial q} &= 0.29 & \text{et} & \quad e_q = 1 \\ \frac{\partial f_T}{\partial \mu} &= 1720 \text{ véh/h} & \text{et} & \quad e_{\mu} = 7.73 \\ \frac{\partial f_T}{\partial \sigma} &= 960 \text{ véh/h} & \text{et} & \quad e_{\sigma} = 0.662. \end{aligned}$$

Les facteurs les plus sensibles sont ici la valeur du temps médiane $\exp(\mu)$ et le différentiel de temps de parcours $\Delta T = T_F - T_T$: un biais de 10% sur le différentiel de temps de parcours ΔT produit un biais de près de 20% sur le trafic et sur la recette !

Analyse de l'erreur. Pour expliciter la distribution de l'erreur sur le résultat, il nous faut des hypothèses sur la distribution des variables exogènes :

- i) sur les temps de parcours T_F et T_T , les erreurs exogènes ε_{T_F} et ε_{T_T} sont gaussiennes centrées de variances respectives $\sigma_{T_F}^2$ et $\sigma_{T_T}^2$; elles sont indépendantes entre elles et vis-à-vis des autres erreurs.
- ii) sur le niveau de demande q , l'erreur exogène ε_q est gaussienne centrée de variance σ_q^2 ; elle est indépendante des autres erreurs exogènes.
- iii) sur les paramètres μ et σ de la distribution des valeurs du temps, les erreurs exogènes ε_{μ} et ε_{σ} sont gaussiennes centrées de variances respectives σ_{μ}^2 et σ_{σ}^2 . Elles sont indépendantes des autres erreurs exogènes, mais corrélées entre elles avec une covariance $\text{Cov}(\varepsilon_{\mu}; \varepsilon_{\sigma})$.

Compte tenu de (10.10) et (10.11), l'erreur exogène propagée sur f_T est gaussienne, centrée, de variance :

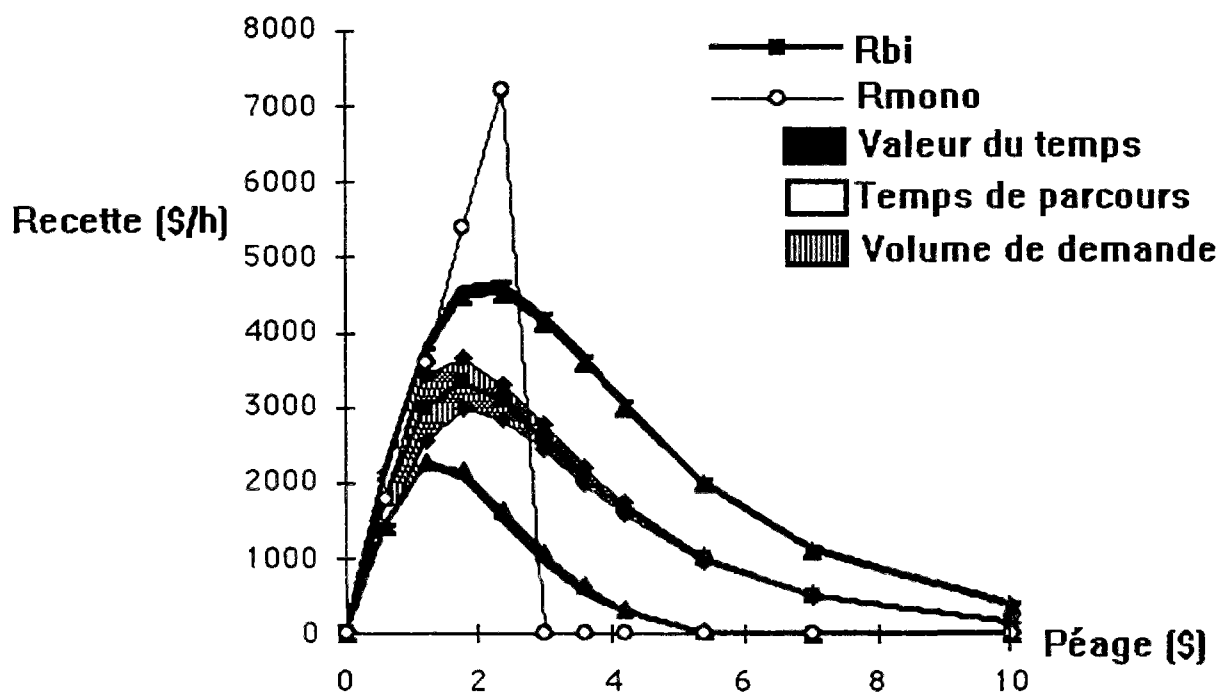
$$\sigma_{f_T}^2 = \left(\frac{\partial f_T}{\partial T_F}\right)^2 (\sigma_{T_F}^2 + \sigma_{T_T}^2) + \left(\frac{\partial f_T}{\partial q}\right)^2 \sigma_q^2 + \left(\frac{\partial f_T}{\partial \mu}\right)^2 \sigma_{\mu}^2 + \left(\frac{\partial f_T}{\partial \sigma}\right)^2 \sigma_{\sigma}^2 + 2 \frac{\partial f_T}{\partial \mu} \frac{\partial f_T}{\partial \sigma} \text{Cov}(\varepsilon_{\mu}; \varepsilon_{\sigma}). \quad (10.14)$$

En supposant que $\sigma_{T_T} / T_T = \sigma_{T_F} / T_F = \sigma_q / q = 10\%$, $\sigma_\mu / \mu = \sigma_\sigma / \sigma = 4\%$ et $Cov(\varepsilon_\mu; \varepsilon_\sigma) / \mu \sigma = -1\%$, nous obtenons que $\sigma_{f_T} / f_T = \sigma_{Recette} / Recette = 37\%$. L'incertitude sur les résultats est au moins double de l'incertitude sur chacune des variables exogènes ! Ce résultat est bien sûr particulier aux valeurs numériques retenues.

La figure 1 représente l'évolution de la recette R_{bi} en fonction du péage P . Nous avons tracé l'intervalle de confiance à 95% autour de la courbe moyenne $R_{bi}(P)$. Pour chaque valeur de P , nous avons décomposé la bande d'incertitude en trois postes qui correspondent à chacune des composantes de l'erreur exogène, selon sa contribution à la variance totale. Pour les valeurs faibles de P , l'incertitude sur le volume de demande (contribution de $\frac{\partial f_T}{\partial q} \delta q$) domine largement tandis que pour les valeurs fortes de P l'erreur sur les temps de parcours (contribution de $\frac{\partial f_T}{\partial T_F} (\varepsilon_{T_F} - \varepsilon_{T_T})$) prévaut sur elle. Dans tous les cas l'incertitude liée à la valeur du temps (contribution de $\frac{\partial f_T}{\partial \mu} \delta \mu + \frac{\partial f_T}{\partial \sigma} \delta \sigma$) reste relativement faible : l'explication tient à ce que l'on considère tout le spectre des valeurs du temps.

On constate que la courbe moyenne $R_{mono}(P)$ du modèle à une seule valeur du temps $M = \exp(\mu + \sigma^2 / 2)$ sort de l'intervalle de confiance à 95% sauf pour les valeurs très faibles de P : pour les petites valeurs (mais pas trop faibles) elle se situe largement au dessus de l'intervalle, tandis que pour les valeurs élevées elle tombe nettement en dessous ! Cela incite à rejeter le modèle à une seule valeur du temps.

Fig. 10d. Modèle à temps fixes et demande fixe.



10.3.1.2 Modèle à temps variables et demande fixe

Au chapitre 8, nous avons donné une formulation en dimension finie pour le modèle

d'affectation bicritère à temps variables. Dans le cas du réseau à deux arcs, si la demande est fixe, l'état d'équilibre du modèle est la solution de l'inéquation variationnelle définie sur l'ensemble $\mathbf{K}_q = \{ \mathbf{f} = (f_F; f_T) \geq \mathbf{0} ; f_F + f_T = q \}$:

$$\text{"trouver } \mathbf{f} \in \mathbf{K}_q \text{ tel que } \forall \mathbf{g} \in \mathbf{K}_q, \mathbf{V}(\mathbf{f}).(\mathbf{g} - \mathbf{f}) \geq 0 \text{"} \quad (10.15)$$

où les composantes du mapping \mathbf{V} sont définies par $V_F(\mathbf{f}) = t_F(\mathbf{f}) - P/H^{-1}(f_F/q)$ et $V_T(\mathbf{f}) = t_T(\mathbf{f})$ et ne diffèrent des fonctions de temps de parcours que par un terme correctif dans V_F .

Pour compléter la définition de V_F , soit $\eta'(\alpha) = 1/H^{-1}(\alpha)$ une fonction de $]0;1[$ vers \mathfrak{R} , avec $H^{-1}(\alpha) = \max\{v ; H(v) < \alpha\}$ la fonction réciproque de H . H^{-1} est bien définie et strictement positive sur $]0;1[$ car H est (faiblement) croissante. Aux points 0 et 1 il peut être nécessaire d'étendre la définition de η' . Si $H^{-1}(0) > 0$ alors $\eta'(0)$ est bien défini, mais si $H^{-1}(0) = 0$ nous posons $\eta'(0) = +\infty$; revenant à (10.15), la condition $H^{-1}(0) = 0$ implique qu'il y a des valeurs du temps très faibles dans la distribution et donc que le chemin gratuit doit être chargé à l'équilibre. Au point 1, si $H^{-1}(1) < +\infty$ alors $\eta'(1)$ est bien défini, sinon nous le posons égal à 0.

En associant un multiplicateur λ à la contrainte $f_F + f_T = q$, à l'équilibre on vérifie que si les deux chemins sont utilisés alors $V_F - \lambda = V_T - \lambda$ donc $V_F = V_T$ ou encore

$$H^{-1}(f_F/q) = P/(t_F(\mathbf{f}) - t_T(\mathbf{f})), \quad (10.16)$$

qui correspond exactement à (10.9) car $H^{-1}(f_F/q) = \hat{v}$.

La formulation en dimension finie a l'avantage d'éliminer le besoin de considérer que chaque valeur du temps est une variable du problème : seule la valeur du temps de coupure \hat{v} est prise en compte, et nous l'exprimons en fonction des variables de flot par chemin grâce à la transformation $f_F/q = H(\hat{v})$. Lorsqu'il y a plus de deux itinéraires, on les range dans l'ordre des prix croissants et l'on considère une valeur du temps de coupure pour chaque couple d'itinéraires adjacents (Cf. §8.2.3.1) : il suffit alors de remplacer, dans la condition (10.8), la variable f_F par la somme des flots des itinéraires de prix inférieur ou égal au prix du premier itinéraire du couple.

Analyse de sensibilité quand les deux chemins sont chargés, i.e. $\alpha = f_F/q \in]0;1[$. Soit $\mathbf{Y} = [f_F; f_T; \lambda]^t$ un vecteur formé d'une solution à (10.15) et du multiplicateur associé, et $\mathbf{J} = [V_F - \lambda; V_T - \lambda; f_F + f_T - q]^t = [t_F - P\eta'(\alpha) - \lambda; t_T - \lambda; f_F + f_T - q]^t$. En anticipant sur le §10.3.2, la solution d'équilibre \mathbf{Y} est une fonction continûment différentiable du vecteur d'input \mathbf{X} , avec un gradient donné par

$$\delta \mathbf{Y} = -(\nabla_{\mathbf{Y}} \mathbf{J})^{-1} \cdot \nabla_{\mathbf{X}} \mathbf{J} \cdot \delta \mathbf{X}. \quad (10.17)$$

Pour tirer parti de cette formule il faut inverser la matrice jacobienne $\nabla_{\mathbf{Y}} \mathbf{J}$:

$$\begin{aligned}
(\nabla_Y \mathbf{J})^{-1} &= \begin{bmatrix} \frac{\partial t_F}{\partial f_F} - \frac{P}{q} \frac{\partial \eta'}{\partial \alpha} & \frac{\partial t_F}{\partial f_T} & -1 \\ \frac{\partial t_T}{\partial f_F} & \frac{\partial t_T}{\partial f_T} & -1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} a & b & -1 \\ c & d & -1 \\ 1 & 1 & 0 \end{bmatrix}^{-1} = \frac{1}{a+d-b-c} \begin{bmatrix} 1 & -1 & d-b \\ -1 & 1 & a-c \\ c-d & b-a & ad-bc \end{bmatrix} \\
&= \frac{1}{\frac{\partial \Delta t}{\partial f_F} - \frac{\partial \Delta t}{\partial f_T} - \frac{P}{q} \frac{\partial \eta'}{\partial \alpha}} \begin{bmatrix} 1 & & & -\frac{\partial \Delta t}{\partial f_T} \\ & -1 & & \frac{\partial \Delta t}{\partial f_F} - \frac{P}{q} \frac{\partial \eta'}{\partial \alpha} \\ \frac{\partial t_T}{\partial f_F} - \frac{\partial t_T}{\partial f_T} & \frac{\partial t_F}{\partial f_F} - \frac{\partial t_F}{\partial f_T} + \frac{P}{q} \frac{\partial \eta'}{\partial \alpha} & (\frac{\partial t_F}{\partial f_F} - \frac{P}{q} \frac{\partial \eta'}{\partial \alpha}) & \frac{\partial t_T}{\partial f_T} - \frac{\partial t_F}{\partial f_T} \frac{\partial t_T}{\partial f_F} \end{bmatrix}, \quad (10.18)
\end{aligned}$$

avec $\Delta t = t_F - t_T$.

Nous considérons à nouveau le vecteur d'inputs $X = [T_F; T_T; q; \theta_i]^t$: une perturbation sur T_F ou T_T correspond à une perturbation additive sur la fonction de temps de parcours t_F ou t_T . Nous obtenons que

$$\nabla_X \mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{J}}{\partial T_F} & \frac{\partial \mathbf{J}}{\partial T_T} & \frac{\partial \mathbf{J}}{\partial q} & \frac{\partial \mathbf{J}}{\partial \theta_i} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{P}{q} \alpha \frac{\partial \eta'}{\partial \alpha} & -P \frac{\partial \eta'}{\partial \theta_i} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \quad (10.19)$$

$$\nabla_X f_T = \begin{bmatrix} \frac{\partial f_T}{\partial T_F} \\ \frac{\partial f_T}{\partial T_T} \\ \frac{\partial f_T}{\partial q} \\ \frac{\partial f_T}{\partial \theta_i} \end{bmatrix} = \frac{1}{\frac{\partial \Delta t}{\partial f_F} - \frac{\partial \Delta t}{\partial f_T} - \frac{P}{q} \frac{\partial \eta'}{\partial \alpha}} \begin{bmatrix} 1 \\ -1 \\ \frac{\partial \Delta t}{\partial f_F} - (1-\alpha) \frac{P}{q} \frac{\partial \eta'}{\partial \alpha} \\ -P \frac{\partial \eta'}{\partial \theta_i} \end{bmatrix} \begin{matrix} (10.20a) \\ (10.20b) \\ (10.20c) \\ (10.20d) \end{matrix}$$

Cette formule aurait pu être obtenue de façon plus directe en considérant une seule variable, f_T , puisque dans le cas de la demande fixe f_F s'en déduit via $f_F = q - f_T$. Cependant notre traitement sert le double objectif d'illustrer celui du §10.3.2 et d'examiner l'influence de chaque variable d'input. Comparons donc le système (10.20) au système (10.13).

Dans les deux cas le temps de parcours de la route gratuite exerce une influence exactement opposée à celle du temps de la route à péage. Si dans (10.20a, b) nous considérons des fonctions de temps de parcours constantes, nous retrouvons (10.13a, b) puisque $\frac{\partial \eta'}{\partial \alpha} = \frac{1}{H'(H^{-1}(\alpha)) \lambda(H^{-1}(\alpha))^2}$. Lorsque les fonctions de temps ne sont pas constantes, l'intuition commande que $\frac{\partial \Delta t}{\partial f_F} \geq 0$ et $\frac{\partial \Delta t}{\partial f_T} \leq 0$: de la croissance de H nous déduisons que $\frac{\partial \eta'}{\partial \alpha} \leq 0$, donc le dénominateur dans (10.20a, b) *majoré* celui dans (10.13a, b). Cela entraîne que la sensibilité du résultat aux inputs est *réduite* dans le cas à temps variables.

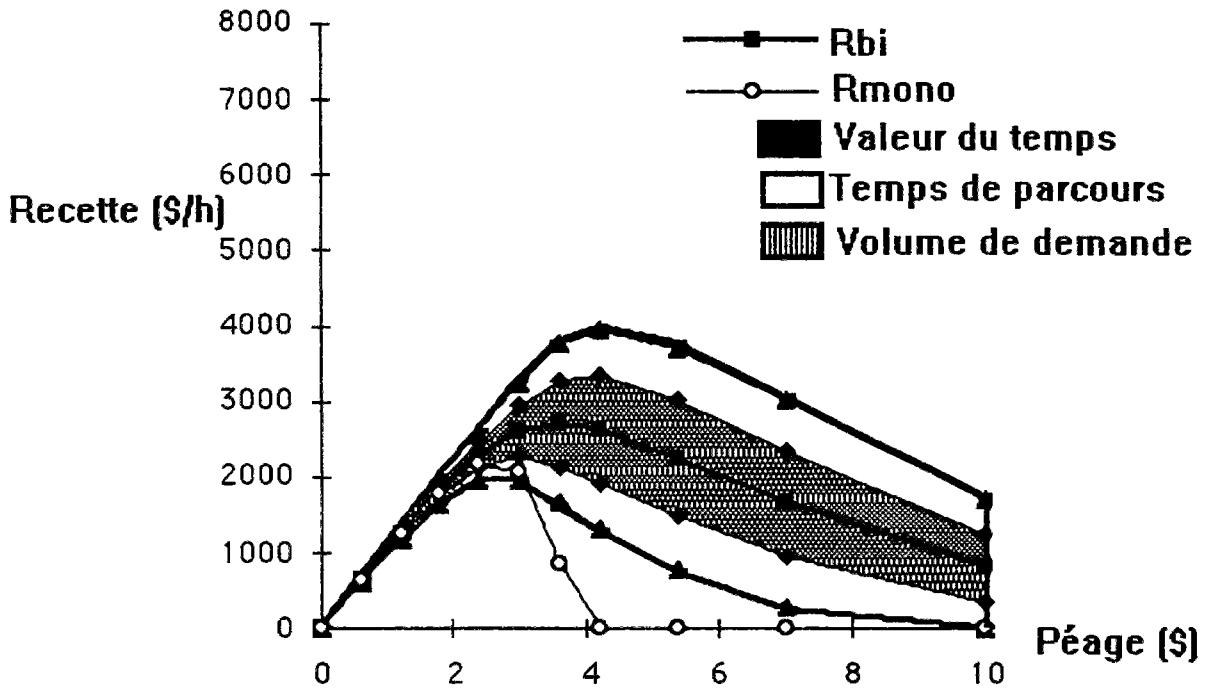
Concernant le volume de demande q , dans le cas de fonctions de temps constantes alors (10.20c) se ramène à (10.13c). Sinon, l'ajout du terme $\partial\Delta t / \partial f_F$ qui est positif (intuitivement !) au terme positif $-(1-\alpha)\frac{P}{q}\frac{\partial\eta'}{\partial\alpha}$ fait croître la valeur absolue du numérateur de (10.20c) par rapport à celui de (10.13c). Nous savons d'après l'examen des temps de parcours que le dénominateur croît également. La sensibilité du résultat au volume de demande est réduite dans le modèle à temps variables si et seulement si le rapport $(\partial\Delta t / \partial f_F) / (\partial\Delta t / \partial f_F - \partial\Delta t / \partial f_T)$ est inférieur à f_T / q : cela dépend des valeurs numériques particulières prises par les fonctions.

Pour étudier l'influence d'un paramètre θ_i de la fonction de distribution cumulée H , nous prenons l'exemple d'une valeur du temps distribuée uniformément sur l'intervalle $[0; A]$. En ce cas, sur $[0; +\infty[$ nous avons $H(v) = \min(v/A; 1)$ et sa réciproque est définie sur $[0; 1]$ par $H^{-1}(\alpha) = A\alpha$: d'où $\eta = 1/\alpha A$. Choisissons $\theta_i = A$: (10.20d) avec des fonctions de temps de parcours constantes livre $\partial f_T / \partial A = q (\partial\eta' / \partial A) / (\partial\eta' / \partial\alpha)$ puis $\partial f_T / \partial A = q\alpha / A = f_T / A$ que l'on obtient également avec (10.13d). Si les temps varient en fonction des débits, l'argument sur le dénominateur de (10.20) implique à nouveau que la sensibilité du résultat est réduite dans le modèle à temps variables.

Application numérique. Nous poursuivons l'exemple du §.1, en supposant maintenant que le temps de parcours de chaque arc dépend du débit local par la formule $t_k(\mathbf{f}) = t_k^0 (1 + \gamma_k + [\alpha_k \cdot (1 - \frac{f_k}{N_k})]^2 + \beta_k^2)^{1/2} - \alpha_k \cdot (1 - \frac{f_k}{N_k}) - \beta_k$ dans laquelle t_k^0 est un temps de parcours libre, N_k est une capacité pratique, α_k et γ_k ($> \alpha_k$) sont des paramètres dont on déduit $\beta_k = \gamma_k(\alpha_k - \gamma_k/2) / (\alpha_k - \gamma_k)$. On fixe les valeurs numériques des paramètres supplémentaires à $t_F^0 = 0.30$ h, $N_F = 5\,000$ véh/h, $\alpha_F = 2.5$ et $\gamma_F = 1.5$ pour le chemin gratuit, et à $t_T^0 = 0.18$ h, $N_T = 1\,000$ véh/h, $\alpha_T = 4.0$ et $\gamma_T = 0.5$ pour le chemin à péage. Au point $P = 3$ \$, le différentiel de temps de parcours vaut encore 0.214 h.

La figure 10e montre l'évolution des recettes R_{bi} et $R_{mono}(P)$ en fonction du péage P , sous les mêmes hypothèses quant à l'erreur exogène a priori et avec les mêmes principes de représentation que pour la figure 10d. Nous observons que la contribution de la valeur du temps demeure faible, comme sur la figure 10d. Mais contrairement au cas à temps fixes, les contributions respectives des temps de parcours et des volumes de demande sont grossièrement équivalentes à toutes les valeurs du péage. Au point $P = 3$ \$ pour lequel les deux modèles ne diffèrent l'un de l'autre que par les coefficients de propagation, ceux du modèle à temps variables sont inférieurs, confirmant l'analyse de (10.20). Enfin, la différence entre les deux courbes de recette R_{bi} et $R_{mono}(P)$ est bien moindre que dans le cas à temps fixes, car par effet de congestion l'affectation s'effectue sur plusieurs chemins même dans le modèle avec une seule valeur du temps.

Fig. 10e. Modèle à temps variables et demande fixe.



10.3.1.3 Modèle à temps variables et demande élastique

Nous supposons désormais que le volume de demande q dépend du niveau de service \bar{S} , défini de façon agrégée comme le temps généralisé en moyenne sur toutes les valeurs du temps :

$$\bar{S} = \int \min\{t_F ; t_T + P / v\} dH(v). \tag{10.21}$$

Soit D la fonction de demande qui relie q à \bar{S} :

$$q = D(\bar{S}). \tag{10.22}$$

Cette dépendance est une *approximation* ; elle n'est parfaitement justifiée que si pour chaque valeur du temps v la densité de demande q_v pour cette valeur du temps obéit à $q_v = D(\min\{t_F ; t_T + P / v\})$ avec une fonction de demande D affine et identique pour toutes les valeurs du temps (Cf. §8.2.2.3).

Finalement, définissons sur l'ensemble des flots positifs ou nuls $\mathbf{K} = \{f = (f_F ; f_T) \geq 0\}$ une application \mathbf{W} par :

$$\mathbf{W}(f) = \begin{bmatrix} W_F(f) \\ W_T(f) \end{bmatrix} = \begin{bmatrix} t_F(f) + P(\eta(1) - \eta(\alpha) - (1 - \alpha)\eta'(\alpha)) - D^{-1}(f_F + f_T) \\ t_T(f) + P(\eta(1) - \eta(\alpha) + \alpha\eta'(\alpha)) - D^{-1}(f_F + f_T) \end{bmatrix}, \tag{10.23}$$

avec $D^{-1}(q) = \min\{x ; D(x) < q\}$, $\alpha = f_F / (f_F + f_T)$ et η une primitive de η' .

L'application \mathbf{W} a pour composantes les fonctions de temps de parcours par chemins, corrigées par des termes qui incorporent la distribution de la valeur du temps et la variabilité du volume de demande.

L'état d'équilibre du modèle bicritère à temps variables et demande variables est la

solution de l'inéquation variationnelle suivante :

$$\text{"trouver } \mathbf{f} \in \mathbf{K} \text{ qui vérifie } \forall \mathbf{g} \in \mathbf{K}, \mathbf{W}(\mathbf{f}) \cdot (\mathbf{g} - \mathbf{f}) \geq 0\text{"}. \quad (10.24)$$

A l'équilibre, si $f_F / q = \alpha \in]0; 1[$ alors $W_F = W_T = 0$ ce qui équivaut à :

i) $t_F(\mathbf{f}) - P \eta'(\alpha) = t_T(\mathbf{f})$ qui revient à (10.8),

ii) $D^{-1} = \alpha(W_F + D^{-1}) + (1 - \alpha)(W_T + D^{-1}) = \int_0^\alpha t_F d\xi + \int_\alpha^1 \{t_T + P \eta'(\xi)\} d\xi$ ce qui, après changement de la variable d'intégration de ξ en v selon $\xi = H(v)$, livre (10.21) puisqu'à l'équilibre les valeurs du temps inférieures à $H^{-1}(\alpha)$ éprouvent un temps généralisé minimum sur la route gratuite, tandis que celles supérieures à cette valeur de coupure sont mieux servies par la route à péage.

Soit $\mathbf{Y} = [f_F; f_T]^t$ une solution de (10.24). Nous appliquons à nouveau (10.17) avec cette fois $\mathbf{J} = \mathbf{W}$. Sous l'hypothèse que $0 < \alpha < 1$, en notant $D^{-1/}$ la dérivée partielle de D^{-1} par rapport à son argument principal q , nous obtenons que :

$$\nabla_{\mathbf{Y}} \mathbf{J} = \begin{bmatrix} \frac{\partial t_F}{\partial f_F} - \frac{P}{q} (1 - \alpha)^2 \frac{\partial \eta'}{\partial \alpha} - D^{-1/} & \frac{\partial t_F}{\partial f_T} + \frac{P}{q} \alpha (1 - \alpha) \frac{\partial \eta'}{\partial \alpha} - D^{-1/} \\ \frac{\partial t_T}{\partial f_F} + \frac{P}{q} \alpha (1 - \alpha) \frac{\partial \eta'}{\partial \alpha} - D^{-1/} & \frac{\partial t_T}{\partial f_T} - \frac{P}{q} \alpha^2 \frac{\partial \eta'}{\partial \alpha} - D^{-1/} \end{bmatrix},$$

$$(\nabla_{\mathbf{Y}} \mathbf{J})^{-1} = \frac{1}{\text{Det}} \begin{bmatrix} \frac{\partial t_T}{\partial f_T} - \frac{P}{q} \alpha^2 \frac{\partial \eta'}{\partial \alpha} - D^{-1/} & -\frac{\partial t_F}{\partial f_T} - \frac{P}{q} \alpha (1 - \alpha) \frac{\partial \eta'}{\partial \alpha} + D^{-1/} \\ -\frac{\partial t_T}{\partial f_F} - \frac{P}{q} \alpha (1 - \alpha) \frac{\partial \eta'}{\partial \alpha} + D^{-1/} & \frac{\partial t_F}{\partial f_F} - \frac{P}{q} (1 - \alpha)^2 \frac{\partial \eta'}{\partial \alpha} - D^{-1/} \end{bmatrix} \quad (10.25)$$

$$\text{où } \text{Det} = |\nabla_{\mathbf{Y}} \mathbf{J}| = \frac{\partial t_F}{\partial f_F} \frac{\partial t_T}{\partial f_T} - \frac{\partial t_T}{\partial f_F} \frac{\partial t_F}{\partial f_T} - D^{-1/} \cdot \left(\frac{\partial \Delta t}{\partial f_F} - \frac{\partial \Delta t}{\partial f_T} - \frac{P}{q} \frac{\partial \eta'}{\partial \alpha} \right) - \frac{P}{q} \frac{\partial \eta'}{\partial \alpha} \left(\alpha^2 \frac{\partial t_F}{\partial f_F} + (1 - \alpha)^2 \frac{\partial t_T}{\partial f_T} + \alpha (1 - \alpha) \left(\frac{\partial t_T}{\partial f_F} + \frac{\partial t_F}{\partial f_T} \right) \right),$$

en utilisant l'abréviation $\Delta t = t_F - t_T$.

Analyse de sensibilité. Nous prenons pour vecteur d'inputs $\mathbf{X} = [T_F; T_T; q_0; \theta_i]^t$, dans lequel q_0 représente un paramètre de la fonction de demande D (donc de la fonction de demande réciproque D^{-1}). On exprime :

$$\nabla_{\mathbf{X}} \mathbf{J} = \begin{bmatrix} \frac{\partial \mathbf{J}}{\partial T_F} & \frac{\partial \mathbf{J}}{\partial T_T} & \frac{\partial \mathbf{J}}{\partial q_0} & \frac{\partial \mathbf{J}}{\partial \theta_i} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -\frac{\partial D^{-1}}{\partial q_0} & P \left[\frac{\partial \eta(1)}{\partial \theta_i} - \frac{\partial \eta(\alpha)}{\partial \theta_i} - (1 - \alpha) \frac{\partial \eta'(\alpha)}{\partial \theta_i} \right] \\ 0 & 1 & -\frac{\partial D^{-1}}{\partial q_0} & P \left[\frac{\partial \eta(1)}{\partial \theta_i} - \frac{\partial \eta(\alpha)}{\partial \theta_i} + \alpha \frac{\partial \eta'(\alpha)}{\partial \theta_i} \right] \end{bmatrix}, \quad (10.26)$$

$$\begin{bmatrix} \frac{\partial f_T}{\partial T_F} \\ \frac{\partial f_T}{\partial T_T} \\ \frac{\partial f_T}{\partial q_0} \\ \frac{\partial f_T}{\partial \theta_i} \end{bmatrix} = \frac{1}{\text{Det}} \begin{bmatrix} \frac{\partial t_T}{\partial f_F} + \frac{P}{q} \alpha(1-\alpha) \frac{\partial \eta'}{\partial \alpha} - D^{-1/} \\ -\frac{\partial t_F}{\partial f_F} + \frac{P}{q} (1-\alpha)^2 \frac{\partial \eta'}{\partial \alpha} + D^{-1/} \\ \left(\frac{\partial \Delta t}{\partial f_F} - (1-\alpha) \frac{P}{q} \frac{\partial \eta'}{\partial \alpha} \right) \frac{\partial D^{-1/}}{\partial q_0} \\ -P \left(\left(\frac{\partial \eta(1)}{\partial \theta_i} - \frac{\partial \eta(\alpha)}{\partial \theta_i} \right) \left(\frac{\partial \Delta t}{\partial f_F} - \frac{P}{q} (1-\alpha) \frac{\partial \eta'}{\partial \alpha} \right) + \frac{\partial \eta'}{\partial \theta_i} \left(\frac{\partial t_F}{\partial f_F} + \alpha \frac{\partial \Delta t}{\partial f_F} - D^{-1/} \right) \right) \end{bmatrix} \begin{matrix} (10.27a) \\ (10.27b) \\ (10.27c) \\ (10.27d) \end{matrix}$$

L'ajout du mécanisme d'équilibre quantitatif entre offre et demande (10.27) renforce le couplage des inputs dans les formules de propagation. Tandis que dans (10.20) et (10.13) leurs effets combinés étaient principalement additifs, dans (10.26) apparaissent également des effets multiplicatifs.

Démontrons que (10.27) se ramène à (10.20) si l'on fixe la demande fixe, dans le cas d'une fonction de demande qui présente une élasticité constante au temps généralisé moyen. Alors $D^{-1}(q) = S_0(q/q_0)^{1/e_0}$ et $D^{-1/}(q) = \bar{S}/e_0q$. Si $e_0 \rightarrow 0^-$ alors $D^{-1/} \rightarrow -\infty$ donc domine tout terme fini auquel on l'ajoute. Ainsi Det se réduit au terme qui contient $D^{-1/}$, puis (10.27a, b, d) se réduisent respectivement à (10.20a, b, d). (10.25c) se réduit à (10.20c) multiplié par q/q_0 parce que $\partial D^{-1/}/\partial q_0 = -q D^{-1/}/q_0$: or $e_0 \rightarrow 0^-$ donc $q/q_0 \approx 1$.

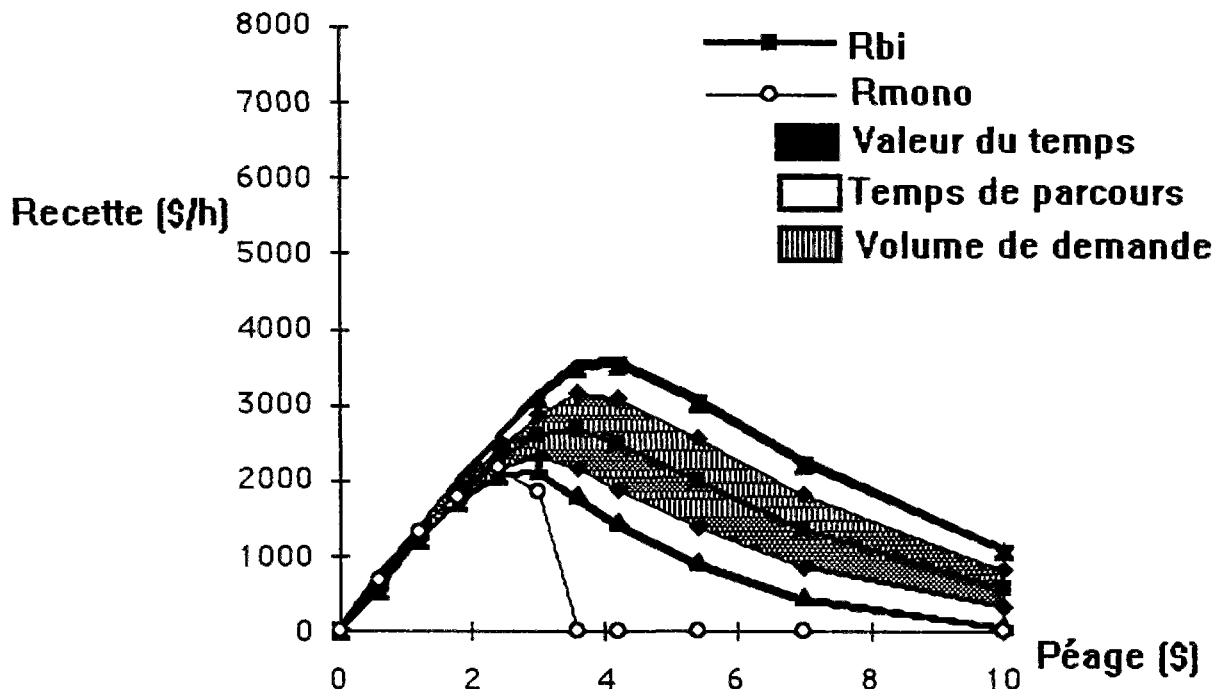
Sous l'hypothèse intuitive que les fonctions de temps ne décroissent pas vis-à-vis du flot, local ou non, et que $\partial t_F/\partial f_F \geq \partial t_T/\partial f_F$ et $\partial t_T/\partial f_T \geq \partial t_F/\partial f_T$, Det est la somme de trois termes positifs ou nuls, à savoir : la différence de deux produits de dérivées de fonctions de temps, un deuxième terme qui contient $D^{-1/}$ et un troisième dans lequel le péage multiplie des dérivées des fonctions de temps de parcours. Dans ces conditions Det est plus grand que son propre deuxième terme. Donc, au point $q = q_0$, la sensibilité à q_0 du modèle à demande variable d'élasticité constante est moindre que la sensibilité du modèle à demande fixe au volume q . De même la sensibilité au temps de parcours de la route à péage décroît : le numérateur de (10.27b) comprend deux termes non négatifs qui équilibrent le terme négatif ou nul $D^{-1/}$ et réduisent la valeur absolue, alors que le dénominateur croît en valeur absolue.

Application numérique. Nous poursuivons l'exemple des §.1 et §.2, en modifiant simplement le volume de demande que nous faisons dépendre du temps généralisé moyen avec une élasticité constante $e_0 = -0.6$, en prenant la valeur $q_0 = 3\,000$ véh/h au point $S_0 = 0.413$ h. Ainsi au point $P = 3$ \$ les données d'application sont identiques aux exemples précédents, à l'exclusion des coefficients de propagation.

La figure 10f montre l'évolution des recettes R_{bi} et $R_{mono}(P)$ en fonction du péage P , sous les mêmes hypothèses quant à l'erreur exogène a priori et avec les mêmes principes de représentation que pour la figure 10d. Nous observons que la contribution de la valeur du temps est légèrement plus importante que précédemment. Comme pour le modèle à demande fixe et temps variables, les contributions respectives des temps de parcours et des volumes de demande sont grossièrement équivalentes à toutes les

valeurs du péage, ainsi que l'avait démontré l'analyse de (10.27). Dans les figures 10e et 10f les modèles à une seule valeur du temps sous-estiment la recette à toutes les valeurs du péage : l'erreur est négligeable pour les faibles valeurs mais considérable pour les valeurs moyennes et fortes.

Fig. 10f. Modèle à temps variables et demande variable.



10.3.2 Le cas général

Au paragraphe précédent, nous avons fourni un traitement analytique complet pour propager l'erreur exogène à travers le modèle d'affectation bicritère, dans le cas du réseau à deux arcs. En ajoutant progressivement des mécanismes explicatifs, nous avons observé qu'un enrichissement conceptuel permet de réduire l'erreur exogène propagée sur le résultat.

Dans le cas général d'un problème d'affectation de taille quelconque, les mêmes formules de propagation existent mais ne peuvent être résolues directement comme pour le réseau à deux arcs. Nous allons les établir pour le modèle d'affectation bicritère à distribution continue de la valeur du temps, demande élastique, temps de parcours variables en fonction du flot et contraintes adjointes.

Les hypothèses du modèle et la caractérisation mathématique ont été données au chapitre 8. Au paragraphe §.1, nous modifions la caractérisation mathématique afin de tirer avantage dans le §.2 des résultats d'analyse de sensibilité établis par Tobin et Friesz (1988). Alors le paragraphe §.3 donne les formules d'analyse de sensibilité pour le modèle bicritère dans le cas général. Enfin le paragraphe §.4 détaille le traitement de certains inputs importants.

10.3.2.1 Une caractérisation mathématique modifiée

Pour appliquer les résultats rappelés au §10.1.3 au modèle bicritère, nous adoptons l'inéquation variationnelle réduite donnée au §8.2.4.2, de manière que seul apparaisse comme variable endogène le vecteur de flots par chemin f , et aussi pour considérer des

contraintes linéaires affines $d_i(\mathbf{f}) = 0$ pour des valeurs de i entre 1 et N_d , afin en particulier de fixer le volume de demande de certaines relations O-D. Alors, dans la définition d'un équilibre bicritère, chaque temps de parcours T_{rs}^k d'un chemin est accru d'un terme $-\sum_i \lambda_i \partial d_i / \partial f_{rs}^k$ dans lequel λ_i est le multiplicateur associé à la contrainte $d_i(\mathbf{f}) = 0$. Lorsqu'une telle contrainte sert à fixer un volume O-D, cela ne modifie pas les différences entre temps de parcours des chemins de l'O-D en question.

Nous supposons dans ce qui suit que :

- i) les fonctions z_b sont convexes et différentiables (c'est le cas notamment pour des contraintes de capacité du type $x_a - C_a \leq 0$),
- ii) il y a des contraintes linéaires affines $d_i(\mathbf{f}) = 0$,
- iii) les gradients $\nabla z_b(\mathbf{f})$ et $\nabla d_i(\mathbf{f})$ sont linéairement indépendants.

Soit $\mathbf{K} = \{\mathbf{f} \geq \mathbf{0} ; z_b(\mathbf{f}) \leq 0 \forall b \in 1..N_b \text{ et } d_i(\mathbf{f}) = 0 \forall i \in 1..N_d\}$ et \mathbf{U} l'application vectorielle allant de \mathbf{K} vers \mathcal{R}^{N_k} où N_k est le nombre de chemins, définie par $\mathbf{U}(\mathbf{f}) = [I_{rs}^k(\mathbf{f}; \mathbf{0})]_{rsk}$.

Théorème de rappel du Théorème (8.53). Un vecteur de flots par chemins $\mathbf{f}^* \geq \mathbf{0}$ participe à un équilibre bicritère soumis aux contraintes $z_b(\mathbf{f}) \leq 0$ et $d_i(\mathbf{f}) = 0$ ssi il est solution de l'inéquation variationnelle suivante (10.28):

$$\text{"trouver } \mathbf{f}^* \in \mathbf{K} \text{ tel que } \forall \mathbf{f} \in \mathbf{K}, \mathbf{U}(\mathbf{f}^*) \cdot (\mathbf{f} - \mathbf{f}^*) \geq 0." \quad (10.28)$$

10.3.2.2 Formules de propagation pour le modèle bicritère

Pour appliquer la formule (AS8) (Cf. §10.1.3.1) au modèle bicritère, il nous faut définir l'ensemble \mathbf{K} , les applications \mathbf{F} , \mathbf{g} et \mathbf{d} , ainsi que les multiplicateurs π et λ . Les notations \mathbf{K} et \mathbf{d} des §10.1.3.1 et §10.3.2.1 sont équivalentes, donc le vecteur \mathbf{f} du §10.3.2.1 correspond au vecteur \mathbf{x} du §10.1.3.1. Les contraintes $\mathbf{g}(\mathbf{x}) \geq \mathbf{0}$ du §10.1.3.1 correspondent à la fois aux conditions $\mathbf{f} \geq \mathbf{0}$ et aux inégalités $\mathbf{z}(\mathbf{f}) \leq \mathbf{0}$, mises sous la forme $-\mathbf{z}(\mathbf{f}) \geq \mathbf{0}$.

Nous faisons coïncider l'application \mathbf{F} du §10.1.3.1 avec l'application vectorielle \mathbf{U} avec autant de composantes que de chemins, la composante d'indice $r-s-k$ associant à \mathbf{f} la valeur $I_{rs}^k(\mathbf{f}; \mathbf{0})$ (Cf. §8.2.3 et §.1). Pour satisfaire aux hypothèses de Tobin and Friesz (1988), nous supposons que \mathbf{U} est monotone ; une condition suffisante est que l'application $\mathbf{f} \mapsto [T_{rs}^k(\mathbf{f}; \mathbf{0})]_{rsk}$ soit monotone, puisque les termes complémentaires

$I_{rs}^k - T_{rs}^k$ dérivent d'une fonction convexe (Leurent, 1993a). Dans ces conditions on obtient l'unicité d'une solution par rapport à chaque flot par classe monétaire. Dans le cas du modèle bicritère, une solution extrême non dégénérée peut compter autant de chemins à flot strictement positif qu'il y a d'arcs et de classes monétaires.

Scindons le vecteur π en deux sous-vecteurs \mathbf{w} et $\bar{\mathbf{f}}$ dont les composantes w_b et \bar{f}_{rs}^k

sont associées respectivement aux contraintes $-z_b \geq 0$ et $f_{rs}^k \geq 0$. Alors, pour une perturbation ε , l'application vectorielle \mathbf{J} a pour composantes :

$$\mathbf{U}(\mathbf{f}|\varepsilon) - \bar{\mathbf{f}} + \sum_b w_b \nabla z_b(\mathbf{f}|\varepsilon) - \sum_{i=1}^{N_d} \lambda_i \nabla d_i(\mathbf{f}|\varepsilon)^t, \quad (10.29a)$$

$$-w_b \cdot z_b(\mathbf{f}|\varepsilon) \text{ pour les valeurs de } b \text{ entre } 1 \text{ et } N_b, \quad (10.29b)$$

$$\bar{f}_{rs}^k f_{rs}^k \text{ pour tous les triplets origine-destination-chemin } r-s-k, \quad (10.29c)$$

$$d_i(\mathbf{f}|\varepsilon) \text{ pour les valeurs de } i \text{ entre } 1 \text{ et } N_d. \quad (10.29d)$$

Pour restreindre le problème comme dans Tobin et Friesz (1988), nous omettons les variables qui correspondent aux chemins sans charge à l'équilibre. Alors le système (10.29) se réduit à :

$$\mathbf{U}(\mathbf{f}|\varepsilon) + \sum_b w_b \nabla z_b(\mathbf{f}|\varepsilon) - \sum_{i=1}^{N_d} \lambda_i \nabla d_i(\mathbf{f}|\varepsilon)^t, \quad (10.30a)$$

$$-w_b \cdot z_b(\mathbf{f}|\varepsilon) \text{ pour les valeurs de } b \text{ entre } 1 \text{ et } N_b, \quad (10.30b)$$

$$d_i(\mathbf{f}|\varepsilon) \text{ pour les valeurs de } i \text{ entre } 1 \text{ et } N_d. \quad (10.30c)$$

Donnons maintenant des formules explicites pour $\nabla_{\mathbf{y}} \mathbf{J}$ et $\nabla_{\varepsilon} \mathbf{J}$. Nous désignons les composantes de \mathbf{J} dans (10.30a) par J_{rsk} , celles dans (10.30b) par J_b et celles dans (10.30c) par J_i . Alors :

$$\nabla_{\mathbf{y}} \mathbf{J} = \begin{bmatrix} \left[\frac{\partial J_{rsk}}{\partial f_{r's'}^{k'}} \right]_{rsk,r's'k'} & \left[\frac{\partial J_{rsk}}{\partial w_b} \right]_{rsk,b} & \left[\frac{\partial J_{rsk}}{\partial \lambda_i} \right]_{rsk,i} \\ \left[\frac{\partial J_b}{\partial f_{rs}^k} \right]_{b,rsk} & \left[\frac{\partial J_b}{\partial w_{b'}} \right]_{b,b'} & \left[\frac{\partial J_b}{\partial \lambda_i} \right]_{b,i} \\ \left[\frac{\partial J_i}{\partial f_{rs}^k} \right]_{i,rsk} & \left[\frac{\partial J_i}{\partial w_b} \right]_{i,b} & \left[\frac{\partial J_i}{\partial \lambda_{i'}} \right]_{i,i'} \end{bmatrix}$$

$$= \begin{bmatrix} \left[\frac{\partial J_{rsk}}{\partial f_{r's'}^{k'}} \right]_{rsk,r's'k'} & \left[\frac{\partial z_b}{\partial f_{rs}^k} \right]_{rsk,b} & \left[-\frac{\partial d_i}{\partial f_{rs}^k} \right]_{rsk,i} \\ \left[-w_b \frac{\partial z_b}{\partial f_{rs}^k} \right]_{b,rsk} & \left[-\delta_{b,b'} z_b \right]_{b,b'} & \left[0 \right]_{b,i} \\ \left[\frac{\partial d_i}{\partial f_{rs}^k} \right]_{i,rsk} & \left[0 \right]_{i,b} & \left[0 \right]_{i,i'} \end{bmatrix}, \quad (10.31)$$

dans lequel $\delta_{b,b'} = 1$ si $b = b'$ ou 0 sinon.

D'après Leurent (1993a) (rappelé au §8.2.4.3) :

$$\frac{\partial J_{rsk}}{\partial f_{r's'}^{k'}} = \left(\sum_a \delta_{rs}^{ak} \frac{\partial t_a}{\partial f_{r's'}^{k'}} \right) + \left(\sum_b w_b \frac{\partial^2 z_b}{\partial f_{r's'}^{k'} \partial f_{rs}^k} \right) - \left(\sum_i \lambda_i \frac{\partial^2 d_i}{\partial f_{r's'}^{k'} \partial f_{rs}^k} \right) + \delta_{r,r'} \delta_{s,s'} \left(-\frac{\partial D_{rs}^{-1}}{\partial q_{rs}} + \sum_{n=1}^{\bar{m}_{rs}-1} \frac{\bar{p}_{rs}^n - \bar{p}_{rs}^{n+1}}{q_{rs}} \cdot (\Gamma_{rs}^{nk} - \frac{q'_{rs,n}}{q_{rs}}) \cdot (\Gamma_{rs}^{nk'} - \frac{q'_{rs,n}}{q_{rs}}) \cdot \frac{\partial \eta'_{rs}}{\partial \alpha} \left(\frac{q'_{rs,n}}{q_{rs}} \right) \right), \quad (10.32)$$

dans lequel $\Gamma_{rs}^{nk} := \partial q'_{rs,n} / \partial f_{rs}^k = 1$ si le prix du chemin k est inférieur ou égal au prix des chemins de la classe isoprix n ou sinon $\Gamma_{rs}^{nk} = 0$.

Concernant $\nabla_{\epsilon} \mathbf{J}$, pour une composante ϵ_c de ϵ nous obtenons que :

$$\frac{\partial J_{rsk}}{\partial \epsilon_c} = \left(\sum_a \delta_{rs}^{ak} \frac{\partial t_a}{\partial \epsilon_c} \right) + \left(\sum_b w_b \frac{\partial^2 z_b}{\partial \epsilon_c \partial f_{rs}^k} \right) - \left(\sum_i \lambda_i \frac{\partial^2 d_i}{\partial \epsilon_c \partial f_{rs}^k} \right) - \frac{\partial D_{rs}^{-1}}{\partial \epsilon_c} (q_{rs}) + \bar{p}_{rs} \bar{m}_{rs} \frac{\partial \eta_{rs}(1)}{\partial \epsilon_c} - \bar{p}_{rs}^1 \frac{\partial \eta_{rs}(0)}{\partial \epsilon_c} + \sum_{i=1}^{\bar{m}_{rs}-1} (\bar{p}_{rs}^i - \bar{p}_{rs}^{i+1}) \left(\frac{\partial \eta_{rs}}{\partial \epsilon_c} \left(\frac{q'_{rs,i}}{q_{rs}} \right) - \frac{q'_{rs,i}}{q_{rs}} \cdot \frac{\partial \eta'_{rs}}{\partial \epsilon_c} \left(\frac{q'_{rs,i}}{q_{rs}} \right) \right) \quad (10.33a)$$

$$\frac{\partial J_b}{\partial \epsilon_c} = -w_b \frac{\partial z_b}{\partial \epsilon_c}, \quad (10.33b)$$

$$\frac{\partial J_i}{\partial \epsilon_c} = \frac{\partial d_i}{\partial \epsilon_c}. \quad (10.33c)$$

Ainsi nous disposons de formules pour analyser la sensibilité du modèle bicritère. Pour obtenir les dérivées partielles de la solution par rapport à une perturbation ϵ_c , on peut inverser la matrice jacobienne (10.29) et multiplier le résultat par le vecteur dont les composantes figurent dans (10.33a, b, c). De façon alternative, si l'on considère seulement quelques paramètres de perturbation, on peut résoudre autant de systèmes linéaires que de paramètres, plutôt qu'inverser la matrice.

10.3.2.4 Traitement de quelques inputs importants

Donnons maintenant des expressions analytiques sous forme fermée pour $\nabla_{\varepsilon} \mathbf{J}$ dans le cas de paramètres ε_c qui apparaissent soit dans une fonction de demande D_{rs} (un tel paramètre est alors noté $q_{rs}^{(0)}$), soit dans la distribution des valeurs du temps H_{rs} (on note alors θ_j), soit dans une fonction de temps de parcours t_a (on note alors T_a) ou dans une contrainte adjointe z_b ou d_i (on note alors dans les deux cas $\theta_{b'}$). On obtient facilement :

$$\frac{\partial J_{rsk}}{\partial T_a} = \delta_{rs}^{ak} \frac{\partial t_a}{\partial T_a} \text{ et se réduit à } \delta_{rs}^{ak} \text{ si une perturbation } T_a \text{ s'applique additivement sur}$$

la fonction de temps de parcours t_a ,

$$\frac{\partial J_{rsk}}{\partial \theta_{b'}} = \left(\sum_b w_b \frac{\partial^2 z_b}{\partial \theta_{b'} \partial f_{rs}^k} \right) - \left(\sum_i \lambda_i \frac{\partial^2 d_i}{\partial \theta_{b'} \partial f_{rs}^k} \right),$$

$$\frac{\partial J_{rsk}}{\partial q_{r's'}^{(0)}} = - \frac{\partial D_{rs}^{-1}}{\partial q_{r's'}^{(0)}},$$

$$\frac{\partial J_{rsk}}{\partial \theta_j} = P_{rs}^{\bar{m}_{rs}} \frac{\partial \eta_{rs}}{\partial \theta_j}(1) - P_{rs}^1 \frac{\partial \eta_{rs}}{\partial \theta_j}(0) + \sum_{i=1}^{\bar{m}_{rs}-1} (\bar{P}_{rs}^i - \bar{P}_{rs}^{i+1}) \left[\frac{\partial \eta_{rs}}{\partial \theta_j} \left(\frac{q'_{rs,i}}{q_{rs}} \right) + \left(\Gamma_{rs}^{ik} - \frac{q'_{rs,i}}{q_{rs}} \right) \frac{\partial \eta'_{rs}}{\partial \theta_j} \left(\frac{q'_{rs,i}}{q_{rs}} \right) \right],$$

avec toujours $\Gamma_{rs}^{nk} := \partial q'_{rs,n} / \partial f_{rs}^k = 1$ si le prix du chemin k est inférieur ou égal au prix de la classe monétaire n ou 0 sinon.

$$\partial J_b / \partial T_a = 0 \text{ et } \partial J_b / \partial q_{rs}^{(0)} = 0 \text{ et } \partial J_b / \partial \theta_j = 0,$$

$$\partial J_b / \partial \theta_{b'} = -\delta_{b,b'} w_b \partial z_b / \partial \theta_b \text{ dans lequel } \delta_{u,v} = 1 \text{ si } u = v \text{ ou } 0 \text{ sinon,}$$

$$\partial J_i / \partial T_a = 0 \text{ et } \partial J_i / \partial \theta_{b'} = \partial d_i / \partial \theta_{b'} \text{ et } \partial J_i / \partial q_{rs}^{(0)} = 0 \text{ et } \partial J_i / \partial \theta_j = 0.$$

Si le volume de la relation O-D $\hat{r} - \hat{s}$ est fixé au niveau $q_{\hat{r}\hat{s}}$, nous imposons une contrainte égalité $d_{\hat{r}} = (\sum_k f_{\hat{r}\hat{s}}^k) - q_{\hat{r}\hat{s}} = 0$ et nous considérons une perturbation $\theta_{\hat{b}'} = q_{\hat{r}\hat{s}}$. Alors :

$$\forall r-s-k, \partial d_{\hat{r}} / \partial f_{rs}^k = \delta_{r,\hat{r}} \delta_{s,\hat{s}} \text{ et } \forall b', \partial^2 d_{\hat{r}} / \partial \theta_{b'} \partial f_{rs}^k = 0,$$

$$\forall b', \partial d_{\hat{r}} / \partial \theta_{b'} = -\delta_{b',\hat{b}'},$$

$$\forall r-s-k, \partial J_{rs}^k / \partial \theta_{\hat{b}'} = 0,$$

$$\forall b, \partial J_b / \partial \theta_{\hat{b}'} = 0,$$

$$\forall i, \partial J_i / \partial \theta_{\hat{i}} = -\delta_{i,\hat{i}}.$$

10.3.3 Un traitement numérique dans le cas général

Afin de résoudre les formules d'analyse de sensibilité développées ci-dessus pour le modèle bicritère, nous avons conçu un algorithme et développé un programme.

10.3.3.1 Des contraintes techniques imposent une résolution "numérique"

Trois voies s'ouvrent a priori pour propager l'erreur exogène par le modèle bicritère prix-temps :

- i) le calcul analytique et l'établissement de formules directes exprimant l'erreur en sortie en fonction de l'erreur en entrée.
- ii) le calcul numérique, en application de formules trop complexes pour en déduire des expressions analytiques.
- iii) la simulation numérique, en appliquant répétitivement le modèle à des données volontairement perturbées, la comparaison des résultats livrant a posteriori la sensibilité du modèle aux facteurs de perturbation.

Nous avons choisi la deuxième voie, celle du calcul numérique, faute de pouvoir exprimer analytiquement $(\nabla_{\mathbf{y}} \mathbf{J})^{-1}$ et donc de poursuivre dans la première voie. La voie du calcul numérique est valable pour des problèmes de dimension petite ou moyenne, car le nombre de variables est un multiple du nombre de chemins donc a fortiori du nombre de relations origine-destination. Comme certaines opérations (inversion de matrice...) coûtent de l'ordre du cube ou du carré du nombre de variables, en espace mémoire comme en temps de calcul, il serait difficile (actuellement !) de traiter sur micro-ordinateur plus de quelques centaines de chemins.

Pour les problèmes de grande taille, la seule solution pratique est la simulation numérique, donc la troisième voie : elle permet d'étudier le rôle de quelques facteurs globaux comme par exemple la valeur du temps moyenne, mais pas d'analyser systématiquement les effets conjoints de multiples facteurs d'importance limitée (par exemple des variations simultanées des temps de parcours de toutes les arêtes du réseau) car il faudrait pour cela des simulations en nombre prohibitif.

10.3.3.2 Organigramme du traitement

La formule de propagation s'applique en un point d'équilibre, pour lequel on dispose des résultats d'affectation¹ c'est-à-dire d'une part le vecteur \mathbf{f}^* des flots par chemins et d'autre part le vecteur \mathbf{w}^* des multiplicateurs associés aux contraintes. Alors les calculs d'analyse de sensibilité et d'erreur peuvent suivre l'organigramme que voici :

- i) calcul de $\nabla_{\mathbf{y}} \mathbf{J}$ et de $\nabla_{\boldsymbol{\varepsilon}} \mathbf{J}$,

¹ obtenu par exemple par un algorithme d'affectation composé d'une boucle extérieure pour traiter les contraintes adjointes par une méthode de Lagrangien augmenté (Cf. §9.1.4.6). A l'intérieur de cette boucle extérieure, on résout des programmes d'optimisation convexe non contraints, en utilisant par exemple la Procédure d'Égalisation par Transvasement.

- ii) inversion de $\nabla_y \mathbf{J}$ par une méthode de pivot partiel,
- iii) calcul de $-\left[\nabla_y \mathbf{J}\right]^{-1} \cdot \nabla_\varepsilon \mathbf{J}$,
- iv) pour une distribution déterminée de l'erreur exogène ε , calculer la distribution de l'erreur exogène propagée sur le flot de chaque chemin.
- v) sur un arc déterminé a_0 , agréger les distributions d'erreur des flots de tous les chemins qui empruntent cet arc.

Ainsi l'on obtient la distribution de l'erreur exogène propagée sur le flot x_{a_0} de l'arc a_0 , exprimé comme une combinaison linéaire des erreurs exogènes ε .

Nous avons programmé la procédure de propagation sur un micro-ordinateur compatible IBM-PC, avec le compilateur Turbo Pascal pour Windows de la société Borland.

10.4 APPLICATION À UNE VRAIE ÉTUDE DE TRAFIC

Pour tester la procédure décrite dans la section précédente, nous l'avons appliquée à une vraie étude de trafic routier en milieu interurbain.

10.4.1 Le site et le modèle de trafic

Nous avons considéré une autoroute interurbaine à péage qui appartient à l'un des itinéraires entre Bordeaux et Pau dans le Sud-Ouest de la France. Les deux villes, distantes de 220 km, sont reliées par quatre itinéraires routiers principaux schématisés sur la figure 10g. Pour les besoins de l'étude de trafic, 180 arcs unidirectionnels et 19 zones de demande avaient été codées par le CETE du Sud-Ouest (1).

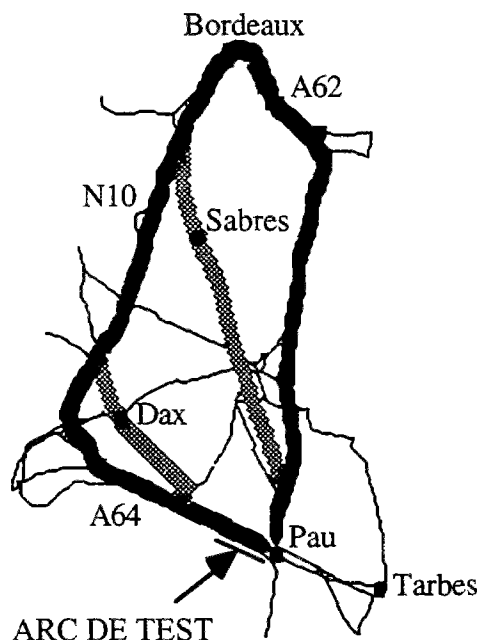
Décrivons le modèle d'affectation employé. Du côté de la demande, nous avons distingué deux classes de véhicules, respectivement les voitures et les poids lourds auxquels nous avons assigné un coefficient d'équivalence de deux unités de voitures particulières. La valeur du temps des voitures suivait une loi log-normale ; la distribution du logarithme naturel de la valeur du temps avait pour moyenne $\mu = 2.48$ et pour écart-type $\sigma = 0.6$, ce qui assurait une valeur du temps moyenne de $\exp(\mu + \sigma^2/2) = 12$ \$/h pour les voitures. Quant aux poids lourds, la distribution du logarithme naturel de leur valeur du temps avait pour moyenne $\mu_h = 3.73$ et pour écart-type $\sigma_h = 0.6$, d'où une valeur du temps moyenne de 50 \$/h. Pour chaque classe de trafic et chaque relation O-D, le volume de demande dépendait du temps généralisé moyen selon une fonction à élasticité constante égale à -0.6, i.e. $q_{rs} = D_{rs}(\bar{S}_{rs}) = q_{rs}^{(0)}(\bar{S}_{rs} / \bar{S}_{rs}^{(0)})^{-0.6}$.

Du côté de l'offre, nous n'avons pas imposé de contraintes adjointes. Sur chaque arc, nous avons spécifié une fonction de temps de parcours selon la formule du BPR, i.e. $t_a(x_a) = t_a^0 \cdot (1 + 0.15(\frac{x_a}{C_a})^4)$ avec t_a^0 temps de parcours libre, C_a capacité pratique et x_a le débit de l'arc en unités de voitures particulières. Le coût de traversée d'un arc

¹merci à Pierre Baillet et Francis Bugeaud pour leur aimable coopération

comprenait les dépenses liées à la consommation de carburant et le péage là où il s'appliquait, c'est-à-dire sur la plupart des arcs autoroutiers (avec un prix kilométrique d'environ 0.08 \$/km pour une voiture particulière). Le prix d'un chemin donné était calculé comme la somme des prix des arcs parcourus le long du chemin ; signalons que cette formule de prix par chemin n'est pas obligatoire dans notre formulation du modèle bicritère, contrairement à celles de Marcotte et Zhu (1994) et de Dial (1996) : cela pourrait se révéler utile dans l'étude d'un schéma de péage variable selon les postes d'origine et de destination sur un cordon.

Fig. 10g. Le site de l'étude.



10.4.2 L'erreur exogène et sa propagation

10.4.2.1 Hypothèses a priori sur l'erreur exogène

Nous avons étudié l'erreur exogène propagée sur le trafic de voitures d'une route particulière a_0 , en supposant que l'erreur exogène tenait aux causes suivantes :

- i) les temps de trajet t_a sur les arcs. Le temps de trajet de chaque arc a subissait une erreur exogène gaussienne centrée ε_{T_a} d'écart-type σ_{T_a} tel que $\sigma_{T_a} / T_a = 10\%$, et indépendante des autres erreurs.
- ii) les volumes de demande par relation O-D pour le trafic de voitures. Le volume de référence $q_{rs}^{(0)}$ de chaque relation O-D $r-s$ subissait une erreur gaussienne centrée $\varepsilon_{q_{rs}^{(0)}}$ d'écart-type $\sigma_{q_{rs}^{(0)}}$ tel que $\sigma_{q_{rs}^{(0)}} / q_{rs}^{(0)} = 20\%$, et indépendante des autres erreurs.
- iii) les paramètres μ et σ d'une distribution de la valeur du temps commune à toutes les relations O-D : μ étant la moyenne et σ l'écart-type du logarithme naturel de la valeur du temps. μ et σ subissaient des erreurs gaussiennes centrées

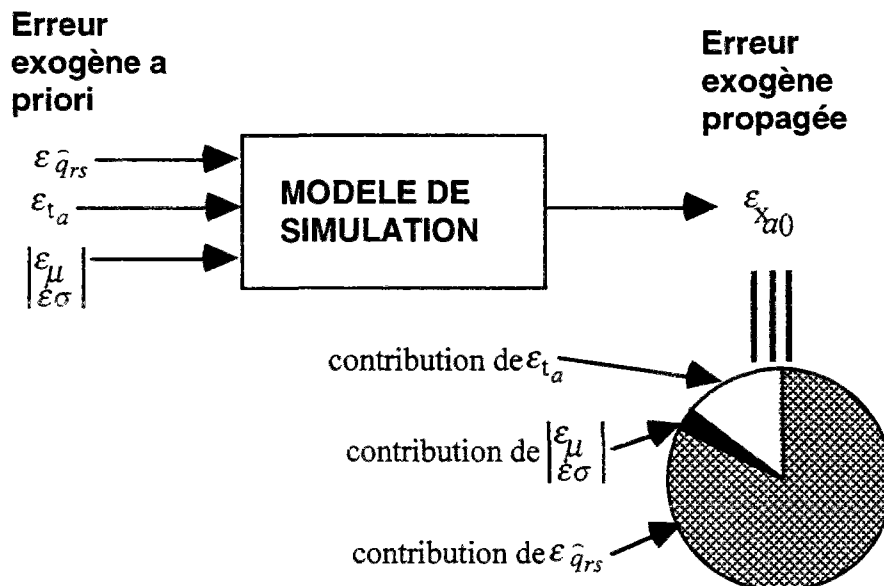
ε_μ et ε_σ , d'écart-types respectifs σ_μ et σ_σ fixés par $\sigma_\mu / \mu = 1\%$ et $\sigma_\sigma / \sigma = 3\%$. ε_μ et ε_σ étaient supposées indépendantes des autres erreurs mais corrélées entre elles avec une covariance fixée par $\text{Cov}(\varepsilon_\mu; \varepsilon_\sigma) / \mu\sigma = -0.1\%$.

10.4.2.2 Résultat de la propagation

Sur la route testée a_0 , pour un débit simulé moyen x_{a0} de 1220 voitures/jour, après propagation de l'erreur exogène a priori nous avons obtenu une erreur a posteriori de moyenne nulle et d'écart-type $\sigma_{x_{a0}} = 123$ voitures/jour, donc $\sigma_{x_{a0}} / x_{a0} = 10.1\%$. Cette incertitude provient principalement des volumes de demande (83%, voir figure 10h) et à un degré moindre des temps de parcours (14%) et de la distribution des valeurs du temps (3%). Les exemples du §10.3.1 suggèrent que les valeurs absolue et relative de l'erreur propagée sont particulières à chaque cas.

Dans le cas présent nous pouvons conclure que l'incertitude exogène propagée est raisonnablement faible, et que les tentatives pour la réduire devraient porter en priorité sur l'amélioration de la connaissance de la matrice O-D. Au sujet de la distribution de la valeur du temps, le faible effet de l'erreur exogène relative à ses paramètres tient à ce que tout le spectre est considéré ; de plus, comme les paramètres ont été estimés simultanément (Cf. §10.2.3), leurs erreurs exogènes respectives sont liées et se contrebalancent (au travers de $\text{Cov}(\varepsilon_\mu; \varepsilon_\sigma)$).

Fig. 10h. L'accumulation des erreurs exogènes.



Chapitre 11

Bilan et perspectives

Au terme de la thèse, nous récapitulons son sujet et indiquons des perspectives d'application et de développement. Après avoir reconnu cinq sources d'erreur dans un modèle : conception, formulation, calcul, erreur d'estimation et erreur exogène a priori, nous avons inscrit l'audit d'un modèle dans quatre dimensions : conceptuelle, formelle, algorithmique et économétrique.

Chaque dimension correspond à des objectifs spécifiques et procède selon une méthode propre, qualitative ou quantitative. L'audit conceptuel explicite le contenu explicatif du modèle afin d'en apprécier la portée explicative. L'audit formel examine la formule caractéristique, synthèse mathématique des composants conceptuels, afin d'en apprécier la cohérence logique et d'établir des propriétés d'existence et d'unicité d'une solution. L'audit algorithmique contrôle la production d'une solution. L'audit économétrique quantifie l'erreur liée au résultat du modèle et la rattache aux erreurs liées aux variables exogènes de la simulation.

Pour démontrer par l'exemple la méthode proposée, nous avons traité le cas particulier d'un modèle d'affectation bicritère ; nous en avons préparé l'audit par une synthèse des principaux éléments relatifs à l'affectation statique du trafic.

Ainsi nous avons traité de l'audit à trois niveaux : principes, préparation (capitaliser le domaine de modélisation), traitement d'un modèle particulier dans ce domaine. La récapitulation du sujet reprend cette distinction entre trois niveaux (section 11.1). Les applications potentielles tiennent à la capitalisation d'un domaine et à l'emploi d'un modèle audité (section 11.2). Nous ouvrons des pistes de recherche relatives d'une part à la méthode d'audit, d'autre part à l'affectation et plus largement à la modélisation des déplacements : une capitalisation nous semble prioritaire (section 11.3).

11.1 RÉCAPITULATION DU SUJET

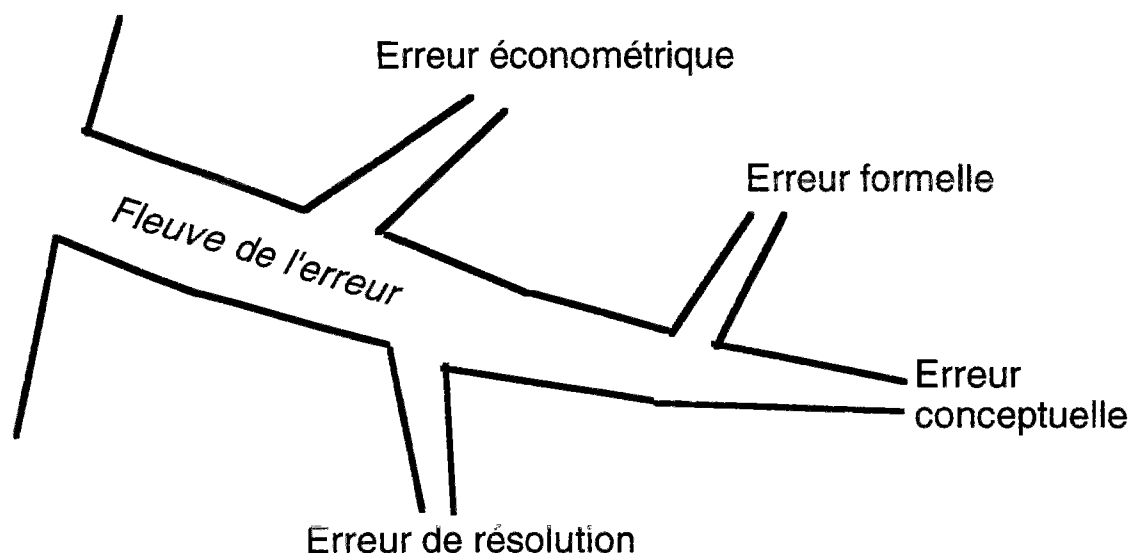
Nous récapitulons le sujet de l'audit en distinguant selon trois niveaux : celui du discours méthodologique d'abord (§11.1.1), puis le niveau d'un domaine spécifique de modélisation (§11.1.2), enfin celui d'un modèle particulier, le modèle bicritère (§11.1.3).

Le niveau intermédiaire, celui d'un domaine spécifique, nous a servi à préparer le traitement du modèle particulier, en capitalisant des concepts opératoires et des méthodes à portée générale et dont on trouve la trace dans le modèle particulier.

11.1.1 Niveau des principes et de la méthode

Reconnaissant la nature hétérogène de l'erreur d'un modèle, nous l'avons décomposée en quatre natures : conception, formulation, erreur de résolution et erreur économétrique (§.1). Aussi nous croyons nécessaire d'auditer un modèle selon quatre dimensions : conceptuelle, formelle, algorithmique, économétrique, avec pour chaque dimension des questions et des concepts opératoires propres (§.2). Nous avons proposé une méthode pour chaque dimension, qualitative ou quantitative, visant à assécher l'erreur ou simplement à l'endiguer, la canaliser (§.3).

Fig. 11a. Schéma de principe du fleuve de l'erreur.



11.1.1.1 Décomposer l'erreur de modélisation en quatre natures

L'erreur d'un modèle se décompose en erreur de conception, erreur de formulation, erreur algorithmique, erreur économétrique, celle-ci recouvrant à la fois l'erreur d'estimation et l'erreur exogène a priori.

L'erreur de conception (ou *sémantique*) tient à la schématisation du système étudié, à la définition et au choix des variables explicatives (exogènes) ou expliquées (endogènes), à la mise en relation de ces éléments par des mécanismes explicatifs.

La formulation du modèle consiste à synthétiser les mécanismes explicatifs par une formule caractéristique, qui relève d'un problème type de mathématique. *L'erreur de formulation* procède d'un manque de conformité entre la formule caractéristique et les mécanismes élémentaires spécifiés dans la conception, ou d'un manque de cohérence logique, ou d'une insuffisante caractérisation.

L'*erreur de résolution* (ou *algorithmique*) tient au dispositif de résolution de la formule caractéristique (support matériel, logiciel, opérateur). Si la résolution procède par simulation numérique, le dispositif peut manifester des imprécisions par arrondi ou un défaut de convergence.

L'*erreur économétrique* d'un modèle de simulation englobe l'*erreur d'estimation* liée à la spécification numérique de la formule caractéristique, et l'*erreur exogène a priori* relative aux variables exogènes. Elle limite l'aptitude empirique du modèle, sa capacité à reproduire des états observés (que des observations aient ou non servi à l'alimentation économétrique du modèle).

11.1.1.2 Projet d'audit : expliciter l'erreur, dégager des concepts opératoires

Notre entreprise d'audit vise à connaître et circonscrire l'erreur dans le modèle de simulation, à définir des garanties techniques qui posent autant d'exigences, à prescrire comment satisfaire à ces exigences.

A cette fin, nous analysons l'erreur d'un modèle, nous en reconnaissons les causes et la nature. La décomposition en quatre natures témoigne de l'analyse menée. Selon sa nature, chaque source tient à des causes spécifiques, ce qui détermine la forme d'audit qui lui est applicable. Nous proposons donc de décliner l'audit selon quatre dimensions : conceptuelle, formelle, algorithmique et économétrique.

A chaque dimension d'audit s'attachent des concepts opératoires spécifiques, valables autant pour l'audit que pour la modélisation.

L'audit conceptuel révèle des *composants sémantiques*, en particulier des *mécanismes explicatifs* qui constituent l'explication théorique du système étudié.

L'audit formel examine la *formule caractéristique* du modèle, qui constitue la synthèse mathématique des mécanismes explicatifs spécifiés dans la conception. Il s'agit de vérifier la logique et la rigueur mathématique, ainsi que la conformité aux mécanismes explicatifs. L'audit formel étudie également l'existence et l'unicité des solutions à la formule caractéristique.

L'erreur de résolution coïncide dans le cas d'une simulation numérique avec l'erreur de calcul, qui tient aux défauts du dispositif de calcul ou de la méthode de calcul. Nous ne traitons pas des défauts du dispositif de calcul (support matériel, implantation logicielle, actions des opérateurs). L'audit algorithmique traque les défauts de la méthode de résolution : il s'agit de détecter les imprécisions par arrondi, et surtout d'examiner la convergence des calculs vers une solution à la formule caractéristique. Pour mesurer la convergence, il faut disposer d'un *critère de convergence théoriquement rigoureux* qui constitue la garantie technique. L'exigence relative à cette garantie consiste à produire une solution potentielle qui réduise suffisamment le critère : pour y parvenir il faut un *algorithme* efficace.

L'audit économétrique détecte l'erreur économétrique en identifiant les objets et les relations auxquels elle s'attache : ces objets et relations, en rapport direct avec ceux de la conception, sont des lieux d'erreur. On qualifie d'exogène a priori l'incertitude qui s'attache aux variables exogènes, liée à une connaissance imprécise ou à une supputation (ex. dans la simulation d'une situation virtuelle). L'erreur d'estimation tient à la procédure d'obtention des paramètres des relations à partir d'observations ; elle revient in fine à une incertitude exogène a priori. L'audit économétrique explicite l'*incertitude*

exogène a priori relative aux variables exogènes et aux paramètres. Il examine et discute les hypothèses statistiques qui déterminent cette incertitude, dont il étudie les *effets sur le résultat* d'une simulation.

11.1.1.3 Une méthode qualitative ou quantitative, toujours systématique

Les concepts opératoires introduits ont une portée générale tant pour la modélisation que pour l'audit. Pour mener à bien celui-ci, il faut appliquer¹ une méthode. Nous avons proposé une méthode particulière qui consiste à dissocier les dimensions d'audit et à traiter chacune d'elles de façon spécifique.

Pour l'audit conceptuel, après une analyse systémique servant à identifier les objets et les relations élémentaires, nous conseillons d'explicitier les mécanismes explicatifs contenus dans le modèle de simulation sous la forme d'une *liste de phénomènes*. Cette liste peut alors être comparée à la *connaissance théorique du système étudié*, elle-même formalisée comme une liste de phénomènes avec pour chacun une *échelle de valeurs* permettant d'apprécier la qualité de représentation.

Pour l'audit formel, nous proposons d'abord de valider la logique et la rigueur mathématiques de la formule caractéristique du modèle, en contrôlant logiquement chaque étape de l'obtention de la formule caractéristique, puis de vérifier que la formule caractéristique implique chacun des mécanismes explicatifs. En cas d'échec, il convient de modifier la formule caractéristique... ou sinon l'on renonce à la validité technique du modèle ! Pour étudier l'existence et l'unicité de solutions à la formule caractéristique, nous conseillons de la rattacher à un *cadre mathématique de référence*, autrement dit à un problème type de recherche opérationnelle, afin de bénéficier de résultats théoriques d'existence et d'unicité.

Pour l'audit algorithmique, nous conseillons d'inventorier des critères de convergence rigoureux et des algorithmes, puis de les tester comparativement afin de choisir un critère et un algorithme sur la base d'applications expérimentales.

Pour l'audit économétrique, nous proposons de *quantifier l'incertitude exogène a priori* liée aux variables exogènes et aux paramètres : pour cela on peut répartir les variables exogènes et les paramètres en plusieurs *classes d'objets d'estimation* et attacher à chaque classe d'objets d'estimation un modèle d'estimation qui permet de préciser les hypothèses statistiques. Alors on peut *propager l'incertitude exogène a priori via le modèle de simulation* pour déterminer l'incertitude liée au résultat et la *justifier* à partir des hypothèses statistiques de la simulation.

Au total, la méthode d'audit proposée consiste à décomposer selon des dimensions relevant chacune d'une discipline principale : analyse systémique pour la conception, mathématiques pour la formulation, physique et algorithmique pour la dimension algorithmique, statistique pour la dimension économétrique. La validité technique d'ensemble suppose la validité dans toutes les dimensions : en particulier la validité de l'audit économétrique est conditionnelle aux validités conceptuelle, formelle et algorithmique. Les dimensions conceptuelle et formelle donnent lieu à des conclusions qualitatives, tandis que les dimensions algorithmique et économétrique permettent des

¹explicitement ou tacitement. L'explicitation offre l'avantage de la clarté, déterminant si l'on cherche à prendre une garantie technique...

considérations quantitatives. Nous conseillons de réduire complètement, d'assécher, l'erreur formelle et l'erreur algorithmique. Pour l'erreur conceptuelle et l'erreur économétrique, il s'agit de prendre en compte l'erreur, de la canaliser, en l'explicitant et en la surveillant ; sa réduction passe par une sophistication de modélisation, soit du modèle de simulation pour l'erreur de conception et la propagation, soit du modèle d'estimation pour l'erreur exogène a priori.

11.1.2 Préparer l'audit dans un domaine, c'est capitaliser !

Avant d'appliquer le discours méthodologique à un modèle particulier, nous avons préparé le terrain en inventoriant et critiquant les ressources du domaine de modélisation dont relève le modèle particulier, en l'occurrence le domaine de l'affectation statique du trafic routier pour le modèle bicritère. Une telle préparation n'est autre qu'une entreprise de capitalisation du domaine, puisqu'elle procède par inventaire des éléments existants et par sélection des concepts les plus pertinents et des méthodes les plus efficaces.

Nous récapitulons la capitalisation du domaine de l'affectation statique du trafic routier, telle qu'effectuée dans les chapitres précédents (§.1) et nous en tirons des enseignements (§.2).

11.1.2.1 Le domaine de l'affectation statique du trafic routier

Pour préparer l'audit conceptuel, nous avons inventorié les composants sémantiques d'une affectation : caractère spatial, temporalité, identification des itinéraires possibles sur un réseau, perception d'un itinéraire comme un agrégat d'attributs, représentation de la congestion, différenciation de types de véhicules, élasticité du volume de demande au niveau de service, rationalité économique de la demande, différenciation de classes de demande. Pour chaque phénomène nous avons élaboré une échelle de valeurs qui classe des solutions possibles de représentation.

Pour préparer l'audit formel, nous avons inventorié les cadres mathématiques de référence auxquels on peut rattacher la formule caractéristique d'un modèle d'affectation, et nous avons donné des exemples de rattachement. Nous avons indiqué les principaux résultats d'existence et d'unicité disponibles dans chacun de ces cadres : inéquation variationnelle, problème de complémentarité non linéaire, problème de point fixe, programme extrémal.

Pour préparer l'audit algorithmique, nous avons inventorié d'une part des critères de convergence selon le cadre mathématique auquel on rattache la formule caractéristique, d'autre part des algorithmes applicables au programme extrémal de Beckmann et souvent aptes à traiter d'autres modèles.

Pour préparer l'audit économétrique, nous avons distingué quatre objets d'estimation : fonctions de temps de parcours, flots origine-destination, fonctions de demande et paramètres comportementaux. Puis nous avons synthétisé diverses ressources : d'une part des modèles d'estimation pour chaque objet, d'autre part une procédure pour propager l'erreur exogène a priori via un modèle d'affectation. Nous avons aussi établi des ordres de grandeur exemplaires pour l'incertitude relative aux temps de parcours, aux flots origine-destination et aux valeurs du temps.

11.1.2.2 Enseignements de la capitalisation menée

Notons d'abord la simplicité de la méthode employée pour l'audit-capitalisation : distinction entre les dimensions, puis, dans chaque dimension, inventaire de ressources, comparaison entre ressources alternatives et sélection des ressources les meilleures ou les plus pertinentes ou les plus efficaces. Cette simplicité autorise l'adaptation à tout domaine de modélisation.

Remarquons aussi que la distinction suivant les dimensions permet de limiter une entreprise d'audit à l'une des dimensions, ou de répartir la tâche entre plusieurs intervenants selon leur spécialité.

Enfin mieux vaut ne pas considérer que la capitalisation est un sous-produit de l'audit. La capitalisation d'un domaine est elle-même un objectif suffisant, dont l'accomplissement facilite l'audit de modèles particuliers. La capitalisation d'un domaine correspond à un coût fixe, tandis qu'un audit particulier correspond à un coût marginal relativement modique.

11.1.3 Etat de développement du modèle bicritère

Le modèle bicritère nous a servi d'exemple pour appliquer notre méthode d'audit. Nous récapitulons ici les principaux éléments établis pour ce modèle au long de la thèse.

11.1.3.1 Contenu sémantique

Le modèle bicritère sert à répartir le trafic entre des itinéraires concurrents qui se différencient en termes de prix et de temps. Côté offre de circulation, les chemins sont bien identifiés, chacun d'eux est décrit par un temps et un prix, le temps de parcours d'une route peut dépendre du trafic local ainsi que des trafics d'autres routes, on peut contraindre les capacités d'écoulement local, et différencier des types de véhicules. Côté demande de déplacement, l'information et la rationalité des auteurs des déplacements sont supposées totales, ces auteurs sont différenciés non seulement par leurs zones d'origine et de destination mais encore par leur valeur du temps et éventuellement par d'autres critères à valeurs discrètes, le volume de demande peut dépendre du niveau de service. Pour cette dernière dépendance, l'élasticité de la demande, nous avons offert deux possibilités : soit une élasticité désagrégée par valeur du temps (Cf. section 8.3), soit une élasticité agrégée sur toutes les valeurs du temps (Cf. section 8.2).

11.1.3.2 Aspects formels

Le modèle bicritère se formule en inéquation variationnelle ou en problème de complémentarité non linéaire, ainsi qu'en programme extrémal lorsque les fonctions de temps de parcours dérivent d'une fonction de potentiel (en particulier quand le temps local dépend uniquement du trafic local). Cela vaut tant pour le cas à demande élastique désagrégée (qui justifie un traitement en dimension infinie) que dans le cas à demande élastique agrégée (pour lequel un traitement en dimension finie est possible, grâce aux notions intermédiaires que sont les classes monétaires de chemins et les valeurs du temps de coupure demandées).

L'existence théorique d'une solution est subordonnée à certaines conditions techniques sur la régularité des fonctions élémentaires, et surtout à la possibilité de satisfaire les contraintes adjointes (il faut qu'il existe un point admissible).

L'unicité théorique d'une solution est subordonnée à des conditions techniques sur les fonctions du flot qui interviennent dans les formules caractéristiques. Avec la formulation caractéristique en inéquation variationnelle, la continuité et la stricte monotonie de la fonction des temps généralisés par chemins suffisent pour une unicité des flots par arcs, et des flots par origine-destination, même des flots par origine-destination et classe monétaire de chemins pour la demande élastique agrégée. Ce dernier résultat d'unicité distingue le modèle bicritère à demande élastique agrégée du modèle de Beckmann et constitue une propriété mathématique remarquable.

11.1.3.3 Résolution algorithmique

Nous avons présenté trois algorithmes pour résoudre le modèle bicritère à demande élastique agrégée : une méthode des moyennes successives (MSA) sans explicitation des chemins, une MSA avec explicitation des chemins (également applicable au cas de la demande élastique désagrégée), et une méthode d'égalisation par transvasement (MET) qui explicite aussi les chemins. La MSA sans explicitation offre la plus simple implantation, tandis que la MET jouit d'une interprétation très intuitive et paraît efficace pour une application de taille moyenne.

Au sujet des critères de convergence, nos méthodes avec explicitation progressive des chemins s'appuient sur la valeur de la fonction objectif, ce qui n'est pas rigoureux. Nous avons proposé un succédané pratique, mais il manque un critère rigoureux dans le cas de la demande élastique¹.

11.1.3.4 Aspects économétriques

Nous avons développé une méthode pour propager l'incertitude exogène via le modèle bicritère, afin de quantifier l'incertitude exogène a posteriori liée au résultat d'une simulation. L'incertitude exogène a priori se répartit entre quatre objets d'estimation : temps de parcours, flots origine-destination, fonctions de demande et paramètres comportementaux comme la valeur du temps. Pour estimer une distribution continue de la valeur du temps, nous avons élaboré une méthode basée sur un modèle de choix discret et la technique du maximum de vraisemblance.

11.1.3.5 Bilan

Au total, le modèle bicritère ainsi composé et outillé constitue un modèle de simulation complet, au même titre que le modèle d'affectation classique de Beckmann qu'il étend et dépasse par la portée explicative (différenciation de la valeur du temps), par les propriétés mathématiques de la formule caractéristique (unicité des flots par classes monétaires de chemins) et par son outillage économétrique (réduction de la variabilité résiduelle liée à la valeur du temps), cela à un prix compétitif dans un problème de dimension moyenne.

¹ dans le cas à demande fixe, Marcotte et Zhu (1994) montrent que l'on dispose de l'écart de dualité

11.2 L'AUDIT EN PRATIQUE

Nous avons traité de l'audit d'un modèle pour répondre à des interrogations sur la validité technique du résultat. Notre réponse est qu'un modèle possède une portée phénoménologique explicitable, et que, dans la limite de cette portée, sous réserve d'une formule caractéristique et d'une résolution correctes, le résultat est une variable aléatoire dont on peut caractériser la distribution statistique. Nous distinguons les applications d'un modèle audité favorablement, certifié (§11.2.1) de celles liées à la capitalisation scientifique (§11.2.2).

11.2.1 Vers un modèle certifié ?

Nous qualifions de "certifié" un modèle dont l'audit conclut favorablement : l'audit porte d'abord in jugement qui, sil est négatif, disqualifie le modèle et devrait conduire à l'abandonner ou à le corriger des défauts relevés dans l'audit. Si l'audit est favorable, il offre une garantie technique à l'exécutant et au commanditaire d'une étude.

La transparence technique d'un modèle audité se manifeste par la livraison non seulement d'un résultat moyen, mais encore d'une incertitude de résultat directement reliée aux incertitudes sur les hypothèses de simulation présentes dans la composition phénoménologique. Les hypothèses absentes ne peuvent être examinées, mais l'audit conceptuel prévient l'exécutant et le commanditaire de leur absence.

Dans ces conditions le commanditaire d'une simulation peut analyser le risque sur le résultat. Dans les limites de l'audit économétrique, il décèle les raisons de ce risque en fonction des hypothèses et il évalue le rôle propre de chaque raison. Il peut commander de diminuer le risque lié au résultat en affinant la simulation : par réduction de l'erreur exogène a priori donc en modifiant les modèles d'estimation qui servent à l'alimentation économétrique, ou par réduction des coefficients de propagation donc en changeant la composition conceptuelle du modèle de simulation.

Pour un chargé d'étude libre de décider le modèle, la disposition de modèles audités permet de choisir en fonction d'arguments techniques : portée explicative et correction de la représentation conceptuelle, exactitude formelle et algorithmique, validité de l'alimentation économétrique.

11.2.2 Fonctions de la capitalisation

Rappelons ici quels services une capitalisation peut rendre, afin de montrer l'importance qu'elle revêt indépendamment de tout objectif d'audit. Il s'agit d'abord de clarifier et synthétiser un fonds commun de modélisation dans un domaine. Cela aide à fabriquer (puis à auditer) un modèle particulier, considéré comme un assemblage de composants standards et peut-être de composants originaux. Le fonds commun de modélisation apparaît alors comme une *boîte à composants*, à outils, qui offre de multiples possibilités.

La présentation des ressources sous la forme d'une boîte à composants facilite la compréhension selon une démarche d'analyse puis de synthèse : décomposer aide à comprendre, donc à assimiler. La capitalisation soutient l'information technique, l'enseignement et l'apprentissage d'une matière.

Autre usage important, fournir un support au débat scientifique et à la discussion technique. La capitalisation met en forme des éléments divers à l'aide d'un *vocabulaire commun*, qui facilite les échanges et les rend certainement plus fructueux. Cela permet d'énoncer clairement des questions clefs, en particulier pour remettre en cause certaines idées établies : car souvent des préjugés techniques compromettent la progression scientifique, par exemple la non explicitation des chemins dans un modèle d'affectation du trafic¹. Cela permet aussi de renvoyer des questions formalisées à d'autres disciplines, par exemple la question de la non unicité des solutions à la formule caractéristique relève de l'optimisation numérique plus que de l'affectation du trafic.

Enfin la capitalisation facilite la comparaison entre modèles, puisqu'elle permet d'identifier les éléments communs ou au contraire spécifiques. Elle prépare l'arbitrage entre les modèles, qui relève lui de l'audit (Cf.section 4.4).

11.3 PISTES DE RECHERCHE

Notre entreprise d'audit reste imparfaite et incomplète. Indiquons donc des approfondissements et des prolongements, autant pour la méthode d'audit (§11.3.1) que pour les modèles de choix de transport pour des flots spatialisés, dont surtout l'affectation (§11.3.2).

11.3.1 Du côté de la méthode d'audit

11.3.1.1 Approfondir ; la question du codage

Nous avons fourni des concepts et des termes à discuter, modifier, adapter. Nous les avons présentés de façon générale, en nous référant au domaine de l'affectation à titre d'exemple et non de support obligé. Cela devrait faciliter le débat méthodologique. Parmi les sujets à débattre, mentionnons :

- la capitalisation de la connaissance théorique en formalisant des mécanismes explicatifs qui sont des composants conceptuels d'un modèle.
- la mise en évidence des dimensions formelle et algorithmique.
- la distinction entre la dimension conceptuelle et la dimension économétrique, plutôt que la position dans un cadre économétrique avec la notion de "jugement a priori" comme issue de secours.
- le pied d'égalité entre les paramètres et les fonctions dans la spécification numérique de la formule caractéristique, plutôt que le choix a priori d'une spécification fonctionnelle suivi d'une estimation de paramètres².

¹Encore s'agit-il là d'un préjugé technique relativement facile à mettre en doute. On constate parfois l'existence de verrous techniques plus complexes, lorsque plusieurs difficultés se liguent. Ainsi, pour le modèle logit linéaire, la conjonction du manque d'un critère de convergence rigoureux et de l'instabilité de l'ensemble des chemins considérés.

²Il s'agit là d'une position chère à Olivier Morellet. Le vocabulaire que nous avons introduit considère de façon identique les paramètres et les fonctions, mais dans l'estimation des objets d'estimation de l'affectation nous avons suivi le séquençage économétrique classique. En termes d'erreur économétrique, et plus spécialement d'analyse de la variance, l'incertitude relative aux paramètres est comprise dans l'erreur d'estimation, tandis que

La question du codage nécessite un examen spécial ; elle renvoie à des opérations de discrétisation et d'agrégation qui déterminent le **format économétrique** du modèle, sous des contraintes de compatibilité avec la spécification conceptuelle.

11.3.1.2 Auditer d'autres domaines pour tester la robustesse des concepts

Pour valider ou adapter le vocabulaire et la méthode d'audit définis dans cette thèse, des essais de transposition à d'autres domaines de modélisation que l'affectation du trafic seraient utiles. A priori une transposition devrait s'effectuer sans trop de heurts, car nous avons considéré non seulement l'affectation mais encore l'économétrie, qui constitue le noyau théorique commun de multiples modèles dans des disciplines variées.

11.3.2 Capitaliser et développer l'affectation dynamique et aux modes non automobiles

Nous avons restreint notre application de la certification à l'affectation statique du trafic routier, en ignorant volontairement les aspects dynamiques et les spécificités des autres modes de transport. Avant d'auditer un modèle d'affectation dynamique ou à d'affectation à un mode d'autre que l'automobile, il convient de capitaliser les développements déjà versés. Nous allons planter les jalons d'une telle capitalisation, en suivant la distinction entre les dimensions d'audit et en indiquant les points qui nécessitent des contributions nouvelles. Notons l'intérêt d'une trame d'ensemble, dans laquelle chaque contribution ponctuelle prend son sens.

11.3.2.1 Conception

Par modèle dynamique d'affectation, on désigne un modèle dans lequel le choix d'itinéraire s'effectue au cours du déplacement, ou interagit avec le choix d'horaire. Par mode non automobile, nous désignons un mode de transport qui peut éventuellement emprunter le réseau routier mais pas dans un véhicule automobile à vocation individuelle tel qu'une voiture ou un poids lourd.

Il existe de nombreux développements à ces sujets : il conviendrait de les capitaliser pour mettre en évidence les mécanismes explicatifs propres au caractère dynamique, à la fois dans le fonctionnement de l'offre et dans le comportement de la demande, ou propres aux modes non automobiles, comme le concept de stratégie pour cheminer sur un réseau de transport en commun (Spiess 1985).

Parmi les autres sujets relatifs à la conception de l'affectation, mentionnons la caractérisation d'ensembles de chemins (définir d'autres ensembles que celui de tous les chemins sans boucle, ou que ceux liés au parcours d'une arborescence), et la segmentation de la demande en plusieurs classes interagissant via le fonctionnement de l'offre.

11.3.2.2 Aspects formels

L'inventaire des cadres mathématiques de référence, présenté pour le cas statique et routier, est exhaustif par rapport à notre connaissance des modèles de choix d'itinéraire

l'incertitude relative aux fonctions est comprise dans l'erreur résiduelle. Considérant l'erreur économétrique globale, ce débat revient à une question d'interprétation, non de quantification ; à condition de définir convenablement l'erreur économétrique globale...

sur un réseau. Pour compléter la capitalisation formelle, les éléments à verser sont des transformations de modèles conceptuels différents en formules caractéristiques, afin d'identifier les notions mathématiques originales employées comme intermédiaires dans les transformations, comme les classes monétaires de chemins dans l'affectation bicritère, ou les stratégies dans l'affectation aux transports collectifs.

11.3.2.3 Algorithmes

Les critères de convergence présentés suffisent par rapport aux cadres mathématiques de référence pour l'affectation. La capitalisation algorithmique se réduit donc aux algorithmes. Du côté des modes non automobiles, l'originalité tient aux algorithmes de sélection d'une stratégie optimale. Du côté des aspects dynamiques, des heuristiques cohabitent avec des applications des mêmes algorithmes que dans le cas statique. Il conviendrait de réaliser une comparaison de ces algorithmes, en termes tant de fonctionnement théorique que de performances sur des exemples standards.

11.3.3.4 Aspects économétriques

La dimension économétrique offre une foule de possibilités qui sont encore très peu exploitées. Elles se répartissent en trois lots.

D'abord quantifier l'erreur exogène liée à des objets d'estimation, et en particulier de l'erreur d'observation liée à une procédure d'enquête. Il faudrait ainsi examiner les comptages usuels des différents modes, les enquêtes de déplacement auprès des ménages, les enquêtes avant et après modification de l'offre de transport...

Puis capitaliser les modèles d'estimation pour les différents objets d'estimation. Le manque est patent pour les fonctions de demande, ainsi que pour les fonctions de décalage d'horaire de déplacement. Au sujet des flots origine-destination, il conviendrait d'inventorier et de synthétiser les multiples modèles d'estimation, et de réaliser un banc d'essai comparatif. Concernant les fonctions de temps de parcours, une synthèse reste à établir, tant pour la route en fonctionnement dynamique, que pour les autres modes, en particulier pour les transports collectifs et les correspondances entre lignes (en distinguant le cas des voyageurs du cas des marchandises). A propos des paramètres du comportement, il serait utile d'étendre la méthode d'estimation d'une valeur du temps distribuée continûment afin de considérer que les attributs des chemins sont des variables aléatoires, comme dans le modèle probit.

Enfin il faudrait définir un cadre statistique intégrateur pour l'estimation d'un modèle d'affectation, en rassemblant les objets d'estimation et les observations disponibles. A priori la ressource statistique existe, c'est la technique des équations simultanées. Un tel cadre intégrateur, muni d'une interprétation bayésienne, serait particulièrement utile pour un modèle dynamique à évolution structurelle de l'offre ou de la demande, afin de mettre en cohérence des observations menées à des époques différentes.

11.3.3.5 Capitaliser les modèles de choix de transport pour des flots spatialisés

L'affectation à un réseau modal conclut le schéma classique d'étude de la demande de transport, auquel participent traditionnellement des modèles de génération, de distribution spatiale, de choix modal, et plus sporadiquement des modèles de choix d'horaire, de choix de fréquence, de motorisation... Ces modèles partagent la représentation d'un choix lié au déplacement, la dimension spatiale car le déplacement va

d'une origine à une destination, l'agrégation des déplacements en flots même si le choix est attribué à un individu représentatif.

Ces modèles sont largement employés et nécessitent un audit autant que l'affectation. La priorité donnée à l'affectation se justifie car la plupart d'entre eux utilisent des fonctions liées à l'affectation, notamment pour calculer les niveaux de service procurés par un réseau modal entre les origines et les destinations. Pour préparer un audit, une capitalisation pourrait procéder selon :

- une synthèse par type de choix de transport.
- un inventaire des "concepts" communs entre divers types de choix et des éléments échangés entre modèles simulant des choix de différents types.
- un examen des modèles à structure plurielle qui combinent plusieurs types de choix, suivi d'une mise au point quant aux possibilités de combinaison dans une structure plurielle.

En conclusion

CONTRIBUTION DE LA THÈSE

Des concepts et une méthode pour décrire et auditer un modèle

La thèse apporte d'abord un vocabulaire pour analyser un modèle, avec les notions de composition sémantique, de formule caractéristique, de dispositif de résolution et d'alimentation économétrique. Ce vocabulaire permet d'explicitier les composants liés aux notions, afin de préciser leurs fonctions techniques. La composition sémantique est un ensemble d'objets et de mécanismes élémentaires qui mettent en relation les objets. La formule caractéristique est une synthèse mathématique des composants sémantiques. Le dispositif de résolution produit une solution à la formule caractéristique. L'alimentation économétrique fixe les valeurs des variables exogènes et la spécification numérique de la formule caractéristique.

Notre vocabulaire sert aussi à qualifier l'erreur de modélisation selon sa nature. Nous décomposons l'erreur en erreur de conception, erreur formelle, erreur algorithmique et erreur économétrique qui englobe l'erreur exogène a priori sur les variables économétriques et l'erreur d'estimation de la formule caractéristique.

Pour examiner la validité technique du modèle, nous avons proposé une méthode d'audit qui reprend la distinction entre conception, formulation, résolution et alimentation. L'audit conceptuel détermine la portée explicative du modèle en fonction des phénomènes qu'il représente ou non par ses mécanismes explicatifs. L'audit formel contrôle la formule caractéristique. L'audit algorithmique, dans le cas d'une résolution par simulation numérique, examine par un test de convergence rigoureux si l'algorithme livre une solution à la formule caractéristique. L'audit économétrique, conditionnel aux autres dimensions, attache au résultat de la simulation une incertitude obtenue en propageant via le modèle une erreur exogène a priori relative aux variables exogènes (élargies aux paramètres).

Dans le domaine de l'affectation

Pour tester notre méthode d'audit, nous l'avons appliquée à un modèle de choix d'itinéraire particulier, le modèle bicritère. Nous avons préparé l'application en

capitalisant le domaine de l'affectation statique du trafic routier : à savoir

- i) en synthétisant des mécanismes explicatifs (dimension conceptuelle),
- ii) en inventoriant des cadres mathématiques de référence auxquels se rattache une formule caractéristique (dimension formelle),
- iii) en recensant des critères de convergence et des algorithmes (dimension algorithmique),
- iv) en présentant des ressources pour quantifier l'erreur exogène a priori et la propager à travers le modèle de simulation (dimension économétrique).

Le modèle bicritère représente des arbitrages différenciés entre deux attributs des chemins parmi les auteurs des déplacements, ce qui influe sur la répartition du trafic entre itinéraires concurrents. Nous l'avons simultanément développé et audité dans chaque dimension. Du côté de la conception, nous considérons la dépendance du temps de parcours envers les flots des chemins, des contraintes adjointes, la dépendance du volume de demande envers le niveau de service (de façon agrégée ou désagrégée). Du côté de la formulation, nous transformons les formules caractéristiques en inéquations variationnelles, et en programme extrémal lorsque les temps de parcours dérivent d'un potentiel. Du côté de la résolution, nous indiquons deux algorithmes qui explicitent les chemins, respectivement une méthode des moyennes successives et une méthode d'égalisation par transvasement ; nous proposons aussi une méthode des moyennes successives sans explicitation des chemins. Du côté économétrique, nous donnons une méthode pour propager l'erreur exogène a priori et une méthode d'estimation de l'erreur exogène a priori relative à une distribution continue de la valeur du temps ; sur l'exemple d'un réseau à deux arcs, nous démontrons que *la sophistication du modèle réduit l'incertitude attachée au résultat*.

Certains apports valent pour d'autres modèles, par exemple la méthode d'égalisation pour le modèle logit linéaire. Cependant le domaine de l'affectation nécessite encore de multiples contributions, en particulier pour les modes non automobiles et les aspects dynamiques.

MODÉLISATION ET AUDIT

Quelles garanties techniques ?

Quelles certitudes ? Nous avons soutenu que l'audit d'un modèle s'étend dans plusieurs dimensions. L'examen de la portée explicative procède d'un jugement qualitatif, relatif à une représentation théorique du système étudié. Le contrôle des aspects formels et algorithmiques fait appel au jugement logique, basé sur les axiomes des mathématiques. La quantification de l'incertitude sur le résultat procède par une mise en forme économétrique des mécanismes explicatifs et des hypothèses sur les variables exogènes. Au total, les certitudes sur le résultat d'un modèle de simulation sont d'ordre économétrique ; elles demeurent relatives à la portée explicative, et conditionnelles à l'exactitude formelle et algorithmique.

Comment prendre des garanties ? Pour obtenir des certitudes économétriques, il faut quantifier l'incertitude économétrique sur le résultat, donc expliciter le fonctionnement statistique du modèle de simulation. On y parvient en caractérisant statistiquement

l'erreur exogène a priori et en la propageant à travers le modèle. Alors le modèle présente une transparence économétrique qui, jointe à une transparence conceptuelle, à une conformité et une cohérence formelles, et à une exactitude algorithmique, offre un fondement rigoureux pour un contrat de garantie entre le commanditaire de la simulation et l'exécutant.

Du bon usage de l'audit

De notre expérience de l'audit, nous dégagons quelques conseils pour une mise en oeuvre raisonnée.

Décomposer pour comprendre, traiter chaque poste d'erreur avec les méthodes dont il relève, et ne négliger aucun poste. L'audit d'un modèle de simulation nécessite l'audit dans chaque dimension.

Réfléchir avant d'agir, non seulement en organisant l'audit d'un modèle, mais encore en le préparant par une capitalisation du domaine. Une telle capitalisation permet des économies d'échelle s'il y a plusieurs modèles à certifier.

Poser clairement les problèmes non résolus : ne pas les contourner en commentant qu'ils sont difficiles, mais les mettre en évidence, les transcrire en questions formalisées à renvoyer aux disciplines scientifiques dont elles sont justiciables par nature.

Pour un modèle alimenté, quantifier l'erreur liée au résultat, ne serait-ce que dans un cas d'école. La position économétrique permet de comparer les modèles, d'examiner si une sophistication sémantique réduit la variabilité résiduelle donc l'incertitude propagée sur le résultat.

Modéliser, avec ou sans garantie...

L'audit permet de mesurer la confiance à accorder au résultat d'une simulation. Il repose fondamentalement sur certaines hypothèses, tout comme la modélisation : hypothèses de représentation théorique, arguments logiques et probabilistes qui découlent d'axiomes mathématiques, enfin des supputations d'ordre finalement subjectif (notamment pour simuler une situation virtuelle).

L'audit nécessite un état d'esprit, un effort continu de lucidité afin de rester conscient des hypothèses et d'étudier la façon dont elles se combinent. Cet esprit de l'audit est identique à l'esprit de la modélisation, qui formalise des hypothèses et simule leurs effets conjoints afin de soutenir la représentation et l'imagination de l'utilisateur du modèle.

Références

- Abraham, C. (1961). La répartition du trafic entre itinéraires concurrents. *Revue Générale des Routes et Aéroports*, Paris, June.
- Abraham, C., & Blanchet, J.D. (1973). Le modèle prix-temps. *Revue de l'Aviation Civile*, Paris.
- Aho., Hopcroft, J.E., & Ullman, J.D. (1974). *The design and analysis of computer algorithms*. Addison Wesley, Reading, Mass.
- Akaike, H. (1973) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19.
- Amemiya, T. (1981). Qualitative Response Models: A Survey. *Journal of Economic Literature*, **19**, 1483-1536.
- Antonisse, R.W., Daly, A.J., & Ben-Akiva, M. (1989). *A Highway Method Based on Behavioral Models of Car Drivers' Route Choice*. Working Paper. Hague Consulting Group, La Haye, Pays-Bas.
- ASF (1990). *Données de trafic sur les autoroutes A7 et A9*. Communication privée.
- Ashley, D.J. (1981). Uncertainty in Interurban Highway Scheme Appraisal, in Stopher, P. R., et al. (eds) *New Horizons in Travel Behavior Research*, Lexington Books, Lexington, 599-615.
- Barbier Saint-Hilaire, F. (1988 à 1996). *Système DAVIS-EQUILIBRE*. Notes internes, INRETS, Arcueil.
- Beckmann, M., McGuire, C.B., & Winsten, C.B. (1956). *Studies in the Economics of Transportation*. Yale University Press, New Haven, Connecticut.
- Beli, M.G.H. (1983). The estimation of an origin-destination matrix from traffic counts. *Transportation Science*, **17/2**, 198-217.
- Bell, M.G.H. (1991). Variances and covariances for origin-destination flows when estimated by log-linear models. *Transportation Research B*, **19/6**, 497-507.
- Beli, M.G.H. (1991). The estimation of OD matrices by constrained generalized least squares. *Transportation Research B*, **25/1**, 13-22.
- Ben-Akiva, M. (1973). *Structure of Passenger Travel Demand Models*. PhD thesis, MIT, Cambridge.
- Ben-Akiva, M., Bergman, M.J., Daly, A.J., & Ramaswamy, R. (1984). Modelling Inter Urban Route Choice Behaviour. *Proceedings of the 9th ISTTT*, VNU Science Press.
- Ben-Akiva, M.E., & Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, MA.
- Bierlaire, M. (1991). *Evaluation de la demande de trafic : quelques méthodes de distribution*. Rapport 91/12, Département de Mathématiques, Facultés Universitaires de Namur, Belgique.
- Blum, J.R. (1954). Multidimensional Stochastic Approximation Methods. *Ann. Math. Stat.* **25**, 734-744.

- Bolduc, D., Dagenais, M.G., & Gaudry, M.J.I. (1989). Spatially autocorrelated errors in origin-destination models: a new specification applied to aggregate mode choice. *Transportation Research B*, **23**, 361-372.
- Bonsall, P.W. (1991). *Feasibility of measuring responses to highway improvements*. TRRL Contractor Report 200. Department of Transport, Crowthorne, England.
- Bourbonnais, R. (1993). *Econométrie : cours et exercices corrigés*. Dunod, Paris.
- Bovy, P.H.L., & Stern, E. (1990). *Route Choice: Wayfinding in Transport Networks*. Kluwer Academic Publishers. Dordrecht, Pays-Bas.
- Bradley, M., Ben-Akiva, M. & Bolduc, D. (1992). Estimation of travel choice models with randomly distributed values of time. *Proceedings of the PTRC Summer Meeting*, **E**, 291-302.
- Bruynooghe, M. (1967). *Affectation du trafic sur un multi-réseau*. Rapport technique, Institut de Recherche sur les Transports, Arcueil, France.
- Bruynooghe, M., Gibert, A. & Sakarovitch, M. (1968). *Une méthode d'affectation du trafic*. Article présenté au 4ème Symposium International sur les Théories de la Circulation Routière, Karlsruhe.
- Burrell, J.E. (1968). *Multiple Route Assignment and its Application to Capacity Restraint*. 4th International Symposium Theory of Traffic Flow, Karlsruhe.
- Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research B*, **18/4-5**, 289-299.
- Cascetta, E., & Nguyen, S. (1988). A unified framework for estimating or updating origin/destination matrices from traffic counts. *Transportation Research B*, **22/6**, 437-455.
- Casey, H.J. (1955). Applications to traffic engineering of the law of retail gravitation. *Traffic quarterly*, **9/1**, 23-35.
- CETE de l'Ouest (1993). *Etude de fonctions A11/RN23 entre Nantes et Angers - Analyse des trafics*. 49 p + illustrations.
- CETE du Sud-Ouest (1993). *Liaison routière Bordeaux-Pau, Etudes Préliminaires*. 90 p.
- Chapleau, R. (1993). *MADITUC, éléments partiels, voire représentatifs, de la démarche totalement désagrégée*. Ecole Polytechnique de Montréal, Montréal.
- Cofiroute (Falaise + Delons sur prix-temps)
- Cohen, G. (1994). *Modélisation des réseaux urbains*. Ecole des Mines de Paris, Fontainebleau.
- Cohen, S. (1990). *Ingénierie du Trafic Routier*. Presses de l'ENPC, Paris.
- Cousot, P. (1991). *Cours d'informatique de l'Ecole Polytechnique*. Ecole Polytechnique, Palaiseau.
- Culioli, J.C. (1994). *Introduction à l'Optimisation*. Ellipses, Paris.
- Curien, N., & Gensollen, M. (1989). *Prévision de la demande de télécommunication*. Eyrolles, Paris.
- Dafermos, S.C. (1971). An Extended Traffic Assignment Model with applications to two-ways traffic. *Transportation Science*, **5**, 366-389.
- Dafermos, S.C. (1972). The Traffic Assignment Problem for Multiclass-User Transportation Networks. *Transportation Science*, **6**, 73-87.
- Dafermos, S.C. (1980). Traffic Equilibrium and Variational Inequalities. *Transportation Science*, **14**, 42-54.
- Dafermos, S.C. (1982). The General Multi-Modal Network Equilibrium Problem with Elastic Demand. *Networks*, **12**, 57-72.
- Dafermos, S.C., & Sparrow, F.T. (1969). The Traffic Assignment Problem for a General Network. *J. Res. Natl. Bur. Stand.*, **73/B**, 91-118.
- Daganzo, C.F. (1979). *Multinomial Probit: the Theory and its Application to Demand Forecasting*. Academic Press, New York.
- Daganzo, C.F. (1983). Stochastic Network Equilibrium with Multiple Vehicle Types and Asymmetric, Indefinite Link Cost Jacobians. *Transportation Science*, **17/3**, 282-300.
- Daganzo, C.F., & Sheffi, Y. (1977). On Stochastic Models of Traffic Assignment. *Transportation Science*, **11/3**, 253-274.

- Daito, T., Okubo T., Moritsu H. and Morikawa T. (1992). *Practical Traffic Assignment for Multiple Highway Routes using Distribution of Values of Time*. Paper presented at the Sixth WCTR, Lyon, France.
- Dalvi, Q. (1978). *Economic Theories of Travel Choices*. In Hensher, D.A., & Dalvi, Q. (eds), *Determinants of Travel Choice*. Saxon House, Farnborough, England.
- Daly, A.J. (1987). Estimating "tree" logit models. *Transportation Research B*, **21/4**, 251-268.
- Delons, J. (1991). *Estimation automatique d'un modèle de trafic*. Communication privée.
- Dial, R.B. (1971). A Probabilistic Multipath Traffic Assignment Model which Obviates Path Enumeration. *Transportation Research* **5**, 83-111.
- DOT 1985. Department of Transport (1985). *Traffic Appraisal Manual*.
- DR (Direction des Routes) (1986). *Directive relative à l'évaluation des avants-projets routiers*. Ministère des Transports, Paris.
- Drissi-Kaïtouni, O. (1993). A variational inequality formulation of the dynamic traffic assignment problem. *Eur. J. of Oper. Res.*, **71**, 188-204.
- Dupuy, G. (1975). *Une technique de planification au service de l'automobile : les modèles de trafic urbain*. Copedith, Paris.
- Easa, S.M. (1985). Shortest Route Algorithms with Movement Prohibitions. *Transportation Research B*, **19/3**, 197-208.
- Etner, F. (1987). *Histoire du calcul économique en France*. Economica, Paris.
- Evans, S.P. (1973). A relationship between the gravity model for trip distribution and the transportation problem in linear programming. *Transportation Research*, **7**, 39-61.
- Evans, S.P. (1976). Derivation and Analysis of Some Models for Combining Trip Distribution and Assignment. *Transportation Research* **10/1**, 37-57.
- Fisk, C. (1977). Note on the maximum likelihood calibration of Dial's assignment method. *Transportation Research*, **11**, 67-68.
- Fisk, C. (1980). Some developments in equilibrium traffic assignment methodology. *Transportation Research B*, **14**, 243-256.
- Florian, M. (1977). A Traffic Equilibrium Model of Travel by Car and Public Transport. *Transportation Science*, **11/2**, 166-179.
- Florian, M., & Nguyen, S. (1976). An application and validation of equilibrium trip assignment methods. *Transportation Science*, **10**, 374-390.
- Florian, M., & Spiess, H. (1982). The convergence of diagonalisation algorithms for asymmetric network equilibrium problems. *Transportation Research B*, **16/6**, 477-484.
- Friesz, T.L. (1985). Transportation Network Equilibrium, Design and Aggregation: key developments and research opportunities. *Transportation Research A*, **19** (5-6), 413-427.
- Fukushima, M. (1984). On the Dual Approach to the Traffic Assignment Problem. *Transportation Research B* **18/3**, 235-245.
- Furness, K.P. (1965). Time function iteration. *Traffic Engineering and Control*, **7/7**, 458-460.
- Gartner, N.H. (1980). Optimal Traffic Assignment with Elastic Demands: A Review, Part 1. *Transportation Science*, **14**, 174-191.
- Gaudry, M.J.I., & Wills, M.I. (1978). Estimating the functional form of travel demand models. *Transportation Research*, **12/4**, 257-289.
- Goodwin, P.B. (1992). A Review of New Demand Elasticities with Special Reference to Short and Long Run Effects of Price Changes. *Journal of Transport Economics and Policy*, **26/2**, 155-170.
- GTM-LASER (1988). *Relation entre la Vitesse Moyenne et le Trafic supporté par un réseau régulé*. Cellule Laser 8, GTM-Entrepose, Nanterre.
- Gunn, H.F., & Bates, J.J. (1982). Statistical aspects of travel demand modeling. *Transportation Research A*, **16** (5/6), 371-382.
- Hachette (1990). *Le dictionnaire du français*. Hachette, Paris.

- Harker P.T. (1985). The state of the art in the predictive analysis of freight transport systems. *Transport Reviews*, 5/2, 143-164.
- Harker, P.T., & Pang, J.S. (1990). Finite-dimensional variational inequality and non linear complementarity problems: a survey of theory, algorithms and applications. *Mathematical Programming*, 48, 161-220.
- Hartgen, D.T. (1981). Uncertainty in the Application of Travel Behavior Models, in Stopher, P. R. & al. (eds) *New Horizons in Travel Behavior Research*, Lexington Books, Lexington, 617-620.
- Heggie, I.G. (1978). *Behavioural Dimensions of Travel Choice*. In Hensher, D.A. & Dalvi, Q. (eds), *Determinants of Travel Choice*. Saxon House, Farnborough, England.
- Horowitz, A.J. (1985). Convergence Properties of Some Iterative Traffic Assignment Algorithms. *TRR* 1220, 21-27.
- Horowitz, J.L. (1981). Sources of Error and Uncertainty in Behavioral Travel-Demand Models, in Stopher, P. R. & al. (eds), *New Horizons in Travel Behavior Research*, 543-558. Lexington Books, Lexington, MA.
- Horowitz, J. (1982). Specification tests for probabilistic choice models. *Transportation Research A*, 16 (5-6), 383-394.
- Janson, B.N., Thint, S.P.T., & Hendrickson, C.T. (1986). Validation and Use of Equilibrium Network Assignment for Urban Highway Reconstruction Planning. *Transportation Research A*, 20/1, 61-73.
- Jones, P., Koppelman, F.S., & Orfeuil, J.P. (1990). Activity analysis: state-of-the-art and future directions. In Jones, P. (ed), *Developments in dynamic and activity-based approaches to travel analysis*. Gower, Avebury, England.
- Judge, G.G., Hill, R.C., Griffith, W.E., Lütkepohl, H., & Lee, T.C. (1988). *Introduction to the Theory and Practice of Econometrics*. Wiley, New York.
- Knuth, D. (1969-1973). *The art of computer programming*, Vol. 1-3. Addison-Wesley.
- Koppelman, F.S. (1976). Methodology for Analyzing Errors in Prediction With Disaggregate Choice Models. *Transportation Research Record* 592, 7-23.
- Lam, W.H.K. (1989). *Integrated Planning of Transport Network and Pricing Policies*. 5th WCTR, Yokohama, 15 p, juillet.
- Lancaster, K.J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74, 132-157.
- Larsson, T., & Patriksson, M. (1992). Simplicial Decomposition with Disaggregate Representation for the Traffic Assignment Problem. *Transportation Science* 26, 4-17.
- Larsson, T. & Patriksson M. (1994). An augmented Lagrangean Scheme for Capacitated Traffic Assignment Problems. In Boilot, F. & al., *Proceedings of the 2nd meeting of the EURO Working Group on Urban Traffic and Transportation, Actes INRETS*, 38, 163-199. INRETS, Arcueil.
- LeBlanc, L.J., & Abdulaal, M. (1982). Combined Mode Split-Assignment and Distribution-Modal Split-Assignment Models with Multiple Groups of Travelers. *Transportation Science*, 16/4, 430-442.
- Leurent, F. (1991a). *Notes sur l'utilisation des modèles statiques d'affectation du trafic routier urbain*. Mémoire de DEA Transport de l'Ecole Nationale des Ponts et Chaussées. ENPC, Paris.
- Leurent, F. (1991b). *Equilibre Multi-Critères sur un Marché*. Note de Travail 91/2. Document interne INRETS. INRETS. Arcueil.
- Leurent, F. (1991c). *Affectation prix-temps sur un réseau : formulation extrémale et algorithme pour un cas simple*. Note de Travail 91/3. Document interne INRETS. INRETS. Arcueil.
- Leurent, F. (1992a). *Présentation des méthodes de Préférences Déclarées*. Note de Travail 92.1. Document technique INRETS. INRETS, Arcueil.
- Leurent, F. (1992b). *Equilibre Offre-Demande et Répartition entre Itinéraires dans les Modèles Statiques d'Affectation du Trafic*. Note de Travail 92.2. Document technique INRETS, Arcueil.
- Leurent, F. (1993a). Cost versus Time Equilibrium over a Network. *Eur. J. of Oper. Res.*, 71, 205-221.
- Leurent, F. (1993b). Modelling Elastic, Disaggregate Demand. In Papageorgiou M. (ed), *Proceedings of the First Meeting of the Euro Working Group on Urban Traffic and Transportation*. Technical

- University of Munich, Munich.
- Leurent, F. (1993c). *Etat de l'art des logiciels d'affectation opérationnels*. Note de Travail 93.4. Document technique INRETS, Arcueil.
- Leurent, F. (1994a). Cost versus Time Equilibrium over a Network. *Transportation Research Record*, **1443**, 84-91.
- Leurent F. (1994b). *Logit-based equilibrium assignment model with elastic demand, capacity constraints and a dual solution algorithm*. In Boillot, F, Bouhri, N., et Leurent, F. (eds) *Actes INRETS* Vol. **38**, pp. 125-162. INRETS, Arcueil.
- Leurent, F. (1994c). *Analyse de sensibilité et d'erreur d'un modèle - le cas de l'affectation prix-temps*. Working paper 94.1. INRETS, Arcueil, France.
- Leurent, F. (1995a). The practice of a dual criteria assignment model with continuously distributed values-of-time. *Proceedings of the 23rd European Transport Forum: Transportation Planning Methods*, **E**, 117-128. PTRC, London.
- Leurent, F. (1995b). Contributions to logit assignment model, *Transportation Research Record*, **1493**, 207-212.
- Leurent, F.M. (1995c). Un algorithme pour résoudre plusieurs modèles d'affectation du trafic : la méthode d'égalisation par transvasement. *Les Cahiers Scientifiques du Transport*, **30**, 31-49.
- Leurent, F. (1995d). *Une Boîte-à-Outils pour opérer diverses affectations statiques du Trafic Routier*. INRETS Research Report 198. INRETS, Arcueil, France.
- Leurent, F. (1995e). *Comparaison de deux principes d'affectation du trafic*. Rapport sur convention SETRA. INRETS, Arcueil.
- Leurent, F. (1995f). *Mode d'emploi du logiciel MARS*. Note de Travail 95.3. Document interne INRETS. INRETS, Arcueil.
- Leurent, F. (1996a). The Theory and Practice of a Dual Criteria Assignment Model with a Continuously Distributed Value-of-Time, in Lesort J.B. (ed), *Transportation and Traffic Theory*, 455-477. Pergamon, Exeter, England.
- Leurent, F. (1996b). An analysis of modeling error, with application to a traffic assignment model with continuously distributed values of time, *Proceedings of the 24th European Transport Forum: Transportation Planning Methods*, **D1**, PTRC, London.
- Leurent, F. (1996d). *Portée et Limites des modèles de trafic*. Rapport sur convention DRAST. INRETS, Arcueil.
- Leurent, F. (1996e). *Sensitivity and error analysis of the dual criteria traffic assignment model*. To be published in *Transportation Research B*.
- Leurent, F. (1996f). *Discrete Choice Models with varying coefficients in mode or route choice problems: the binary case*. INRETS technical report.
- Leurent, F. [1997]. Curbing the computational difficulty of the logit equilibrium assignment model. *Transportation Research B*, **31/4**, 315-326.
- Loos, A.H., Bovy, P.H.L., & Van der Hoorn, T.I.J.M. (1992). *The M10 Amsterdam Beltway: effects of opening upon travel behavior*. Paper presented at the 71st Annual Meeting of the TRB. TRB, Washington DC, 27 p.
- Madre, J.L. (1989). *Projection du trafic automobile sur les routes nationales et les autoroutes françaises*. Proceedings of the 5th WCTR, Yokohama, Japan.
- Maher, M.J. (1983). Inferences on Trip Matrices From Observations on Link Volumes: a Bayesian Statistical Approach. *Transportation Research B*, **17/6**, 435-447.
- Maher, M.J. & Hughes, P.C. (1995). *A probit-based stochastic user equilibrium assignment model*. Submitted to *Transportation Research*.
- Manheim, M.L. (1979). *Fundamentals of Transportation Systems Analysis*. MIT Press, Cambridge, Mass.
- Marche R. (1973). *Valuation of travel time, French inter-city values*. Paper presented at the Symposium on Transportation and Environment, Southampton, April.

- Marche R. (1988). *Notes de Travail LASER 88.1 à 88.7*. Documents internes INRETS. INRETS, Arcueil.
- Marche R. (1991). *Rapport d'expertise sur le Boulevard Périphérique Nord de Lyon*. Document interne INRETS. INRETS, Arcueil.
- Marcotte, P., & Zhu, D. (1994). An efficient algorithm for a bicriterion traffic assignment problem. *Preprints of the 3rd Capri Seminar*, 2. 891-897. Capri, June.
- Marcotte, P., & Zhu, D. (1996). On Continuous Multiclass Problems. To be published in the *Proceedings of a conference on Non-linear Complementarity Problems* held at the John Hopkins University.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In Zarembka (ed), *Frontiers in Econometrics*, 105-142. Academic Press, New York.
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3, 303-328.
- McFadden, D. (1978). Modeling the Choice of Residential Location. *Transportation Research Record* 673, 72-77.
- Morellet, O. (1980). *Un modèle de choix des usagers entre itinéraires concurrents*. Ministère des Transports, Paris.
- Morellet O. (1982). La modélisation de la demande de trafic de VP en milieu interurbain, in *Bilan de la modélisation de la demande*, CEMT, Paris, pp. 10-20.
- Morellet O., & Julien H. (1990). *MATISSE, un modèle intégrant étroitement induction et partage modal fin du trafic*. INRETS Research Report 129. INRETS, Arcueil, France.
- Morellet O., & Marchal, P. (1996). *MATISSE, un modèle intégrant étroitement induction et partage modal fin du trafic*. INRETS Research Report 203. INRETS, Arcueil, France.
- Murchland, J.D. (1966). *Some remarks on the gravity model of traffic distribution and on equivalent maximisation formulation*. LSE-TNT-38. Transport Network Theory Unit, London Graduate School of Business Studies.
- Ortuzar, J. de Dioz, & Willumsen, L.G. (1990). *Modeling Transport*. Wiley, Londres.
- Papon, F. (1988). *Analyse des valeurs du temps dans les Alpes Maritimes*. Notes de travail n°88.3, 88.4 et 88.8. INRETS, Arcueil.
- Papon, F. (1992). *Les routes de première classe*. Paradigme, Caen, France.
- Powell, W.B., & Sheffi, Y. (1982). The convergence of equilibrium algorithms with predetermined step sizes. *Transportation Science*, 16/1, 45-55.
- Patriksson, M. (1993). A unified description of some iterative algorithms for traffic equilibria. *Eur. J. of Oper. Res.* 71/2, 154-176.
- Patriksson, M. (1994). *The Traffic Assignment Problem, Models and Methods*. VSP, Utrecht, Pays-Bas.
- Planco (1986). *Evaluation macro-économique des investissements dans l'infrastructure de transport*. Publications du ministre fédéral des transports (n°69). Bonn, Allemagne.
- Robillard, P. (1974). Calibration of Dial's Assignment Method. *Transportation Science* 8, 117-125.
- Robillard, P. (1975). Estimating the O-D matrix from observed link volumes. *Transportation Research*, 9, 123-128.
- Safwat, K., & Magnanti, T (1988). A combined trip generation, trip distribution, modal split and trip assignment model. *Transportation Science* 22, 14-30.
- Saporta, G. (1990). *Probabilités, Analyse des données et Statistique*. Technip, Paris.
- Schittenhelm, H. (1990). *On the integration of an effective assignment algorithm with path and path-flow management in a combined trip distribution and traffic assignment algorithm*. Proceedings of the 18th Summer Annual Meeting of the PTRC, H. PTRC, London.
- Schneider, M. (1959). Gravity models and trip distribution theory. *Papers and Proceedings of the Regional Science Association*, 5/1, 51-56.
- SETRA (1984). *Temps de parcours des véhicules sur itinéraire interurbain*. Note de synthèse. Bagnaux, France.

- Sheffi, Y. (1984). *Urban Transportation Networks*. Prentice Hall, Englewood Cliffs, New Jersey.
- Sheffi, Y., & Daganzo, C.F. (1978). Hypernetworks and Supply-Demand Equilibrium with Disaggregate Demand Models. *Transportation Research Record*, **673**, 113-121.
- Smith, M.J. (1979). The Existence, Stability and Uniqueness of Traffic Equilibria. *Transportation Research B*, **13**, 295-304.
- Spiess, H. (1985). *Contributions à la Théorie et aux Outils de Planification des Réseaux de Transport Urbain*. Publication du Centre de Recherche sur les Transports, Université de Montréal.
- Stopher, P.R., & Meyburg, A.H. (1975). *Urban Modelling and Planning*. Lexington Books, Lexington, MA.
- Stover, V.G., Benson, J.D., & Buechler, J. (197?). A sensitivity evaluation of the Traffic Assignment Process. *Transportation Research Record*, 50-52.
- Tagliacozzo, F., & Pirzio, F. (1973). Assignment models and urban path selection criteria: results of a survey of the behaviour of road users. *Transportation Research*, **7**, 313-329.
- Taylor, M.A.P. (1977). Parameter estimation and sensitivity of parameter values in a Flow-Rate/Travel-Time Relation. *Transportation Science*, **11/3**, 275-292.
- Taylor, M.A.P., & Young, W. (1988). *Traffic Analysis: new technology and new solutions*. Hargreen, Victoria, Australie.
- Thomas, R. (1991). *Traffic Assignment Techniques*. Avebury Technical, Aldershot, England.
- Tobin, R.L. (1986). Sensitivity Analysis for Variational Inequalities. *J. Optim. Theory Appl.* **48**, 191-204.
- Tobin, R.L., & Friesz, T.L. (1988). Sensitivity Analysis for Equilibrium Network Flow. *Transportation Science*, **23/4**, 242-250.
- TRB (1983). *Forecasting Inputs to Transportation Planning*. NCHRP Report 266. TRB, Washington DC.
- Tukey, J. W. (1957). The Propagation of Errors, Fluctuations and Tolerances: Basic Generalized Formulas. *Dpt of Mathematics, Technical Report 10*, Princeton University.
- USAP (1992). *Comportement des automobilistes en matière de choix d'itinéraires*. Paris, 73 p. + annexes.
- Van Zuylen, H.J., & Willumsen, L.G. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research B*, **14**, 281-293.
- Voorhees, A. (1955). In *Proceedings of the ITE Annual Meeting*.
- Walter, E., & Pronzato, L. (1994). *Estimation de modèles paramétriques*. Masson, Paris.
- Wardrop, J.G. (1952). Some Theoretical Aspects of Road Traffic Research. *Proc. Inst. Civil Engineers*, Part II, 325-378.
- Watling, D.P., & Maher, M.J. (1992). A statistical procedure for estimating a mean origin-destination matrix from a partial registration plate survey. *Transportation Research B*, **26/3**, 171-193.
- Williams, H.C.W.L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, **9**, 285-344.
- Wilson, A.G. (1967). A statistical theory of spatial distribution models. *Transportation Research* **1**, 253-269.
- Wonnacott, T.H., & Wonnacott, R.J. (1991). *Statistique. Economica*, Paris.
- Wyley, W.J., Hardcastle, D.J., & Fearon, E.J. (1981). Traffic data variability and its incorporation into model validation procedures. *Traffic Engineering and Control (August/September)*, 473-477.
- Wynter, L.M. (1994). La Valeur du temps de transport de fret en France : Estimation à partir d'une enquête sur les préférences déclarées. *RTS*, **45**, 29-38.

English Summary

AN ANALYSIS OF MODELING ERROR, WITH APPLICATION TO A TRAFFIC ASSIGNMENT MODEL WITH CONTINUOUSLY DISTRIBUTED VALUES OF TIME

INTRODUCTION

Background. A model is a representation of a system, which it assumes to be comprised of separate elements which behave in a certain way and which are linked together through relationships, specially of a causal nature. The model allows us to analyze and study the interactions between these elements. The model is used for qualitative or quantitative simulation: as it assumes that there is a causal linkage between elements, it is possible to deduce the "rational" consequences of hypotheses concerning the state variables of the system or control variables or perturbation variables. The nature of the causal linkage is determined by the model and its designer.

Is it possible to assess, or even better to measure, how well the model conforms to reality? In other terms can a model be certified - can we guarantee that if its hypotheses prove to be true then its result will closely approximate reality? This question becomes more and more vital the greater what is at stake in the studied system. For example in planning a transport system, the high cost of the major actions (like construction of new infrastructure) not only encourages the use of a model in order to assist decision making, but also a careful examination of its predictive quality.

Objective. This paper aims to describe and analyze the error of a model and to provide detection criteria and corrective treatments. The paper puts forward a systematic certification method which is valid for any model. The method is applied to a route choice model on a transport network, with a continuous distribution of user values of time.

The discrepancy between the model and the studied system can have several causes: it may be an error in design (semantic error), or in formulation, or in computation, or in estimation, or even an exogenous error which is propagated through the model. These different causes all require specific detection and treatments, it is therefore necessary for each of them to be clearly identified and defined.

Structure. The paper contains three sections. The first section distinguishes between the different causes of error and shows how they combine with each other. Next, we provide numerical details concerning the exogenous error in a traffic assignment model. This error could relate to origin-destination flows, travel time functions, or behavioural parameters such as the value of time. Lastly, we define a method for tracing the propagation of the exogenous error through a dual criteria assignment

model. We show, in the context of a realistic interurban road traffic study, how the exogenous errors contribute to the uncertainty in the results of the model.

1. ANALYSIS OF ERROR

1.1 Terminology for identifying error

As a schematic representation of elements and relationships, the model formalizes theoretical knowledge of the system. We shall use the term **conceptual content** to describe this formalized knowledge: it includes the elements and the relationships between them. The elements are organized into subsystems. The relationships (eg causality, simultaneity of occurrence) are **explanatory mechanisms** which fix the values of endogenous variables (= explained by the model) on the basis of exogenous variables (= fixed outside the model).

The formalization of the explanatory mechanisms gives rise to the **characteristic formula**. This can be obtained by specifying a mathematical notation for each element and then expressing each explanatory mechanism as an elementary relationship which links some elements (variables which are endogenous to the relationship) to others (variables which are exogenous to the relationship) and finally by synthesizing the elementary relationships.

Solving the characteristic formula involves a dedicated process, which we shall call a solver or an **algorithm** (we shall extend the meaning of this term to include the hardware and software support as well as the application team).

Lastly, quantitative application requires data concerning the studied case: we shall use the term **econometric supply** to describe the data and parameters and the specification of the mathematical functions involved in the characteristic formula.

In the light of these distinctions we can perceive five potential sources of error:

- **design** or **semantic** errors (**sem**) may occur in the representation of the studied system, in the definition and choice of explanatory variables (exogenous) and the variables which are explained (endogenous), in the linking of these elements, and thus in the identification and formulation of causal relationships.

- **formulation** errors (**for**) result from either a lack of agreement between the characteristic formula and the conceptual content or a lack of logical coherence or alternatively an inadequate description of the model output. This class of errors includes basic mathematical hypotheses concerning the functions which link endogenous and exogenous variables (for example continuity or monotonicity).

- **solution** or **algorithm** errors (**sol**) in the solution process itself. If the process involves a computation, algorithm errors may occur because the method is heuristic and imprecise or simply because of technical constraints: termination within a finite period of time (truncation of an iterative algorithm), or because of the limited precision of operations (rounded numbers).

- **estimation** errors (**esti**) affecting the specification of mathematical functions and the assigning of numerical values to model parameters, in other terms in the

entire numerical description of the links between different endogenous variables and between endogenous variables and exogenous variables.

- **exogenous error (exo)**, in the application, as variables are affected by bias or uncertainty ("input error").

In order to clarify these ideas, Table 1 shows the location where each of these sources of error occurs in a model whose characteristic formula is $Y = F(X)$, where the effect (or response) Y is the vector of endogenous variables, the cause (or argument) X is the vector of exogenous variables and the mechanism F is the vector of the relationships of dependency.

Tab. 1. Types of error in a model.

Type	Location
Semantic	choice of X and Y , ideas concerning the elements in F
Formulation	formulation of F (relationships of dependency between explanatory and explained variables, basic mathematical hypotheses)
Algorithm	error in the solution process of F
Econometric (estimation)	mathematical and numerical expression of F (observation error affecting estimation data...)
Econometric (propagation)	error affecting specific values of X

Table 2 classes hypotheses regarding the dual criteria assignment model on the basis of the type of error which they are likely to generate.

1.2 The addition of errors

These five types of errors combine together. Let us demonstrate this by considering the model with the equation $Y = F(X)$ and using a specific symbol to denote each type of error. This is a question of clarifying the expression $Y = F(X) + \varepsilon$ where ε is the total error of the model, and Y and X are the observed variables.

Firstly there is the risk that the variables have been inaccurately observed: $Y = Y^* + \varepsilon_{\text{réf}}$, where Y^* is the "true" dependent variable and $\varepsilon_{\text{réf}}$ is a "reference" error which occurs in the observation of Y . Similarly, $X = X^* + \varepsilon_{\text{exo}}$, where X^* is the "true" exogenous variable and ε_{exo} is an exogenous error (bias or uncertainty) prior to propagation by the model.

Let us use $\varepsilon_{\text{sém}}$ to denote the semantic error of the "true" model between Y^* and X^* : this gives $Y^* = F^*(X^*) + \varepsilon_{\text{sém}}$ where F^* is the "true" dependency mechanism.

To this is added the formulation error ε_{for} which results from simplifications and approximations as regards F^* : $F^{(\text{for})} = F^* + \varepsilon_{\text{for}}$, thus $Y^* = F^{(\text{for})}(X^*) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}}$.

The algorithm error related to the actual solution of $F^{(\text{for})}$, is as follows $F^{(\text{sol})} = F^{(\text{for})} + \varepsilon_{\text{sol}}$, thus $Y^* = F^{(\text{sol})}(X^*) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{sol}}$.

To this we add the estimation error given by $\mathbf{F} = \mathbf{F}^{(\text{sol})} + \varepsilon_{\text{esti}}$, thus $\mathbf{Y}^* = \mathbf{F}(\mathbf{X}^*) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{sol}} - \varepsilon_{\text{esti}}$.

At last we add the exogenous error ε_{exo} relating to the exogenous variables \mathbf{X}^* : i.e. $\mathbf{Y}^* = \mathbf{F}(\mathbf{X} - \varepsilon_{\text{exo}}) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{sol}} - \varepsilon_{\text{esti}}$ which we approximate as follows $\mathbf{Y}^* \approx \mathbf{F}(\mathbf{X}) - \nabla_{\mathbf{X}}\mathbf{F} \cdot \varepsilon_{\text{exo}} + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{sol}} - \varepsilon_{\text{esti}}$, in which the total net modeling error $\varepsilon_{\text{net}} \approx -\nabla_{\mathbf{X}}\mathbf{F} \cdot \varepsilon_{\text{exo}} + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{sol}} - \varepsilon_{\text{esti}}$ can be identified.

The total gross modeling error also includes the reference error $\varepsilon_{\text{réf}} = \mathbf{Y} - \mathbf{Y}^*$ that arises between the observed \mathbf{Y} and the true dependent variable \mathbf{Y}^* : thus

$$\varepsilon_{\text{gross}} \approx -\nabla_{\mathbf{X}}\mathbf{F} \cdot \varepsilon_{\text{exo}} + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{sol}} - \varepsilon_{\text{esti}} + \varepsilon_{\text{réf}}.$$

Tab. 2. Analysis of error in the dual criteria model.

Class	Location
Design	<ul style="list-style-type: none"> - flows of trip-makers can be considered as continuous variables - trips from origin zones r to destination zones s - identification of paths k on the basis of the arcs a in the network - the travel time T_a on a link a depends on the local flow x_a, according to the travel time function $T_a = t_a(x_a)$ - full information to trip-makers concerning paths, prices and travel times, correct subjective perception of objective journey times T_a - rational behaviour of the users: each trip-maker optimizes his generalized travel time made of two cost criteria viz. time and price - differentiation of travellers on the basis of value of time - the volume of demand q_{rs} depends on the mean generalized time \bar{S}_{rs}, on the basis of a demand function $q_{rs} = D_{rs}(\bar{S}_{rs})$
Formulation	<ul style="list-style-type: none"> - the travel time function t_a is weakly increasing - values of time have a continuous probability distribution with CDF H_{rs} - the demand function D_{rs} is weakly decreasing
Algorithm	<ul style="list-style-type: none"> - nature of the algorithm - correctness and precision of the calculations, in the case of an iterative process
Estimation	<ul style="list-style-type: none"> - in the form and the parameters of the functions t_a, D_{rs}, H_{rs}
Propagation	<ul style="list-style-type: none"> - error with respect to exogenous variables (for example, the parameters for the functions t_a, D_{rs}, H_{rs})

Tab. 3. The addition of errors.

Combined types of error	Model formula
Sem	$Y^* = F^*(X^*) + \varepsilon_{\text{sém}}$
Sem & For	$Y^* = F^{(\text{for})}(X^*) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}}$
Sem & For & Sol	$Y^* = F^{(\text{rés})}(X^*) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{rés}}$
Sem & For & Sol & Esti	$Y^* = F(X^*) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{rés}} - \varepsilon_{\text{esti}}$
Sem & For & Sol & Esti & Exo	$Y^* = F(X - \varepsilon_{\text{exo}}) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{rés}} - \varepsilon_{\text{esti}}$
Net error + reference error	$Y = F(X - \varepsilon_{\text{exo}}) + \varepsilon_{\text{sém}} - \varepsilon_{\text{for}} - \varepsilon_{\text{rés}} - \varepsilon_{\text{esti}} + \varepsilon_{\text{réf}}$

1.3 The measurement and treatment of error

Each type of error must be given specific consideration.

As regards **semantic error**, which exists between the model and the studied system, we are only able to perceive the difference between the model and a theoretical representation of the system! Any measurement is therefore **relative**, as is any way of dealing with the error. A possible way of detecting this type of error is to sum up, and make consistent, knowledge about the system: this yields a theoretical reference with which specific hypotheses can be compared in order to assess their relevance. In a situation where a hypothesis is not relevant one solution is to improve the conceptual content of the model, or avoid using the model in question in an application where the hypothesis plays an important role!

In order to detect and correct **formulation error** we suggest (i) validating the mathematical logic and rigour of the model's characteristic formula, (ii) checking that the characteristic formula indeed contains each of the explanatory mechanisms assumed in the conceptual content. In the event of a problem the characteristic formula must be modified (a failure to do so means that the model cannot be technically valid!). In order to address the existence and uniqueness of outputs of the characteristic formula it is helpful to place it within a mathematical framework of reference, in other words to cast it into a standard operational research problem for which theoretical results are available.

Solution error, if we exclude technical failures in the technical system (hardware, software, operator error), this involves inaccuracies which result from rounded numbers and, in particular, which relate to achieving algorithm convergence when an iterative technique is used. In order to measure convergence a theoretically rigorous convergence test is required. If positive the rigorous test provides a technical guarantee, whereas the theoretical convergence of the algorithm is only a favourable assumption for a practical application. Solution error is dealt with by changing the algorithm or by modifying the convergence test.

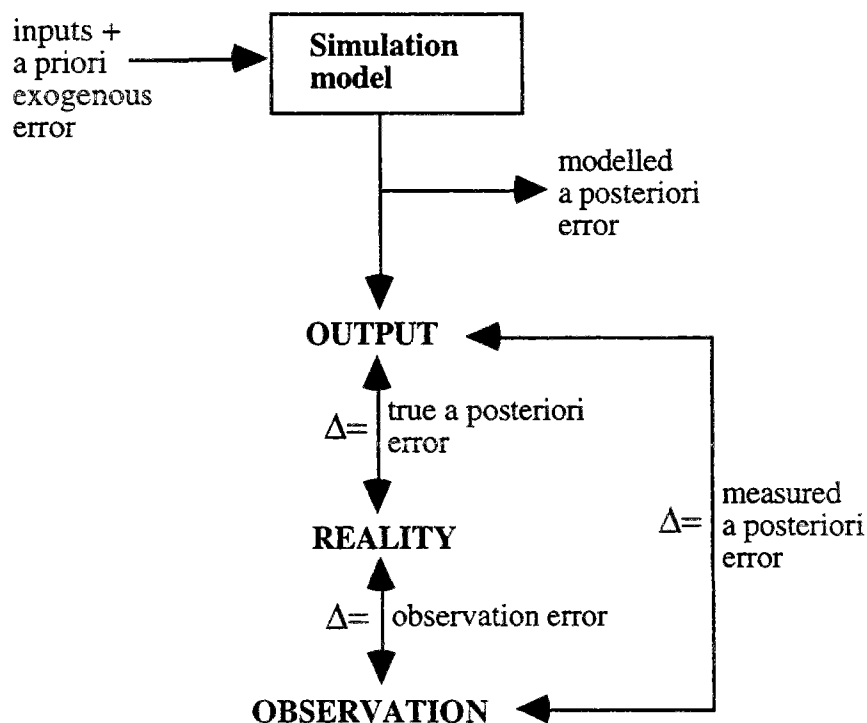
Econometric error affects exogenous variables, and the functions which give endogenous variables. The first is termed exogenous error and the second estimation

error. As the parameters are variables which are exogenous to the application all econometric errors can be considered to be exogenous errors. Uncertainty as regards exogenous variables is linked either to imprecise knowledge or an assumption (for example in the case of a hypothetical situation). It is propagated in the outputs of the model through the characteristic formula: a distinction is made between a priori exogenous error and propagated exogenous error. A priori exogenous error is quantified on the basis of one or more estimation models, which must not be confused with the simulation model. An estimation model generates parameter values from observations; it can also be used to quantify the associated estimation error, which contributes to the exogenous error of any simulation based on the estimation outputs. It should be noted that the estimation model may be explicitly probabilistic or remain intuitive and simply reflect the knowledge or intuition of the user.

Propagated exogenous error is quantified by propagating the a priori exogenous error through the characteristic formula: this is the principle of sensitivity analysis. Quantification is of assistance to the user in dealing with error as it provides a warning.

The other way of dealing with this type of error is to reduce a priori error by improving the estimation model or to reduce the sensitivity of the simulation model by increasing the sophistication of the characteristic formula.

Fig. 1. Comparison between the output of the model and reality.



1.4 Review

We have broken down modeling error into four different types: semantic error, formulation error, computation error and exogenous error. Measuring and dealing with these requires specific methods for each type. Design errors involve system analysis, formulation errors involve mathematical logic and mathematical programming, solution

errors involve the art and science of algorithms, and exogenous errors involve econometrics.

Our examination of modeling error allows us to make some remarks concerning other approaches to the subject. General econometric theory (eg Judge et al., 1988) distinguishes between semantic error (linked to "a priori judgment") and estimation error, does not deal with either formulation or computation and makes no distinction between a simulation model and an estimation model. In the area of traffic modelling, we should note several studies which relate to computation error (Horowitz, 1986, Rose et al., 1988), estimation error and propagation error (Ashley, 1981, TAM, 1995, Marche, 1991), and some general coverage in training manuals (eg Ortuzar and Willumsen, 1990). Our presentation provides clear vocabulary and categories, and a systematic method of detection and treatment. It reveals the importance of formulation error, the potential effects of which are largely unsuspected (see the two problems in the logit assignment model that were pointed out by Leurent, 1995b).

2. EXOGENOUS ERROR IN AN ASSIGNMENT MODEL: A FEW EXAMPLES

We must be careful not to overstress the importance of econometric error within modeling error as a whole. However, in the case of a model with a given conceptual content, with a characteristic formula which has been certified and which can be computed with the desired level of accuracy, econometric error is the only type of error which it is possible to alter. For this purpose it must be quantified and if possible reduced.

This section presents quantifications of a priori exogenous errors which affect three inputs of an assignment model. These are firstly a travel time function, secondly origin-destination flows and thirdly a distribution of the value of time. In each case an estimation model has been used in order to quantify exogenous error.

2.1 Travel time function

In the case of a static assignment model we shall consider a dependency of the local travel time y with respect to the local flow x in the absence of any queueing effect. We shall assume that there is a dependency $y = f_{\alpha, \beta}(x)$, where α and β are the parameters to be estimated. We wish to determine, for a specific road, the proportion of the variability of travel time which is explained by variations in flow. What we need to do is to break down total variability into explained and residual variability.

We shall index the n available observations $(x_i; y_i)$ by i . Let $\bar{y} = \frac{1}{n} \sum_i y_i$ and $\hat{y}_i = f_{\alpha, \beta}(x_i)$. The total variability is thus given by $\text{VT} = \sum_i (y_i - \bar{y})^2$, the explained variability is given by $\text{VE} = \sum_i (\hat{y}_i - \bar{y})^2$ and the residual variability by $\text{VR} = \sum_i (y_i - \hat{y}_i)^2$.

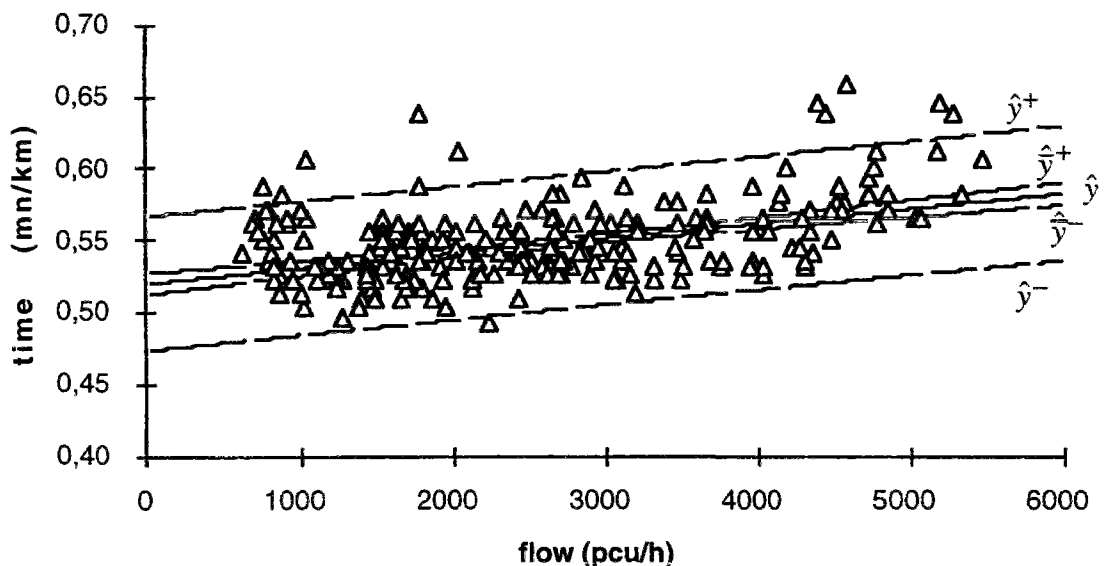
The purpose of the estimation is to minimize residual variability according to the parameters α and β , and, at the point $(\hat{\alpha}; \hat{\beta})$ where minimization occurs, to quantify the uncertainty associated with a prediction $y = f_{\hat{\alpha}, \hat{\beta}}(x)$.

Data. Our observations consist of spot speed and flow pairs which were measured during one day in 1990 on one three-lane carriageway of the A7 motorway between Lyon and Valence. The measurements were taken once every six minutes. The spot speeds v_i were transformed into spot kilometer journey times $y_i = 1/v_i$.

Assuming a standard linear regression model $y = f_{\alpha, \beta}(x) = \alpha + \beta x$, we obtained estimates of the ordinary least squares type: $\hat{\beta} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$ for β and

$\hat{\alpha} = \bar{y} - \bar{x}\hat{\beta}$ for α , in which $\bar{x} = \frac{1}{n} \sum_i x_i$. The example data set yielded: $\hat{\beta} = 1,04 \cdot 10^{-5} \frac{\text{mn}}{\text{km.pcu}}$ and $\hat{\alpha} = 0,52 \text{ mn/km}$.

Fig. 2. The set of observations. Source : ASF (1990).



Assuming that each observation y_i is a random variables in that it is a sample of y , the estimators $\hat{\alpha}$ et $\hat{\beta}$ are themselves random variables. According to the conventional assumptions concerning y_i , $\hat{\beta}$ has an approximately normal distribution whose mean is the true value of β and whose standard deviation is $SE_{\hat{\beta}} = \sigma / \sqrt{\sum_i (x_i - \bar{x})^2}$, where σ is the standard deviation of the residual uncertainty $y - \hat{y} = y - f_{\alpha, \beta}(x)$. σ^2 is estimated by the residual variance $s^2 = \left(\sum_i (\hat{y}_i - y_i)^2 \right) / (n - 2)$. The example data set yielded: $s = 2,43 \cdot 10^{-2} \frac{\text{mn}}{\text{km}}$ and $SE_{\hat{\beta}} = 1,22 \cdot 10^{-6} \frac{\text{mn}}{\text{km.pcu}}$.

Thus the confidence interval at the $1-z$ level of the slope β is $\beta = \hat{\beta} \pm t_{z/2}^{(n-2)} SE_{\hat{\beta}}$, where $t^{(n-2)}$ is a Student variable with $n-2$ degrees of freedom. For $1-z = 95\%$, our

data set yielded $\beta = \hat{\beta} \pm t_{z/2}^{(n-2)} \text{SE}_{\hat{\beta}} = 1,04 \cdot 10^{-5} \pm 2,32 \cdot 10^{-6}$. When the number of observations n is large, β 's distribution is approximately normal with mean zero and a standard deviation of $\text{SE}_{\hat{\beta}}$. The exogenous error which affects α is deduced from $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$.

At a point x_0 , the confidence interval at the $1-z$ confidence level for the mean value of y is $\hat{y}^{+/-} = \hat{\alpha} + \hat{\beta}x_0 \pm t_{z/2}^{(n-2)} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$ and only accounts for the uncertainty

which is linked to the estimation of parameters. The confidence interval at the $1-z$ level for a specific value of y at point x_0 also includes the residual uncertainty due to

the observation of y and is $\hat{y}^{+/-}(x_0) = \hat{\alpha} + \hat{\beta}x_0 \pm t_{z/2}^{(n-2)} \cdot s \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} + 1}$.

The credibility of the model is measured by comparing the explained variability to the residual variability. In the case of the linear regression model, an equivalent measure is to evaluate the critical probability of the null, "no slope" hypothesis ($\beta = 0$), which is the absence of a dependency between y and x : the critical probability p.c. is the probability of observing $\{t^{(n-2)} = \hat{\beta} / \text{SE}_{\hat{\beta}}\}$ with the hypothesis $\beta = 0$. This p.c. is measured by the area of the tail of a distribution $t^{(n-2)}$ above the value $t = \hat{\beta} / \text{SE}_{\hat{\beta}}$.

For our set of observations, we can see that $t = 8.49$ which gives p.c. $< 10^{-3}$. At a 0,999 confidence level we can reject the null, "no slope" hypothesis and conclude that the travel time really does increase with the growth of traffic.

If we consider a travel time function with a slope $\hat{\beta}$, as opposed to a zero slope with constant journey times, it is possible to reduce the unexplained variability from VT to VR. In a fixed times assignment, the unexplained variability associated with the travel time is VT, whereas in a variable times assignment it is reduced to VR. Continuing the example, at the confidence level $1-z = 95\%$ we have in the fixed times case $y = \bar{y} \pm t_{z/2}^{(n-1)} \sigma_{tf} = 0,549 \pm 0,052$ using $(n-1)\sigma_{tf}^2 = \text{VT}$, whereas in the variable times case $\hat{y}^{+/-}(x_0) = \hat{\alpha} + \hat{\beta}x \pm t_{z/2}^{(n-2)} \sigma_{tv} = 0,549 \pm 0,046$ using $(n-2)\sigma_{tv}^2 = \text{VR}$.

With fixed times, $\frac{\sigma_y}{y} \approx 5,0\%$ whereas with variable times $\frac{\sigma_y}{y} \approx 4,4\%$.

2.2 O-D Flows

Static assignment relates to a reference period during which traffic conditions are considered to be unchanged. The reference period corresponds to an average situation, for example in the case of an interurban road, the average day of the year. There is probably some variability around this average situation: in the case of origin-destination (O-D) flows, this variability influences the flows on the arcs, which in turn influence journey times, which in turn modify route choice.

Data. In order to measure the scale of the phenomenon we analyzed daily O-D flow measurements taken during 1994 at 10 toll points on the Motorway between Paris and Tours, which is operated by the company Cofiroute. From this data we deduced the average annual daily traffic and the covariance between flows. Such extensive knowledge is not available in the case of a dedicated traffic survey, which merely has O-D survey results covering a few days.

We have considered 3 O-D pairs with contrasting average daily flows. For each of these we computed the average daily flow μ , the standard deviation σ , and the covariance between the O-D flow and the flows for all other O-D in relation to μ and the average total flow Q (table 4). It should be noted that the relative dispersion does not decrease with the average, rather the contrary. The correlations between O-D pairs seem to be high for O-D pairs with high flow, possibly due to highly seasonally-peaked long-distance traffic.

Tab. 4. Moments of the true distribution and of several sample distributions (with respect to the sample size N).

O-D pair	μ (cars/day)	σ (cars/day)	Cov / μQ	N = 1, SE/ $\mu = \sigma/\mu$	N = 5, SE/ μ	N = 20, SE/ μ
Paris-Tours	5 727	4 163	0.231	0.73	0.33	0.17
Orléans-Tours	1 311	763	0.202	0.58	0.26	0.13
Arthenay-Orléans	123	47	0.011	0.38	0.17	0.09

2.3 Behavioural parameters

To measure the exogenous error associated with the value of time (VoT), the estimation error affecting the parameters of its distribution must be estimated. The maximum likelihood method can be used to estimate a VoT distribution, using disaggregate, individual data (Leurent, 1995a, e). Assume that, to trip-maker i , the generalized cost of travelling along path k is

$$G_k(i) = \left\{ \sum_b r_{i,b} \cdot X_{b,k} \right\} + v_i T_k = P_k + v_i T_k$$

where b is an attribute index, $X_{b,k}$ the value of the b -th attribute along path k (an attribute may be the monetary expense, or distance, or discomfort etc, but it cannot be the travel time), $r_{i,b}$ are trade-off coefficients, v_i is the VoT and T_k is the travel time along path k .

There are $B+2$ parameters to be estimated: (i) the B coefficients $r_{i,b}$, hereafter assumed not to depend on i , (ii) the 2 parameters of the distribution of the VoT v_i , hereafter assumed to follow a log-normal distribution characterized by means of the mean μ and standard-deviation σ of its logarithm.

We assume that, on a given O-D pair, the alternative paths k were surveyed so that their flows f_k are known. The proposed estimation method relies on the maximization of the log-likelihood of the sample, evaluated as

$$L(\Theta) = \log \prod_i \pi_{k_i}(\Theta) = \sum_k f_k \log(\pi_k(\Theta))$$

where $\Theta = \{\theta_n\}_n = \{r_b\}_b \cup \{\mu; \sigma\}$ denotes the vector of parameters, $\pi_k(\Theta)$ is the modelled probability of travelling on path k , and k_i is the observed, chosen path of trip-maker i .

If we assume that the alternative paths are ranked in order of increasing travel prices and that they are all efficient, then $\pi_k(\theta) = H_{\mu, \sigma}(U_k) - H_{\mu, \sigma}(L_k)$, where U_k and L_k are the supply-related frontier VoTs between alternatives k and $k+1$, and $k-1$ and k , respectively.

Maximization of $L(\theta)$ can be performed using a first- or second- order descent algorithm: it is easy to calculate the partial derivatives of L with respect to the θ_n from the partial derivatives of the π_k with respect to the θ_n . From the value of the second order derivatives of L at the optimum, we can estimate the covariance matrix of the estimators.

Table 5 provides an application with Revealed Preference data: three O-D pairs in the west of France, each with two competing paths, were surveyed during one day in 1990. The travel times T_k were measured, and the travel prices P_k include the toll if applicable and the travel costs.

Tab. 5. Modeled vs. observed flows in the Angers-Nantes O-D survey.

Origin-destination pair		P_k (FF)	T_k (h)	veh/day observed	veh/day modeled
Nantes- Angers	Free highway N23	65.4	0.967	1500	1522
	Toll motorway A11	86.7	0.617	3500	3477
Nantes- Ancenis	Free highway N23	25.4	0.40	2600	2621
	Toll motorway A11	33.9	0.30	1300	1279
Angers- Ancenis	Free highway N23	40.1	0.567	500	456
	Toll motorway A11	50.8	0.417	450	494

A log-normal distribution of the VoT was estimated, yielding $\hat{\mu} = 4.284$, $\hat{\sigma} = 0.348$, $\hat{\sigma}_{\hat{\mu}} = 0.0045$, $\hat{\sigma}_{\hat{\sigma}} = 0.01$, $\text{Cov}(\hat{\mu}; \hat{\sigma}) / \hat{\mu}\hat{\sigma} = -7.10^{-6}$, hence an estimated mean VoT of $\exp(\mu + \sigma^2 / 2) = 77.1$ FF/hour.

2.4 Provisional conclusion

We have given examples for the exogenous error for three fundamental inputs to an assignment model. According to the input and the observations which are available for estimation, the ratio between exogenous error and the mean value of the input appears to be extremely variable: $\sigma_X / X \approx 1\%$ for behavioural parameters, $\approx 5\%$ for travel time data and $\approx 20\%$ for O-D flow data. Let us note that the exogenous error still includes exogenous assumption error, which is always possible when a hypothetical situation is simulated.

3. PROPAGATION OF EXOGENOUS ERROR THROUGH THE SIMULATION MODEL

In order to assess the effect of econometric error on the output of the simulation model, it is not only necessary to know the a priori exogenous error and the estimation error, but also examine the way in which these errors are propagated through the model. This is the purpose of sensitivity analysis which is of crucial practical importance in order to appraise the statistical significance of the result.

Sensitivity analysis may be performed either by numerical simulation, by applying the model to different sets of inputs, or analytically, when the characteristic formula of the model is of a type which can be used to develop propagation formulae for small-scale exogenous errors (Tukey, 1957).

There is such a formula in the case of the dual criteria assignment model (Leurent, 1994c). We have developed a procedure for propagating econometric error through this model, in accordance with a principle which is described in sub-section 3.1. Sub-section 3.2 presents an application of this procedure to a traffic study.

3.1 Principle of error propagation

Our aim is to measure the error on the output vector Y of a model, given the exogenous error ε_X in the input vector X . In the case of a model with a straightforward dependence expressed by $Y = F(X)$ with a differentiable F , for small ε_X it holds that $\varepsilon_Y = (\nabla_X F) \varepsilon_X$, from which we can deduce the mean and standard deviation of ε_Y , given the distribution of ε_X .

In the case of an implicit model, i.e. $F(X, Y) = 0$, if $\nabla_Y F$ can be inverted, then it follows that $\varepsilon_Y = [\nabla_Y F]^{-1} [-\nabla_X F] \varepsilon_X$ (Tobin 1986). Let us now recall the result which is useful for our purpose. We can express the set of all possible model input combinations as follows: $\mathbf{K} = \{ \mathbf{x} \in \mathfrak{R}^n ; \mathbf{g}(\mathbf{x}) \geq \mathbf{0}, \mathbf{d}(\mathbf{x}) = \mathbf{0} \}$, where $\mathbf{g}: \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ is a differentiable mapping and $\mathbf{d}: \mathfrak{R}^n \rightarrow \mathfrak{R}^p$ is a linear affine mapping.

We shall assume that any equilibrium state \mathbf{x}^* of the model solves the following variational inequality, where \mathbf{F} is a continuous mapping of $\mathfrak{R}^n \rightarrow \mathfrak{R}^n$:

$$\forall \mathbf{x} \in \mathbf{K}, \mathbf{F}(\mathbf{x}^*)^t \cdot (\mathbf{x} - \mathbf{x}^*) \geq \mathbf{0},$$

Let ε be a disturbance error standing for the exogenous error: we shall write $\mathbf{F}(\mathbf{x}; \varepsilon)$, $\mathbf{g}(\mathbf{x}; \varepsilon)$ and $\mathbf{d}(\mathbf{x}; \varepsilon)$ to take ε into account. Let also $\mathbf{y} = [\mathbf{x}^t; \boldsymbol{\pi}^t; \boldsymbol{\mu}^t]^t$ where $\boldsymbol{\pi} \in \mathfrak{R}^m$ and $\boldsymbol{\mu} \in \mathfrak{R}^p$. Finally, let $\mathbf{J}(\mathbf{y}; \varepsilon)$ express the mapping of $\mathfrak{R}^{n+m+p} \rightarrow \mathfrak{R}^{n+m+p}$ the components of which are as follows:

$$\mathbf{F}(\mathbf{x}; \varepsilon) - \sum_{i=1}^m \pi_i \nabla g_i(\mathbf{x}; \varepsilon)^t - \sum_{i=1}^p \mu_i \nabla d_i(\mathbf{x}; \varepsilon)^t,$$

$$\pi_i g_i(\mathbf{x}; \varepsilon) \text{ for values of } i \text{ between } 1 \text{ and } m,$$

$$d_i(\mathbf{x}; \varepsilon) \text{ for values of } i \text{ between } 1 \text{ and } p.$$

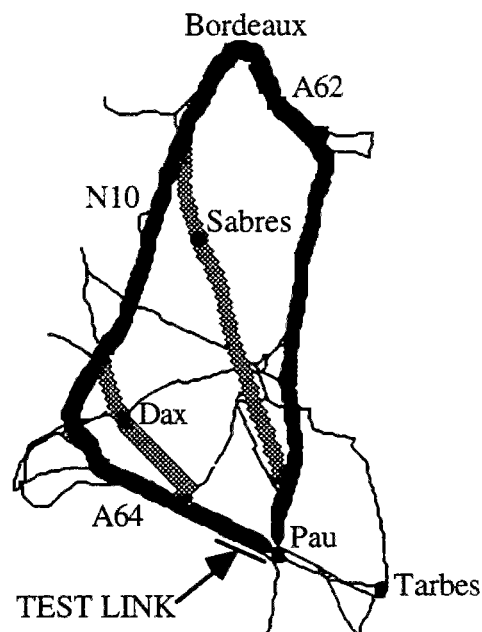
Under some technical assumptions, at a solution point $\mathbf{x}^* \in \mathbf{K}$ there exist $\pi^* \in \mathfrak{R}^m$ and $\mu^* \in \mathfrak{R}^p$ such that $\mathbf{J}(\mathbf{y}^*; \mathbf{0}) = \mathbf{0}$ and $\nabla_{\mathbf{y}} \mathbf{J}$ is invertible in a local neighbourhood of $(\mathbf{y}^*; \mathbf{0})$, and that make the partial derivatives of \mathbf{y}^* with respect to ε satisfy:

$$\nabla_{\varepsilon} \mathbf{y}^* = [\nabla_{\mathbf{y}} \mathbf{J}(\mathbf{y}^*; \mathbf{0})]^{-1} [-\nabla_{\varepsilon} \mathbf{J}(\mathbf{y}^*; \mathbf{0})].$$

3.2 Case study

We shall consider an interurban toll motorway which contributes to communications between Bordeaux and Pau in the south-west of France. These two cities are 220 km apart and are served by four main routes (figure 3). For the purposes of the traffic study 180 one-way arcs and 19 demand zones were coded.

Fig. 3. Road network structure for the Bordeaux-Pau traffic study.



We have applied a dual criteria model in which journey times vary with traffic, in which O-D volumes vary with respect to the mean generalized travel time on the basis of $q_{rs} = D_{rs}(\bar{S}_{rs}) = \hat{q}_{rs}(\bar{S}_{rs} / \hat{S}_{rs})^{-0.6}$. We have considered two classes of vehicles, cars and heavy freight vehicles. The equilibrium was computed by means of the transfer-and-equalize procedure (Leurent 1995c) which handles the paths.

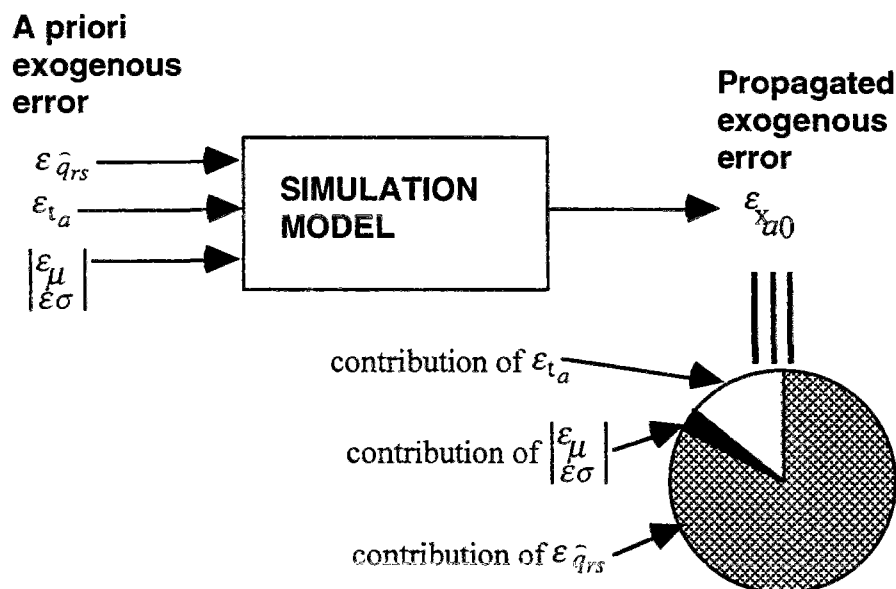
We have studied the exogenous error $\varepsilon_{x_{a0}}$ which is propagated for car traffic on the toll road by considering the a priori exogenous errors which relate to the following:

- the journey times t_a on the arcs. Each arc a has a centred gaussian error ε_{t_a} with a standard deviation σ_{t_a} such that $\sigma_{t_a} / t_a = 10\%$, and which is independent of other errors.

- the O-D volumes for car traffic. Each O-D pair $r-s$ has a centred gaussian error $\varepsilon_{\hat{q}_{rs}}$ with a standard deviation $\sigma_{\hat{q}_{rs}}$ such that $\sigma_{\hat{q}_{rs}} / \hat{q}_{rs} = 20\%$, and which is independent from other errors.
- the parameters M and σ of a log-normal distribution of the value of time : M is the mean and σ the standard deviation of the logarithm of the value of time. M and σ are assumed to be subject to the centred gaussian uncertainties ε_M and ε_σ , and have the standard deviations σ_M and σ_σ respectively such that $\sigma_M / M = 1\%$ and $\sigma_\sigma / \sigma = 3\%$. ε_M and ε_σ are assumed independent from other errors but correlated with each other according to $\text{Cov}(\varepsilon_M; \varepsilon_\sigma) / M\sigma = -0,1\%$.

On the studied road $a0$, for average modeled traffic $x_{a0} = 1220$ cars/day, after propagation of the a priori exogenous error we obtain an a posteriori error with mean zero and standard deviation $\sigma_{x_{a0}} = 123$ cars/day, hence $\sigma_{x_{a0}} / x_{a0} = 10,1\%$. This uncertainty is mainly due to demand volumes (83%, see figure 4) and to a lesser degree to journey times (14%) and the distribution of values of time (3%).

Fig. 4. The accumulation of exogenous uncertainties.



3.3 Discussion

We have proved that it is possible in practice, in the context of a real traffic study, to provide a relatively complete analysis of econometric error, by distinguishing between the role of errors with regard to times, volumes and values of time. This proof of feasibility should encourage practitioners to conduct a rigorous examination of error, and not merely conduct a few tests on only a few variables. A better understanding of a priori exogenous error is desirable: section 2 contains examples of this for development and completion.

The procedure used to propagate error requires knowledge of the model output with respect to paths, and it is thus necessary to resolve the assignment problem using an

algorithm which handles paths. The propagation procedure also demands the inversion of a matrix which is larger than the number of O-D pairs: this limits the size of the problems which can be currently dealt with to a few tens of demand zones.

4. CONCLUSION

Exogenous error in a simulation model can be studied quantitatively, which is necessary in order to appraise the statistical significance of the model's outputs. However, the error of the model does not consist only of the econometric error, if only because the validity of the sensitivity analysis depends entirely on the validity of the characteristic formula and the resolution algorithm and requires that formulation error and computation error have already been detected and treated.

Finally, the model design is a source of error which should never be ignored, in particular in view of the fact that appraising it is always dependent on a particular theoretical representation. As a preventive action against design error it is possible to sum up, and make consistent, the empirical facts and theoretical concepts relating to the modelled system.

Glossaire des sigles

ARIANE	logiciel d'affectation statique en milieu routier interurbain, développé par le SETRA
ASF	Autoroutes du Sud de la France (société concessionnaire)
CETE	Centre d'Etudes Techniques de l'Equipement
COBA	COst Benefit Analysis, guide méthodologique du DOT
COFIROUTE	Compagnie Financière des Autoroutes (société concessionnaire)
DAVIS	logiciel d'affectation routière statique en milieu urbain, développé par l'INRETS
DOT	Department of Transport, ministère Britannique des transports
EMME/2	logiciel d'étude de déplacement (affectation etc.) développé par INRO
INRETS	Institut National de Recherche sur les Transports et leur Sécurité
INRO	bureau d'études canadien
LASER	Liaison Autoroutière Souterraine Express Régionale, projet développé par GTM-Entrepose (société de bâtiment et travaux publics)
MARS	logiciel polyvalent d'affectation, développé par l'INRETS
MATISSE	modèle de transport interrégional de voyageurs, développé par l'INRETS
MSA	Method of Successive Averages (un algorithme)
MVA	Martin Voorhees Associates (Consultant)
NCP	acronyme anglais pour Problème de Complémentarité Non linéaire
PCC	Plus Court Chemin
PET	Procédure d'Egalisation par Transvasement (un algorithme)
POLYDROM	logiciel d'étude de déplacement (affectation etc.) développé par De Rham Consultant
TRIP	logiciel d'étude de déplacement (affectation etc.) développé par MVA
SETRA	Service d'Etudes Techniques des Routes et Autoroutes
TAM	Traffic Appraisal Manual, guide méthodologique du DOT
USAP	Union des Sociétés d'Autoroutes à Péage

Liste des figures

Fig. 1a. Principes du schéma à quatre étapes.....	9
Fig. 1b. Données pour le cas d'école.....	13
Fig. 1c. Trafics et recettes, en fonction du péage.	15
Fig. 1d. Déplacements de l'équilibre offre-demande.	16
Fig. 1e. Flot sur l'arc diagonal en fonction du coefficient antidispersif.	16
Fig. 2a. Qualité d'une étude par simulation.....	37
Fig. 2b. Un processus décentralisé d'audit.....	39
Fig. 5a. La méthode d'exploration systématique.....	85
Fig. 5b. Schéma de principe de l'audit économétrique.....	104
Fig. 7a. Mécanisme d'ensemble du modèle bicritère.	130
Fig. 8a. Egalité des valeurs de coupure respectivement offerte et demandée.	148
Fig. 8b. Egalité des valeurs de coupure, cas des temps variables.....	151
Fig. 8c. Effets d'une contrainte de capacité.....	152
Fig. 8d. Un équilibre contraint de type A.Ba : décalage de la valeur de coupure offerte.....	154
Fig. 8e. Une fonction de demande.....	155
Fig. 8f. Les itinéraires H-efficaces.....	158
Fig. 8g. Les ensembles H-efficaces.	160
Fig. 8h. Les fonctions de temps de parcours des trois modes.....	176
Fig. 8i. Distribution des valeurs du temps.	176
Fig. 9a. Equilibre entre offre et demande d'acheminement.	186
Fig. 9b. Principe des calculs.....	187
Fig. 9c. Schéma du réseau des premières études A86 ouest.....	206
Fig. 9d. Evolution de $\log(J^{(n)} / J^* - 1)$ en fonction du nombre d'itérations n.....	208
Fig. 9e. RMSE relatif normé, en fonction du nombre d'itérations (échelle logarithmique).....	208
Fig. 9f. RMSE absolu normé, en fonction du nombre d'itérations (échelle logarithmique).....	209
Fig. 9g. Taux de convergence de la PET entre les itérations 10 et 36.	210

Fig. 9h. Taux de convergence de la MSA-iti.	210
Fig. 9i. RMSE relatif normé vs. RMSE absolu normé ou vs. $J(n)/J^*-1$,	211
Fig. 10a. Le jeu d'observations.	224
Fig. 10b. Bandes de confiance.	226
Fig. 10c. La concurrence A11-N23 entre Nantes et Angers.	235
Fig. 10d. Modèle à temps fixes et demande fixe.	241
Fig. 10e. Modèle à temps variables et demande fixe.	245
Fig. 10f. Modèle à temps variables et demande variable.	248
Fig. 10g. Le site de l'étude.	255
Fig. 10h. L'accumulation des erreurs exogènes.	256
Fig. 11a. Schéma de principe du fleuve de l'erreur.	258

Liste des tableaux

Tab. 1A. Résultats attendus et composants qui leur correspondent.....	12
Tab. 3A. Synthèse de l'état de la pratique et de la recherche.....	57
Tab. 4A. Classes d'erreur nette dans un modèle.....	63
Tab. 4B. Hypothèses dans le modèle bicritère.....	63
Tab. 4C. Cumul des erreurs.....	65
Tab. 5A. Récapitulation des quatre postes d'erreur de calcul.....	98
Tab. 7A. Phénomènes et solutions de représentation.....	126
Tab. 7B. Variables exogènes d'identification physique et comportementale.....	130
Tab. 7C. Variables exogènes de spécification fonctionnelle.....	131
Tab. 7D. Recensement des variables endogènes.....	131
Tab. 7E. Relations de dépendance.....	132
Tab. 7F. Audit sémantique du modèle bicritère.....	133
Tab. 8A. Variables exogènes pour les dimensions physiques et comportementales.....	138
Tab. 8Abis. Variables exogènes pour la spécification fonctionnelle.....	138
Tab. 8B. Recensement des variables endogènes.....	139
Tab. 8C. Relations de dépendance.....	139
Tab. 8D. Notation vectorielle des variables et fonctions.....	140
Tab. 8E. Forme vectorielle des relations de dépendance.....	141
Tab. 8F. Les principales notions originales de l'équilibre bicritère.....	157
Tab. 8G. Caractères des trois modes.....	175
Tab. 8H. Résultats avant et après la mise en service.....	177
Tab. 9A. Evaluation technique des principaux critères de convergence.....	195
Tab. 9B. Applicabilité des critères aux algorithmes, selon le modèle.....	200
Tab. 10A. Moments de la (vraie) distribution empirique.....	227
Tab. 10B. Moments de quelques distributions d'échantillons.....	228
Tab. 10C. Concurrence entre itinéraires, déplacements à environ 150 km.....	231
Tab. 10D. Déplacements VL, motif privé, heure creuse, à distance de 150 km.....	235
Tab. 10E. Attributs "objectifs" des itinéraires et résultats de l'enquête.....	236
Tab. 10F. Comparaison entre observation et simulations, liaison Angers-Nantes.....	236