



D.R.A.S.T

MODELISATION DES FLUX DE TRANSPORTS

Mesurabilité et modélisation dans le transport de marchandises à l'échelle européenne

Appui sur le cas des échanges entre l'Europe et le
Royaume-Uni

Mars 2003



SOMMAIRE

I.	OBJET DE LA RECHERCHE.....	3
II.	PROBLEMATIQUE.....	4
II.1.	OPERATIONNALITE DU MODELE.....	4
II.2.	PROBLEMATIQUE.....	5
II.3.	LE MODELE DEVELOPPE PAR ISIS POUR EUROTUNNEL	7
III.	ORGANISATION DU RAPPORT.....	8
IV.	RECHERCHE ET IDENTIFICATION DES BASES DE DONNEES FRET.....	9
IV.1.	GENERALITES	9
IV.2.	PRODUCTION DE DONNEES ET SOURCES DE DONNEES	9
IV.3.	PRESENTATION DES RESULTATS SOUS FORME DE TABLEAUX	10
IV.4.	ELEMENTS SOUS-JACENTS AUX RESULTATS	12
V.	ARTICULATION ENTRE MESURABILITE ET COHERENCE / PERTINENCE DANS LA MODELISATION DES FLUX DE MARCHANDISES	49
V.1.	IDENTIFICATION DES SOURCES DE DONNEES EXISTANTES ET MISE EN OEUVRE.....	49
V.2.	PROBLEMES DE VALIDATION DES NOUVELLES BASES.....	57
V.3.	CONSTRAINTES FONCTIONNELLES ASSOCIEES AUX BASES DE DONNEES	64
VI.	LA RELATION MESURABILITE – COHERENCE	68
VI.1.	TEST DE MODIFICATION DU ZONAGE.....	69
VI.2.	TEST DE MODIFICATION DU NIVEAU DES TARIFS	72
VI.3.	TEST DE MODIFICATION DU NOMBRE DE ZONES ET DU NIVEAU DES PRIX	75
VI.4.	TEST DE NON PRISE EN COMPTE DE LIAISONS FERRIES.....	77
VI.5.	PRISE EN COMPTE D’UN NOMBRE DE LIAISONS FERRIES REDUITS ET D’UN ZONAGE DIFFERENT	79
VI.6.	. PRISE EN COMPTE D’UN NOMBRE DE LIAISONS FERRIES REDUITS, D’UN ZONAGE DIFFERENT ET DE NIVEAUX DE TARIFS DIFFERENTS.....	81
VI.7.	CONCLUSIONS SUR LES TESTS	84
VII.	CONCLUSION GENERALE	85

I. OBJET DE LA RECHERCHE

La présente recherche s'inscrit dans un appel à proposition lancé par le groupe thématique "Modélisation des flux de transports" de la DRAST (Ministère des Transports et de l'Équipement).

Elle s'intéresse aux aspects relatifs à la « mesurabilité » dans le domaine de la modélisation en transport de marchandises; à caractère international.

La mesurabilité, qui sera définie par la suite, renvoie à **la production de données**, terminologie qui recouvre plusieurs acceptions relatives à la « génération » de données par enquête, à la « reprise » de données existantes, ou à la « création » de données (utilisation de sources diverses pour proposer de nouvelles valeurs).

L'objet de cette recherche est de tenter d'apporter quelques éléments de réponses quant à l'articulation entre la production de données et les autres dimensions relatives au développement d'une modélisation fret.

Il convient de préciser que cette recherche d'une durée de 1,1 homme-mois a pour prétention d'ouvrir un champ pour les échanges et la réflexion sur les aspects relatifs aux données dans le domaine qui nous occupe. Elle ne constitue donc pas une fin en soit.

II. PROBLEMATIQUE

II.1. OPERATIONNALITE DU MODELE

Cette recherche s'appuie sur les principes énoncés par le Professeur Bonnafous¹.

Sachant qu'un modèle consiste en une construction théorique et une formalisation mathématique², se pose la question de sa qualité.

« Un modèle apte à réaliser les performances que l'on attend de lui sera dit opérationnel. L'opérationnalité du modèle constitue donc sa qualité fondamentale » (A. Bonnafous).

Or, la notion même d'opérationnalité d'un modèle renvoie aux objectifs qui lui sont assignés. Le modèle ne peut répondre qu'aux questions qui lui sont posées, dans un contexte particulier. Ainsi un modèle de type prix-temps permettant d'analyser la concurrence rail / avion pour les voyageurs, ne sera pas forcément à même de traiter la concurrence rail / route pour les marchandises.

Selon A. Bonnafous, trois conditions président l'opérationnalité d'un modèle : la pertinence, la cohérence et la mesurabilité. Ces trois conditions qui sont contradictoires, nous poussent à nous interroger sur les erreurs intervenant dans la conception ou l'utilisation des modèles.

Reprenons, ici, P. Bonnel, qui s'appuie sur A. Bonnafous :

- *« La pertinence correspond à la capacité du modèle à être conforme à ce que l'on croit savoir de la réalité, de l'état des choses. Un modèle non pertinent ne saurait être conforme à l'état des choses. L'absence de pertinence trouve principalement sa source dans deux dimensions. La première concerne le choix des variables explicatives, soit que certaines soient omises ou soient mal définies. La seconde concerne les relations entre variables, que certaines ne soient pas prises en compte ou soient mal définies ou spécifiées.*
- *La cohérence du modèle recouvre deux notions : tout d'abord une condition de cohérence interne du modèle, ensuite une condition de cohérence d'objectifs du modèle. La cohérence interne du modèle correspond à une condition de cohérence logico-mathématique. C'est-à-dire que le modèle ne doit pas comporter de contradiction interne que ce soit au niveau de la définition des variables ou des valeurs prises par celles-ci ou au niveau des relations de causalité retenues. Toutefois, sa mise en oeuvre peut parfois soulever quelques difficultés. La seconde condition est celle de cohérence d'objectifs du modèle. L'ajout de nouvelles variables explicatives pour répondre à un objectif particulier pourra nécessiter l'introduction d'équations nouvelles ou d'hypothèses supplémentaires pour lever*

¹ A compléter

² Pour plus détail, cf. P. Bonnel : Prévion de la demande de transport, Manuel d'analyse et de modélisation de la demande de déplacements urbains de personnes. LET / ENTPE / CNRS. Janvier 2001

une indétermination. Elle nécessitera également le plus souvent des tests d'indépendances entre variables.

- *La mesurabilité constitue la troisième condition de l'opérationnalité du modèle. Cette exigence se décline selon trois registres:*
 1. *La mesurabilité des relations causales correspond à la nécessité de pouvoir les formaliser. Une absence de formalisation ne permettrait pas d'utiliser le modèle en prévision.*
 2. *La mesurabilité des quantités mises en jeu dans le modèle est bien évidemment nécessaire (ex : mesurabilité d'une des variables « input » du modèle. Si telle variable doit croître sous l'effet de la variation d'une autre variable le moins que l'on puisse attendre est de pouvoir confronter cette relation à l'état des choses).*
 3. *Enfin, la mesurabilité des paramètres de ses équations est également indispensable. L'indétermination d'un coefficient ne permettra pas d'estimer le modèle qui restera... indéterminé.*

*C'est pour répondre à ces trois dimensions de la mesurabilité que **la production de données** est nécessaire. En l'absence de données d'enquête, les quantités des variables risquent fort de ne pouvoir être définies. Mais surtout, la formalisation ne pourra pas être confrontée à la « réalité » représentée par le jeu de... données. Enfin, la détermination des coefficients du calibrage sera impossible.*

Les trois conditions précédentes sont nécessaires à l'opérationnalité du modèle, et elles sont aussi suffisantes (Bonnaïfous, Gérardin, 1976). Mais elles sont contradictoires. Ainsi l'amélioration de l'une des conditions se fait le plus souvent au détriment des deux autres, constituant ainsi la problématique du modèle ».

II.2. PROBLEMATIQUE

La contradiction entre les trois composantes précédentes, s'opère de différente manière sur lesquelles nous ne nous attarderons pas ³.

Dans cette recherche on s'intéresse aux aspects relatifs à la **production des données** (au sens plus large que celui évoqué ci-avant, et défini en IV.2.), données qui permettent de satisfaire les conditions de mesurabilité.

Considérons un modèle qui est pertinent car les variables explicatives utilisées sont suffisantes pour expliquer les phénomènes modélisés et que les relations entre les variables ont été testées et jugées comme satisfaisantes.

Considérons, aussi, un modèle cohérent, car il utilise des fonctions mathématiques largement éprouvées.

³ On renvoi le lecteur aux travaux de P. Bonnel à ce sujet.

L'objet de la recherche concerne **la relation entre mesurabilité et cohérence / pertinence**. Si un modèle est jugé comme cohérent et pertinent c'est parce qu'il a été nécessaire d'effectuer plusieurs opérations relatives à la **production de données** pour la modélisation. La solution obtenue à l'issue de ce processus participe à l'atteinte d'un **état d'équilibre dans la mesurabilité**, qui se retrouvera ensuite au niveau de la cohérence et de la pertinence.

L'objet de la présente recherche concerne aussi **la qualité de la mesurabilité**. Si l'on modifie certaines données, alors on peut tester l'impact sur les résultats. Une modification du zonage, des tarifs, du réseau, ou toute autre modification, correspondent à des erreurs possibles qui peuvent être testées. Dans ce cas on raisonne ceteris-paribus, c'est-à-dire que les hypothèses prises en projection sont considérées comme n'intervenant pas, car elles ne sont pas le fait du modélisateur. Ceci n'a donc rien à voir avec les tests de sensibilité sur certaines variables explicatives, mais renvoie bien à la modification de certains « fondamentaux ».

Un modèle sera considéré comme opérationnel à deux niveaux d'un point de vue pratique, selon l'état d'équilibre atteint dans la mesurabilité et selon la qualité de la mesurabilité. Le modèle opérationnel sera conforme à ce que le technicien croit savoir des phénomènes (en ce cas les résultats reconstitués doivent être bons), le modèle sera aussi opérationnel quant le client se « l'appropriera ».

Les deux niveaux sont donc les suivantes :

- Micro-opérationnalité : les trafics trouvés en reconstitution sont bons, et les tests d'élasticités réalisés conformes aux « canons » de la modélisation. On considère, aussi, que les trafics sont corrects en projection. Même si cet état n'est pas connu, il a été « approché » à l'aide de scénarios contrastés divers qui seront ensuite exploités par le client.
- Macro-opérationnalité : le client trouve que le modèle est conforme à la vision qu'il a de son marché, et il s'en sert comme un outil d'aide à la décision / réflexion stratégique et commerciale.

La mesurabilité est l'élément clef du système que représente un modèle. En effet, cohérence et pertinence peuvent toujours être traités « à minima » pour ne pas prendre de risque, et ce dans le cadre de procédures éprouvées (c'est ce qui se passe généralement dans les bureaux d'études). Le problème concerne donc la production des données. Un modèle cohérent et pertinent atteindra un certain degré d'opérationnalité en fonction de l'état d'équilibre dans la mesurabilité.

La problématique de cette recherche est donc double :

Comment assurer une bonne articulation entre mesurabilité et cohérence / pertinence dans la modélisation des flux de marchandises ?

Pour répondre à cette question et sachant que l'on se situe au niveau du fret **dans son ensemble**, le cas des échanges de flux de marchandises entre l'Europe et le Royaume-Uni est pris comme appui.

Le développement des pratiques de modélisation marchandises étant un domaine en expansion, et ce encore plus du fait de la croissance de certains types de flux de transit ou d'échanges, le cas de la mesurabilité méritait d'être abordé.

II.3. LE MODELE DEVELOPPE PAR ISIS POUR EUROTUNNEL

Une partie nos réflexions d'appuient sur un modèle développée par ISIS pour le compte de la société Eurotunnel. Ce modèle fret spatialisé à l'échelle européenne est composé de deux sous-modèles. Un modèle pour les produits dits « unitarisés », c'est-à-dire tous les produits qui sont « conditionnés », et un modèle pour les produits « non unitarisés ». Le premier modèle met en concurrence le train dit « intermodal » (ou combiné) et le camion, sachant que ses derniers peuvent traverser la manche soit par le tunnel (navettes fret) soit par le ferry, l'un est l'autre étant un prolongement de la route. Les camions, accompagnés ou non accompagnés peuvent utiliser plus de 30 liaisons ferries, regroupées dans 7 lignes écran portuaires. Le second modèle met en concurrence les trains dits automobiles (transport de voiture), les trains conventionnels et le bateau.

Le découpage comprend 83 zones sur les 15 pays de l'Union Européenne, plus 5 pays de l'Est et la Suisse. Les flux entre le Royaume-Uni et les pays du continent sont modélisés pour 13 groupes de produits, en exportations et en importations. Ces produits (demande) se déclinent sur un réseau ferroviaire, un réseau routier (composé de 3 niveaux), un réseau ferries et un réseau maritime. Les réseaux du modèle unitarisé comportent environ 4 700 arcs et 1900 nœuds, ceux du modèle non unitarisé 2000 arcs et 738 nœuds. Ces réseaux sont intégrés et se situent tous à l'échelle européenne. Les paramètres des fonctions de coût ont été calculées dans une logique de croisement de mode x produit.

III. ORGANISATION DU RAPPORT

Le présent rapport est organisé autour de trois thématiques ;

1. Dans un premier temps, les sources de données fret sont recherchées et identifiées. Le travail a consisté à recenser le plus exhaustivement possible les informations dont on peut disposer sur le sujet. Cinq thèmes ont été définis pour chaque source identifiée : nature de l'information, intérêts pour une modélisation fret à caractère international, caractéristique de la source de données, acquisition de la source, contacts. Les éléments sous-jacents aux résultats du recensement sont alors abordés (quantité de l'information, qualité de l'information, autres sources de données, le cas des données géographiques, le progrès dans la diffusion de l'information).
2. Dans un deuxième temps, on s'interroge sur l'articulation entre mesurabilité et cohérence / pertinence dans la modélisation. On considère, à ce stade, que les sources d'information ont été identifiées et que le modèle va être développé et utilisé à partir des bases de données issues de celles-ci. On montre, alors, que les potentielles d'erreurs peuvent être nombreuses et impacter fortement les résultats. Le propos s'appuie sur plusieurs exemples concrets, parfois volontairement caricaturaux.
3. Dans un troisième temps, on essaie de « quantifier » la relation mesurabilité – cohérence en testant différentes erreurs possibles à partir d'un modèle existant. Six types d'erreurs sont simulées. Un avis est porté quant à leur impact sur les résultats de trafic. On remarque que le risque le plus important concerne sur le cumul d'erreurs.

IV. RECHERCHE ET IDENTIFICATION DES BASES DE DONNEES FRET

Dans ce chapitre, les aspects relatifs à l'identification et à la recherche des sources de données fret sont abordés. Après quelques généralités, la notion de « production » de sources de données est qualifiée, celles-ci renvoyant à plusieurs dimensions. Le résultat du travail de recherche des différentes sources de données existantes est ensuite présenté. Enfin, les éléments sous-jacents aux résultats trouvés sont abordés à travers quatre thématiques.

IV.1. GENERALITES

L'objet de ce chapitre consiste en deux points :

1. Recenser et présenter les partenaires gestionnaires et propriétaires de sources de données concernant le fret, utilisables pour des modélisations à caractère « international ».
2. Proposer un état détaillé du contenu des sources rattachées à chaque partenaire au regard de plusieurs critères

Les informations concernent les études de modélisation fret à caractère international (en Europe), pour l'élaboration de modèles plus ou moins complexe. Seule l'information qui semble utile est présentée. Il se peut, ainsi, que certains organismes n'apparaissent pas. Ceci est normal. Par exemple, en matière de projections macro-économiques spatialisées à l'échelle européenne, les experts du Cambridge Econometrics proposent des produits élaborés et adaptés à la demande, qui sont d'une finesse plus élevée que ceux du London Economics, et qui tiennent compte des logiques d'import / export par « produit » et destination (ce qui n'est pas le cas du BIPE en France). Ce dernier n'apparaît donc pas (ce qui ne veut pas dire qu'il n'a pas d'intérêt)..

IV.2. PRODUCTION DE DONNEES ET SOURCES DE DONNEES

Le terme générique « production » de données recouvre plusieurs acceptions, que l'on peut qualifier de la manière suivante au regard d'une modélisation fret ;

- La production de données **originelles**, qui correspond à la réalisation d'enquête (i.e. pour les besoins d'une étude). On classera dans cette catégorie tout ce qui est enquêtes origine-destination, comptages, enquête préférence déclarées, etc.
- La production de données **par reprise**, qui correspond au transfert, pour l'étude, d'informations issues d'une source de données existante (quelle qu'elle soit, que ces informations aient ou non nécessité une transformation simple de la part de l'éventuel organisme diffuseur). On classera dans cette catégorie toutes les données issues des organismes publics, para-publics ou privés.

- La production de données **par création**, qui correspond à l'utilisation d'une ou de plusieurs sources de données différentes ou de tout autre support, afin de proposer de nouvelles valeurs. On classera dans cette catégorie toutes les données issues d'une transformation complexe faisant appel aux mathématiques (type recherche opérationnelle), à la statistique, l'analyse des données ou l'économétrie, et ce de la part du modélisateur.

En ce qui concerne le présent recensement, le terme « source de données » renvoie à tout ce qui permet de produire des données par « reprises » ou par « création ». Il est considéré de manière large, qu'il s'agisse de données structurées, à caractère quantitatif ou qualitatif.

IV.3. PRESENTATION DES RESULTATS SOUS FORME DE TABLEAUX

Le recensement effectué s'est appuyé à la fois sur des bases documentaires et l'outil internet. **L'exercice est réalisé en relation avec une problématique de flux principalement orientée vers le Royaume-Uni.** Trente tableaux ont été dressés. Ils sont classés par série thématique :

Thème 1 : Nature de l'information.

Pour chaque source de données, le contenu de l'information disponible est présenté.

Thème 2 : Intérêts pour une modélisation fret à caractère international

Un avis est porté quant à la valeur ajoutée de la source de données pour une modélisation du type précédent. Les remarques importantes sont faites.

Thème 3 : Caractéristiques de la source de données

Ce thème regroupe les informations relatives au « profil » de la source de données : mode(s) de transport concerné, date de référence, période d'actualisation.

Thème 4 : Acquisition de la source

Ce thème regroupe les informations relatives à l'acquisition de la source de données : le type de prestations (dédiées ou normatives), le support, le coût (quand cela est nécessaire et/ou connu).

Thème 5 : Contacts

Ce thème regroupe des informations diverses sur les contacts : le site internet, la personne à contacter, téléphone et fax.

Les organismes diffuseurs ou producteurs des données qui apparaissent dans les tableaux sont les suivants :

CGP : Commissariat Général au Plan

DAEI - SES : Direction des Affaires Economiques et Internationales du Ministère de l'Equipement. Services de l'Economie et des Statistiques.

DTMPL : Direction du Transport maritime des Ports et du Littoral

SETRA : Service d'Etude Techniques des Routes et Autoroutes

DETR : Department of Transport and....

IGN : Institut Géographique National

VNF : Voies Navigables de France

CE : Commission Européenne

CEMT : Conférence Européenne des Ministres des Transports

IWW / INFRAs : Centres de recherches allemands et Suisses

ASECAP : Association des Sociétés Européennes Concessionnaires d'Autoroutes à Péages

ASFA : Association des Sociétés Françaises d'Autoroutes

Eurotunnel : Opérateur du service transmanche par le tunnel

Cambridge Econometrics : Institut de prévision macro-économique

CNR : Conseil National Routier

Ports de la Manche : France, Belgique, Hollande, mais aussi de l'Espagne

INSEE ; Institut National de la Statistique et des Etudes Economiques

SNCF : Société Nationale des Chemins de Fer

UIC : Union Internationale des Chemins de fer

Mappy, Michelin

EUROSTAT : Services des statistiques européennes

IV.4. ELEMENTS SOUS-JACENTS AUX RESULTATS

Plusieurs éléments sont sous jacents aux tableaux présentés dans les pages qui suivent.

IV.4.1. La quantité d'information disponible

Plus de 70 sources de données différentes ont été recensées. Ceci permet de fournir un aperçu général certain, qui met en évidence qu'une information importante existe et que le problème est plus souvent de la « localiser » que de l'obtenir. Il est difficile, en dehors de certains besoins très spécifiques, de ne pas trouver, en 2003, l'information souhaitée pour réaliser des études dans le transport de marchandises. Le frein le plus important est peut-être le coût dans certains cas (cf. BDD GISCO de la Commission Européenne).

On retiendra que de l'information plus précise est aussi disponible auprès d'organismes divers tels que l'Association des Utilisateurs de Transport de Fret (AUFT) en France, les exploitants ferroviaires en Europe (SB Cargo, CFF,...).

Par ailleurs, on consultera avec utilité le site d'information sur la recherche en europe CORDIS (www.cordis.lu) et en France, celui du PREDIT (www.predit.rd.fr).

IV.4.2. La qualité de l'information disponible

Il est difficile de se prononcer sur la qualité de l'information recensée. Cependant, on précisera que des sources de données importantes pour l'établissement d'un modèle comme SITRAM (flux de transports par mode x produits), COMEXT (données à caractère macro-économique), enquêtes Transit DAEI/SES (données de trafics PL (et fer sur Cd-Rom)) d'échange et de transit avec la France), CNR (coûts PL), sont de bonne facture. D'autres, comme GISCO sont très complètes, mais aussi très lourdes et certainement pas au point sur les aspects trafics. Les projections macro-économiques par branches établies par le CAMBRIDGE ECONOMETRICS s'inscrivent dans le cadre d'un modèle macro-économique éprouvé sans qu'il soit possible de juger de la qualité même du modèle (cette source a pour intérêt de ne pas laisser la responsabilité des hypothèses macro-économiques au modélisateur). Les sources statistiques des ports sont souvent décevantes, et l'information la plus importante (flux par produit et liaison) difficile à obtenir, ne serait-ce que parce que les ports ne font pas les enquêtes nécessaires. Les données de la source REGIO (caractéristiques régionales) sont souvent manquantes pour certaines zones, et ce même pour un pays comme la France ; alors que les cartes sur les trafics du SETRA sont généralement périmées a leur sortie. **On retiendra, aussi, qu'en 2002, malgré tous les projets de recherche financées par la Commission Européenne, qu'il n'existe pas, de matrices de flux de transports publiques, par mode et par zone, à l'échelle européenne (alors qu'il s'agit d'une information de source).**

IV.4.3. Autres sources de données

Les tableaux suivants ne prennent pas en compte les sources de données importantes qui sont celles des consultants. Que ce soit pour certains clients ou leurs propres besoins, ils produisent des données originelles dans différents domaines. Le domaine le plus important en économie des transports concerne les coûts du transport sous toute ses formes (différents modes) et selon différentes déclinaisons (coûts spatialisés⁴, problème de ruptures de charge, coûts de maintenance et d'exploitation selon le type de matériel,...)

D'autres type de données produites par les consultants consistent en des ratios divers et variés qu'ils utilisent dans les études : nombre de jours d'exploitation par an (différents selon les modes de transports), capacités annuelles, tonnages nettes de marchandises par mode selon le mode de transport, temps d'arrêts, de chargement et de déchargement, nombre de wagons par trains,...ces ratios sont généralement spatialisés.

Les consultants ont, en effet, accès à nombre élevés d'études ou de rapports d'études d'organismes divers qui leur confient des prestations. A ce titre, ils sont en mesure de capitaliser l'information.

IV.4.4. Les cas des données géographiques

Compte tenu des sources de données à disposition, il est dorénavant possible de présenter les résultats, ou des données d'entrée pour la modélisation grâce aux Systèmes d'Information Géographiques, qui ont aussi pour particularité de gérer l'information.

On insistera plus particulièrement sur les données issues de la source AND, et qui concernent l'Europe entière. Chargeable sous SIG on peut obtenir les différents types de réseaux (différents niveaux à disposition pour la route), ainsi les délimitations administratives, ou la localisation de certains éléments comme les aéroports, les villes,... Des fonctionnalités peuvent alors être utilisées pour le calcul d'isochrones ou de temps de parcours,...

Le coût de ces sources est cependant difficilement amortissable sur une étude (il faut compter de l'ordre de 2 400 euros pour les réseaux routiers de niveau 2 à l'échelle européenne et 530 euros pour l'abonnement annuel concernant l'actualisation des données).

Dans le domaine de « l'illustration » des études, la Commission européenne a mis récemment à disposition sur son site Web les cartes des réseaux trans-européens par mode et par pays, ainsi que les cartes des ports et aéroports. On consultera avec profit l'adresse suivante :

<http://europa.eu.int/comm/transport/themes/network/french/ten-t-fr.html>

⁴ C'est-à-dire qu'ils peuvent être différents selon la zone géographique retenue (exemple : traversée de frontière ou rase campagne,...).

IV.4.5. Progrès dans la diffusion de l'information

On constate, depuis quelques années un progrès considérable dans la diffusion de l'information. Ce progrès concerne les données des institutions publiques, c'est-à-dire hors entité type Grandes Entreprises Nationales, ces dernières ne souhaitant pas pour des raisons commerciales mettre à disposition une grande partie de leurs données. L'existence des supports informatiques à grande capacité que sont les CD-Rom, ainsi que la mise en place de sites internet performants permettant des téléchargements en ligne ont largement favorisé cette diffusion.

Ainsi, le Ministère de l'Equipeement, via le service statistique (SES) de la Direction aux Affaires Economiques et Internationales, publie plusieurs CD-Rom importants pour les études transports : les enquêtes dites « Transit », les statistiques sur les transports en France (cf. tableaux). On peut regretter que les informations sur les ports (DTPML) ne soient que sur support papier, mais sur simple demande téléphonique il est généralement possible d'obtenir quelques fichiers informatiques via courriel.

L'autre progrès important est qu'il est dorénavant possible d'effectuer ses recherches de données quasi-exclusivement via-internet. Les sites sont nombreux et les liens entre sites se sont fortement développés (par exemple : le site de l'INSEE renvoie à tous les Instituts Nationaux de Statistiques Européens).

En ce qui concerne le coût de l'information, on retiendra que celle-ci correspond généralement au coût de mise à disposition uniquement (traitements voir développements correspondants). Pour certains services dédiés, il convient de prévoir des coûts assez élevés du fait qu'il s'agit de réelles prestations intellectuelles (sorties des modèles macro-économiques).

THEME 1 : NATURE DE L'INFORMATION

[Tableaux 1a à 1f]

**THEME 2 : INTERETS POUR UNE MODELISATION FRET A CARACTERE
INTERNATIONAL**

[Tableaux 2a. à 2f.]

THEME 3 : CARACTERISTIQUES DE LA SOURCE DE DONNEES

[Tableaux 3a. à 3f.]

THEME 4 : AQUISITION DE LA SOURCE

[Tableaux 4a. à 4 f.]

THEME 5 : CONTACTS

[Tableaux 5a. à 5f.]

V. ARTICULATION ENTRE MESURABILITE ET COHERENCE / PERTINENCE DANS LA MODELISATION DES FLUX DE MARCHANDISES

L'étude de l'articulation entre mesurabilité et cohérence / pertinence dans la modélisation des flux de marchandises, renvoie à diverses questions qu'il est nécessaire de se poser à différentes étapes du travail. Celles-ci peuvent être rattachées aux étapes suivantes de la production de données:

- identification des sources de données existantes, et mise en œuvre,
- validation des sous-sources créées,

Rappelons que la production de données renvoie à plusieurs notions. Dans tous les cas, le produit « fini » est une information généralement quantitative.

V.1. IDENTIFICATION DES SOURCES DE DONNEES EXISTANTES ET MISE EN OEUVRE

L'identification de sources de données existantes passe par plusieurs étapes, selon le processus évoquée au chapitre précédent :

1. Recherche auprès de la Maîtrise d'Ouvrage. Cette dernière dispose en effet souvent d'informations sur le projet à étudier. Les informations à caractère commercial sont généralement les plus intéressantes, car elles s'appuient sur des données de trafic (enquêtes ou estimations).
2. Recherche auprès des services statistiques des Ministères des pays concernés.
3. Recherche auprès des autres organismes (Cf chapitre précédent)

Une fois ces étapes passées, la modélisation est construite, étape par étape. Dans les sections qui suivent, le cas des phases de génération / distribution et choix-modal / affectation est abordé à partir d'exemples concrets, qui servent de base de discussion x démonstration des aspects relatifs à l'interface mesurabilité x modélisation.

V.1.1. Génération /distribution ⁵

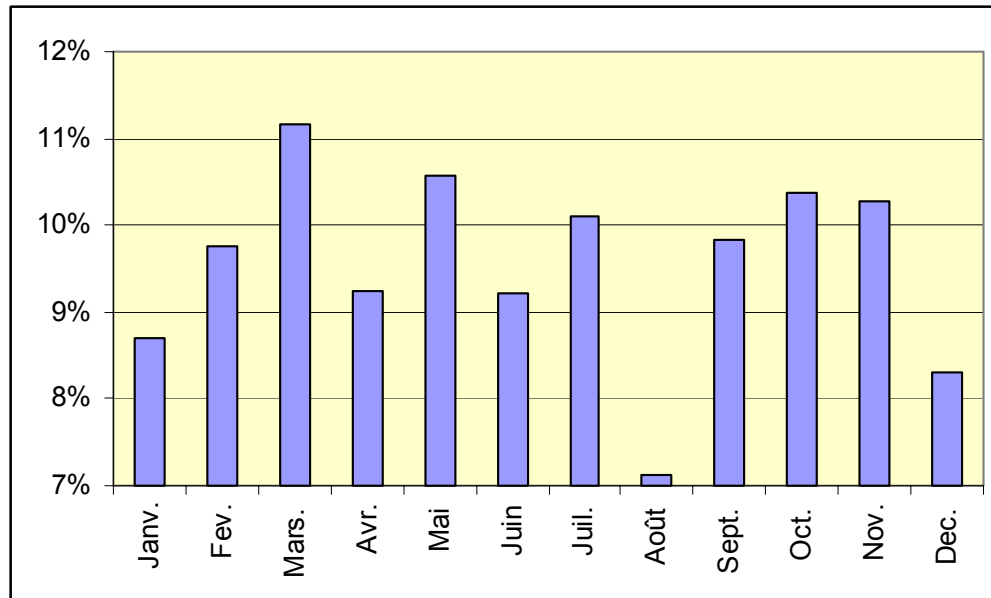
Ces données sont les fondements d'une modélisation. L'idéal est de développer la modélisation à partir de données d'enquêtes origine-destination « ad-hoc ».

Cependant, il peut-être nécessaire de prendre en compte la saisonnalité des trafics, afin d'analyser les problèmes liés aux « pointes ». Une solution alternative à la réalisation de plusieurs enquêtes est d'effectuer une enquête sur une période donnée de l'année, choisie judicieusement, et d'utiliser les données de trafic sur section (comptages poids lourds, par exemple) pour compléter l'analyse (estimation des trafics en période de pointe).

⁵ La phase de distribution est considérée à la fois en V1.1. et en V.1.2.

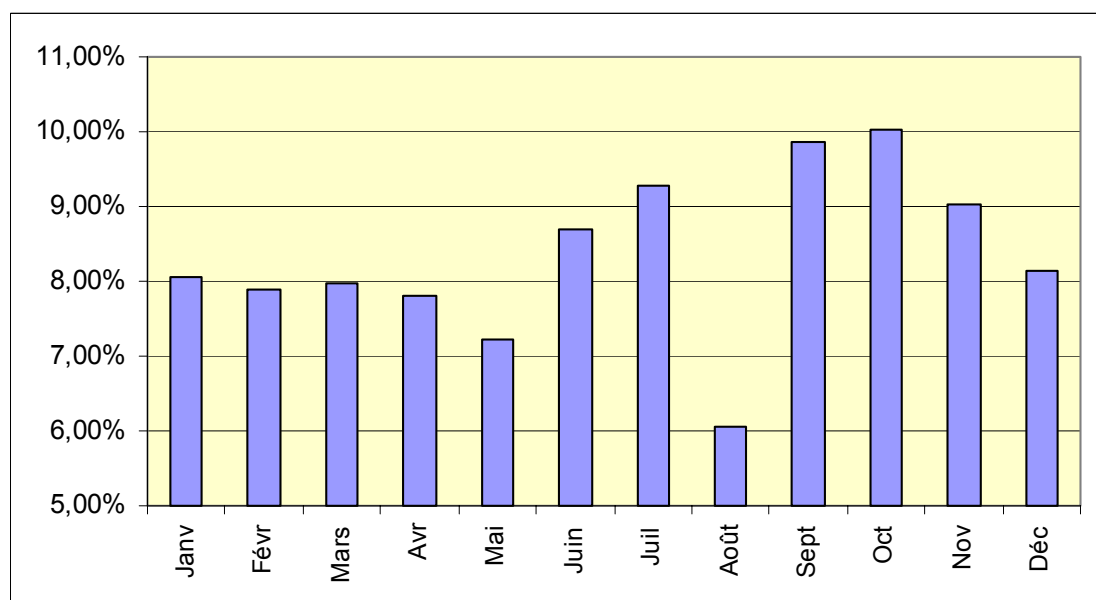
Les figures suivantes montrent qu'en transport de marchandises, la distribution des trafics est bien différente selon les périodes de l'année. Cette distribution peut avoir un intérêt, par exemple, pour la mise au point d'une tarification d'usage de l'infrastructure différenciée selon les périodes de l'année.

Figure 1 : Distribution annuelle des poids lourds, arc Fréjus-Brenner, année 2002 (poids lourds comparables). Total = 100%



Source : ISIS, à partir de données SFTRF, DETEC et Ministère de l'économie et des finances autrichien

Figure 2 : Distribution annuelle des tonnages ferroviaires, façade Méditerranéenne des Pyrénées, année 2002". Total = 100%



Source : ISIS, à partir des enquêtes des « Flux terrestres de marchandises en transit ou en échanges à travers les alpes françaises et les Pyrénées en 1999 ». MELT.

Ce type d'enquête est d'autant plus intéressante qu'elle permettra de définir correctement le zonage. Ce dernier doit en effet assurer un bon compromis entre taille des zones (pas trop grandes) et flux émis (qui doivent être assez importants).

Si l'on ne possède pas d'enquête origine-destination, où que le client ne juge pas opportun d'en réaliser, il est nécessaire de s'appuyer sur les bases existantes. Dans le cas d'une modélisation fret nationale (France), SITRAM est une très bonne base de données ⁶. Cependant, pour l'estimation des flux OD à caractère internationaux, cette base est moins performante. On lui préférera la base de données des enquêtes dites « transit » du Ministère de l'Équipement (1992/93 et 1999). Ces dernières existent aux barrières pyrénéennes et alpines, mais pas pour le nord-est de la France. Elles peuvent être utilement complétées par les bases de données suisse et autrichiennes pour ce qui concerne l'arc alpin.

Si l'on prend le cas des flux transmanches, des enquêtes OD existent effectivement côté Royaume-Uni, mais celles-ci ne sont pas récentes (DETR, 1996). Or, comme pour toute enquête à disposition, des tests doivent être effectués afin de vérifier la fiabilité de l'information. Dans le cas présent, l'enquête a été réalisée par envoi de courrier, principalement à l'attention de chargeurs anglais, alors même que l'île est plus importatrice qu'exportatrice. Le faible nombre de réponses (7500) fait que toute désagrégation pose problème. Ainsi, si les flux pays-pays peuvent être considérés comme assez fiables, ce n'est plus le cas dès que l'on se situe à un niveau zonal plus fin. Cependant, la répartition entre produits dits « unitarisés » et « non unitarisés » a un intérêt indéniable, car il s'agit d'une variable de stratification importante, qui peut-être utilisée de manière indépendante de tout affinage géographique (on la retient dans une approche pays-pays) ⁷.

La seule alternative, face à des enquêtes de ce type, consiste en l'utilisation des données de la base COMEXT (Eurostat) sur les échanges extérieurs entre pays. Cependant, si ces données peuvent être très détaillées (classification NST en 9, 52 et plus de 200 postes), elles ne fournissent pas d'information sur les destinations finales des produits et posent quelques problèmes de codage sur les modes de transport (surtout le Ro-Ro). Elles ont cependant l'intérêt d'apporter un éclairage important sur la dynamique des flux car elles peuvent être étudiées sur « longue » période.

L'existence de données pays-pays, par produit, pose un problème important pour la modélisation des phases de génération / distribution. Il faut en effet trouver les variables explicatives de la dite génération / distribution des flux, permettant de spatialiser les données (éclatement des données pays-pays en données zone-zone).

⁶ SITRAM fournit des origine-destination à l'échelle des départements, par mode, par nature de marchandise, par type de véhicule et de conditionnement. SITRAM s'appuie à la fois sur les données issues des enquêtes TRM (Transports Routiers de Marchandises) et les transports internationaux (Douanes) et sur le fichier « rail » fourni par la SNCF (le transport de colis de détail de la SERNAM est exclu).

⁷ Les produits unitarisés correspondent à tous les produits « conditionnables », c'est-à-dire transportable par camions ou par transport combiné, à la différence des produits non unitarisés.

Ce qui est « évident » en transport de voyageurs longue distance (population), l'est beaucoup moins dans le cas des marchandises. Eurostat (REGIO) peut fournir, sur la base d'un zonage NUTS, des caractéristiques de zones (population, valeur ajoutée marchande, valeur ajoutée non marchande, valeur ajoutée industrielle, emplois par secteurs, superficies agricoles selon cultures, etc.). Le problème est que la plupart de ces données sont souvent mal renseignées.

La recherche des variables explicatives peut donc s'effectuer à partir de données SITRAM pour la France, qui fournit des flux OD département / département que l'on croise avec les données NUTS2 d'Eurostat (pour les mêmes départements).

Le travail consiste à établir des corrélations entre niveaux de trafics et données potentiellement explicatives de flux de trafics. Il est généralement possible de mettre en évidence des corrélations, que l'on peut considérer comme des liens de causalité dans certains cas.

L'absence de corrélation, permet de mettre en évidence l'existence d'un facteur important de la génération des trafics (en origine généralement, mais quelquefois aussi en destination) : la localisation des activités.

Les analyses montrent que la localisation des activités est le facteur explicatif principal des flux de marchandises pour les produits suivants ⁸:

- Les automobiles
- Les produits chimiques
- Les produits métallurgiques
- Les produits non périssables
- Les verres / verreries
- Les papiers /cartons
- Les minéraux (aussi vrai à la destination)
- Les matériaux de construction

Le problème consiste donc en la manière de « quantifier » cette localisation. Trois sources peuvent être mises à contribution : l'annuaire KOMPASS ; l'annuaire EUROPAGES et l'annuaire géographique d'EUROSTAT. Les deux premières sources sont consultables via internet. Les sites de production recensés peuvent alors être rattachés aux zones du découpage et les données de trafic Sitram corrélées à la quantité de sites identifiés. Ensuite, l'analyse est généralisée sur la zone d'étude. Ce travail est particulièrement long et fastidieux, mais il permet d'obtenir des résultats utilisables pour la modélisation.

⁸ Les éléments sont fournis, ici, de manière agrégée

En guise de conclusion, le tableau suivant fournit une appréciation du type de données à utiliser pour la phase de génération / distribution dans le transport de marchandises.

+++	Enquêtes origine-destination « ad-hoc »
++	Enquêtes origine-destination in situ menées par les Etats
++	Enquêtes origine-destination par courrier (sous réserve d'une validation par tests)
+	Données Eurostat spatialisées avec recherche des variables explicatives de la spatialisation

V.1.2. Distribution, partage modal et choix d'itinéraire

La demande par produit va se décliner sur différents modes de transport, qui choisiront ensuite un itinéraire déterminé. Le choix de mode x itinéraire peut-être réalisé de manière combiné ou hiérarchique, selon l'architecture du modèle mis au point.

Dans cette sous-section nous abordons deux aspects, l'un concernant la segmentation de la demande et la désagrégation des flux (zonage), l'un concernant les fonctions de coûts. A chaque fois des exemples simples sont pris pour mettre en évidence les éléments clefs sous-jacents.

V.1.2.1. Segmentation de la demande versus flux en jeu

La difficulté principale réside dans la manière dont la demande est segmentée : mode et/ou produits. Deux approches sont généralement valorisées dans les fonctions de coûts / utilité : préférences révélées ou préférences déclarées. Mettons pour l'instant de côté ces dernières, car elles renvoient à des problématiques particulières difficilement généralisables⁹, et plaçons nous dans le cadre d'une modélisation classique, dans laquelle le « coût généralisé » détermine le choix de mode x itinéraire.

La segmentation de la demande doit être mise en relation avec le découpage retenu. Dans le cas, a priori simple, d'un modèle à 30 zones, 9 classes de produits (ex : NST), et 4 modes de transports, 31 320 flux sont à prendre en compte. Avec 200 zones, c'est 1 432 800 flux qui sont à considérer. Comme les données pour les zonages les plus fins sont obtenus généralement à partir d'un « éclatement » de données de zones plus grossières sur la base des niveaux de population (par exemple), on apprécie aisément les problèmes de « fiabilité » sous-jacents, qui ne font que s'amplifier en projection.

On retrouve ici deux écoles de planificateurs : les adeptes des découpages plutôt « agrégés », et ceux des découpages « fins ». Les premiers considèrent qu'il vaut mieux raisonner sur les OD les plus importantes, quitte à traiter les autres « à part ». Les seconds préfèrent ne pas sacrifier le niveau de finesse qui peut avoir une utilité en matière de mise en évidence de la spécificité de certaines zones, mais aussi dans le détail des caractéristiques de l'offre (de zone à zone).

⁹ Nous abordons le cas des enquêtes SP sous un angle encore peu étudié en V.2.2. alors qu'il est déterminant en matière de résultats.

Le tableau suivant, présente un croisement correspondant au nombre de zones d'un modèle avec le nombre de produits pris en compte pour 4 modes de transports. Les chiffres dans chaque case correspondent au nombre de flux concernés, sachant que le résultat final attendu est un trafic sur section :

Zones Produits	30	50	70	90	110	150	200	250
4	13 920	39 200	77 280	128 160	191 840	357 600	636 800	996 000
9	31 320	88 200	173 880	288 360	431 640	804 600	1 432 800	2 241 000
12	41 760	117 600	231 840	384 480	575 520	1 072 800	1 910 400	2 988 000
20	69 600	196 000	386 400	640 800	959 200	1 788 000	3 184 000	4 980 000

On remarque que le risque de mauvaise représentativité croît fortement avec le nombre de zones et de produits pris en compte.

V.1.2.2. Composition des fonctions de coût

Les fonctions de coût vont permettre de reconstituer le partage modal en année de base, et seront utilisées en simulation / projection.

Le principal problème, en terme de mesurabilité, concerne la « plausibilité » de ce que l'on mesure , donc le contenu même des fonctions de coûts.

Rappelons en effet, que les raisonnements en transport de marchandises concernent le coût du transport, alors que la pratique conduirait plutôt vers des raisonnements en terme de prix du transport (payé par le chargeur). Or, la notion même de prix est difficilement appréciable surtout pour le transport ferroviaire (pour lequel le prix peut d'ailleurs être inférieur au coût dans certains cas).

Pour ce qui est de la route, les principaux postes de coûts sont les coûts variables et les coûts fixes, qui peuvent être différents selon les produits, via le type de camion apte les transporter. Les données du CNR sont tout à fait adaptées pour définir ce type de coût. Pour le rail, il n'existe pas de source « officielle » pour le calcul des coûts, mais les consultants ont pu établir des bases de données en accumulant de l'information dans le temps sur ce thème.

Que ce soit pour la route ou pour le rail, la construction d'une fonction de coût ne saurait se limiter à la prise en compte de ces postes uniquement, dans la mesure où cette approche est éloignée de la réalité. Les logiques « économiques » doivent être mieux intégrées qu'elles ne le sont dans les études de trafic. Des enquêtes menées par ISIS sur l'articulation entre structure des coûts et choix de mode x itinéraire, il ressort que le poste « coût de personnel » doit être isolé dans la fonction de coût ¹⁰ sur la base d'une journée de travail (sachant que tout dépassement de cette durée se traduira par un surcoût correspondant à un équivalent journée supplémentaire, puisque le conducteur est immobilisé ¹¹).

¹⁰ Surtout pour le transport routier, mais ceci doit aussi être effectué par soucis d'homogénéité pour le transport ferroviaire (afin de d'apprécier les logiques concurrentielles)

¹¹ Cet aspect est important, puisque l'une des remarques faites les chargeurs-transporteurs de l'association TLF su le non attrait de la liaison de ferroustage Aiton-Orbassano, et que, par rapport à un trajet par route, le conducteur sera immobilisé par l'obligation de respecter la fin de la longue pose réglementaire de 9 heures, une fois le parcours en ferroustage réalisé.

Le coût du péage doit être pris en compte à part dans la fonction et pas mélangé dans le coût variable. (car des sensibilités sur ce poste peuvent être devoir testées).

Le dernier élément important constitutif d'une fonction de coût est le « coût d'opportunité ». Cette notion est mise en évidence par les professionnels lors des entretiens. Ce coût est lié au temps, mais il ne s'agit pas d'une « valeur du temps » au sens traditionnel du terme, mais du coût lié à la non disposition d'un bien car il est transporté. Par exemple, un transporteur/ chargeur cherchera à transporter le plus rapidement et de manière la plus fiable possible un bien dont la valeur ajoutée est élevée. Si l'on se place du côté de la demande, un client se sentira plus pénalisé si sa voiture neuve n'est pas livrée à temps que si le retard concerne des biens alimentaires (pour lesquels il y a profusion d'offre et/ou substitution, et une grande disponibilité). C'est en ce sens que l'on parle de coût d'opportunité : le transport a une importance relative plus forte pour certains types de bien ¹². Ce type de coût peut être apprécié à l'aide des bases de données Eurostat, sur la base du rapport valeur / tonnes, et pris en compte dans une logique de pondération du temps de trajet OD (comme une valeur du temps).

Le dernier élément est, bien entendu, que le coût kilométrique variera en fonction de la distance, puisque les coût fixes s'amortissent mieux plus le déplacement est long (par exemple, les transporteurs routiers ne refusent pas le niveau des péages aux tunnels du Mont-Blanc et du Fréjus, mais sont très sensibles aux augmentations de ces péages du fait que l'essentiel des trafics sont de courtes à moyenne distance (le transit ne représente que 30% des trafics)).

Les éléments précédents montrent que deux approches sont possibles. Prenons un exemple à **caractère illustratif**, d'un choix d'itinéraire routier avec application de la formule dite « d'Abraham » ($T1/T2 = (C2/C1)^n$ avec $n=10$), et pour deux relations de point de choix à point de choix. L'une gratuite, l'autre payante. Deux approches peuvent être développées :

1. Une approche traditionnelle (établissant une fonction de coût type circulaire n°98-99 du 20/10/98, à laquelle on rajoutera éventuellement un « malus » de calage),
2. Une approche dite « économique », dans laquelle on raisonne sur la base des éléments de coût fixe, de coût variable, de coûts salariaux et de coûts d'opportunité. Dans ce cas il existe deux façons de procéder : soit on essaie de recalculer la valeur n de la fonction d'Abraham (si l'on possède un nombre suffisant de points de choix pour avoir une bonne significativité statistique), soit, on recherche la valeur du Malus qui permet de reconstituer le choix d'itinéraire.

Supposons que l'on s'intéresse à des camions chargés de 16 tonnes nettes en moyenne, qui effectuent un parcours de 900 km, et qui transportent des pièces détachées ou des produits non périssables. Ils, sont, à un moment de leur parcours, confrontés à un choix d'itinéraire entre une route nationale à deux fois deux voies gratuite, sur laquelle ils

¹² On comprend mieux, ainsi que le transport puisse jouer un rôle de « stockage temporaire » pour certains produits en vrac (exemple : transport maritime et voie d'eau).

roulent à 70 km/h de moyenne, et une autoroute à péage où ils roulent à 90 km/h de moyenne et qui leur coûtera 0,11 Euros du km.¹³

Le tableau suivant présente une simulation de l'impact d'une variation du péage de +10%, sur le choix d'itinéraire, autour de trois cas théoriques. L'impact est mesuré en terme d'élasticité (réduction du trafic PL autoroutier) :

- ♣ Cas 1 : fonction de coût « traditionnelle » selon circulaire des routes, avec introduction d'un malus
- ♣ Cas 2 : fonction de coût « économique » pour un camion transportant des pièces détachées, avec introduction d'un malus
- ♣ Cas 3 : Identique au cas 2, mais le camion transporte des produits non périssables

Tableau 1 : Test illustratif simplifié, fonction de coûts

	Cas 1	Cas 2	Cas 3
Formule de coût	Traditionnelle	Economique	Economique
Type de bien	Non défini	Pièces détachées	Produits non périssables
n	10	10	10
Malus route (fr/h)	18	23,5	33,5
Elasticité au péage (trafic autoroute)	-1,13	-1,15	-1,18

Le résultat de ce petit exercice montre que l'élasticité du choix d'itinéraire du trafic ne sera pas la même selon les cas, conduisant à des résultats d'affectation sur l'autoroute différents de plus de 5%. Par ailleurs, on notera que l'utilisation d'une valeur de n différente dans la formule d'Abraham (plus faible ou plus forte) conduira à des sensibilités beaucoup plus importantes.

La composition des fonctions de coût est donc un élément essentiel sur lequel on ne s'est certainement pas assez penché dans les études de trafic x revenus, tout comme celle de la « valeur standard » utilisée dans la formule d'affectation d'Abraham, toujours prise « par défaut ».

En matière de « mesurabilité », on ne serait donc recommander que plus d'attention sur ce thème.

¹³ On présuppose que l'on connaît le choix d'itinéraire pour l'exercice. La valeur du « malus » route sert à retrouver ce choix.

V.2. PROBLEMES DE VALIDATION DES NOUVELLES BASES

Les bases de données consolidées, le modélisateur est confronté à la validation de celles-ci. La notion de validation renvoie à celle de fiabilité : quel usage pour le processus de modélisation. Derrière ce thème ce cache bien évidemment l'impact attendu sur les résultats finaux (trafic sur section et éventuellement revenus).

Nous nous attardons, dans cette section, sur trois types de démarches de validation, en fournissant dans tous les cas des exemples pratiques.

V.2.1. Fiabilité du recueil de l'information

C'est la première forme de fiabilité, lorsque des enquêtes sont réalisées.

La première attention à apporter est celle relative au taux de sondage, sur laquelle nous ne nous attarderons pas car elle est généralement bien maîtrisée par les sociétés d'enquêtes ¹⁴.

L'attention doit surtout porter sur la réalisation même de l'enquête. Lorsque le plan d'enquête a été mis au point. Concentrons nous sur deux aspects :

- ♣ **Le contenu du questionnaire et sa compréhension.** Il est important que le questionnaire soit tout d'abord rédigé par un expert du domaine. Ensuite, la rédaction du questionnaire sera reprise par la société spécialisée et le bouclage se fera avec l'expert.
- ♣ **Le déroulement des enquêtes.** La présence impromptu d'un expert lors de la passation des questionnaire est un élément fondamental, trop souvent négligé. Vérifier la façon dont l'information est codée par l'enquêteur est très important. La transcription des réponses peut en effet souvent être mal réalisée, pour des raisons de mauvaise compréhension, de délais, mais aussi quelquefois, de négligence. En terme d'exemple concrets auxquels on a pu assister on pourra citer dans le cadre d'enquêtes internationales, avec un questionnaire rédigé en plusieurs langues : une modalité mal traduite et inversée, des questions sur les tarifs (souvent importantes) qui s'avèrent inadaptées car la notion de tarifs utilisée n'est pas en phase avec celle du marché. Apprécier le déroulement d'une enquête permet de mieux faire face à d'éventuels résultats aberrants. Prendre, au hasard, une pile d'une centaine de questionnaire remplis par les enquêteurs avant codage de la société d'enquête, et les étudier est souvent un très bon indicateur.

Ce dernier point conduit à s'attarder sur des pratiques peu à la mode en France, mais qui tendent à se développer, alors qu'elles sont très utilisées par les anglo-saxons : les enquêtes préférences déclarées. Ces enquêtes posent en effet un problème crucial relatif à l'articulation entre données d'enquêtes et modélisation (des préférences déclarées). Cet aspect est traité dans le point suivant.

¹⁴ On suppose, aussi que la demande « cible » est bien définie et stratifiée.

V.2.2. Fiabilité des réponses des enquêtes

Les enquêtes préférences déclarées sont un bon exemple de l'articulation mesurabilité / cohérence interne. Lors de ce type d'enquête, on demande aux personnes interrogées (chargeurs, transporteurs, logisticiens par exemple), de se prononcer sur un ensemble de caractéristiques relatives à un nouveau service ou à un nouveau mode de transport par exemple.

En ce qui concerne l'introduction d'un nouveau service / mode de transport, il existe, essentiellement, deux façons de l'aborder :

- En se basant sur les caractéristiques d'une offre existante, que l'on modélise, et que l'on améliore pour simuler un niveau de « qualité » qui serait celui du nouveau service x mode. On regarde, ensuite, l'impact correspondant.
- Réaliser des enquêtes préférences déclarées, dans lesquelles on « simule » le choix des individus au regard de différents modes (dont un nouveau mode) ayant des caractéristiques diverses. Plusieurs possibilités de choix sont offertes afin d'obtenir de la « variance », qui servira à établir le modèle (fonctions d'utilité estimées à partir du maximum de vraisemblance).

La question clef concerne la transcription des résultats issus des enquêtes de préférences déclarées dans la modélisation, et plus précisément la pondération des réponses. Or, curieusement, cet aspect n'est pas traité dans la littérature existante. L'application de sets de pondération des réponses différents, **peut conduire à des écarts sur les résultats de trafic allant de 1 à 2.**

Dans le cadre de travaux récents, nous avons interrogé des chercheurs au niveau international, pour apprécier le type de pondération qu'il faudrait choisir. Les recommandations des chercheurs sont que les sets de pondération fournis par les logiciels de modélisation (type Alogit) constituent des points de départ qu'il convient ensuite de « dégrader » jusqu'à se caler sur un set considéré comme le plus défavorable, mais recommandé pour la prévision de trafic, compte tenu des tendances des individus à sur-estimer leurs réactions (qui elles sont bien connues).

Le tableau suivant présente une comparaison d'un set de pondération des réponses dit « par défaut » (logiciel de modélisation) et d'un set de pondération « recommandé pour la prévision de trafic ». Celui-ci se lit de la manière suivante (deuxième colonne). Si l'individu a répondu au questionnaire de préférence déclaré qu'il utiliserait probablement le nouveau mode, alors il convient de considérer au niveau du fichier de calibrage du modèle une probabilité de 50% qu'il l'utilise dans la réalité. La probabilité associée à une personne qui a répondu qu'elle utiliserait certainement le nouveau mode est de 75%.

Tableau 2 : Quelle fiabilité accorder aux réponses issues des enquêtes SP : le cas de la pondération des réponses

	Par défaut	Recommandé pour prévision de trafic
Certainement mon mode actuel	0,10	0,00
Probablement mon mode actuel	0,30	0,05
Probablement le nouveau mode	0,70	0,50
Certainement le nouveau mode	0,90	0,75
Ne sait pas	0,50	0,05

Le set de pondération choisi se retrouvera automatiquement dans la valeur de la constante du nouveau mode (plus cette dernière sera élevée, plus il y a de chances que le set choisi ne soit pas adapté).

En conclusion sur ce point, on remarque que la validation de nouvelles bases (issues d'enquêtes SP) passe par une analyse de l'interface enquêtes x modélisation.

V.2.3. Cohérence avec le passé

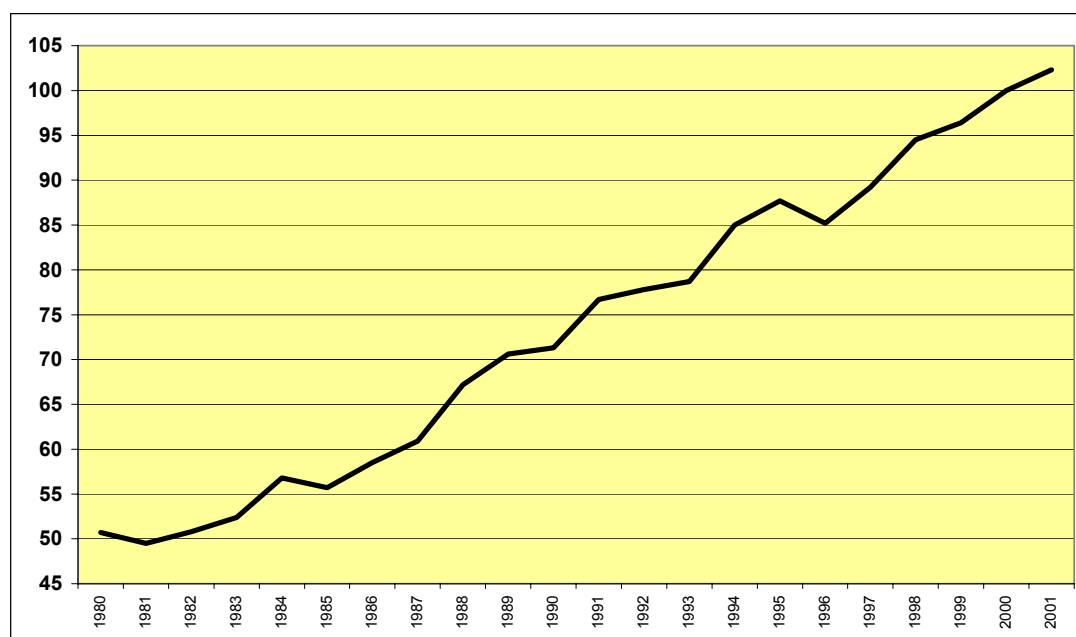
Une méthode bien connue pour projeter les niveaux de demande globale (quelle soit fret ou voyageurs), consiste à étudier ce qui s'est déroulé dans le passé, et à le modéliser. La principale méthode utilisée est celle des moindres carrées. A partir de la base de données créée, des facteurs explicatifs de l'évolution sont recherchés. Ceux-ci sont généralement d'ordre macro-économique : PIB, Commerce extérieur, etc..

Or, si il y a un domaine où les erreurs sont fatales, c'est bien celui de la projection de la demande globale. En effet, celle-ci « dimensionne » les résultats, quelle que soit la chaîne de modélisation qui est développée par la suite (choix de mode x itinéraire). Une estimation de 100, au lieu de 80, aura un impact quasi proportionnel sur les résultats. Ce constat peut aussi être vrai dans l'autre sens. En effet, il est possible que l'on utilise des données en année de base qui sont inférieures à celles qui auraient dû être utilisées (par exemple dans le cas d'un marché en forte croissance, un écart de 2 ans sur les données en année de base peut conduire à un différentiel dépassant les 15%. Celui-ci se retrouvera logiquement en projection, et ce de manière plus que proportionnelle).

Toute la difficulté réside dans les hypothèses qui seront retenues : passé versus futur. Prenons l'exemple d'une « traversée » pour laquelle des séries chronologiques sur longue période sont disponibles : la traversée alpine, s'étendant du Fréjus/Modane au Brenner.

On remarque, qu'en 20 années (1980-2000), le tonnage global est passé de 50 à 100 millions de tonnes.

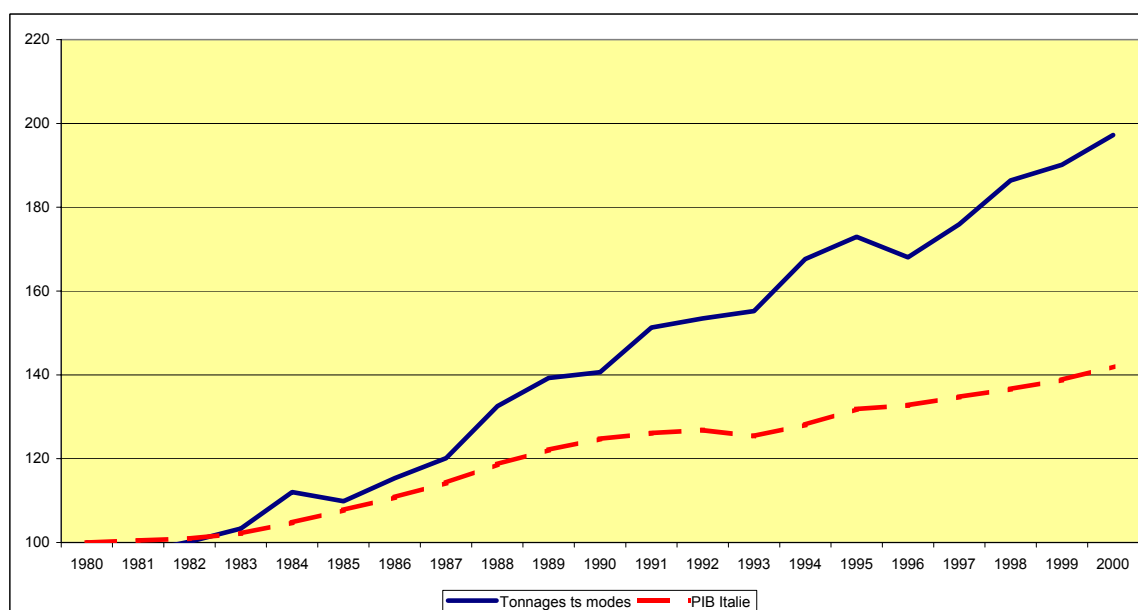
Figure 3 : Evolution du tonnage tous modes, arc Fréjus Brenner : 1980 - 2001



Source : ISIS, à partir des données du DETEC

Si l'on cherche à corréliser les niveaux de trafic avec une variable macro-économique, on obtient les résultats présentés sur le graphique suivant. On considère ici, pour simplifier l'exercice, le PIB italien (en toute rigueur il conviendrait d'étudier les évolutions à partir des données de commerce extérieur (car il s'agit d'échanges internationaux)).

Figure 4 : Croissance du PIB et croissance des tonnages tous modes, arc Fréjus Brenner : 1980 – 2001. Indice base 100 en 1980



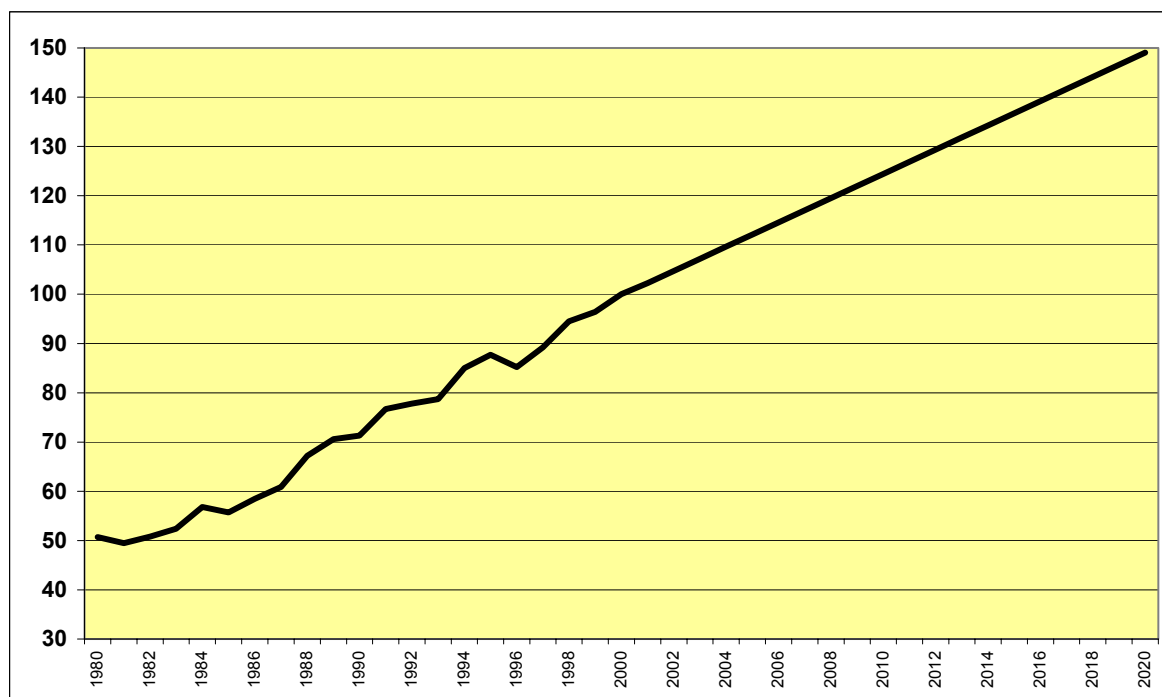
Une modélisation double logarithmique montre que « l'élasticité » sur longue période (1980-2000) est parfaitement égale à 2. Sur une période plus courte 1993-2000, l'élasticité s'élève à 1,75.

La question de fond est double :

- ♣ Elle concerne tout d'abord le niveau d'élasticité à retenir pour les projections : l'élasticité des sept dernières années est-elle d'ordre « conjoncturelle ». Sera-t-elle amenée à tendre de nouveau vers 2 lorsque les contraintes du pacte de stabilité seront assimilées par l'Italie ? A contrario ne doit-on pas s'attendre, compte tenu de la part croissante des services dans le PIB à une élasticité tendant plutôt vers 1,5 ?.
- ♣ Elle concerne ensuite la valeur du taux de croissance du PIB à retenir. Cependant, sur ce dernier point, l'analyse dans le temps montre que le taux de croissance moyen du PIB Italien a été de 1,7% par an environ entre 1980 et 2000, et qu'il a dépassé les 2% par an entre 1990 et 1998, mais a été de 1,3% par an entre 1990 et 2000. Retenir un taux supérieur à 2% pourrait déjà apparaître, pour certains, comme un pari en soit.

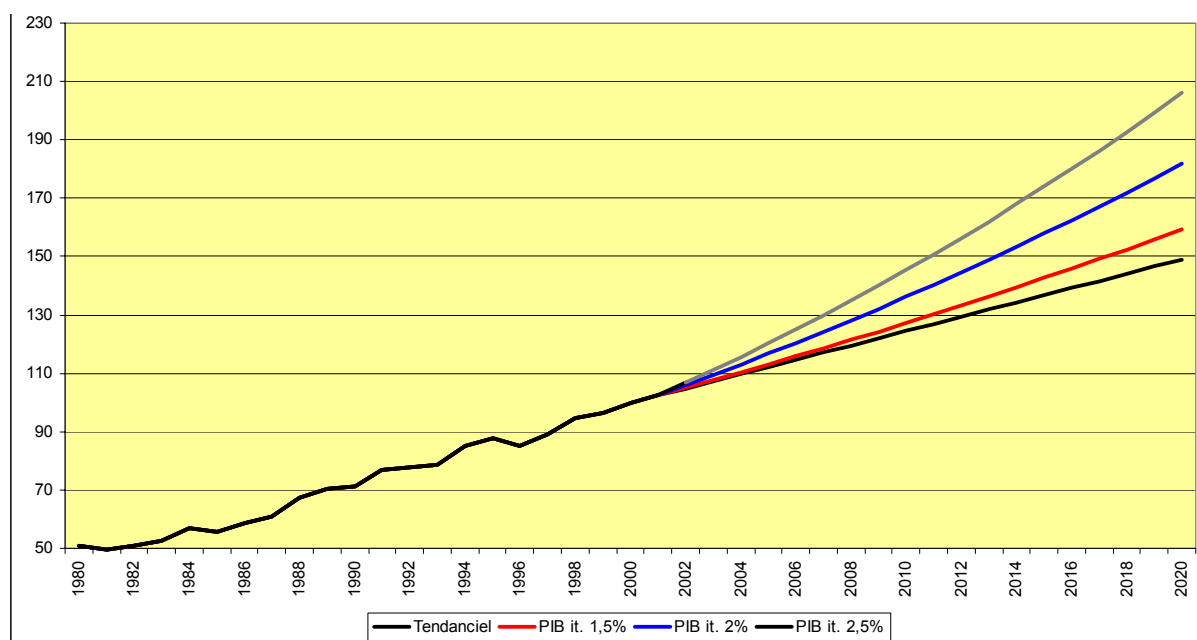
La première approche consiste donc à rechercher la valeur « naturelle » des tonnages en situation de projection. Or, les croissances de long terme dans le transport de marchandises ont pour particularité d'être « linéaires ». Le graphique suivant le montre bien. Si l'arc Fréjus-Brenner concentrerait 50 millions de tonnes en 1980 et 100 en 2000, il devrait en concentrer, toutes choses égales par ailleurs, 150 millions en 2020. Cette valeur « plancher » pour certains peut être une valeur « probable » pour d'autres.

Figure 5 : Evolution du tonnage tous modes, arc Fréjus Brenner : croissance linéaire 1980 - 2020



Si l'on poursuit l'exercice en retenant une élasticité du trafic au PIB Italien de 1,5 par an, pour différentes hypothèses de croissance annuelle moyenne du PIB de 1,5%, 2% et 2,5%, on obtient les courbes suivantes¹⁵: (tonnages tous modes sur l'arc Fréjus-Brenner) :

Figure 6 : Evolution du tonnage tous modes, arc Fréjus Brenner : quatre scénarios



On remarque que plus le taux de croissance du PIB est important, plus la courbe est concave¹⁶, et surtout, qu'avec une élasticité « prudente », les tonnages de long terme se situent dans une fourchette de 150 à 210 millions de tonnes, soit un écart de 40% !

Par ailleurs, à la visualisation du graphique précédent (étude de la pente de long terme : 1980-2020), on remarque qu'une hypothèse à 2,5% peut à l'évidence être qualifiée « d'optimiste » et que les tonnages ont une probabilité plus forte de se retrouver en réalité entre 150 et 190 millions de tonnes.

A titre illustratif, sachant que le poids du transport par rail à Modane représentait 9,4% des 100 millions de tonnes de l'arc Fréjus-Brenner, et en supposant que cette part reste constante (ce qui est une hypothèse forte puisque cette part était de 19,5% en 1980), les tonnages ferroviaires à Modane en 2020 seraient compris entre 14 et 18 millions de tonnes (soit un écart correspondant à près de la moitié des tonnages actuels, alors qu'en volume il ne passe pas plus de tonnes en 2000 qu'il n'en passait vingt ans plus tôt).

¹⁵ Application d'un modèle double logarithmique

¹⁶ Ceci est du à la formulation du modèle appliqué. Il faudrait, logiquement l'appliquer en double logarithmique sur 10 à 15 ans, puis lisser la courbe ensuite.

Deux conclusions peuvent être déduites de l'exercice précédent :

- La première concerne le positionnement que l'on doit avoir en terme d'hypothèses de taux de croissance et d'élasticité de la demande globale à ce taux. Il ne paraît pas illogique de préconiser de regarder le résultat de scénarios de projection en regard de la tendance passée de long terme. La forme des courbes au delà de l'année de base constitue en soit un bon indicateur de « fiabilisation » du faisceau. Compte tenu des enjeux financiers sur les grandes infrastructures, et compte tenu des incertitudes inhérentes à tout exercice de ce type (plus les hypothèses de croissances sont « hautes » plus on effectue un « pari »), on réduit le risque en se plaçant sur les parties basses des courbes. C'est ce type d'approche que retiendra tout investisseur privé. Une autre possibilité, qui n'est pas antinomique avec la précédente, est d'appliquer la méthode de Monte-Carlo sur la base d'une moyenne raisonnable (qui peut-être celle constatée sur le long terme) et d'un écart type de l'ordre de 10%.
- La deuxième consiste à faire travailler les spécialistes du transport et ceux de la macro-économie. L'application de modèles macro-économiques permettant de fixer les hypothèses de croissance par produit peut être intéressant. Les coefficients déformateurs des matrices de produits dans le temps sont alors fournis par les macro-économistes à l'issue d'une étude et utilisés par les experts transports dans leur travaux. Plutôt que de fixer des hypothèses, l'idée consiste à chercher à les bâtir avec des méthodes rigoureuses.

V.3. CONTRAINTES FONCTIONNELLES ASSOCIEES AUX BASES DE DONNEES

De quoi parle-t-on ? Qu'utilise-t-on ?

Lorsque l'on traite de problématiques transport, voici le genre de questions qu'il faudrait systématiquement se poser, et ce à toutes les phases de l'analyse.

Dans cette section nous allons nous intéresser aux « contraintes fonctionnelles » des bases de données. Cette terminologie recouvre des éléments relatifs aux difficultés rencontrées dans le cadre de l'utilisation de bases de données qu'elles soient liées à des aspects :

- De compréhension : contenu de la base bien identifié ?
- Techniques (sous-estimations des travaux d'adaptation nécessaires à l'appropriation et à la mise en forme de la base de données)
- De cohérence interne de la base (contenu adapté ?)

V.3.1. Compréhension des bases de données

Ce n'est pas parce que l'on traite d'une unité que l'on est sûr quelle est « générique ». En V.2.3. nous avons ainsi évoqué les tonnages de marchandises traversant les alpes du nord.

Mais qu'est-ce qu'une tonne ?

Plusieurs définition de la « tonne » existent selon le type de base de données que l'on utilise.

V.3.1.1. Le domaine routier

En ce qui concerne le domaine routier, c'est plus précisément trois dimensions sont présentes. Celles-ci sont reprises dans le tableau suivant, qui concerne la traversée des Pyrénées (année 1999) :

Tableau 3 : Tonnages routiers, frontière pyrénéenne, selon différentes sources

	Eurostat	Enquêtes transit PL > 7,80 m	Comptages Classes 3 et 4
TMJA		13 186	13 860
Tonnage net avec emballage (poids total - poids à vide)			15,2
Tonnage net avec emballage (feuille de route)		14,24	
Tonnages hors emballages et hors échanges irréguliers	57 000 000	68 556 996	76 895 280

On remarque que le tonnage qui peut être annoncé varie entre 57 et 77 millions de tonnes, soit un écart de 35%. Ces tonnages correspondent à des réalités différentes, mais il n'est pas rare de voir citer ces derniers sans références à celles-ci.

Dans le cas d'Eurostat, il s'agit de tonnages hors emballages et hors échanges « irréguliers » (dits de nature passagères). Dans le cas des enquêtes transit il s'agit des tonnages correspondant à ceux indiqués sur la feuille de route sachant que seuls les points de passage du Perthus et du Biriadou sont pris en compte. Dans le cas des comptages il s'agit de données relevées par les sociétés concessionnaires et les DDE (tonnes nettes avec emballage correspondant au différentiel entre poids total et poids du camion à vide (tare)).

A ces tonnages en volume correspondent des tonnages moyens par camion logiquement différents. Attardons nous sur ce point, et prenons un exemple caricatural, relatif à une étude dans laquelle on ne dispose que de volumes de tonnages pour une raison x ou y ¹⁷.

Supposons que l'on retienne un chiffre supérieur à 15,2t (chiffre adéquat pour l'exercice). Augmentons ce tonnage de 15,2t par pas de 0,5t correspondant à la prise en compte de degrés d'erreurs différents.

Tableau 4 : Simulation impact tonnage moyen par camion

Avec 15,2 t	0,5	1	1,5	2
Erreur de x%	3,29%	6,58%	9,87%	13,16%
Impact année de base en t	-2 529 450	-5 058 900	-7 588 350	-10 117 800
Impact année de base en PL	-161 111	-312 278	-454 392	-588 244
Impact projection en PL 15 ans (3% par an)	-251 006	-486 519	-707 928	-916 465
% trafic actuel	-5,0%	-9,6%	-14,0%	-18,1%
Impact projection en PL 30 ans (3% par an)	-391 060	-757 980	-1 102 929	-1 427 823
% trafic actuel	-7,7%	-15,0%	-21,8%	-28,2%

On remarque qu'une sur estimation du tonnage moyen de 1 tonne (soit +6,58%) correspond à une sous-estimation de plus de 310 000 poids lourds en année de base. Ceci représente quand même à 15% du nombre de poids lourds passant au passage du Biriadou ou presque 2 fois plus de camions que ceux qui étaient passé au tunnel routier du Mont-Blanc entre janvier et mars 1999.

Dans le cas d'une projection, la valeur croît logiquement (ici 3% par an), ce qui pourra poser des problèmes non quant au poids de cette erreur (car elle représentera toujours une part faible du nombre total de camions ou de tonnages projetés), mais en terme d'impact marginal, si il s'agit de traiter de problèmes de saturation par exemple.

V.3.1.2. Le domaine ferroviaire

En ce qui concerne le domaine ferroviaire, la notion de tonne renvoie elle aussi à plusieurs dimensions qui peuvent conduire à des malentendus, ou des erreurs :

- ♣ Le tonnage brut correspond au poids d'un train, tout compris. C'est lui qui va servir à « dimensionner » l'infrastructure, compte tenu de l'importance qui va lui être accordé (il peut-être préférable de conserver une pente pas trop forte qui élimine les trains lourds car ils peu nombreux).
- ♣ Le tonnage net déduit du tonnage brut les tares ferroviaires (locomotives et wagons)

¹⁷ On utilise par exemple les données Eurostat que l'on redresse par le coefficient 1,35 permettant d'approcher des tonnages issus de comptages.

- ♣ Le tonnage net-net qui déduit du tonnage net les tares des véhicules routiers (cas du ferroutage) ou des caisses mobiles (cas du transport combiné).

Si l'on associe aux éléments précédents, la prise en compte des mesures utilisées pour les calculs à l'année, il convient d'être attentif. Les raisonnements peuvent se faire en JOB (jour ouvrable de base, généralement 250 jours par an), en combinaison JOB-PH (pointe hebdomadaire) pour la définition d'un JMA (Jour Moyen Annuel), et intégrer ou non des coefficients d'abattement intégrant la notion de « taux de remplissage ».

Les éléments précédents peuvent ne pas être neutre dans la pratique, car dans la jungle des chiffres il est souvent difficile de s'y « retrouver », d'autant plus que d'une manière générale les définitions ne sont généralement pas fournies. Ce peut-être le cas dans les rapports, mais aussi lorsqu'il s'agit d'enquêter les générateurs spécifiques de trafic ferroviaire.

V.3.1.3. Le cas des tarifs

Le problème des tarifs est surtout vrai pour le routier. La notion même de tarif est difficile à cerner, car il faut être capable de raisonner en « recette moyenne pondérée » (par la structure des trafics) et non en « tarif affiché ». Ceci est souvent moins facile à réaliser qu'on ne le croît. Si l'on arrive à passer ce cap, reste à apprécier l'impact des variations tarifaires sur les niveaux de trafic. Le problème c'est que l'on raisonne sur des moyennes (et non sur des distributions), donc que l'on peut sous-estimer ou sur-estimer cet impact. L'idéal est de segmenter la demande le plus possible pour avoir des niveaux de tarifs « moyen » par catégorie de demande. On réduit ainsi l'incertitude, puisque les classes sont plus homogènes. Le problème c'est que l'on peut commettre des erreurs car les ajustements / calage seront moins bons ou plus difficiles à réaliser, plus le nombre de catégories est élevé.

On retiendra que les erreurs liées à la « mesurabilité » du tarif sont certainement beaucoup plus nombreuses, avec un impact plus fort, dans le transport de voyageurs que dans le transport de marchandises.

V.3.2. Gestion technique des bases de données

L'appropriation et la mise en forme d'une base de données ne pose habituellement pas de problèmes particuliers, si l'on a pris soin d'apprécier la plupart des éléments évoqués jusqu'ici.

Cependant, des difficultés importantes peuvent résulter de la transformation de données pour la mise en place de réseaux de transport. Cette information peut effectivement ne pas être à disposition. Lorsque le réseau doit couvrir un périmètre important, les problèmes peuvent apparaître. On ne peut en effet se satisfaire d'arcs dont la longueur est calculée « manuellement » à partir d'une carte puis ensuite transposées sous le logiciel de modélisation avec des attributs (de capacité par exemple) affectés aux différents arcs.

Une solution alternative à la non disposition de réseaux est de récupérer ceux-ci à partir de bases SIG (SIG de la Commission Européenne par exemple). La difficulté principale est alors de transformer les identifiants SIG pour les transposer dans le logiciel de

modélisation. L'identifiant de base est le lien, puisque dans le SIG les réseaux sont décrits comme une somme de liens plus ou moins nombreux selon l'échelle retenue.

Le réseau une fois extrait du SIG se transformera en une succession de lien x nœuds intermédiaires très nombreux. Ce sont ces nœuds qui poseront le plus de problèmes car il génèrent une taille très importante de fichier, qui peut s'avérer incompatible avec la réalisation des affectations et générer des temps de calculs très élevés, voire impossibles.

Le temps nécessaire pour trouver la bonne « échelle » pour les données issues du SIG et la suppression des nœuds « indésirable » peut être élevé, voire conduire à l'échec.

V.3.3. Cohérence interne de la base

Comme nous avons pu le constater jusqu'ici, les erreurs possibles en terme de mesurabilité sont nombreuses, mais d'autres sont encore possibles.

Celles-ci concernent la cohérence interne de la base de données finale qui a été créée.

Supposons que l'on possède une base de données complète, sur les aspects demande, tarifs et caractéristiques des offres concurrentes. Le problème relatif à une unité (Cf ci-avant : la tonne ou les tarifs par exemple), peut se retrouver de manière transversale.

Le cas le plus délicat concerne certainement la segmentation de la demande, pour ce qui concerne le transport de marchandise à caractère international. On le retrouve à deux niveaux :

- Le premier concerne les déclarations faites dans les enquêtes sur la destination finale de la marchandise, point certainement encore mal apprécié. Que ce soit pour les traversées alpines, les traversées pyrénéennes ou les traversées transmanches, le poids important des trafics des régions limitrophes (toujours supérieur à 50 voire 60%), pose question (nous n'avons pas la réponse !). Le risque est bien évidemment de confondre destination intermédiaire de la marchandise et destination finale (qui est celle qui nous intéresse in fine), alors que souvent les conducteurs peuvent se limiter à mentionner la destination finale du camion et non de la marchandise.
- Le second concerne l'adéquation entre les bases de demande originelles (enquêtes, eurostat,...) et les données de matrices de produit. Ceci se complique lorsqu'il faut combiner deux sources (par exemple les chapitres STIC du Royaume-Uni et les NST utilisées par Eurostat ou Sitram), voire trois sources, si l'une d'elle est nécessaire pour la définition du zonage par exemple (produits issus d'une étude ad-hoc par exemple et mis à la disposition du consultant).

VI. LA RELATION MESURABILITE – COHERENCE

L'objet de ce chapitre consiste en la réalisation de tests permettant de simuler différents types d'erreurs relatives à la mesurabilité.

Le tableau suivant présente le type d'erreurs simulées et les tests réalisés avec le modèle d'ISIS développé sur les traversées transmanches.

Tableau 5 : Type d'erreur simulée et tests

Type d'erreur simulée	Test
a. Manque d'information aux niveaux zonale obligeant à agréger les données zonales	Prise en compte d'un zonage plus réduit : passage de 87 à 43 zones
b. Problèmes dans l'estimation des niveaux de tarifs	Prise en compte de niveaux de tarifs différents
Combinaison des deux tests précédents (a x b)	Prise en compte combinée d'un zonage plus réduit et de niveaux de tarifs différents
c. Manque d'information sur l'existence de certaines liaisons ferries	Prise en compte d'un nombre de liaisons ferries réduits
Combinaison des tests a et c (a x c)	Prise en compte d'un nombre de liaisons ferries réduits et d'un zonage différent
Combinaisons des tests a, b et c (a x b x c)	Prise en compte d'un nombre de liaisons ferries réduits, d'un zonage différent et de niveaux de tarifs différents

Dans tous les cas, les résultats des tests appliqués sur l'année 2010, sont comparés à la situation correspondant aux résultats du modèle sur la base du scénario de référence ou « fil de l'eau » (année 2010 aussi).

Ce choix est fait car il est plus illustratif de se placer en projection qu'en situation de reconstitution des trafics. On apprécie alors mieux l'incidence de l'erreur au regard des résultats qui sont ceux qui intéressent le client.

Dans un objectif d'interprétation des résultats, les « lignes écrans portuaires » du modèle ont été regroupées en quatre groupes dont la dénomination correspond au positionnement géographique par rapport à la zone de traversée « centrale » de la Manche qu'est DOCA(Douvres Calais, regroupant à la fois la traversée par ferries ou par navettes et la traversée du tunnel en transport combiné).

Le regroupement a été opéré de sorte de conserver des niveaux de part de marché en année de base, jamais inférieurs à 15% dans le cas le plus faible, et d'isoler la traversée DOCA

Tableau 6 : Groupes de ports créés pour présenter le résultat des tests

	Lignes écrans portuaires du modèle	Nombre de lignes ferries concernées départ côté anglais
Ouest proche	Sussex & Hants Bristol, W & Wales	10
DOCA (PL + TC)	Douvres Calais (Poids Lourds + Trains combinés)	1 + tunnel camions + tunnel trains
Est proche	Haven & Wash Thames & Kent	12
Est éloigné	NE & Scotland Humber	10

VI.1. TEST DE MODIFICATION DU ZONAGE

VI.1.1. Méthodologie

La modification du zonage a porté sur les principaux pays générateurs / attracteurs de trafics avec le Royaume-Uni. Le tableau suivant montre que ces pays représentent 93% du total des flux (en tonnes)

Tableau 7 : répartition des flux (imports + exports) avec le RU, année 1998 (Source Eurostat)

France	22%
Allemagne	21%
Hollande	16%
Belgique + Lux	12%
Espagne	10%
Italie	9%
Portugal	2%
Autriche	1%
TOTAL	93%
Autres pays	7%
TOTAL GENERAL	100%

La modification du découpage a consisté à passer de 83 à 47 zones, sachant que 7 zones n'étaient pas modifiées (pays de l'Est ou de faibles échanges).

Les principaux critères de regroupement étaient les suivants :

- le niveau des flux émis / reçu devait croître ou rester consistant.

- le regroupement devait être le plus important possible compte tenu des caractéristiques du pays (exemple : Italie subdivisée entre nord et sud (50% des flux dans chaque nouvelle zone), Espagne : maintien des zones de Barcelone et de Madrid, puis séparation nord / sud).

Le tableau suivant montre les modifications qui ont été apportées. Le nombre total de zones a été réduit de 43%, passant de 83 à 47 zones.

Tableau 8 : Modification du zonage réalisé pour les tests

	Zonage d'origine	Nouveau zonage	Réduction (en %)
France	16	10	-38%
Allemagne	17	11	-35%
Hollande	5	1	-80%
Belgique-Luxembourg	4	1	-75%
Espagne	8	4	-50%
Italie	8	2	-75%
Portugal	3	1	-67%
Autriche	2	1	-50%
Royaume-Uni	11	7	-36%
Danemark	1	1	0%
Grèce	1	1	0%
Zones extérieures	7	7	0%
	83	47	-43%

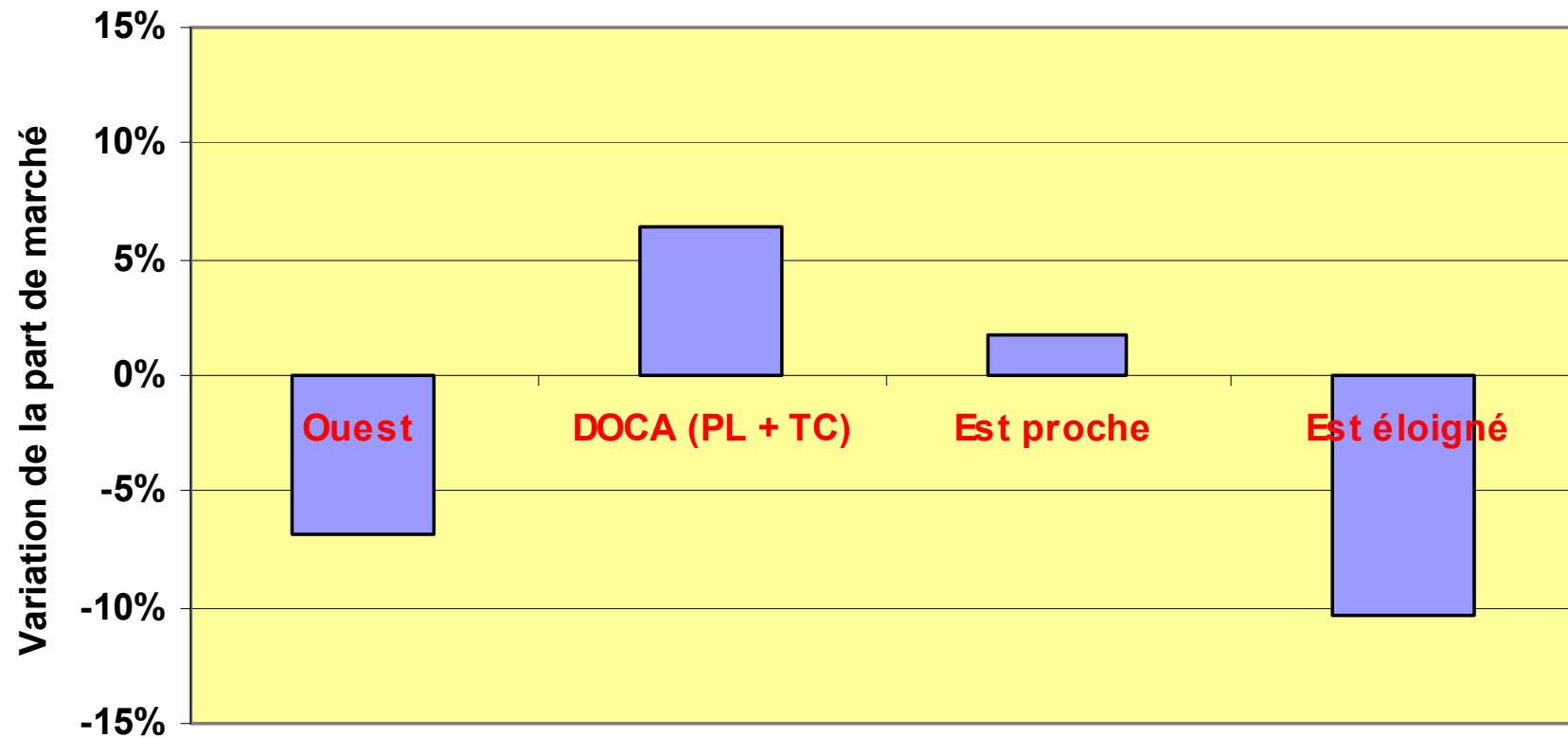
VI.1.2. Résultats

On remarque que la réduction du zonage perturbe les résultats.

La figure suivante, sur laquelle sont calculés les écarts de part de marché entre la situation de base 2010 et le test, montre que l'écart le plus important est de l'ordre de 10%.

La réduction du zonage, en modifiant les centroïdes de zone (le centroïde d'une zone regroupée avec un ou des autres se rattache à celui de la zone principale), perturbe les accès aux réseaux et modifie les logiques d'affectation (par rapport à la situation initiale). D'un point de vue pratique, on remarque que l'accès central au RU par la Manche (DOCA) se renforce au détriment du groupe de port « ouest », alors que la zone Est éloignée perd du trafic au détriment de la zone Est proche, mais aussi de l'accès central. Ceci est en grande partie due au regroupement opéré sur la Hollande et la Belgique.

Tests modification du zonage uniquement (comparaison avec la situation de base)



Ces résultats en valeur relative en cachent d'autres plus importants, en valeur absolue : les quantités de camions. On se rend alors compte que l'écart peut concerner plusieurs centaines de camions par jour. L'enjeu est relativement faible sur la traversée centrale lorsqu'on le compare au nombre total de camions, il l'est plus sur le groupe « Est éloigné ».

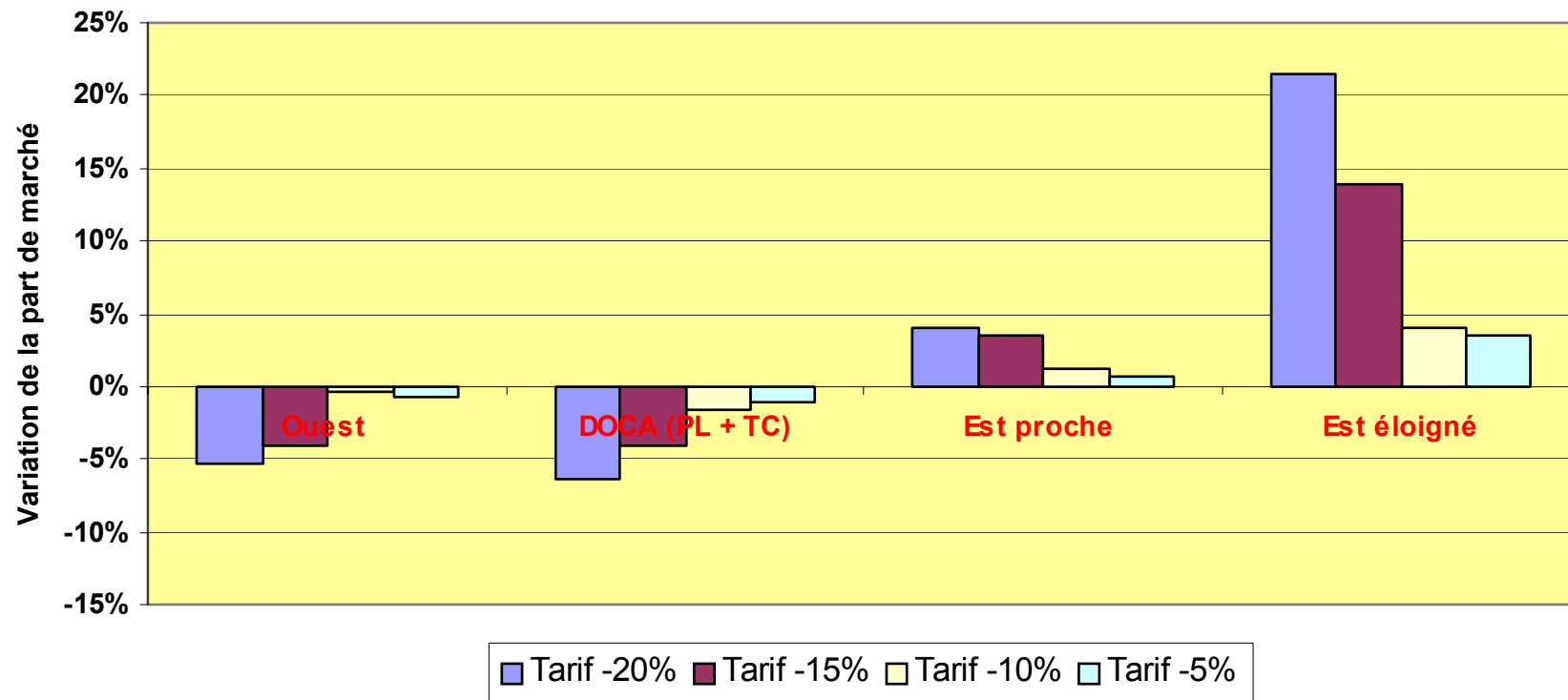
A titre conclusif sur l'effet de la réduction du zonage, il est possible de qualifier ce type d'erreur de « moyenne », ce terme générique recouvrant des différences assez sensibles selon les groupes de ports.

VI.2. TEST DE MODIFICATION DU NIVEAU DES TARIFS

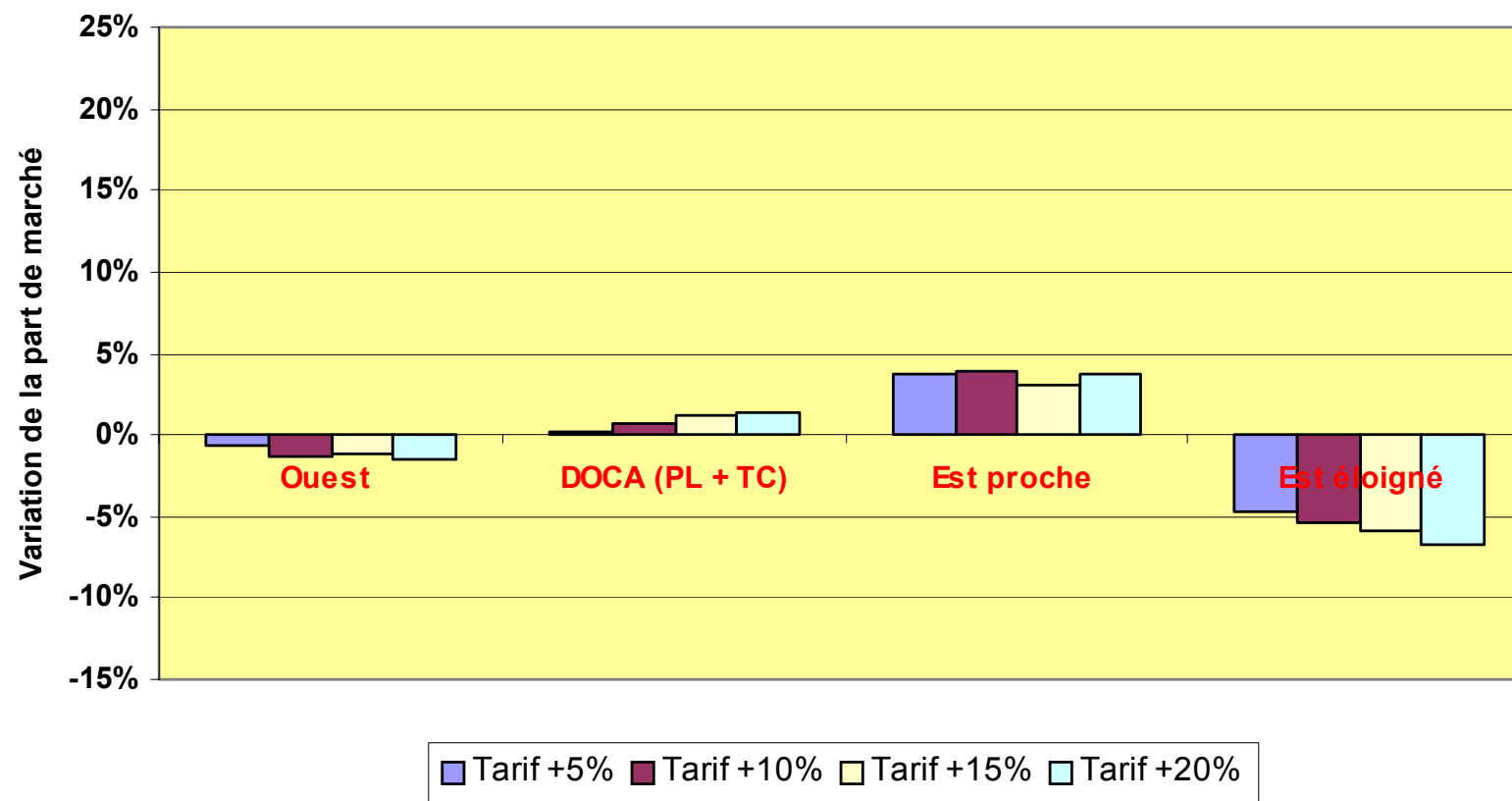
L'objectif de ce test est de voir quel est l'impact d'une erreur sur l'estimation des niveaux de tarifs. On part du principe que les tarifs qui ont été considérés pour le calage du modèle (fournit pour chaque liaison ferries et le tunnel par l'exploitant), sont fiables. Ces tarifs sont alors déformés de -20% à $+20\%$, par pas de 5 points. On considère que l'erreur sur la mesurabilité des tarifs ne peut concerner que des écarts relatifs toutes traversées confondues. En effet, la concurrence entre les services ferries conduit à une certaine homogénéité dans les niveaux de tarifs **par groupe de ports**. Par ailleurs, des tests sur un seul groupe de liaisons conduiraient au calcul d'élasticités, ce qui n'est pas l'objet de la recherche.

On remarque, à la lecture des graphiques suivants qu'une diminution de tarif favorise les groupes de l'Est au détriment de ceux de l'Ouest.

Tests de modifications des niveaux de tarifs (comparaison avec la situation de base)



Tests modification des niveaux de tarifs (comparaison avec la situation de base)



Une faible diminution des tarifs (-5%) n'a qu'un impact mineur, juste perceptible. Une diminution un peu plus importante (-10%) a un effet qui reste somme toute négligeable. Une diminution de -15%, puis de -20% qui correspondent à des degrés d'erreurs qui deviennent importants se traduit par une **perturbation du processus d'affectation**. En effet les niveaux de tarifs en valeur absolue sont différents, et une baisse importante en valeur relative entraîne des réductions qui font que des « effets de seuils » apparaissent.

Dès que l'on dépasse le seuil de -10%, les zones situées à l'Est deviennent plus attractives. Il faut dire que les tarifs sur ces liaisons sont élevés, comparativement aux autres, le gain en valeur absolue devient très intéressant.

La sensibilité à des tarifs plus élevés est moins marquée, et profite plus à la traversée centrale (DOCA) et à la zone Est Proche, qui se trouve, à la vue de ces résultats et des précédents, une véritable « zone tampon » entre le centre et l'Est éloigné.

Quoiqu'il en soit, les écarts sont bien moins importants que dans le cas de tarifs plus faibles, puisque l'on dépasse difficilement les -5% de variation ce qui correspond à un faible nombre de camions en valeur absolue. Des tarifs plus élevés agissent comme la prise en compte de pénalités appliqués à des tarifs ayant des niveaux absolus différents. Les tarifs déjà élevés de la zone Est éloigné, le deviennent encore plus, ce qui pénalise cette zone au profit de la zone Est proche, qui elle même est moins pénalisée par l'augmentation du fait de tarifs moyens de l'ordre de 25% moins élevé que la zone précédente.

En guise de conclusion, sur ce test, on retiendra qu'une erreur relative uniforme sur les tarifs a un impact « moyen » sur les résultats.

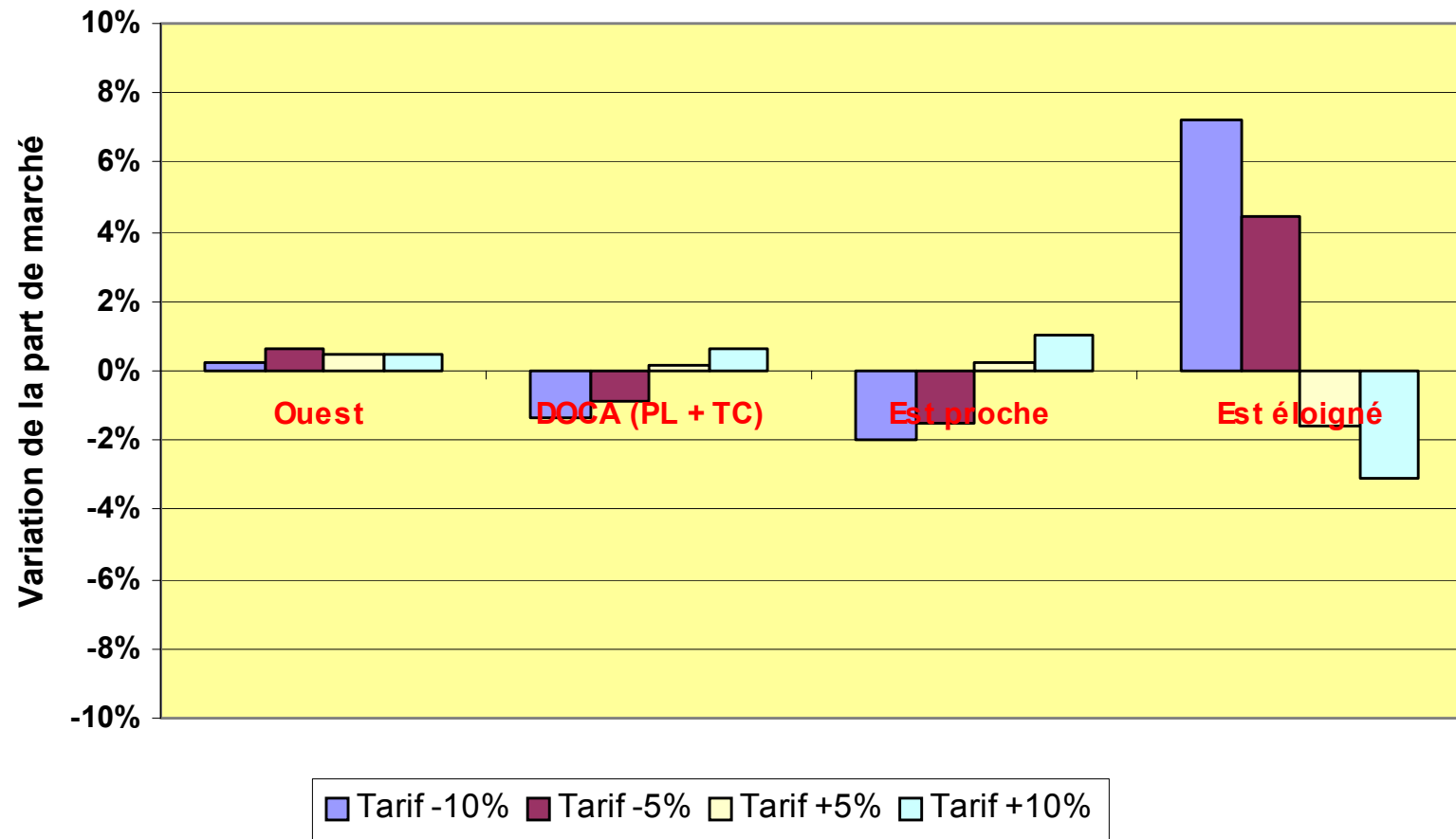
VI.3. TEST DE MODIFICATION DU NOMBRE DE ZONES ET DU NIVEAU DES PRIX

L'objectif de ce test est d'étudier la prise en compte d'une erreur combinée sur le zonage et sur les tarifs. A la réduction de 43% du nombre de zones, a été combiné une variation des tarifs comprise entre - 10% / + 10%. La présentation des résultats consiste en une comparaison avec la situation où le zonage est modifié et non pas la situation de base 2010.

Si l'on compare aux tests précédents, on ne remarque pas de grands changements, si ce n'est pour la zone ouest qui reste toujours dans le positif alors qu'elle était préalablement dans le négatif (mais les écarts sont extrêmement faibles).

L'ajout d'une erreur supplémentaire est donc relativement pénalisant pour un type de zone (Est éloigné), ce qui pose un réel problème, puisque l'on a vu que c'est précisément cette zone qui était la plus touchée par la modification du zonage.

Tests modification du zonage et des niveaux de tarifs (comparaison avec la situation de modification du zonage)



En conclusion de ce test on retiendra qu'une erreur sur le zonage et sur les niveaux de tarifs de traversée à la fois, aura une forte incidence sur les résultats. En d'autres termes, une erreur sur le zonage sera fortement aggravée par une erreur sur le tarif.

VI.4. TEST DE NON PRISE EN COMPTE DE LIAISONS FERRIES

La prise en compte des lignes ferries existantes et de la reconstitution de l'information les concernant constitue un aspect essentiel du travail de modélisation. En effet, les produits émis / attirés par les zones respectives traversent la manche par ces « voies » qui constituent un prolongement de la route. Pour un camion, la concurrence sur un itinéraire porte donc à la fois sur la possibilité de passer en surface (ferries) ou en souterrain (tunnel) et sur celle de choisir entre différentes liaisons de ferries en surface.

D'un point de vue pratique le calage du modèle ne peut se faire par ligne ferry puisqu'il est impossible d'ajuster les paramètres pour reconstituer l'information aussi finement. Le calage se fait donc par « ligne écran » portuaire correspondant à un regroupement homogène de liaisons ferries et dont nous avons parlé en début de chapitre.

Le nombre de liaisons ferries modélisées est relativement important dans le modèle ISIS, ce qui veut dire qu'il existe de nombreuses alternatives d'une manière générale, mais aussi au sein d'une ligne écran (environ 5 par ligne écran).

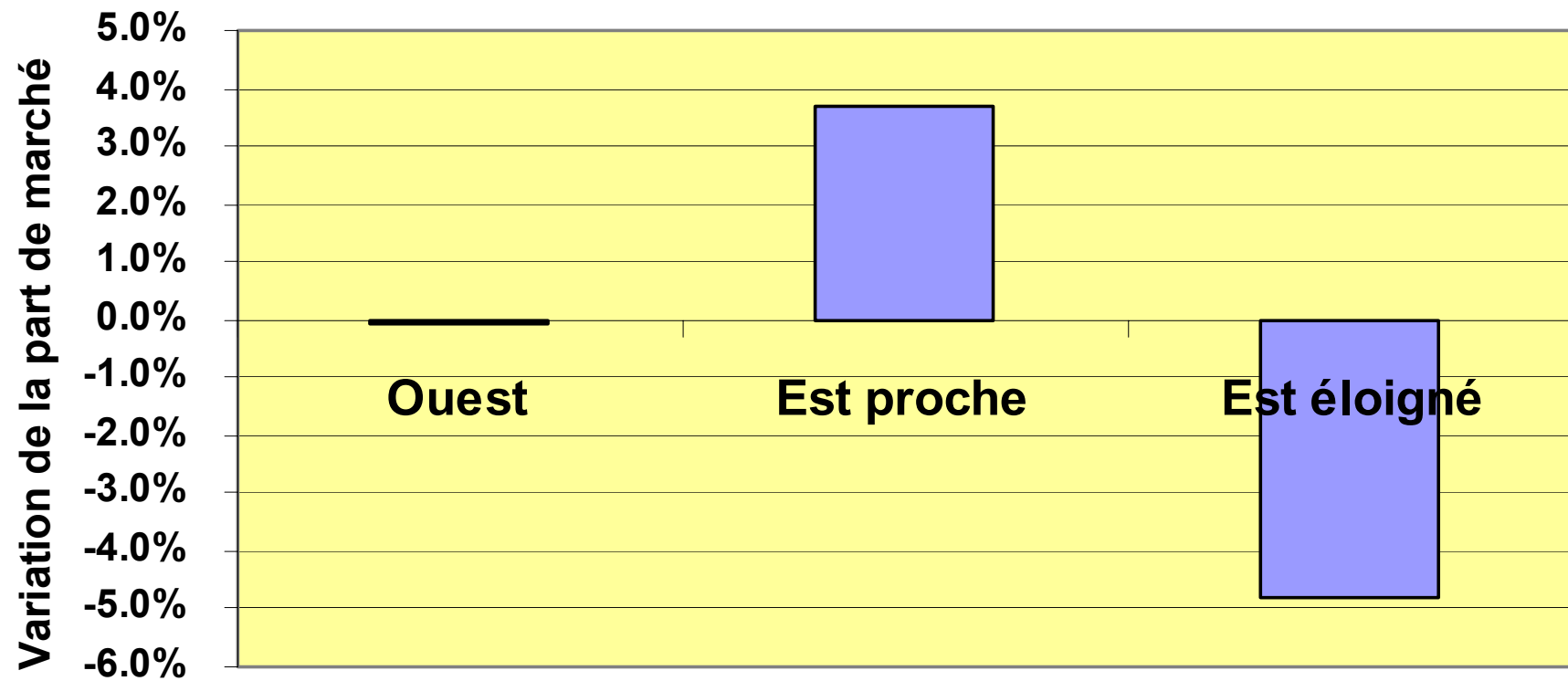
L'erreur sur la prise en compte des liaisons ferries peut provenir de deux sources : pas de connaissance sur l'existence même de certaines liaisons, pas d'information disponibles ou reconstituables sur des liaisons existantes. Dans les deux cas, les liaisons ne peuvent être prises en compte dans la modélisation.

Deux tests ont été menés:

- Suppression de la liaison qui possède le trafic le plus important de chaque ligne écran portuaire
- Suppression de 3 liaisons dans chaque ligne écran portuaire (situées soit à l'Est, soit à l'Ouest, afin de simuler un retrait aléatoire)

Le premier graphique montre le résultat du premier test. On remarque que la traversée centrale n'est pas présentée, car elle n'est tout simplement pas concernée par ce test (il n'y a qu'une ligne de ferry qui est Douvres-Calais). L'impact de la suppression de la ligne la plus chargée n'est pas très important ni d'un point de vue relatif ni d'un point de vue absolu (le nombre de camions concernés ne remet pas en cause les résultats de la modélisation). On remarquera, dans notre cas, que le glissement se fait de la zone Est éloignée vers la zone Est proche. En fin de compte seule la zone Est éloignée est touchée et voit ses affectations perturbées par un basculement de trafic vers l'Ouest.

Supression de la liaison ferry la plus chargée (Comparaison avec la situation de base)



Les résultats du second test, dans ses deux composantes, n'est pas présenté sous forme de graphique du fait des variations minimales (proches de zéros) qui ont été trouvées. Le retrait aléatoire simulé (la liaison ferry la plus chargée de chaque ligne écran portuaire n'est pas retirée), montre que les autres liaisons encore actives jouent bien leur rôle d'itinéraire alternatif.

En conclusion de ce test, on retiendra que la notion de « ligne écran portuaire » est fiable, dans le sens où elle permet de compenser une possible erreur sur les liaisons ferries prises une à une. Si, comme c'est le cas dans le modèle qui sert de base aux tests, le nombre de lignes écrans est assez élevé (par rapport à la zone du littoral couverte), il ne sera nécessaire de conserver un niveau de détail important au niveau des liaisons que s'il convient de restituer de l'information détaillée par ligne au Maître d'Ouvrage (les variations relatives par ligne écran sont alors en suite appliquées par liaison ferry, ce qui ne s'avère qu'un simple exercice de style !).

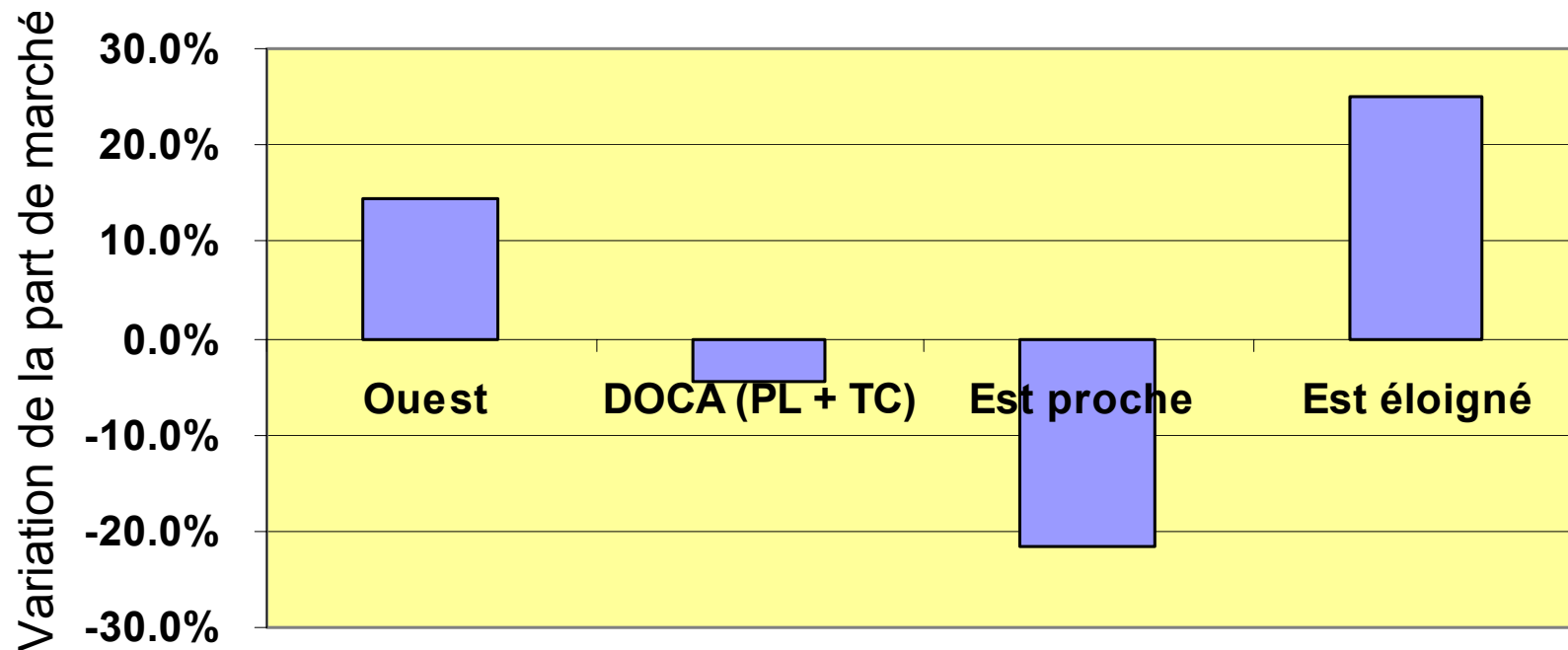
VI.5. PRISE EN COMPTE D'UN NOMBRE DE LIAISONS FERRIES REDUITS ET D'UN ZONAGE DIFFERENT

Ce test combine deux types d'erreurs : l'un sur le zonage et l'autre sur les liaisons ferries. Seuls les résultats de la suppression de la liaison la plus importante sont présentés. En effet, les autres tests apportent des résultats tout à fait comparables (dans les ordres de grandeur).

On remarque qu'une double erreur combinant zonage et liaisons ferries perturbe le processus d'affectation et entraîne des basculements de trafic d'un groupe de ports vers un autre. Ainsi, la zone Est proche est la plus touchée, ainsi que DOCA, au profit respectifs des zones Est éloigné et Ouest.

En conclusion de ce test on retiendra qu'une double erreur sur deux éléments importants de la modélisation, à savoir le zonage et le réseau, ont un impact négatif assez important sur les résultats. Tout comme pour le test combiné déjà réalisé, il n'est pas possible de se satisfaire de deux erreurs.

**Supression de la liaison ferry la plus chargée et modification
du zonage**
(comparaison avec la situation de modification du zonage)



VI.6. . PRISE EN COMPTE D'UN NOMBRE DE LIAISONS FERRIES REDUITS, D'UN ZONAGE DIFFERENT ET DE NIVEAUX DE TARIFS DIFFERENTS

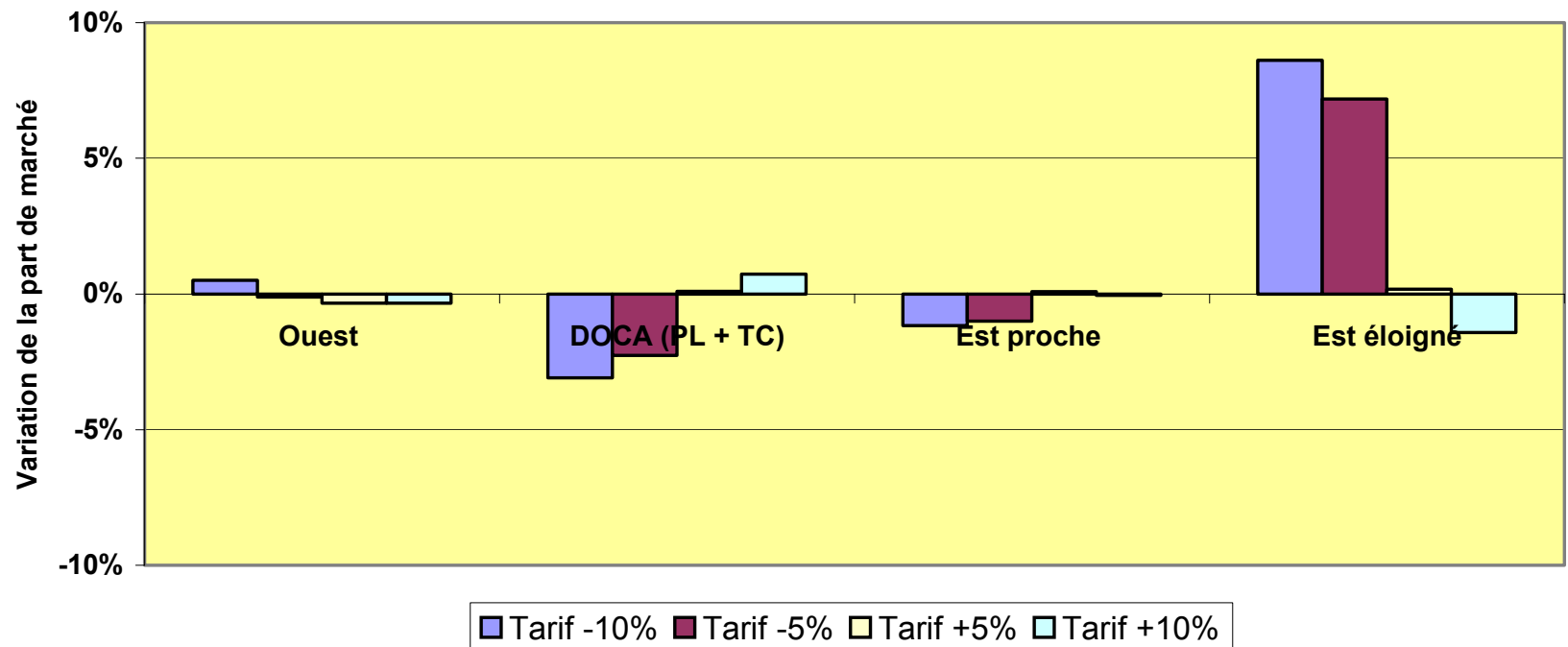
La prise en compte de trois types d'erreurs, ou plutôt de deux erreurs supplémentaires (tarifs et liaisons ferries) à l'erreur faite sur le zonage, est assez perturbatrice. Nous ne présentons, ici, qu'une situation où trois liaisons sont oubliées par « ligne écran » portuaire.

Le premier graphique montre quel est l'impact de l'erreur marginale, alors que le second compare l'effet des trois erreurs par rapport à la situation de base

Dans le premier cas, on remarque que l'erreur marginale reste dans des proportions convenables. Dans le second cas on remarque que les erreurs cumulées conduisent à des distorsions importantes.

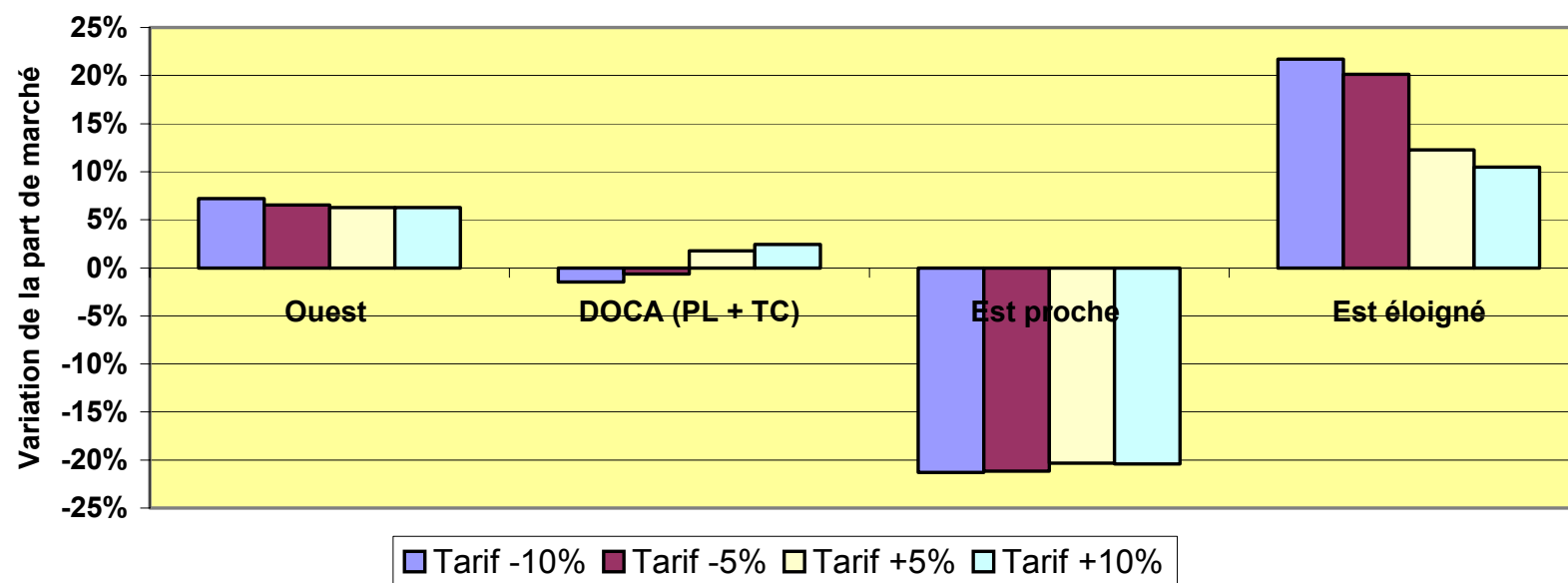
Tests de modifications des niveaux de tarifs

(comparaison avec la situation de modification du zonage et suppression de la liaison ferry la plus chargée)



Tests de modifications des niveaux de tarifs

(comparaison avec la situation de modification du zonage et suppression de la liaison ferry la plus chargée)



VI.7. CONCLUSIONS SUR LES TESTS

Le tableau suivant présente une synthèse des tests qui ont été menés.

Type d'erreur	Simulation	Impact quantifié sur les résultats par groupe de ports : modification des résultats en %	Avis concernant l'impact sur les résultats
Erreur simple			
Zonage	Réduction de presque la moitié de la taille du zonage	Jusqu'à 10%	Moyenne
Tarifs	Variation comprise entre [-5% et +5%]	Jusqu'à 5%	Faible
	Variation comprise entre [-10% et +10%]	Jusqu'à 10%	Moyenne
	Variation comprise entre [-20% et +20%]	Jusqu'à 20%	Importante
Réseau (liaisons ferries)	Suppression aléatoire de 3 liaisons par ligne écran portuaire	Jusqu'à 5% (rq : la technique des lignes écrans compense l'erreur)	Faible
Erreurs combinées (2)			
Zonage + tarifs	Cf. ci-avant	Impact supplémentaire jusqu'à 10%	Très importante
Zonage + Réseau (liaisons ferries)	Cf. ci-avant	Impact supplémentaire jusqu'à 20%	Très importante
Erreurs combinées (3)			
Zonage + réseau + tarifs	Cf. ci-avant	Impact supplémentaire jusqu'à 10%	Très importante

VII. CONCLUSION GENERALE

Outre les éléments relatifs au recensement des sources existantes, de leur intérêt pour une modélisation fret, ainsi que de leurs caractéristiques et des aspects pratiques les concernant (acquisition, contacts), nous espérons que la réflexion présentée dans ce rapport a attiré l'attention du lecteur sur certains aspects relatifs la « mesurabilité ».

Si l'on peut « faire dire aux chiffres ce que l'on veut », il est tout aussi possible « que les chiffres nous fassent dire ce que l'on ne veut pas », et cela sans le savoir !.

Le problème ne consiste donc pas à construire une base de données lorsque l'on établit un modèle, mais bien à se poser les (bonnes) questions lorsqu'on le fait. La confiance à accorder à nos modèles et aux méthodes de calcul est certainement beaucoup plus importante que celle à accorder à ce qui les alimente, et surtout à l'interface entre données et modèles.

Cependant, outre l'aspect « input », les bases de données recèlent tout un ensemble de pièges pouvant conduire à la prise d'hypothèses impactant *de facto* les résultats en projection, et ce de manière très forte.

Enfin, les erreurs ne se compensent certainement pas, comme le laisse penser un adage souvent évoqué dans les études transports, et elles ont certainement plus tendance à se cumuler.