L'utilisation de modèles dans la statistique publique relative aux transports

FRANÇOIS-XAVIER DE ROSSI* et ANTONIO F. GUALTIEROTTI**

1. INTRODUCTION: PROBLÈME ET DONNÉES

La connaissance que l'on a du transport de personnes résulte de plusieurs sources parmi lesquelles on compte en particulier le recensement fédéral de la population, l'enquête auprès des ménages sur les transports, et les comptages automatiques sur les routes. Voici un exemple de questions auxquelles on aimerait ou devrait pouvoir répondre:

1. *Combien de personnes se déplacent chaque jour de la commune A vers la commune B?*
2. *Comment évolue ce nombre?*
3. *Quelles voies et quels modes de transport sont utilisés à cette fin?*
4. *Comment varient et évoluent ces nombres?*

Ce genre de problèmes et de questions est central quand on veut faire de l'évaluation et du *monitoring* de politiques publiques, en l'occurrence celle des transports et de l'environnement. L'instrument statistique est un élément incontournable pour de telles évaluations.

Il y a environ 3000 communes en Suisse, ce qui donne 9 millions de couples „origine-destination". Or l'enquête sur les transports touche 15 à 20 000 personnes et on dispose de quelques 300 stations de comptage. Quant au recensement, il ne couvre que les trajets pour se rendre au travail. Il faut donc, pour produire les nombres requis, recourir à d'autres moyens. C'est ce que font depuis longtemps les ingénieurs des transports: il font appel à des modèles qu'ils calent à l'aide des données disponibles. L'objectif est donc de rendre cette approche statistiquement compatible. En effet, les ingénieurs s'intéressent peu (ORTÚZAR, 1994) à la signification statistique de leurs calculs et leur premier souci est celui d'arriver au bout du calcul, si bien que les méthodes qu'ils mettent en œuvre sont un mélange indissociable[1] de modélisation, d'expédients de calcul et d'expérience pratique. Les besoins de la statistique officielle sont tout autres et toute statistique issue de modèles doit pouvoir être comprise au niveau de ces modèles eux-mêmes.

* Office fédéral de la statistique, Neuchâtel
** Institut de hautes études en administration publique, Chavannes-près-Renens

1. La plupart du temps non documenté!

Ces modèles s'articulent sur

1. une description explicite de l'offre et de la demande,
2. des matrices Origine-Destination (O-D)
3. une opération d'affectation aux réseaux de transports.

Chaque élément de la liste précédente fait appel à une modélisation en soi et à des données spécifiques. La qualité des données résultant d'un tel processus dépend non seulement de la qualité de chaque étape du processus, mais encore de celle du processus lui-même.

Dans ce qui suit, on décrit rapidement deux grandes familles de modèles qui sont utilisés dans le domaine des transports, puis on aborde la question de la description statistique des comptages horaires. On termine par quelques considérations sur la nature des problèmes rencontrés.

2. LE „MODÈLE STANDARD" EN MATIÈRE DE TRANSPORTS

Peut-être faudrait-il parler de „méthode standard" plutôt que de „modèle standard" tant ce modèle varie d'un auteur à l'autre. En fait, le plus souvent, on ne trouve dans la littérature que des fragments de modèles, car les divers composants des modèles relèvent de techniques très diverses, les deux grandes familles étant la régression et l'optimisation. Ce qui suit s'inspire d'un ouvrage (ORTÚZAR, 1994) didactique qui survole le domaine sans cependant entrer dans les aspects de mise-en-œuvre.

2.1 Zones et données

La modélisation commence par la création d'un système de zones qui sont des regroupements territoriaux de ménages et d'entreprises et qui constituent les nœuds de départ et d'arrivée du réseau des transports. Ce sont donc des entités artificielles qui représentent, comme un député le fait d'une circonscription, une partie du territoire, que ce soit du point de vue des personnes, de l'économie ou encore des infrastructures. Les informations associées à ces zones sont démographiques et socio-économiques. Par contre, concernant les habitants de ces zones, on s'efforce de connaître les itinéraires empruntés, leurs origines et destinations, les modes de transport utilisés, les heures de départ et d'arrivée, et la fréquence des trajets. Les zones de prédilection en Suisse sont évidemment les communes.

Il va de soi que le système des zones ne peut être conçu indépendamment des données disponibles qui ont également une influence sur les techniques mathématiques ou statistiques auxquelles il est fait appel. Par ailleurs, la taille du réseau des zones joue un rôle primordial car l'efficacité des techniques mathématiques et statistiques que l'on peut mobiliser en est très tributaire. C'est presqu'une règle que des données adéquates font toujours défaut et que les créateurs de modèles recourent à des calages successifs basés sur des données dont la qualité doit augmenter au fur et à mesure de ces calages. Ainsi l'Office fédéral du développement territorial (ODT)[2] commence par obtenir une matrice O–D à partir d'un modèle gravitaire simple, basé sur la population et les distances kilométriques, ce qui exige l'estimation d'un paramètre. Une seconde matrice O–D est obtenue à partir de la première et d'un calage aux marges. Une troisième matrice est enfin obtenue en utilisant la répartition des distances parcourues.

2.2 La matrice des origines et des destinations (O–D)

C'est une matrice dont les lignes représentent les zones O_i d'origine du trafic et les colonnes, les zones D_j de sa destination: à la croisée de la ligne i et de la colonne j, on trouve le nombre de trajets issus de O_i et aboutissant en D_j. On note ce nombre $T_{i,j}$. Cette matrice a tendance à être une matrice creuse car $T_{i,j} = 0$ est fréquent.

L'opération qui produit les O_i et les D_j est appelée „génération des déplacements".[3] Dans le meilleur des cas, il est possible d'observer les nombres O_i et D_j, sinon il faut les estimer. Dans ce dernier cas, des techniques de régression sont très fréquemment utilisées. Le choix des variables explicatives utilisées dépend d'une part des données disponibles et d'autre part de la perception qu'ont les auteurs des modèles des facteurs qui génèrent prioritairement le trafic.

L'opération suivante, dite de „répartition des déplacements",[4] consiste à estimer les $T_{i,j}$ à partir des „marges" O_i et D_j. Il y a semble-t-il deux catégories principales d'approches. La première s'inspire des modèles de gravité: on ajustera, par exemple, en assurant que les contraintes

$$T_{i,\bullet} = O_i \text{ et } T_{\bullet,j} = D_j \tag{1}$$

soient satisfaites, un modèle de la forme

$$T_{i,j} = A_i O_i B_j D_j f(c_{i,j}), \tag{2}$$

où

A_i et B_j	sont	des paramètres à estimer,
$f(x) = x^n e^{-\beta x}$	est	la fonction de dissuasion,
$c_{i,j}$	est	le „coût" d'un déplacement de la zone i vers la zone j.

2. ODT, Division de la coordination des transports. Jusqu'en 1999: SET/GVF
3. Soit: *Trip generation.*
4. Soit: *Trip distribution.*

La seconde approche recourt à des critères de maximisation d'entropie de la forme

$$-\sum_{i,j}\left[T_{i,j}\ln(T_{i,j})-T_{i,j}\right], \tag{3}$$

toujours avec une contrainte sur les marges.

2.3 Le partage modal

C'est l'opération qui „ventile" le total $T_{i,j}$ selon les divers modes k de transport: on en tire les valeurs $T_{i,j}^{(k)}$. Les techniques utilisées relèvent de la régression logistique.

2.4 L'attribution au réseau

C'est l'opération qui lie à chaque arête k du réseau la part de $T_{i,j}$, soit $T_{i,j}^{(k)}$, qui lui est propre. Quand on dispose de comptages routiers, on est amené à résoudre des systèmes de la forme suivante (les inconnues sont les $T_{i,j}$):

$$V_l=\sum_{i,j}T_{i,j}p_{i,j}^{(l)}, \tag{4}$$

où

$p_{i,j}^{(l)}$	:=	proportion de déplacements de la zone i vers la zone j, transitant par l'arête l,
V_l	:=	comptage sur l'arête l.

Ce dernier point illustre bien les difficultés pratiques de l'utilisation de modèles: tout d'abord l'ensemble $\{l\}$ est un ensemble beaucoup plus petit que l'ensemble $\{(i,j)\}$, et ensuite les valeurs obtenues pour $T_{i,j}$ devront être compatibles avec celles issues des étapes précédentes!

3. UN MODÈLE PROBABILISTE POUR LES MATRICES ORIGINES-DESTINATIONS

L'avantage de tels modèles, par rapport aux „modèles standard", est que l'incertitude qui les caractérise en fait explicitement partie, ce qui permet le calcul direct de matrices de covariance, et donc d'intervalles de confiance pour les paramètres d'intérêt. L'hypothèse de base est la suivante: $T_{i,j}$ devient l'espérance d'une variable aléatoire de Poisson $N_{i,j}$. Quand on pose[5]

5. La quantité $\underline{c}_{i,j}$ est encore une valeur de „coût".

$$T_{i,j} = E\big[N_{i,j}\big] = A_i B_j e^{\langle \underline{\theta}, \underline{c}_{i,j} \rangle_{\mathbb{R}^m}}, \tag{5}$$

il est possible d'estimer $\underline{\theta}$ par moindres carrés ou maximum de vraisemblance et de produire la matrice de covariance des estimations (SEN, 1995). La limite de ces méthodes tient au fait que le plus souvent il y a une hypothèse d'indépendance des $N_{i,j}$ qui semble difficile à justifier.

Remarque: Dans le domaine probabiliste des modèles des transports, on peut constater l'émergence de méthodes manifestement statistiques, comme celles recourant à l'analyse spatiale et aux méthodes bayésiennes (ICKSTADT, 1998).

4. LA DESCRIPTION STATISTIQUE DES COMPTAGES HORAIRES

Quelle que soit la modélisation choisie, il est nécessaire de comprendre la nature de la variation rencontrée, soit pour pouvoir correctement la prendre en compte, soit pour (in-)valider les modèles que l'on a l'intention d'utiliser. On décrit ci-dessous une étude des comptages horaires routiers disponibles.

4.1 Données et objectifs de l'analyse

Il y a quelques 300 stations de comptage, que l'on peut classer selon divers critères, comme par exemple le type de route (autoroute, route nationale) ou le genre de lieu (urbain, rural). L' étude se limite aux années 1990–1998, soit 9 années de comptages horaires, ce qui fait des séries chronologiques contenant environ 78 840 observations ($9 \times 24 \times 365$) par station de comptage.

Soit $N_j^{(i)}$ le nombre de véhicules recensés à l'heure j à la station de comptage i. Typiquement, on aura:

$$N_j^{(i)} = \psi_{i,j}(\underline{X}_i, \underline{Y}_j, \underline{Z}_{i,j}), \tag{6}$$

où

$\underline{X}_i$	décrit	la localisation et l'environnement de la station de comptage,
$\underline{Y}_j$	décrit	la structure temporelle des comptages,
$\psi_{i,j}$ et $\underline{Z}_{i,j}$	décrivent	la structure probabiliste des comptages.

L'analyse a principalement deux objectifs: identifier $\psi_{i,j}$ et $\underline{Z}_{i,j}$, et estimer les fluctuations et l'évolution de $N_j^{(i)}$. $\underline{X}_i$ est donné, et $\underline{Y}_j$ est la partie „modèle" de l'expression.

En pratique, il convient de choisir une expression explicite de $N_j^{(i)}$ que l'on s'efforce ensuite d'ajuster et de valider.

4.2 *Le modèle de Zeger*

L'un des premiers modèles de régression pour les séries chronologiques de comptages est dû à Zeger (ZEGER, 1988). Il est typique des modèles de cette famille et des modèles qu'il faut utiliser dans le cadre de la modélisation des transports. Pour simplifier la notation, on ne prend en compte qu'une seule station de comptage.

Soit $\underline{B} = \{B_n,\ n \in \mathbb{Z}\}$, un processus stationnaire tel que

$$E[B_n] = 1,\ \operatorname{cov}(B_n, B_{n+p}) = \sigma_B^2 \Gamma_B(p), \tag{7}$$

et soit $\{\underline{Y}_n,\ n \in \mathbb{Z}\} \in \mathbb{R}^m$, un processus d'observables (ici l'heure, le type de jour et l'année). On pose

$$\ln\left[\lambda_j(\underline{\theta})\right] = \langle \underline{\theta}, \underline{Y}_j \rangle_{\mathbb{R}^m}. \tag{8}$$

On définit alors:

$$P\left(N_1 = k_1,\ \dots\ N_p = k_p \mid \underline{B}\right) = \prod_{j=1}^{p} \frac{\left[\lambda_j(\underline{\theta}) B_j\right]^{k_j}}{k_j!}\ e^{-\left[\lambda_j(\underline{\theta}) B_j\right]}. \tag{9}$$

L'identification consiste à estimer $\underline{\theta}$, σ_B^2 et Γ_B.

On a que:

$$\operatorname{corr}\left(N_j, N_{j+q}\right) = \frac{\Gamma_B(q)}{\sqrt{\left(1 + \frac{1}{\lambda_j \sigma_B^2}\right)\left(1 + \frac{1}{\lambda_{j+q} \sigma_B^2}\right)}}, \tag{10}$$

si bien que, quand les comptages sont grands, la corrélation est celle du processus stationnaire latent, alors que quand ils sont petits, la contribution du processus latent est moindre et elle est modulée par les comptages eux-mêmes.

Soit $\underline{N}_p$ le vecteur des variables $N_1, \dots, N_p$, et $\underline{\lambda}_p$ celui des espérances $\lambda_1(\underline{\theta}), \dots \lambda_p(\underline{\theta})$.[6] La méthode d'estimation proposée par Zeger est une méthode qui relève de celle des équations d'estimation généralisées et qui consiste à résoudre l'équation (en $\underline{\theta}$)

$$\left\{\frac{\partial}{\partial \underline{\theta}}\left[\underline{\lambda}_p\right]\right\}^t \Sigma_{\underline{N}_p}^{-1}\left(\underline{N}_p - \underline{\lambda}_p\right) = \underline{0}. \tag{11}$$

Conditionnellement aux paramètres de nuisance, le paramètre $\underline{\theta}$ se calcule récursivement à l'aide des deux formules suivantes:

6. On omettra le plus souvent dans la suite le paramètre $\underline{\theta}$.

$$\underline{Z}_n = \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_p] \right\} \underline{\theta} + (\underline{N}_p - \underline{\lambda}_p), \tag{12}$$

$$\underline{\hat{\theta}}_{m+1} = \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_p] \right\}^t \Sigma^{-1}_{\underline{N}_p} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_p] \right\} \right\}^{-1} \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_p] \right\}^t \Sigma^{-1}_{\underline{N}_p} \underline{Z}_n \right\} \left[\underline{\hat{\theta}}_m \right], \tag{13}$$

où le „terme" $\left[\underline{\hat{\theta}}_m\right]$ de cette dernière égalité indique que ce qui précède est évalué en $\underline{\hat{\theta}}_m$. La matrice de covariance asymptotique de $\underline{\hat{\theta}}$ est donnée par l'expression

$$\Sigma_{\underline{\hat{\theta}}} = \lim_p \left(p \left\{ \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_p] \right\}^t \Sigma^{-1}_{\underline{N}_p} \left\{ \frac{\partial}{\partial \underline{\theta}} [\underline{\lambda}_p] \right\} \right\}^{-1} \right). \tag{14}$$

4.3 Problèmes d'application

La pratique de l'ajustement de modèles du type de ceux de Zeger, dans le cadre des transports, fait rapidemment émerger une série de difficultés que l'on va maintenant passer en revue.

4.3.1 Choix de la fonction de régression

Un examen visuel des séries fait apparaître des régularités „fines" si bien qu'il est difficile de choisir les fonctions de régression qui permettent la dérivation standard des résultats asymptotiques (polynômes, fonctions trigonométriques) (Davis, 1999). Les méthodes basées sur les ondelettes (Nievergelt, 1999), qui permettent plus facilement la modélisation des „détails," présentent aussi des difficultés: par exemple, le nombre d'observations doit être une puissance de deux, et, par nature, sauf peut-être les ondelettes de Haar, semblent mal adaptées aux comptages.[7] C'est pourquoi c'est un modèle saturé qui a été retenu: il y a donc $1 + 23 + 6 + 8$ paramètres de régression.[8] Un modèle avec des catégories de routes conduit donc à des modèles avec de nombreux paramètres.

4.3.2 Valeurs initiales

Typiquement, le choix des valeurs initiales est celui que l'on obtient en ignorant la structure de corrélation des comptages. Cela donne des estimateurs consistants et asymptoti-

7. Exigeant en particulier un nombre élevé de coefficients.
8. Niveau général, heures, jours, années.

quement normaux même sous les hypothèses de corrélation. On a utilisé, dans les essais qui ont été faits, un lissage non paramétrique.[9]

On a recouru, pour l'estimation, à la méthode de Zeger et à celle du maximum de vraisemblance. Cette dernière présente la particularité que les estimations obtenues sont souvent très proches des valeurs initiales, ce qui n'est pas le cas pour les estimateurs de Zeger.

4.3.3 Biais et autres insuffisances

Davis, Dunsmuir et Wang (Davis, 2000) font non seulement remarquer que la méthode de Zeger produit des estimateurs biaisés, mais offrent encore un remède. Zeger lui-même avait d'ailleurs fait remarquer que son estimateur de la covariance peut produire une matrice qui n'est pas positive définie.

Les premières expériences faites avec les comptages routiers semblent indiquer que le remède proposé n'induit que des changements „imperceptibles" par rapport aux estimations avec biais.

4.3.4 Inversion des matrices de covariance

Soit Λ_p la matrice diagonale dont les éléments diagonaux sont $\lambda_1, \ldots \lambda_p$. On a que la matrice de covariance de $\underline{N}_p$ a la forme suivante:

$$\Sigma_{\underline{N}_p} = \sigma_B^2 \Lambda_p \Sigma_B \Lambda_p. \tag{15}$$

Les diverses formules exigent donc le calcul de

$$\Sigma_{\underline{N}_p}^{-1} = \frac{1}{\sigma_B^2} \Lambda_p^{-1} \Sigma_B^{-1} \Lambda_p^{-1}. \tag{16}$$

On a vu que la taille des séries est de l'ordre de 75 000 à 80 000 observations: inverser une matrice de covariance de cette taille pose des problèmes évidents. C'est pourquoi Zeger propose de simplifier la structure de covariance du vecteur des comptages de la manière suivante.

Soit D_p la matrice diagonale dont les éléments diagonaux sont respectivement $\lambda_i + \sigma_B^2 \lambda_i^2$, $1 \leq i \leq p$, et soit $\Sigma_q^{(p)}$ la matrice de corrélation d'un processus auto-régressif d'ordre q. On pose alors

$$\Sigma_{\underline{Y}_p} \approx D_p^{\frac{1}{2}} \Sigma_q^{(p)} D_p^{\frac{1}{2}}. \tag{17}$$

9. 4253H, twice.

Si $L_{p,q}$ désigne la matrice suivante, de dimensions $(p-q,p)$:

$$L_{p,q} = \begin{bmatrix} 1 & -\alpha_1 & \cdots & -\alpha_q & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & -\alpha_1 & \cdots & -\alpha_q \end{bmatrix}, \tag{18}$$

on peut affirmer que

$$\left\{\Sigma_q^{(p)}\right\}^{-1} \approx L_{p,q}^t L_{p,q}. \tag{19}$$

L'expérience montre que même le cas $q = 1$ peut poser problème.[10]

4.4 Résultats

Voici une liste des premières constatations que l'on peut faire:

1. L'examen des comptages moyens pour les 24 heures de la journée, par type de jour, et par année, indique que les types de jours semblent se regrouper „naturellement" en classes:

   $$\{LU\},\ \{MA, ME, JE, VE\},\ \{SA\},\ \{DI\},$$

   et qu'il y a une légère augmentation annuelle.
2. Les paramètres pour les heures montrent bien, pour les valeurs initiales, l'effet du lissage. La procédure d'estimation change assez peu les valeurs initiales. Elle fait surtout apparaître les variations fines qui sont à l'origine de la forme des trajectoires moyennes.
3. Les écarts les plus sensibles (relativement, mais aussi qualitativement) entre valeurs initiales et valeurs estimées se trouvent au niveau des paramètres qui „décrivent" le type de jour: on observe une „symétrie" par rapport à l'horizontale.
4. Pour ce qui est des années, les différences visibles sont relatives aux années 1997 et 1998.
5. Les résidus d'Anscombe ne sont pas normaux, mais cela n'est pas surprenant puisqu'on ignore, en les calculant, la corrélation entre comptages successifs. On remarque aussi la présence d'un certain nombre de résidus positifs excentriques. Sinon l'histogramme est remarquablement symétrique.
6. L'examen des „courbes" superposées des comptages et de la moyenne estimée, sur des périodes de 7 jours plus ou moins consécutifs montrent que pour certains types de jours (typiquement les jours ouvrables) la moyenne est au dessus des comptages,

10. Pour une seule série de comptages, dans une configuration PC avec 512 M de mémoire vive, et recours aux matrices creuses de MATLAB.

et pour d'autres types de jours (typiquement samedis et dimanches) la moyenne est au dessous des comptages. Ce fait est confirmé par le calcul de courbes horaires moyennes, par jour et par année. Les différences ont surtout lieu dans la partie „active" de la journée, mais les courbes moyennes donnent une image plus variée des différences rencontrées.

5. CONCLUSIONS

L'application de la méthode de Zeger donne des descriptions acceptables des comptages routiers bien qu'il y ait, à l'issue des calculs, nombre de phénomènes dont on peine à identifier la cause.

Les difficultés les plus immédiates que l'on a sont relatives à la taille des matrices de covariance que l'on doit usuellement calculer. Le recours à des techniques numériques et informatiques très spécifiques (matrices creuses, par exemple) semble impératif.

La validation des résultats obtenus semble exiger un important travail et ce d'autant plus que les instruments de diagnostic disponibles sont rares. Il y a un très grand nombre de méthodes en concurrence dont les avantages et les inconvénients restent à établir. Les comptages routiers constituent à cette fin un banc d'essais particulièrement utile.

REFERENCES

DAVIS, R. A., Y. WANG et W. T. M. DUNSMUIR (1999), „Modeling Time Series of Count Data", dans: GHOSH, S. (Editeur) *Asymptotics, Nonparametrics, and Time Series*, Dekker, New York.

DAVIS, R. A., W. T. M. DUNSMUIR, et Y. WANG (2000), „On autocorrelation in a Poisson regression model", *Biometrika*, 87, 491–505.

ICKSTADT, K., R. L. WOLPERT et X. LU (1998), „Modeling travel demand in Portland", Oregon, dans: Dey, P. Müller et D. Sinha (Editeurs), *Practical Nonparametric and Semiparametric Bayesian Statistics*, Lecture Notes in Statistics 133, New York.

NIEVERGELT, Y. (1999), *Wavelets Made Easy*, Boston.

ORTÚZAR, J. DE D. et L. G. WILLUMSEN (1994), *Modelling Transport*, 2e édition, Chichester, UK.

SEN A. et T. E. SMITH (1995), *Gravity Models of Spatial Interaction Behaviour*, Berlin.

ZEGER, (1988), „A regression model for time series of counts", *Biometrika*, 75, 621–629.

RESUME

L'information statistique issue des recensements et des enquêtes ne couvre pas toujours les besoins en information qui s'expriment et il n'est pas toujours possible non plus d'augmenter, pour les satisfaire, la taille des enquêtes: c'est le cas des statistiques relatives aux transports. Mais, dans ce cas précis, il y a une longue pratique, parmi les ingénieurs des transports, de recours à des modèles mathématiques explicites pour obtenir les nombres nécessaires. Par contre, les pratiques de cette modélisation ignorent le plus souvent l'aspect statistique de ces questions, et en particulier tout ce qui relève de la variation. Ce rapport est une contribution à une approche statistique des modélisations dans le domaine des transports.

ZUSAMMENFASSUNG

Die statistische Information, die aus Erfassungen und Umfragen hervorgeht, deckt nicht immer alle artikulierten Bedürfnisse. Es ist nicht immer möglich, deren Umfang zu erhöhen, um sämtlichen Bedürfnissen gerecht zu werden. Dies ist auch der Fall für das Transportwesen. Aber in diesem präzisen Fall haben die Transportingenieure unter Zuhilfenahme von expliziten mathematischen Modellen eine langjährige Erfahrung, um die benötigten Zahlen zu erreichen. In der Praxis ist es jedoch so, dass diese Modelisation häufig den statistischen Aspekt dieser Fragen ausser Acht lässt, und zwar ganz speziell was die Variationen anbelangt. Dieser Bericht ist ein Beitrag zu einer statistischen Auslegung der Modelisation im Bereich der Transporte.

SUMMARY

The statistical information produced by censuses and surveys does not always cover the information needs that exist and neither is it possible to always increase the size of the surveys to gather that information: such is the case for transport statistics. However, in this very area, there is a long tradition, among transport engineers, to use explicit mathematical models to produce the necessary data. But this practice ignores most often the statistical aspects of such modeling and particularly all that pertains to variation. This report is a contribution to the statistical modeling of transport problems.