

J.-B. KAZMIERCZAK

## **Migrations interurbaines dans la banlieue sud de paris**

*Les cahiers de l'analyse des données*, tome 3, n° 2 (1978),  
p. 203-218

[http://www.numdam.org/item?id=CAD\\_1978\\_\\_3\\_2\\_203\\_0](http://www.numdam.org/item?id=CAD_1978__3_2_203_0)

© Les cahiers de l'analyse des données, Dunod, 1978, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

## MIGRATIONS INTERURBAINES DANS LA BANLIEUE SUD DE PARIS

### [TRAFFIC]

#### 1 Objet de l'étude. Les données.

par J.-B. Kazmierczak (1)

##### 1.1 Prévision de trafic

Etant donné une zone géographique (un département, une région... ) divisée en sous-zones assez homogènes quant à la superficie ou au nombre d'habitants, on se propose de déterminer les trafics de sous-zone à sous-zone à l'aide de paramètres facilement accessibles (emplois et actifs par sous-zone, durée du trajet entre deux sous-zones...)

Attitude actuelle : Ainsi que nous le montrerons, tous les calculs effectués jusqu'à ce jour reposent sur des modèles "a priori".

Dénonçant cette démarche, nous montrerons qu'il est possible grâce à l'analyse factorielle des correspondances (A.F.C.) de déterminer les principaux facteurs explicatifs de trafic sans aucune hypothèse a priori (\*) puis par les formules de reconstitution des données à partir des facteurs - formules adaptées aux tableaux symétriques - de retrouver les trafics avec une bonne précision sans avoir à utiliser de nouveaux modèles.

##### 1.2 Plan de l'exposé

Nous adopterons la démarche suivante :

- présentation des modèles classiques
- premiers résultats de l'A.F.C.
- intérêt d'étudier un tableau symétrisé
- reconstitution des données

mais avant toute chose présentons les données en notre possession.

##### 1.3 Les données

Celles-ci concernent la banlieue Sud-Est de Paris, la planche I montre le découpage en sous-zones au nombre de 15. Les trafics du type domicile-travail, enregistrés en 1968 sont consignés en un tableau 15 x 15 (planche II).

Notons que tout ce qui ne fait pas partie de la zone d'étude a été divisé en 12 sous-zones (dites extérieures) et que si l'on connaît les trafics entre deux sous-zones d'un type différent, i.e. une intérieure et une extérieure, on ne connaît pas les trafics entre deux sous-zones extérieures.

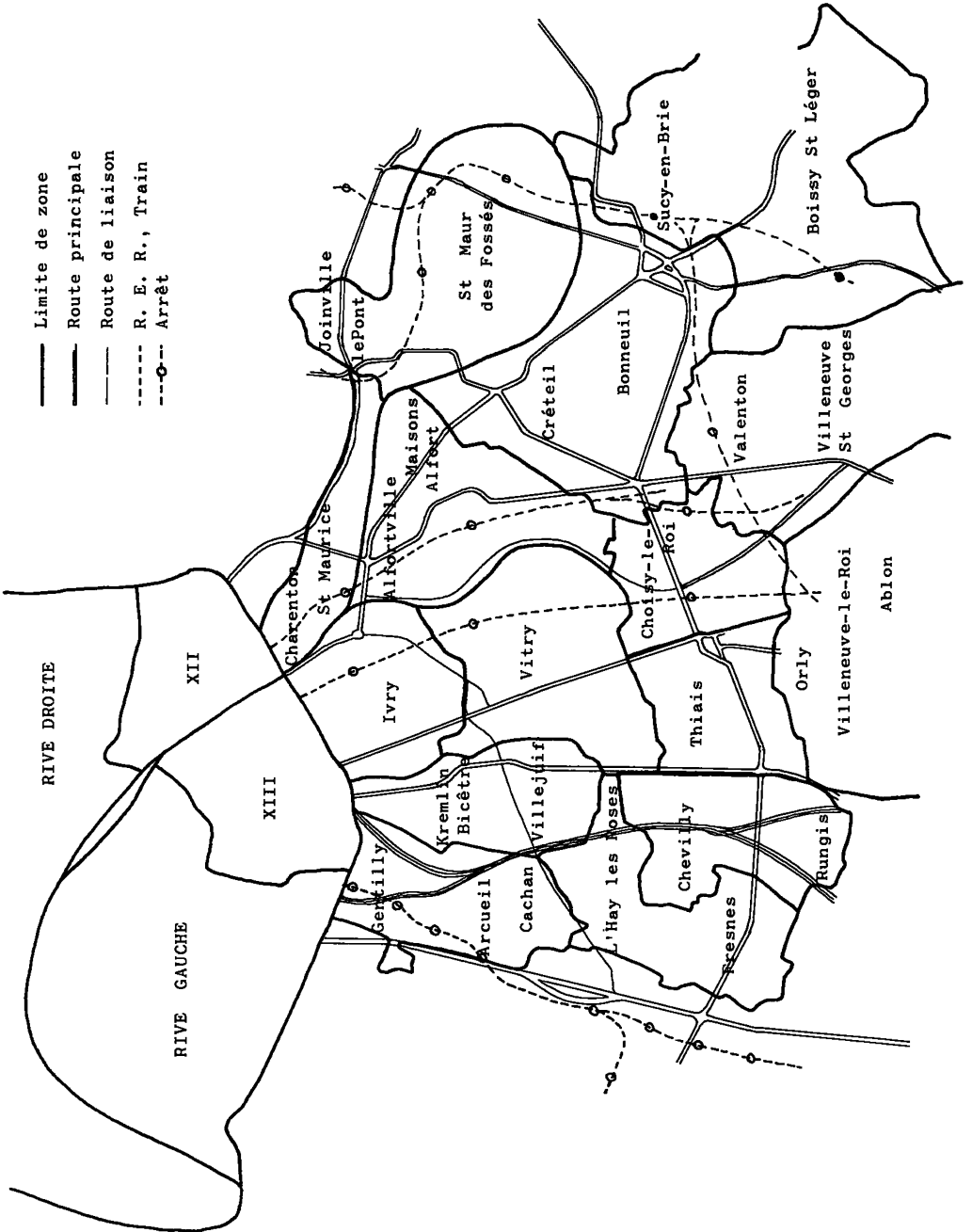
Pour cette raison nous sommes limités presque exclusivement à l'étude du tableau 15 x 15.

Regrettons l'absence de données concernant la même zone pour d'autres années que 1968. Seule sera dégagée une représentation statique du phénomène ; nous reviendrons plus tard sur ce point.

(\*) Autres que celles très peu restrictives de l'A.F.C. (le principe d'équivalence distributionnelle est parfaitement adapté à notre cas)

(1) Ecole Centrale, Paris.

PLANCHE I - LA ZONE D'ETUDE



[TRAFFIC]

PLANCHE II - LES DONNEES

INSEE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	TOTAL
	301	302	303	304	310	311	312	313	401	402	403	404	493	506	608	
DE VERS	Charenton	Ivry	Kremlin	Gentilly	Vitry	Alfort	Choisy	Bonneuil	Valenton	Orly	Rungis	Fresnes	Thiais	Joinville	Sucy	
1	6238	269	45	14	204	1824	57	250	70	76	16	36	0	403	189	9691
2	270	11268	1113	257	2483	1450	530	708	166	878	166	205	281	457	174	20406
3	34	585	11353	1001	1493	32	143	62	133	207	327	549	226	133	0	16278
4	0	106	1389	10695	425	100	99	220	27	111	215	1037	26	152	117	14719
5	186	667	894	281	11263	1009	1577	148	123	1021	154	265	860	314	90	18852
6	713	258	134	75	632	16420	595	1675	593	250	29	0	118	507	297	22266
7	0	181	78	41	763	148	5590	24	396	964	104	38	745	25	87	9184
8	51	81	68	0	133	1094	109	9235	107	92	0	28	39	1831	491	13359
9	31	34	34	28	34	316	271	148	6161	628	0	0	59	83	228	8055
10	14	108	492	177	353	104	528	209	568	6461	315	408	551	191	130	10609
11	0	21	160	83	81	33	23	20	64	248	1455	110	106	21	0	2425
12	0	53	310	260	156	0	0	0	0	82	481	3889	131	0	0	5362
13	0	66	21	0	151	40	421	24	43	248	26	0	1498	25	0	2563
14	327	43	0	63	206	801	42	1362	0	40	54	90	35	17045	774	20882
15	0	0	0	26	26	20	28	159	591	102	0	0	0	403	5624	6979
	7864	13740	16091	13001	18403	23391	10013	14244	9012	11408	3342	6655	4675	24590	8201	181630

## 2 Les modèles classiques

### 2.1 Les facteurs de croissance

C'est un modèle purement dynamique. On suppose connus les trafics à une époque donnée et on se contente de les modifier en fonction des modifications de population des sous-zones. Ce modèle étant totalement abandonné nous n'insisterons pas davantage.

### 2.2 Le modèle gravitaire

Ainsi appelé parce que sa première formulation est calquée sur la loi de gravitation :

$$T_{ii'} = k A_i E_{i'} / (d_{ii'})^2$$

$T_{ii'}$ , étant le trafic de  $i$  vers  $i'$

$A_i$  et  $E_{i'}$ , le nombre d'actifs de  $i$  et d'emplois de  $i'$

$d_{ii'}$ , la distance entre  $i$  et  $i'$

$k$  un coefficient approprié.

Les résultats obtenus sont assez médiocres aussi a-t-on modifié une première fois ce modèle en écrivant :

$$T_{ii'} = A_i E_{i'} f(c_{ii'})$$

$c_{ii'}$ , étant le "coût généralisé" du trajet entre  $i$  et  $i'$  (c'est-à-dire une combinaison linéaire de la durée du trajet et de la distance)

$f$  une fonction à ajuster de façon à minimiser un certain critère, généralement :

$$\sum_{ii'} (T_{ii'}^{\text{réel}} - T_{ii'}^{\text{calculé}})^2$$

Signalons que classiquement on recherche  $f$  sous la forme :

$$f(c) = k(c_0 + c)^{-b}$$

Sous les formes ci-dessus, le modèle gravitaire présentait le défaut de ne pas satisfaire aux "contraintes de marges" c'est-à-dire :

$$\sum_i T_{ii'} = E_{i'} \quad \text{et} \quad \sum_{i'} T_{ii'} = A_i$$

On a tourné la difficulté en écrivant :

$$T'_{ii'} = T_{ii'} + T_{i'i} = X_i X_{i'} f(c_{ii'})$$

$(X_i)$  étant un vecteur à ajuster en même temps que  $f$ .

Notons enfin qu'une version simplifiée suppose :

$$X_i = A_i + E_i$$

### 2.3 Le modèle d'opportunité

Ce modèle, appelé encore modèle de Chicago ou modèle Pittsburg du nom des villes où il a été expérimenté pour la première fois, est né il y a une dizaine d'années.

Un voyageur moyen effectue un classement préférentiel d'occupation des sous-zones pour une origine donnée. Le critère généralement adopté est celui par temps d'accès croissant.

On suppose de plus que tous les emplois situés dans une même sous-zone  $i$  sont équiprobables et possèdent tous une probabilité  $p_i$  de satisfaire au but du déplacement du voyageur moyen.  $E_i$  étant le nombre d'emplois de la sous-zone  $i$ , on a :

$$0 \leq p_i E_i \leq 1$$

Les sous-zones étant numérotées suivant le classement défini plus haut, on a pour une origine donnée :

$$Q_k = 1 - (1/A) \sum \{T_i | i = 1, \dots, k\} = \pi \{ (1 - p_i E_i) | i = 1, \dots, k \}$$

$Q_k$  désignant la proportion de voyageurs continuant au-delà de la sous-zone  $k$ .

Pour plus de commodité on note :

$$b_i = -(1/E_i) \text{Log}(1 - p_i E_i) \# p_i \text{ si } p_i E_i \ll 1$$

il vient alors :

$$Q_k = \exp(- \sum \{b_i E_i | i = 1, \dots, k\})$$

que l'on transforme une dernière fois en posant :

$$V_k = \sum \{E_i | i = 1, \dots, k\}$$

$$L_k = (1/V_k) \sum \{b_i E_i | i = 1, \dots, k\}$$

$$\text{d'où } Q_k = \exp(-L_k V_k)$$

C'est sous cette dernière forme que certains auteurs introduisent le modèle d'opportunité et interprètent  $L_k$  comme un coefficient d'atténuation.

Deux attitudes ont été adoptées pour l'ajustement de ce modèle :

- 1 - On suppose  $p_i$  constant sauf pour la sous-zone d'origine. C'est ce qui avait été fait par les auteurs de ce modèle.
- 2 - On étudie les variations de  $L_k$  en fonction de divers paramètres ( $V_k$ ,  $E_k$ , temps d'accès...). C'est la tendance actuelle.

Pour terminer précisons que les contraintes de marges sont difficiles à respecter.

#### 2.4 Les résultats obtenus

Les deux derniers modèles proposés fournissent, suivant les conditions d'emploi, des résultats de précision comparable.

Le modèle d'opportunité n'a été, à notre connaissance, testé que sur des zones peu étendues. Le modèle gravitaire ne fournit des résultats intéressants que pour des zones très vastes englobant plusieurs départements ; au voisinage des grandes villes, et principalement Paris, les résultats restent médiocres.

Nous avons testé le modèle gravitaire sur les données de l'étude que nous présentons aujourd'hui. A cette époque nous utilisons comme indice de précision la quantité :

$$\left( (1/n^2) \sum_{ii'} (T_{ii'}^{\text{réel}} - T_{ii'}^{\text{calculé}})^2 \right)^{1/2} / (1/n^2) \sum_{ii'} T_{ii'}^{\text{réel}}$$

Suivant le nombre de paramètres utilisés (de 3 à plus de 30) la précision oscille entre 30% et plus de 100%. Il faut cependant noter que si l'on ne tient compte que des déplacements extérieurs à la sous-zone de départ, une erreur de 50% ne représente que 130 déplacements ce qui ne semble pas excessif.

### 3 Analyse factorielle des correspondances

#### 3.1 Les tableaux analysés

Nous pouvons regrouper en trois classes les tableaux que nous avons analysés :

- 1 - Les données brutes : tableau  $15 \times 15$  des trafics  $T_{ii}$ , depuis l'origine  $i$  jusqu'à la destination  $i'$ .
- 2 - Les tableaux  $(15 + 12) \times 15$  et  $15 \times (15 + 12)$ , les 12 nouvelles lignes ou colonnes représentant les sous-zones extérieures.

Les analyses ont été refaites en plaçant ces éléments en éléments supplémentaires.

- 3 - Le tableau  $15 \times (15 + 15)$  construit en juxtaposant au tableau des données brutes le tableau transposé.

Toutes les analyses donnent des résultats comparables, seules apparaissent de très légères différences qui ne semblent pas explicables.

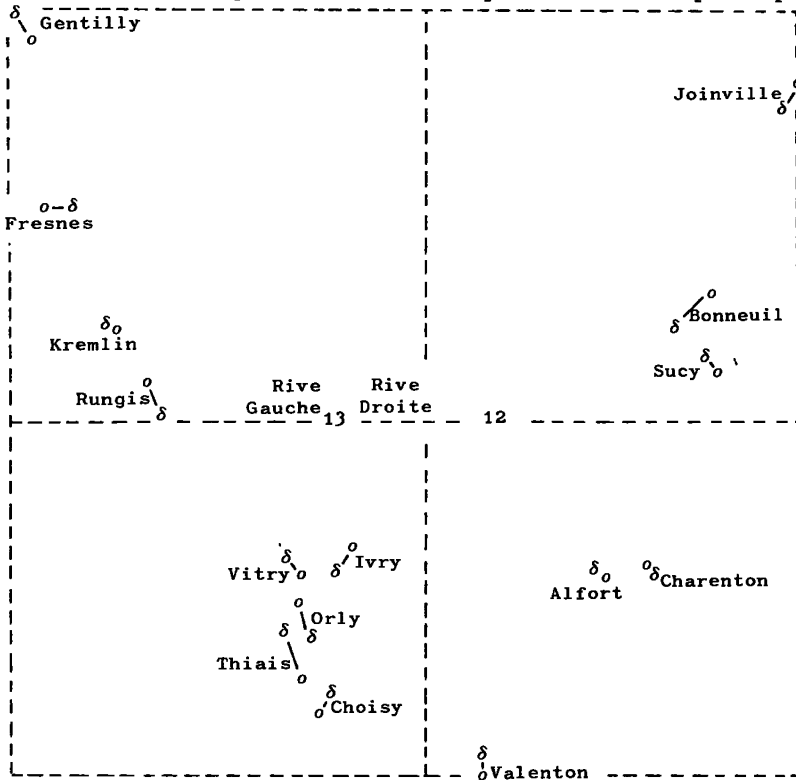


PLANCHE III : Analyse des correspondances; tableau  $15 \times (15 + 15)$  (tableau juxtaposé à son transposé) des flux trafic; on a noté  $o \times$  le point figurant la colonne des flux ayant  $x$  pour origine;  $\delta x$  la colonne des flux à destination de  $x$ .

Les résultats que nous indiquons ci-dessous sont issus de l'analyse du tableau 3 où ils apparaissent le plus clairement mais peuvent s'appliquer *in extenso* aux autres analyses.

### 3.2 Premiers résultats

Cinq points retiendront notre attention :

- 1 - Les valeurs propres sont assez élevées :  
.7839 .6492 .6080 .5471 .4745
- 2 - Les valeurs propres décroissent lentement :  
13.2% 10.9% 10.2% 9.2% 8.0%
- 3 - Les points "origine" sont proches des points "destination".
- 4 - Effet Guttman dans le plan (1-2) (second degré) que l'on retrouve dans le plan (1-4) (troisième degré)
- 5 - Effet attractif de Paris sur l'axe 3.

On aurait pu s'attendre à observer un effet Guttman bidimensionnel caractérisant les positions géographiques des sous-zones (et observer dans le plan 1-2 une carte de la région). L'effet unidimensionnel que l'on constate traduit en fait les difficultés de communication entre les différentes sous-zones bien plus que leurs relations géographiques: les sous-zones traversées par un même axe de circulation ont des valeurs assez proches sur le premier axe. Il en est ainsi de Gentilly et de Fresnes (RN 20), Vitry, Ivry, Thiais, Choisy et Orly (RN 7), Alfort, Charenton, Bonneuil et Sucy (RN 19).

On met donc en évidence les difficultés des liaisons transversales. Il est beaucoup plus facile d'aller à Paris que de tourner autour de la capitale.

Le troisième axe montre ce que nous avons appelé l'effet attractif de Paris. Il oppose Sucy-en Brie, Valenton puis Orly (éloignés de Paris) à Ivry, Alfortville et enfin Charenton (qui en sont proches).

Précisons cependant que si l'on met Paris en élément supplémentaire, celui-ci se trouve placé à l'origine. Ainsi tout semble se passer comme si les actifs éloignés de la capitale tenaient le raisonnement : "s'il faut se déplacer loin, allongeons un peu notre route et allons à Paris".

Le fait de trouver Paris proche de l'origine n'est pas propre à l'axe 3. On peut faire la même remarque pour les 6 premiers axes. Ainsi Paris est peu éloigné du centre de gravité du nuage :

#### PLANCHE IV - Comparaison des valeurs propres et taux

$f_{II}$  le tableau initial 15 x 15

$f_{I.I+I}$  le tableau et son transposé 15 x (15 + 15)

$f_{II}^S$  le tableau symétrisé 15 x 15

$\alpha$	$f_{II}$		$f_{I.I+I}$		$f_{II}^S$	
	$\lambda_\alpha$	$\tau_\alpha(\%)$	$\lambda_\alpha$	$\tau_\alpha(\%)$	$\lambda_\alpha$	$\tau_\alpha(\%)$
1	.7827	13.3	.7839	13.1	.7828	13.4
2	.6492	11.0	.6492	10.9	.6473	11.1
3	.6004	10.1	.6080	10.2	.6007	10.3
4	.5446	9.2	.5471	9.2	.5449	9.3
5	.4595	7.8	.4745	8.0	.4674	8.0



## Le premier facteur

	$f_{II}$	$f_{I.I+I}$	$f_{II}^s$
Cha	831	802	810
Ivr	- 304	- 293	- 273
Kre	-1109	-1107	-1107
Gen	-1457	-1427	-1426
Vit	- 466	- 459	- 460
Alf	627	620	623
Cho	- 338	- 362	- 368
Bon	928	970	968
Val	201	205	198
Orl	- 411	- 431	- 434
Run	- 963	- 943	- 946
Fre	-1291	-1311	-1311
Thi	- 493	- 486	- 495
Joi	1292	1302	1300
Suc	980	1032	1018

$$f_I^{\text{Paris}} \# f_I$$

Finalement les quatre premiers facteurs peuvent être interprétés . Néanmoins l'inertie relative aux facteurs non expliqués reste importante et l'objectif que nous nous sommes fixé, reconstruire les données à partir des facteurs expliqués, ne sera pas atteint avec une grande précision. Nous verrons plus loin comment y remédier.

Nous avons déjà signalé que les points "origine" étaient voisins des points "destination". Cette remarque - mais aussi l'importance logarithmique de la quantité  $T_{ii} + T_{i,i}$ , trafic global entre  $i$  et  $i'$  - nous incite à construire le tableau symétrisé :

$$T'_{ii'} = T_{ii} + T_{i,i}$$

#### 4 Analyse du tableau symétrisé

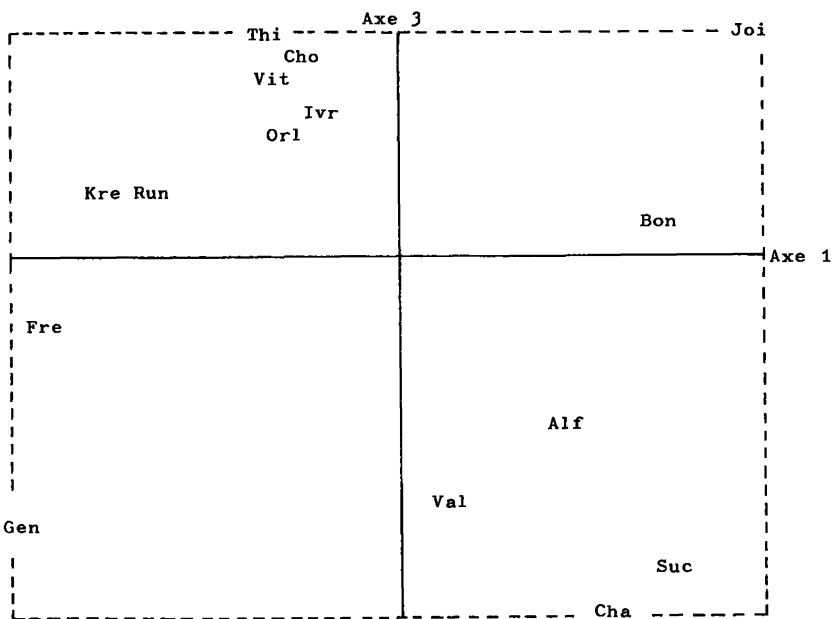
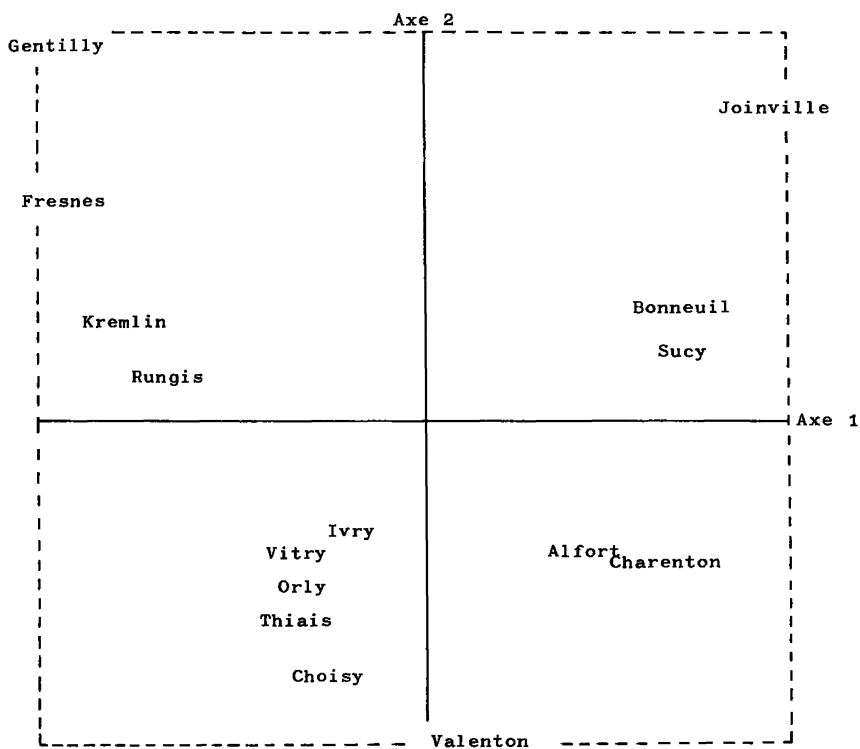
Dans ce § on utilise les propriétés des correspondances symétriques telles qu'elles sont exposées dans le traité sur l'Analyse des Données (TII B n° 9 [Corr. Sym.]) et rappelées assez en détail dans ce même cahier (cf [Pol. Sym.], ce cahier, pp 219-234).

##### 4.1 Les résultats

L'analyse factorielle du tableau symétrisé ( $T'_{ii'}$ ) conduit aux mêmes résultats que les analyses précédentes (pour les valeurs propres et le premier facteur on se reportera planche IV). Ajoutons que tous les facteurs sont directs, ce qui ne nous surprend pas étant donné que la diagonale est très chargée.

Il nous a cependant semblé utile de construire le convexe associé à la correspondance symétrique que nous étudions. On note (comme dans [Corr. Sym.] et dans [Pol. Sym.] ) :

PLANCHE V (Analyse du tableau symétrisé)



$$f_{ii'} = T'_{ii'}/T' \quad ; \quad \text{où } T' = \sum_{ii'} T'_{ii'}$$

$f_I$  la loi marginale

$$f_{II}^d = (\delta_I^I \times \delta_I^I) \text{ of } f_I : \text{loi diagonale} : f_{ii'}^d = \delta_i^{i'} f_i \quad ;$$

$$f_{II}^P = f_I \times f_I \quad : \text{loi produit} \quad : f_{ii'}^P = f_i f_{i'} \quad ;$$

on définit  $f_{II}^{uv}$  comme combinaison linéaire de  $f_{II}$  donné et des lois  $f_{II}^d$  et  $f_{II}^P$  (diagonale et produit) :

$$f_{II}^{uv} = u f_{II} + v f_{II}^d + (1-u-v) f_{II}^P$$

On sait que  $\{(u,v) \mid \forall i,i' : f_{ii'}^{uv} \geq 0\}$  est dans  $R^2$  un polygone convexe dont chaque point  $(u,v)$  peut être identifié à la loi  $f_{II}^{uv}$  correspondante.

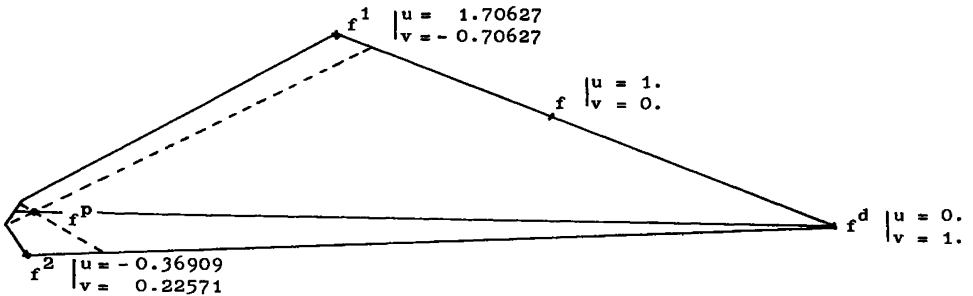


PLANCHE VI - Convexe associé à la correspondance symétrique  $f_{II}$ .

Les droites en pointillé délimitent la région où tous les facteurs sont directs (région contenant  $f_{II}$  et  $f_{II}^d$ ) et celle, très petite où ils sont tous inverses.

$f_{II}$			$f_{II}^1$			$f_{II}^2$		
$\alpha$	$\epsilon_\alpha$	$\lambda_\alpha^{1/2}$	$\alpha$	$\epsilon_\alpha^1$	$\lambda_\alpha^{1/2}$	$\alpha$	$\epsilon_\alpha^2$	$\lambda_\alpha^{2/2}$
1	+	.88480884	1	+	.80345271	1	-	.10086773
2	+	.80456680	2	+	.66653808	14	+	.09214981
3	+	.77506895	3	+	.61620675	2	-	.07125115
4	+	.73820692	4	+	.55331015	3	-	.06036377
5	+	.68369000	5	+	.46028953	13	+	.05535763
6	+	.65852466	6	+	.41735066	12	+	.04693075
7	+	.64363514	7	+	.39194510	4	-	.04675834
8	+	.63219563	8	+	.37242620	11	+	.03009590
9	+	.59959842	9	+	.31680655	5	-	.02663666
10	+	.57946416	10	+	.28245204	6	-	.01734837
11	+	.52998095	11	+	.19802030	7	-	.01185279
12	+	.48436923	12	+	.12019436	10	+	.01183211
13	+	.46153777	14	-	.08884906	8	-	.00763058
14	+	.36185442	13	+	.08123771	9	+	.00440074

PLANCHE VII - Tableau pour  $f, f^1, f^2$  des valeurs propres avec leur parité ; on a rangé chaque colonne dans l'ordre décroissant des valeurs absolues.

On a représenté ce convexe planche VI, en signalant les deux sommets  $f_{II}^1$  et  $f_{II}^2$ , points qui s'éloignent le plus de la ligne  $f_{II}^p$ ,  $f_{II}^d$  (la métrique utilisée étant celle définie dans [Corr. Sym.] et construite dans [Pol. Sym.]).

4.2 Analyse des correspondances  $\delta_{II}^1$  et  $\delta_{II}^2$

Il est acquis que les vecteurs propres issus de l'analyse de  $f_{II}^{uv}$  sont identiques à ceux issus de l'analyse de  $f_{II}$ , seules changent les valeurs propres associées. En numérotant de 1 à 14 les facteurs de  $f_{II}$ , on a indiqué planche VII leur nouveau rang, leur parité et la valeur propre associée.

L'analyse de  $f_{II}^1$  n'apporte rien de nouveau à notre connaissance mais confirme l'importance des quatre premiers facteurs. Celle de  $f_{II}^2$  attire notre attention sur deux nouveaux axes : le n° 14 (dans la numérotation définie par  $f_{II}$ ) qui oppose Thiais à toutes les autres sous-zones et le n° 13 qui cette fois oppose Rungis à toutes les autres sous-zones. Nous mettons ainsi en évidence le rôle particulier que jouent Thiais et Rungis. Cette constatation ne doit pas nous surprendre, ces deux sous-zones présentant des irrégularités de comportement que nous avons signalées dans une étude antérieure (ajustement des paramètres de mauvaise qualité dans le modèle d'opportunité).

4.3 La reconstitution des données

Les notations étant celles de 4.1 nous avons :

$$f_{II}^{uv} = u f_{II} + v f_{II}^d + (1 - u - v) f_{II}^p$$

$(u, v) \in C$ ,  $C$  étant un polygone convexe de  $R^2$  tel que :

$$\forall i, i' : f_{ii'}^{uv} \geq 0$$

Pour  $u \neq 0$ , nous pouvons écrire :

$$f_{II} = (1/u) (f_{II}^{uv} - (1 - u - v) f_{II}^p - v f_{II}^d)$$

L'analyse du tableau  $f_{II}$  (resp.  $f_{II}^{uv}$ ) nous donne les valeurs propres  $\lambda_\alpha$  (resp.  $\lambda_\alpha^{uv}$ ) avec la parité  $\epsilon_\alpha$  (resp.  $\epsilon_\alpha^{uv}$ ) où  $\alpha \in A$ , et les vecteurs propres  $\varphi_\alpha^i$ .

On a ainsi :

$$f_{ii'} = (1/u) f_i f_{i'} + (1 + \sum \{ \epsilon_\alpha^{uv} (\lambda_\alpha^{uv})^{1/2} \varphi_\alpha^i \varphi_\alpha^{i'} \mid \alpha \in A \}) - (1/u) (1 - u - v) f_i f_{i'} - (v/u) f_i \delta_i^{i'}$$

soit encore : pour  $i \neq i'$  :

$$f_{ii'} = f_i f_{i'} + (v/u) + \sum \{ (\epsilon_\alpha^{uv} (\lambda_\alpha^{uv})^{1/2} / u) \varphi_\alpha^i \varphi_\alpha^{i'} \mid \alpha \in A \}$$

et  $f_{ii} = (f_i)^2 (1 + (v/u) + \sum \{ \epsilon_\alpha^{uv} (\lambda_\alpha^{uv})^{1/2} / u) (\varphi_\alpha^i)^2 \mid \alpha \in A \}) - (v/u) f_i$

Etant donné la relation :

$$\epsilon_\alpha^{uv} (\lambda_\alpha^{uv})^{1/2} = u \epsilon_\alpha (\lambda_\alpha)^{1/2} + v = u (\epsilon_\alpha (\lambda_\alpha)^{1/2} + (v/u))$$

on constate que dans les formules ci-dessus n'intervient que le rapport  $\rho = (v/u)$ .

Construisons maintenant une correspondance  $f_{II}^*$  en n'utilisant que quelques facteurs issus de l'analyse du tableau  $f_{II}$ . Notons  $A^*$  l'ensemble tel que les indices des facteurs utilisés pour la reconstitution soient des éléments de  $C_A A^*$  et  $E_\rho(A^*)$  la quantité :

$$E_\rho(A^*) = \sum \{ (f_{ii} - f_{ii}^*)^2 / f_i f_i \mid i \in I, i' \in I \}$$

Nous avons :

$$f_{ii} - f_{ii}^* = f_i f_i \sum \{ (\epsilon_\alpha (\lambda_\alpha)^{1/2} + \rho) \varphi_\alpha^i \varphi_\alpha^{i'} \mid \alpha \in A^* \}$$

et  $(f_{ii} - f_{ii}^*)^2 / f_i f_i = f_i f_i \sum \{ (\epsilon_\alpha (\lambda_\alpha)^{1/2} + \rho) (\epsilon_\beta (\lambda_\beta)^{1/2} + \rho) \varphi_\alpha^i \varphi_\alpha^{i'} \varphi_\beta^i \varphi_\beta^{i'} \}$   
la somme étant étendue pour  $\alpha$  et  $\beta$  décrivant  $A^*$

Alors :  $\sum_i (f_{ii} - f_{ii}^*)^2 / f_i f_i = f_i \sum \{ (\epsilon_\alpha (\lambda_\alpha)^{1/2} + \rho)^2 (\varphi_\alpha^i)^2 \mid \alpha \in A^* \}$

puisque :

$$\sum \{ f_i \varphi_\alpha^i \varphi_\beta^{i'} \mid i' \in I \} = \begin{matrix} 0 & \text{si } \alpha \neq \beta \\ 1 & \text{si } \alpha = \beta \end{matrix}$$

Enfin :

$$E_\rho(A^*) = \sum \{ (\epsilon_\alpha (\lambda_\alpha)^{1/2} + \rho)^2 \sum \{ f_i (\varphi_\alpha^i)^2 \mid i \in I \} \mid \alpha \in A^* \}$$

et finalement :

$$E_\rho(A^*) = \sum \{ (\epsilon_\alpha (\lambda_\alpha)^{1/2} + \rho)^2 \mid \alpha \in A^* \}$$

En adoptant les notations suivantes :

$$R(A^*) = \sum \{ \epsilon_\alpha (\lambda_\alpha)^{1/2} \mid \alpha \in A^* \}$$

$$S(A^*) = \sum \{ \lambda_\alpha \mid \alpha \in A^* \}$$

nous pouvons écrire :

$$E_\rho(A^*) = S(A^*) + 2\rho R(A^*) + \rho^2 |A^*|$$

où l'on note  $|A^*|$  pour  $\text{card}(A^*)$

On remarque que  $E_\rho(A^*)$  est minimum pour :  $\rho = -R(A^*) / |A^*|$

et vaut :

$$e(A^*) = S(A^*) - R^2(A^*) / |A^*| \quad (\text{on suppose : } |A^*| \neq 0)$$

Nous verrons dans le paragraphe suivant l'intérêt qu'il peut y avoir à rechercher le minimum de  $e(A^*)$  lorsque  $A^*$  décrit  $\mathcal{P}(A)$  pour  $|A^*|$  donné.

Notons au passage deux propriétés élémentaires de la quantité  $e$  :

- 1 -  $e(\emptyset) = 0$
- 2 -  $|X| = 1 \Rightarrow e(X) = 0$

La première propriété découle de la définition de  $E_\rho(X)$  ; en effet : si  $X = \emptyset$  alors  $f_{II}^* = f_{II}$  et  $E_\rho(X) = 0$ .

Dans le second cas une seule valeur propre n'est pas utilisée dans la reconstitution, soit  $\lambda$ . Alors :

$$|X| = 1, S(X) = \lambda \text{ et } R(X) = \epsilon \lambda^{1/2}$$

d'où :  $e(X) = \lambda - (\lambda/1) = 0$ . ce qui était évident puisqu'en choisissant  $\rho$  tel que  $\lambda^{uv} = 0$ , on reconstruira  $f_{II}$  en n'utilisant que  $\text{card}(I)-2$  facteurs au plus<sup>a</sup>.

## 5. Reconstitution des données

### 5.1 Qualité de la reconstitution

Nous venons de voir qu'en reconstruisant un tableau symétrique à partir de quelques facteurs seulement il était possible de diminuer l'inertie résiduelle.

Dans notre cas,  $T'_i$  variant peu, nous avons :  $T'_i \# T'/n$ .

Ainsi :

$$\begin{aligned} e(A^*) &= \sum_{ii} \{ (f_{ii} - f_{ii}^*)^2 / f_i f_i \} = \sum_{ii} \{ (T'_{ii} - T'^*_{ii})^2 / T'_i T'_i \} \\ &= \sum_{ii} \{ (T'_{ii} - T'^*_{ii})^2 \} / (T'^2 / n^2) \\ &= (1/n^2) \sum_{ii} \{ (T'_{ii} - T'^*_{ii})^2 \} / (1/n^4) T'^2 \end{aligned}$$

$e(A^*)^{1/2}$  sera peu différent de la quantité introduite en 2.4 représentant l'écart quadratique moyen rapporté à la moyenne des déplacements. Il est commode d'utiliser  $e(A^*)^{1/2}$  exprimé en % pour estimer la qualité de la reconstruction.

Nous avons reproduit dans le tableau planche VIII quelques résultats concernant la qualité de la reconstitution.

Notre première remarque portera sur l'amélioration obtenue en utilisant les formules du § 4.3 par rapport aux formules classiques ( $\rho=0$ ). L'erreur moyenne est réduite entre 5 et 7 fois et compte-tenu des facteurs que l'on sait expliquer restera dans des limites acceptables.

Une seconde remarque s'impose : en recherchant pour  $k$  donné le minimum de  $e(X)$  pour  $x \in \mathcal{P}(A)$ ,  $\text{card}(X) = k$ , on constate que les facteurs à utiliser en priorité ne sont pas les premiers fournis par l'analyse de  $f_{II}$  mais ceux qui apparaissent en analysant la correspondance  $f_{II}^2$  (bien que subsistent encore quelques inversions). Ceci confirme l'importance de l'étude des points remarquables  $f^1$  et  $f^2$  du polygone associé à la correspondance.

Remarquons également que si l'on passe de  $k = 3$  à  $k = 4$  le facteur n° 2 disparaît ! Il est même possible *a priori* que tous les facteurs soient remplacés par d'autres lorsqu'on passe de l'ordre  $k$  à l'ordre  $k+1$  comme on peut le constater sur l'exemple ci-dessous :

Une analyse factorielle nous fournit seulement 5 valeurs propres non nulles :

$$\lambda_1 = .99 \quad \lambda_2 = .98 \quad \lambda_3 = .03 \quad \lambda_4 = .02 \quad \lambda_5 = .01$$

On constatera aisément que pour  $k = 2$  il faut prendre les facteurs 1 et 2 alors que pour  $k = 3$  on choisira les facteurs 3, 4 et 5 pour minimiser l'inertie résiduelle.

5.2 Méthode pratique de reconstruction

En se donnant pour but de reconstruire les trafics en n'utilisant qu'un minimum de paramètres, il nous paraît utile de procéder comme suit :

- 1 - Déterminer le 1° facteur (on peut à la limite le supposer acquis mais cela exige d'introduire 15 paramètres, un par sous-zone).
- 2 - Construire les facteurs 2 et 3 à partir du premier (effet Guttman).
- 3 - Se donner les valeurs propres correspondantes ainsi que le coefficient  $\rho$ .
- 4 - Reconstruire les données à l'aide de la formule du § 4.

Examinons successivement les étapes proposées :

Le premier facteur : Nous avons déjà noté que ce facteur est lié au temps d'accès, nous serons plus précis :

PLANCHE VIII - *Qualité de la reconstruction.*

*On a exprimé  $e(A^*)^{1/n}$  en %*

(1) k	(2)	(3)	(4)	(5)
0	241.8	51.7	51.7	
1	225.0	44.5	43.5	14
2	210.1	39.4	36.1	1 14
3	195.3	35.0	31.3	1 2 14
4	180.8	30.7	26.2	1 12 13 14
5	167.4	28.0	21.7	1 2 12 13 14
6	153.9	25.5	17.0	1 2 3 12 13 14
7	139.8	22.8	12.8	1 2 3 4 12 13 14
8	124.7	19.5	8.6	1 2 3 4 11 12 13 14
9	109.3	16.3	6.2	1 2 3 4 10 11 12 13 14
10	92.7	12.3	3.9	1 2 3 4 9 10 11 12 13 14
11	76.1	9.2	1.9	1 2 3 4 5 9 10 11 12 13 14
12	58.6	7.1	0.8	1 2 3 4 5 6 9 10 11 12 13 14
13	36.2	0.0	0.0	arbitraire
14	0.0	0.0	0.0	1 2 3 4 5 6 7 8 9 10 11 12 13 14

- colonne 1 : k = Nombre de facteurs utilisés pour la reconstruction
- colonne 2 : On utilise la formule classique de reconstitution des données à partir des facteurs.
- colonne 3 : On utilise la formule du § 4. Les facteurs sont - comme précédemment - pris dans l'ordre où ils apparaissent en analysant la correspondance  $f_{II}$ .
- colonne 4 : Pour un nombre de facteurs donné et égal à k, on choisit parmi l'ensemble de tous les facteurs k d'entre eux de sorte que  $e(A^*)$  soit minimum. Les facteurs utilisés sont indiqués en colonne 5.

$\varphi_1(i)$  est fortement corrélé ( $\geq 90$ ) à la quantité  $d_1(i) - d_2(i)$  où  $d_1(i)$  (resp.  $d_2$ ) est la durée du trajet depuis Gentilly (resp. Joinville) jusqu'à la sous-zone i. Gentilly et Joinville ne sont pas pris au

hasard mais sont les deux sous-zones les plus éloignées (en temps de parcours) l'une de l'autre - et occupent de ce fait les positions extrêmes sur le premier axe-.

Nous calculerons  $\varphi_1(i)$  de la façon suivante :

- (1) on pose  $x_i = d_1(i) - d_2(i)$
- (2) centrage :  $\sum_i \{f_i(x_i - m)\} = 0$
- (3) normalisation :  $k \sum_i \{f_i(x_i - m)^2\} = 1$
- (4) on a enfin :  $\varphi_1(i) = k^{1/2}(x_i - m)$

Les facteurs 2 et 3 : Nous écrivons :

$$\varphi_2(i) = a\varphi_1^2(i) + b\varphi_1(i) + c$$

$$\varphi_3(i) = d\varphi_1^3(i) + e\varphi_1^2(i) + f\varphi_1(i) + g$$

Pour déterminer les 7 coefficients a, b, ..., g nous écrivons les 7 équations :

- (1) centrage de  $\varphi_2$
- (2) orthogonalité  $\langle \varphi_1, \varphi_2 \rangle = 0$
- (3) normalité  $\langle \varphi_2, \varphi_2 \rangle = 1$
- (4) centrage de  $\varphi_3$
- (5), (6) orthogonalité  $\langle \varphi_1, \varphi_3 \rangle = \langle \varphi_2, \varphi_3 \rangle = 0$
- (7) normalité  $\langle \varphi_3, \varphi_3 \rangle = 1$

Les valeurs propres : Jusqu'à présent nous n'avons introduit aucun paramètre arbitraire, il faut nous résigner à le faire. Nous avons besoin des 3 valeurs propres correspondant aux trois facteurs pris en compte :

$$\lambda_1 = .7829 \quad \lambda_2 = .6473 \quad \lambda_3 = .5449$$

Il nous faut encore fixer la valeur de  $\rho$  qui va minimiser  $E_\rho$  pour les 3 facteurs considérés. En utilisant les résultats du § 4 on trouve :

$$\rho = .5827$$

La reconstruction : Les résultats obtenus par la méthode exposée ci-dessus sont satisfaisants. L'erreur moyenne (36%) est pratiquement égale à celle que l'on aurait obtenu en utilisant les valeurs exactes des facteurs. La précision est très bonne sur la diagonale (10%) sauf pour Thiais et Rungis (40%) ce qui était attendu.

Comme on a pu le constater nous n'avons pas tenu compte de "l'effet d'attraction de Paris". La raison en est que nous ne disposions pas de l'estimation du temps de parcours entre Paris et les différentes sous-zones. Les résultats théoriques obtenus nous permettent cependant de penser qu'il serait possible de ramener l'erreur de 36% à un peu plus de 30%.

Signalons enfin que les contraintes de marge (§ 2.2) sont automatiquement satisfaites, les facteurs étant centrés.

### 5.3 Conclusion

Finalement, il apparaît qu'une théorie maintenant classique, l'analyse factorielle des correspondances, permet avec une économie de moyens surprenante d'obtenir des résultats au moins équivalents à ceux qu'apportent les modèles spécialement élaborés pour le problème.



Regrettons que l'absence de données pour d'autres périodes ne nous autorise pas à faire réellement de la *prévision* de trafic. Seule une analyse dynamique (faisant intervenir le temps - époque à laquelle sont prélevées les données) peut fournir une solution pleinement satisfaisante au problème de la prévision.

## BIBLIOGRAPHIE

- BARBIER M. : Cours de circulation routière et études économiques ; E.N.C.P. - SETRA (ex-SERC)
- BENZECRI J.P. et coll. : L'Analyse des Données. TI et TII. DUNOD (1973) 2° éd. (1976).
- BLUMENFELD H. : Orientation future des recherches sur les transports urbains. Rapport O.C.D.E.
- KAZMIERCZAK J.B. : Le modèle d'opportunité. Séminaire Math. Appli. Ecole Centrale des Arts et Manufactures.