

**DEPLACEMENTS VOYAGEURS
INTER-REGIONAUX A LONGUE DISTANCE
MODELES D'INTERACTIONS SPATIALES
ET
RECHERCHE DE PARTITIONNEMENT DE L'ESPACE**

Mars 1997

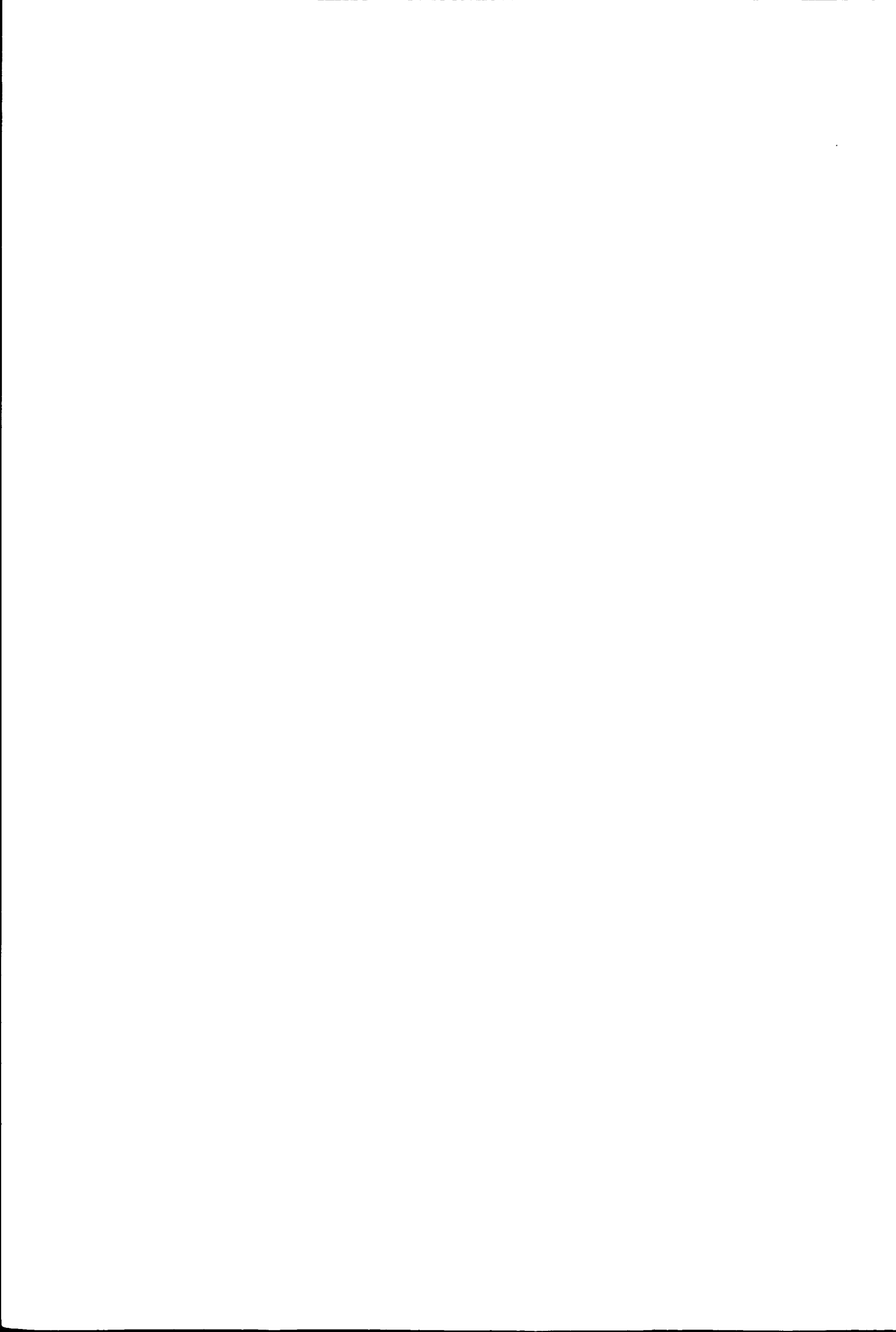
SYNTHESE DES RESULTATS

Jean-Marc VINCENT
François-Gaël OTTOGALLI
*Institut IMAG - Laboratoire de Modélisation et Calcul
Equipe Parallélisme*

Claude GRASLAND
Equipe PARIS - Université Paris I

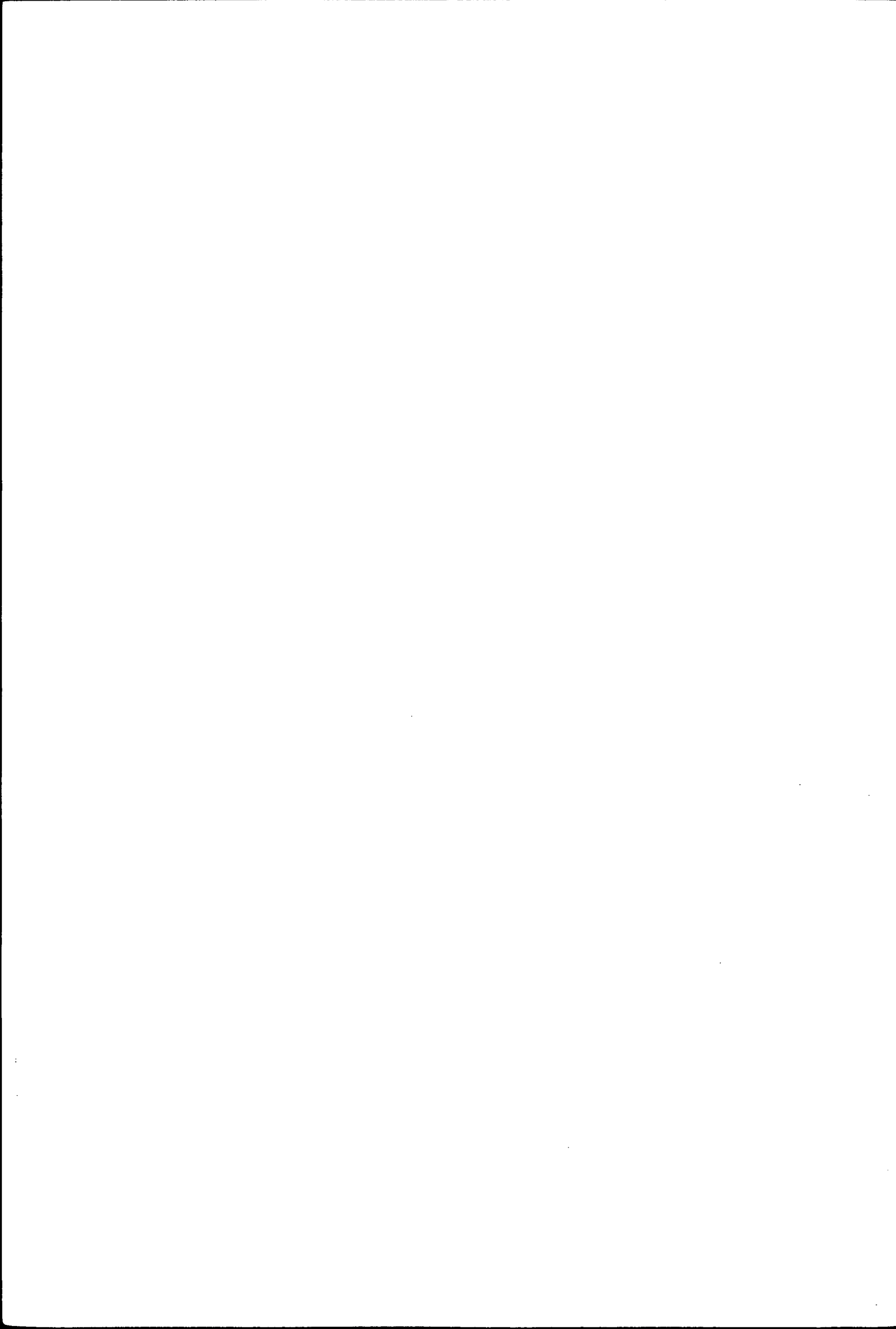
Christian CALZADA
METL/SES/DEE

11361

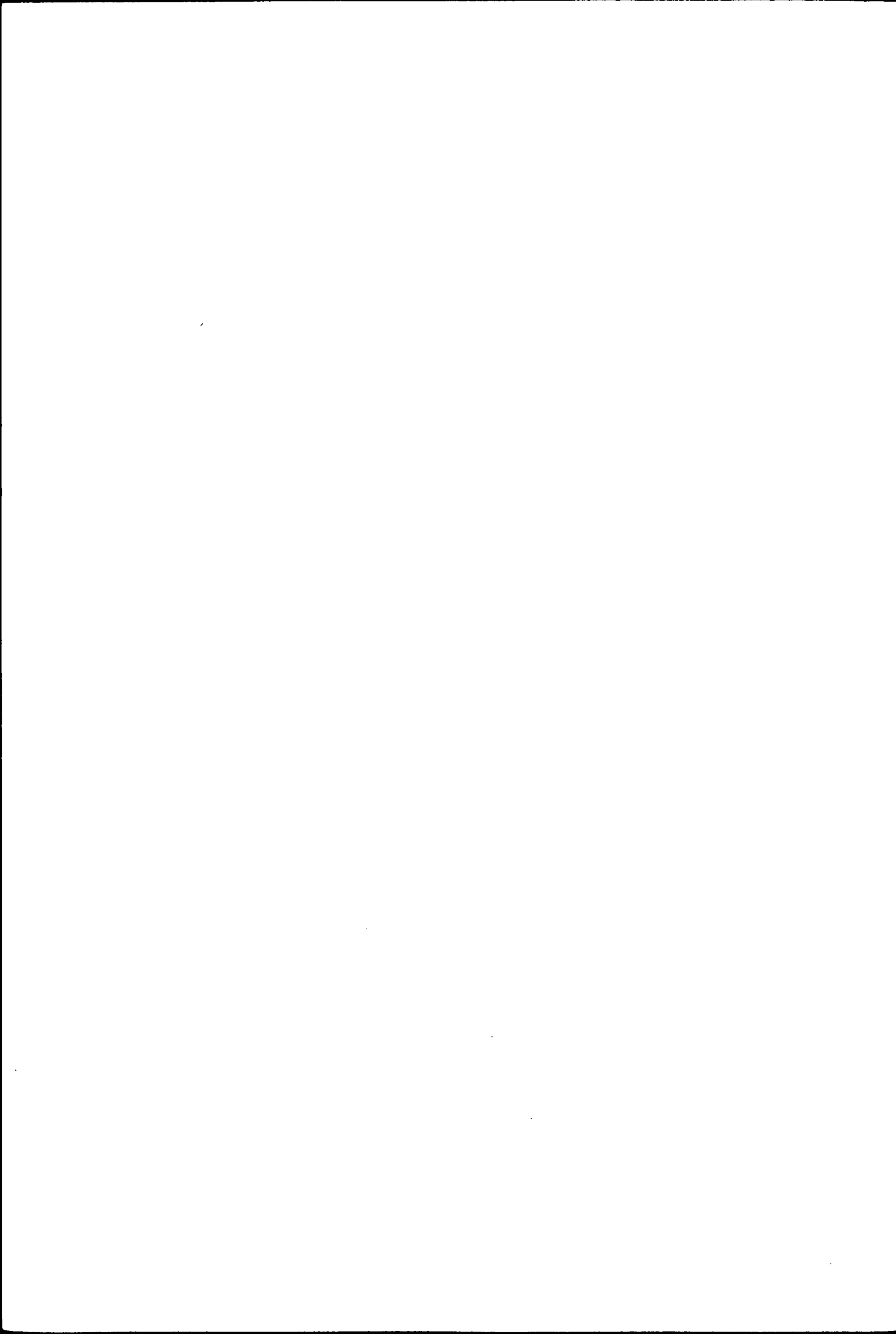


AVERTISSEMENT AU LECTEUR

Cette note réalise une rapide synthèse des principaux résultats de l'étude issus du rapport général de l'étude.



AVERTISSEMENT AU LECTEUR	p. 3
OBJECTIFS DE L'ETUDE	p. 4
DONNEES UTILISEES	p. 5
COMPARAISONS DES SOURCES	p. 6
BILAN	p. 6
ZONAGES	p. 6
EVALUATION DE L'INFLUENCE DE PARIS	p. 7
DISTANCES EMPLOYEES	p. 8
SELECTION DE LA MATRICE DE DISTANCE	p. 9
SPECIFICATIONS	p. 11
MESURE DE LA QUALITE DES MODELES	p. 12
PLATE-FORME EXPERIMENTALE UTILISEE	p. 13
EVALUATION DES PARAMETRES DU MODELE	p. 14
RECHERCHE DE PARTITIONNEMENT DE L'ESPACE GEOGRAPHIQUE	p. 15
CLASSIFICATION EXHAUSTIVE SUR L'ENSEMBLE DES PARTITIONS	p. 16
PRINCIPAUX RESULTATS	p. 18



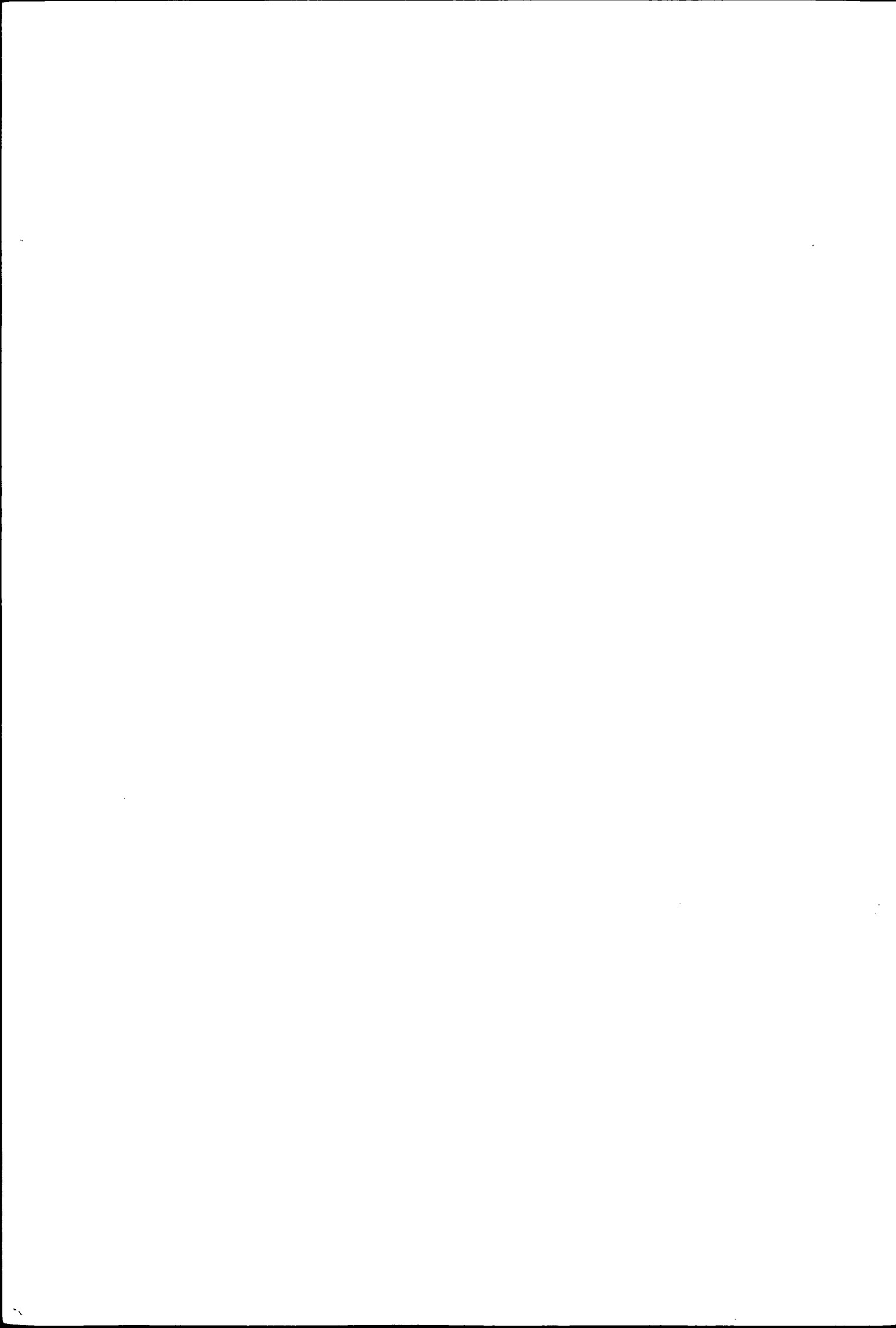
OBJECTIFS DE L'ETUDE

Cette étude qui associait des chercheurs de disciplines différentes (statistique (SES/DEE), géographique quantitative (Equipe P.A.R.I.S.), informatique (LMC (IMAG))) visait un double objectif:

* l'estimation d'un modèle à doubles contraintes pour les déplacements inter-régionaux de voyageurs et de véhicules à longue distance par mode et tous motifs, à partir de différentes sources et pour différentes spécifications de la fonction de résistance,

* ces modèles étant dans un deuxième temps généralisés, en incluant de possibles effets d'appartenance (barrières ou anti-barrières géographiques). L'idée est de tester l'hypothèse d'un espace français des transports spatialement homogène et isotrope.

Ce dernier paramètre sera utilisé sous une forme multiplicative dans le modèle à doubles contraintes, on recherchera la partition en 2 optimale pour le critère du χ^2 pour l'ensemble connexe considéré, en neutralisant ou non l'effet de Paris; une plate-forme informatique de recherche de partitionnement étant 'implémentée' à cet effet.



DONNEES UTILISEES

Les matrices de flux (voyageurs et trafics) utilisées sont issues de plusieurs sources:

* Source n° 1: *Modèle de prévision MATISSE*

Matrices multimodales de trafics en voyageurs pour le trafic français domestique en 1992, les zones origine-destination correspondant aux départements de la métropole (sauf les Hauts-de-Seine et le Val de Marne agrégés à Paris). Tous les couples de zones dont les centres sont distants de plus de 50 km à vol d'oiseau sont traités hors trajets internes à un département.

Les flux ne sont pas différenciés par sens et ont été agrégés au niveau régions administratives.

Particularités de la matrice obtenue:

La matrice 1992 est celle extrapolée par le modèle MATISSE de l'année 1988 sur la base des évolutions de contexte socio-économique et d'offre de transport observées pendant cette période.

La cohérence de cette matrice s'appuie sur les éléments suivants:

- calage sur les marges par classes de distance et moyens de transports avec les enquêtes ménages Transports INSEE, 1981-82 et 1993-94.

- calage sur le nombre de trajets ferroviaires par relation avec les statistiques de ventes de billets gare à gare de la SNCF (1988)

- calage avec ce que donnerait un modèle gravitaire simple (voir plus haut)

- conformité du partage des trajets entre la VP, le train et l'avion avec ce que donne le modèle MATISSE compte tenu de l'offre de transport multimodale existant sur la relation.

De plus une matrice régionale de trafics en nombre de véhicules a pu être mise en oeuvre.

* Source n° 2: *Enquête Transports et Communications (ETC) 1993-94 (INSEE)*

Cette enquête transports ménages s'est déroulée sur la période allant du 03/05/93 au 30/04/94. Les données ici utilisées proviennent du bloc « déplacements à longue distance (+ 80 km à vol d'oiseau du lieu de domicile ». L'interview de la personne B pour ses voyages et déplacements pondérés porte sur les trois derniers mois. Les « moyens de transports principaux » considérés sont: la VP (3 ou 4 roues), le train.

* Source N° 3: *Matrices du SETRA*

Matrices de flux de véhicules VP intérieurs interrégionaux pour « une année 1993 » en TMJA, les deux sens confondus

Cette matrice provient de la construction d'une matrice interdépartementale, réalisée à partir des enquêtes de circulation menées dans les CETE au cours des 5 dernières années. L'exploitation des enquêtes a permis de renseigner 50% de la matrice, le reste a été modélisé à partir des relations connues à l'aide du logiciel TRIPS (MVA) suivant la spécification suivante:

$$f(D) = D^{X_1} * \exp^{X_2 D} \quad \text{avec } D: \text{ distance entre préfectures}$$

$$\text{Flux (i j)} = A_i * A_j * f(D)$$

Les flux manquants sont alors recalculés avec les paramètres précédemment obtenus.

Les flux d'échanges VL entre régions limitrophes sont à prendre avec réserve, car les relations entre départements limitrophes quand elles sont disponibles par enquêtes, ont été appréhendées essentiellement sur les grands axes, la méthode de reconstitution des données manquantes n'étant pas adaptée pour ce type de flux très lié à l'environnement local. Deux versions étaient disponibles.

COMPARAISONS DES SOURCES

En raison de la difficulté de mise en place d'une matrice multimodale propre, le recours à plusieurs sources s'est révélé nécessaire.

En effet la matrice de l'ETC 1993-94 est soumise à certains biais:

- biais résultant de l'appel à la mémoire sur les trois mois précédant l'interview
- biais introduit pour limiter l'effet de lassitude résultant de la description de nombreux déplacements par les « hyper-mobiles » (détail sur un mois des déplacements des « grands voyageurs »: + 6 voyages au cours des 3 derniers mois)
- problème de comparaison des matrices de résidents avec les sources externes des opérateurs notamment qui incluent aussi le champ des non-résidents
- problème des absents de longue durée non recensés dans l'enquête 1993-94 a contrario de celle de 1981-82.
- problèmes liés au champ de « déclaration des distances » par rapport aux « distances estimés » provenant d'autres sources.

D'où un certain nombre de précautions d'utilisation:

- une sous-estimation des trajets par train dans l'enquête alors que les matrices routières semblent correctes
- le problème des relations avec l'Ile-de-france: un faible nombre de déplacements par route émis ou reçus sur des distances courtes (0-250 km) et un taux faiblement négatif pour les distances très longues (> 750 km). Ces résultats pour le moins inquiétants apparaissent d'autant plus problématiques au vu de que l'on sait de l'évolution réelle constatée entre les deux enquêtes.

BILAN

L'ensemble des matrices régionales (9) couvrent les deux sens de trajets confondus, elles ont en effet été 'symétrisées' ($f_{ij} + f_{ji}$) et le mode aérien a été mis de côté.

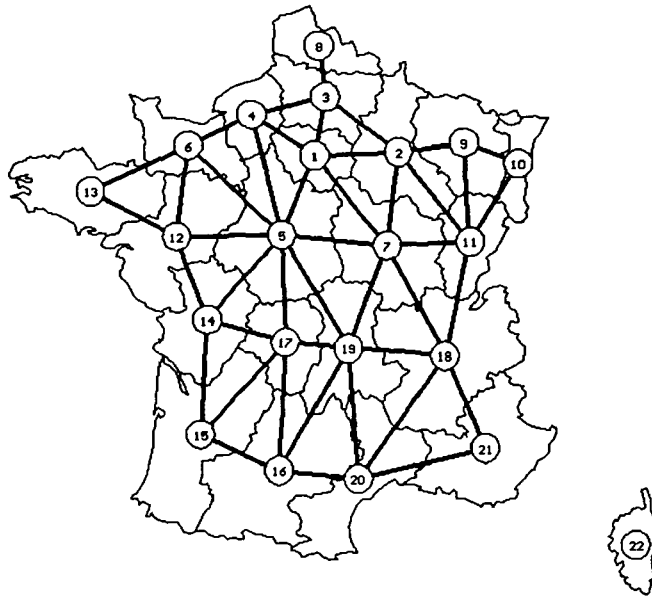
On dispose donc de 9 types de matrices:

- * 3 matrices issues de MATISSE: VP, Train et Total des flux,
- * 3 matrices issues de l'ETC 93-94: VP, Train et Total des flux,
- * 3 matrices routières en nombre de véhicules: VP-SETRA 1^{ère} version, VP-SETRA 2^{ème} version, VP-MATISSE.

ZONAGES

. un zonage spécifique a été défini à partir du fichier des zones d'emplois (INSEE).

. Topologie du graphe de contiguïté: un graphe de contiguïté [voir graphe] a été élaboré dont une région La Corse a été volontairement exclue. D'autre part l'option de non-rattachement des régions N°8 (Nord) et N°2 (Champagne) a été retenue même s'il s'en faut de quelques communes et compte tenu de l'allongement de la région Champagne-Ardenne.



EVALUATION DE L'INFLUENCE DE PARIS

L'influence de l'Île de France a été décomposée en trois parties :

- * cette région jouant un rôle identique à n'importe quelle autre région,
- * une région neutralisée (PARIS NEUTRE), au sens où les flux peuvent transiter par la région mais sans s'arrêter,
- * une région déconnectée des autres unités géographiques: aucun flux ne peut s'arrêter ni transiter par cette région.

DISTANCES EMPLOYEES

La distance la plus simple à employer est la distance euclidienne ($DEucl$). Elle représente la distance « à vol d'oiseau » entre deux unités géographiques.

$$DEucl = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

(x_i, y_i) : centre de gravité de l'unité i

(x_j, y_j) : centre de gravité de l'unité j

Les coordonnées citées ci-dessus représentent le centre de gravité des unités i et j , mais elles peuvent tout aussi bien identifier la ville la plus importante, le chef lieu administratif, etc...

Bien que la distance euclidienne puisse paraître tout à fait adaptée à notre cas, il est possible d'employer n'importe quel type de distance, d'autant plus que les matrices de flux utilisées concernent les déplacements en moyenne à plus de 80 km à vol d'oiseau du lieu de domicile. A cette occasion trois nouvelles distances ont été construites. Elles permettent de prendre en compte la population incluse dans les zones d'emplois INSEE. Cette subdivision des unités géographiques en zones d'emplois permet de pondérer l'action des flux en fonction de la population.

1. D_0 : se base sur la définition de la distance euclidienne pondérée par la proportion de la population des zones d'emplois.

Soit les unités géographiques i et j et les zones d'emplois k ($k \in i$) et l ($l \in j$). La distance $D_{0i,j}$ est définie par:

$$D_{0i,j} = \sum_{k \in i} \sum_{l \in j} \frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]}$$

avec

$pop_i[k]$: population de la zone d'emploi k de l'unité géographique i

2. D_{50} : se base sur la définition de la distance euclidienne pondérée par la proportion de la population des zones d'emploi séparées d'au moins 50 km.

Soit les unités géographiques i et j et les zones d'emplois k ($k \in i$) et l ($l \in j$). La distance $D_{50i,j}$ est définie par:

$$D_{50i,j} = \sum_{k \in i} \sum_{l \in j} \frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]} * \delta_{[DEucl_{k,l} \geq 50]}$$

avec

$pop_i[k]$: population de la zone d'emploi k de l'unité géographique i

$\delta_{[DEucl_{k,l} \geq 50]}$: vaut 1 si $DEucl_{k,l} \geq 50$, 0 sinon.

3. D_{80} : se base sur la définition de la distance euclidienne pondérée par la proportion de la population des zones d'emploi séparées d'au moins 80 km.

Soit les unités géographiques i et j et les zones d'emplois k ($k \in i$) et l ($l \in j$). La distance $D_{80i,j}$ est définie par:

$$D_{80i,j} = \sum_{k \in i} \sum_{l \in j} \frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]} * \delta_{[DEucl_{k,l} \geq 80]}$$

avec

$pop_i[k]$: population de la zone d'emploi k de l'unité géographique i

$\delta_{[DEucl_{k,l} \geq 80]}$: vaut 1 si $DEucl_{k,l} \geq 80$, 0 sinon.

SELECTION DE LA MATRICE DE DISTANCE

L'étude exhaustive croisant matrices de flux (9) * fonctions de distance (4) * distances interrégionales (5) correspondrait à 180 configurations possibles !. Il s'est vite avéré nécessaire d'effectuer un choix sur l'une de ces contraintes (flux, fonctions de distance, distances). A cet effet une étude complète des distances disponibles a été réalisée grâce au modèle à doubles contraintes associé à la fonction de la distance de Pareto [voir tableau].

L'analyse des erreurs présentée dans les tableaux ci-joints nous orientent vers le **choix de la distance D_0** comme distance la plus appropriée.

SELECTION DE LA MATRICE DE DISTANCE

Modèle de Pareto

	DEucl	D0	D50	D80	D50+Prop	D80+Prop	MIN	MEDIANE
	Err	Err	Err	Err	Err	Err		
M31	0,0163	0,0080	0,0080	0,0587	0,0081	0,0724	0,0080	0,0122
M32	0,0197	0,0191	0,0194	0,0363	0,0189	0,0410	0,0189	0,0195
M34	0,0299	0,0279	0,0286	0,0709	0,0278	0,0843	0,0278	0,0292
MS15	0,0482	0,0507	0,0522	0,1675	0,0506	0,1978	0,0482	0,0515
MS16	0,0265	0,0247	0,0248	0,0301	0,0246	0,0316	0,0246	0,0256
MS19	0,0467	0,0499	0,0511	0,1303	0,0497	0,1517	0,0467	0,0505
VLS	0,0250	0,0105	0,0109	0,0981	0,0107	0,1232	0,0105	0,0180
VLS2	0,0163	0,0080	0,0080	0,0587	0,0081	0,0724	0,0080	0,0122
VLS3	0,0312	0,0250	0,0251	0,0770	0,0250	0,0598	0,0250	0,0282

SPECIFICATIONS

Afin d'aboutir au plus degré de généralité côté spécification des équations testées, il a été convenu de partir d'un modèle de Wilson à doubles contraintes sans conditionner l'ajustement dans la première phase par la disponibilité de variables économiques et démographiques pour qualifier les facteurs d'émission et d'attraction, ce qui est souvent le cas, pour dans des phases ultérieures permettre, si nécessaire, des procédures de réajustement aux marges plus larges.

La forme générale des modèles utilisés ci-après est celle des Modèles d'Interactions Spatiales dits à doubles contraintes. Cette dernière permet de combiner le pouvoir de rétention et d'attraction des unités géographiques, dans un objectif plus de prévision que de caractérisation.

On a:

$$F_{ij}^* = a_i O_i b_j D_j h(c_{ij})$$

avec:

$$a_i = 1 / \left[\sum_{i,j} b_j D_j h(c_{ij}) \right] : \text{contrainte de la conservation des départs}$$

$$b_j = 1 / \left[\sum_{i,j} a_i O_i h(c_{ij}) \right] : \text{contrainte de la conservation des arrivées}$$

O_i : somme des arrivées en i ,

D_j : somme des départs de j .

$h(c_{ij})$: fonction d'interaction spatiale (résistance).

☞ Pour ce qui suit quatre types de spécification dont une dite de 'référence' ont été retenus:

- * modèle dit de « référence » (I): $F_{ij}^* = a_i O_i b_j D_j h(c_{ij})$ avec $h = Id$
- * modèle dit de *Pareto* (II): $F_{ij}^* = a_i O_i b_j D_j h(c_{ij})$ avec $h(c_{ij}) = c_{ij}^{-\alpha}$
- * modèle dit de *Pareto+barrière* (III): $F_{ij}^* = a_i O_i b_j D_j h(c_{ij}) \gamma^{A_{ij}}$ avec $h(c_{ij}) = c_{ij}^{-\alpha}$
- * modèle dit de *barrière seule* (IV): $F_{ij}^* = a_i O_i b_j D_j \gamma^{A_{ij}}$

$\gamma^{A_{ij}} / A_{ij} = 0$ (flux inter-régions), $A_{ij} = 1$ (flux intra-régions)

Les erreurs de chaque type de modèle seront exprimées par rapport au modèle dit de référence (mod. I)

$$Err(F, F^*) = \sum_{ij} (F_{ij} - F_{ij}^*)^2 / F_{ij}^* \quad (\text{critère du type CHI2: } \chi^2_{F, F^*})$$

- ◆ cinq types de distances sont prises en compte (segmentation par nombre de paires se situant à plus de ...): distance euclidienne, distances à vol d'oiseau entre zones d'emploi: seuil=0, =50 km, =80. Ces distances ont été établies à partir du fonds de carte des zones d'emploi 1994 (INSEE).

- ◆ à titre de test deux nouvelles variables avaient été définies ($prop_{ab}$, d_{ab}):

* $S_{ij} = 0$ ou $= 1$ (si zones d'emploi sont entre elles à un seuil de distance d_{ij})

$$* prop_{ab} = \left[\sum_{i \in a} \sum_{j \in b} P_i P_j S_{ij} \right] / \left[\sum_{i \in a} \sum_{j \in b} P_i P_j \right]$$

(proportion de paires d'OD qui vérifient la condition S_{ij})

$$* d_{ab} = \left[\sum_{i \in a} \sum_{j \in b} P_i P_j S_{ij} d_{ij} \right] / \left[\sum_{i \in a} \sum_{j \in b} P_i P_j S_{ij} \right]$$

Les tests d'introduction des dites variables n'ont pas amélioré statistiquement les résultats d'estimation et donc celles-ci ont été de facto écartées.

MESURE DE LA QUALITE DES MODELES

La qualité du modèle est déterminée par l'erreur entre la matrice de flux observés et la matrice des flux estimés: $Err(F, F^*)$. Cette erreur permet d'apprécier l'adéquation du modèle utilisé par rapport au critères étudiés. Toute norme matricielle est utilisable pour mesurer cette erreur. Trois types de mesure sont communément employées:

1. la mesure en *valeur absolue*:

$$M1 = Err(F, F^*) = \sum_{ij} F_{ij} - F_{ij}^*$$

2. la mesure du *carré de l'erreur*:

$$M2 = Err(F, F^*) = \sum_{ij} (F_{ij} - F_{ij}^*)^2$$

3. la mesure du χ^2_{F, F^*} :

$$M3 = Err(F, F^*) = \sum_{ij} \frac{(F_{ij} - F_{ij}^*)^2}{F_{ij}^*}$$

Cette dernière mesure M3 est plus coûteuse en temps de calcul que la seconde, elle-même plus coûteuse que la première ($M3 > M2 > M1$). Il faut en outre choisir une mesure permettant une interprétation claire des résultats. Les deux premières mesures (M1 et M2) seront utilisées en présence de données homogènes, alors que la troisième (M3) le sera pour des données hétérogènes. En effet, lorsqu'il s'agit de données hétérogènes, les deux premières mesures négligeront l'action des petites valeurs au profit des grosses, alors que la mesure du χ^2 (M3) réalisera une normalisation des données avant traitement. Cette normalisation s'avérera utile dans le cas d'unités géographiques très éloignées dont les valeurs de flux seront souvent faibles. Les résultats ainsi obtenus serviront à définir une fonction de coût dont la minimisation permettra l'ajustement des paramètres du modèle.

PLATE-FORME EXPERIMENTALE UTILISEE

La plate-forme expérimentale se base sur une première version de la mesure de qualité des modèles réalisée par les géographes de l'équipe PARIS. Dans cette première version la formulation de l'erreur (Err) était:

$$MA(\chi^2) = \sum_{ij} \frac{(F_{ij} - MA(F))^2}{MA(F)} \quad \chi^2_{F,F^*} = \sum_{ij} \frac{(F_{ij} - F_{ij}^*)^2}{F_{ij}^*}$$

$$Err = \chi^2_{F,F^*} / MA(\chi^2)$$

avec :

$$MA(F): \text{Moyenne arithmétique des valeurs de la matrice des flux observés } F = \frac{1}{n^*(n-1)} \sum_j F_{ij}$$

n: dimension de la matrice carrée.

Nous voyons ici que le χ^2 calculé est normalisé par un ' χ^2 moyen', Ce dernier nous permet d'apprécier la dispersion des flux observés par rapport à la moyenne.

Dans un même ordre d'idée, il nous a paru intéressant de calculer une erreur propre à la matrice de distance employée. Cette approche nous permet d'apprécier l'apport de la 'complexification' de la fonction de distance introduite. Pour se faire, nous calculons une *erreur de référence* pour chaque matrice en utilisant le modèle à doubles contraintes avec la fonction d'identité: χ^2_{F,F^*id} . Cette valeur nous servira à normaliser les erreurs calculées avec d'autres types de fonction de la distance.

La formulation de l'erreur du modèle (ϵ) devient:

$$\chi^2_{F,F^*id} = \sum_i \frac{\sum_j (F_{ij} - F^* Id_{ij})^2}{F^* Id_{ij}} \quad \epsilon = \chi^2_{F,F^*} / \chi^2_{F,F^*id}$$

La définition de cette erreur nous permet de construire une fonction coût représentative de chaque option de modélisation.

EVALUATION DES PARAMETRES DU MODELE

L'évaluation des paramètres est effectuée à l'aide d'une fonction d'optimisation. Celle qui est actuellement disponible sur la plate-forme est nommée 'Hooke'. Elle a été implantée par Mark G. Johnson (déc.1994), d'après la méthode de 'recherche directe' développée par Hooke et Jeeves. Le principe général de l'algorithme mis en place repose sur la 'recherche directe' d'une solution à des problèmes numériques ou statistiques. La méthode consiste à examiner un ensemble de solutions triviales afin de déterminer une stratégie de construction permettant de former un nouvel ensemble de solutions, inclus dans le premier, de meilleure qualité. Si nous prenons l'exemple d'une fonctionnelle f à minimiser, ceci revient à évaluer sa valeur en un point initial x_0 et à trouver un nouveau point x_1 tel que:

$$x_1 = x_0 \pm \delta_0 * f(x_1) < f(x_0)$$

La valeur de δ_0 (pas de recherche de la solution) est fixée à l'initialisation. δ_0 est multiplié par un facteur $\rho < 1$ si $f(x_0 \pm \delta_0) \leq f(x_0)$. De cette façon, x_i converge vers \underline{x} qui minimise la fonctionnelle f . La condition d'arrêt de l'optimisation est réalisée lorsque $f(x_{i+1}) - f(x_i) < \nu$ ou lorsque le nombre d'itérations maximum est atteint.

Cette fonction d'optimisation a été choisie car elle ne nécessite pas de conditions particulières sur la linéarité ou la dérivabilité de la fonctionnelle considérée. Néanmoins, l'usage a montré que son comportement est très sensible aux conditions initiales (valeur de x_0 , δ_0 et ρ) et à la forme de la fonctionnelle étudiée (si l'on se trouve au milieu d'un plateau, de diamètre supérieur à $2 * \delta_0$, la fonction Hooke ne convergera pas). De plus l'optima trouvé peut être local ou global sans qu'on puisse le savoir. Le principal défaut de cette fonction est son caractère itératif qui la rend très coûteuse en temps de calcul. Dans le cas qui nous concerne (utilisation du χ^2 comme fonction de coût), le problème de la dérivabilité ne se pose pas, néanmoins comme la formulation du modèle de Wilson est intrinsèquement de nature itérative et donc rend l'expression de la dérivée difficilement exploitable pour la détermination des paramètres optimaux.

RECHERCHE DE PARTITIONNEMENT DE L'ESPACE GEOGRAPHIQUE

- Rappels sur la problématique des MIS :

Classiquement il s'agira de fixer une règle de mobilité f permettant de reconstituer au mieux une matrice de flux estimés $[F^*]$ à partir de E , R et D .

E et R : vecteurs des sommes de départ et d'arrivée.

D : matrice de coût généralisé moyen des déplacements.

$$[F^*] = f(E, R, [D])$$

“ *L'hypothèse centrale de ces modèles (MIS) est que la distance, c'est à dire la proximité spatiale mesurée en kilomètres, temps ou coûts a une influence déterminante sur la probabilité de relation entre les habitants ou groupes sociaux répartis sur un territoire* ”

La modélisation de l'effet d'appartenance doit permettre de mettre en évidence “ *des formes d'intégration territoriale spécifique, autrement dit des portions de territoires dans lesquelles les habitants adoptent un comportement migratoire significativement différent de celui des habitants d'autres portions du territoire* ”.

[Claude Grasland, 1996].

Questions posées :

- l'introduction d'une ou plusieurs partitions territoriales *a priori* dans le modèle entraîne-t-elle un gain d'ajustement ?

- *recherchons la partition optimale a posteriori* qui garantisse le meilleur ajustement, ce qui nous conduit à résoudre des problèmes de partition d'ensembles connexes. La recherche de la configuration optimale parmi l'ensemble de toutes les configurations possibles peut vite s'avérer lourde et très coûteuse en temps de calcul. Il a donc fallu recourir aux techniques dites de *parallélisation*, dont schématiquement on peut décrire les différentes phases :

- * phase n°1 de parallélisation : répartition des calculs sur un nombre de processeurs p ,

- * phase n°2 : application de méthodes de répartition de charges afin d'éviter les déséquilibres entre les processeurs,

- * phase n°3 : Classification des différentes partitions par le biais de la fonction de coût associée au modèle.

La partition optimale, est celle dont la valeur de la fonction de coût est minimale (e).

avec $e = \chi^2_{F,F^*} / \chi^2_{F,F^*id}$

CLASSIFICATION EXHAUSTIVE SUR L'ENSEMBLE DES PARTITIONS

Une étude exhaustive vise à rechercher la configuration optimale parmi l'ensemble de toutes les configurations possibles. Cette démarche s'avère très vite lourde et coûteuse en temps de calcul, en effet selon la configuration considérée le nombre d'expérimentations peut varier exponentiellement. Ainsi un graphe totalement connecté de taille 10 permet de construire $2^{10-1} = 512$ partitions un-deux, $3^{10-1} = 19\ 683$ partitions en trois, $4^{10-1} = 262\ 144$ partitions en quatre, Dans le même ordre d'idée, il est intéressant de noter que les trois formes de découpage de la France, ci-après retenues ont permis de générer respectivement: 1 085 partitions France entière, 1 116 partitions France sans Paris. 30 103 partitions France avec Paris neutre.

Une fois les résultats obtenus, il devient possible de classer les différentes partitions par la biais de la fonction de coût associée au modèle. Ceci permet de définir la partition optimale, c'est à dire la partition dont la valeur de la fonction de coût est minimale.

Le principe de cette méthode de classification oblige à prendre en compte l'ensemble des cas possibles sans distinction. La complexité du traitement d'une structure géographique sera donc fonction de la taille de son « partitionnement ».

La complexité algorithmique de la fonction coût est en $O(n^2)$ (calcul d'un χ^2), donc celle de la fonction d'optimisation sera du même ordre. Afin de mieux appréhender l'impact de la complexité algorithmique sur les limites de la plate-forme expérimentale, celle-ci sera traduite en un temps de calcul moyen (les temps peuvent varier d'un facteur 2 selon la partition traitée) pour un modèle donné.

Le tableau permet de matérialiser le problème du temps de calcul nécessaire pour mener à bien cette expérimentation. Il apparaît clairement qu'un nombre excessif de partitions entraîne des temps de calcul ne permettant pas une utilisation aisée de la plate-forme. Afin de remédier à cela, une première « phase de parallélisation » a été effectuée, elle correspond à une répartition des calculs sur un nombre de processeurs p . Les partitions sont réparties sur chacun des processeurs uniformément. Ces derniers réalisent les opérations nécessaires pour chaque partition et enregistrent les résultats dans un fichier personnel. C'est lors du traitement de ces résultats que les fichiers sont fusionnés et triés. Chaque processeur traite donc un problème de taille inférieure. En première approximation, les temps de calculs sont en $T_p = O(T/p)$, pour tout p inférieur ou égal au nombre total de partitions. L'avantage d'une telle technique est qu'elle ne présuppose aucune caractéristique particulière pour le matériel employé, ce qui lui confère une 'portabilité' maximale.

Dans un deuxième temps, il est nécessaire d'appliquer les méthodes de répartition de charges afin d'éviter les déséquilibres entre les processeurs. En effet les calculs nécessaires à l'optimisation de la fonction coût varient selon la partition qui lui est associée. Des situations où certains processeurs ont fini de traiter les partitions qui lui sont attribué, alors que d'autres sont en cours de calcul peuvent survenir. Une redistribution des partitions restantes permettrait d'augmenter l'efficacité de la parallélisation.

Une troisième voie de développement porterait sur la parallélisation de la fonction de coût. Celle-ci se base sur un calcul du χ^2 dont la complexité algorithmique est en $O(n^2)$ sur machine séquentielle et qui passe en $O(\log n)$ sur machine parallèle. Bien que les temps de calcul puissent être diminués par ces différentes options de parallélisme, certaines situations obligent à sélectionner les partitions à traiter lorsque celles-ci sont trop nombreuses.

[voir aussi partie ALGORITHMES en annexe]

***Temps moyen de calcul par type de modèle utilisé
en secondes et heures/minutes/secondes***

<i>Nombre de partitions en 2 parties connexes</i>				
	<i>Référence</i>	<i>avec PARIS</i>	<i>sans PARIS</i>	<i>PARIS neutre</i>
	1	2 085	1 116	3 103
BARRIERE	2	4 170 01:09:30	2 232 00:37:12	6 206 01:43:26
PARETO+BARRIERE	8	16 680 04:38:00	8 928 02:28:48	2 484 06:53:44

PRINCIPAUX RESULTATS

Les principaux résultats présentés [pour l'ensemble se reporter au rapport général] ne concernent que les meilleurs découpages de l'espace géographique, partitions de rang 1.

Ces résultats seront commentés en regard de plusieurs éléments:

- comparaisons pour un même mode entre matrices d'origine statistique différente,
- cartes modales des charges de réseau,
- 'cartes de potentiels' de population et de densité de population en France métropolitaine en 1990,

* une première remarque générale: le modèle simple de Pareto explique en moyenne plus de 80% des flux, avec une différence notable pour la matrice fer issue de l'ETC (56% de la variance expliquée).

* l'introduction de l'effet de barrière dans le modèle de Pareto apparaît significative pour les matrices fer, d'autant plus si l'on regarde les spécifications anti-barrière.

* si l'on cumule l'ensemble des partitions de toutes les matrices par type de modèle et si l'on regarde à la fois les cartes d'affectation des trafics et les cartes dites de potentiel de population établies par l'équipe PARIS, des analogies apparaissent immédiatement [voir carte].

* la mesure des effets frontières à partir de la partition n°1 laisse apparaître une faible dispersion intermodale des valeurs [voir tableau].

Une analyse des paramètres de la fonction de distance (α) par type de modèle et par matrice pour les dix meilleurs découpages a été effectuée. Il en a été fait de même pour le paramètre de découpage (γ). L'ensemble de ces résultats est repris in-extenso dans le rapport général.

La plate-forme informatique développée dans le cadre de cette étude devrait servir de support à d'autres travaux. Notamment le groupe de travail inter-disciplinaire monté à cette occasion a en effet décidé de poursuivre cette première réflexion au cas du rôle des effets de réseaux sur la réduction dans le temps de l'effet barrière pour les échanges physiques de marchandises entre la France, l'Allemagne et la Belgique.

TSRESUL

Types de modèles		REFERENCE		PARETO		IS SEULE				PARETO+BARRIERE					
		I		II		IV		III							
		ERR		ϵ	α	BARRIERE		ANTI-BARRIERE		BARRIERE			ANTI-BARRIERE		
		ϵ	$\gamma > 1$	ϵ	$\gamma < 1$	α	ϵ	$\gamma > 1$	α	ϵ	$\gamma < 1$	α	ϵ	$\gamma < 1$	
MATISSE ROUTE	M31	0,35	0,06	-2,21	0,581	3,336	0,965	0,6669	-2,191	0,048	1,426	-2,325	0,049	0,681	
MATISSE FER	M32	0,11	0,18	-1,41	0,548	2,535	0,967	0,7503	-1,355	0,142	1,54	-1,754	0,131	0,607	
MATISSE TOTAL	M34	0,29	0,10	-1,91	0,597	2,969	0,966	0,6569	-1,912	0,071	1,523	-2,18	0,077	0,654	
ETC ROUTE	MS15	0,57	0,09	-2,19	0,604	3,502	0,978	0,0847	-2,111	0,066	1,552	-2,442	0,068	0,586	
ETC FER	MS16	0,06	0,45	-1,04	0,624	2,149	0,9	0,5766	-1,103	0,391	1,451	-1,373	0,353	0,555	
ETC TOTAL	MS19	0,42	0,12	-1,92	0,607	3,3	0,977	0,3907	-1,857	0,097	1,449	-2,245	0,091	0,601	
	VLS	0,44	0,02	-3,19	0,581	4,311	0,972	0,2398	-3,085	0,015	1,797	-3,28	0,019	0,735	
	VLS2	0,23	0,04	-2,59	0,544	3,649	0,943	0,6151	-2,517	0,031	1,427	-2,643	0,031	0,771	
	VLS3	0,19	0,13	-2,29	0,581	4,311	0,972	0,2398	-2,226	0,113	1,934	-2,431	0,118	0,711	
PARIS NEUTRE															
	M31				0,669	4,201	0,975	0,7073	-2,106	0,047	1,672	-2,283	0,056	0,636	
	M32				0,669	3,654	0,948	0,5296	-1,392	0,142	1,689	-1,482	0,167	0,742	
	M34				0,668	4,012	0,977	0,6758	-1,813	0,076	1,728	-1,987	0,089	0,639	
	MS15				0,636	4,359	0,99	0,7489	-2,048	0,069	1,707	-2,219	0,084	0,49	
	MS16				0,708	3,484	0,928	0,4595	-1,028	0,384	1,94	-1,057	0,41	0,652	
	MS19				0,631	4,753	0,985	0,7659	-1,786	0,091	1,775	-1,964	0,109	0,435	
	VLS				0,684	5,023	0,982	0,6728	-3,134	0,016	1,768	-3,249	0,019	0,69	
	VLS2				0,681	4,341	0,948	0,1011	-2,474	0,031	1,404	-2,635	0,029	0,727	
	VLS3				0,738	4,066	0,949	0,4961	-2,239	0,117	1,537	2,2986	0,118	0,679	
SANS PARIS															
	M31				0,669	4,201	0,975	0,7073	-2,106	0,047	1,672	-2,29	0,058	0,65	
	M32				0,669	3,654	0,945	0,5301	-1,344	0,148	1,653	-1,482	0,167	0,742	
	M34				0,668	4,012	0,977	0,6758	-1,813	0,076	1,728	-1,951	0,09	0,505	
	MS15				0,636	4,359	0,99	0,7489	-2,047	0,069	1,707	-2,219	0,084	0,49	
	MS16				0,708	3,484	0,928	0,4595	-1,075	0,388	2,018	-1,145	0,413	0,634	
	MS19				0,631	4,753	0,985	0,7659	-1,786	0,091	1,775	-1,964	0,109	0,435	
	VLS				0,684	5,023	0,982	0,6728	-3,134	0,016	1,768	-3,234	0,02	0,722	
	VLS2				0,681	4,341	0,948	0,1011	-2,474	0,031	1,404	-2,713	0,03	0,592	
	VLS3				0,738	4,066	0,949	0,4961	-2,25	0,118	1,66	-2,341	0,119	0,673	

Part expliquée par le modèle
(en référence au modèle dit de référence)

<i>Types de modèles</i>	REFERENCE	PARETO	IS SEULE		PARETO+BARRIERE		GAINS III/II	
	I	II	IV		III		<i>en points</i>	
			B	AB	B	AB	B	AB
	ERR	ε	ε	ε	ε	ε		
ROUTE [MATISSE]	0,355	93,8%	41,9%	3,6%	95,2%	95,2%	1,4%	1,4%
FER [MATISSE]	0,108	82,3%	-90,8%	-216,8%	86,1%	87,0%	3,8%	4,7%
TOTAL [MATISSE]	0,288	90,3%	40,3%	3,5%	93,1%	92,4%	2,7%	2,0%
ROUTE [ETC]	0,569	91,1%	39,5%	2,3%	93,5%	93,1%	2,4%	2,1%
FER [ETC]	0,055	55,5%	38,5%	9,5%	60,2%	63,8%	4,7%	8,3%
TOTAL [ETC]	0,422	88,2%	39,3%	2,2%	90,3%	90,8%	2,1%	2,6%
VLS	0,444	97,7%	41,9%	2,8%	98,4%	98,2%	0,8%	0,5%
VLS2	0,228	96,5%	45,5%	5,5%	96,9%	96,9%	0,4%	0,4%
VLS3	0,194	87,1%	43,8%	7,7%	88,7%	88,1%	1,5%	1,0%

ERR

Types de modèles	REFERENCE	PARETO	Part expliquée par le modèle										GAINS III/II		
			I	II	IS SEULE		PARETO+BARRIERE		II	IS SEULE		PARETO+BARRIERE		points	
					IV		III			IV	III	B	AB		
					B	AB	B	AB						B	AB
ERR	ERR	ERR	ERR	ERR	ERR	ε	ε	ε	ε	ε	B	AB			
MATISSE ROUTE	M31	0,355	0,022	0,206	0,342	0,017	0,017	93,8%	41,9%	3,6%	95,2%	95,2%	1,4%	1,4%	
MATISSE FER	M32	0,108	0,019	0,206	0,342	0,015	0,014	82,3%	-90,8%	-216,8%	86,1%	87,0%	3,8%	4,7%	
MATISSE TOTAL	M34	0,288	0,028	0,172	0,278	0,020	0,022	90,3%	40,3%	3,5%	93,1%	92,4%	2,7%	2,0%	
ETC ROUTE	MS15	0,569	0,051	0,344	0,556	0,037	0,039	91,1%	39,5%	2,3%	93,5%	93,1%	2,4%	2,1%	
ETC FER	MS16	0,055	0,025	0,034	0,050	0,022	0,020	55,5%	38,5%	9,5%	60,2%	63,8%	4,7%	8,3%	
ETC TOTAL	MS19	0,422	0,050	0,256	0,413	0,041	0,039	88,2%	39,3%	2,2%	90,3%	90,8%	2,1%	2,6%	
	VLS	0,444	0,010	0,258	0,432	0,007	0,008	97,7%	41,9%	2,8%	98,4%	98,2%	0,8%	0,5%	
	VLS2	0,228	0,008	0,124	0,215	0,007	0,007	96,5%	45,5%	5,5%	96,9%	96,9%	0,4%	0,4%	
	VLS3	0,194	0,025	0,109	0,179	0,022	0,023	87,1%	43,8%	7,7%	88,7%	88,1%	1,5%	1,0%	
PARIS NEUTRE													GAINS / REF.		
	M31	0,355		0,2370	0,3460	0,0170	0,0200	33,2%	2,5%	95,2%	94,4%	94,4%	0,0%	-0,8%	
	M32	0,108		0,0720	0,1020	0,0150	0,0180	33,3%	5,5%	86,1%	83,3%	83,3%	0,0%	-2,8%	
	M34	0,288		0,1920	0,2810	0,0220	0,0260	33,4%	2,5%	92,4%	91,0%	91,0%	-0,7%	-2,1%	
	MS15	0,569		0,3620	0,5630	0,0390	0,0480	36,4%	1,0%	93,1%	91,6%	91,6%	-0,4%	-1,9%	
	MS16	0,055		0,0390	0,0510	0,0210	0,0230	29,4%	7,7%	62,0%	58,4%	58,4%	1,8%	-1,8%	
	MS19	0,422		0,2660	0,4160	0,0380	0,0460	37,0%	1,4%	91,0%	89,1%	89,1%	0,7%	-1,2%	
	VLS	0,444		0,3040	0,4360	0,0070	0,0080	31,6%	1,9%	98,4%	98,2%	98,2%	0,0%	-0,2%	
	VLS2	0,228		0,1550	0,2160	0,0070	0,0070	31,9%	5,1%	96,9%	96,9%	96,9%	0,0%	0,0%	
	VLS3	0,194		0,1430	0,1840	0,0230	0,0230	26,3%	5,1%	88,1%	88,1%	88,1%	-0,5%	-0,5%	
SANS PARIS													GAINS / REF.		
	M31	0,355		0,2370	0,3460	0,0170	0,0210	33,2%	2,5%	95,2%	94,1%	94,1%	0,0%	-1,1%	
	M32	0,108		0,0720	0,1020	0,0160	0,0180	33,3%	5,5%	85,2%	83,3%	83,3%	-0,9%	-2,8%	
	M34	0,288		0,1920	0,2810	0,0220	0,0260	33,4%	2,5%	92,4%	91,0%	91,0%	-0,7%	-2,1%	
	MS15	0,569		0,3620	0,5630	0,0390	0,0480	36,4%	1,0%	93,1%	91,6%	91,6%	-0,4%	-1,9%	
	MS16	0,055		0,0390	0,0510	0,0210	0,0230	29,4%	7,7%	62,0%	58,4%	58,4%	1,8%	-1,8%	
	MS19	0,422		0,2660	0,4160	0,0380	0,0460	37,0%	1,4%	91,0%	89,1%	89,1%	0,7%	-1,2%	
	VLS	0,444		0,3040	0,4360	0,0070	0,0090	31,6%	1,9%	98,4%	98,0%	98,0%	0,0%	-0,5%	
	VLS2	0,228		0,1550	0,2160	0,0070	0,0070	31,9%	5,1%	96,9%	96,9%	96,9%	0,0%	0,0%	
	VLS3	0,194		0,1430	0,1840	0,0230	0,0230	26,3%	5,1%	88,1%	88,1%	88,1%	-0,5%	-0,5%	

TOUS

Types de modèles		REFERENCE	PARETO		IS SEULE				PARETO+BARRIERE					
		I	II		IV				III					
					BARRIERE		ANTI-BARRIERE		BARRIERE			ANTI-BARRIERE		
ERR	ϵ	α	ϵ	$\gamma > 1$	ϵ	$\gamma < 1$	α	ϵ	$\gamma > 1$	α	ϵ	$\gamma < 1$		
MATISSE TOTAL	M34	0,29	0,10	-1,91	0,597	2,969	0,966	0,6569	-1,912	0,071	1,523	-2,18	0,077	0,654
MAT. TOT. P.N.	M34				0,668	4,012	0,977	0,6758	-1,813	0,076	1,728	-1,987	0,089	0,639
MAT. TOT. S.P.	M34				0,668	4,012	0,977	0,6758	-1,813	0,076	1,728	-1,951	0,09	0,505
ETC TOTAL	MS19	0,42	0,12	-1,92	0,607	3,3	0,977	0,3907	-1,857	0,097	1,449	-2,245	0,091	0,601
ETC TOTAL P.N.	MS19				0,631	4,753	0,985	0,7659	-1,786	0,091	1,775	-1,964	0,109	0,435
ETC TOTAL S.P.	MS19				0,631	4,753	0,985	0,7659	-1,786	0,091	1,775	-1,964	0,109	0,435

ROUTE

<i>Types de modèles</i>		REFERENCE	PARETO		IS SEULE				PARETO+BARRIERE					
		I	II		IV		III			ANTI-BARRIERE				
		ERR	ϵ	α	ϵ	$\gamma > 1$	ϵ	$\gamma < 1$	α	ϵ	$\gamma > 1$	α	ϵ	$\gamma < 1$
<i>nombre de voyageurs</i>														
MATISSE	M31	0,35	0,06	-2,21	0,581	3,336	0,965	0,6669	-2,191	0,048	1,426	-2,325	0,049	0,681
MAT. PARIS NEUTRE	M31				0,669	4,201	0,975	0,7073	-2,106	0,047	1,672	-2,283	0,056	0,636
MAT. SANS PARIS	M31				0,669	4,201	0,975	0,7073	-2,106	0,047	1,672	-2,29	0,058	0,65
ETC ROUTE	MS15	0,57	0,09	-2,19	0,604	3,502	0,978	0,0847	-2,111	0,066	1,552	-2,442	0,068	0,586
ETC PARIS NEUTRE	MS15				0,636	4,359	0,99	0,7489	-2,048	0,069	1,707	-2,219	0,084	0,49
ETC SANS PARIS	MS15				0,636	4,359	0,99	0,7489	-2,047	0,069	1,707	-2,219	0,084	0,49
<i>nombre de véhicules</i>														
SETRA 1	VLS	0,44	0,02	-3,19	0,581	4,311	0,972	0,2398	-3,085	0,015	1,797	-3,28	0,019	0,735
SETRA 1 - P.N.	VLS				0,684	5,023	0,982	0,6728	-3,134	0,016	1,768	-3,249	0,019	0,69
SETRA 1 - S.P.	VLS				0,684	5,023	0,982	0,6728	-3,134	0,016	1,768	-3,234	0,02	0,722
SETRA 2	VLS2	0,23	0,04	-2,59	0,544	3,649	0,943	0,6151	-2,517	0,031	1,427	-2,643	0,031	0,771
SETRA 2 - P.N.	VLS2				0,681	4,341	0,948	0,1011	-2,474	0,031	1,404	-2,635	0,029	0,727
SETRA 2 - S.P.	VLS2				0,681	4,341	0,948	0,1011	-2,474	0,031	1,404	-2,713	0,03	0,592
MATISSE	VLS3	0,19	0,13	-2,29	0,581	4,311	0,972	0,2398	-2,226	0,113	1,934	-2,431	0,118	0,711
MATISSE - P.N.	VLS3				0,738	4,066	0,949	0,4961	-2,239	0,117	1,537	2,2986	0,118	0,679
MATISSE - S.P.	VLS3				0,738	4,066	0,949	0,4961	-2,25	0,118	1,66	-2,341	0,119	0,673

FER

Types de modèles		REFERENCE	PARETO		IS SEULE				PARETO+BARRIERE					
		I	II		IV				III					
						BARRIERE		ANTI-BARRIERE		BARRIERE			ANTI-BARRIERE	
	ERR	ϵ	α	ϵ	$\gamma > 1$	ϵ	$\gamma < 1$	α	ϵ	$\gamma > 1$	α	ϵ	$\gamma < 1$	
MATISSE FER	M32	0,11	0,18	-1,41	0,548	2,535	0,967	0,7503	-1,355	0,142	1,54	-1,754	0,131	0,607
MATISSE FER - P.N.	M32				0,669	3,654	0,948	0,5296	-1,392	0,142	1,689	-1,482	0,167	0,742
MATISSE - S.P.	M32				0,669	3,654	0,945	0,5301	-1,344	0,148	1,653	-1,482	0,167	0,742
ETC FER	MS16	0,06	0,45	-1,04	0,624	2,149	0,9	0,5766	-1,103	0,391	1,451	-1,373	0,353	0,555
ETC - P.N.	MS16				0,708	3,484	0,928	0,4595	-1,028	0,384	1,94	-1,057	0,41	0,652
ETC - S.P.	MS16				0,708	3,484	0,928	0,4595	-1,075	0,388	2,018	-1,145	0,413	0,634

Mesure de l'effet frontière

$\gamma > 1$

		-	PARIS NEUTRE	SANS PARIS
MATISSE ROUTE	M31	1,4	1,7	1,7
ETC ROUTE	MS15	1,6	1,7	1,7
MATISSE FER	M32	1,5	1,7	1,7
ETC FER	MS16	1,5	1,9	2,0
MATISSE TOTAL	M34	1,5	1,7	1,7
ETC TOTAL	MS19	1,4	1,8	1,8
SETRA ¹	VLS	1,8	1,8	1,8
SETRA ²	VLS2	1,4	1,4	1,4
MATISSE	VLS3	1,9	1,5	1,7

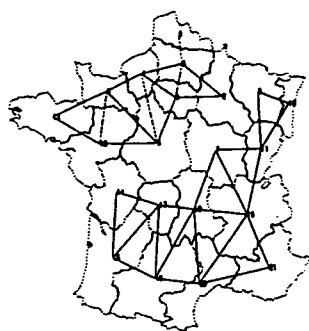
Spécification: **PARETO+BARRIERE**

$\gamma > 1$

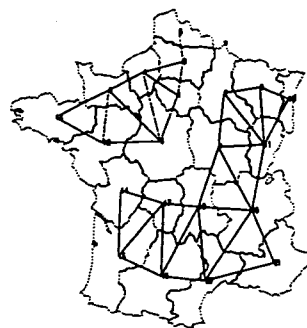
		-	PARIS NEUTRE	SANS PARIS
MATISSE ROUTE	M31	3,3	4,2	4,2
ETC ROUTE	MS15	3,5	4,4	4,4
MATISSE FER	M32	2,5	3,7	3,7
ETC FER	MS16	2,1	3,5	3,5
MATISSE TOTAL	M34	3,0	4,0	4,0
ETC TOTAL	MS19	3,3	4,8	4,8
SETRA ¹	VLS	4,3	5,0	5,0
SETRA ²	VLS2	3,6	4,3	4,3
MATISSE	VLS3	4,3	4,1	4,1

Spécification: **BARRIERE**

M31



MS15



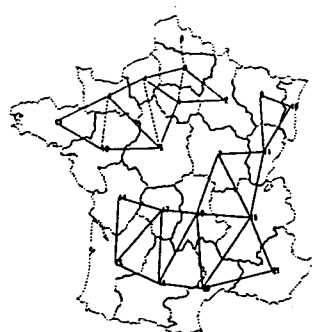
M32



MS16



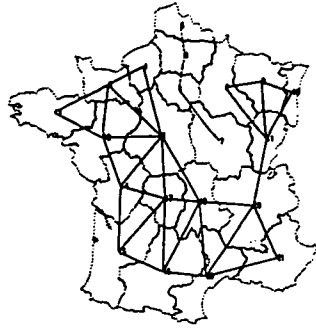
M34



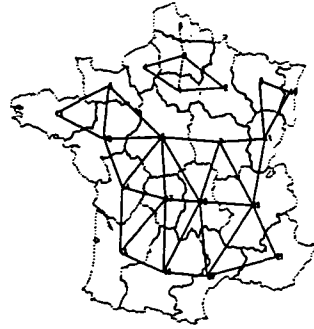
MS19



M31



MS15



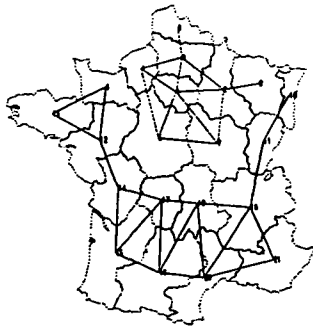
M32



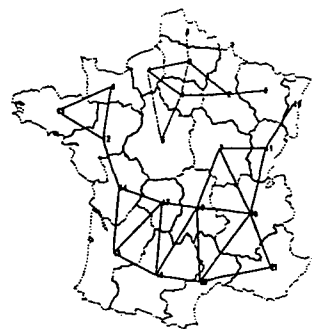
MS16



M34



MS19



M31



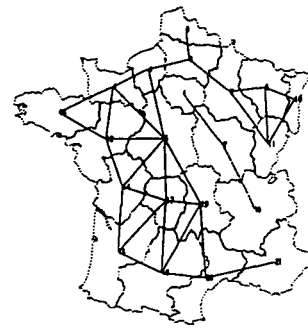
MS15



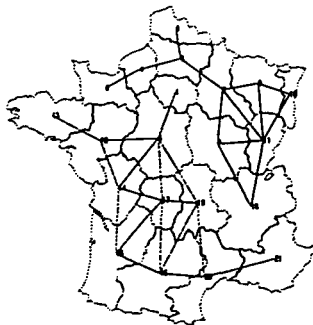
M32



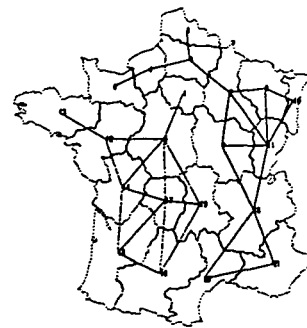
MS16



M34



MS19

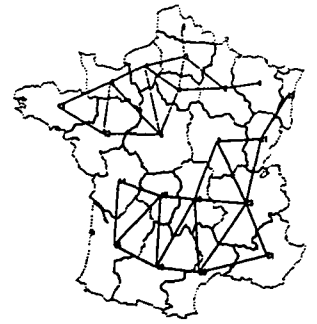


VLS

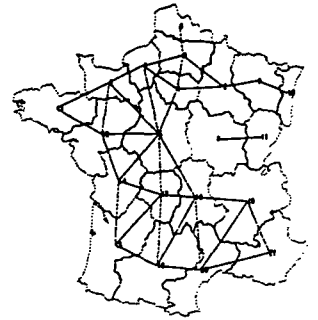
VLS 2

VLS 3

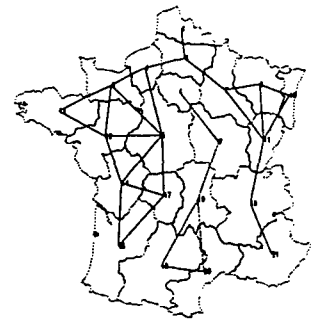
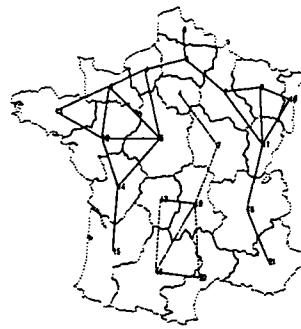
Barriere



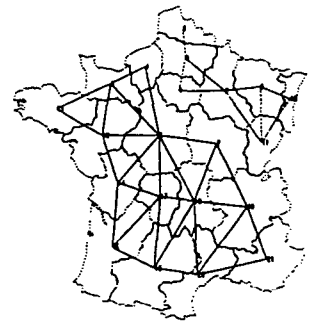
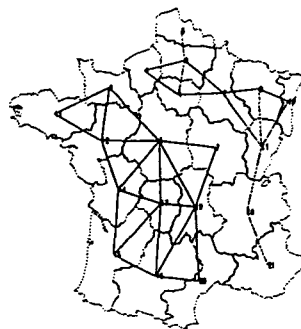
PARETO + Barriere



Anti-Barriere



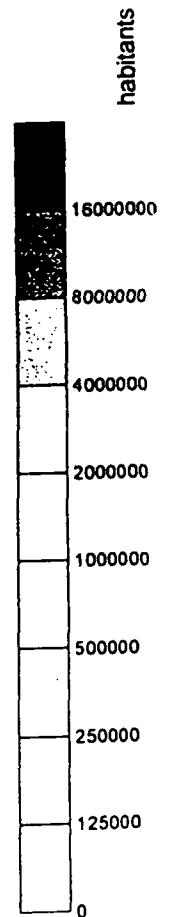
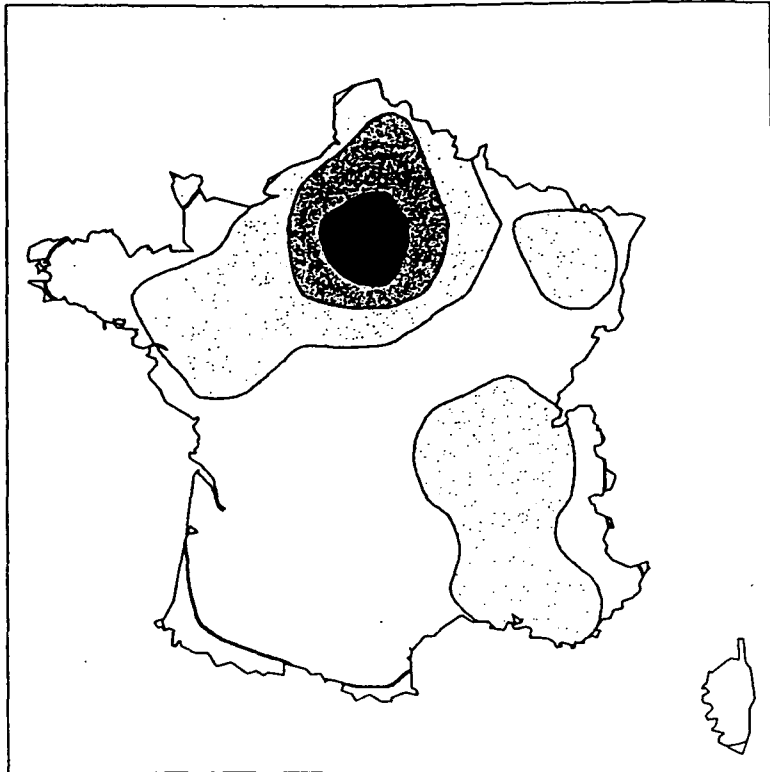
PARETO + Anti-Barriere



MATRICES DE TRAFICS (VL)

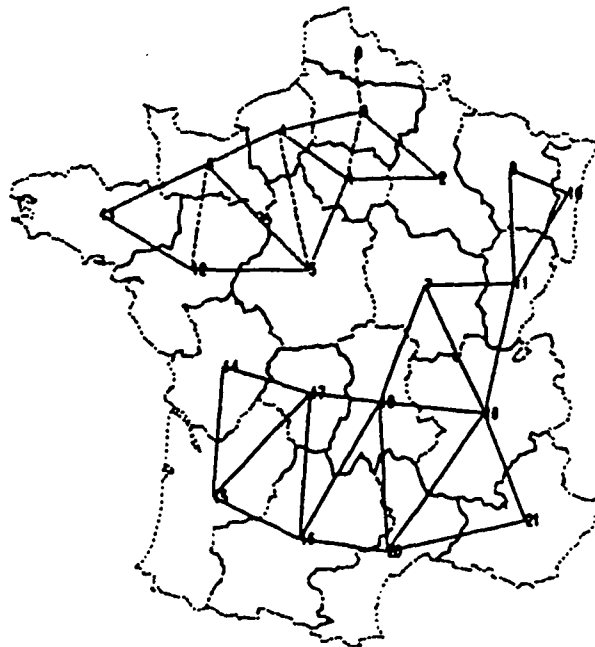
Potentiel de population en France (1990)

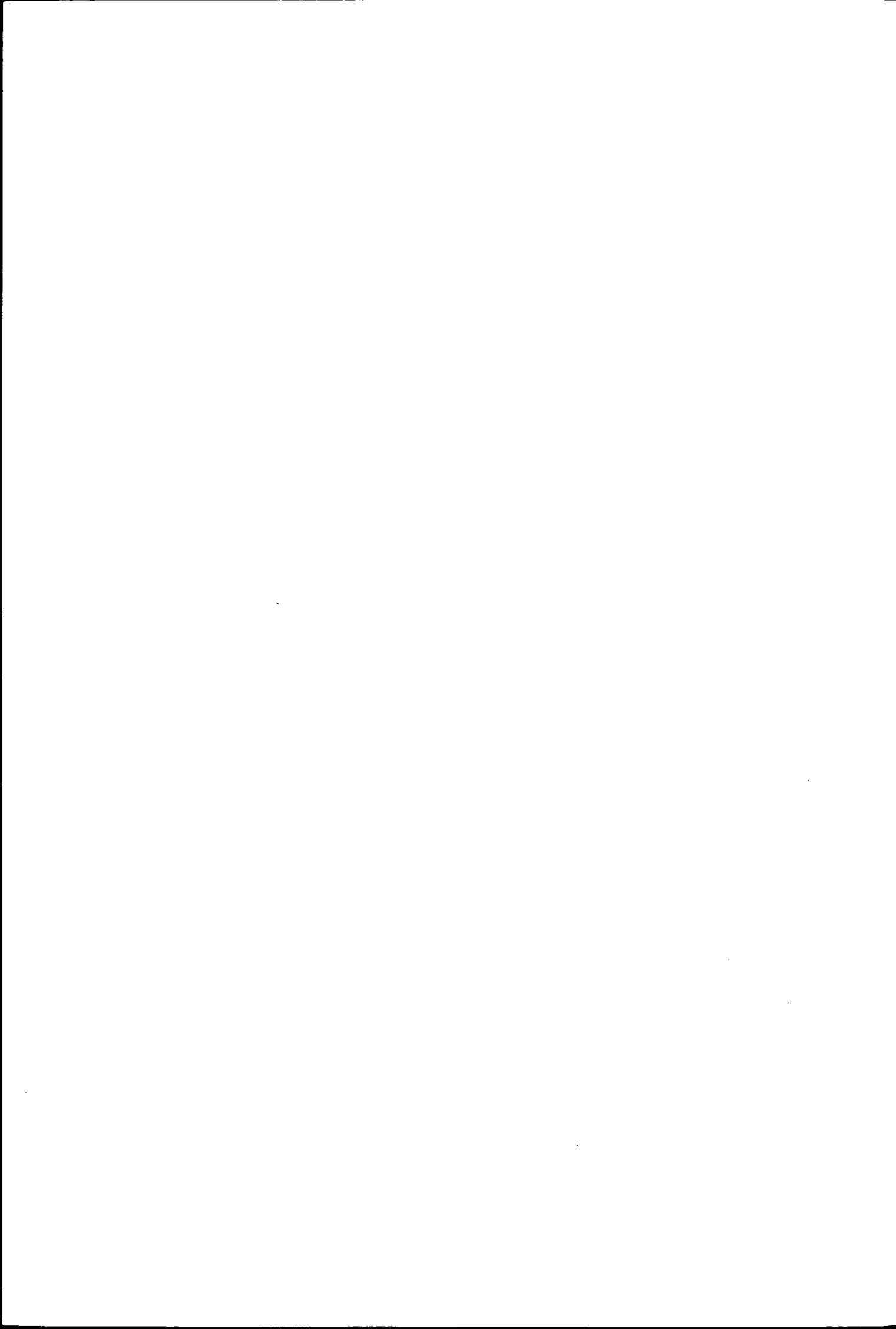
voisinage gaussien de portée 80 km



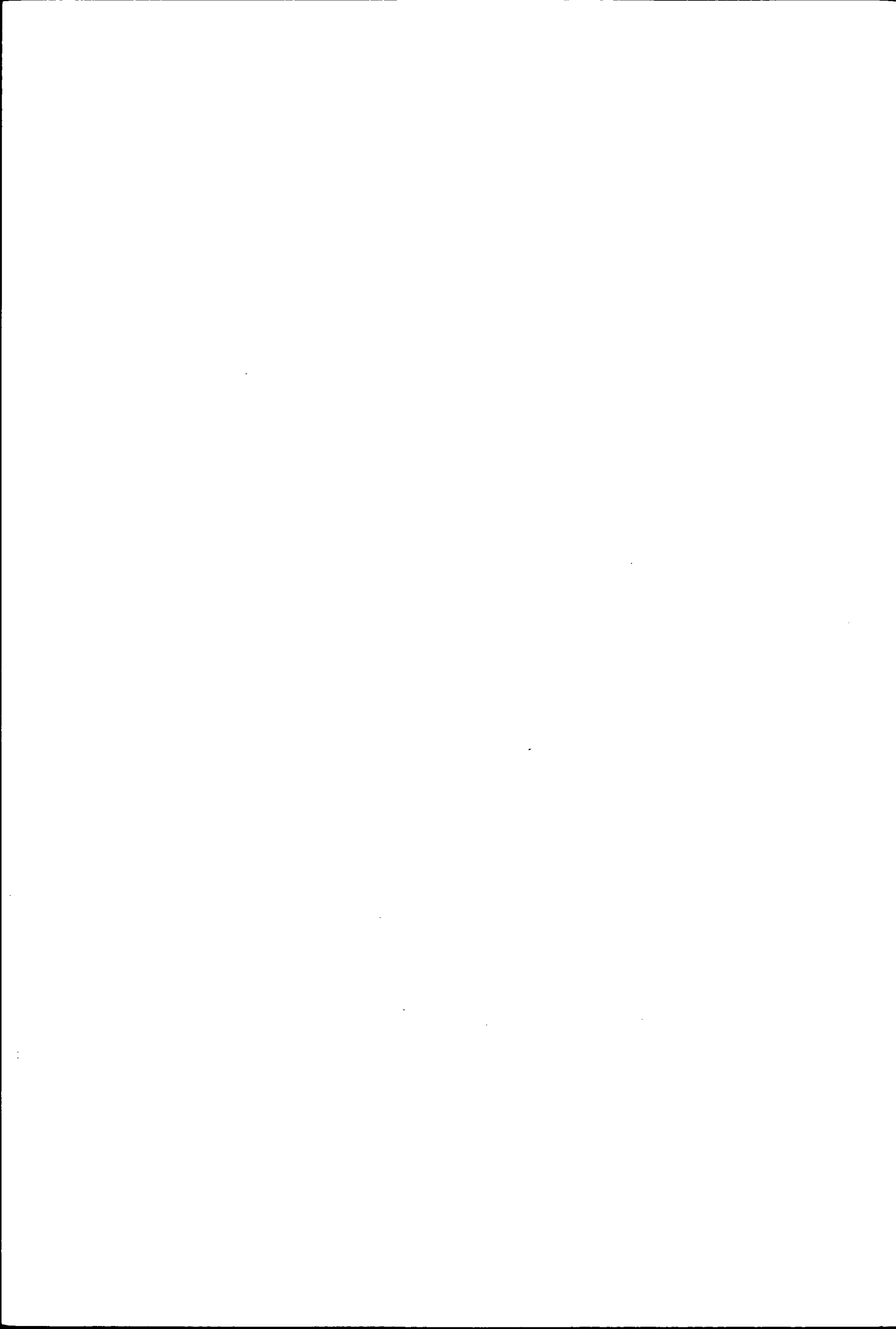
C. GRASLAND
CNRS-Equipe P.A.R.I.S.
(1997)

MODELE COMPILE (BARRIERE)





ANNEXES



QUELQUES AUTRES EXEMPLES DE MODELES 'GRAVITAIRES' EN COURS DE FONCTIONNEMENT

OUTIL POLYDROM

Forme générale du modèle de demande

Le trafic total T_{ij} entre deux pôles générateurs de trafic i et j se décompose de la manière suivante:

$$\text{avec } T_{ij}^{\lambda} = K^{\lambda} * [(P_i * P_j) * (R_{mij}^{\lambda} * d_{ij} + VT^{\lambda} * d_{ij}/V_{ij})^{a\lambda}] * PMT_{ij}^{\lambda} * PMC_{ij}^{\lambda} * E_{fij}^{\lambda}$$

λ : indice de classe tarifaire

a : exposant à caler (=2 par défaut)

P_i et P_j : populations des pôles concernés (éventuellement corrigés d'un effet de fréquentation touristique)

R_{mij}^{λ} : prix moyen payé par un voyageur de la classe λ considérée pour un trajet ij (revenu moyen pour l'opérateur transport)

VT^{λ} : valeur du temps pour les voyageurs de la classe considérée

d_{ij} et V_{ij} : distance et vitesse moyenne du voyage entre les pôles i et j

PMT_{ij}^{λ} : coefficient tenant compte de la concurrence entre transports (fer/avion)

PMC_{ij}^{λ} : coefficient tenant compte de la concurrence entre transports (fer/route)

E_{fij}^{λ} : effet frontière entre deux pays avec comme valeurs par défaut:

EF, étranger-étranger: 0

EF, étranger-France: 0,2 (id. EF, France-étranger)

EF, France-France: 1,0

Sources:

* Rapport ISIS (03/06/96): « Conception d'un outil d'évaluation socio-économique d'un réseau de transport ferroviaire de voyageurs », DTT.

* Rapport DTT « Contribution théorique à l'élaboration d'un schéma de service public de transport collectif intervilles de voyageurs », 1995, ISIS.

* POLYDROM - Modèle de transport intermodal de personnes et de marchandises, Manuel de l'utilisateur, version 7.1., février 1995, SYSTEMS CONSULT.

OUTIL REFLET II

Formulations

$$F_{ij} = k E_i A_j F(C_{ij})$$

avec :

E_i : populations ou emplois tertiaires de l'agglomération j (émission)

A_i : emplois tertiaires, populations ou nombre d'étudiants de l'agglomération i (attraction)

C_{ij} : coût généralisé de transports entre l'agglomération i et j

$F(C_{ij}) = 1/C_{ij}^n$ ou $\exp(-xC_{ij})$ suivant le motif (travail, professionnel, étudiant, personnel)

Paramètres du modèle « Motif personnel » :

$$F(C_{ij}) = 1/C_{ij}^n \text{ ou } \exp(-xC_{ij})$$

$$1/C^2 = 0,5 (1/C_{\text{fer}}^2 + C_{\text{route}}^2)$$

$$\begin{aligned} \text{avec } C_{\text{fer}} &= a1 * c1 * (d_i + d_j) + (c1 * T1_{ij} + h1 * P1_{ij}) * (1 + d1 / \sqrt{S_{ij}}) + (b1 * c1) * (1 + e1 / \sqrt{S_{ij}}) \\ &= 5,5 * 1 * (d_i + d_j) + (T_{ij} + 0,8 * P1_{ij}) * (1 + 1200 / \sqrt{S_{ij}}) + 27 * (1 + 7,5 / \sqrt{S_{ij}}) \\ C_{\text{route}} &= [a2 * T_{ij} (1 - e2 * V_{ij} / 100) + b2 * D_{ij}] * [1 + (c2 / t_i^2 - c2 / t_0^2)] + d2 * P2_{ij} \\ &= [1,1 * T_{ij} (1 - 0,15 * V_{ij} / 100) + 0,52 * D_{ij}] + P2_{ij} \end{aligned}$$

d_i et d_j : distances moyennes à la gare

$T1_{ij}$ et $T2_{ij}$: temps de transport par mode entre les agglomérations i et j

S_{ij} : nombre de possibilités quotidiennes de séjour dans l'agglomération j

$P1_{ij}$ et $P2_{ij}$: prix (billet train, péage autoroutier)

V_{ij} : % de 2 x 2 voies entre les agglomérations i et j

t_i et t_0 : taux de motorisation de l'agglomération i et moyenne

D'autres valeurs des paramètres sont définies pour les motifs travail, professionnel, étudiant.

Sources:

* « *Modèle REFLET 2 (Voyageurs), Modèle de Reconstitution des Flux de Voyageurs intercités* », A. GAUDEFROY, A. HUCHULSKI, Fiche F, ORT Pays de La Loire.

* « *Schéma régional transports des Pays de la Loire (Etude des trafics), dossier de base* », ORT Pays de la Loire, version projet, juin 1996.

* « *Schéma régional transports - Pays de la Loire, Etude des trafics voyageurs* », ORT Pays de la Loire, Région des Pays de la Loire, juin 1996.

OUTIL MODELE INTERURBAIN (Cofiroute/ENPC)

Contraintes sur la constance d'un certain nombre de variables: nombre de déplacements par habitant, nombre de kilomètres parcourus par habitant, distribution des distances de déplacements.

Modèle avec réajustement aux marges:

* étape n° 1: génération d'une matrice OD initiale

* étape n° 2: normalisation une à une des lignes (resp. colonnes) de telle sorte que la somme de chaque ligne (resp. colonne) corresponde à la marge fixée initialement. On itère sur les lignes et colonnes jusqu'à ce que l'écart maximum entre la somme de chaque ligne (resp. colonne) soit inférieur à une constante ε fixée. Les valeurs prises comme marges sont les sommes réelles des émissions et réceptions à l'exception des termes diagonaux.

* étape n° 3: modélisation avec réajustement aux marges: l'idée sous-jacente est que plutôt que de conditionner l'ajustement dans la première phase, par la disponibilité de variables économiques et démographiques pour qualifier les facteurs d'émission et d'attraction, on cherche ici après ajustement à relier ('ajuster') les dites marges à ces mêmes variables explicatives.

1. $T_i = k * P_i^\alpha$ avec P_i^α : population de l'unité spatiale i

2. $T_i = k * P_i^\alpha * \prod_n \%PCS_{n,i}^{\beta_n}$ avec $\%PCS$: part relative des actifs de PCSⁿ dans l'unité spatiale i

$\Leftrightarrow T_i = k * \prod_n [M(P_{pcs})/P_{tot}]^{\beta_n} * P_i^\alpha * \prod_n [(P_{pcs})/M(P_{pcs})]^{\beta_n}$
(minimisation de la somme des carrés des écarts à la réalité)

$M(P_{pcs})$ = population active de PCSⁿ qu'aurait l'unité spatiale si elle respectait la moyenne nationale

P_{tot} : population active de PCSⁿ dans le département i

P_{tot} : population totale dans l'unité spatiale i

Source:

* « Développement d'un modèle de trafic interurbain », J.B. GOUESSE, projet de fin d'études ENPC, Cofiroute, février-juin 1995, sous la direction de M. E. QUINET.

MODELE MATISSE

CRITERE DE CALIBRAGE

Le critère de calibrage est la minimisation de la somme sur l'ensemble des relations entre zones i et j des écarts définis comme suit :

- . si $N^{mij}/N^{t_{ij}} \geq 1$ alors l'écart est égal à $N_{ij}^t * [(\min(N_{mij}/N^{t_{ij}}, 10)^2 - 1)]$
- . si $N^{mij}/N^{t_{ij}} < 1$ alors l'écart est égal à $N_{ij}^t * [(1/\max(N_{mij}/N^{t_{ij}}, 0, 1))^2 - 1]$

où $N^{t_{ij}}$ est le nombre de trajets tous moyens de transport confondus pour le motif considéré,
 N^{mij} est le nombre de trajets donné par le modèle simple pour le motif considéré.

Un tel critère tend à réduire les très forts écarts entre modèle et trafic à reconstituer, dans un sens ou dans l'autre. Le modèle obtenu est tel que le nombre de trajets N^{mij} entre les zones i et j est proportionnel à :

tous motifs de trajets

$$F_{ij} = k * \alpha_{ij} * [P_i * P_j]^{0,60} * [R_i * R_j]^{0,75} * [1+0,25*T_i] * [1+0,25*T_j] * \Phi(\text{offre}_{ij}) \\ + \beta_{ij} * [E_i * E_j]^{1,10} * [S_i * S_j]^4 * \Psi(\text{offre}_{ij})$$

avec

- . P_i, j : nombre d'habitants de chaque zone
- . R_i, j : PIB / tête de la zone
- . $T=0, 1$ ou 2 (i, j): coefficient d'attractivité touristique
- . Φ : fonction d'impédance des trajets « personnels », qui indique comment varie le nombre de trajets avec l'offre plurimodale existant entre les deux zones
- . Ψ : fonction d'impédance des trajets « personnels », qui indique comment varie le nombre de trajets avec l'offre plurimodale existant entre les deux zones

** avec Paris comme zone d'émission*

$$F_{ij} = 1,50 * [P_i * P_j]^{0,60} * [R_i * R_j]^{0,75} * [1+0,25*T_i] * [1+0,25*T_j] * \Phi(\text{offre}_{ij}) \\ + 2,1 * [E_i * E_j]^{1,10} * [S_i * S_j]^4 * \Psi(\text{offre}_{ij})$$

** sans Paris comme zone d'émission*

$$F_{ij} = [P_i * P_j]^{0,60} * [R_i * R_j]^{0,75} * [1+0,25*T_i] * [1+0,25*T_j] * \Phi(\text{offre}_{ij}) \\ + [E_i * E_j]^{1,10} * [S_i * S_j]^4 * \Psi(\text{offre}_{ij})$$

pour les trajets personnels

$$N^{mij} = k * \alpha_{ij} * [P_i * P_j]^{0,60} * [R_i * R_j]^{0,75} * [1+0,25*T_i] * [1+0,25*T_j] * \Phi(\text{offre}_{ij})$$

où α_{ij} est un coefficient qui vaut 1,50 si l'une des zones est Paris et 1,00 sinon.
48% relations de rapport $< 0,5$, 51% relations de rapport entre $0,5 < \dots < 10$,
1% de relations de rapport > 10 .

pour les trajets professionnels

$$N^{mij} = k * \beta_{ij} * [E_i * E_j]^{1,10} * [S_i * S_j]^4 * \Phi(\text{offre}_{ij})$$

où β_{ij} est un coefficient qui vaut 2,10 si l'une des zones est Paris et 1,00 sinon.
31% relations de rapport $< 0,4$, 62% relations de rapport entre $0,4 < \dots < 5$,
7% de relations de rapport > 5 .

Mesure de l'effet frontière

$$EF(\text{couple de pays}) = \sum_{ij} N^{mij} = \sum_{ij} N_{ij} \quad ij \in \text{couple de pays}$$

Valeurs pour 1988: France-Allemagne (18,48), France-Benelux (5,70), ... Moyenne France-pays étrangers (12,85).

Source:

* « Application du modèle MATISSE à l'étude multimodale des Schémas Directeurs pour le transport interrégional de voyageurs », Rapport technique, version provisoire, INRETS-SES, Mars 1997.

ALGORITHMES DE PARTITIONNEMENT D'UN ESPACE GEOGRAPHIQUE

On s'intéresse ici aux effets de la partition géographique sur les flux migratoires. On cherche à mettre en évidence le rôle joué par le découpage territorial sur les flux migratoires, ce qui permet de développer une analyse tout autant sur la notion d'appartenance que sur celle de distances inter-unités. Pour répondre à ces attentes, le LMC (IMAG) s'est penché sur les problèmes de partition d'un ensemble connexe. L'objectif est de développer un algorithme qui génère automatiquement tous les découpages, en k parties connexes, de la structure géographique étudiée.

Analyse du problème

Le problème de la connexité d'un ensemble à n éléments a été largement traité dans la littérature algorithmique \cite{bk:aho,bk:graham,bk:knuth}. Par contre il est beaucoup plus rare de trouver des ouvrages ou articles traitant de la génération exhaustive des partitions d'un ensemble.

Nous avons choisi de représenter la structure géographique étudiée par son graphe de contiguïté (Figure 1). De cette façon, nous pouvons appliquer toutes les théories et les algorithmes portant sur les graphes planaires \cite{bk:graham,bk:sakarovitch}. En nous basant sur un algorithme récursif de parcours en profondeur du graphe, nous avons développé notre propre générateur automatique de partitions à partir de:

- X un ensemble de sommets modélisant les unités territoriales.
- E un ensemble d'arêtes modélisant la contiguïté entre unités territoriales.

Ce qui permet de définir:

qu'une partition du territoire, représentée par les ensembles de sommets $P_1 \dots P_k$, est cohérente (du point de vue des géographes), si et seulement si les sous-graphes définis par ces ensembles sont connexes.

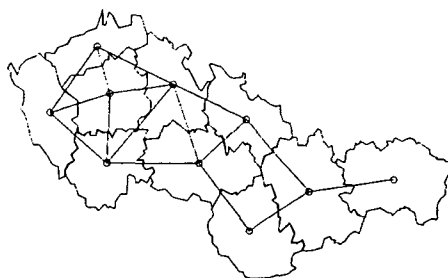


Figure 1: Graphe de contiguïté représentant la structure géographique étudiée

Cette définition permet de ne pas avoir à tester la totalité des partitions possibles (soit k^{n-1} pour une partition en k parties d'un graphe de dimension n) mais de ne retenir que celles dites "cohérentes".

Présentation de l'algorithme de génération des partitions

Principe de fonctionnement de l'algorithme

L'algorithme est basé sur le principe de parcours en profondeur récursif d'un graphe (graphe de contiguïté associé à l'ensemble X (exemple Figure 1), de type "Depth-First Search" \cite{bk:aho}).

Définitions:

- Comme l'algorithme est basé sur un parcours en profondeur d'un graphe G, il faudra comprendre par descente dans le graphe l'examen d'une branche de l'arbre de récursions.
- L'ensemble des voisins du noeud x_i V_{x_i} représente l'ensemble des noeuds contigus à x_i .
- L'ensemble des noeuds visités: *Visit*, représente l'ensemble des noeuds rencontrés lors du parcours du graphe.

Soit un graphe G et un noeud $\{y\}$ ($\{y\} \in G$). $\{y\}$ déconnecte le sous-graphe g ($g \subset G$) si il n'existe pas de sous graphe connexe g' ($g' \subset g$) tel que:

$$g' = g \setminus \{y\}$$

Le parcours du graphe initial G va permettre de générer des sous-graphes définissant les parties de la structure géographique étudiée. Ces sous-graphes sont représentés par des ensembles de noeuds rencontrés lors du parcours du graphe (X_k). Prenons l'exemple de la construction des partitions en deux parties connexes. L'algorithme va visiter tous les noeuds du graphe à partir d'un noeud d'initialisation $\{x_0\}$. Le passage du noeud $\{x_i\}$ au noeud $\{x_j\}$ est défini grâce aux ensembles V_{x_i} et *Visit*. Le premier ensemble permet de déterminer les noeuds candidats, le second garantit de ne pas visiter deux fois le même noeud. Les noeuds rencontrés seront placés dans X_1 lors de la descente dans le graphe, puis dans X_2 lors de la remontée. Pour qu'une partition soit générée, il faut que l'ensemble V_{x_i} soit vide, c'est-à-dire qu'il n'existe plus de noeud voisin non visité. Dans ce cas, X_1 définira la première partie, et son complémentaire la deuxième. La partition sera définitivement validée si les deux sous-graphes, représentés par X_1 et X_2 , sont connexes (Ex: Figure 2).

Le parcours du graphe initial nous assure de générer la totalité des partitions possibles. De plus, puisqu'un noeud ne peut faire parti des deux ensembles en même temps, et qu'il ne peut être visité qu'une seule fois par descente dans le graphe, nous sommes certains de ne pas construire deux fois la même partition.

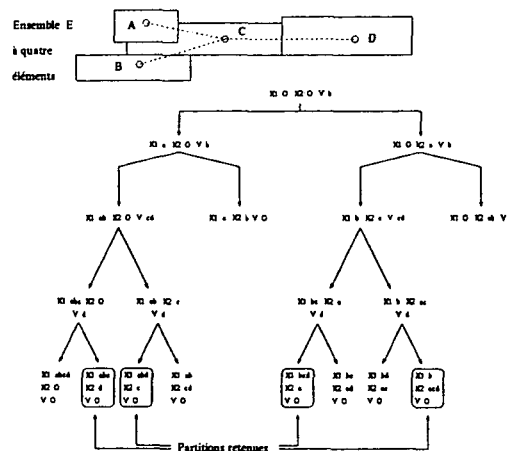


Figure 2 : Arbre d'exécution du deuxième algorithme sur un ensemble E à quatre éléments

Pour un graphe fortement connexe, la complexité de cette algorithmme est exponentielle (fonction de la dimension de la structure géographique étudiée). Néanmoins, puisque la construction des partitions repose sur le parcours en profondeur du graphe de contiguïté, nous pouvons supposer que le nombre de partitions construites par l'algorithme sera fonction de la dimension du graphe, et du degré moyen de ces noeuds.

Construction récursive optimale

Le principe de l'algorithme est globalement équivalent au précédent. Si l'on reprend l'exemple précédent, la différence s'établit sur le mode de construction des deux parties.

L'initialisation s'effectue toujours grâce à un noeud de $\{x_0\}$ pris au hasard. Mais, lorsque l'on considère un noeud $\{x_j\}$ appartenant à V_{x_i} , on cherche à savoir s'il déconnecte ou non la deuxième partie X_2 (la première partie X_1 est constituée de noeuds voisins les uns des autres, elle est donc connexe par construction). Si x_j ne déconnecte pas X_2 , on cherche à l'englober dans un sous-graphe connexe de E . Si un tel sous-graphe existe, X_1 sera redéfini par son complémentaire. Une fois ces étapes réalisées, la descente dans le graphe se poursuit avec un noeud de V_{x_j} . Si par contre x_j déconnecte X_2 , la descente dans le graphe sera arrêtée et $\{x_j\}$ sera placé dans X_2 . L'algorithme reprendra la construction des partitions avec un autre élément de V_{x_i} .

De cette manière, dès qu'un noeud déconnecte X_2 , tout le parcours du graphe qui s'y rapporte est supprimé (Figure 3). De même, lorsque l'on a identifié le sous-graphe connexe incluant X_2 , nous pouvons ré-initialiser X_1 par son complémentaire: $\text{Compl}X_2$, puisque les noeuds de $\text{Compl}X_2$ ne pourront pas faire partie de X_2 (Figure 4).

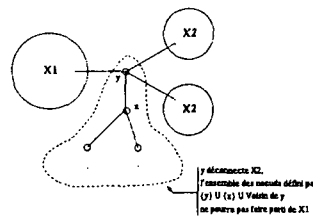


Figure 3 : Branche de parcours du graphe coupée par l'algorithme.

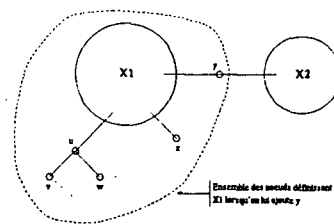


Figure 4 : Construction de X_1 .

L'exemple représenté par la Figure 5 illustre les deux situations types rencontrées lors de l'application de cet algorithme. E est équivalent à $X_1 \cup X_2 \cup \{y'\} \cup \{y''\}$. L'algorithme va traiter successivement $\{y'\}$ puis $\{y''\}$. Le premier noeud ne déconnecte pas X_2 dans la mesure où il existe un sous-graphe connexe de E , $X_2 \cup \{y''\}$, qui inclut X_2 . $\{y'\}$ sera donc ajouté à X_1 afin de former X_1' . Par contre le deuxième noeud déconnectera X_2 puisqu'il n'existe pas de sous-graphe connexe de E qui inclut X_1' ne sera donc pas construit par l'algorithme. $\{y''\}$ sera ajouté à X_2 , et une nouvelle descente dans le graphe sera effectuée à partir de X_1' et $X_2 \cup \{y''\}$.

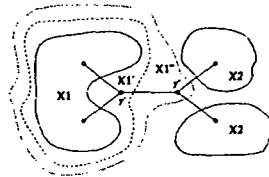


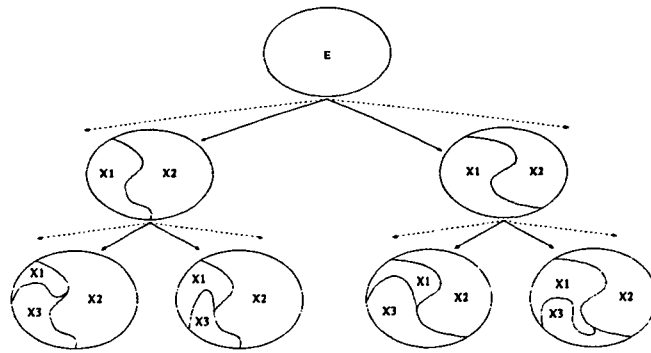
Figure 5 : Représentation du déroulement du troisième algorithme.

L'optimisation effectuée permet la suppression des branches de la récursion n'aboutissant pas à deux sous-graphes connexes. L'algorithme génère uniquement des partitions cohérentes au sens des géographes. Il possède donc une complexité équivalente au nombre de partitions générées.

Généralisation: partition en k parties connexes

La généralisation à la construction de k ($k > 2$) parties connexes s'effectue en subdivisant les sous-parties déjà formées. Ainsi, la génération des partitions en trois parties connexes s'effectue en appliquant l'algorithme une première fois sur l'ensemble E , puis en le réitérant une deuxième fois sur l'une des parties. Pour les partitions quatre parties, l'algorithme sera ré-appliqué sur les deux parties précédemment calculées, etc....

Partition en trois parties connexes



Partition en quatre parties connexes

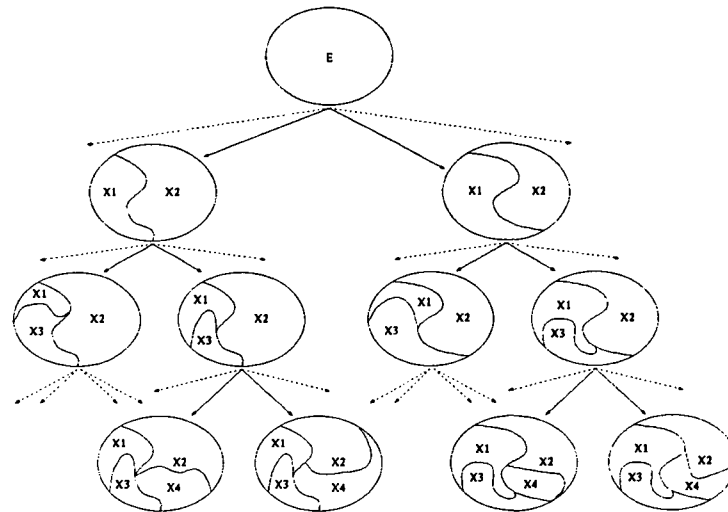


Figure 6 : Construction de k ($k > 2$) parties connexes.

Le tableau ci-dessous représente le nombre de partitions en 2, 3 ou 4 parties générées par le troisième algorithme. Les résultats fournis sont tirés de l'étude de l'ex-Tchécoslovaquie subdivisé en 10 régions ou 109 districts.

Nombre de partitions construites

10 unités	2	3	4
109 unités	48	557	760

ANNEXE

Algorithme optimal de construction récursive des partitions en deux parties connexes:

ConstRecPartOpt (X₁, X₂, Voisin, Visité) {

/ Condition de fin de récursion */*

SI (Voisin = \emptyset) **ALORS** {

SI (Connexe(X₂))

X₁ et X₂ sont connexes, fin.

/ car X₁ est connexe par construction */*

SINON

la partition n'est pas retenue, fin.

} FIN SI

/ Construction des deux parties X₁ et X₂ */*

SINON {

{e} = PremierElément(Voisin)

/ Ajout de l'élément {e} à l'ensemble X₁ */*

Voisin = Voisin - {e}

/ Enregistrement de l'état du système */*

DupliqueEtatSys(X₁, NouvX₁, X₂, NouvX₂, Voisin, NouvVoisin, Visité,

NouvVisité)

/ Ajout de l'élément {e} à l'ensemble NouvX₁ */*

NouvX₁ = NouvX₁ \cup {e}

NouvX₂)

/ SI (il existe une sous-partie connexe du complémentaire de NouvX₁ contenant*

inclus == 1 (cette sous-partie est définie par X₃)

*SINON inclus = 0 */*

inclus = EnsembleConnexe(NouvX₁, NouvX₂, X₃)

SI (inclus == 1) **ALORS** {

Complémentaire(X₃, ComplX₃)

/ NouvX₁ et NouvVisité sont équivalent au complémentaire de X₃ */*

NouvX₁ = ComplX₃

NouvVisité = ComplX₃

NouvX₁ */

/ Calcul des nouveaux ensembles de noeuds visités et de noeuds voisins de*

NouveauVoisin(NouvX₁, NouvX₂, NouvVoisin, NouvVisité);

/ Appel récursif */*

ConstRecPartOpt (X₁, X₂, NouvVoisin, NouvVisité)

} FIN SI

```

/* Retablissement de la valeur initiale des ensembles de noeuds */
DétruitSys( NouvX1, NouvX2, NouvVoisin, NouvVisité )
DupliqueEtatSys( X1, NouvX1, X2, NouvX2, Voisin, NouvVoisin, Visité,
NouvVisité )

/* Ajout de l'élément {e} à l'ensemble X2 */
X2 = X2 ∪ {e}

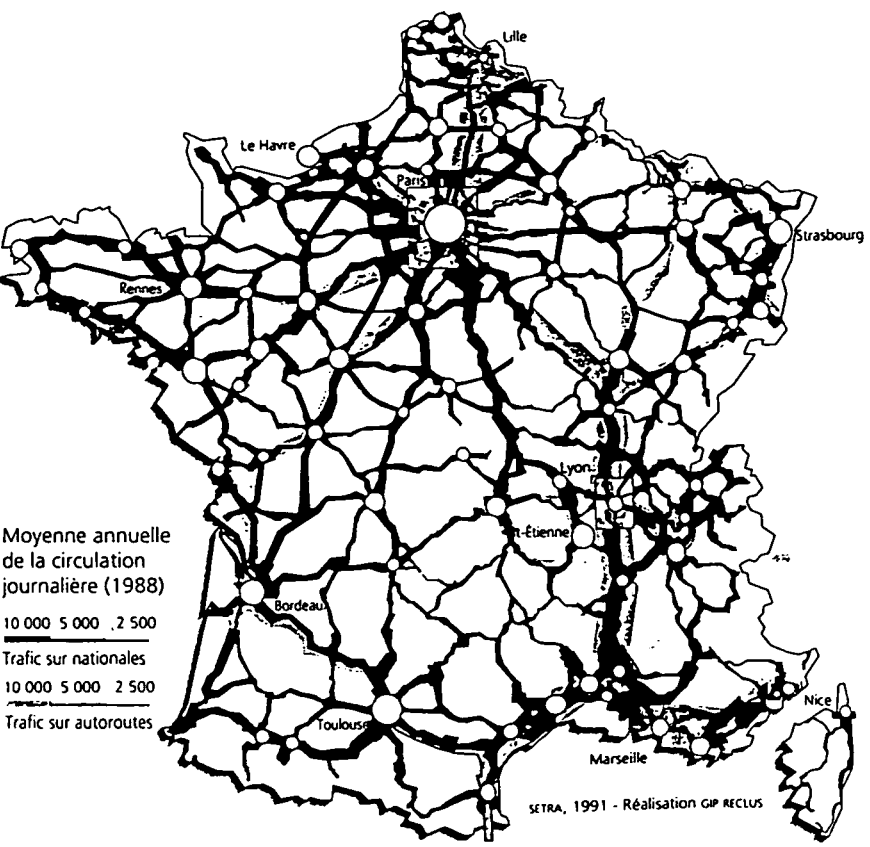
/* Appel récursif */
ConstRecPartOpt ( X1, X2, NouvVoisin, NouvVisité )

/* Destructions des ensembles créés */
DétruitSys(NouvX1, NouvX2, NouvVoisin, NouvVisité )

/* Suppression de l'élément {e} de l'ensemble X2 */
X2 = X2 \ {e}

} FIN SINON
} FIN ConstRecPartOpt

```



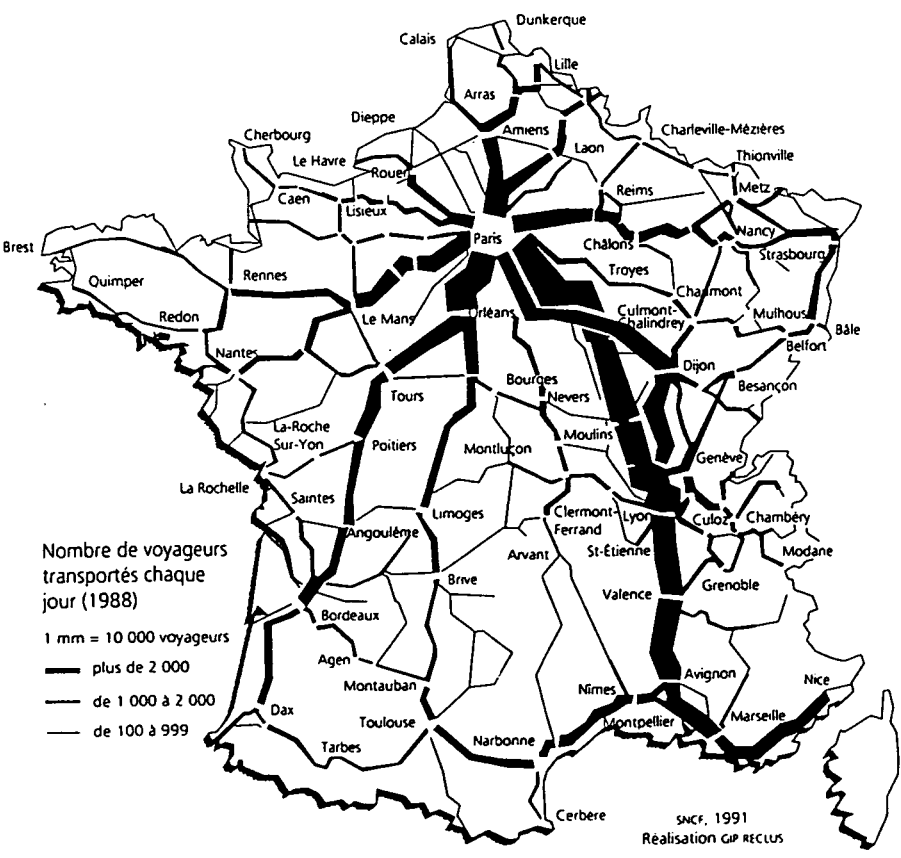
La circulation routière

L'axe majeur est Lille-Provence, mais ici l'Ouest apparaît chargé, en raison de ses densités et de ses activités, et de l'obligation où il se trouve de compenser les déficiences de la desserte par fer et par eau.

Moyenne annuelle de la circulation journalière (1988)
 10 000 5 000 2 500
 Trafic sur nationales
 10 000 5 000 2 500
 Trafic sur autoroutes

SETRA, 1991 - Réalisation GIP RECLUS

DATAR - RECLUS - LA DOCUMENTATION FRANÇAISE © ATLAS DE FRANCE 1993



Nombre de voyageurs transportés chaque jour (1988)
 1 mm = 10 000 voyageurs
 — plus de 2 000
 — de 1 000 à 2 000
 — de 100 à 999

SNCF, 1991
 Réalisation GIP RECLUS

Les points nodaux

Une classification selon une dizaine de critères de nodalité met en évidence les potentialités de Lille, de Lyon et du grand Bas Rhône, en sus de l'agglomération parisienne. La densité des points nodaux est plus forte dans la moitié est que dans la moitié ouest, plus éloignée des grands trafics de la mégapole européenne.

