

**IMPLANTATION ET EVALUATION D'UNE PLATE-FORME
D'ANALYSE QUANTITATIVE
DE FLUX DE DEPLACEMENTS VOYAGEURS
A LONGUE DISTANCE EN FRANCE**

François-Gaël OTTOGALLI
Rapport général
Avril 1997

Sous la direction de:
Jean Marc VINCENT
Institut IMAG - Laboratoire de Modélisation et Calcul
Equipe Parallélisme

Claude GRASLAND
Equipe PARIS - Université Paris I

Christian CALZADA
MELTT/SES/DEE

**CDAT
11791**

Table des matières

Introduction	7
① Problème et données	11
I Définition du problème	13
II Données statistiques	15
② Méthodologie	17
I Configuration des modèles	21
I.1 Description des modèles	21
I.1.1 Introduction et définition	21
I.1.2 Les modèles mathématiques de flux migratoires	22
I.1.3 La notion d'appartenance	27
I.2 Les mesures de la qualité des modèles	28
I.3 Evaluation des paramètres du modèle	30
II Algorithmes de partitionnement d'un espace géographique	33

II.1	Analyse du problème	33
II.2	Présentation de l'algorithme de génération des partitions	34
II.2.1	Principe de fonctionnement de l'algorithme	34
II.2.2	Construction récursive optimale	35
II.2.3	Généralisation: partition en k parties connexes	38
III	Classification	41
III.1	Exhaustive sur l'ensemble des partitions	41
III.2	Sélection sur un ensemble de partitions	43
IV	Forme des résultats	45
IV.1	Fichiers de résultats	45
III	Résultats	47
I	Sélection de la matrice de distance	51
II	Définition des erreurs de référence	53
III	Etude complète de la distance D_0	55
IV	Etude des paramètres de la fonction de la distance	93
IV.1	Paramètre de la distance: α	93
IV.2	Paramètre de découpage: γ	97
V	Influence de Paris	105
V.1	Paris neutre	105

TABLE DES MATIÈRES

5

V.2 Sans Paris 143

IV Conclusion 181

A Algorithmes 185

B Documentation technique des programmes réalisés 189

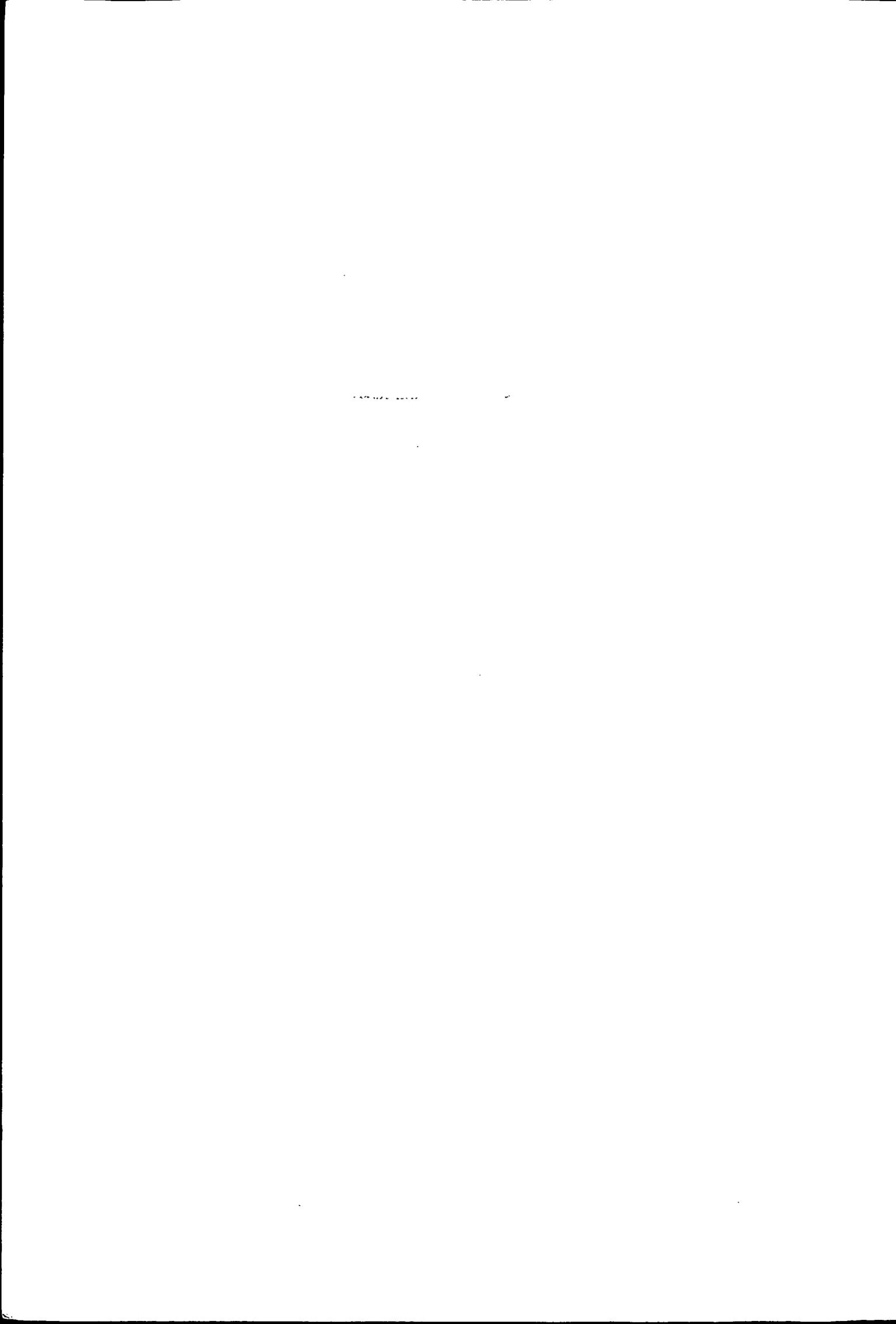
B.1 Présentation du travail réalisé 189

B.2 Documentation technique 190

Index 215

partie

Introduction



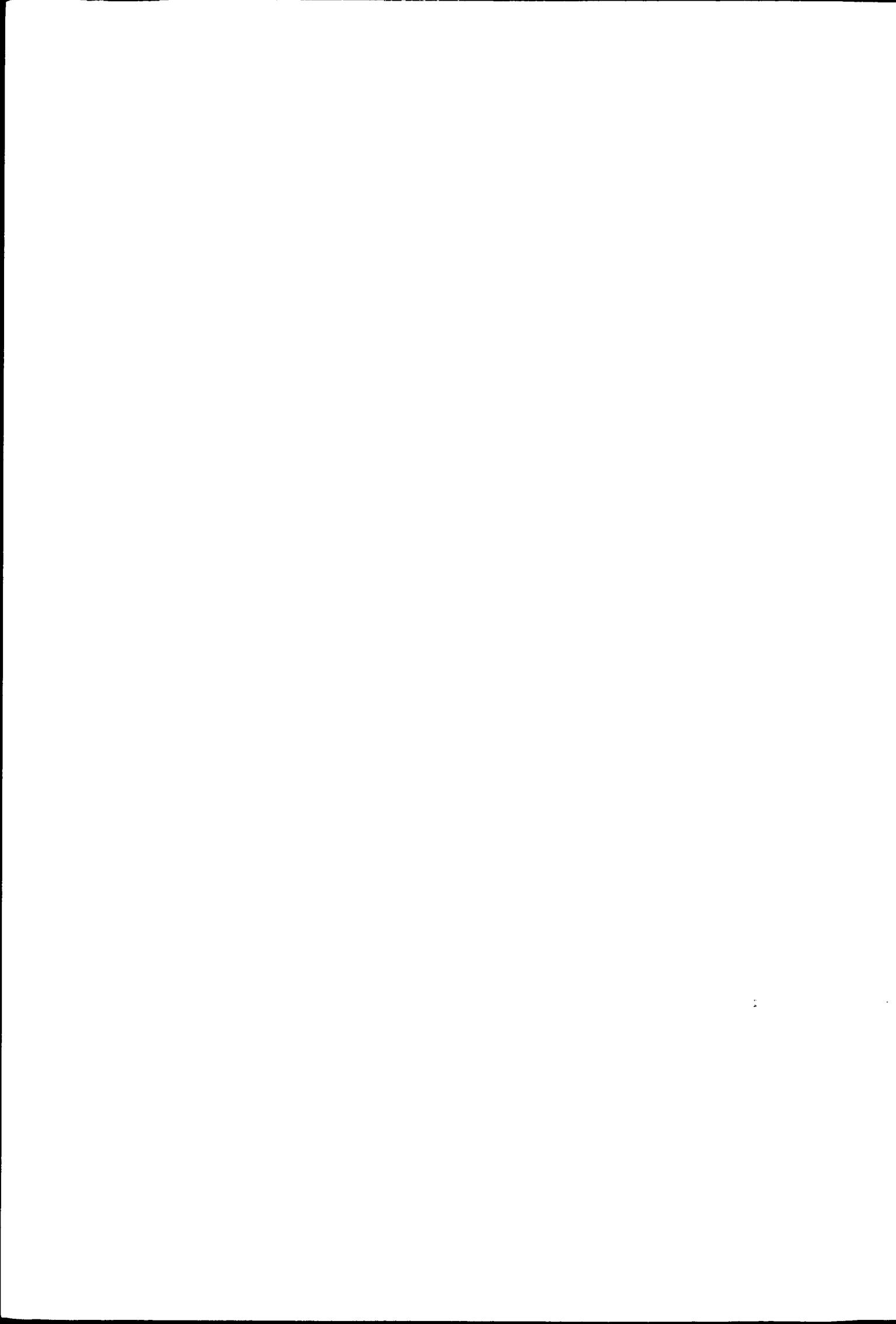
A la suite du projet réalisé par le Laboratoire de Modélisation et Calcul (L.M.C.) de Grenoble, équipe parallélisme, portant sur les problèmes de flux migratoires [3], le Département des Etudes Economiques du ministère des transports a commandé une application des algorithmes développés sur des données de flux de déplacements inter régionaux Français. Il s'agit donc de reprendre les techniques utilisées pour le cas de l'ex-Tchécoslovaquie, et de les adapter aux contraintes des nouvelles données.

Le document qui vous est présenté s'inspire très largement du rapport de stage, réalisé dans le cadre de la troisième année d'IUP "Mathématiques Appliquées et Industrielles", cité en référence [3].

Dans la suite de ce rapport, une première partie sera consacrée à la définition du problème et des données qui s'y rapportent. La deuxième partie, quant à elle, traitera de la présentation des modèles et de l'algorithme de découpage d'un espace géographique. Ce qui permettra d'aborder la classification et la forme des résultats. La dernière partie sera consacrée à la présentation des résultats obtenus lors du traitement de flux de déplacements Français.

Première partie

Problème et données



Chapitre I

Définition du problème

Le problème mis à jour porte sur l'analyse des flux de déplacements voyageurs à longue distance entre unités territoriales. Pour mener à bien cette l'analyse de ces données, les géographes font appel aux techniques de modélisation d'interactions spatiales. Ils donnent une expression mathématique aux phénomènes observés en fonction d'hypothèses formulées a priori. Par exemple, une simple observation des déplacements piétonniers en fonction de la distance met à jour le caractère de frein que joue la longueur du déplacement. Outre l'emploi de la distance, une nouvelle approche consiste à introduire la notion d'appartenance dans les modèles employés. L'intérêt de cette « réduction mathématique » est de pouvoir mesurer l'écart entre les flux observés et les flux modélisés. Cette différence permet de définir l'erreur de modélisation et donc le niveau d'adéquation des hypothèses par rapport aux observations.

Concrètement cette étude visait donc un double objectif:

* l'estimation d'un modèle à doubles contraintes pour les déplacements inter-régionaux de voyageurs à longue distance par mode 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.

Chapitre II

Données statistiques

Les matrices de flux utilisées sont issues de plusieurs sources:

* *Source: Modèle MATISSE*

Matrices multimodales de trafics en voyageurs pour le trafic français domestique en 1992, les zones origine-détermination 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: Enquête Transports et Communications 1993-94*

Cette enquête 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: Matrices 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.

COMPARAISON 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 précédents 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 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 1ère version, VP-SETRA 2ème version, VP-MATISSE.

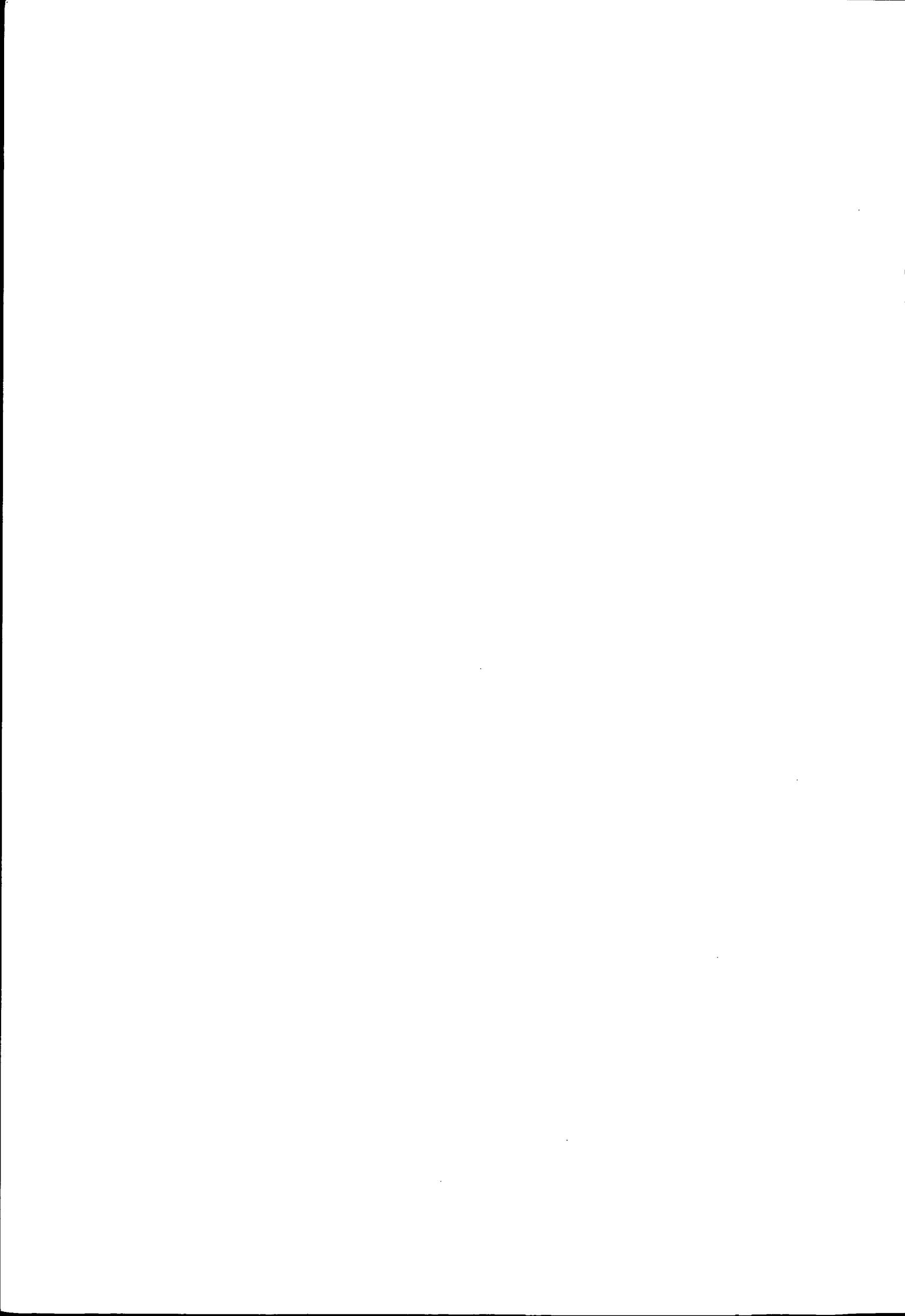
ZONAGES

. A 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.

Deuxième partie

Méthodologie



La démarche méthodologique qui a permis la réalisation de la plate-forme expérimentale se décompose en plusieurs parties. Tout d'abord, il a fallu choisir, dans les modèles de géographie quantitative, ceux dont la forme convenait aux contraintes posées. A la suite de quoi, il a fallu déterminer une méthode d'estimation des modèles. Cette étape a débouché sur la réalisation d'un algorithme de génération des sous-parties du territoire, afin de valider l'hypothèse de partage de la France. Ce qui a permis d'effectuer une classification des découpages possibles.

Chapitre I

Configuration des modèles

I.1 Description des modèles

I.1.1 Introduction et définition

Il peut paraître vain de vouloir prédire le comportement migratoire des hommes puisqu'il est à priori imprévisible au niveau individuel. Malgré tout, il est improbable qu'un Cherbourgeois décide d'aller acheter son melon à Cavaillon le dimanche matin. Sans entrer dans de telles évidences, cet exemple permet de comprendre l'importance de la distance dans les déplacements. Importance qui varie selon le motif du déplacement. De même, il est possible de caractériser un potentiel d'attraction intrinsèque aux lieux de destination. Le potentiel d'attraction du Sud-Est de la France sera plus important que celui du Nord-Est pour quiconque cherche la mer et le soleil. Ces quelques exemples mettent en évidence un ensemble de critères, tels que la distance, l'ensoleillement, la religion... permettant de caractériser certains types de flux migratoires. Leur analyse permet de formaliser les observations effectuées, et ainsi de définir des modèles mathématiques. A la simple vue de l'évolution des flux migratoires en fonction de la distance, il apparaît clairement que celle-ci jouera un rôle de frein en fonction du déplacement réalisé.

Mais, outre le critère de la distance, les géographes, tout particulièrement ceux de l'équipe P.A.R.I.S., s'intéressent à la notion d'appartenance à un territoire. Cette notion d'appartenance s'applique autant au concept de nationalité qu'à celui de confrérie, ou tout autre caractère permettant de regrouper des individus. Il est donc possible de rassembler les habitants d'un pays selon leur profil politique, professionnel, religieux..., dans le but de modéliser leur comportement en fonction du caractère étudié.

Il s'agit ici de comprendre et d'analyser les interactions spatiales observées. Nous pouvons définir le concept d'interaction spatiale comme le mouvement ou la communication à travers un espace qui résulte d'un processus de décision [2]. Ceci regroupe tout aussi bien les trajets professionnels, que les départs en vacances ou les communications téléphoniques. Il s'agit d'un

large concept qui pourra tout aussi bien s'appliquer aux flux migratoires qu'à l'étude de réseaux (routier, informatique, de télécommunications...), pourvu que l'on sache définir mathématiquement ce que l'on cherche à étudier. Cette définition commence par celle des unités territoriales, et des flux migratoires.

Une unité territoriale se définit comme une zone géographique possédant un caractère singulier (limite administrative, zone d'influence industrielle ou religieuse...). Ce même caractère peut se retrouver dans différentes unités, mais jamais dans deux unités contiguës. Cette définition donne la possibilité de "partitionner" un espace géographique en sous parties, et ainsi d'entrer dans une observation plus fine des phénomènes rencontrés en son sein. Lorsqu'il s'agit d'un mouvement d'informations, de populations, de marchandises... entre une unité géographique d'origine i et une unité géographique de destination j , nous définirons ce mouvement comme un flux migratoire inter-unité. Les statistiques issues de l'observation de ces flux permettront de construire des matrices numériques (F), qui serviront de support au paramétrage des modèles employés.

I.1.2 Les modèles mathématiques de flux migratoires

L'approche mathématique de la géographie quantitative donne la possibilité de formaliser les notions étudiées. Elle permet de définir des modèles dont l'objectif est d'appuyer ou d'infirmier les hypothèses énoncées. Lors de l'étude des flux migratoires, ces derniers sont l'écho de déplacements d'individus, d'informations, de marchandises... Les paramètres de modélisation de ces déplacements peuvent être:

- la distance inter-unité (euclidienne, de plus court chemin...).
- l'infrastructure présente permettant le déplacement.
- l'opportunité de ce déplacement.
- le coût financier associé...

Les objectifs et les contraintes (définies ultérieurement) de la géographie quantitative permettent de réduire le nombre de paramètres intervenant dans la modélisation des déplacements. Les modèles d'interactions spatiales [6] se décrivent par l'équation (cf I.1):

$$F^* = f(O, D, I) \quad (I.1)$$

avec, pour une structure géographique de dimension n ,

- F^* une matrice de dimension $[n, n]$ décrivant les mouvements entre les n origines et les n destinations.

- f une fonction des paramètres d'entrée.
- D un vecteur de dimension $[n]$ décrivant le pouvoir d'émission des lieux d'origine (somme des départs).
- O un vecteur de dimension $[n]$ décrivant le pouvoir d'attraction des lieux de destination (somme des arrivées).
- I une matrice de dimension $[n, n]$ décrivant le degré de séparation des lieux d'origine et de destination. La distance exprimée par cette ou ces matrices peut être euclidienne, de plus court chemin ou tout autre selon l'objet d'étude des géographes (pour modéliser la rentabilité d'une ligne de transport, on pourra par exemple, associer à chaque trajet son coût financier en fonction de l'amortissement du véhicule, du salaire du chauffeur, de la cargaison transportée...).

Dans le cadre de notre étude, seules les fonctions de type multiplicatif sont abordées, (restriction instruite par l'équipe P.A.R.I.S). Il existe néanmoins une deuxième forme d'équation de type additif [7] (ex: le modèle de Tobler I.2):

$$F^* = (A_i + B_j) / f(I_{ij}) \quad (I.2)$$

Dans un même esprit de restriction, le nombre d'unités d'origine est égale au nombre d'unités de destination. La théorie développée dans ce chapitre, et les équations mathématiques qui s'y rapportent, se généralisent au problème de dimension $n \times m$ (n origines et m destinations).

Dans la suite de ce rapport F représentera la matrice de flux observés, et F^* la matrice de flux estimés. Ces matrices sont de dimension $n \times n$ (n : dimension de l'espace de départ et d'arrivée).

Les modèles mathématiques présentés se distinguent par les contraintes qu'ils mettent en cause.

Les modélisations des contraintes de flux

- Contrainte d'origine:

Ce modèle permet d'étudier la force de rétention d'une zone par le biais d'un coefficient proportionnel au nombre de départs.

$$F_{ij}^* = b_j \cdot D_j \cdot f(I_{ij}) \quad (I.3)$$

avec

$b_j = \frac{1}{\sum_i f(I_{ij})}$: contrainte de la conservation du total des départs.

D_j : somme des départs de i.
 $f(I_{ij})$: fonction de la distance entre i et j.

– *Contrainte de destination:*

Il permet d'étudier la force d'attraction d'une zone par le biais d'un coefficient proportionnel au nombre d'arrivées.

$$F_{ij}^* = a_i \cdot O_i \cdot f(I_{ij}) \quad (I.4)$$

avec

$a_i = \frac{1}{\sum_j f(I_{ij})}$: contrainte de la conservation du total des arrivées.

O_i : somme des arrivées de j.

$f(I_{ij})$: fonction de la distance entre i et j.

– *Contrainte de flux total:*

Il permet de combiner le pouvoir de rétention et d'attraction des unités géographiques par le biais d'un coefficient proportionnel au nombre de départs et d'arrivées.

$$F_{ij}^* = k \cdot O_i \cdot D_j \cdot f(I_{ij}) \quad (I.5)$$

avec

$k = \frac{\sum_{ij} F_{ij}}{\sum_{ij} O_i \cdot D_j \cdot f(I_{ij})}$: contrainte de la conservation du total des départs et des arrivées.

O_i : somme des arrivées en j.

D_j : somme des départs de i.

$f(I_{ij})$: fonction de la distance entre i et j.

– *Double contraintes:*

Tout comme le modèle de contrainte de flux total, il permet de combiner le pouvoir de rétention et d'attraction des unités géographiques, mais dans un objectif plus de prévision que de caractérisation.

$$F_{ij}^* = a_i \cdot O_i \cdot b_j \cdot D_j \cdot f(I_{ij}) \quad (I.6)$$

avec

$a_i = \frac{1}{\sum_j b_j \cdot D_j \cdot f(I_{ij})}$: contrainte de la conservation des départs.

O_i : somme des arrivées en j.

$b_j = \frac{1}{\sum_i a_i \cdot O_i \cdot f(I_{ij})}$: contrainte de la conservation des arrivées.

D_j : somme des départs de i.

$f(I_{ij})$: fonction de la distance entre i et j.

Tous les modèles employés font intervenir une fonction de la distance. Trois d'entre elles vont être présentées.

Les fonctions de la distance

– *La fonction identité:*

L'emploi de cette fonction rend le modèle insensible à la distance. Ce qui correspond à une répartition uniforme, sur toute la structure géographique étudiée, des flux migratoires observés.

$$f(I_{ij}) = 1 \quad (I.7)$$

avec

I_{ij} : distance entre i et j .

– *Fonction gravitaire:*

La formule de la gravitation universelle de Newton a servi de base au développement des premiers modèles de géographie quantitative (Zipf K.J., 1946; Steward J.Q., 1947). Par analogie avec la théorie de la gravitation, le modèle gravitaire a été construit en associant à chaque unité spatiale un potentiel d'attraction de la même forme que celui attribué au corps céleste. L'usage de cette fonction vise à établir une distribution théorique dépendante de la distance inter-unité. Elle permet de rendre compte de près de quatre vingt pour-cent de la variation des flux migratoires entre unités géographique (cf [5]);

$$f(I_{ij}) = \frac{1}{I_{ij}^2} \quad (I.8)$$

avec

I_{ij} : distance entre i et j .

– *Généralisation de la fonction gravitaire: fonction de Paréto ou de puissance:*

Au cours de l'étude des modèles, basés sur une fonction de la distance de type gravitaire, il est apparu que de meilleurs résultats, en terme de qualité du modèle, peuvent être obtenus en modifiant la valeur de l'exposant associé à la distance. Ainsi, le rôle de l'exposant peut s'assimiler à celui d'un frein dont la rigidité varie selon la valeur qui lui est donnée [8]. Les modèles de Paréto (Zipf K.J., 1946; Steward J.Q., 1947) [11] se sont développés sous deux formes:

– puissance :

$$f(I_{ij}) = I_{ij}^{-\alpha_p} \quad (I.9)$$

– exponentiel :

$$f(I_{ij}) = e^{-\alpha_e \cdot I_{ij}} \quad (I.10)$$

avec

I_{ij} : distance entre i et j .

α : paramètre du modèle.

Les études empiriques sur les migrations de population montrent que, d'une manière générale, le modèle exponentiel décrit mieux le comportement des migrants sur courte distance, tandis que le modèle puissance est plus approprié pour les longues distances. Il est à remarquer que l'on peut passer de l'emploi de l'une ou l'autre des deux fonctions par une simple transformation (exponentielle ou logarithmique) de la matrice des distances, les deux fonctions sont donc mathématiquement similaires.

Les fonctions présentées font intervenir la notion de distance. Cette distance peut se trouver sous de multiples formes.

Les distances employées

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

$$DEucl_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (I.11)$$

(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, la plus riche, ou le chef lieu administratif...

Bien que la distance euclidienne puisse paraître tout à fait adaptée dans notre cas, il est possible d'employer n'importe quel type de distance une fois celle-ci clairement définie. Par exemple, les distances ne sont pas forcément exprimées avec l'unité conventionnelle du mètre, mais peuvent être représentatives du coût du transport associé au déplacement, ou du nombre de cultures ethniques à traverser.

Pour les besoins de l'étude, trois nouvelles distances ont été construites. Elle permettent de prendre en compte la population incluse dans les zones d'emploi. Cette subdivision des unités géographiques en zones d'emploi 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'emploi.

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

$$D_{0,i,j} = \sum_{k \in i} \sum_{l \in j} \left(\frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]} \right) \quad (I.12)$$

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'emploi $k(k \in i)$ et $l(l \in j)$. La distance $D_{50,k,l}$ est défini par:

$$D_{50,i,j} = \sum_{k \in i} \sum_{l \in j} \left(\left(\frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]} \right) * \delta_{[DEucl_{k,l} \geq 50]} \right) \quad (I.13)$$

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 zone d'emploi $k(k \in i)$ et $l(l \in j)$. La distance $D_{80,k,l}$ est défini par:

$$D_{80,i,j} = \sum_{k \in i} \sum_{l \in j} \left(\left(\frac{DEucl_{k,l} * pop_i[k] * pop_j[l]}{\sum_{s \in i} pop_i[s] * \sum_{t \in j} pop_j[t]} \right) * \delta_{[DEucl_{k,l} \geq 80]} \right) \quad (I.14)$$

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

Pendant, il convient de ne pas se limiter à la notion de distance dans l'étude des flux migratoires, pour cela un nouveau terme multiplicatif a été ajouté aux modèles permettant de prendre en compte la notion d'appartenance à un territoire.

I.1.3 La notion d'appartenance

Comme il a été écrit dans l'introduction de cette partie, les géographes de l'équipe P.A.R.I.S. s'intéressent à la notion d'unité sociale (appartenance à une même région, à un même territoire...)[8].

Cette notion permet de réaliser une partition de la structure géographique étudiée en regroupant les individus suivant un caractère commun. Elle a été introduite dans les modèles multiplicatifs par le terme $\gamma^{P_{ij}}$, transformant la formulation générale de ces derniers en :

$$g(I_{ij}) = f(I_{ij}) \cdot \gamma^{P_{ij}} \quad (\text{I.15})$$

avec

$f(I_{ij})$: fonction de la distance définie ci-dessus (cf I.7, I.8, I.9, I.10).

γ : paramètre du modèle à estimer.

P_{ij} : matrice de dimension $[n,n]$ définissant la notion d'appartenance, dont les éléments valent 0 si i et j appartiennent au même territoire, 1 dans les autres cas.

Remarque: l'emploi de modèles multiplicatifs est une restriction que nous pourrions lever ultérieurement en adaptant la notion d'appartenance aux autres types de modèles (ex: on ajoute $\gamma * P_{ij}$ au modèles de type additif).

Ainsi, dans le cadre de notre étude, nous définirons la notion de partition géographique comme le regroupement d'unités territoriales répondant au même critère (appartenance à une même région, à un même système financier...). Elle sera représentée par la matrice de binaire P , dont les éléments P_{ij} valent 0 si les unités i et j appartiennent au même territoire, ou 1 dans les autres cas.

Cette nouvelle formulation permet de se pencher sur les problèmes d'interactions spatiales et territoriales. L'étude des modèles employés devient fonction de la partition testée. En comparant les résultats obtenus entre une partition initiale de référence et des partitions fictives définies par l'expérimentateur, il est possible de déterminer la partition qui fournira la matrice de flux estimés la plus en adéquation avec celle de flux observés.

Afin de quantifier la pertinence du modèle utilisé, une erreur de modèle est calculée traduisant l'écart entre la matrice de flux observés et la matrice de flux estimés.

I.2 Les mesures de la qualité des modèles

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

1. la valeur absolue:

$$Err(F, F^*) = \sum_{ij} |F_{ij} - F_{ij}^*| \quad (\text{I.16})$$

2. le carré de l'erreur:

$$Err(F, F^*) = \sum_{ij} (F_{ij} - F_{ij}^*)^2 \quad (I.17)$$

3. le χ^2_{F, F^*} :

$$Err(F, F^*) = \sum_{ij} \frac{(F_{ij} - F_{ij}^*)^2}{F_{ij}^*} \quad (I.18)$$

La troisième mesure (à savoir celle du χ^2 I.18) est plus coûteuse en calcul que la deuxième (cf I.17), elle-même plus coûteuse que la première (cf I.16). Il faut donc choisir une mesure permettant une interprétation claire des résultats. Les deux premières mesures seront utilisées dans le cadre de données homogènes, alors que la troisième 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 le χ^2 réalisera une normalisation des données avant de les traiter. Cette normalisation sera utile dans le cas des unités géographiques très éloignées dont les valeurs de flux seront sans doute faibles. Les résultats ainsi obtenus serviront à définir une fonction coût dont la minimisation permettra l'ajustement des paramètres du modèle.

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

$$\overline{\chi^2_F} = \sum_{ij} \frac{(F_{ij} - \overline{F})^2}{\overline{F}} \quad (I.19)$$

$$\chi^2_{F, F^*} = \sum_{ij} \frac{(F_{ij} - F_{ij}^*)^2}{(F_{ij}^*)} \quad (I.20)$$

$$Err = \frac{\chi^2_{F, F^*}}{\overline{\chi^2_F}} \quad (I.21)$$

\overline{F} : moyenne des valeurs de la matrice des flux observés F

$$\overline{F} = \frac{\sum_{ij} F_{ij}}{n \cdot (n-1)}$$

où n est la dimension de la matrice carrée F .

Nous voyons ici que le χ^2 calculé est normalisé par un " χ^2 moyen " ($\overline{\chi^2_F}$, cf I.19). Ce " χ^2 moyen " nous permet d'apprécier la dispersion des flux observés par rapport à leur 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 la distance. Pour se faire, nous calculons une erreur de référence pour

chaque matrice en utilisant le modèle à double contraintes avec la fonction Identité: $\chi_{F, F_{Id}^*}^2$ (cf I.22) . Cette valeur nous servira à normaliser les erreurs calculées avec d'autres fonctions de la distance. La formulation de l'erreur du modèle (ε) devient:

$$\chi_{F, F_{Id}^*}^2 = \sum_i \frac{\sum_j (F_{ij} - F_{Id,ij}^*)^2}{F_{Id,ij}^*} \quad (I.22)$$

$$\varepsilon = \frac{\chi_{F, F^*}^2}{\chi_{F, F_{Id}^*}^2} \quad (I.23)$$

La définition de cette erreur nous permet de construire une fonction coût représentative de chaque option de modélisation. Cette fonction servira à déterminer les paramètres optimaux des modèles.

I.3 Evaluation des paramètres du modèle

L'évaluation des paramètres est effectuée par une fonction d'optimisation. Celle qui est actuellement disponible sur la plate-forme est nommée "Hooke". Elle a été implantée par Mark G. Johnson (12/02/94), d'après la méthode de "recherche directe" développée par Hooke et Jeeves¹ [9].

Le principe de l'algorithme mis en place repose sur la "recherche directe" d'une solution à des problèmes numériques ou statistiques. La méthode de "recherche directe" 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 meilleur 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:

$$x_1 = x_0 \pm \delta_0 f(x_1) < f(x_0) \quad (I.24)$$

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

1. Les sources de cette fonction ont été trouvées à l'adresse électronique: <http://www.netlib.org/opt/hooke.c>

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'optimal trouvé peut être local ou global, nous n'avons pas de moyen de 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 coût) le problème de la dérivabilité ne se pose pas. Néanmoins, la formulation du modèle de Wilson (I.5) est intrinsèquement itérative, ce qui rend l'expression de la dérivée difficilement exploitable pour la détermination des paramètres optimaux. Il faudra donc envisager l'utilisation d'une nouvelle fonction d'optimisation parallélisable pour gagner en efficacité!!!! proposition!!!!

Les tests de la plate-forme n'ont pu être menés à terme faute de réponse de la fonction Hooke. Il est donc indispensable de mener plus loin la recherche de fonctions d'optimisation plus souples d'utilisation. Une bibliothèque de fonctions d'optimisation doit être constituée afin de répondre à toutes les situations.

Chapitre II

Algorithmes de partitionnement d'un espace géographique

L'équipe P.A.R.I.S., s'intéresse aux effets de la partition géographique sur les flux migratoires. Elle cherche à mettre en évidence le rôle joué par le découpage territorial sur les flux migratoires, ce qui lui permet de développer son analyse tout autant sur la notion d'appartenance que sur celle de distances inter-unités. Pour répondre à ses attentes, nous avons dû nous pencher 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.

II.1 Analyse du problème

Le problème de la connexité d'un ensemble à n éléments a été largement traité dans la littérature algorithmique [1, 4, 10]. 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 II.1). De cette façon, nous pouvons appliquer toutes les théories et les algorithmes portant sur les graphes plannaires [4, 12]. 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.

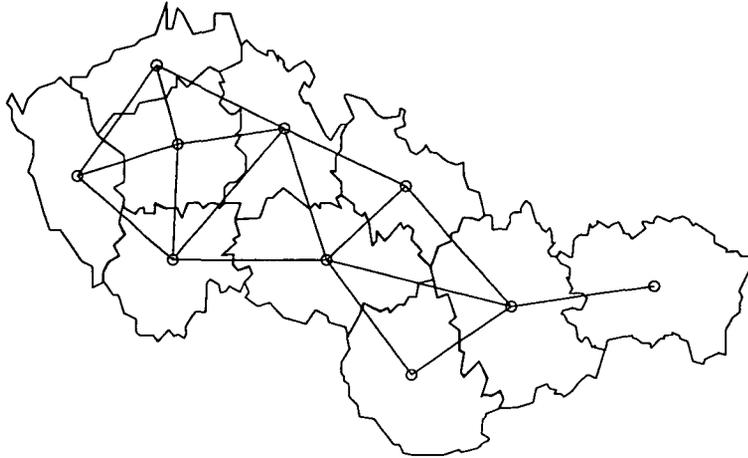


FIG. II.1 – Graphe de contiguïté représentant la structure géographique étudiée

Cette définition permet de ne pas avoir à tester la totalité des partitions possibles (soit k^{n-1} pour une partition en k parties), mais de ne retenir que celles dites "cohérentes" lors de leur génération par les différents algorithmes.

II.2 Présentation de l'algorithme de génération des partitions

II.2.1 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 II.1)), de type "Depth-First Search" [1].

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 récursion.
- L'ensemble des voisins du nœud x_i : Vx_i , représente l'ensemble des nœuds contigus à x_i .
- L'ensemble des nœuds visités: $Visit$, représente l'ensemble des nœuds rencontrés lors du parcours du graphe.

- Soit un graphe G et un nœud 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:

$$g' = g - y$$

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 nœuds 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 nœuds du graphe à partir d'un nœud d'initialisation x_0 . Le passage du nœud x_i au nœud x_j est défini grâce aux ensembles Vx_i et $Visit$. Le premier ensemble permet de déterminer les nœuds candidats, le second garantit de ne pas visiter deux fois le même nœud. Les nœuds 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 Vx_i soit vide, c'est-à-dire qu'il n'existe plus de nœud 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 II.2).

Le parcours du graphe initial nous assure de générer la totalité des partitions possibles. De plus, puisqu'un nœud 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.

Pour un graphe fortement connexe, la complexité de cette algorithme reste exponentielle par rapport à 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 nœuds.

II.2.2 Construction récursive optimale

Le principe de l'algorithme (Annexe A, p 185) est globalement équivalent au précédent. Si l'on reprend l'exemple précédent, la différence s'établit sur le mode de construction des deux parties.

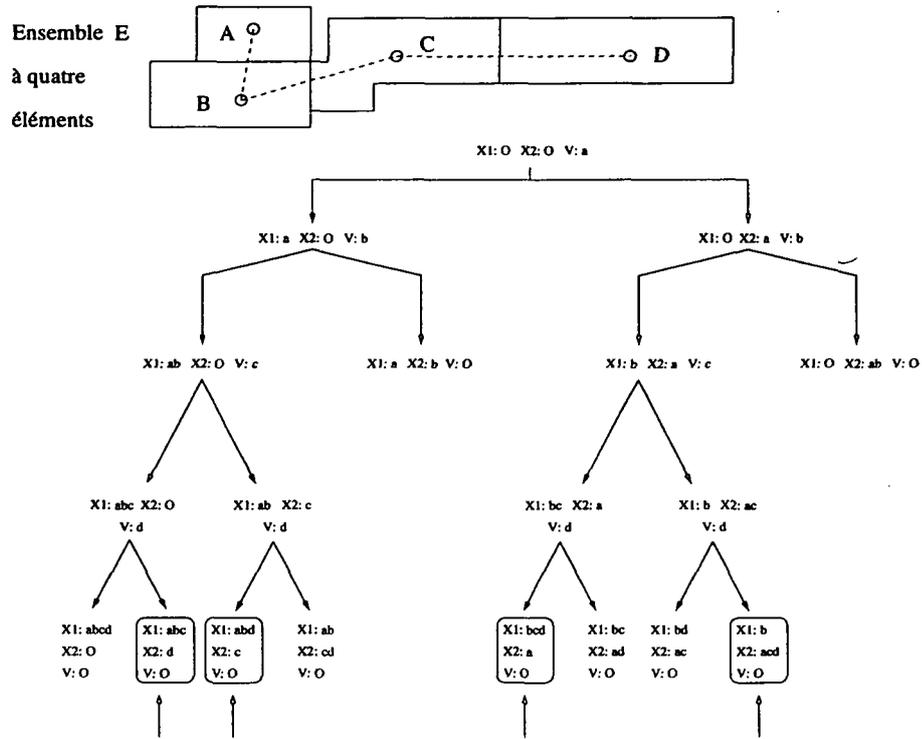


FIG. II.2 – Arbre d'exécution du deuxième algorithme sur un ensemble E à quatre éléments

L'initialisation s'effectue toujours grâce à un nœud de x_0 pris au hasard. Mais, lorsque l'on considère un nœud x_j appartenant à Vx_i , on cherche à savoir s'il déconnecte ou non la deuxième partie X_2 (la première partie X_1 est constituée de nœuds 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 nœud de Vx_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 Vx_i .

De cette manière, dès qu'un nœud déconnecte X_2 , tout le parcours du graphe qui s'y rapporte est supprimé (figure II.3). De même, lorsque l'on a identifié le sous-graphe connexe incluant X_2 , nous pouvons réinitialiser X_1 par son complémentaire: $Compl X_2$, puisque les nœuds de $Compl X_2$ ne pourront pas faire partie de X_2 (figure II.4).

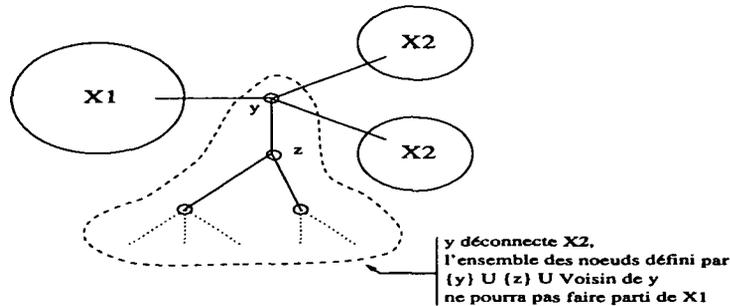


FIG. II.3 - Branche de parcours du graphe coupée par l'algorithme

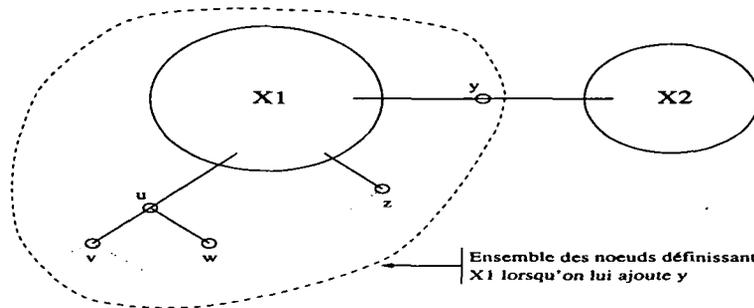


FIG. II.4 - Construction de X1

L'exemple représenté par la figure II.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 nœud 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 nœud déconnectera X_2 puisqu'il n'existe pas de sous-graphe connexe de E qui l'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 $X_2 \cup \{y''\}$.

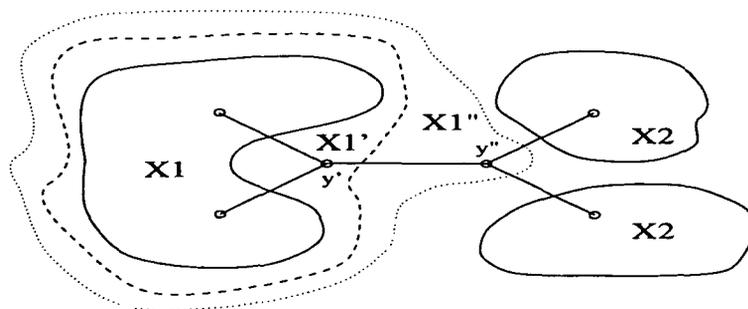


FIG. II.5 - Représentation du déroulement du troisième algorithme

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.

II.2.3 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 réappliqué sur les deux parties précédemment calculées, etc.... (figure II.6).

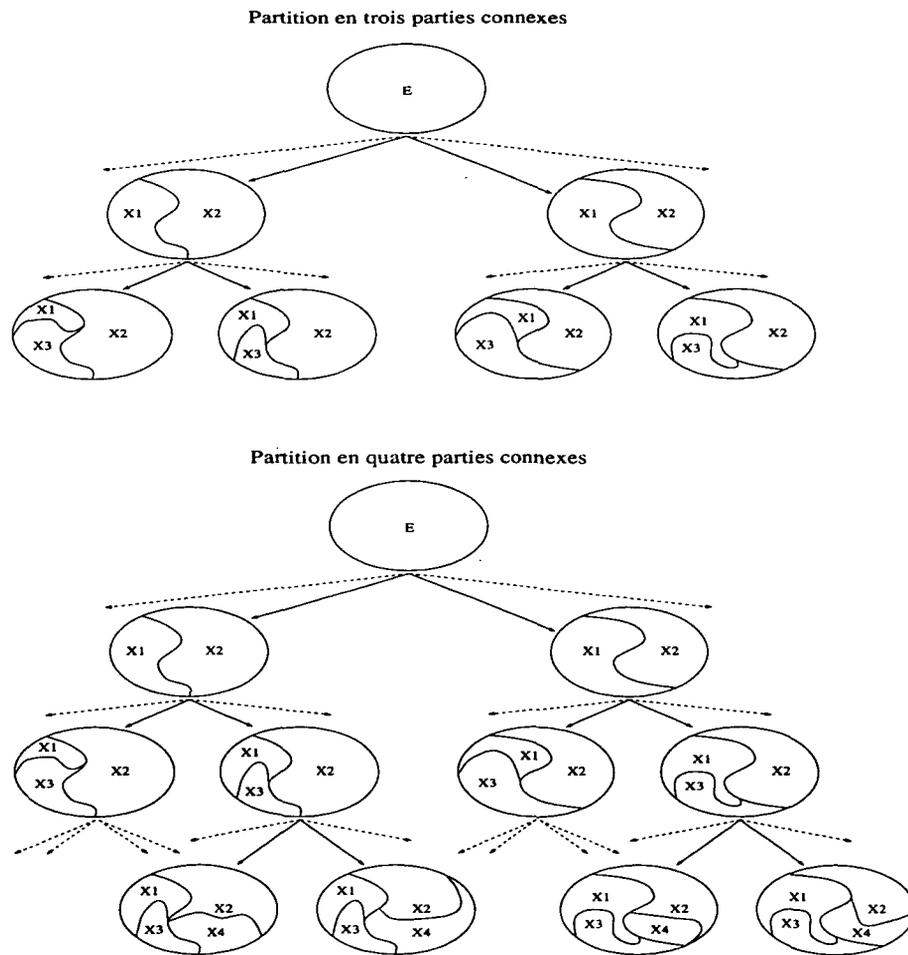


FIG. II.6 - Construction de k ($k > 2$) parties connexes

Le tableau II.1 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.

	2	3	4
10 unités	48	557	760
109 unités	$\simeq 10^5$

TAB. II.1 - *Nombre de partitions construites.*

Chapitre III

Classification

III.1 Exhaustive sur l'ensemble des partitions

Une étude exhaustive de l'ensemble des partitions permet de comparer toutes les configurations possibles, et ainsi déterminer la configuration optimale. En contre-partie cela demande un traitement lourd et coûteux des fichiers de résultats. En effet, selon la configuration considérée le nombre d'expérimentations peut varier exponentiellement. Pour fixer les idées, il faut remarquer qu'un graphe totalement connecté de taille 10 permet de construire $2^{10-1} = 512$ partitions un deux, $3^{10-1} = 19683$ partitions en trois, $4^{10-1} = 262144$ partitions en quatre... Dans le même esprit, il est intéressant de remarquer que les trois formes de découpage de la France, qui seront étudiées dans la partie III, ont permis de générer:

- France avec Paris: 2085 partitions.
- France sans Paris: 1116 partitions
- France avec Paris neutre: 30103 partitions.

De telles variations dans le nombre de cas à traiter devra conduire à un choix judicieux du découpage utilisé.

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

Le principe de cette méthode de classification oblige à prendre en compte l'ensemble des cas possibles sans distinctions. La complexité du traitement d'une structure géographique sera donc fonction de la taille de son "partitionnement". D'autre part, les mêmes calculs seront effectués pour chacune des partitions. 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 moyen¹ pour un modèle donné.

Modèle	Nombre de partitions en deux parties connexes			
	Référence 1	avec Paris 2085	sans Paris 1116	Paris neutre 3103
BARRIERE	2s	4170s = 1h 9mn 30s	2232 = 37mn 12s	6206 = 1h 43mn 26s
PARETO + BARRIERE	8s	16680s = 4h 38mn	8928s = 2h 28mn 48s	24824s = 6h 53mn 44s

TAB. III.1 - Temps moyen de calcul pour les modèles BARRIERE et PARETO + BARRIERE sur RS 6000-370

Le tableau III.1 permet de matérialiser le problème du temps de calcul nécessaire pour mener à bien une expérimentation (hors traitement des résultats). Il apparaît clairement qu'un nombre excessif de partitions entraîne des temps de calculs 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 partitions 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 $T_p = O(\frac{T_1}{p})$, 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 répartitions de charges afin d'éviter les déséquilibres entre processeurs. En effet, les calculs nécessaires à l'optimisation de la fonction coût varient selon la partition qui lui est associée. Il apparaît donc des situations où certains processeurs ont fini de traiter les partitions qui lui sont attribuées, alors que d'autres sont en cours de calcul. 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 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 parallélisation, certaines situations obligent à sélectionner les partitions à traiter lorsque celles-ci sont trop nombreuses.

1. Les temps peuvent varier d'un facteur deux selon la partition traitée, sur une même machine.

III.2 Sélection sur un ensemble de partitions

Les limites expérimentales évoquées précédemment amène à développer des stratégie d'étude propre aux structures géographiques trop denses (dont le nombre de partitions est supérieur à quelques milliers). Les deux principales stratégies envisagées sont:

1. une sélection au hasard des partitions à traiter.
2. une sélection des partitions selon les résultats obtenus sur un problème plus simple.

La première technique permet de débiter une étude lorsque l'on a aucune idée à priori sur le(s) découpage(s) optimal(aux) de la structure géographique étudiée. Le fait de réduire le nombre de partitions permet donc une approche du problème à résoudre, mais la part de "hasard" introduite augmente considérablement la probabilité de ne pas trouver les paramètres optimaux.

La deuxième approche se base sur une pré-étude de la structure géographique "simplifiée". Cette simplification a pour but de diminuer le nombre de partitions à considérer. A la suite de quoi, la sélection des partitions se fera à partir des découpage optimaux trouver à la première étape. Il sera possible, par exemple, de faire varier un jeux d'unités territoriales d'un côté puis de l'autre d'une frontière afin de visualiser leur influence sur les paramètres.

L'utilisation de l'une ou l'autre de ces méthodes pose le problème de l'interprétation des résultats. Dans le cas présent, il sera possible de parler d'optimal local mais la généralisation à l'optimal global ne pourra se faire que par l'étude systématique de l'ensemble des partitions, ou grace aux propriétés mathématiques des modèles (certaines partitions ne seront jamais opti- males).

Chapitre IV

Forme des résultats

IV.1 Fichiers de résultats

Les résultats des différentes expérimentations sont sauvegardés dans des fichiers de type texte. La présentation dépend de la fonction de la distance employée. Il existe néanmoins une partie commune qui comprend le calcul:

- de "Err" l'erreur du modèle définie par les géographes.
- du χ^2 .

A la suite de quoi s'ajoute:

- α pour les modèles fonction de la distance.
- γ et le vecteur définissant le découpage de la France, pour les modèles fonction de l'appartenance.
- α , γ et le vecteur définissant le découpage de la France, pour les modèles fonction de la distance et de l'appartenance.

De ces fichiers il est possible d'extraire les informations nécessaires pour tracer des courbes, ou bien visualiser les différents découpages de la France. Quelques exemples sont présentés dans la partie suivante.

Rg	Err	ϵ	Partition																	α			
0	0.342	0.965	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0.6669
1	0.343	0.968	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0.4537
2	0.345	0.972	0	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	0.6703

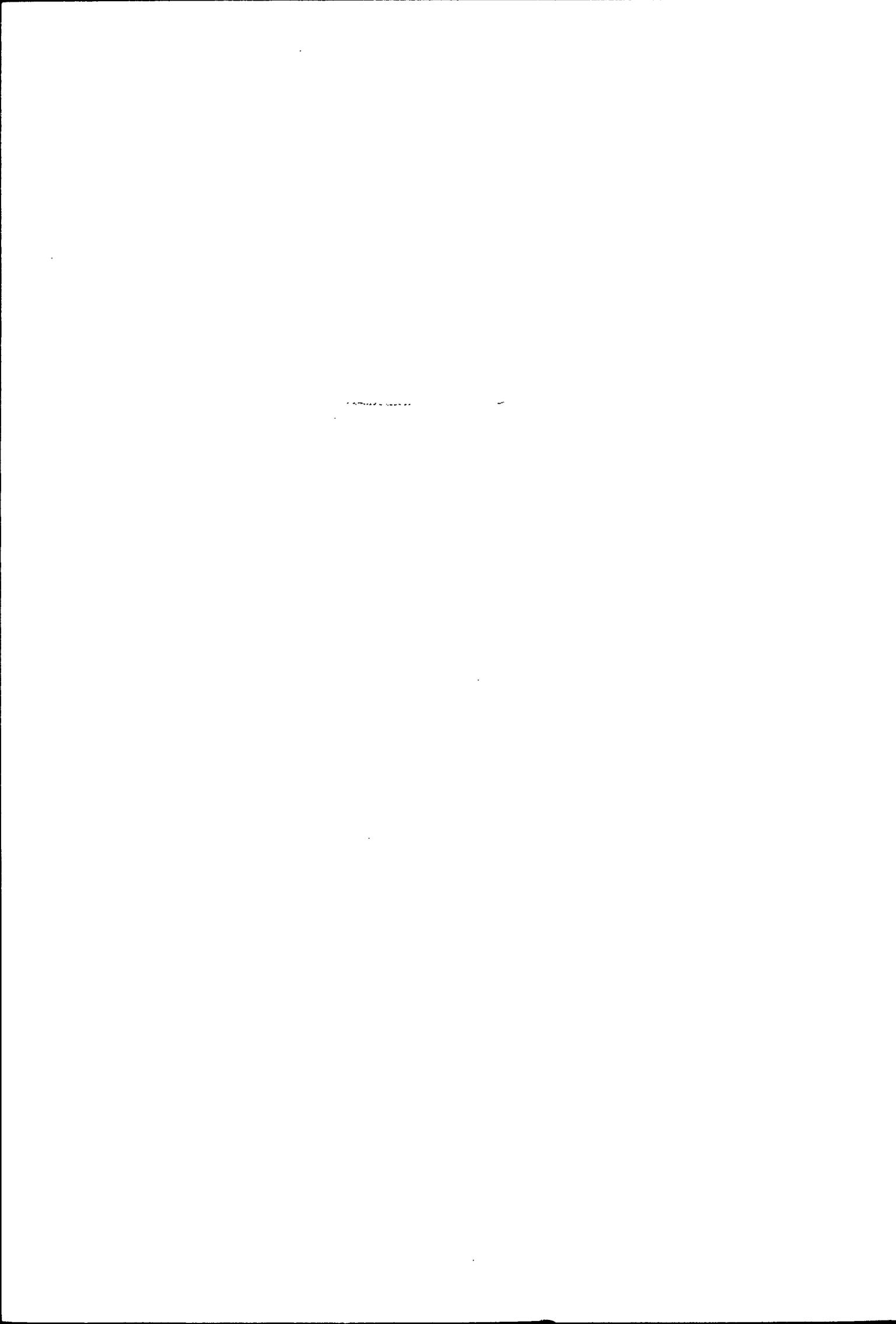
TAB. IV.1 - Exemple de fichier résultats pour le modèle Barrière

Rg	Err	ϵ	Partition																	α	γ			
0	0.017	0.049	0	1	0	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	-2.3254	0.6808
1	0.018	0.050	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	-2.3672	0.6951
2	0.018	0.051	0	1	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	-2.3599	0.7082

TAB. IV.2 - Exemple de fichier résultats pour le modèle Paréto + Barrière

Troisième partie

Résultats



La plate-forme expérimentale va être utilisée pour l'étude de flux de déplacements inter-régionaux français. Cette étude se décomposera en cinq parties. La première permettra de choisir parmi les différentes distances la plus appropriée pour la modélisation des flux observés. En second lieu, une erreur de référence sera calculée pour chaque matrice de flux. Ces valeurs serviront à normaliser les résultats obtenus par la suite. La troisième partie portera sur l'étude de l'influence des différentes fonctions de la distance sur la modélisation. Ce travail sera complété par l'analyse des paramètres des modèles. Pour finir, l'influence de la région Parisienne sur l'évolution du découpage de la France sera analysée.

Chapitre I

Sélection de la matrice de distance

Trois sources de données sont disponibles:

- les matrices de flux observés (neuf matrices).
- les fonctions de la distance (quatre fonctions sont disponibles).
- les distances inter-régionales (cinq distances ont été construites).

Une étude exhaustive porterait à combiner l'ensemble des paramètres, ce qui correspondrait à 180 configurations possibles. Toute fois, un travail préliminaire à montrer le rôle prépondérant de la distance dans l'explication des matrices de flux. Ces résultats permettent de limiter le nombre des expérimentations en fixant la distance utilisée. Une étude complète des distances disponibles a été réalisée grâce au modèle à double contraintes associé à la fonction de la distance de Paréto (cf!!!). Les résultats obtenus, présentés dans les tableaux I.1, I.2 et I.3, doivent permettre de choisir la distance qui sera retenue pour la suite des expérimentations.

Distance	M31		M32		M34	
	Err	α	Err	α	Err	α
DEucl	0.016302	-2.3565	0.019739	-1.3310	0.029877	-1.8009
D0	0.00801	-2.59304	0.01912	-1.40489	0.02794	-1.91002
D50	0.00795	-2.57755	0.01935	-1.39271	0.02856	-1.89309
D80	0.05869	-1.72812	0.03625	-1.07485	0.07090	-1.42237
D50+Prop	0.008084	-2.5964	0.018947	-1.4120	0.027785	-1.9128
D80+Prop	0.072378	-1.6940	0.041014	-1.1020	0.084343	-1.4295

TAB. I.1 - Matrice de flux M31, M32 et M34

Distance	MS15		MS16		MS19	
	Err	α	Err	α	Err	α
DEucl	0.048178	-2.0419	0.026490	-0.96827	0.046702	-1.8053
D0	0.05071	-2.18996	0.02468	-1.03702	0.04994	-1.92043
D50	0.05222	-2.17169	0.02476	-1.02869	0.05113	-1.90422
D80	0.16753	-1.52644	0.03013	-0.82746	0.13026	-1.39640
D50+Prop	0.050629	-2.1895	0.024623	-1.0471	0.049724	-1.9220
D80+Prop	0.197783	-1.5129	0.031561	-8.7094	0.151707	-1.3904

TAB. I.2 – Matrice de flux MS15, MS16 et MS19

Distance	VLS		VLS2		VLS3	
	Err	α	Err	α	Err	α
DEucl	0.025002	-2.8751	0.016302	-2.3565	0.031813	-2.1172
D0	0.01050	-3.18992	0.00801	-2.59304	0.025020	-2.2870
D50	0.01093	-3.16415	0.00795	-2.57755	0.025131	-2.2681
D80	0.09812	-2.04108	0.05869	-1.72812	0.077048	-1.4412
D50+Prop	0.010657	-3.1844	0.008084	-2.5964	0.024994	-2.2902
D80+Prop	0.123195	-1.9509	0.072378	-1.6940	0.059762	-1.6442

TAB. I.3 – Matrice de flux VLS, VLS et VLS3

L'analyse des erreurs présentées dans les tableaux ci-dessus permet de distinguer la distance D_0 comme étant la mieux adaptée pour la modélisation des flux observés. Dans la suite du rapport la distance D_0 sera systématiquement employée dans les expérimentations.

Chapitre II

Définition des erreurs de référence

Afin de pouvoir comparer les différents modèles de flux, une erreur, basée sur la distance du χ^2 (cf I.18), est prise comme référence. Cette erreur est calculée par le biais du modèle à double contraintes (cf!!!) sans l'influence de la distance. Les erreurs calculées pour chaque matrice de flux sont reportés dans le tableau II.1.

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

Matrice	χ^2	Err
m31	1224265	0.354720
m32	114458.546	0.107950
m34	1232204.875	0.288126
ms15	974475200	0.568950
ms16	62000376	0.055243
ms19	1019435968	0.422089
vls	3164687	0.444352
vls2	1333445.5	0.227548
vls3	1162357.625	0.193970

TAB. II.1 – Erreurs du modèle Identité

La définition de ces erreurs permettra de normaliser les résultats ultérieurs. Cette normalisation a pour but d'apprécier le gain apporté par la "complexification" de la fonction de la distance. Une étude complète de la distance choisie au chapitre I va montrer les possibilités offertes par cette méthode.

Chapitre III

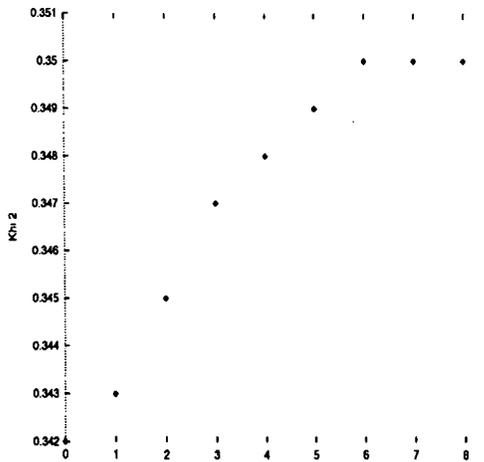
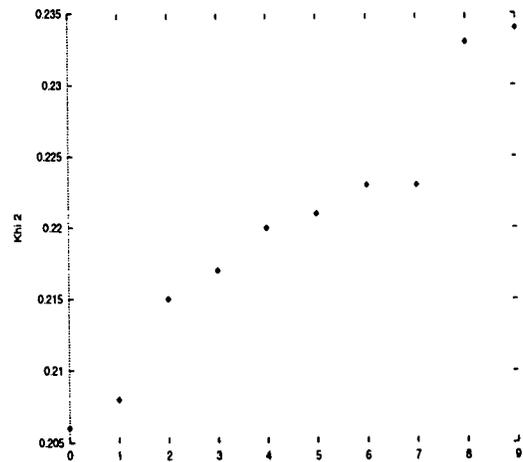
Etude complète de la distance D_0

La distance D_0 a été choisie au chapitre I comme étant la plus appropriée à la modélisation des flux observés. Une étude exhaustive de cette distance va être présentée. C'est à dire que la distance D_0 sera associée à toute les matrices de flux, et toutes les fonctions de la distance. L'ensemble des résultats sont présentés dans les tableaux et graphiques ci-dessous.

Matrice	ϵ	Err	α
m31	0.06	0.0219	-2.207
m32	0.18	0.0191	-1.405
m34	0.97	0.0279	-1.909
ms15	0.09	0.0507	-2.190
ms16	0.45	0.0246	-1.037
ms19	0.12	0.0499	-1.920
vls	0.02	0.0104	-3.190
vls2	0.04	0.0080	-2.593
vls3	0.13	0.0250	-2.287

TAB. III.1 - Erreurs du modèle Paréto sur les différentes matrices de flux (D_0)

M31 Rang	$\gamma < 1$			$\gamma > 1$		
	ϵ	Err	γ	ϵ	Err	γ
0	0.965	0.342	0.666	0.581	0.206	3.335
1	0.968	0.343	0.453	0.585	0.208	3.413
2	0.972	0.345	0.670	0.607	0.215	3.257
3	0.978	0.347	0.721	0.612	0.217	3.172
4	0.982	0.348	0.469	0.619	0.220	3.086
5	0.985	0.349	0.766	0.623	0.221	3.335
6	0.986	0.350	0.429	0.628	0.223	3.010
7	0.987	0.350	0.374	0.629	0.223	3.276
8	0.987	0.350	0.134	0.657	0.233	2.774
9	0.989	0.351	0.582	0.661	0.234	2.822

TAB. III.2 – *Ecart des dix meilleures partitions pour le modèle à effet de barrière (M31, D_0)*FIG. III.1 – “Err” ($\gamma < 1$, M31, D_0)FIG. III.2 – “Err” ($\gamma > 1$, M31, D_0)

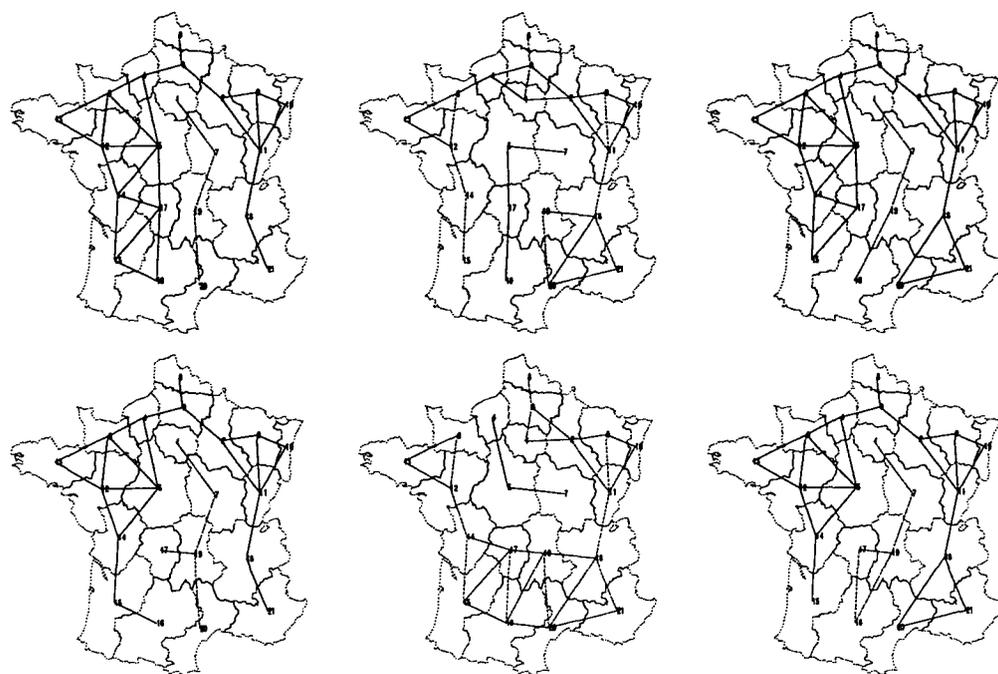


FIG. III.3 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux M31 (D_0)

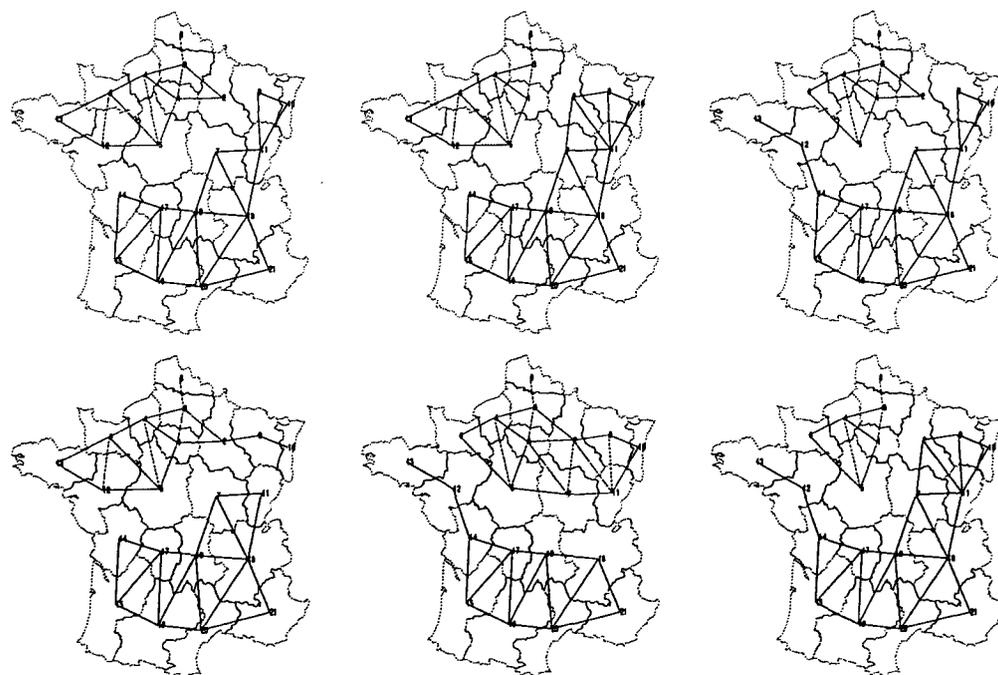
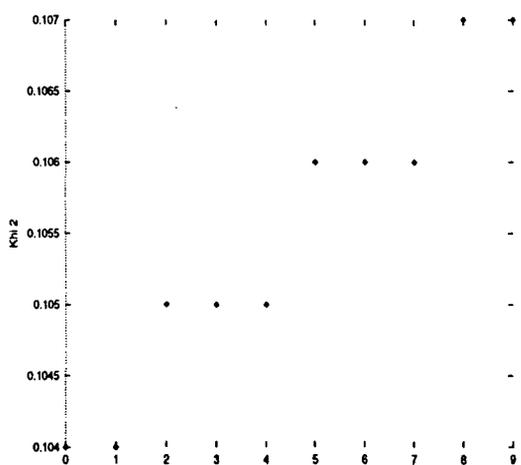
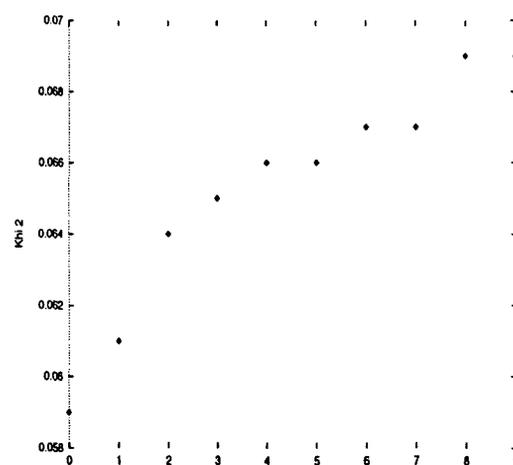


FIG. III.4 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux M31 (D_0)

M32	$\gamma < 1$			$\gamma > 1$		
Rang	ϵ	Err	γ	ϵ	Err	γ
0	0.965	0.342	0.666	0.581	0.206	3.335
1	0.968	0.343	0.453	0.585	0.208	3.413
2	0.972	0.345	0.670	0.607	0.215	3.257
3	0.978	0.347	0.721	0.612	0.217	3.172
4	0.982	0.348	0.469	0.619	0.220	3.086
5	0.985	0.349	0.766	0.623	0.221	3.335
6	0.986	0.350	0.429	0.628	0.223	3.010
7	0.987	0.350	0.374	0.629	0.223	3.276
8	0.987	0.350	0.134	0.657	0.233	2.774
9	0.989	0.351	0.582	0.661	0.234	2.822

TAB. III.3 – *Ecart des dix meilleures partitions pour le modèle à effet de barrière (M32, D_0)*FIG. III.5 – “Err” ($\gamma < 1$, M32, D_0)FIG. III.6 – “Err” ($\gamma > 1$, M32, D_0)

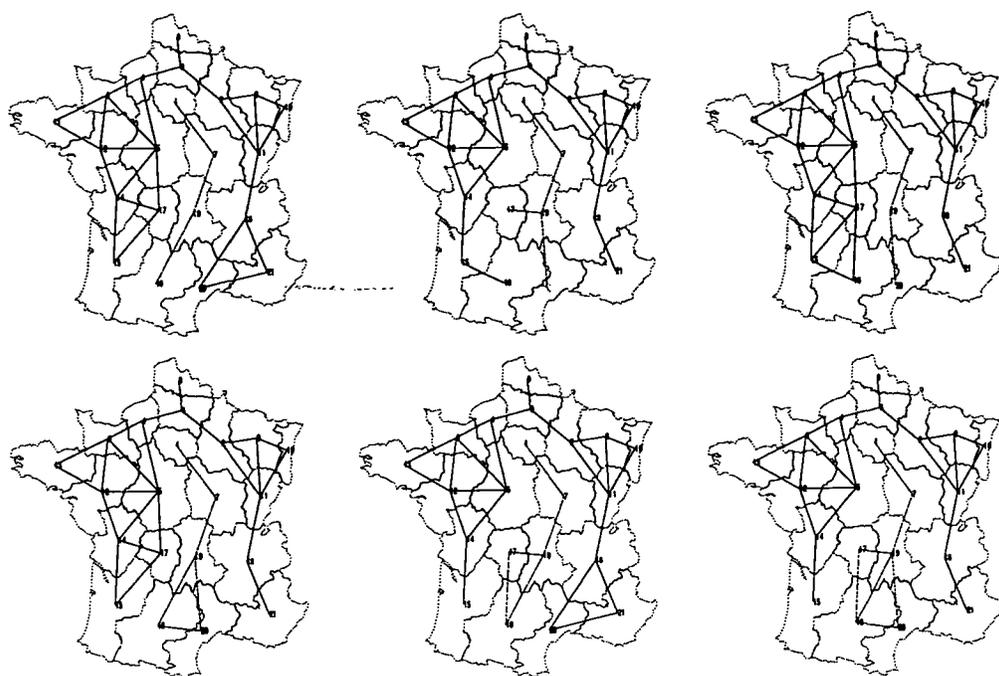


FIG. III.7 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux M32 (D_0)

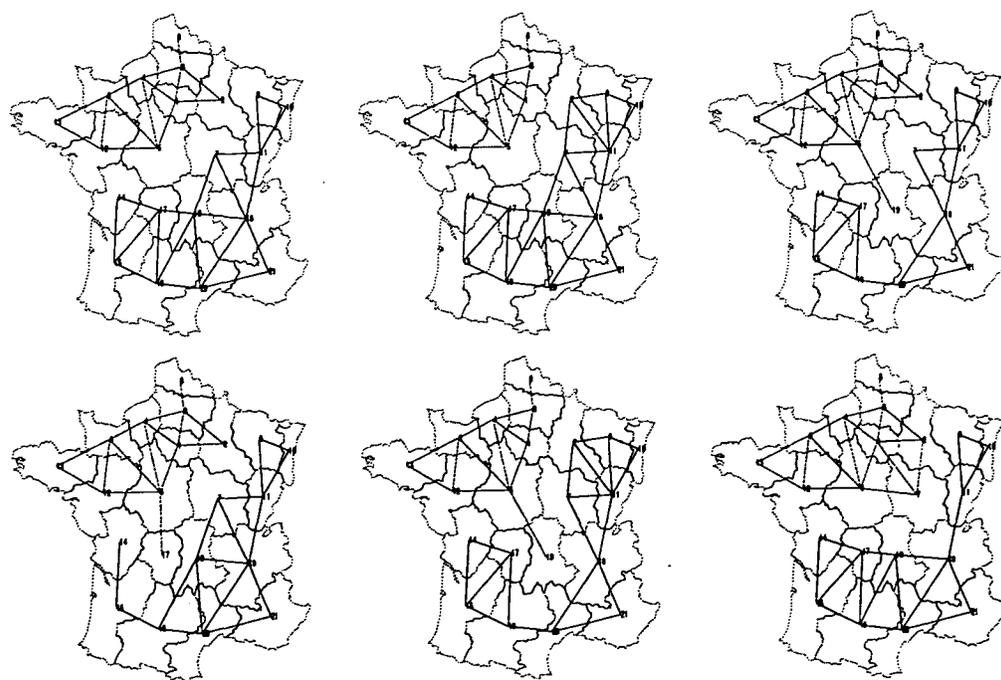
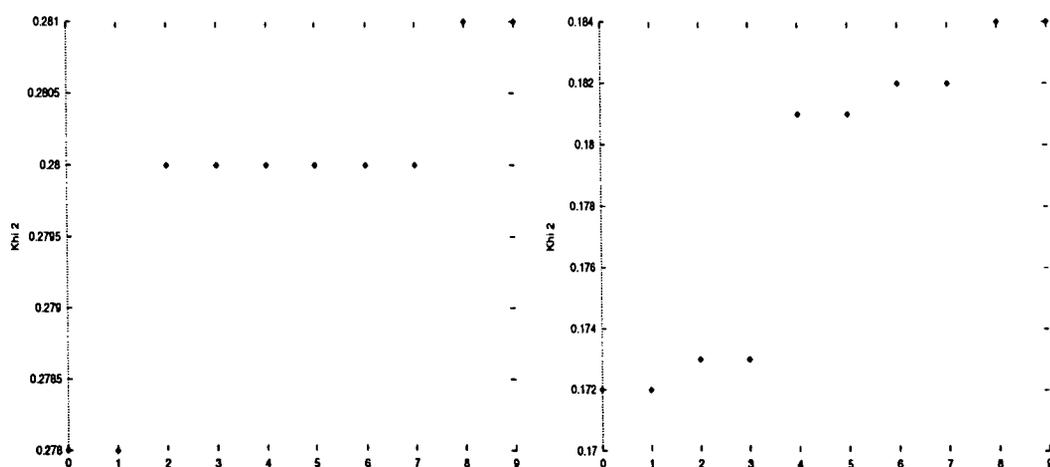


FIG. III.8 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux M32 (D_0)

M34	$\gamma < 1$			$\gamma > 1$		
Rang	ϵ	Err	γ	ϵ	Err	γ
0	0.966	0.278	0.656	0.597	0.172	2.968
1	0.966	0.278	0.656	0.597	0.172	2.968
2	0.970	0.280	0.675	0.599	0.173	3.035
3	0.970	0.280	0.675	0.599	0.173	3.035
4	0.971	0.280	0.417	0.628	0.181	3.082
5	0.971	0.280	0.417	0.628	0.181	3.082
6	0.972	0.280	0.375	0.632	0.182	2.829
7	0.972	0.280	0.375	0.632	0.182	2.829
8	0.976	0.281	0.732	0.639	0.184	2.909
9	0.976	0.281	0.732	0.639	0.184	2.909

TAB. III.4 - *Ecart des dix meilleures partitions pour le modèle à effet de barrière (M34, D_0)*FIG. III.9 - "Err" ($\gamma < 1$, M34, D_0)FIG. III.10 - "Err" ($\gamma > 1$, M34, D_0)

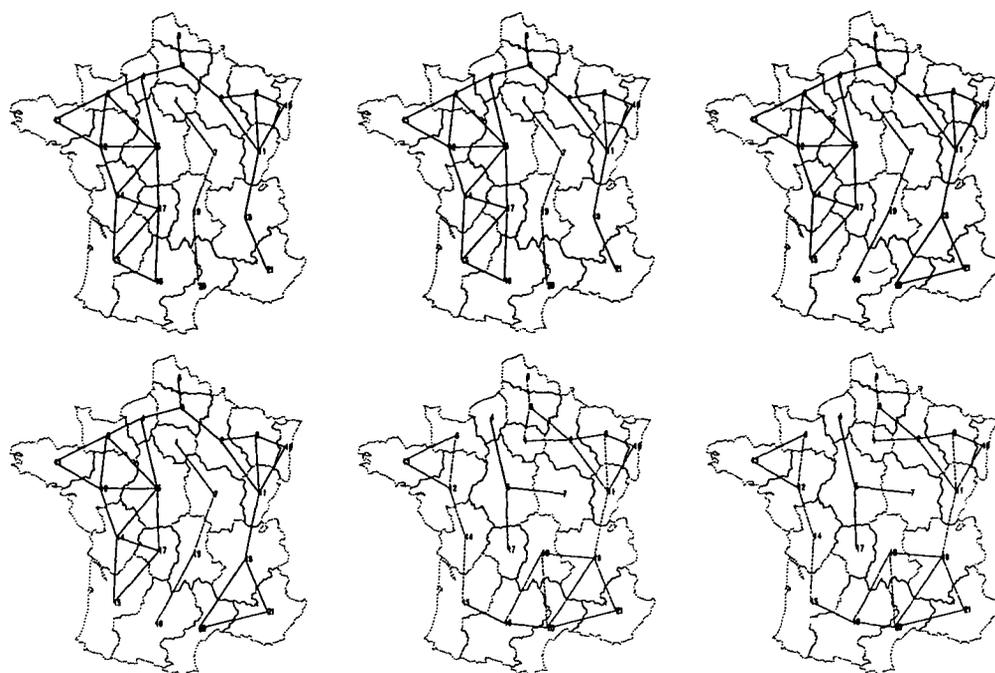


FIG. III.11 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux M34 (D_0)

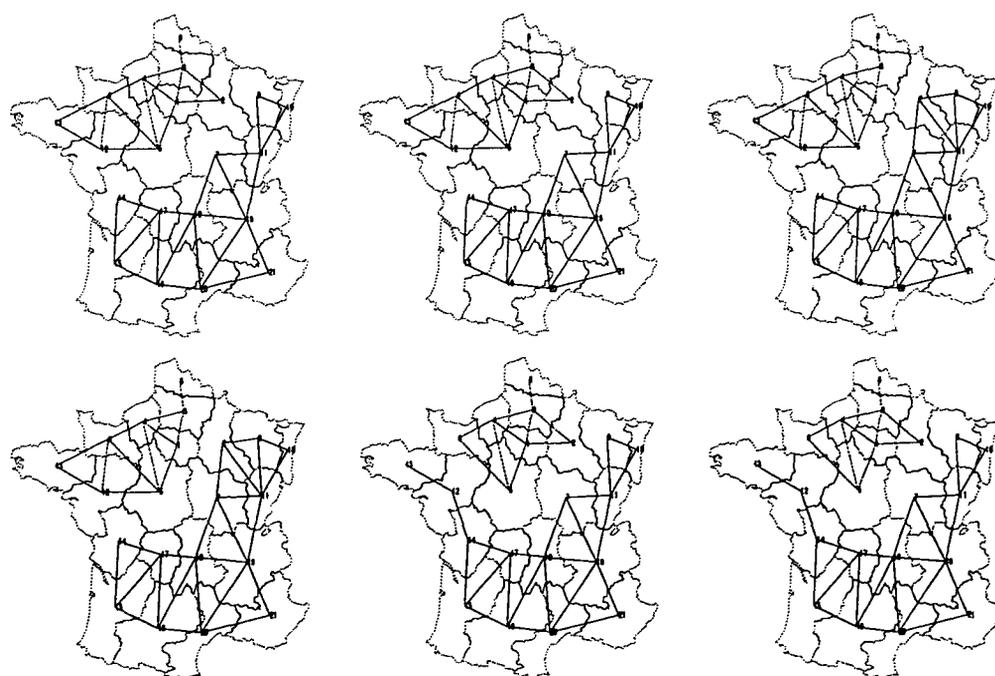
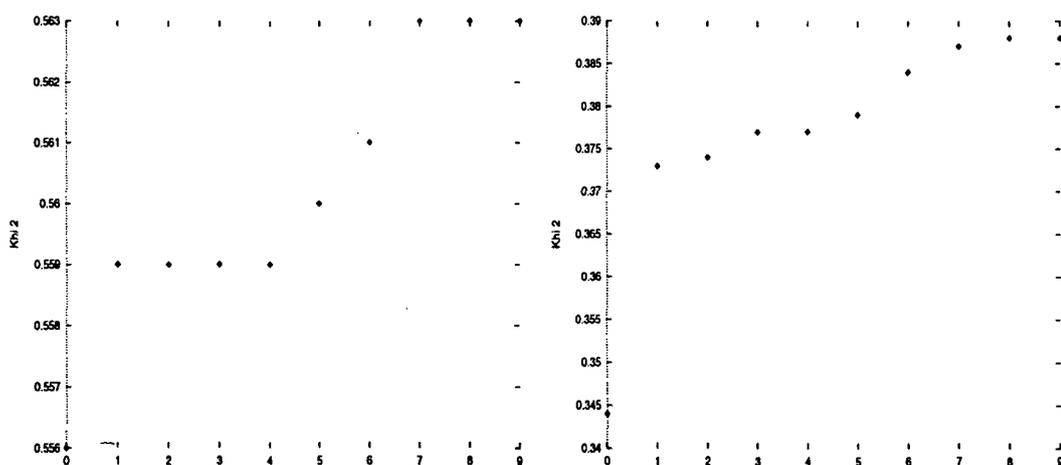


FIG. III.12 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux M34 (D_0)

MS15	$\gamma < 1$			$\gamma > 1$			
	Rang	ϵ	Err	γ	ϵ	Err	γ
0		0.978	0.556	0.084	0.604	0.344	3.502
1		0.982	0.559	0.113	0.655	0.373	3.194
2		0.982	0.559	0.658	0.658	0.374	3.007
3		0.983	0.559	0.282	0.662	0.377	3.021
4		0.983	0.559	0.527	0.663	0.377	2.965
5		0.985	0.560	0.685	0.667	0.379	3.027
6		0.986	0.561	0.620	0.675	0.384	2.882
7		0.989	0.563	0.805	0.681	0.387	2.829
8		0.990	0.563	0.162	0.682	0.388	2.821
9		0.990	0.563	0.178	0.683	0.388	2.827

TAB. III.5 - *Ecart des dix meilleures partitions pour le modèle à effet de barrière (MS15, D_0)*FIG. III.13 - "Err" ($\gamma < 1$, MS15, D_0) FIG. III.14 - "Err" ($\gamma > 1$, MS15, D_0)

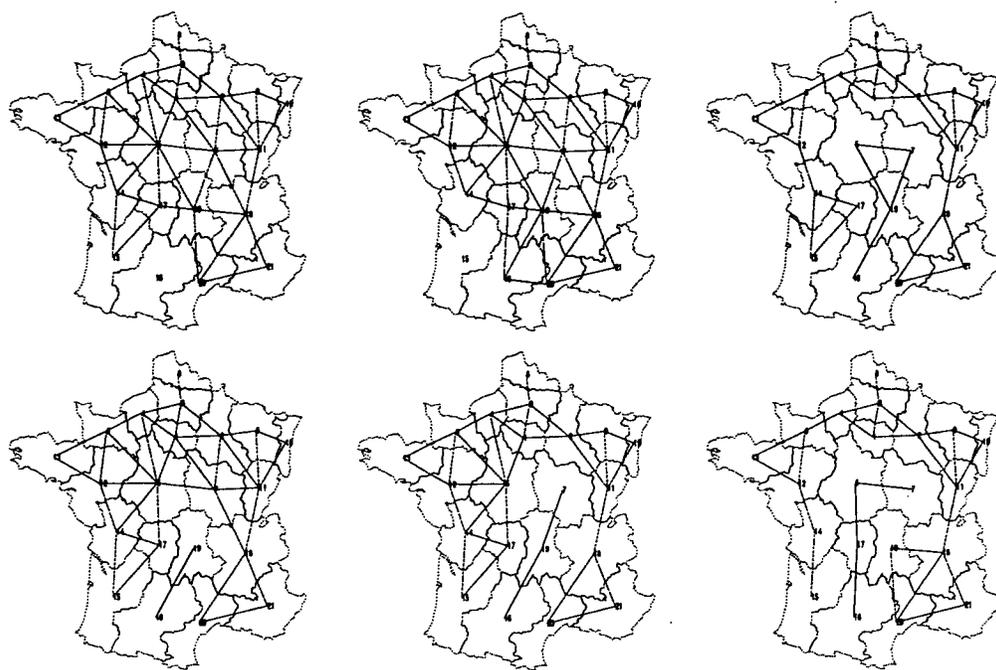


FIG. III.15 – Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux MS15 (D_0)

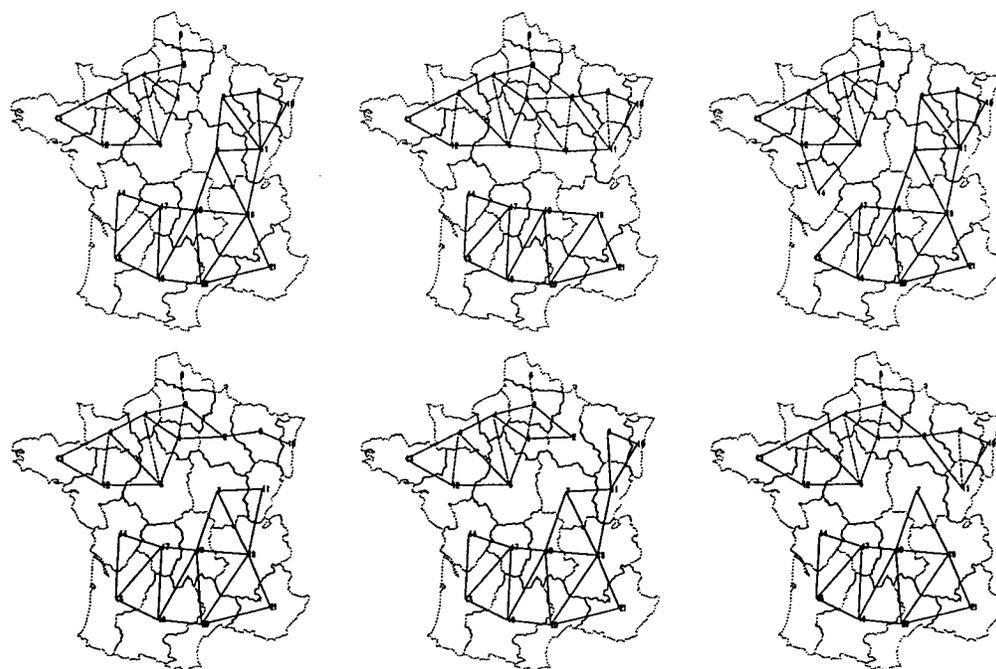
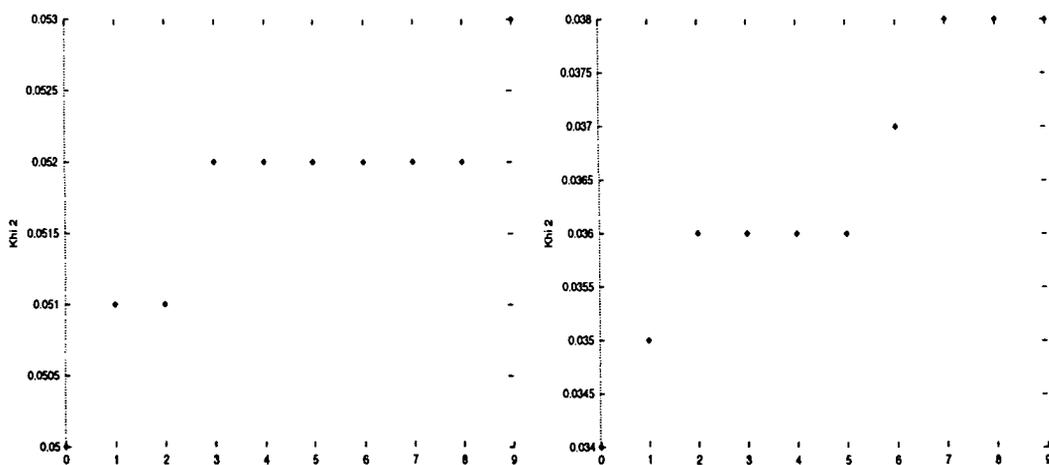


FIG. III.16 – Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux MS15 (D_0)

MS16	$\gamma < 1$			$\gamma > 1$		
	Rang	ϵ	Err	γ	ϵ	Err
0	0.900	0.050	0.576	0.624	0.034	2.148
1	0.920	0.051	0.708	0.629	0.035	2.149
2	0.928	0.051	0.720	0.643	0.036	2.095
3	0.940	0.052	0.741	0.655	0.036	2.069
4	0.941	0.052	0.726	0.655	0.036	2.100
5	0.942	0.052	0.742	0.656	0.036	2.148
6	0.946	0.052	0.753	0.675	0.037	2.025
7	0.946	0.052	0.742	0.686	0.038	1.979
8	0.948	0.052	0.679	0.687	0.038	2.044
9	0.953	0.053	0.684	0.693	0.038	2.238

TAB. III.6 - *Ecart des dix meilleures partitions pour le modèle à effet de barrière (MS16, D_0)*FIG. III.17 - "Err" ($\gamma < 1$, MS16, D_0)FIG. III.18 - "Err" ($\gamma > 1$, MS16, D_0)

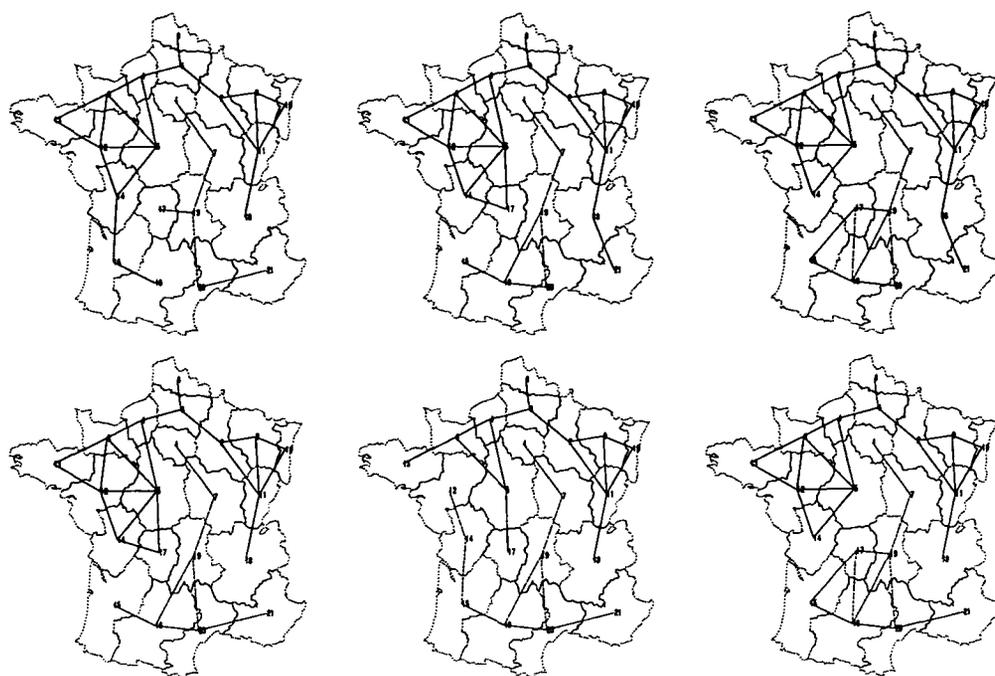


FIG. III.19 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux MS16 (D_0)

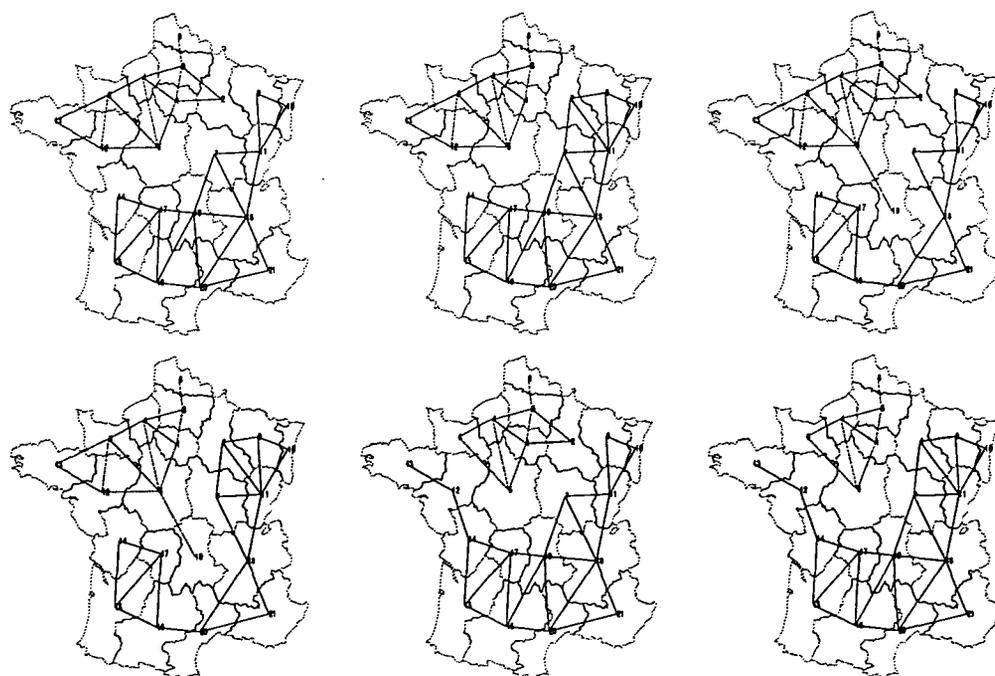
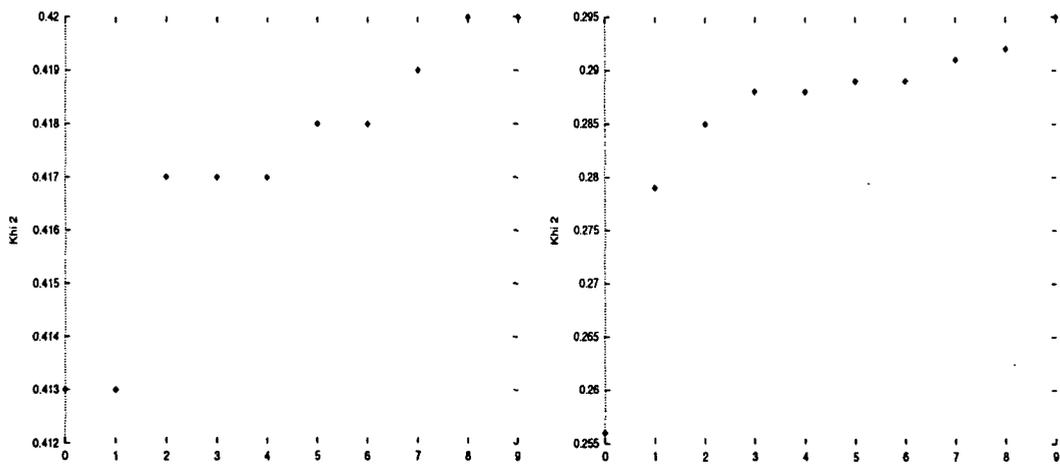


FIG. III.20 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux MS16 (D_0)

MS19 Rang	$\gamma < 1$			$\gamma > 1$		
	ϵ	Err	γ	ϵ	Err	γ
0	0.977	0.413	0.390	0.607	0.256	3.299
1	0.979	0.413	0.455	0.662	0.279	2.761
2	0.987	0.417	0.638	0.674	0.285	2.871
3	0.989	0.417	0.616	0.682	0.288	2.654
4	0.989	0.417	0.782	0.683	0.288	2.708
5	0.989	0.418	0.296	0.685	0.289	2.734
6	0.991	0.418	0.692	0.685	0.289	2.610
7	0.992	0.419	0.718	0.688	0.291	2.574
8	0.996	0.420	0.772	0.692	0.292	2.758
9	0.996	0.420	0.824	0.699	0.295	2.557

TAB. III.7 – *Ecart des dix meilleures partitions pour le modèle à effet de barrière (MS19, D_0)*FIG. III.21 - "Err" ($\gamma < 1$, MS19, D_0) FIG. III.22 - "Err" ($\gamma > 1$, MS19, D_0)

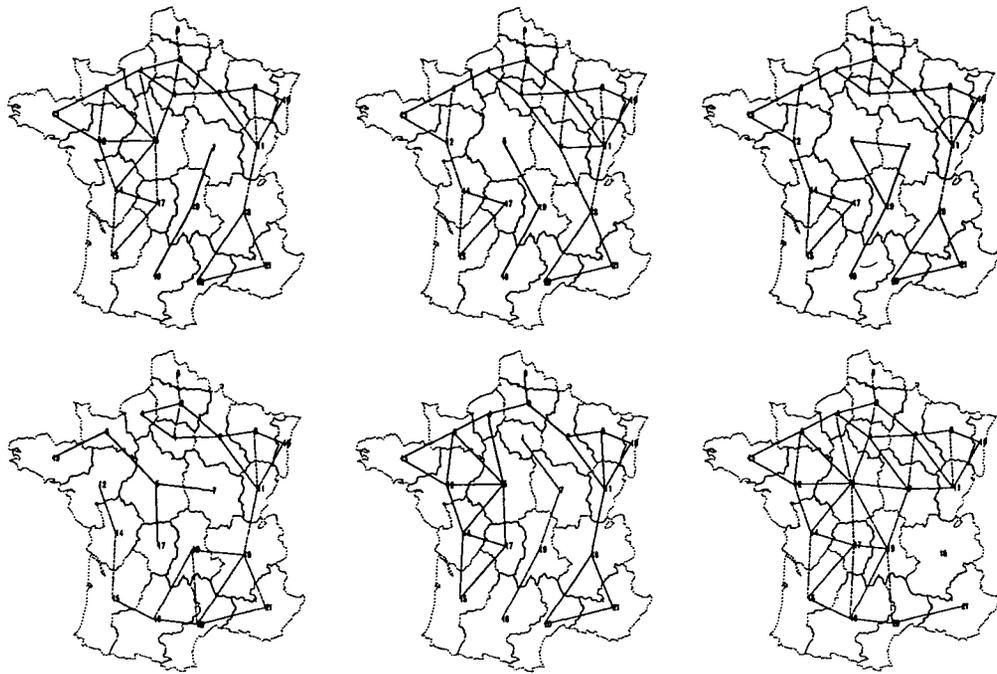


FIG. III.23 – Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux MS19 (D_0)

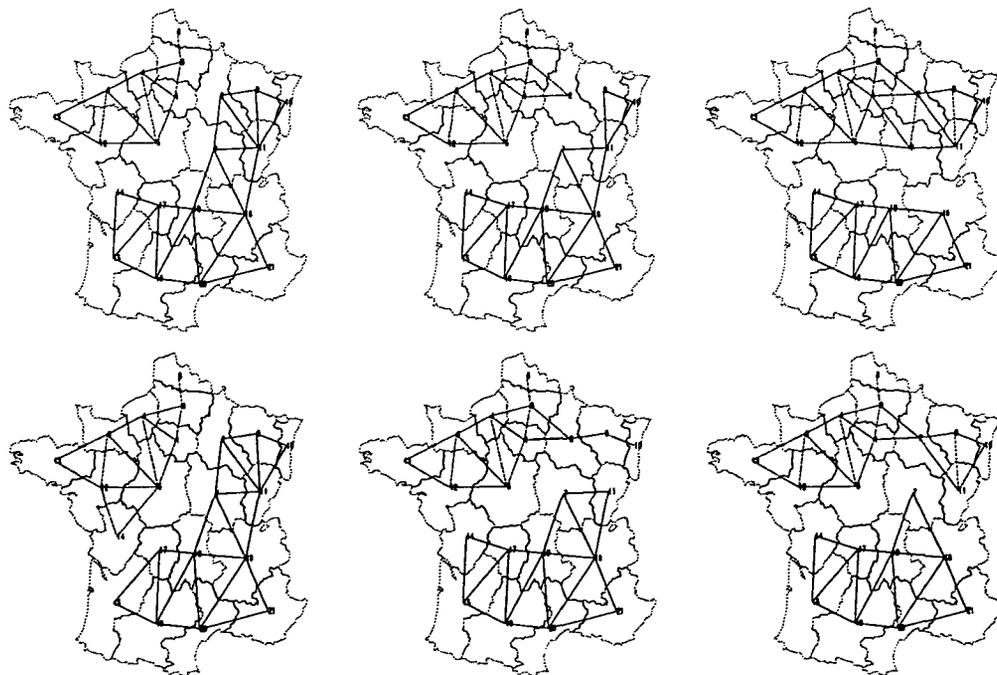
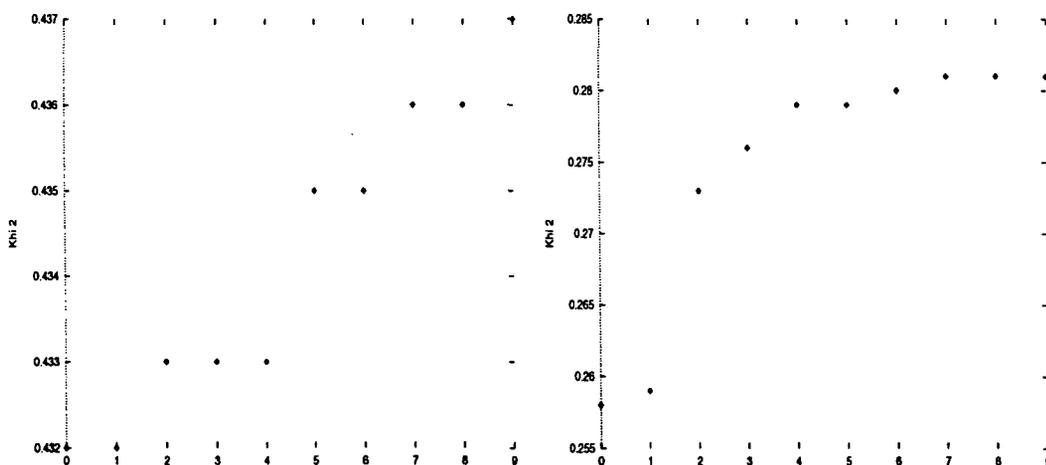


FIG. III.24 – Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux MS19 (D_0)

VLS	$\gamma < 1$			$\gamma > 1$		
	ϵ	Err	γ	ϵ	Err	γ
0	0.972	0.432	0.239	0.581	0.258	4.311
1	0.972	0.432	0.510	0.584	0.259	4.454
2	0.974	0.433	0.729	0.614	0.273	4.105
3	0.975	0.433	0.724	0.620	0.276	4.071
4	0.975	0.433	0.636	0.627	0.279	3.660
5	0.978	0.435	0.687	0.627	0.279	4.085
6	0.978	0.435	0.626	0.630	0.280	3.876
7	0.980	0.436	0.566	0.632	0.281	4.015
8	0.981	0.436	0.760	0.632	0.281	3.759
9	0.983	0.437	0.749	0.633	0.281	3.696

TAB. III.8 - *Ecart des dix meilleures partitions pour le modèle à effet de barrière (VLS, D_0)*FIG. III.25 - "Err" ($\gamma < 1$, VLS, D_0)FIG. III.26 - "Err" ($\gamma > 1$, VLS, D_0)

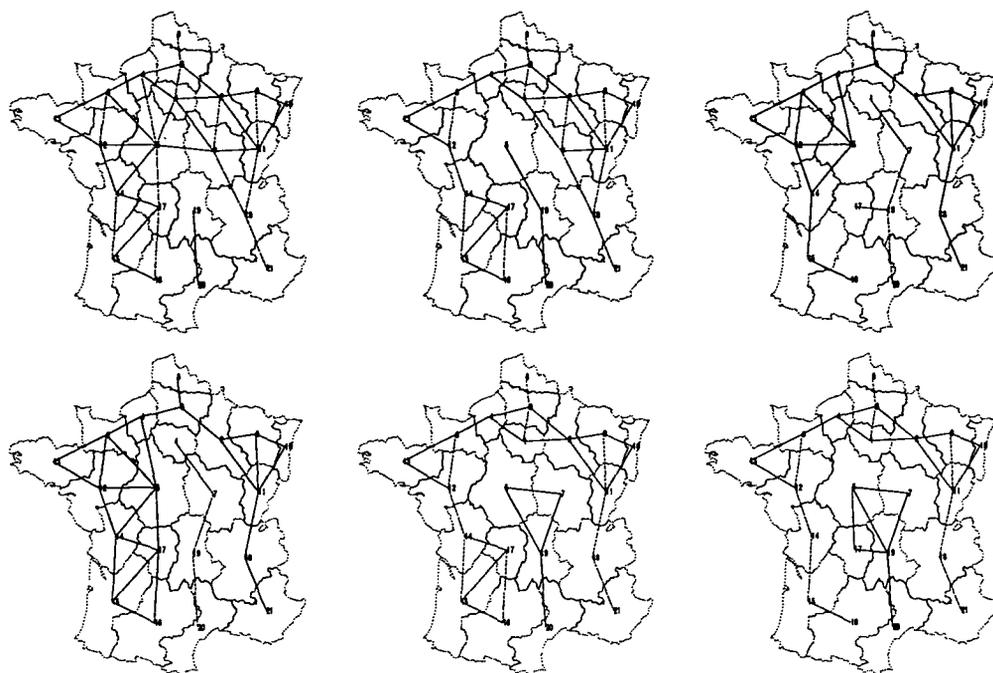


FIG. III.27 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux VLS (D_0)

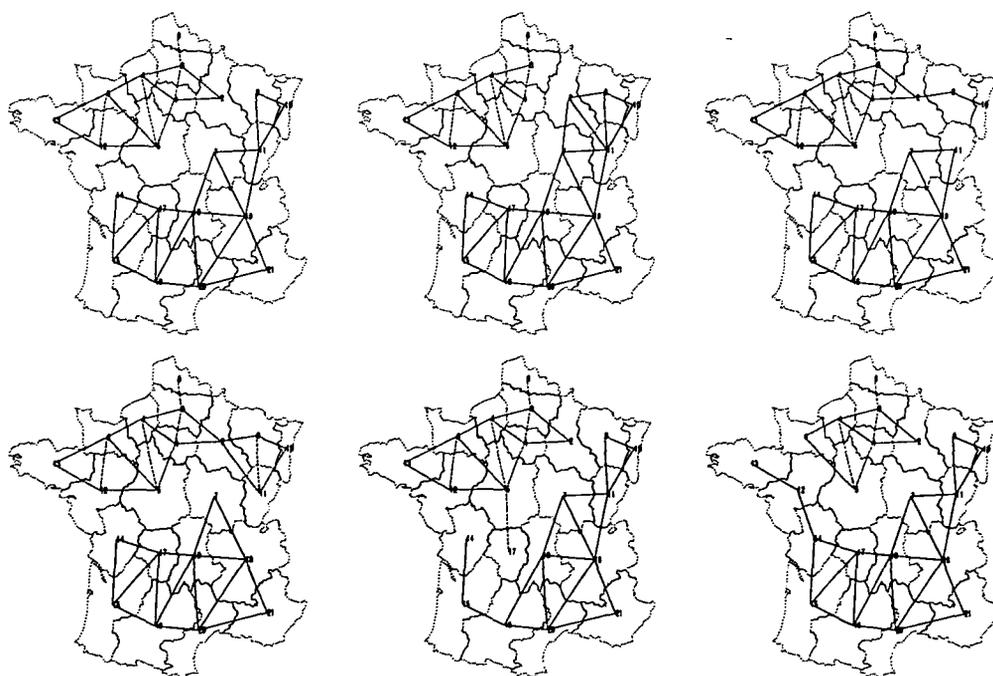
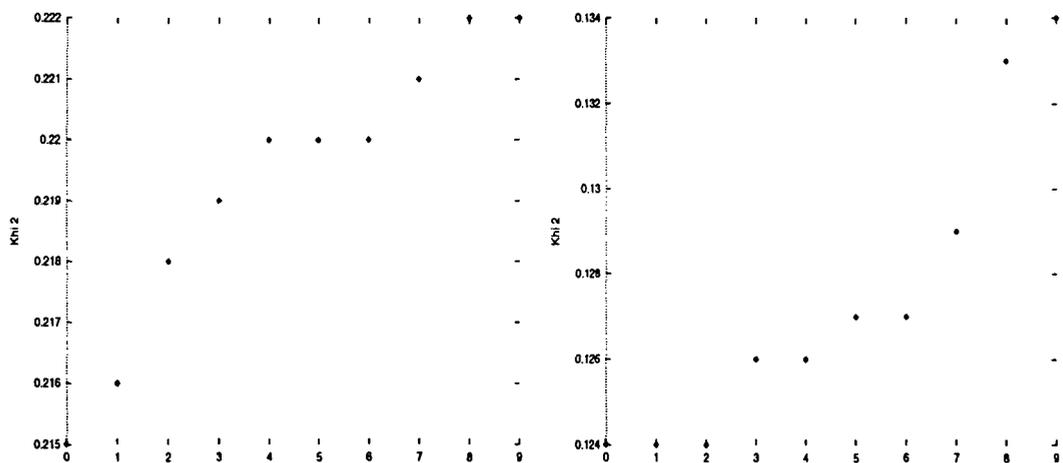


FIG. III.28 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux VLS (D_0)

VLS2 Rang	$\gamma < 1$			$\gamma > 1$		
	ϵ	Err	γ	ϵ	Err	γ
0	0.943	0.215	0.615	0.544	0.124	3.649
1	0.950	0.216	0.623	0.545	0.124	3.526
2	0.959	0.218	0.620	0.546	0.124	3.741
3	0.964	0.219	0.652	0.552	0.126	3.469
4	0.967	0.220	0.395	0.555	0.126	3.496
5	0.968	0.220	0.644	0.558	0.127	3.424
6	0.969	0.220	0.422	0.560	0.127	3.476
7	0.973	0.221	0.140	0.567	0.129	3.575
8	0.974	0.222	0.756	0.584	0.133	3.379
9	0.975	0.222	0.754	0.588	0.134	3.244

TAB. III.9 - *Ecart des dix meilleures partitions pour le modèle à effet de barrière (VLS2, D_0)*FIG. III.29 - "Err" ($\gamma < 1$, VLS22, D_0) FIG. III.30 - "Err" ($\gamma > 1$, VLS2, D_0)

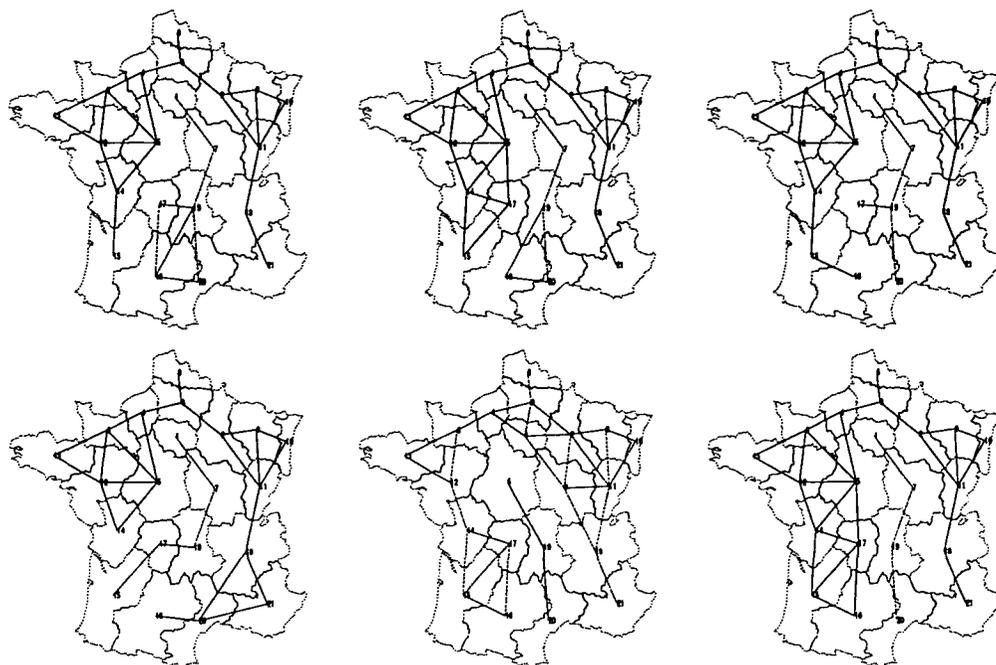


FIG. III.31 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux VLS2 (D_0)

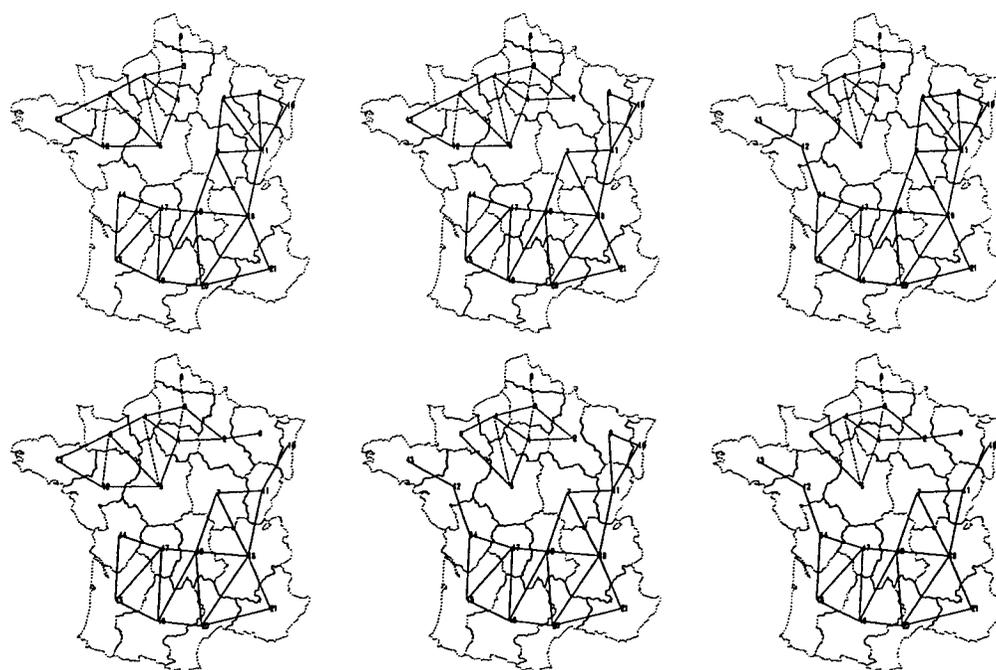
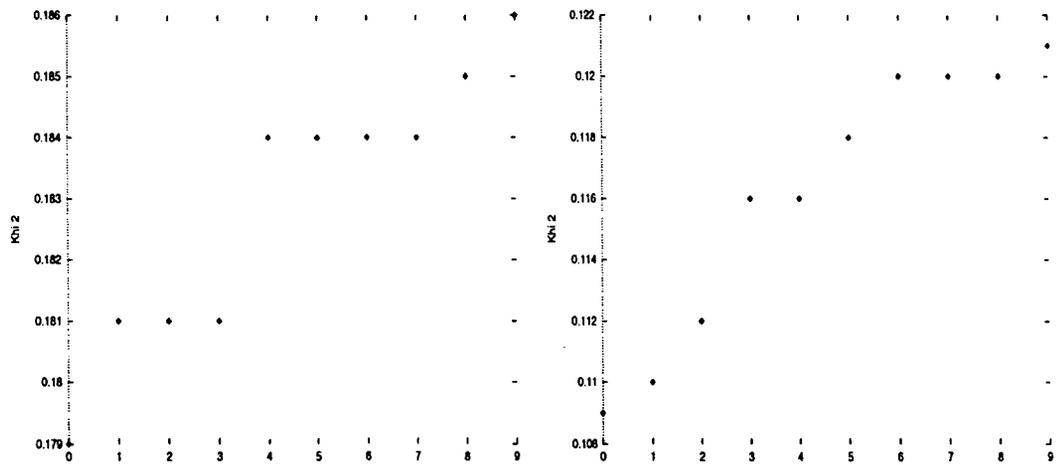


FIG. III.32 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux VLS2 (D_0)

VLS3	$\gamma < 1$			$\gamma > 1$		
	Rang	ϵ	Err	γ	ϵ	Err
0	0.925	0.179	0.591	0.560	0.109	3.629
1	0.931	0.181	0.563	0.569	0.110	3.483
2	0.935	0.181	0.595	0.575	0.112	3.483
3	0.935	0.181	0.585	0.597	0.116	4.088
4	0.946	0.184	0.552	0.598	0.116	3.255
5	0.948	0.184	0.589	0.610	0.118	3.159
6	0.950	0.184	0.620	0.617	0.120	3.452
7	0.950	0.184	0.618	0.620	0.120	3.175
8	0.952	0.185	0.691	0.621	0.120	3.136
9	0.960	0.186	0.604	0.624	0.121	3.142

TAB. III.10 - *Ecart des dix meilleures partitions pour le modèle à effet de barrière (VLS3, D_0)*FIG. III.33 - "Err" ($\gamma < 1$, VLS3, D_0) FIG. III.34 - "Err" ($\gamma > 1$, VLS3, D_0)

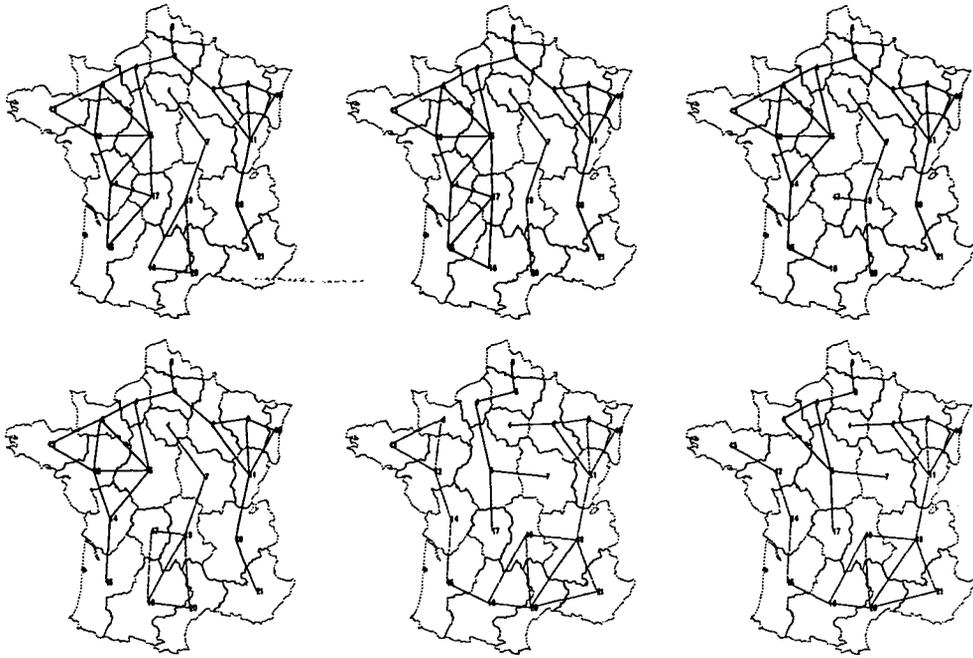


FIG. III.35 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux VLS3 (D_0)

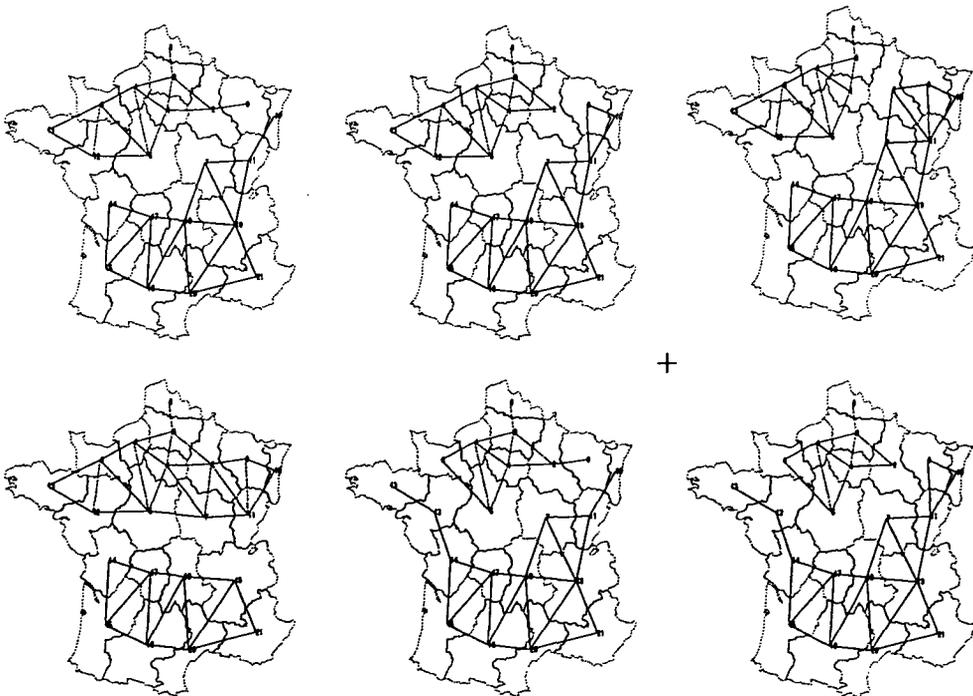
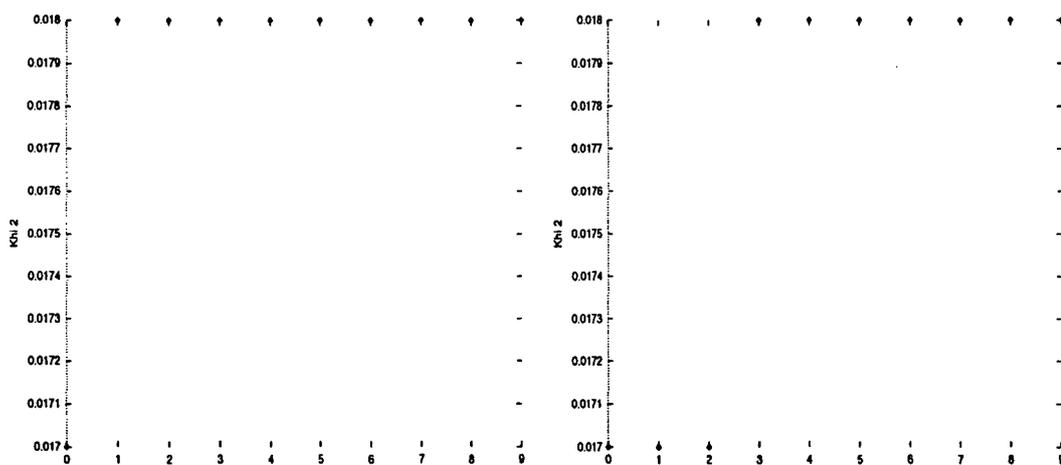


FIG. III.36 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux VLS3 (D_0)

M31 Rang	$\gamma < 1$				$\gamma > 1$			
	ϵ	Err	α	γ	ϵ	Err	α	γ
0	0.049	0.017	-2.32	0.680	0.048	0.017	-2.19	1.426
1	0.050	0.018	-2.36	0.695	0.048	0.017	-2.15	1.404
2	0.050	0.018	-2.33	0.705	0.049	0.017	-2.15	1.388
3	0.051	0.018	-2.35	0.708	0.050	0.018	-2.18	1.394
4	0.051	0.018	-2.34	0.660	0.050	0.018	-2.17	1.379
5	0.052	0.018	-2.43	0.698	0.051	0.018	-2.14	1.363
6	0.052	0.018	-2.39	0.705	0.051	0.018	-2.15	1.350
7	0.052	0.018	-2.36	0.721	0.052	0.018	-2.15	1.339
8	0.052	0.018	-2.36	0.721	0.052	0.018	-2.17	1.333
9	0.052	0.018	-2.37	0.714	0.052	0.018	-2.16	1.342

TAB. III.11 – *Ecart des dix meilleures partitions pour le modèle Paréto + Barrière (M31, D_0)*FIG. III.37 – “Err” ($\gamma < 1$, M31, D_0)FIG. III.38 – “Err” ($\gamma > 1$, M31, D_0)

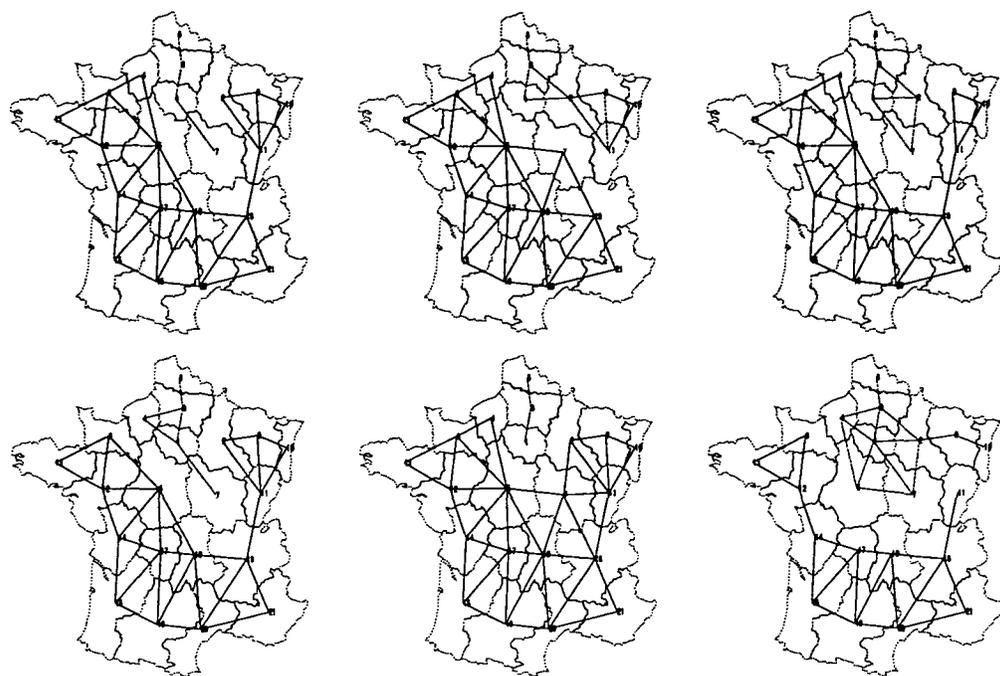


FIG. III.39 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma < 1$) appliqué à la matrice de flux M31 (D_0)

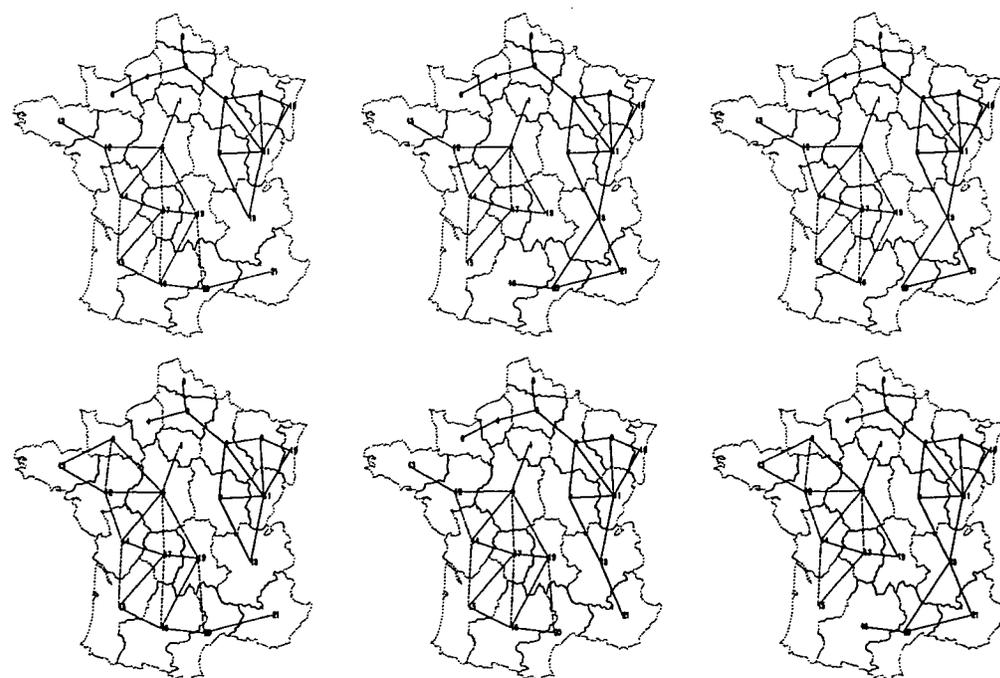


FIG. III.40 - Cartes des six meilleurs découpages de la France pour le modèle Barrière ($\gamma > 1$) appliqué à la matrice de flux M31 (D_0)

M32	$\gamma < 1$				$\gamma > 1$			
	Rang	ϵ	Err	α	γ	ϵ	Err	α
0	0.126	0.014	-1.76	0.592	0.142	0.015	-1.35	1.540
1	0.130	0.014	-1.72	0.621	0.144	0.016	-1.36	1.506
2	0.131	0.014	-1.75	0.607	0.145	0.016	-1.38	1.454
3	0.132	0.014	-1.70	0.626	0.146	0.016	-1.39	1.433
4	0.133	0.014	-1.66	0.640	0.147	0.016	-1.37	1.463
5	0.133	0.014	-1.76	0.609	0.148	0.016	-1.38	1.439
6	0.134	0.014	-1.67	0.628	0.150	0.016	-1.39	1.469
7	0.135	0.015	-1.63	0.653	0.150	0.016	-1.38	1.441
8	0.138	0.015	-1.59	0.648	0.151	0.016	-1.43	1.344
9	0.138	0.015	-1.69	0.646	0.152	0.016	-1.43	1.338

TAB. III.12 - *Ecart des dix meilleures partitions pour le modèle Parèto + Barrière (M32, D_0)*

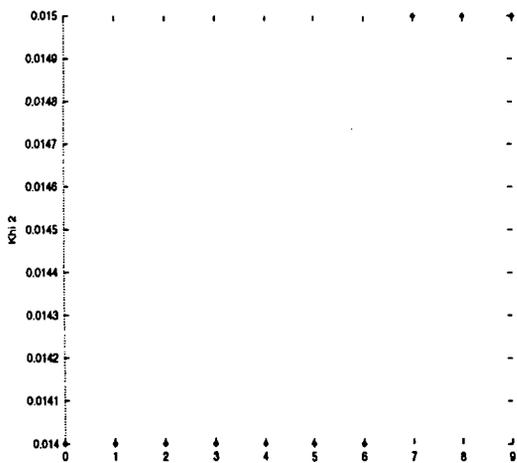


FIG. III.41 - "Err" ($\gamma < 1$, M32, D_0)

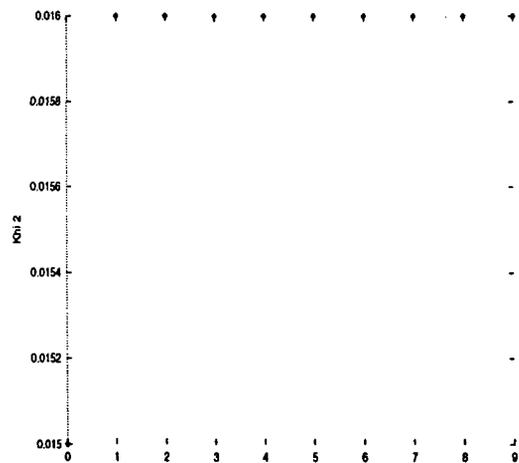


FIG. III.42 - "Err" ($\gamma > 1$, M32, D_0)

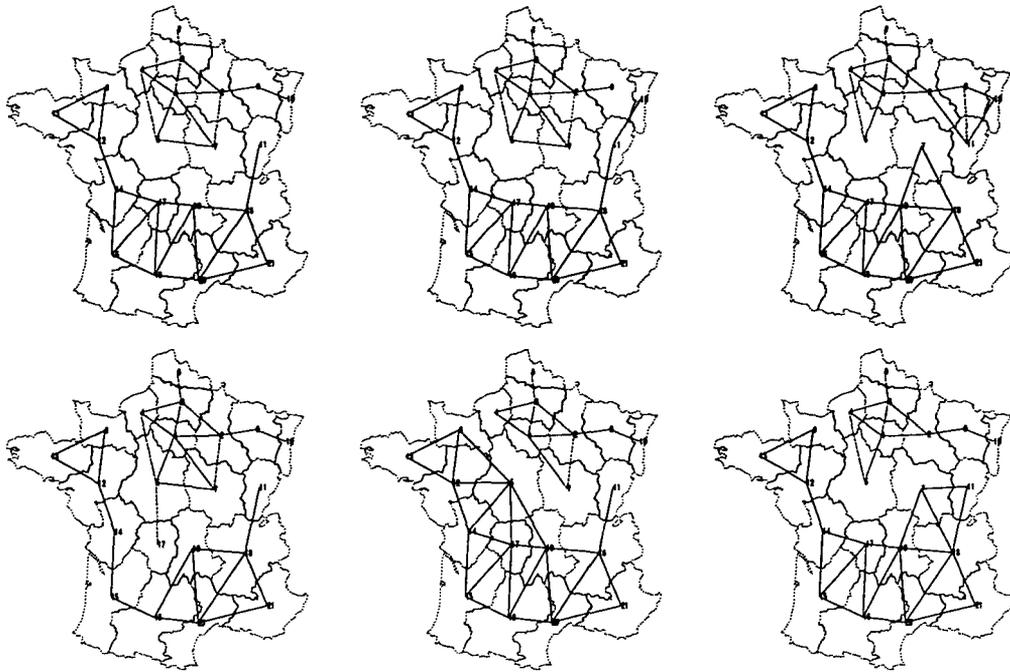


FIG. III.43 – Cartes des six meilleurs découpages de la France pour le modèle Pareto + Barrière ($\gamma < 1$) appliqué à la matrice de flux M32 (D_0)

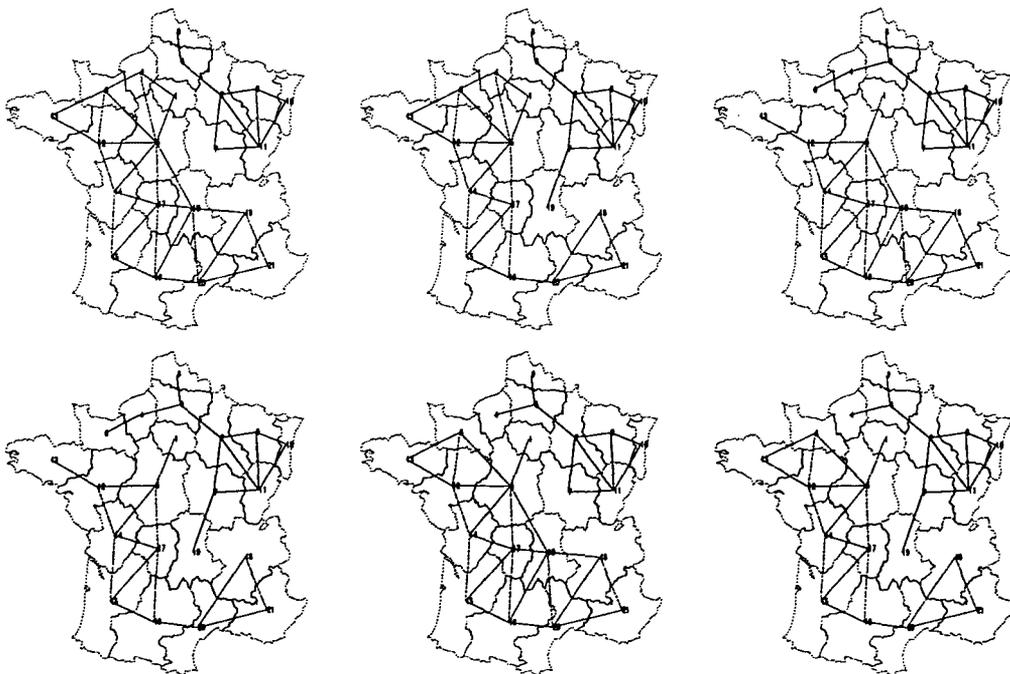
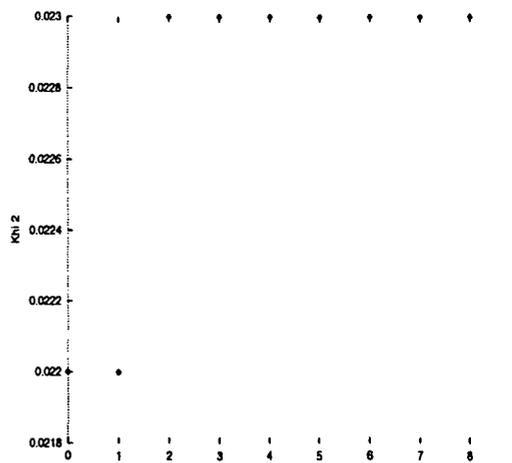
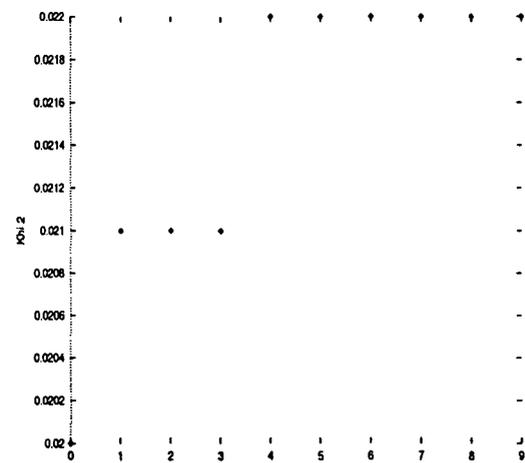


FIG. III.44 – Cartes des six meilleurs découpages de la France pour le modèle Pareto + Barrière ($\gamma > 1$) appliqué à la matrice de flux M32 (D_0)

M34	$\gamma < 1$				$\gamma > 1$			
Rang	ϵ	Err	α	γ	ϵ	Err	α	γ
0	0.077	0.022	-2.17	0.654	0.071	0.020	-1.91	1.523
1	0.077	0.022	-2.20	0.641	0.074	0.021	-1.90	1.486
2	0.078	0.023	-2.04	0.663	0.074	0.021	-1.89	1.507
3	0.078	0.023	-2.11	0.674	0.074	0.021	-1.90	1.457
4	0.078	0.023	-2.06	0.681	0.075	0.022	-1.90	1.459
5	0.079	0.023	-2.11	0.677	0.075	0.022	-1.89	1.440
6	0.079	0.023	-2.09	0.675	0.075	0.022	-1.90	1.471
7	0.079	0.023	-2.06	0.677	0.076	0.022	-1.91	1.431
8	0.079	0.023	-2.12	0.674	0.077	0.022	-1.89	1.483
9	0.081	0.023	-2.19	0.663	0.078	0.022	-1.89	1.421

TAB. III.13 - *Ecart des dix meilleures partitions pour le modèle Paréto + Barrière (M34, D_0)*FIG. III.45 - "Err" ($\gamma < 1$, M34, D_0)FIG. III.46 - "Err" ($\gamma > 1$, M34, D_0)

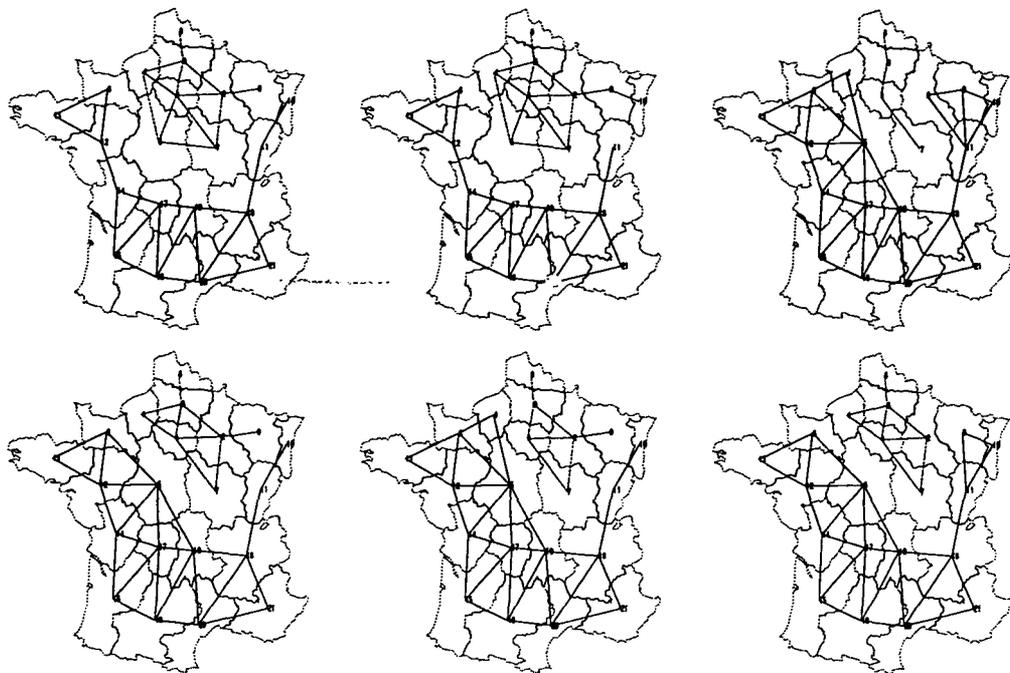


FIG. III.47 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma < 1$) appliqué à la matrice de flux M34 (D_0)

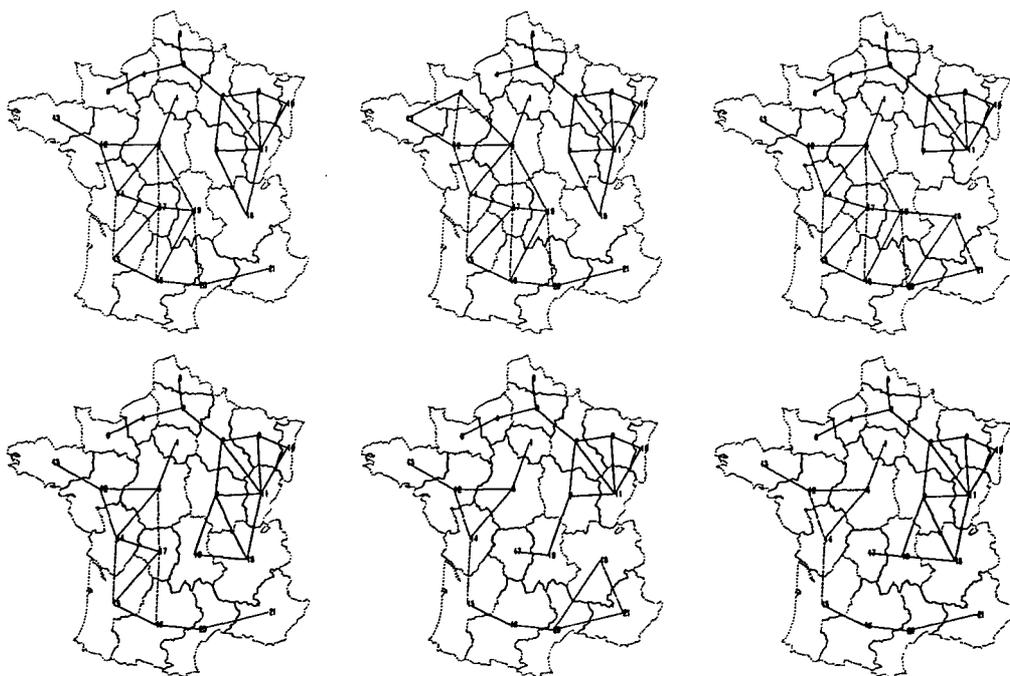


FIG. III.48 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma > 1$) appliqué à la matrice de flux M34 (D_0)

MS15 Rang	$\gamma < 1$				$\gamma > 1$			
	ϵ	Err	α	γ	ϵ	Err	α	γ
0	0.068	0.039	-2.44	0.586	0.066	0.037	-2.11	1.552
1	0.068	0.039	-2.50	0.601	0.067	0.038	-2.09	1.549
2	0.069	0.039	-2.48	0.617	0.067	0.038	-2.05	1.555
3	0.070	0.040	-2.49	0.617	0.072	0.041	-2.14	1.435
4	0.070	0.040	-2.42	0.615	0.072	0.041	-2.11	1.449
5	0.070	0.040	-2.50	0.615	0.072	0.041	-2.12	1.450
6	0.071	0.041	-2.43	0.626	0.073	0.041	-2.10	1.449
7	0.072	0.041	-2.42	0.633	0.073	0.042	-2.13	1.433
8	0.074	0.042	-2.39	0.619	0.073	0.042	-2.12	1.436
9	0.074	0.042	-2.42	0.672	0.073	0.042	-2.09	1.442

TAB. III.14 - *Ecarts des dix meilleures partitions pour le modèle Paréto + Barrière (MS15, D_0)*

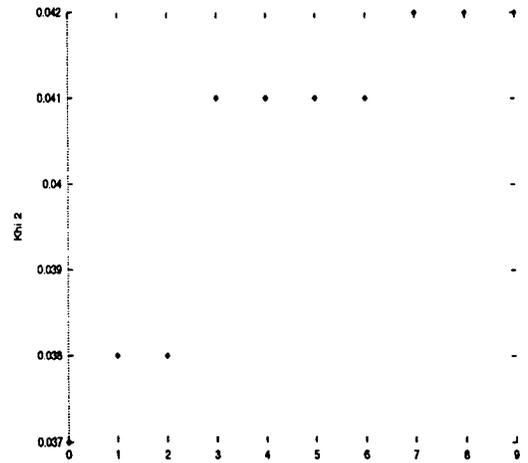
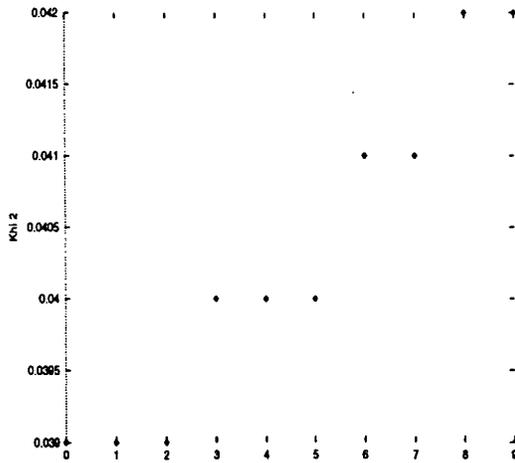


FIG. III.49 - "Err" ($\gamma < 1$, MS15, D_0) FIG. III.50 - "Err" ($\gamma > 1$, MS15, D_0)

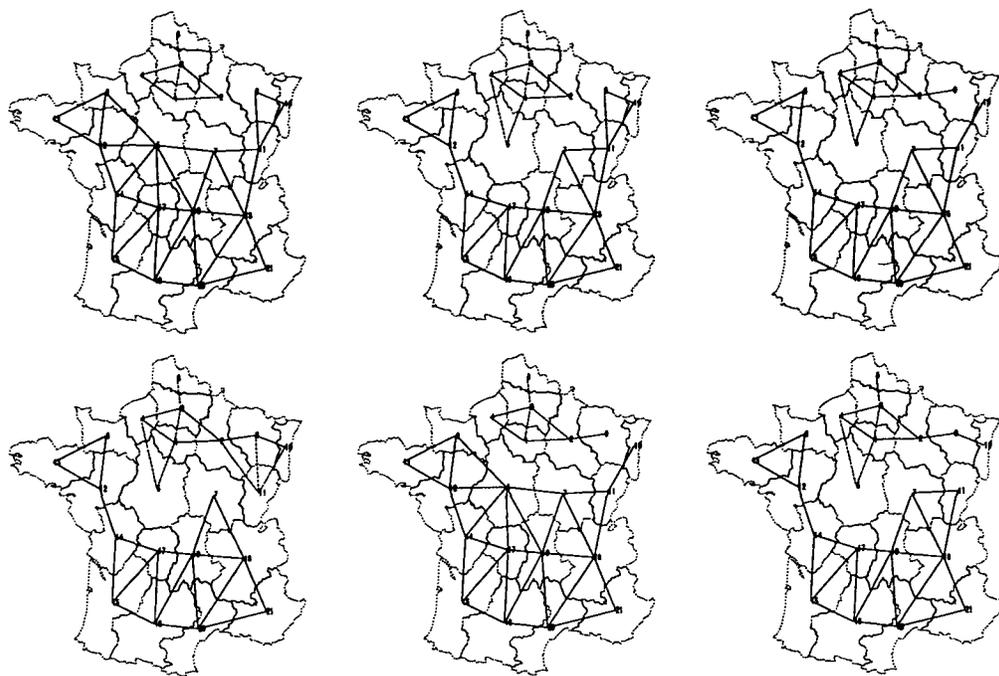


FIG. III.51 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma < 1$) appliqué à la matrice de flux MS15 (D_0)

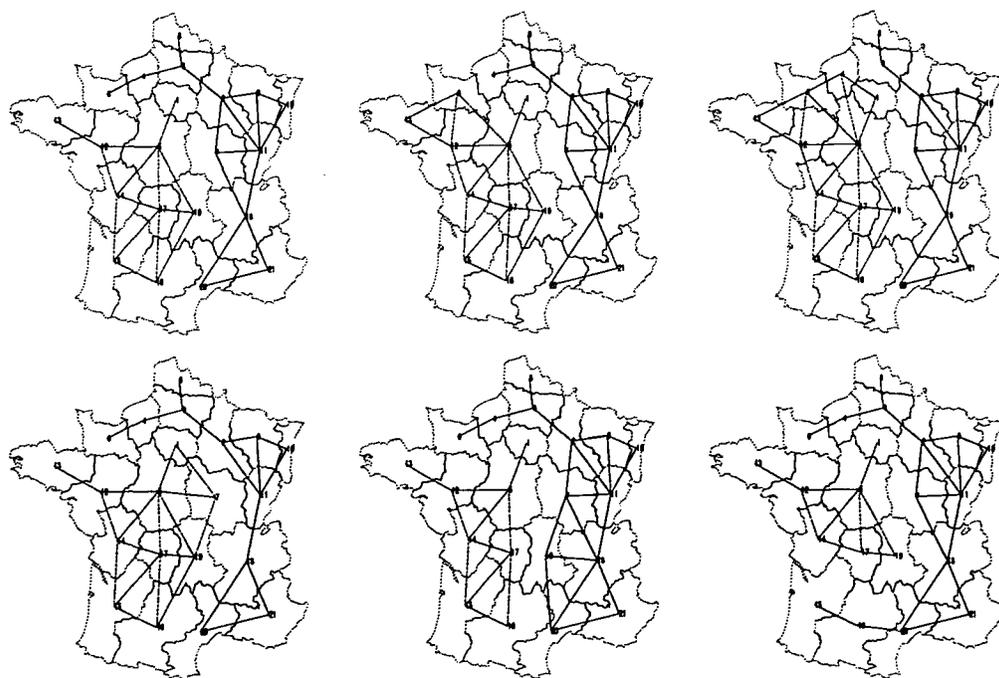


FIG. III.52 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma > 1$) appliqué à la matrice de flux MS15 (D_0)

MS16	$\gamma < 1$				$\gamma > 1$			
	Rang	ϵ	Err	α	γ	ϵ	Err	α
0	0.353	0.020	-1.37	0.554	0.391	0.022	-1.10	1.450
1	0.355	0.020	-1.35	0.569	0.392	0.022	-1.04	1.396
2	0.377	0.021	-1.28	0.638	0.394	0.022	-1.04	1.385
3	0.378	0.021	-1.30	0.604	0.394	0.022	-1.09	1.427
4	0.382	0.021	-1.22	0.669	0.398	0.022	-0.88	1.431
5	0.383	0.021	-1.28	0.650	0.400	0.022	-1.03	1.353
6	0.386	0.021	-1.22	0.676	0.402	0.022	-1.08	1.384
7	0.390	0.022	-1.23	0.676	0.404	0.022	-1.08	1.367
8	0.390	0.022	-1.29	0.662	0.404	0.022	-0.90	1.386
9	0.395	0.022	-1.27	0.680	0.406	0.022	-0.92	1.356

TAB. III.15 - *Ecart des dix meilleures partitions pour le modèle Paréto + Barrière (MS16, D_0)*

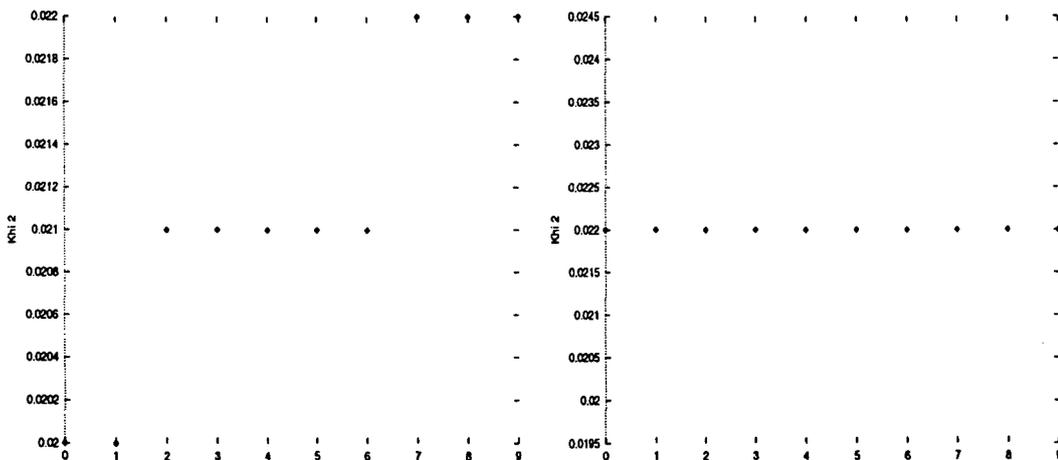


FIG. III.53 - "Err" ($\gamma < 1$, MS16, D_0) FIG. III.54 - "Err" ($\gamma > 1$, MS16, D_0)

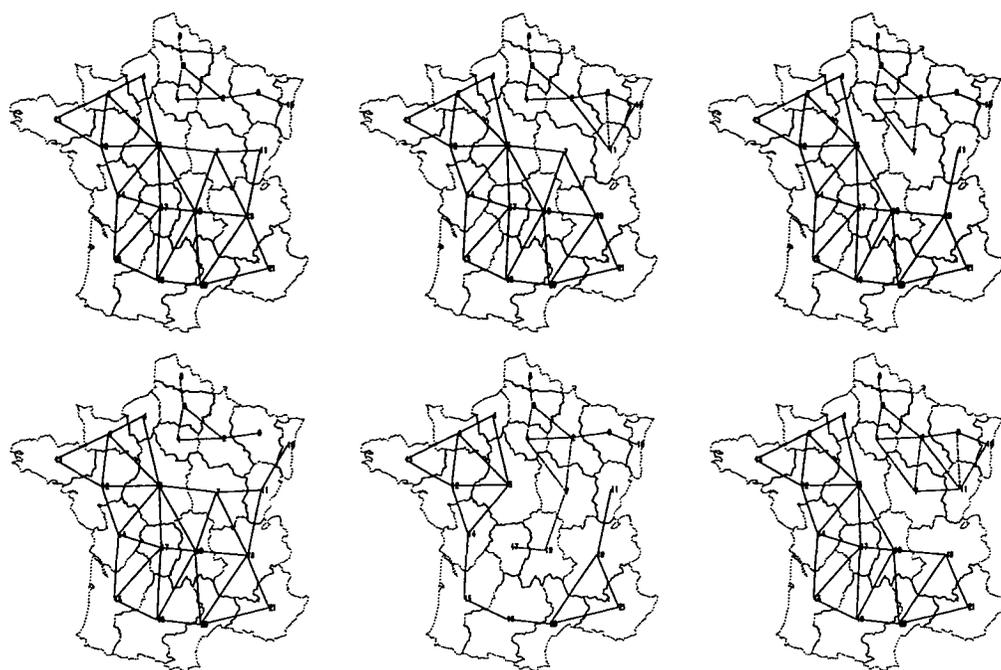


FIG. III.55 – Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma < 1$) appliqué à la matrice de flux MS16 (D_0)

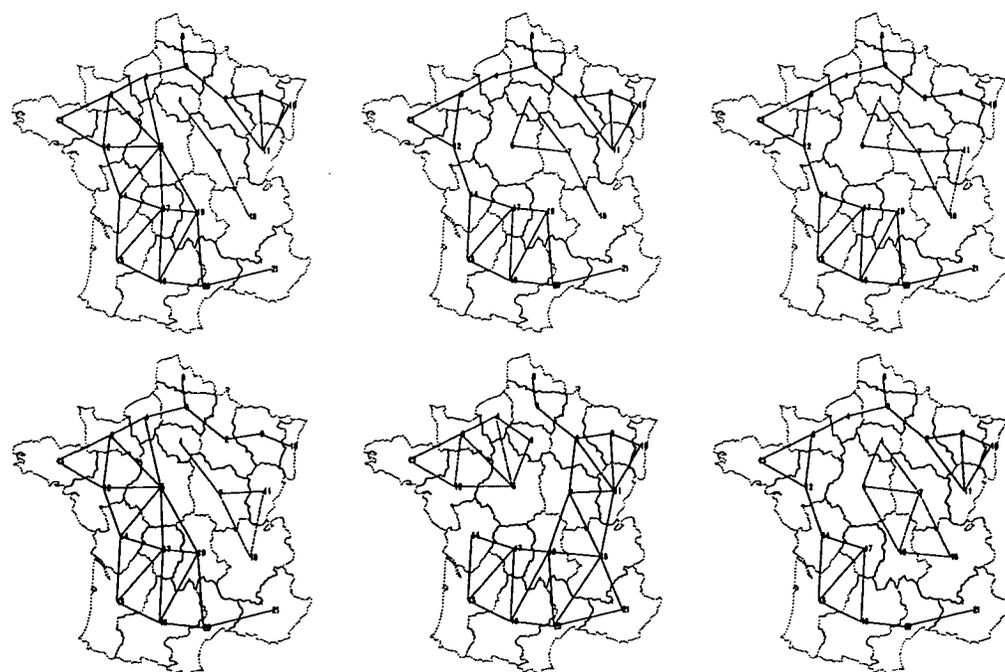


FIG. III.56 – Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma > 1$) appliqué à la matrice de flux MS16 (D_0)

MS19	$\gamma < 1$				$\gamma > 1$			
	Rang	ϵ	Err	α	γ	ϵ	Err	α
0	0.091	0.039	-2.24	0.600	0.097	0.041	-1.85	1.448
1	0.092	0.039	-2.26	0.598	0.097	0.041	-1.84	1.452
2	0.092	0.039	-2.27	0.593	0.098	0.041	-1.80	1.452
3	0.093	0.039	-2.16	0.607	0.099	0.042	-1.85	1.536
4	0.093	0.039	-2.17	0.595	0.099	0.042	-1.89	1.429
5	0.093	0.039	-2.24	0.604	0.100	0.042	-1.88	1.439
6	0.094	0.040	-2.19	0.635	0.100	0.042	-1.86	1.461
7	0.095	0.040	-2.20	0.632	0.100	0.042	-1.88	1.449
8	0.095	0.040	-2.18	0.615	0.100	0.042	-1.87	1.457
9	0.096	0.040	-2.17	0.625	0.101	0.042	-1.85	1.561

TAB. III.16 - *Ecart des dix meilleures partitions pour le modèle Paréto + Barrière (MS19, D_0)*

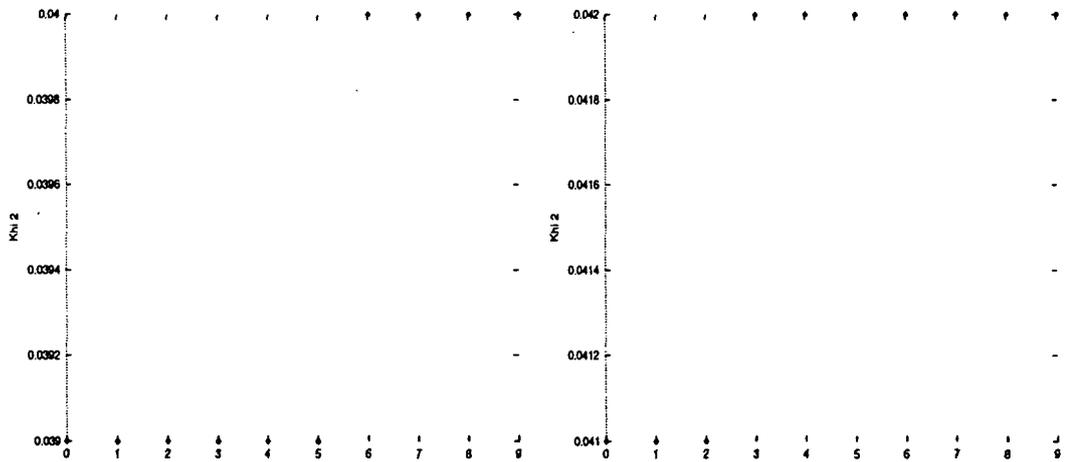


FIG. III.57 - "Err" ($\gamma < 1$, MS19, D_0) FIG. III.58 - "Err" ($\gamma > 1$, MS19, D_0)

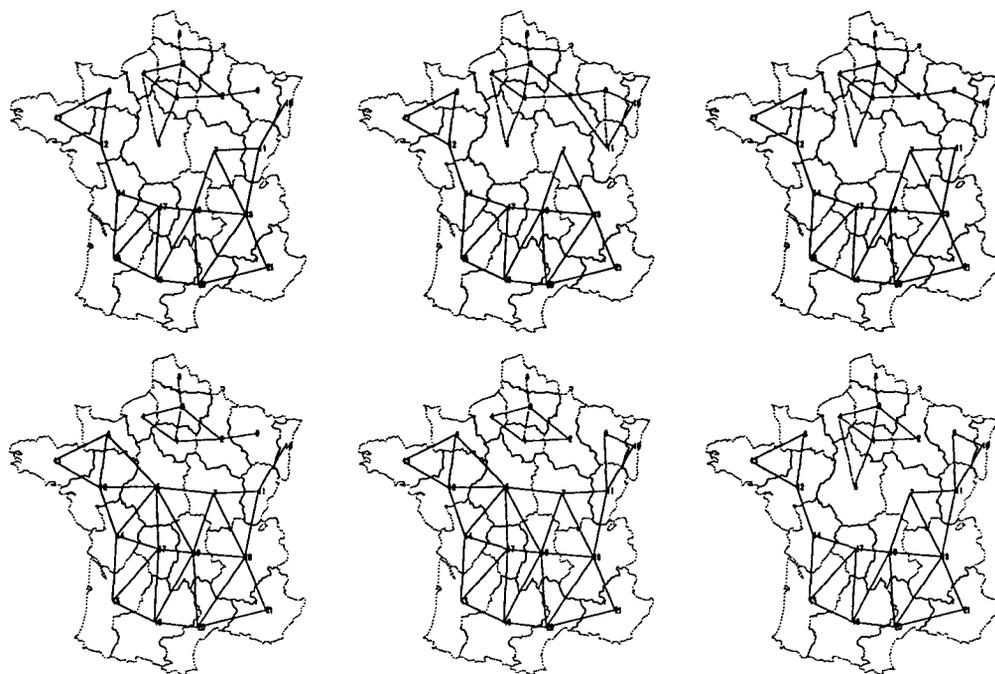


FIG. III.59 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma < 1$) appliqué à la matrice de flux MS19 (D_0)

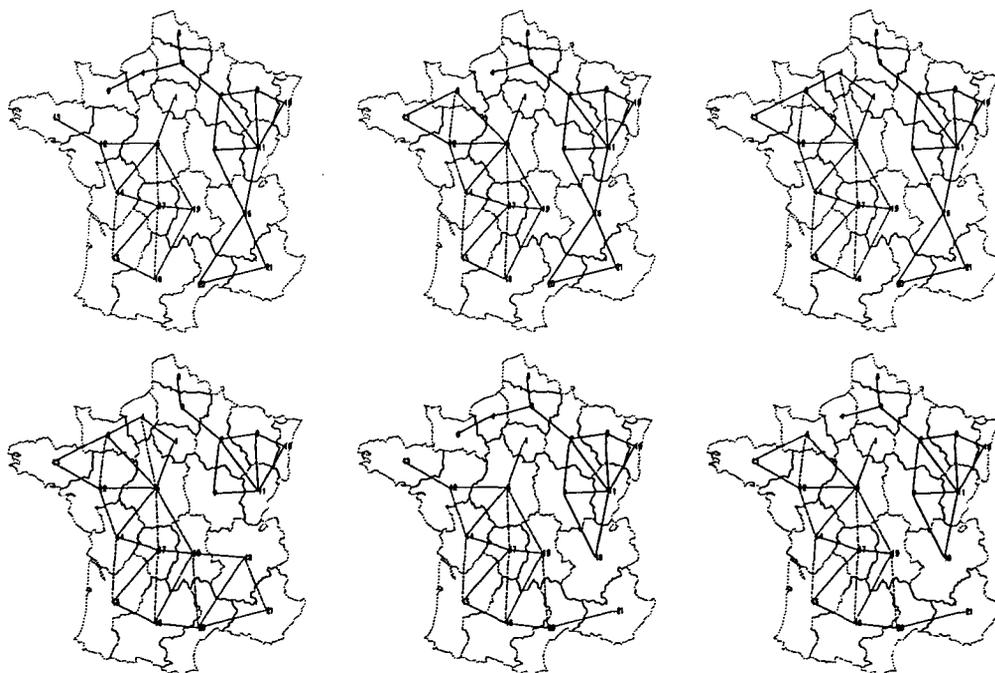
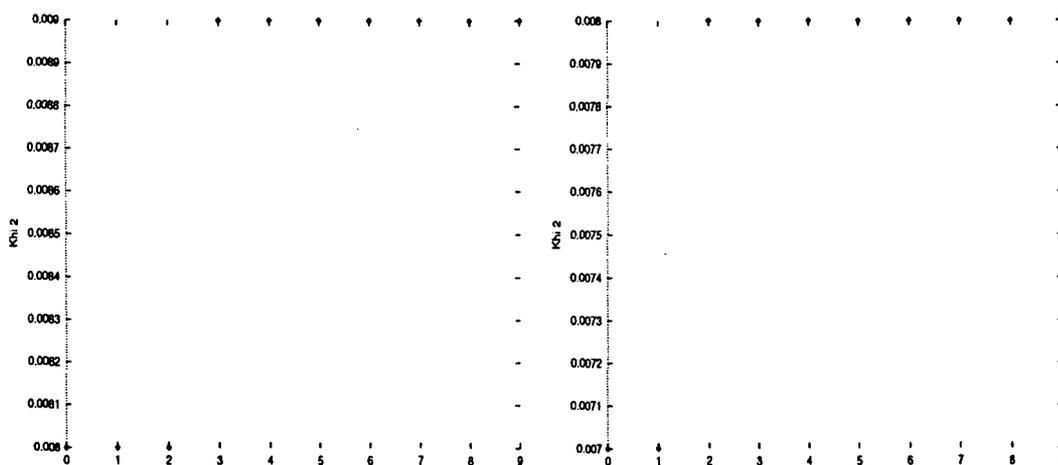


FIG. III.60 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma > 1$) appliqué à la matrice de flux MS19 (D_0)

VLS	$\gamma < 1$				$\gamma > 1$			
Rang	ϵ	Err	α	γ	ϵ	Err	α	γ
0	0.019	0.008	-3.28	0.734	0.015	0.007	-3.08	1.796
1	0.019	0.008	-3.26	0.741	0.017	0.007	-3.07	1.555
2	0.019	0.008	-3.25	0.743	0.018	0.008	-3.11	1.554
3	0.019	0.009	-3.28	0.741	0.018	0.008	-3.10	1.372
4	0.020	0.009	-3.33	0.699	0.018	0.008	-3.10	1.369
5	0.020	0.009	-3.28	0.755	0.019	0.008	-3.14	1.328
6	0.020	0.009	-3.22	0.716	0.019	0.008	-3.16	1.389
7	0.020	0.009	-3.27	0.753	0.019	0.008	-3.16	1.377
8	0.020	0.009	-3.19	0.743	0.019	0.008	-3.07	1.353
9	0.021	0.009	-3.21	0.808	0.019	0.008	-3.14	1.484

TAB. III.17 - *Ecart des dix meilleures partitions pour le modèle Paréto + Barrière (VLS, D_0)*FIG. III.61 - "Err" ($\gamma < 1$, VLS, D_0)FIG. III.62 - "Err" ($\gamma > 1$, VLS, D_0)

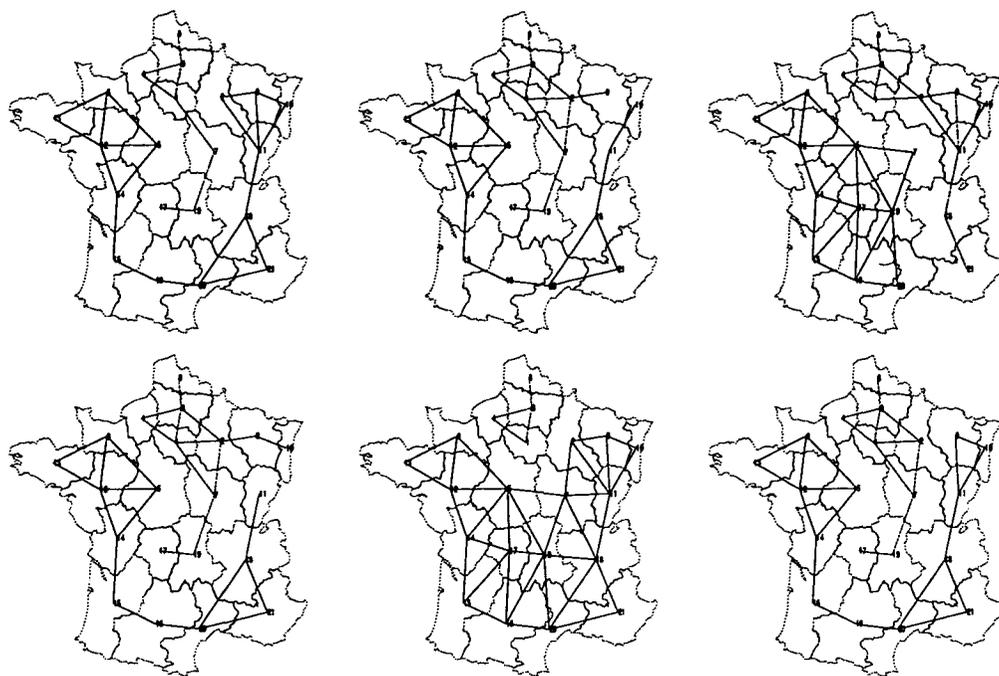


FIG. III.63 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma < 1$) appliqué à la matrice de flux VLS (D_0)

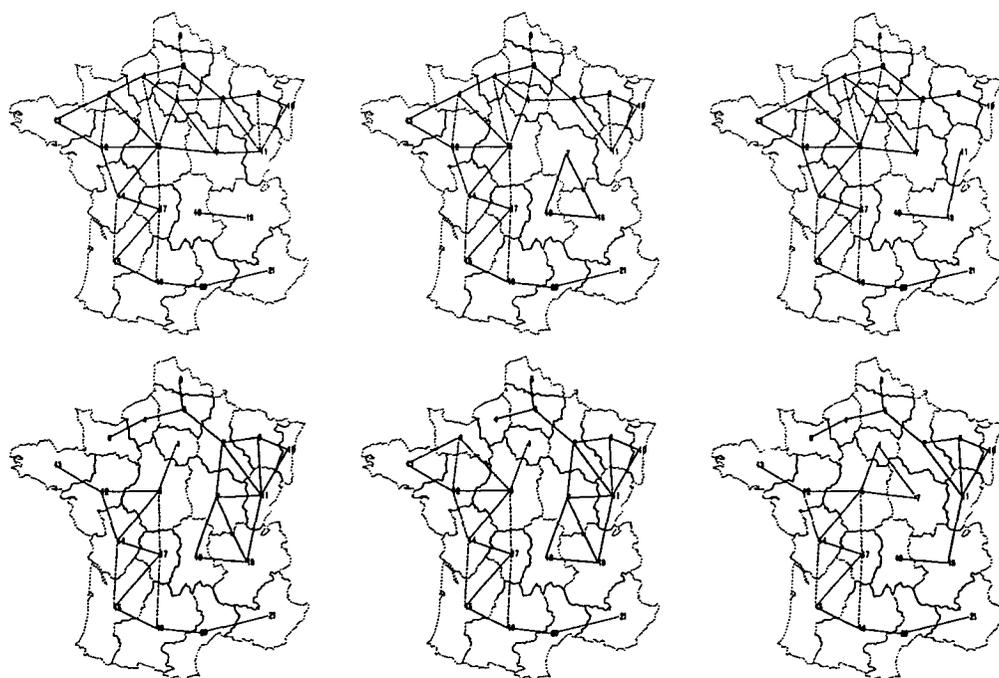
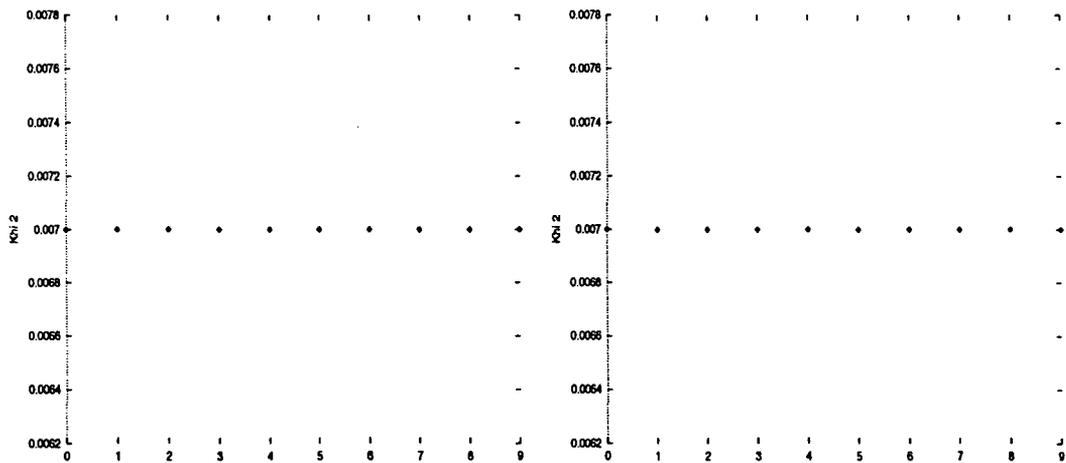


FIG. III.64 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma > 1$) appliqué à la matrice de flux VLS (D_0)

VLS2	$\gamma < 1$				$\gamma > 1$			
	Rang	ϵ	Err.	α	γ	ϵ	Err.	α
0	0.030	0.007	-2.66	0.779	0.031	0.007	-2.51	1.426
1	0.031	0.007	-2.64	0.770	0.032	0.007	-2.54	1.207
2	0.031	0.007	-2.65	0.800	0.032	0.007	-2.50	1.242
3	0.032	0.007	-2.59	0.818	0.032	0.007	-2.51	1.213
4	0.032	0.007	-2.65	0.815	0.032	0.007	-2.55	1.276
5	0.033	0.007	-2.68	0.816	0.032	0.007	-2.47	1.251
6	0.033	0.007	-2.62	0.792	0.032	0.007	-2.56	1.171
7	0.033	0.007	-2.63	0.794	0.033	0.007	-2.47	1.217
8	0.033	0.007	-2.60	0.733	0.033	0.007	-2.55	1.161
9	0.033	0.007	-2.59	0.721	0.033	0.007	-2.54	1.197

TAB. III.18 – *Ecart des dix meilleures partitions pour le modèle Paréto + Barrière (VLS2, D_0)*FIG. III.65 – “Err” ($\gamma < 1$, VLS22, D_0) FIG. III.66 – “Err” ($\gamma > 1$, VLS2, D_0)

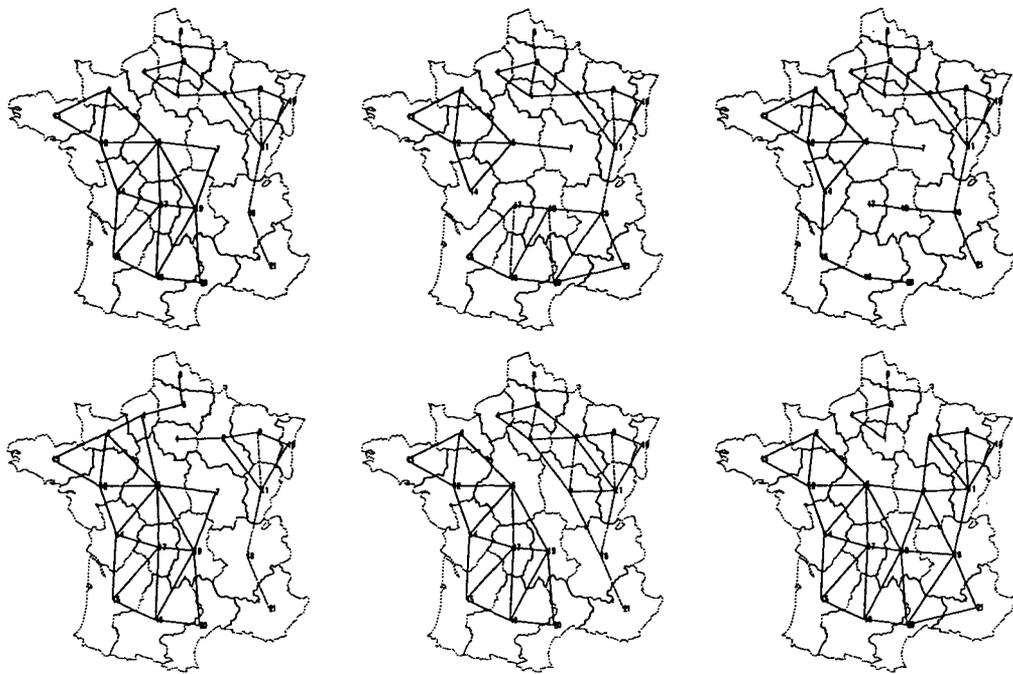


FIG. III.67 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma < 1$) appliqué à la matrice de flux VLS2 (D_0)

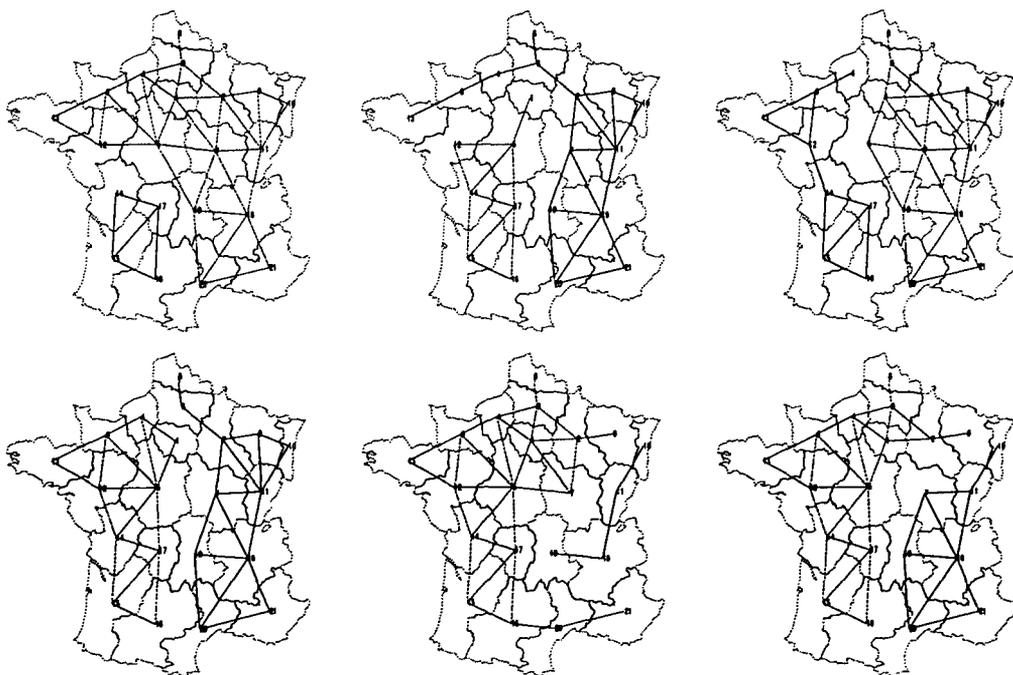


FIG. III.68 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma > 1$) appliqué à la matrice de flux VLS2 (D_0)

VLS3 Rang	$\gamma < 1$				$\gamma > 1$			
	ϵ	Err	α	γ	ϵ	Err	α	γ
0	0.118	0.023	-2.43	0.711	0.113	0.022	-2.22	1.933
1	0.119	0.023	-2.33	0.723	0.113	0.022	-2.20	1.779
2	0.120	0.023	-2.46	0.727	0.115	0.022	-2.21	1.642
3	0.120	0.023	-2.40	0.751	0.115	0.022	-2.23	1.704
4	0.122	0.024	-2.42	0.766	0.116	0.022	-2.21	1.551
5	0.122	0.024	-2.36	0.784	0.119	0.023	-2.22	1.373
6	0.122	0.024	-2.27	0.777	0.120	0.023	-2.24	1.424
7	0.123	0.024	-2.34	0.788	0.120	0.023	-2.26	1.315
8	0.124	0.024	-2.39	0.802	0.121	0.023	-2.26	1.293
9	0.124	0.024	-2.45	0.770	0.121	0.023	-2.26	1.280

TAB. III.19 – Ecart des dix meilleures partitions pour le modèle Paréto + Barrière (VLS3, D_0)

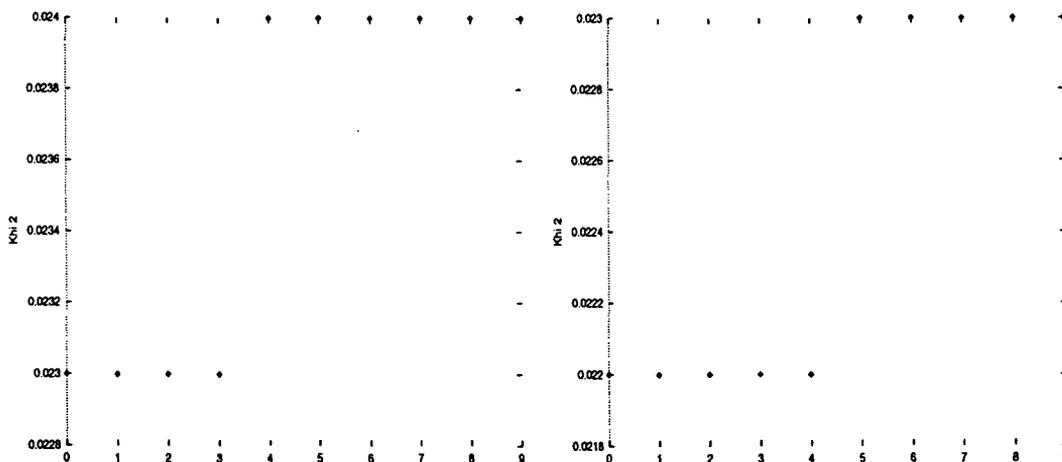


FIG. III.69 – “Err” ($\gamma < 1$, VLS32, D_0) FIG. III.70 – “Err” ($\gamma > 1$, VLS3, D_0)

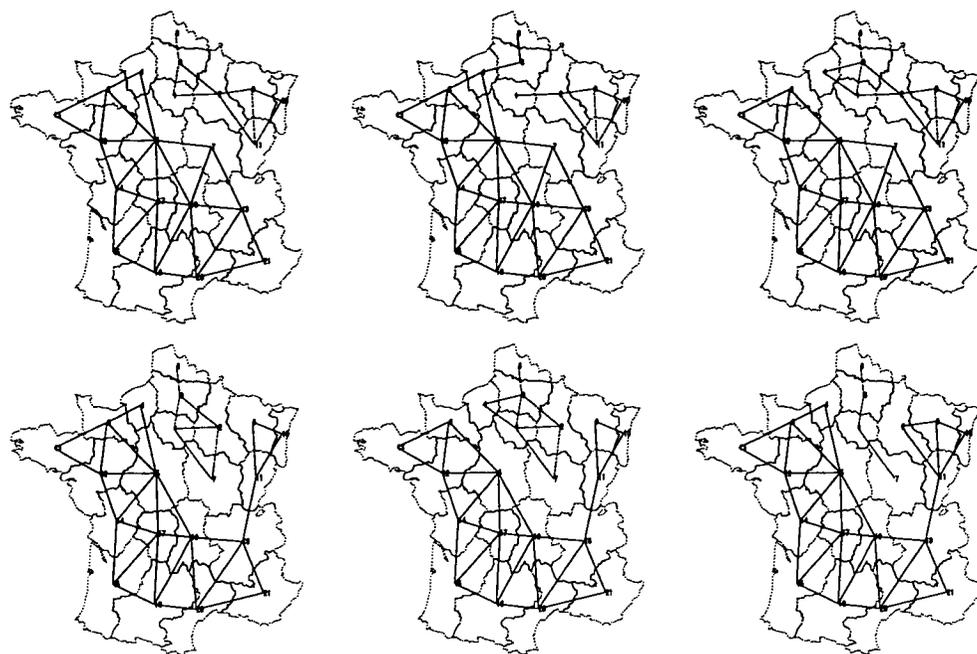


FIG. III.71 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma < 1$) appliqué à la matrice de flux VLS3 (D_0)

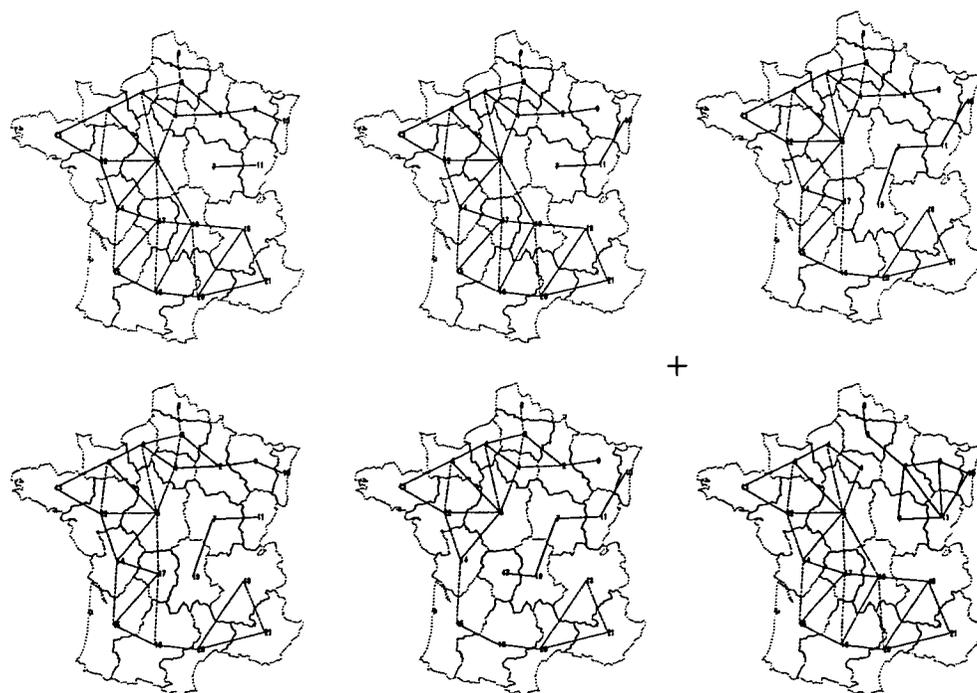


FIG. III.72 - Cartes des six meilleurs découpages de la France pour le modèle Paréto + Barrière ($\gamma > 1$) appliqué à la matrice de flux VLS3 (D_0)

Chapitre IV

Etude des paramètres de la fonction de la distance

IV.1 Paramètre de la distance: α

L'emploi du paramètre α dans les fonctions de la distance permet de modifier le rôle joué par la distance dans les modèles (cf I.1.2, p 25). Une valeur négative indiquera que la distance freine les déplacements (plus la distance augmente, et plus les flux de déplacement diminuent). Tout au contraire, une valeur positive de α montrera l'influence de la distance dans l'augmentation de ces flux (si l'on considère les flux aérien, plus les distances s'allongent, et plus l'on est tenté d'utiliser les lignes aériennes).

$$F_{ij}^* = k.e^{\alpha.d_{ij}} \text{ avec } d_{ij} \text{ distance entre les unités géographiques } i \text{ et } j$$

Fonction de la distance de type exponentiel, ou Paréto

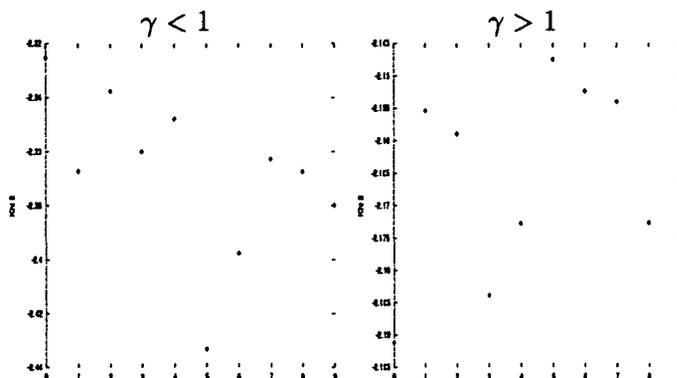


FIG. IV.1 - Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Barrière appliqué à la matrice de flux M31 (D_0)

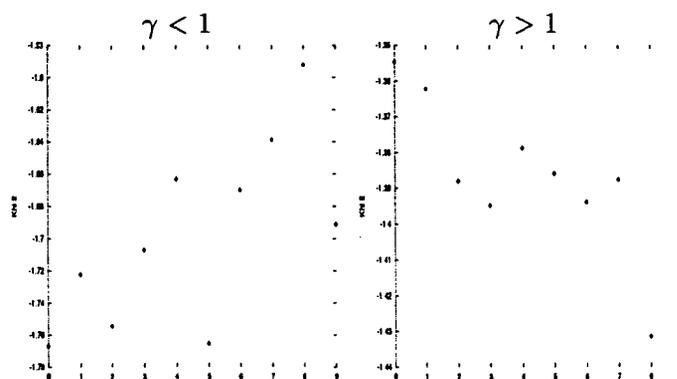


FIG. IV.2 - Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Barrière appliqué à la matrice de flux M32 (D_0)

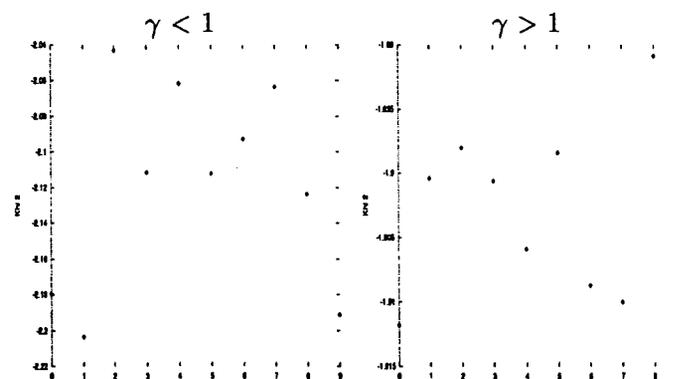


FIG. IV.3 - Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux M34 (D_0)

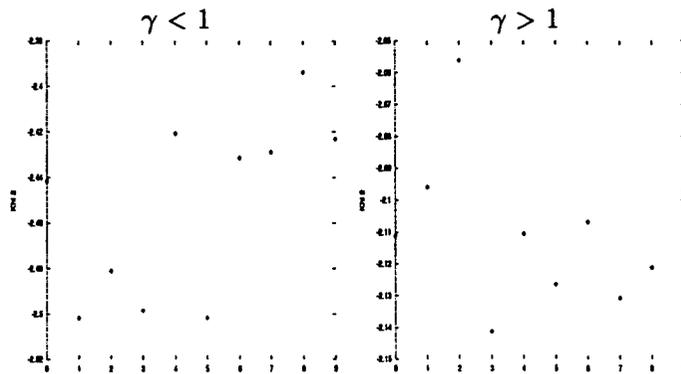


FIG. IV.4 – Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS15 (D_0)

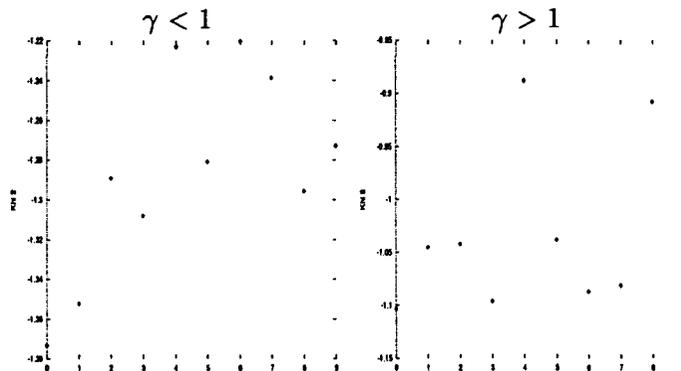


FIG. IV.5 – Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS16 (D_0)

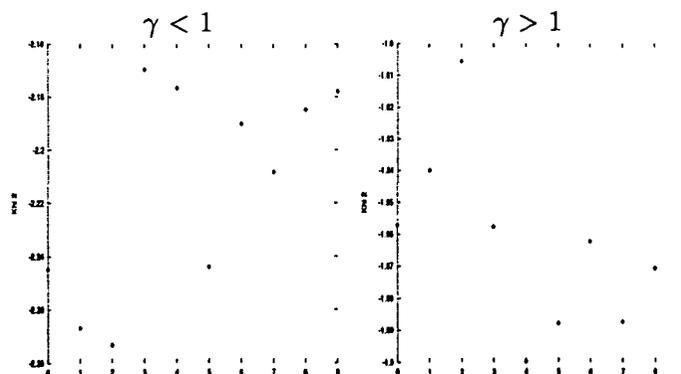


FIG. IV.6 – Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS19 (D_0)

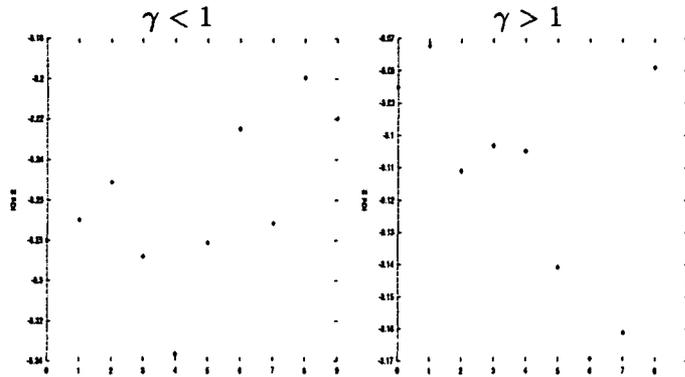


FIG. IV.7 – Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS (D_0)

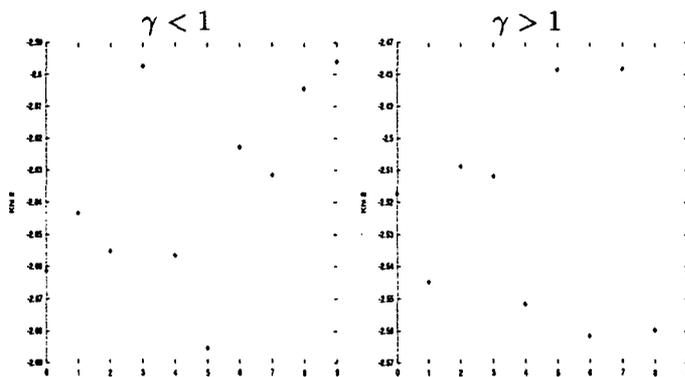


FIG. IV.8 – Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS2 (D_0)

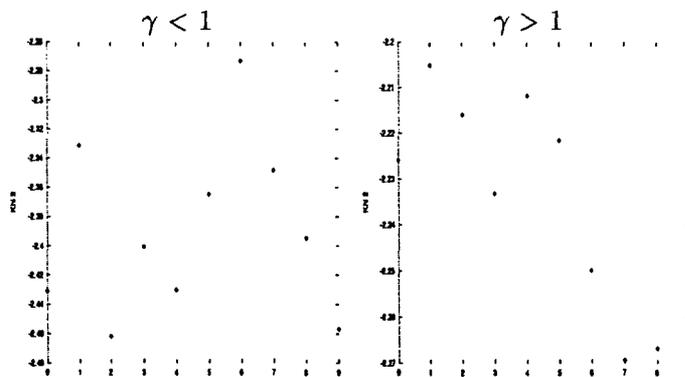


FIG. IV.9 – Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS3 (D_0)

IV.2 Paramètre de découpage: γ

γ permet de modéliser l'idée d'appartenance à un groupe (cf I.1.3, p 27). Lorsque sa valeur est inférieure à 1, ce facteur tend à minimiser les flux entre les unités géographiques étrangères. Il s'agit d'un effet de barrière. Alors qu'un γ supérieur à 1, favorisera les relations entre unités géographiques n'appartenant pas au même groupe. Ceci définit l'effet d'anti-barrière.

$$F_{ij}^* = k.f(d).\gamma^{a_{ij}} \text{ avec } a_{ij} = 0 \text{ si } i \text{ et } j \text{ appartiennent au même groupe, } 1 \text{ sinon.}$$

Forme générale du modèle multiplicatif utilisant la notion d'appartenance

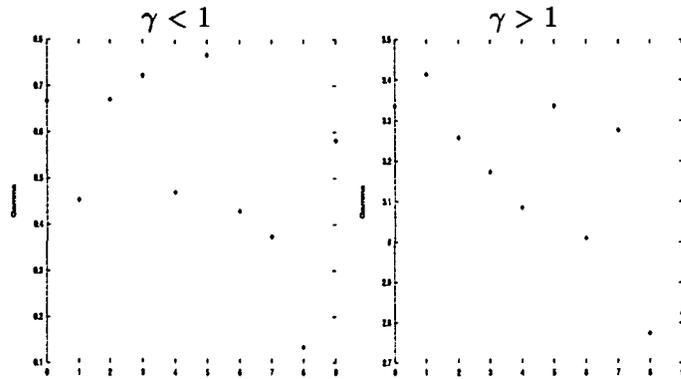


FIG. IV.10 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Barrière appliqué à la matrice de flux M31 (D_0)

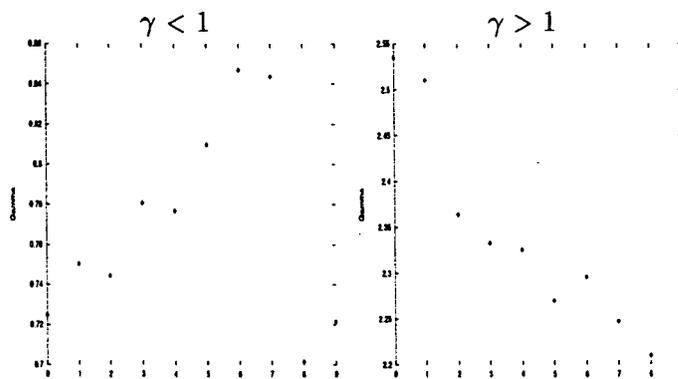


FIG. IV.11 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Barrière appliqué à la matrice de flux M32 (D_0)

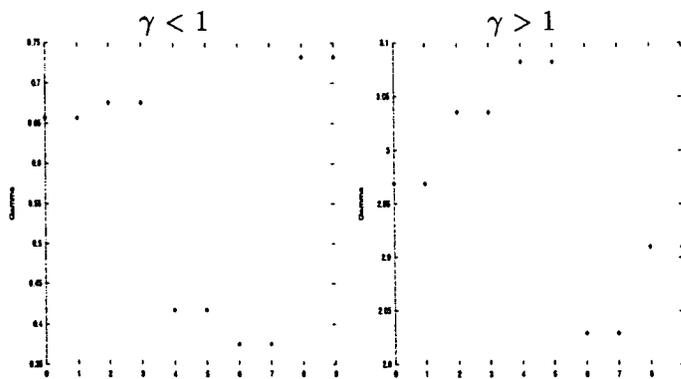


FIG. IV.12 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux M34 (D_0)

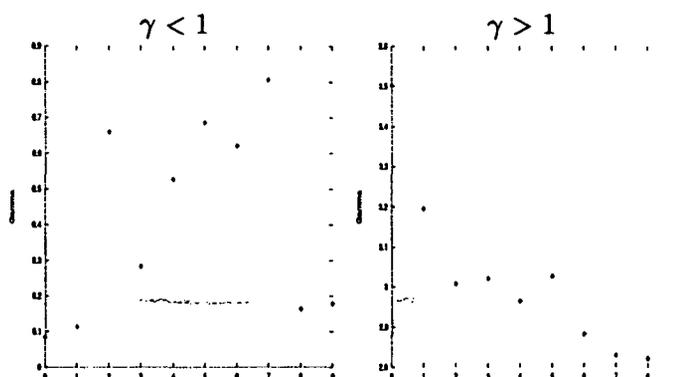


FIG. IV.13 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS15 (D_0)

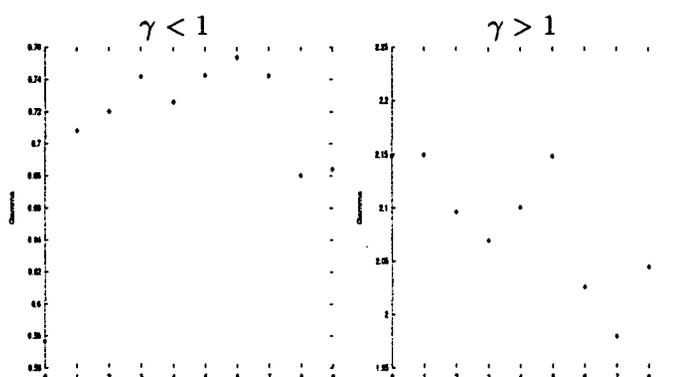


FIG. IV.14 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS16 (D_0)

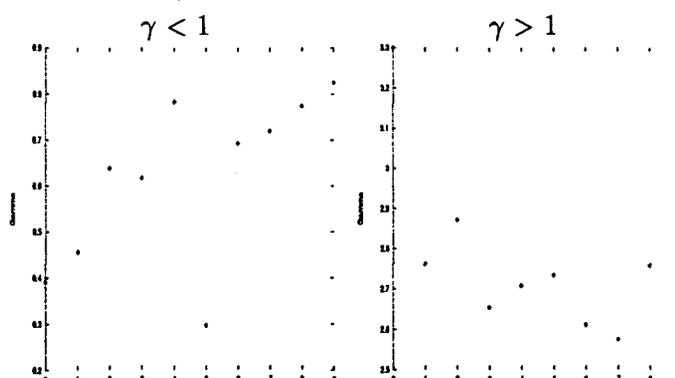


FIG. IV.15 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS19 (D_0)

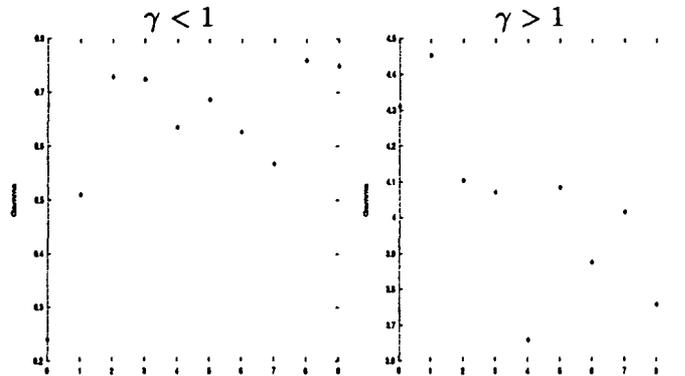


FIG. IV.16 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS (D_0)

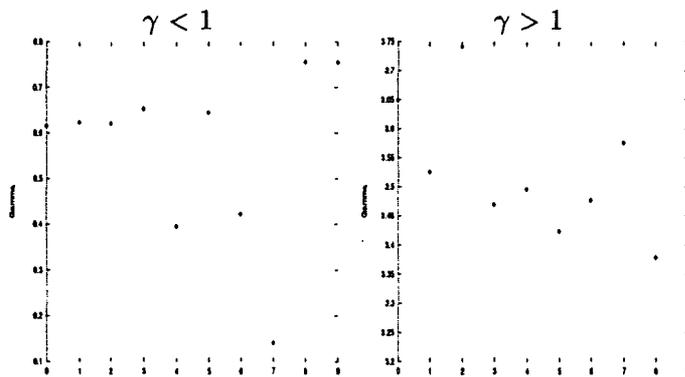


FIG. IV.17 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS2 (D_0)

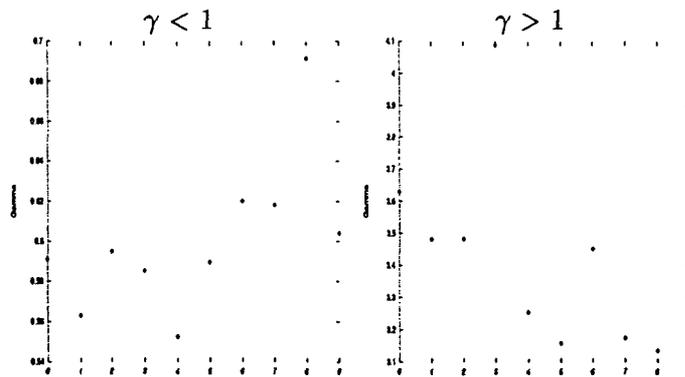


FIG. IV.18 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS3 (D_0)

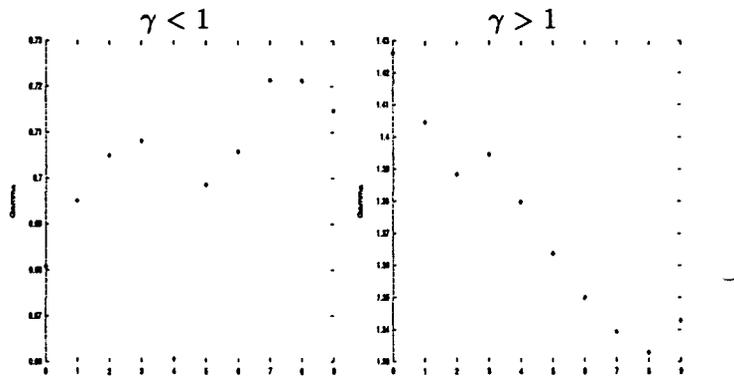


FIG. IV.19 – Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Barrière appliqué à la matrice de flux M31 (D_0)

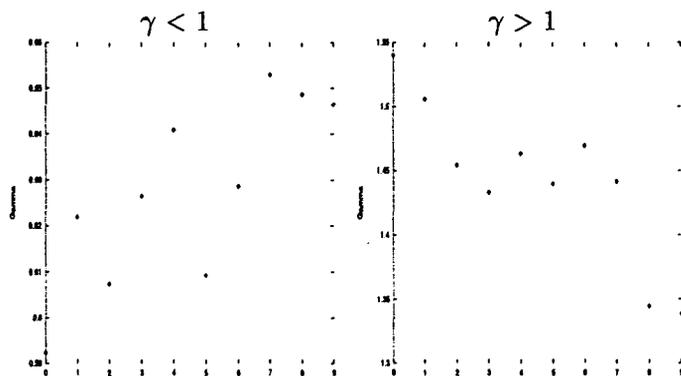


FIG. IV.20 – Valeur du paramètre α pour les dix meilleurs découpages de la France, modèle Barrière appliqué à la matrice de flux M32 (D_0)

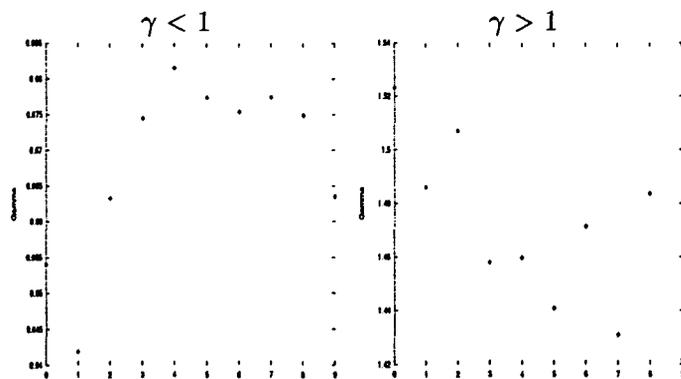


FIG. IV.21 – Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux M34 (D_0)

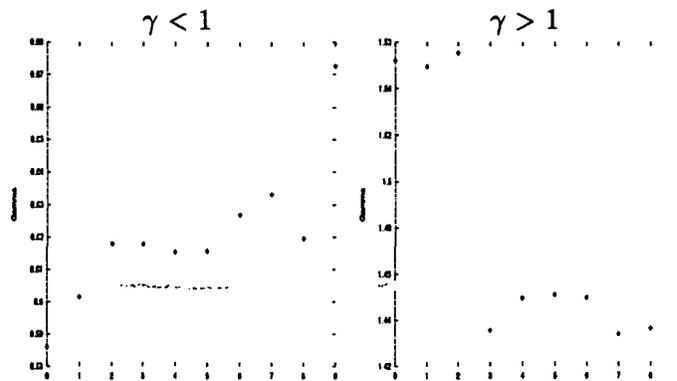


FIG. IV.22 – Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS15 (D_0)

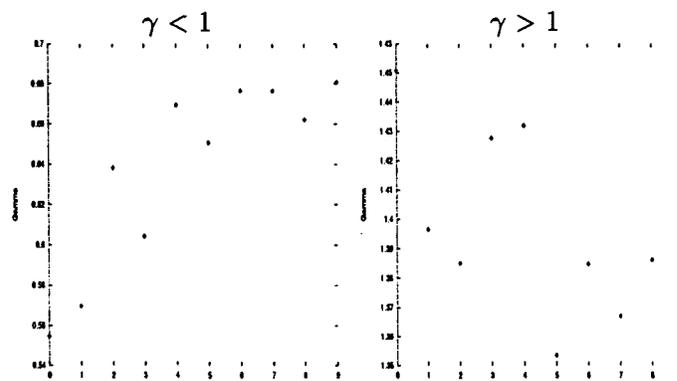


FIG. IV.23 – Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS16 (D_0)

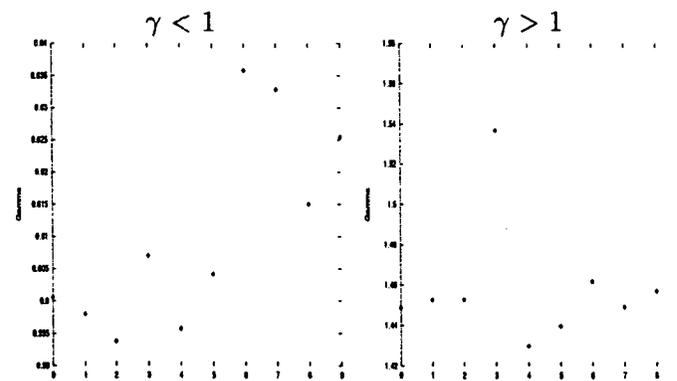


FIG. IV.24 – Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux MS19 (D_0)

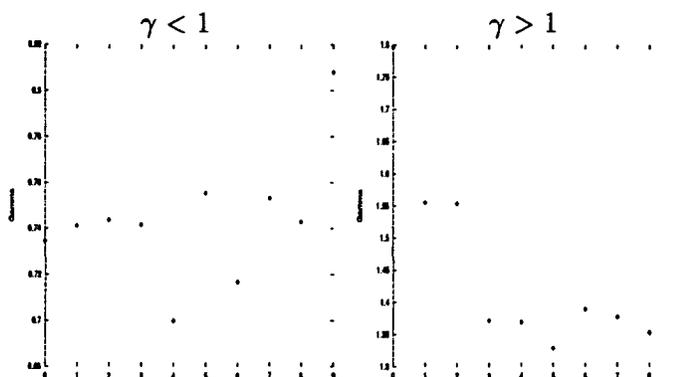


FIG. IV.25 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS (D_0)

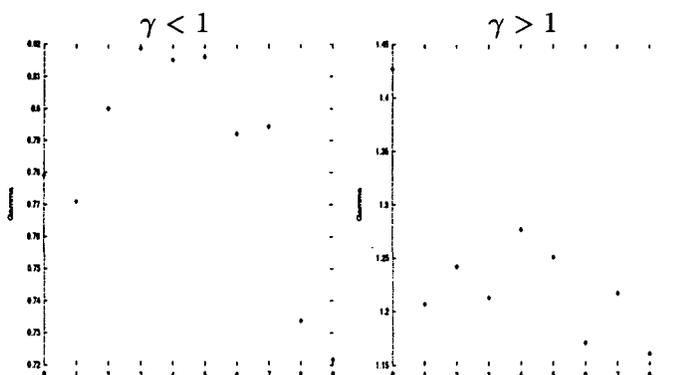


FIG. IV.26 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS2 (D_0)

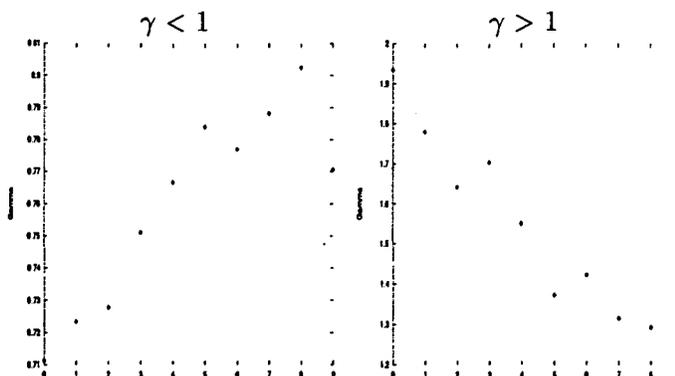


FIG. IV.27 - Valeur du paramètre γ pour les dix meilleurs découpages de la France, modèle Pare'to + Barrière appliqué à la matrice de flux VLS3 (D_0)

