

COMMISSION DES COMMUNAUTÉS EUROPÉENNES

INFORMATIONS INTERNES sur
L'AGRICULTURE

**Analyse régionale des structures
socio-économiques agricoles**

– Essai de typologie régionale pour la Communauté des Six

Partie I : Rapport

COMMISSION DES COMMUNAUTÉS EUROPÉENNES

DIRECTION GÉNÉRALE DE L'AGRICULTURE

Direction Economie Agricole – Division Bilans, Etudes, Informations Statistiques

*La reproduction, même partielle, du contenu de ce rapport est subordonnée
à la mention explicite de la source*

COMMISSION DES COMMUNAUTÉS EUROPÉENNES

INFORMATIONS INTERNES sur
L'AGRICULTURE

Analyse régionale des structures
socio-économiques agricoles

- Essai de typologie régionale pour la Communauté des Six

Partie I : Rapport

AVANT-PROPOS

Cette étude a été réalisée dans le cadre des programmes d'études de la Direction Générale de l'Agriculture et de la Direction Générale de la Politique Régionale par :

P. RAINELLI

Chargé de Recherches

INRA - RENNES

en collaboration avec

M. KERBAOL

Ingénieur - Analyste

Aux travaux ont collaboré également la Division "Objectifs et méthodes de la programmation régionale" de la Direction Générale de la Politique Régionale, les Divisions "Structures de production et environnement agricole", "Structures sociales en agriculture, problème foncier", "Bilans, études, informations statistiques", "Analyse de la situation des exploitations agricoles" de la Direction Générale de l'Agriculture et avec la participation de la Division "Etudes" de la Direction Générale des Affaires Sociales.

x

x

x

Cette étude ne reflète pas nécessairement les opinions de la Commission des Communautés Européennes dans ce domaine et n'anticipe nullement sur l'attitude future de la Commission en cette matière.

SOMMAIRE (1)

	<u>Pages</u>
<u>Introduction</u>	7
<u>CHAPITRE I</u> : Problèmes méthodologiques	10
<u>CHAPITRE II</u> : Définition de l'espace des observations et de l'espace des variables	33
Liste des symboles et abréviations utilisés dans les graphiques	57
<u>CHAPITRE III</u> : Présentation des résultats	67
<u>CHAPITRE IV</u> : Récapitulation	139

Annexes

<u>Annexe 1</u> : Présentation mathématique simplifiée de l'analyse des données	153
<u>Annexe 2</u> : Résultats des analyses préliminaires	185

<u>Annexe 3</u> : Données statistiques de base par circonscription (2)	
--	--

(1) Une table des matières figure à la fin du volume.

(2) Seront publiées ultérieurement dans la série "Informations Internes sur l'Agriculture".

INTRODUCTION : LES OBJECTIFS DE L'ETUDE

Cette étude porte sur la situation de l'agriculture au plan régional compte-tenu du contexte économique. Entreprise pour les six états membres de la Communauté Economique Européenne initiale, elle répond essentiellement à deux objectifs :

- d'une part, la caractérisation de l'agriculture selon les circonscriptions de base. Ceci doit permettre la mise en évidence de zones souffrant de déséquilibres structurels plus ou moins graves sur le plan strictement agricole.
- d'autre part, la différenciation des régions de la Communauté en fonction de leurs caractéristiques économiques générales. Cela conduit à voir comment se situent les disparités régionales au sein de l'Europe.

Ces deux objectifs étroitement interdépendants débouchent sur l'élaboration d'une politique régionale au sein de laquelle une attention particulière est réservée à l'agriculture. Mais ce n'est pas notre rôle d'ébaucher ici quelle pourrait ou devrait être une telle politique. Cette étude vise simplement à fournir un panorama de la situation des régions tant pour l'agriculture que pour le niveau général d'activité.

1. Objectif régional agricole

Le souci de voir les agriculteurs les moins avantagés bénéficier de conditions de vie meilleures conduit à considérer avec soin leurs conditions de production. En effet, en l'état actuel des choses, toute mesure susceptible d'aider les exploitants agricoles dans leur effort de rattrapage en matière

de revenu, repose sur une amélioration des structures agricoles. Or, il est incontestable qu'il existe dans ce domaine une très grande diversité de situations à travers l'Europe.

Une agriculture de type "capitaliste" faisant appel en abondance au capital et à la main d'oeuvre salariée, cotoie des formes autarciques. Dans le premier cas existent des entrepreneurs combinant des facteurs de production en fonction d'un marché, et donc très sensibles aux prix. Dans le second cas, on est en présence de personnes âgées ou d'ouvriers-paysans, obéissant à d'autres motivations que le profit, et donc peu influencées par les cours des produits.

Entre ces deux modes de production extrêmes se situe la multitude des exploitations artisanales de polyproduction. Mais les unes sont conduites par des chefs dynamiques ayant un excellent niveau technique et qui ont intensifié au maximum leur système de production, alors que d'autres sont menées par des agriculteurs pour qui la marge de progrès possible est considérable.

Une politique agricole qui se veut adaptée à toutes ces situations doit donc tenir compte de cette diversité. Les études déjà entreprises sur ce thème montrent que certains types sont en général prédominants dans des zones géographiques déterminées. Aussi peut-on estimer que l'élaboration d'une typologie des régions agricoles où sont regroupées des zones ayant des problèmes similaires, et ceci en raison de structures de production identiques, est à la base de toute définition d'une politique agricole régionale qui se veut efficace.

Tout ceci implique la nécessité de proposer une typologie allant au-delà de la simple description. Il faut une classification rendant compte non seulement de la diversité de l'agriculture, mais ayant aussi un caractère explicatif. Il ne suffit pas de constater que deux régions sont différentes, mais il importe de savoir pourquoi, afin de pouvoir agir le cas échéant.

2. Objectif de développement régional

Le troisième programme de politique économique à moyen terme de la Communauté stipule que "le développement équilibré de la Communauté exige que, à la responsabilité des Etats-Membres s'ajoute une responsabilité de la Communauté à l'égard de certains problèmes régionaux d'intérêt commun" (J.O. n° L 49 du 1er mars 1971). Indépendamment de la politique agricole commune apparaît ainsi le souci du développement régional.

Mais les zones les plus défavorisées de la Communauté sont précisément celles où l'activité agricole est dominante. D'autre part, les problèmes les plus aigus se posent là où il existe un fort excédent de population active agricole. Ceci se traduit par des structures agricoles particulièrement exigües. Aussi toute politique de développement régional suppose que l'on accorde une attention spéciale aux problèmes posés par la reconversion des agriculteurs (mutations professionnelles et mutations géographiques).

Une approche des problèmes de développement régional suppose donc une évaluation satisfaisante de l'état d'avancement socio-économique des régions. Ceci peut s'envisager en se référant aux théories usuelles qui appréhendent le niveau de développement d'une zone à partir du poids respectif des principales activités classées selon la nomenclature habituelle en trois secteurs. Il n'en résulte pas obligatoirement, bien sûr, qu'une zone à dominante agricole est défavorisée. Tel est le cas de certaines circonscriptions, à peuplement très lâche, où des exploitations de grande étendue bénéficient de conditions suffisamment favorables pour avoir des revenus élevés.

Par ailleurs, toujours dans cette optique, il importe de caractériser la population agricole sur le plan démographique. Une densité agricole relativement forte peut avoir des conséquences différentes si dans un cas la pyramide des chefs d'exploitation indique une forte proportion d'exploitants âgés, alors que dans l'autre cas, la pression est due surtout aux jeunes. Dans la première éventualité, la situation n'est

pas bloquée. A l'aide de mesures de type Indemnité Viagère de Départ, on peut espérer une amélioration à moyen terme. Dans la seconde éventualité le déséquilibre est beaucoup plus grave, et l'on doit chercher des solutions hors de l'agriculture.

Ainsi les deux objectifs de l'étude sont étroitement complémentaires puisque :

- le développement régional des zones les plus défavorisées passe en premier lieu par des mesures touchant l'activité agricole et les agriculteurs,
- toute possibilité d'évolution de l'agriculture est liée au développement de l'économie non agricole de la région.

La typologie des régions agricoles de la Communauté en fonction des problèmes économiques généraux des régions suppose donc une démarche à deux niveaux :

1er niveau : l'agriculture, avec une définition des structures de production tenant compte des problèmes spécifiques aux exploitations des régions, et des possibilités d'évolution des structures.

2è niveau : l'activité économique générale, avec l'accent mis sur l'emploi et la population afin de voir dans quelle mesure les mutations professionnelles sont envisageables pour les agriculteurs en surnombre.



CHAPITRE I - PROBLEMES METHODOLOGIQUES

La méthodologie ne pouvant jamais, selon l'expression de Max WEBER (1), "être autre chose qu'une réflexion sur les moyens qui se sont vérifiés dans la pratique", la manière de résoudre le problème posé est étroitement liée à l'état de la statistique. Toute notre démarche est donc très dépendante des données disponibles. Mais bien sûr, le choix de la méthode

(1) WEBER (Max), Essais sur la théorie de la science, 1922, traduction française, Paris, Plon 1965, p. 220.

à utiliser est fondamental. Partant de la démarche classique, nous en montrerons les limites, puis nous examinerons les possibilités offertes par les développements récents de la statistique dans le domaine de l'analyse des données. Enfin, en troisième lieu, nous présenterons de façon simplifiée sur quelles bases mathématiques repose l'analyse des données.

A - Approche classique

Désirant élaborer une typologie répondant aux objectifs précédemment formulés, nous nous trouvons confrontés, dans une approche classique, à deux sortes de difficultés : quels critères retenir en vue de différencier de façon significative des régions que l'on désire regrouper entre elles ? et corollaire : comment fixer les classes à l'intérieur des critères retenus ?

Si l'on retient comme objectif général de toute politique agricole, l'atténuation des disparités de niveau de vie entre agriculteurs et autres catégories socio-professionnelles, on va centrer son attention sur le revenu, celui-ci étant supposé fournir la meilleure indication en ce domaine.

On cherchera donc à apprécier les disparités entre le revenu agricole d'une région et le revenu des autres secteurs de cette même région. Puis on verra comment se situe le revenu agricole de la région par rapport au revenu moyen de la communauté pour les activités non agricoles ; puis le revenu agricole d'une région par rapport au revenu agricole moyen communautaire.

Enfin, il pourra s'avérer nécessaire de faire intervenir des critères de densité de la région par rapport à la densité communautaire, pour l'ensemble des activités, ainsi que pour l'agriculture uniquement.

En ce qui concerne les classes se pose la question du nombre et des seuils. Doit-on retenir pour chaque critère deux possibilités (bon ou mauvais) ou en faut-il trois (bon, mauvais et moyen). Quel que soit le choix, reste à choisir

le seuil. Ainsi pour définir le niveau de vie des diverses catégories socio-professionnelles, on se réfère à la notion de revenu, plus précisément dans notre cas, compte tenu des données disponibles, de valeur ajoutée. De ce fait, il n'est pas évident que la parité de niveau de vie entre agriculteurs et autres activités corresponde à une valeur ajoutée par tête équivalente. Si l'on retient trois classes pour chaque critère, il est délicat de déterminer le stade où une situation cesse d'être moyenne pour devenir bonne ou mauvaise. Il apparaît donc nécessaire d'étudier au préalable la distribution du caractère. Ceci conduit à des classes différentes selon les variables.

La démarche classique consiste donc à définir un nombre aussi restreint que possible d'indices, qui par voie de conséquence doivent être hautement significatifs. Des seuils sont ensuite déterminés, de façon approximative. Afin d'avoir un nombre de groupes suffisamment restreint pour que la typologie ait un caractère opérationnel, on est contraint à réduire au maximum la liste des variables et le nombre de classes. En effet, pour n indices, chacun subdivisé en p classes, on se trouve face à un nombre de possibilités théoriques égal à p^n .

C'est-à-dire que pour 5 critères et 2 possibilités (situation bonne ou mauvaise), on aboutit à 25 types. Si toujours pour 5 critères, on envisage 3 possibilités (situation bonne, moyenne ou mauvaise) on se trouve, en théorie, face à 125 cas distincts. Ce qui signifie l'élaboration d'un outil quasiment inutilisable. En réalité, on ne trouverait pas 125 types de régions, car certaines classes seraient vides. Mais même si seulement la moitié des cas correspondent à des situations existantes, la typologie reste lourde à manier.

La limitation obligatoire dans le choix des critères et des classes à l'intérieur de ceux-ci, restreint les possibilités d'analyse socio-économique. C'est d'ailleurs la raison pour laquelle dans ce type d'approche, le travail doit être complété par des analyses ultérieures faisant intervenir d'autres critères que l'on considère a priori comme moins importants.

Pour tourner cette limitation qui conduit à sélectionner des aspects prioritaires et des aspects secondaires, on peut envisager une approche moins systématique. Celle-ci consiste à comparer les régions deux à deux en s'attachant aux informations essentielles et à regrouper les circonscriptions a posteriori sans disposer d'une grille préalable. On définit plus ou moins empiriquement une ressemblance entre les régions que l'on regroupe ensuite. Si l'on désire une typologie conduisant à un nombre restreint de groupes, la notion de ressemblance sera plus large que si l'on souhaite un petit nombre de possibilités de classement.

Notons que c'est ainsi qu'a procédé J. KLATZMANN dans son travail au niveau des petites régions naturelles françaises en vue d'étudier les phénomènes de localisation de la production (1). Pour opérer plus rationnellement et simplifier sa tâche, il a établi des typogrammes qui permettent de réduire à un espace à 2 dimensions, des phénomènes exprimés dans un espace à 3 dimensions.

Toujours dans le même ordre d'idée, signalons la méthode employée par M. BERTIN, directeur du Laboratoire de cartographie de l'Ecole Pratique des Hautes Etudes à Paris. Le procédé qu'il a imaginé, consiste à représenter visuellement sur un tableau à double entrée, l'ensemble des données pour les zones étudiées. Chaque circonscription est décrite à l'aide de variables subdivisées en 4 ou 5 catégories. Celles-ci apparaissent grâce à un système de couleurs ou par l'intermédiaire de procédés divers. Au moyen de matrices construites par l'auteur de la méthode, il est possible d'arriver par essais successifs, à regrouper les circonscriptions de base présentant des caractéristiques voisines.

Cette dernière méthode pour intéressante qu'elle soit, constitue en fait une systématisation du procédé artisanal. Elle suppose aussi que le problème des seuils définissant les classes soit réglé correctement.

(1) KLATZMANN (Joseph) La localisation des cultures et des productions animales en France, Paris, Imprimerie Nationale, 1955.

L'existence des ordinateurs, permet aujourd'hui de faire appel à des techniques beaucoup plus raffinées, relevant essentiellement de l'analyse des données.

B - L'analyse des données

Avant d'exposer les principes généraux de l'analyse des données puis dans la section C les bases mathématiques sur lesquelles repose cette branche de la statistique, nous voudrions présenter un travail de Richard STONE.

§1. L'utilisation du concept de distance par STONE

Cherchant une méthode permettant de comparer la structure économique des régions, Richard STONE se réfère au concept mathématique de distance (1). Ce travail est intéressant en ce qu'il présente un caractère précurseur par rapport à notre essai de typologie, et parce qu'il dégager les grandes lignes de l'analyse des données telle qu'elle est pratiquée aujourd'hui.

L'auteur définit chacune des 12 régions du Royaume-Uni à l'aide des agrégats économiques tels qu'ils apparaissent dans les comptes régionaux de 1948 (compte de production, de consommation et de capital). Ces agrégats homogènes par rapport à un système de mesure, puisqu'il s'agit de termes monétaires, sont au nombre de 14. Mais ils se réduisent à 11 si l'on tient compte des seules variables indépendantes (non liées comptablement).

STONE va tenter de regrouper les régions en classes de manière à ce que les unités à l'intérieur d'un groupe se ressemblent plus que les unités appartenant à des classes différentes. Pour cela il a recours au concept classique de distance euclidienne.

si l'on a n régions	1, 2, ..., j, k, ... n
et m agrégats indépendants	1, 2, ..., r, s ... m

(1) STONE (Richard) A comparison of the economic structure of regions based on the concept of "distance" in Mathematics in the Social Sciences and other essays. Ed. Chapman and Hall, 1966, p. 152-171.

x_{rj} désigne l'agrégat r dans la région j et la distance entre les régions j et k s'écrit :

$$d_{jk}^2 = \sum_r (x_{rj} - x_{rk})^2 = (x_j - x_k)' (x_j - x_k)$$

avec ()' représentant la matrice transposée.

Mais si une région est très importante sur le plan économique, elle risque de se situer à une grande "distance" des autres régions, même si elle leur ressemble. Cela en raison du type de distance utilisé. Pour éviter cela, STONE envisage trois possibilités.

Dans l'analyse A, toutes les variables retenues sont divisées par la population de chaque région. Mais les régions riches qui ont un produit par tête élevé s'éloigneront des régions pauvres. Si l'on souhaite éviter que la distinction se fasse sur ce critère, il vaut mieux diviser les agrégats régionaux plutôt par le montant du produit régional. On pourrait même envisager de rapporter les agrégats à des valeurs différentes, ce qui pourrait conduire à des variables du type des coefficients des tableaux d'échanges inter-industriels.

Pour éviter que la distance entre deux régions soit affectée par la taille des agrégats, STONE propose d'adopter une mesure "normalisée". En effet celle-ci est souhaitable car le produit par tête par exemple est une grandeur susceptible de varier considérablement d'une région à l'autre, tandis qu'en comparaison l'épargne par tête est moins élevée, et moins fluctuante. Le choix d'une mesure "normalisée" évite qu'un agrégat ait plus de poids qu'un autre dans l'analyse.

Dans l'analyse B on pose que la somme des carrés des agrégats par tête à l'intérieur d'une région équivaut à 1.

Si x_{rj}^* représente l'agrégat n ramené au nombre d'habitants dans la région j , on a :

$$\sum (x_{rj}^*)^2 = 1$$

et en mesure normalisée :

$$x_{rj}^{**} = \frac{x_{rj}^*}{\sum_j (x_{rj}^*)^2}$$

Enfin dans l'analyse C, on se ramène par rapport aux autres analyses à une base orthogonale en supposant que les variables économiques sont statistiquement indépendantes.

On a la transformation linéaire $x_j^{**} = A z_j$

où A est une matrice non singulière d'ordre m.

Si l'on calcule la distance en termes de z, on a :

$$d_{jk}^2 = (x_j^{**} - x_k^{**})' R^{-1} (x_j^{**} - x_k^{**}) \quad \text{où } R^{-1} = (AA')^{-1}$$

R est une matrice d'ordre m des corrélations entre agrégats.

Cette expression, semblable à la distance généralisée de MAHALANOBIS, indique que le carré de la distance entre deux régions j et k est une forme quadratique des différences des valeurs observées des agrégats "normalisés". Alors que dans l'analyse B, la normalisation se fait par rapport à l'ensemble des agrégats d'une région, ici la normalisation se fait par rapport à l'ensemble des régions.

Pour classer les régions, on calcule la matrice des corrélations entre agrégats (R) et les $\frac{n}{2}(n-1)$ distances entre les n régions. Puis on regroupe les régions en classes de façon que la valeur moyenne des d^2 à l'intérieur d'une classe soit inférieure à la valeur moyenne des d^2 entre classes.

On peut noter que la matrice de la formule $x_j^{**} = A z_j$ est semblable à celle de l'analyse factorielle en composantes principales et où z_j peut être interprété comme un facteur orthogonal hypothétique des m agrégats pouvant permettre de décrire les agrégats réels.

Ainsi à travers la démarche de STONE apparaissent les bases de l'analyse des données avec référence à la classification automatique et à l'analyse factorielle. Le détail des difficultés lié à un type de distances nous éclaire sur l'importance de ce problème et met en évidence l'intérêt de l'analyse des correspondances où la question est résolue de façon élégante.

§2. Principes généraux de l'analyse des données

Le terme d'analyse des données recouvre cette partie de la statistique qui s'appuyant sur un outil mathématique purement algébrique, cherche à décrire, réduire et classer des observations multidimensionnelles. Cette branche de la statistique se refuse à faire aucune hypothèse sur les phénomènes observés contrairement aux écoles traditionnelles qui visent à formaliser l'induction.

Notons à la suite de MORLAT (1) que parmi les praticiens de la statistique mathématique ne se référant pas à l'analyse des données, on peut distinguer les "anciens" et les "classiques" :

- " les anciens", ne pouvaient accepter l'idée d'une inférence qui ne fût point basée sur des connaissances a priori exprimables sous forme de distributions de probabilité concernant les paramètres des lois des phénomènes étudiés, et sur l'emploi de la formule de BAYES,
- "les classiques" adoptaient les vues des statisticiens anglosaxons (R.A. FISHER, J. NEYMAN, A. WALD) et considéraient le premier devoir du statisticien d'établir des méthodes d'inférence qui ne laissent point de place aux idées a priori, sauf pour ce qui regarde la forme analytique des lois de probabilité utilisées.

En fin de compte, si l'on reprend la conclusion de MORLAT "l'analyse des données n'est pas autre chose que la forme moderne de la statistique descriptive".

(1) MORLAT (G.), in ouvrage collectif CAILLIEZ, MAILLES, NAKACHE, PAGES, Analyse des données multidimensionnelles. Centre d'Etudes Economiques d'Entreprises, Conservatoire National des Arts et Métiers, Paris, 1971, tome 1, avant-propos, p. 11.

Il importe de souligner que les développements récents de l'analyse des données sont directement liés à l'avènement des ordinateurs qui permettent de traiter des masses considérables de chiffres. Et c'est en cela que les progrès conjugués de la statistique mathématique et des moyens de calcul mettent aujourd'hui à notre disposition des outils remarquables en vue de résoudre le type de problème qui nous est posé.

En effet, on a vu que pouvant difficilement confronter pour chaque région (surtout si leur nombre est de l'ordre de deux cents), l'ensemble des variables qui caractérisent une zone, on était réduit à limiter le nombre d'indices. D'où un appauvrissement de l'analyse, que l'on essayait de combler ultérieurement par l'étude d'aspects jugés secondaires a priori.

Il est très gênant sur un plan scientifique d'éliminer certaines variables sans être sûr que celles-ci ne jouent aucun rôle. L'idéal est donc de prendre en compte toutes les variables susceptibles d'apporter une information intéressante. Et c'est là que le recours à l'analyse des données et à l'ordinateur s'avère irremplaçable, puisque cette démarche autorise le traitement simultané d'un grand nombre d'informations.

C'est ce qu'exprime J. P. BENZECRI lorsqu'il déclare (1) que "si la notion de cause nous paraît incertaine, on ne peut la rejeter ; et c'est justement à l'observation qu'il faut recourir pour la préciser et la fonder aussi honnêtement que possible. Encore faut-il que le champ d'observation soit assez ample". D'où le principe selon lequel il convient de traiter simultanément des informations concernant le plus grand nombre possible de dimensions.

"Du coup, précise BENZECRI, le problème de la validité (du "test") troublant parfois, avouons-le, passe au second plan. Nul ne sait si l'inégalité $0,5 \neq 0,7$ peut dans tel cas pratique être considérée comme un résultat empirique sûr, ou si ce n'est qu'un jeu de hasard. Tandis que de trouver dans un espace de

(1) BENZECRI (J. P.), Les principes de l'analyse des données. Publication ronéotée du Laboratoire de Statistique Mathématique des Universités de Paris, p. 8.

dimension deux, 50 points approximativement rangés sur un cercle est sûrement une découverte (à moins que la méthode de calcul ne soit une duperie)".

L'analyse des données permet ainsi de résoudre simultanément le double problème du nombre de variables à retenir, et de la fixation des classes à l'intérieur de celles-ci.

C - Les bases de l'analyse des données (1)

Supposons que l'on ait rassemblé pour chacune des circonscriptions un ensemble de données permettant de caractériser de façon valable la situation tant sur le plan économique général, que sur le plan agricole. Admettons que l'on dispose en tout d'une cinquantaine de variables.

L'ensemble des données peut être rassemblé dans un grand tableau rectangulaire où l'on trouve en lignes les 200 régions environ de la C.E.E. et en colonne les 50 variables retenues. Chaque région est donc répertoriée dans un espace à 50 dimensions.

On est en présence d'un ensemble I de régions et d'un ensemble J de variables caractérisant ces régions. Le problème de typologie revient à chercher une représentation simplifiée de ces deux ensembles.

Schématiquement, on peut affirmer qu'il y a deux façons de traiter ce tableau rectangulaire de I x J nombres $k(i, j)$ afin d'en obtenir une représentation simplifiée : soit à l'aide d'un procédé qualifié d'ascendant, soit par un procédé dit descendant. Dans le premier cas, il s'agit de réunir les éléments les plus proches au moyen de systèmes de classification automatique semblables à ceux utilisés par les taxinomistes. Dans le deuxième cas, on procède par réduction de l'ensemble en un nombre aussi faible que possible de sous-ensembles

(1) Pour plus de détails sur le plan méthodologique, cf. annexe "Présentation mathématique simplifiée de l'analyse des données".

homogènes ; c'est ce que permet l'analyse factorielle (1). Ces deux points seront examinés successivement, et complétés par la présentation de l'analyse des correspondances. Une annexe à caractère plus mathématique précise certains points.

§1. La classification automatique

Sous ce terme très général, on regroupe les méthodes et algorithmes qui donnent une partition d'un ensemble E en classes C telles que le "degré de ressemblance" des éléments de C_i est plus élevé que le "degré de ressemblance" entre éléments de classes différentes.

Dans la plupart des procédés, le point de départ consiste en une mesure de proximité sur l'ensemble E. La proximité entre éléments est définie à partir de la notion de distance (euclidienne ou non), ou à partir d'indices de similarité, d'indices de dissimilarité ou d'indices de distance.

Ainsi x et y se ressemblent plus que z et t s'écrit :

$d(x, y) < d(z, t)$ si d est un indice de dissimilarité.

$d(x, y) > s(z, t)$ si s est un indice de similarité.

Pratiquement le nombre d'indices de proximité que l'on peut construire est très élevé, ce qui rend le choix difficile.

En pratique, au lieu d'avoir recours à des procédés ne fournissant qu'une partition, on fait appel à des procédés conduisant à une hiérarchie de partitions. Ces méthodes, bien que plus coûteuses en temps de calcul ont l'avantage de laisser plus de souplesse à l'interprétation. Celle-ci est facilitée par la visualisation à l'aide d'arbres de classification, appelés aussi dendrogrammes.

Au-delà de la classification proprement dite, on est amené dans un deuxième stade à décrire la "cohésion" des classes, ainsi que la proximité entre classes. La plus ou moins grande homogénéité des données se manifeste ainsi.

(1) Pour être plus exact, il convient de préciser que certaines méthodes de classification automatique ont un caractère descendant.

Par ailleurs, on distingue les procédés donnant des classes monothétiques des procédés donnant des classes polythétiques.

Dans le premier cas, on dit que la classe C_i est monothétique relativement au caractère j si tous les éléments de C_i prennent la même valeur pour ce caractère. L'intérêt de ces classifications se manifeste lorsqu'on doit affecter a posteriori un élément dans une classe, car on a alors des clefs d'identification.

Dans le second cas, une classe est polythétique si les éléments qui y figurent se ressemblent relativement à l'ensemble des caractères. On considère la ressemblance des éléments pour la classe prise globalement. Ceci a pour conséquence qu'un caractère donné ne prend pas obligatoirement la même valeur chez tous les individus que contient la classe.

Cette méthode, assez prisée des taxinomistes est peu utilisée par les économistes. Elle n'a fait l'objet, à notre connaissance, que d'un très petit nombre d'applications dans les sciences humaines. Citons quelques travaux concernant la construction de nomenclatures d'activité (1) et surtout un travail en cours de J. KLATZMANN, plus proche de nos préoccupations, car concernant la typologie des régions agricoles françaises (2).

Ce procédé ne nous paraît pas être le plus judicieux en raison de l'importance des problèmes soulevés. En effet, doit-on choisir le diamètre des classes, la distance moyenne entre classes, ou bien le saut minimum entre classes en se basant sur la distance entre les deux éléments de chaque classe

(1) cf. BENIER, DEHEDIN, GIUDICELLI, PIETRI, "La nomenclature des activités du secteur : travail mécanique du bois. Publication ronéotée ISUP 1969 - MASSON (M.), Construction d'une nomenclature économique d'activités, Thèse 3^e cycle, Faculté des Sciences de l'Université de Paris, ISUP, 1970 - GROS MANGIN (A.), Construction d'une nomenclature économique d'activités. Thèse 3^e cycle, Faculté des Sciences de l'Université de Paris, ISUP, 1970.

(2) KLATZMANN (Joseph), La typologie des régions agricoles françaises. Problèmes méthodologiques. Communication au congrès de l'Union de Géographie Internationale. Hamilton (Canada) 1972, Texte aimablement communiqué par l'auteur.

les plus proches ? Autant de questions faisant l'objet de réponses divergentes selon les auteurs, puisqu'il existe 8 ou 9 programmes de classification automatique, dont aucun n'est vraiment plus satisfaisant que les autres. D'autre part, la hiérarchie construite se fait pas à pas sans que l'on puisse revenir en arrière pour modifier une branche déjà adoptée. Le choix d'une mesure de proximité, fonction du problème traité, introduit également un certain arbitraire.

Sur un plan pratique, ces inconvénients se traduisent de plusieurs manières. J. KLATZMANN dans l'article précité les expose ainsi : "Si l'on compare des critères de même nature, la différence entre deux nombres quelconques... a une signification parfaitement claire. Mais que faire s'il faut comparer une différence mesurée en termes monétaires à une différence entre deux pourcentages"...

"La seconde remarque concerne l'importance relative à donner aux divers caractères pris en considération. Si l'on répartit le territoire en vingt natures de cultures, et le bétail en quatre catégories seulement, il y a bien des chances pour que la classification quelle que soit la méthode utilisée soit plus influencée par l'utilisation des terres que par l'importance du bétail".

Ces difficultés propres aux méthodes traditionnelles de classification automatique n'apparaissent pas dans la méthode des nuées dynamiques de DIDAY brièvement exposée en annexe. En fait, il n'est pas sûr que les nuées dynamiques puissent être considérées comme relevant des techniques de classification automatique.

Reste enfin un dernier reproche que l'on peut formuler ainsi. Les techniques de classification aboutissent à des regroupements de l'ensemble des individus sans que les caractères jouent un rôle particulier, en vue de l'interprétation. Et réciproquement on a une agrégation des variables qui se manifeste à partir des arborescences, indépendamment des observations qu'elles caractérisent. L'aspect descriptif prédomine au détriment

de l'aspect explicatif. Ceci explique pourquoi la plupart des auteurs préconisent l'utilisation de la classification automatique en vue de compléter les résultats obtenus avec l'analyse factorielle.

§2. L'analyse factorielle

Historiquement (1), l'analyse factorielle est apparue avec SPEARMAN en 1904. Face à une population de sujets à classer en fonction de leur degré de réussite à des épreuves psychométriques, cet auteur a proposé de faire appel au modèle mathématique suivant :

$$z_{ij} = a_j F_i + e_{ij} \quad \left\{ \begin{array}{l} i = 1, \dots, m \\ j = 1, \dots, n \end{array} \right.$$

avec z_{ij} désignant la réussite du sujet i au test j .

où F_i est une variable hypothétique, un "facteur" qui explique la réussite au test j , et qui donc est commun à l'ensemble des épreuves.

a_j mesure l'importance de l'aptitude dans la réussite au test j ,

e_{ij} mesure l'aptitude de l'individu i par rapport à des facteurs de réussite propres au test j ; les e_{ij} sont spécifiques pour chacun des tests.

La résolution de ce modèle permet l'estimation des paramètres a_j . Ceux-ci sont intéressants car ils permettent de voir comment les tests sont liés à la variable hypothétique F , supposée par SPEARMAN correspondre à la notion d'intelligence. De même l'estimation des F_i permet de voir comment les individus sont liés au facteur F . D'où la possibilité d'une classification des tests par rapport à la variable supposée être l'intelligence ainsi qu'une classification des individus par rapport à cette même variable.

Mais l'utilisation de ce modèle, dit unifactoriel, puisque se référant à un seul facteur (facteur commun unique),

(1) BOUDON (Raymond), L'analyse mathématique des faits sociaux, Paris, Plon, 1967, p. 206 et suivantes.

a très vite révélé son insuffisance. C'est à THURSTONE que l'on doit l'élaboration d'une méthode expliquant les corrélations entre les divers tests psychologiques utilisés dans des épreuves et plusieurs facteurs communs.

THURSTONE s'en explique ainsi (1) "Le problème central auquel SPEARMAN consacra principalement son attention en analyse factorielle fut celui-ci : les coefficients d'une table expérimentale donnée de corrélations sont-ils en accord avec l'hypothèse d'un facteur général unique ?... Il semble étonnant que le problème de SPEARMAN n'ait jamais été formulé en fonction du rang de la matrice des corrélations, bien que SPEARMAN ait eu la collaboration de mathématiciens éminents.

En 1930, nous avons posé la question différemment : combien de facteurs sont nécessaires pour rendre compte des corrélations ? Nous avons appelé ceci analyse multifactorielle. La question de savoir si, dans un cas particulier quelconque, un seul facteur suffirait pour rendre compte des inter-corrélations, devenait alors une question de fait expérimental".

Notons que la solution rigoureuse sur le plan mathématique, du modèle multifactoriel orthogonal est due à HOTELLING en 1933.

La méthode directement issue de ce courant est connue sous le nom général d'analyse en composantes principales, et correspond à la notion classique d'analyse factorielle.

En raisonnant en termes géométriques, on peut assez simplement indiquer comment l'analyse factorielle s'avère un excellent outil dans l'élaboration d'une typologie.

Notre tableau des 200 régions définies à l'aide de 50 variables correspond à un nuage de points dans un espace à 50 dimensions. Dans son principe l'analyse factorielle permet de déterminer un nombre d'axes relativement restreint, les facteurs, tels que l'on puisse reconstituer la position du nuage

(1) Communication de THURSTONE au colloque de 1955 du CNRS. cf. : l'analyse factorielle et ses applications, Paris, CNRS, 1956, p. 32.

de points avec le minimum possible de coordonnées. Pour cela, on projette le nuage sur des sous-espaces choisis de façon à ce que la déformation soit aussi faible que possible. En effet, plus la déformation est réduite et moins la perte d'information est grande. L'ajustement en sous-espaces se fait en déterminant successivement certaines directions d'allongement du nuage, auxquelles sont ensuite associés les facteurs.

L'observation, c'est-à-dire ici la région, est considérée comme une réalisation d'un vecteur aléatoire qui aurait 50 composantes. De ce fait, il est nécessaire que les variables retenues prennent des valeurs numériques continues et que leur distribution obéisse à la loi normale. Souvent pour remplir cette condition, on raisonne sur des variables centrées et réduites, ou le cas échéant, ayant subi une transformation logarithmique.

Notons par ailleurs, sans entrer dans les détails mathématiques, que l'analyse en composantes principales nécessite des rotations de vecteurs en vue de faciliter l'interprétation des résultats. Malgré l'utilisation de nombreuses méthodes, telle le varimax, ces rotations revêtent un caractère arbitraire pouvant obscurcir l'interprétation des résultats.

L'analyse factorielle classique n'est pas sans poser des problèmes logiques assez délicats que BOUDON présente ainsi (1). "En fait, cette méthode n'est autre qu'un modèle de type causal dont les hypothèses sont extrêmement particulières et qui, pourtant, peut être appliqué en théorie à tout ensemble de variables. En conséquence, on doit, à moins d'adopter un nominalisme radical se demander, dans tous les cas, si les résultats d'une analyse factorielle traduisent une réalité ou représentent un artefact mathématique".

Dans la mesure où la méthode proposée par le Professeur J. P. BENZECRI, sous le nom d'analyse factorielle des correspondances, permet d'éliminer l'hypothèse restrictive concernant la normalité des variables, et autorise même l'utili-

(1) BOUDON (Raymond), L'analyse mathématique des faits sociaux, op. cit. p. 207.

sation de variables discrètes ou booléennes, il semble que l'on soit assuré d'un meilleur résultat qu'avec les composantes principales. La critique de fond de BOUDON reste néanmoins sans réponse. A-t-on ou non un artefact mathématique ? Ce n'est que par une interprétation très poussée des résultats qu'il est possible de déterminer si la structure décrite correspond ou non à un phénomène réel. En fait, une bonne connaissance du problème étudié permet à une personne compétente de répondre assez rapidement.

§3. Les principes de base de l'analyse des correspondances

Pour exposer aussi simplement que possible les principes de l'analyse des correspondances, nous partirons d'un exemple de LEBART (1) relatif à l'étude des catégories socio-professionnelles (C.S.P.) en fonction des causes de décès à l'intérieur d'une même classe d'âge. L'auteur a considéré 13 C.S.P. et 8 causes de décès.

a - La correspondance

A l'intersection de la ligne de la catégorie i appartenant à l'ensemble I des C.S.P. et de la colonne j appartenant à l'ensemble J des causes de décès, se trouve le nombre $k(i, j)$ de décédés. On est en présence d'un ensemble avec répétition de couples (i, j) ce qui correspond à un corpus. Il s'agit donc d'une correspondance statistique. La somme :

$$k = \sum_{(i,j) \in I \times J} k(i, j)$$

est l'effectif du corpus.

Pour que la comparaison entre causes de décès pour 2 C.S.P. soit intéressante, il vaut mieux confronter la part de chaque cause de décès dans l'ensemble des décès, plutôt que les nombres absolus. Ceci revient à dire que l'on s'intéressera plutôt aux probabilités conditionnelles de décès des C.S.P. $p(i)$ et $p'(j)$.

(1) LEBART (Ludovic), Cours de statistique, ISUP, mars 1970.

$$p(i) = \sum_{j \in J} \frac{k(i, j)}{k} \quad p'(j) = \sum_{i \in I} \frac{k(i, j)}{k}$$

l'analyse des C.S.P. et des causes de décès est menée symétriquement.

Dans notre exemple, $p(i)$ caractérise l'importance de la C.S.P. i dans le total des 13 C.S.P. et $p'(j)$ l'importance de la cause de décès j dans le total des 8 causes de décès.

On est passé d'une correspondance statistique à une correspondance aléatoire puisque la mesure de probabilité sur $I \times J$ est décrite par une fonction p à valeurs réelles et positives satisfaisant à la condition :

$$\sum_{(i, j) \in I \times J} p(i, j) = 1$$

on montre que cette correspondance définit sur I une notion de proximité. Si i et i' s'associent de façon semblable aux éléments de J , ils sont proches.

L'analyse des correspondances permet de décrire les proximités pouvant exister entre les lignes et les colonnes du tableau de correspondances (proximités entre formes, indépendamment des niveaux ou tailles), compte tenu des poids différents de ces lignes et de ces colonnes.

Tableau de correspondances

	1	J	j	13
1				
i			$p(i,j)$	
I				
90				

$$\sum_{j \in J} p(i,j) = p(i)$$

$$\sum_{j \in J} \frac{p(i,j)}{p(i)} = 1$$

$$\sum_{i \in I} p(i,j) = p'(j)$$

$$\sum_{i \in I, j \in J} p(i,j) = 1$$

$$\sum_{i \in I} \frac{p(i,j)}{p'(j)} = 1$$

b - La distance

Dans l'espace R^J (ici espace à 8 dimensions) se trouve un nuage de I points. Chaque point est affecté, par analogie avec la mécanique, d'une masse égale à sa probabilité d'apparition $p(i)$. Chaque point définit un vecteur que l'on note P_j^i et dont les J composantes sont $(\frac{p(i,j)}{p(i)})$ $j \in J$ les I points P_j^i sont tous situés dans un sous-espace à J-1 dimensions.

Si deux points i et i' sont proches dans cet espace, cela signifie que les "profils" des lignes représentées par ces points sont voisins. Cette proximité est définie à l'aide de la distance pondérée suivante :

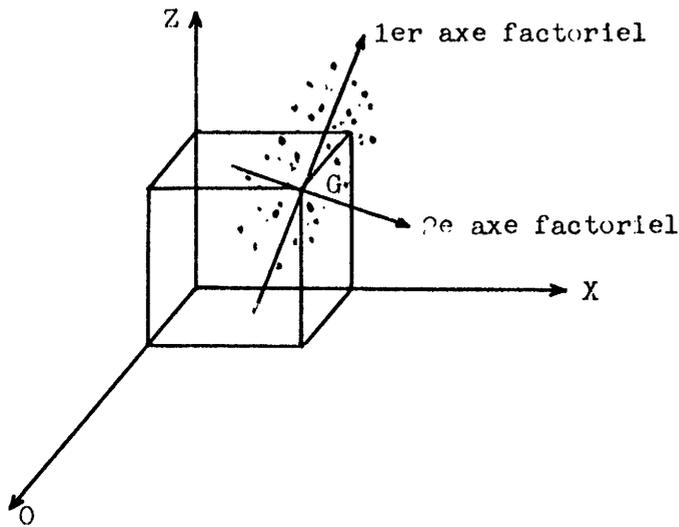
$$d^2(i, i') = \sum_{j \in J} \frac{1}{p(j)} \left[\frac{p(i, j)}{p(i)} - \frac{p(i', j)}{p(i')} \right]^2$$

Cette distance vérifie le principe "d'équivalence distributionnelle". Ceci signifie que si pour deux points i et i' on a $\frac{p(i, j)}{p(i)} = \frac{p(i', j)}{p(i')}$ alors il est possible de considérer un point unique i_0 tel que $p(i_0, j) = p(i, j) + p(i', j)$, et les distances entre les éléments de J ne sont pas changées. Cette propriété explique la stabilité des résultats. Ainsi, si la classification des causes de décès confond alcoolisme et cirrhose du foie au sein d'une seule catégorie, ce regroupement ne modifie pratiquement pas les résultats.

c - Résolution du problème

On a défini un nuage de point R^J muni d'une distance particulière dite de Chi-Carré. Pour représenter ce nuage dans un espace euclidien de dimension réduite en conservant les distances, on effectue une analyse factorielle classique. Celle-ci permet de définir sur I et J des coordonnées qui rendent compte de la correspondance aléatoire. Dans un plan, l'ensemble I rapporté aux axes factoriels est représenté par des points M_i dont les distances varient en sens inverse des proximités que nous avons définies ci-dessus.

Représentation du nuage de points dans un espace à 3 dimensions



Le centre G du nuage a pour projection les moyennes des projections de tous les points du nuage. Si le nuage a une forme sphérique autour de G, on ne peut pas extraire le facteur. Si, au contraire, on a un nuage très allongé le long d'un segment passant par le Centre, on peut déterminer une direction dénommée axe factoriel. La perpendiculaire à cet axe (dans le plan du nuage si on peut assimiler celui-ci à une galette) est le 2^e axe.

Dans un espace à n dimensions, on peut extraire plusieurs facteurs. Mathématiquement, on détermine successivement les axes principaux d'inertie du nuage dans l'ordre décroissant des moments principaux. On arrête l'extraction des facteurs quand leur grandeur est de l'ordre des aléas statistiques, ce que l'on apprécie à l'aide du test du Chi-Carré.

On a vu comment se situaient les points de l'ensemble I pour une correspondance sur I x J. On a représenté les proximités entre les éléments de I dans le plan des deux premiers axes factoriels. On pourrait faire d'une façon symétrique la représentation de l'ensemble J ce qui donnerait

sur le même graphique les proximités entre les éléments de J. Il est démontré qu'entre les deux nuages de points et entre leurs facteurs, il existe une dualité.

Ceci signifie que la catégorie i est d'autant plus près de la variable j que celle-ci intervient avec beaucoup d'intensité dans le profil de cette catégorie.

Ainsi, par rapport à l'essai de Richard STONE, on voit que grâce à l'analyse des correspondances, on dispose d'un outil autrement plus performant. Toutes les difficultés pratiques liées à l'utilisation d'une distance euclidienne sont parfaitement résolues. D'autre part, la mise en évidence de la dualité entre espace des observations et espace des variables est d'un très grand secours pour l'interprétation des résultats.

. . .

CHAPITRE II - DEFINITION DE L'ESPACE DES OBSERVATIONS ET
DE L'ESPACE DES VARIABLES

Choix des observations et choix des variables sont étroitement liés. Le recours à un certain type de données impose un découpage géographique déterminé. En général, plus les circonscriptions de base sont étendues, et plus les données existantes sont nombreuses et de meilleure qualité. A la limite une analyse exigeant des variables trop précises et en trop grand nombre, ne pourrait se faire qu'en considérant les pays. Ceci tient évidemment à ce que l'appareil statistique est conçu d'abord en fonction d'entités nationales, puis secondairement en fonction des régions.

Ce lien inverse entre taille des circonscriptions et renseignements disponibles est particulièrement aigu pour les données non agricoles : statistiques sociales essentiellement. Ainsi la connaissance de la structure de l'emploi, son niveau, le chômage ... est essentielle pour l'étude envisagée. Or, l'enquête de l'Office Statistique de la Communauté Européenne sur les forces de travail est fondamentale de ce point de vue, puisqu'elle donne selon une nomenclature identique pour tous les pays des chiffres sur l'emploi par statut professionnel et secteur d'activité. Mais comme il s'agit d'une enquête par sondage, on a retenu 11 Länder pour l'Allemagne, 8 zones d'études et d'aménagement du territoire (Z.E.A.T.) pour la France, 11 zones pour l'Italie, 4 pour les Pays-Bas, 9 pour la Belgique et 1 pour le Luxembourg. Si l'on souhaite se référer au type de renseignements fournis par cette enquête, on fera une typologie portant sur 44 circonscriptions seulement.

L'analyse des 44 circonscriptions précédemment définies aurait l'avantage d'intégrer des variables très significatives, et surtout bien homogènes. Dès lors que l'on souhaite un découpage géographique plus fin, au moins pour ces éléments on va perdre en homogénéité et qualité. Il s'agit de trouver un compromis harmonieux entre finesse du découpage et type de renseignements disponibles, leur valeur et leur harmonisation entre pays.

On visera dans un premier point à faire la synthèse des analyses préliminaires auxquelles on s'est livré pour cette étude de façon à déterminer quelles sont les variables les plus intéressantes à retenir. Puis, compte tenu des résultats obtenus et des contraintes statistiques, on déterminera le champ des variables et celui des circonscriptions.

A - Enseignement des analyses préliminaires

Notre but n'est pas de reprendre le détail des résultats auxquels on a pu parvenir au cours des analyses préliminaires. Ces résultats, qui figurent dans l'annexe 2, concernent plusieurs analyses :

- typologie de la France découpée en 21 régions de programme;
- typologie de la C.E. découpée en 94 circonscriptions (21 régions françaises, 21 provinces du Bénélux, 32 Regierungsbezirke et 20 regioni italiennes).

D'autres essais ont été effectués à un niveau plus agrégé (Länder au lieu de Regierungsbezirke), ou au contraire avec un découpage plus fin (regioni subdivisées en zones altimétriques, départements au lieu des régions françaises). Il s'agit ici de tirer les conclusions pratiques de toutes ces tentatives, tant pour ce qui est de la nature des variables, et de leur nombre, ainsi que de leur type.

§ 1. Nature des variables

Au cours des premières analyses sur la France, il a été fait appel à des variables de nature très différente en vue de caractériser l'agriculture, ainsi que le niveau général d'activité d'une région.

a - variables caractérisant l'agriculture

Pour définir le système de production d'une zone, on a utilisé :

- des variables monétaires : valeur des principaux postes de la production finale agricole pour ce qui est des céréales, plantes sarclées (betteraves et pommes de terre), légumes, fruits, vin, gros bovins, lait, ovins-caprins...
- des variables physiques : surfaces occupées par les prairies permanentes, les céréales, les plantes sarclées, les cultures arboricoles... ; le nombre d'unités de bétail bovin, ovin-caprin, porcin...
- des variables sous forme de pourcentage : pourcent du lait, de la viande de gros bovins, des céréales... dans la valeur de la production finale agricole.

Il semble bien que, quel que soit le type de variables retenu, on ait des résultats identiques (1). Les légères différences que l'on peut enregistrer sont dues au fait qu'il n'y a pas équivalence parfaite entre variables monétaires et variables physiques. Ainsi au groupe de critères surface toujours en herbe, cultures fourragères, nombre d'unités de bétail bovins, porcins, ovins-caprins, correspondent les rubriques monétaires gros bovins, lait, veau, aviculture, porc, ovins-caprins. Néanmoins, il semble que les résultats en termes physiques donnent une structure générale du graphe plus éclatée. Ceci provient probablement de la plus grande variance de ce type de données par rapport à la structure en pourcent de la production finale.

Pour définir la structure des exploitations, on s'est servi du nombre d'exploitations par classe de surface, ainsi que des superficies occupées par les exploitations par classe de surface. Là aussi, les résultats sont identiques, les variables exprimant les nombres et les surfaces étant parfois confondues pour certaines classes quand on emploie simultanément

(1) Le rapprochement des graphiques 7 et 8 (cf. annexe 2, première analyse sur les 21 régions de programme françaises) montre une très grande similitude dans la position respective des régions dans le plan 1-2 de l'analyse faisant appel à la structure de la production finale (graphe 7), et de l'analyse en termes physiques (graphe 8).

ces deux sortes de données (1). Cela ne signifie pas qu'il soit indifférent de se référer au nombre d'exploitations ou aux surfaces occupées. Ainsi une analyse en nombre diverge légèrement d'une analyse menée uniquement en terme de superficie. Ceci tient à ce que les régions où coexistent un grand nombre de petites exploitations et une faible proportion de très grandes exploitations sont différemment représentées selon la solution choisie. L'analyse en surface est de ce point de vue plus intéressante puisqu'elle permet de mieux cerner ce type de structure.

b - variables caractérisant le niveau général d'activité

Le niveau général d'activité peut être saisi à travers des critères d'emploi, ou à partir de la répartition de la valeur ajoutée globale entre grands secteurs : le primaire censé représenté essentiellement l'agriculture ; le secondaire, l'industrie ; et le tertiaire, les services.

Les valeurs monétaires qui ont l'avantage de donner une mesure de la contribution de chaque secteur à la production de la richesse régionale sont d'emploi délicat. En effet, ces données sont disponibles et significatives seulement pour des zones géographiques de taille relativement importante. Elles ont aussi l'inconvénient d'être élaborées selon les pays à partir de méthodes et en recourant à des définitions qui ne sont pas absolument identiques, d'où des difficultés dans les comparaisons. Enfin, sur un plan plus conceptuel la nécessité d'adopter une unité monétaire unique n'est pas sans poser des problèmes.

Il est donc important de voir si l'utilisation de variables caractérisant le type d'activité prépondérante conduit à des résultats analogues selon que l'on se réfère à la valeur ajoutée par secteur ou à la population active par secteur. L'intérêt de ce rapprochement tient d'une part à ce que les chiffres de population active sont mieux connus et d'autre

(1) L'étude des graphiques 4 et 5 (annexe 2, 1ère analyse France), où figurent à la fois les données en surface et en nombre, montre que pour chaque classe de surface la variable nombre d'exploitation et la variable surface occupée sont très proches. Il n'y a de décalage que dans les classes extrêmes (petites et grandes surfaces).

part à ce qu'ils existent à un échelon administratif beaucoup plus fin. Enfin, malgré quelques différences de définitions par pays, il s'agit de données strictement comparables puisque les unités de mesure sont les mêmes.

De la même manière que pour le système de production, l'analyse a montré que la répartition des observations sur les graphiques était semblable avec la population active, ou avec la valeur ajoutée. Une tentative faite avec le nombre de salariés et la masse des salaires perçus a indiqué que le nombre de salariés pouvait remplacer les actifs du secondaire et du tertiaire, tandis que la masse des salaires donnait les mêmes indications que la valeur ajoutée du secondaire et du tertiaire (1).

Notons que l'utilisation d'un type de données n'exclut pas le recours à l'autre type. Il peut s'avérer intéressant d'introduire les deux sortes de renseignements si l'on souhaite se référer au critère de productivité du travail tant au niveau global qu'au niveau sectoriel.

§2. Nombre et type de variables

Pour des raisons pratiques, on a intérêt à raisonner avec un espace des variables de dimension aussi restreinte que possible. Ces raisons tiennent au coût du traitement qui est une fonction exponentielle de la taille de la matrice de départ, et aux difficultés d'interprétation qui résultent d'un trop grand nombre de caractères retenus.

Dans les analyses préliminaires, on a cherché à caractériser la structure des exploitations par un découpage en classes de surface agricole utilisée aussi fin que possible en distinguant les exploitations sans sol, celles inférieures à 1 ha, 1 à 2, 2 à 5, 5 à 10, 10 à 20, 20 à 50, 50 à 100 et celles supérieures à 100 ha. L'interprétation des résultats nous a permis de constater que les exploitations sans sol

(1) Annexe 2 - 1ère analyse France. Les graphiques donnent une répartition des régions françaises le long du premier facteur très comparable selon que l'on se réfère à l'analyse avec la valeur ajoutée par secteur (graphe 2) et ou à l'analyse avec le nombre d'actifs par secteur (graphes 4 et 5).

n'avaient aucun poids dans l'analyse. D'autre part, on constate que les variables propres aux petites surfaces sont en général très proches les unes des autres (1). Cela signifie que les régions se comportent de manière identique pour ces critères. En conséquence, il est possible de remplacer ces critères par un seul qui est égal à leur somme, sans modifier la répartition des observations et cela en vertu du principe d'équivalence distributionnelle.

Afin de définir la nature économique du système de production, un essai a été fait incluant des données très élaborées : coefficients de terre, de travail, de consommations intermédiaires, d'amortissements et d'intérêts du capital. Ces coefficients, exprimés en pourcent de la production finale agricole brut, donnent une idée de la combinaison productive prépondérante d'une région. Il s'avère que ces éléments, théoriquement fort intéressants, sont peu instructifs car ils n'interviennent pas pour expliquer comment se répartissent les circonscriptions. De même, les taux de remplacement des chefs d'exploitation et le pourcentage des terres libérées dans les années prochaines ne jouent aucun rôle (2).

Le peu d'influence de ces critères ne peut pas être attribué à la forme sous laquelle ils sont exprimés. En effet, on a vu que la description du système de production en pourcent de la production finale donnait des résultats convenables. C'est donc le type de variable employée qu'il faut incriminer. Des données présentées sous une forme trop élaborée semblent apporter une information peu intéressante dans le traitement des observations. Les manipulations qu'elles ont subi les rendent moins aptes à exprimer les phénomènes qu'elles sont censées représenter de manière synthétique. Ce type de variable n'a pas une grande puissance discriminante.

L'utilisation de critères complexes en vue de caractériser une structure conduit à des résultats décevants. Il est préférable de faire directement appel aux éléments de

(1) cf. annexe 2 - 2^e analyse. 94 régions, graphes 1 et 2.

(2) Tous ces critères ont été utilisés au niveau des 21 régions de programme françaises, cf. annexe 2, 1^{ère} série, graphiques n^o 2, 3 et 7.

base qui ont été combinés, car ils décrivent le phénomène étudié sans perte d'information. D'autre part, l'utilisation de certaines conventions dans le calcul et l'élaboration de ces critères composites est susceptible de "brouiller" la représentation de la structure étudiée, ou de faire écran. La variable synthétique étant déjà un début d'interprétation des faits, mieux vaut se référer aux caractéristiques originales sans idée préconçue sur la manière dont elles se combinent, puisque la méthode l'autorise.

A l'inverse, une démarche basée uniquement sur des critères trop analytiques est à éviter (1). Une variable simple qui conduit à caractériser un phénomène sans portée générale, car restreint à un petit nombre de circonscriptions, est d'une utilité restreinte. Cette variable peut apporter du point de vue purement statistique une information importante, mais elle se révélera sans intérêt pour la recherche en cours. C'est ce qui risque de se produire si l'on tente de représenter le système de production, à l'aide d'une nomenclature de produit (de type physique, monétaire ou autre) trop fine. Quelle est la portée, du point de vue de la typologie à élaborer, de la caractérisation d'une ou plusieurs régions par la culture de petits pois par exemple ? Au cours des précédentes analyses, il est arrivé qu'apparaisse un facteur constitué, pour partie, par la variable ovins-caprins à laquelle étaient associées certaines régions, dont Sardegnna et Midi-Pyrénées (2).

(1) Ainsi dans l'analyse 64 variables portant sur 94 régions au niveau de l'ancienne Communauté, la surface agricole utilisée a été subdivisée en 11 rubriques : cultures permanentes dont cultures fruitières, vignes, céréales, pommes de terre, betteraves sucrières, surface toujours en herbe, plantes industrielles, fleurs et plantes ornementales, cultures sous verre, cultures de fourrages verts des terres arables. A ces catégories avaient été jointes les surfaces boisées faisant partie des exploitations agricoles. Sur ces 12 variables seules les cultures permanentes, vignes, céréales, surfaces toujours en herbe et surfaces boisées, soit 5 variables, interviennent de façon marquée dans la définition des facteurs. Les betteraves sucrières, les plantes industrielles, les cultures fourragères et les fruits ont un poids limité. Les pommes de terre n'interviennent presque pas. Quant aux cultures sous verre, ainsi que fleurs et plantes ornementales, leur poids dans l'analyse est négligeable. - Cf. Annexe 2 - 2^e analyse.

(2) Il s'agit du deuxième facteur de l'analyse 64 variables 94 régions. cf. Annexe 2 - 2^e analyse - graphiques n° 1 et n° 4.

Un tel renseignement n'a pas de valeur opérationnelle. Mieux vaut une variable plus générale permettant de distinguer le système herbager extensif, qu'il soit accompagné de la production ovine, caprine ou bovine. Pour cela, le nombre d'ha couverts par la surface toujours en herbe associé à de grandes exploitations est satisfaisant.

Il est apparu que certaines variables très significatives sur le plan économique, répartition des chefs d'exploitation par classes d'âge, présence ou non d'un successeur potentiel, ont un poids limité dans l'analyse quand elles sont introduites telles quelles. Par ailleurs, nous avons constaté que dans tous les traitements, les éléments de structure agraire (distribution des exploitations par classe de surface, en nombre ou en superficie) interviennent toujours avec intensité, déterminant une répartition des régions de type parabolique. Ceci nous a conduit à penser que si l'on souhaitait voir jouer un rôle dans l'analyse aux critères âge et successeur, il était nécessaire de les présenter par classe de surface.

Enfin, on constate de façon générale une très forte liaison entre système de production et structure agraire. Aux petites surfaces correspond un système intensif alors que les grandes surfaces sont associées aux céréales ou aux surfaces toujours en herbe. Ceci autorise un choix restreint de productions pour définir le système productif dès lors que l'on fait intervenir la structure agraire (1).

Ainsi l'analyse des données se révèle un précieux outil puisqu'elle permet de voir quelles sont les variables dont l'action est prépondérante, et celles qui font double emploi, ou sont sans effet. Les hypothèses de départ peuvent donc être testées. D'où l'intérêt de la méthode qui est l'amorce d'un processus expérimental dans les sciences sociales.

En effet, l'hypothèse d'une très forte dépendance entre système de production et structure a pu être aisément vérifiée. Il a suffi de supprimer les données sur le nombre

(1) cf. Annexe 2 - 2^e analyse - graphiques n° 1 avec 64 variables et n° 6 avec 23 variables (critères nombre d'exploitations par classe de SAU supprimés).

d'exploitations par classe de surface. Aucune modification n'étant apparue dans les résultats, il était évident que le lien entre ces deux types de données est très étroit. De la même manière, on a pu juger de l'importance des variables relatives à la population active non agricole. Leur suppression a totalement changé le classement des circonscriptions (1).

B - Données de base pour l'analyse définitive

§ 1. L'espace des observations

Compte tenu des objectifs de la typologie et de la disponibilité de statistiques pour un découpage géographique donné, deux sortes d'analyses ont été effectuées. L'une porte uniquement sur des variables agricoles et concerne pour l'Italie les zones altimétriques. L'autre porte essentiellement sur des variables générales et concerne pour l'Italie les provinces. Le seul recoupement possible pour l'Italie a trait aux regioni.

1) Les circonscriptions retenues dans l'analyse portant uniquement sur des variables agricoles

- Pour la France, on a retenu 89 départements (la Seine et l'ancienne Seine-et-Oise étant confondues en une seule circonscription), et cela au lieu des 21 régions de programme considérées au cours des analyses préliminaires.
- Pour le Bénélux, on a conservé le découpage déjà utilisé en 21 zones : 9 provinces belges, 11 provinces néerlandaises et le Grand Duché du Luxembourg, celui-ci comptant pour une circonscription.
- Pour la R.F. d'Allemagne, on s'est référé aux Regierungsbezirke. Mais les villes-états Bremen, Berlin et Hambourg ont été éliminées. Leur rejet se justifie dans la mesure où l'on cherche à élaborer une typologie des régions tenant compte des problèmes de l'agriculture, même si c'est en relation avec le reste de l'économie.

(1) Ceci apparaît quand on compare dans l'analyse avec 47 variables relative aux 21 régions de programme françaises (1ère analyse en Annexe 2) les graphes 2 et 7. Ainsi la région Nord proche dans le graphe 2 de la région Parisienne, se trouve avec la Franche-Comté et le Poitou-Charentes dans le graphe 7.

Or, ces villes-états ont très peu d'agriculture. Par ailleurs, on a tenu compte de la fusion de certains Regierungsbezirke : Darmstadt-Wiesbaden, Koblenz-Montabaur et Rheinhessen-Pfalz. Au total, on dispose de 32 zones.

- Pour l'Italie, les unités prises en compte correspondent au découpage de l'enquête communautaire sur la structure des exploitations agricoles. Il s'agit des zones altimétriques : plaines, collines et montagnes. Ces zones sont des subdivisions à l'intérieur des "Regioni", ce qui nous donne en tout 51 circonscriptions. Au niveau des "Province", il n'existe aucun résultat agricole pour 1967. Seuls, certains éléments du Recensement Général de l'Agriculture de 1970 sont aujourd'hui disponibles, mais de façon globale. Ces renseignements n'autorisent pas la distinction entre les trois niveaux d'activité retenus (1). Le découpage le plus fin susceptible d'être pris en compte pour l'Italie est celui des zones altimétriques, si l'on souhaite intégrer dans l'analyse un nombre suffisamment grand de variables agricoles.

Au total, on dispose de 193 observations auxquelles il faut ajouter les 20 "Regioni" italiennes. Celles-ci que l'on retrouve également dans l'analyse avec des variables générales, servent de trait de liaison pour l'interprétation des résultats italiens avec les circonscriptions de base. Le découpage géographique ainsi adopté conduit à des unités de base plus comparables entre elles du point de vue agricole, que dans l'analyse préliminaire portant sur 94 zones en tout.

2) Les circonscriptions retenues dans l'analyse économique générale

Ces circonscriptions se situent au niveau des province italiennes. Seules les province Isernia et Campobasso n'ont pas été prises en considération en raison de l'absence de données dans certains domaines pour ces circonscriptions. Nous avons donc dû prendre en compte l'unité Molise qui regroupe Isernia et Campobasso.

(1) On entend par "niveau d'activité" : - les exploitations dont le chef n'a aucune activité extérieure, ce qui correspond à une notion d'agriculture à temps plein; - les exploitations dont le chef a une activité extérieure égale ou inférieure à 50 % de son temps, ce qui correspond à une notion d'agriculture à titre principal; - l'ensemble des exploitations, quelle que soit l'importance de l'activité extérieure du chef.

Dans cette analyse intégrant surtout des variables générales, on est en présence des 89 départements, 21 provinces Bénélux, 32 Regierungsbezirke et 93 Province soit 235 observations auxquelles il faut toujours joindre les 20 regioni italiennes, soit au total 255 observations. En réalité, Valle d'Aosta et Molise apparaissent deux fois en province et en regioni.

3) Les circonscriptions communes aux deux analyses

Elles sont au nombre de 162, à savoir : 89 départements pour la France, 21 provinces pour le Bénélux, 32 Regierungsbezirke pour la R.F. d'Allemagne et 20 Regioni pour l'Italie.

+

+

+

On trouvera en annexe à la fin de ce chapitre (p. 57 à 66) la liste des circonscriptions et des variables ainsi que les symboles et abréviations utilisés dans les graphiques du Chapitre III.

+

+

+

§ 2. L'espace des variables

La nature des variables et leur présentation diffèrent très sensiblement selon que l'on considère l'analyse purement agricole ou l'analyse générale. L'analyse agricole constituant le prolongement direct d'un précédent essai (1), nous donnerons quelques indications le concernant, puis nous présenterons les variables retenues en fin de compte. Ensuite, nous verrons les variables de l'analyse générale.

1) La première phase de l'analyse uniquement en termes agricoles

Le matériau de base pour cette approche est constitué des seuls éléments conterus dans l'enquête sur la structure des exploitations effectuée en 1967 dans l'ensemble des pays de la Communauté, sous l'égide de l'Office Statistique des Communautés Européennes. Cette enquête communautaire a l'avantage de fournir des résultats très homogènes d'un pays à l'autre puisque les définitions sont identiques. Les résultats concernant l'Italie sont représentatifs au niveau des Regioni et de leurs subdivisions en zones altimétriques, d'où l'obligation imposée dans le choix des observations.

L'analyse dont il est question a été menée en favorisant deux aspects qui nous ont paru primordiaux. D'une part, les relations de l'agriculture avec les autres activités économiques de la région. D'autre part, les possibilités d'évolution interne de l'agriculture en fonction de critères démographiques.

a - La prise en compte des interrelations entre agriculture et autres activités

Celle-ci s'est faite à partir de l'importance de l'activité extérieure du chef d'exploitation. Trois niveaux ont été distingués :

- les exploitations dont le chef n'a aucune activité extérieure ce qui correspond à une notion d'agriculture à temps plein (V_1)(2);

(1) Cet essai a fait l'objet d'un document de travail peu élaboré, aussi n'a-t-on pas fait figurer de résultats en annexe. Le très grand nombre d'observations retenues rendait difficile toute représentation graphique.

(2) cf. tableau p. 48 et 49.

- les exploitations dont le chef n'a aucune activité extérieure ou a une activité extérieure égale ou inférieure à 50 % de son temps, ce qui correspond à une notion d'agriculture à titre principal ($V_1 + V_3$) (1);
- l'ensemble des exploitations, quelle que soit l'importance de l'activité extérieure du chef ($V_1 + V_2 + V_3$) (1).

Sur un plan pratique ces trois possibilités ont été envisagées comme si l'on avait pour une région 3 observations distinctes, ce qui revient à définir pour la circonscription i : i_1 comprenant les seules exploitations à temps plein, i_2 avec les exploitations à titre principal et i_3 avec l'ensemble des exploitations.

Une approche des problèmes régionaux basée sur cette subdivision paraît à première vue susceptible de fournir des indications intéressantes quant à l'élaboration d'une politique agricole adaptée. En effet, il est probable que les mesures à prendre différeront selon la prédominance d'un type ou l'autre d'agriculture (type exclusif, à titre principal, ou toutes exploitations confondues).

Si l'agriculture paraît défavorisée dans un ensemble de circonscriptions où les possibilités d'activité extérieure sont nombreuses, la situation n'offre pas un caractère de gravité tel qu'il urge de prendre des mesures de caractère économique. Même si le revenu agricole régional est relativement faible, grâce aux emplois secondaires, il se peut que le niveau de vie des agriculteurs soit décent. Peut-être, une intervention sur le plan social pour les agriculteurs les plus désavantagés est-elle suffisante.

A l'inverse, dans un type de régions où l'agriculture à temps plein prédomine et où se manifestent des difficultés, il conviendra d'appliquer d'autres mesures à caractère essentiellement économique : mesures concernant les investissements et les structures, mesures relatives à l'économie régionale entière, afin de favoriser la création d'emplois non agricoles. Encore, faudrait-il distinguer parmi les exploitants à activité extérieure nulle deux catégories bien différentes :

(1) cf. tableau p. 48 et 49.

- les exploitants qui ne recherchent pas de revenu d'appoint, soit parce que leur exploitation leur rapporte assez, soit parce qu'il s'agit d'exploitations de subsistance,
- les exploitants qui souhaiteraient exercer une activité secondaire, mais ne le peuvent pas, car ils ne trouvent pas de travail à leur convenance, ou parce que leur qualification professionnelle est insuffisante.

b - Les possibilités d'évolution interne de l'agriculture

Celles-ci ont été saisies à partir :

- de l'importance du nombre d'exploitants âgés de plus de 50 ans et ceux âgés de 50 ans et moins,
- de l'existence ou non chez les exploitants âgés de plus de 50 ans d'un successeur potentiel.

Ceci donne 4 variables démographiques qui ont été individualisées pour six classes de surface et en distinguant les résultats en nombre (par exemple, nombre d'exploitants sans successeur dans la classe 5-10 ha) des résultats en surface (nombre d'ha occupés par les exploitants de la classe 5-10 ha par exemple).

Tous ces résultats peuvent être synthétisés comme provenant d'un croisement entre niveaux d'activité extérieure par présence d'un successeur et ceci par classe de surface, complété par un croisement entre niveaux d'activité extérieure par l'âge et ceci également par classe de surface. Tous ces résultats sont dédoublés en tenant compte du nombre d'exploitants et des surfaces occupées.

On a pour raison des traitements en nombre et en surface, 6 valeurs (classes de surface) pour chacune des 4 variables de nature démographique (âge et successeur), soit 24 données par circonscription. Celles-ci sont multipliées par 3 fois (trois niveaux d'activité extérieure) 213 circonscriptions.

Peut-être eut-il été souhaitable d'introduire plus de finesse dans les variables relatives à l'âge du chef

d'exploitation. Le choix qui a été fait repose sur l'hypothèse qu'en deçà de 50 ans des possibilités d'adaptation et de modification des structures existent, alors qu'au-delà les choses sont figées. D'autre part, la limite de 50 ans permet d'avoir un ensemble d'exploitations résultant de l'addition des deux sous-ensembles avec et sans successeur, puisque ces deux sous-ensembles sont relatifs aux chefs de plus de 50 ans.

Le recours à un critère d'âge défini selon 4 classes, telles que chefs d'exploitations ayant moins de 35 ans, 35 à 50 ans, 50 à 65 ans et plus de 65 ans eût conduit à accroître la taille du tableau initial de façon très importante. En effet, cela revient à passer d'une matrice de 213 observations x 48 variables à une matrice de 213 x 60 variables, ce qui pose des problèmes de taille de la mémoire centrale de l'ordinateur nécessaire aux calculs.

Les éléments nécessaires à l'élaboration de ces données proviennent d'un dépouillement particulier de l'enquête structure. Les deux tableaux de centralisation que nous avons constitué pour cette analyse figurent p. 48 et 49.

2) Les variables retenues dans l'analyse agricole définitive

Le principal enseignement de cette première phase (1) de l'analyse uniquement en termes agricoles, est la grande identité des résultats selon que l'on travaille avec le nombre des exploitations ou la surface occupée. Ceci à la différence près, déjà signalée, que la surface permet d'identifier les régions caractérisées par un petit nombre de très grandes exploitations.

Le second enseignement est le faible impact dans la répartition du nuage de points de la notion d'activité extérieure du chef d'exploitation, alors que les autres variables sont assez discriminantes.

D'autre part, une lacune se manifeste due à ce que les regroupements de régions selon les classes de surface ne

(1) Voir p. 44 § 2. 1) la première phase de l'analyse uniquement en termes agricoles.

Exploitants de plus de 50 ans

Province :

	avec successeur potentiel						sans successeur potentiel					
	< 2 ha	2-5	5-10	10-20	20-50	> 50	< 2 ha	2-5	5-10	10-20	20-50	> 50
nombre d'exploitations par classe												
exploitants sans activité extérieure (V ₁)												
activité extérieure ≤ 50 % (V ₃)												
agriculture à titre principal (V ₁ + V ₃)												
activité extérieure > 50 % (V ₂)												
ensemble exploitation (V ₁ + V ₂ + V ₃)												
surface occupée												
exploitants sans activité extérieure (V ₁)												
activité extérieure ≤ 50 % (V ₃)												
agriculture à titre principal (V ₁ + V ₃)												
activité extérieure > 50 % (V ₂)												
ensemble exploitation (V ₁ + V ₂ + V ₃)												

Province :

nombre d'exploitations par classe	Exploitants ayant plus de 50 ans (1)						Exploitants ayant 50 ans et moins (1)					
	< 2 ha	2-5	5-10	10-20	20-50	> 50	< 2 ha	2-5	5-10	10-20	20-50	> 50
exploitants sans activité extérieure (V ₁)												
activité extérieure ≤ 50 % (V ₃)												
agriculture à titre principal (V ₁ + V ₃)												
activité extérieure > 50 % (V ₂)												
ensemble exploitation (V ₁ + V ₂ + V ₃)												
surface occupée												
exploitants sans activité extérieure (V ₁)												
activité extérieure ≤ 50 % (V ₃)												
agriculture à titre principal (V ₁ + V ₃)												
activité extérieure > 50 % (V ₂)												
ensemble exploitation (V ₁ + V ₂ + V ₃)												

(1) ou'ils aient ou non un successeur potentiel.

tiennent pas compte des grandes différences dans les systèmes de production qui peuvent exister.

Pour améliorer l'analyse, il apparaît nécessaire d'introduire sous une forme aussi simple que possible le système de production. C'est ce que l'on a fait en relevant les surfaces cultivées, plus précisément les terres labourables, désignées par les lettres tl, et en distinguant parmi celles-ci les surfaces en céréales, indiquées par les lettres cér. Comme autres critères significatifs, nous avons retenu les surfaces toujours en herbe (sth) qui sont le propre des systèmes extensifs, et les cultures permanentes (perm) qui caractérisent bien les types de culture intensive.

Toutefois, l'introduction du système de production n'est possible qu'au niveau de l'ensemble des exploitations, car le dépouillement particulier de l'enquête communautaire dans lequel figure les variables, présence ou non d'un successeur potentiel, ne fournit aucune indication sur le système de production, selon l'activité extérieure du chef d'exploitation. Ainsi nous n'avons pas retenu pour l'analyse définitive le niveau d'activité, qui n'avait pas donné de résultats très probants, au profit des variables indicatives du système de production (1).

Le choix auquel nous avons été conduits (introduction des surfaces cultivées par grandes catégories), nous a incité à raisonner pour l'ensemble des autres variables également en termes de surfaces occupées. Les données traitées sont donc d'une parfaite homogénéité.

(1) Notons qu'en individualisant plus précisément les niveaux d'activité, exploitants à activité extérieure nulle, à activité extérieure inférieure à 50 %, à activité extérieure supérieure à 50 % au lieu des notions d'agriculture à temps plein, à titre principal et ensemble des exploitations (voir p. 44), on aurait probablement obtenu de meilleurs résultats.

Cette démarche aurait l'avantage de caractériser les régions ayant un profil particulier quant à l'activité extérieure du chef d'exploitation. Toutefois, nous avons préféré l'autre formule en raison de la nécessité de faire correspondre, dans la phase ultérieure, les résultats tirés du dépouillement spécial, avec des chiffres relatifs au système de production. Or, de tels chiffres ne sont disponibles que pour l'ensemble des exploitations, ce qui aurait obligé à retenir cette catégorie. Inclure celle-ci en supplément des trois niveaux (activité extérieure nulle, inférieure à 50 % ou supérieure à 50 %), eût conduit à augmenter très sensiblement le volume des données à traiter (213 x 4 soit 852 observations au lieu de 213 x 3 soit 639 observations, ce qui est déjà considérable).

3) Récapitulation des variables retenues pour l'analyse agricole

Pour l'analyse agricole définitive ont été retenues, pour les 213 circonscriptions, les variables suivantes :

- l'âge du chef d'exploitation selon deux classes (50 ans et plus; moins de 50 ans),
- parmi les chefs d'exploitation de 50 ans et plus, deux groupes selon la présence ou non d'un successeur potentiel,
- le système de production défini par quatre modes d'utilisation du sol (terres labourées, terres cultivées en céréales, superficie toujours en herbe et superficie en cultures permanentes),
- les catégories de surface au nombre de 6 :
 - moins de 2 ha (classe désignée par le chiffre arabe 1)
 - de 2 à 5 ha (classe désignée par le chiffre arabe 2)
 - de 5 à 10 ha (classe désignée par le chiffre arabe 3)
 - de 10 à 20 ha (classe désignée par le chiffre arabe 4)
 - de 20 à 50 ha (classe désignée par le chiffre arabe 5)
 - de plus de 50 ha (classe désignée par le chiffre arabe 6)

A chacune de ces classes de surface correspond donc la superficie occupée par :

- les exploitations dont le chef âgé de plus de 50 ans est pourvu d'un successeur potentiel (variable désignée par l'abréviation Eas),
- les exploitations dont le chef âgé de plus de 50 ans n'est pas pourvu d'un successeur potentiel (variable désignée par l'abréviation Ess),
- les exploitations dont le chef a plus de 50 ans qu'il y ait ou non un successeur (variable désignée par Ev (Ev = Eas + Ess)),
- exploitations dont le chef a moins de 50 ans sans qu'intervienne la notion de successeur (variable désignée par Ej),
- les terres labourables (tl),
- les céréales (cer),
- les surfaces toujours en herbe (sth),
- les cultures permanentes (perm).

Ceci conduit à un total de 6 x 8, soit 48 variables indiquées par leur code suivi d'un chiffre arabe fixant leur appartenance à une classe de surface (Par exemple Ess 6 signifie : les surfaces occupées par les exploitations de plus de 50 ha dont le chef est sans successeur potentiel).

Cette façon de procéder combine les aspects de structure, vus de manière statique, ainsi que les aspects dynamiques perçus à l'aide d'un critère démographique, âge des chefs d'exploitation, et du critère présence ou non d'un successeur potentiel.

4) Les variables retenues pour l'analyse économique générale

Rappelons que cette analyse caractérise 255 observations, dont les 93 provinces italiennes et les 20 Régioni. Le facteur limitant dans le choix des variables se situe au niveau des provinces pour lesquelles on ne dispose que d'un très petit nombre d'éléments agricoles. Ceux-ci proviennent du recensement général de l'agriculture en 1970 qui a fait l'objet de publications contenant un volume restreint d'informations.

Trois types de données ont été utilisées :

- données sur la population,
- données sur la valeur ajoutée,
- données générales sur l'agriculture.

a - données concernant la population

On a fait appel aux données disponibles les plus complètes, tant en matière de population active que de population totale. Il s'agit de chiffres qui nous ont été fournis par la Direction Générale de la politique régionale et dont un certain nombre ont fait l'objet d'une publication récapitulative (1). Dans certains cas, notamment les province, la répartition des actifs entre secteur secondaire et secteur tertiaire a dû faire l'objet d'estimations de la part des services de la Communauté.

Cette répartition des actifs en trois secteurs (primaire, assimilé à l'agriculture ; secondaire assimilé à l'industrie et tertiaire assimilé aux services) a été complétée par l'utilisation de données sur la population totale, la superficie des circonscriptions et la densité.

En ce qui concerne la R.F. d'Allemagne, se posait le problème des villes-états Bremen, Berlin et Hamburg, dans lesquelles l'activité agricole est négligeable. Mais ces villes sauf Berlin, dont la situation est particulière, exercent sur le plan de l'activité économique générale une action sur les régions avoisinantes, principalement pour ce qui est des possibilités d'emploi. Afin de tenir compte de cela, l'action de Bremen a été supposée s'exercer pour un tiers sur Stade, un tiers sur Hannover et un tiers sur Oldenburg. De la même manière, Hamburg exerce une influence que l'on a supposée être répartie par tiers également, sur Schleswig-Holstein, Luneburg et Stade. Ces proportions ont été fixées par les services compétents de la Communauté. Quant à la ville-état de Berlin elle a été éliminée de l'analyse.

Dans l'estimation de la population totale et de la densité pour la circonscription Seine, qui comprend le département de la Seine proprement dit ainsi que l'ancienne Seine-et-Oise, nous avons ôté la population parisienne qui eût provoqué un gonflement trop important ayant quelque chose d'artificiel.

(1) cf. L'évolution régionale dans la Communauté. Bilan analytique, 1971. Commission des Communautés Européennes, Annexe statistique, p. 169-314.

b - données concernant la valeur ajoutée

La Direction générale de la politique régionale s'est livrée à un délicat travail d'estimation de la valeur ajoutée, et de sa ventilation par secteur, au niveau de l'ensemble des circonscriptions. Grossièrement, on peut dire que la méthode suivie consiste à définir à l'aide de toutes les statistiques régionales disponibles dans chaque pays, des coefficients de répartition du Produit Intérieur Brut (PIB) par circonscription et par secteur. Pour ce faire, on s'est référé au niveau français aux comptes régionaux de 1962 et 1967, ainsi qu'à la répartition de la population active par secteur en 1968. Au niveau italien, on a fait appel aux chiffres par province établis par Tagliacarne ("Il reddito prodotto nelle provincie italiane 1963-1970"). Pour le Bénélux, les données de base concernent 1965 et 1966, et pour la R.F. d'Allemagne 1966.

L'Office Statistique des Communautés Européennes fournit une estimation globale du PIB par pays élaborée selon une base homogène. Le passage des valeurs monétaires nationales en Unités de Compte communes se fait à l'aide des taux pivots arrêtés à Washington les 18 et 19 décembre 1972. L'année de base considérée est 1970. Aux estimations nationales en U.C. sont enfin appliqués les coefficients de répartition précédemment définis.

De la même manière que pour la population, les circonscriptions allemandes : Schleswig-Holstein, Lüneburg et Stade ont vu leur produit accru du tiers de la valeur ajoutée de Hamburg. Stade, Hannover et Oldenburg ont également vu leur produit augmenté du tiers de la valeur ajoutée de Bremen.

Enfin, puisque les données existaient, on a estimé la valeur ajoutée par actif. Le PIB sur le nombre total des actifs conduit à une notion de productivité de l'ensemble des travailleurs. Le rapport valeur ajoutée de l'agriculture sur nombre d'actifs agricoles donne une estimation de la productivité des travailleurs de ce secteur. Ont été aussi calculées les productivités des travailleurs du secondaire, du tertiaire et de l'ensemble des travailleurs non agricoles.

c - données concernant l'agriculture

Compte tenu des données publiées par l'Institut de Statistique d'Italie (ISTAT) sur le recensement général de l'agriculture de 1970, nous avons retenu les variables suivantes :

- nombre d'exploitations par classe de surface agricole utilisée : moins de 2 ha, 2 à 5 ha, 5 à 10, 10 à 20, 20 à 50 et plus de 50 ha,
- répartition du territoire agricole en terres labourables, surfaces toujours en herbe et cultures permanentes. On dispose également des surfaces boisées. Mais un rapprochement au niveau des regioni entre les chiffres de l'enquête communautaire de 1967 et ceux du recensement général de 1970, met en évidence des écarts trop importants pour qu'on les attribue simplement à l'évolution naturelle. Il est probable que les différences de définition interviennent. Pour les terres labourables, surfaces toujours en herbe et cultures permanentes, l'écart constaté entre statistiques de 1967 et statistiques de 1970 portant sur les regioni est moins important.

d - récapitulation des variables retenues pour l'analyse économique générale

Pour l'analyse économique générale ont donc été retenues les variables suivantes pour les 255 circonscriptions :

- population totale (indiquée par Ptot)
- surface totale (indiquée par S)
- densité de la population (indiquée par d)

- actifs totaux (désignés par Pat)
- actifs du secteur agricole (désignés par Paa)
- actifs du secteur industriel (désignés par Pai)
- actifs du secteur des services (désignés par Pas)

- le produit intérieur brut (indiqué par Rt)
- la valeur ajoutée de l'agriculture (Ra)
- la valeur ajoutée de l'industrie (Ri)
- la valeur ajoutée des services (Rs)

- le rapport entre PIB et actifs totaux (Fr)
- le rapport entre BAa et Aa (Fra)
- le rapport entre VAI et Ai (Pri)
- le rapport entre VAs et As (Prs)

- le rapport entre valeur ajoutée de l'industrie et des services sur la population active de l'industrie et des services (Frna)

- nombre d'exploitations de moins de 2 ha (n_1)
- nombre d'exploitations de 2 à 5 ha (n_2)
- nombre d'exploitations de 5 à 10 ha (n_3)
- nombre d'exploitations de 10 à 20 ha (n_4)
- nombre d'exploitations de 20 à 50 ha (n_5)
- nombre d'exploitations de plus de 50 ha (n_6)

- surface occupée par les terres labourables (tl)
- surfaces toujours en herbe (prairies permanentes) (sth)
- surface occupée par les cultures permanentes (perm)

Ceci conduit à 16 variables économiques générales et 9 variables agricoles.

ANNEXE AU CHAPITRE II :

Liste des symboles et abréviations utilisés dans les graphiques

Afin de faciliter la lecture et l'interprétation des graphiques joints à la présentation des résultats des analyses (Chapitre III), il est donné ci-après :

- la liste des circonscriptions,
- la liste des variables.

LISTE DES CIRCONSCRIPTIONS COMMUNES AUX DEUX ANALYSES

BENELUX - Dans les graphiques en couleur les abréviations et symboles pour les pays du BENELUX sont en rouge.

Les circonscriptions sont désignées par 3 lettres précédées du symbole suivant :

- Belgique, un cercle ○
- Pays-Bas, un cercle avec une croix ⊕
- G.D. de Luxembourg, deux cercles concentriques ⊙

<u>Grand Duché du Luxembourg</u>		<u>Pays-Bas</u>	
	D-LX	Groningen	GRO
<u>Belgique</u>		Friesland	FRS
Antwerpen	ANT	Drenthe	DRN
Brabant belge	BBB	Overijssel	OVR
Hainaut	HAI	Gelderland	GLD
Liège	LIE	Utrecht	UTR
Limburg belge	LBB	Noord-Holland	N-HL
Luxembourg belge	LXB	Zuid-Holland	Z-HL
Namur	NAM	Zeeland	ZEE
Oost-Vlaanderen	O-VL	Brabant	BBN
West-Vlaanderen	W-VL	Limburg	LBN

R.F. d'ALLEMAGNE - Dans les graphiques en couleur les symboles et abréviations sont en jaune.

Les circonscriptions sont désignées par 3 lettres précédées d'un carré pointe en bas ◇

Schleswig-Holstein	SCH	Darmstadt-Wiesbaden	DR-W
Saarland	SAA	Kassel	KAS
Hannover	HAN	Koblenz-Montabaur	KB-M
Hildesheim	HIL	Trier	TRI
Lüneburg	LUN	Rheinhessen-Pfalz	RH-P
Stade	STA	Nordwürttemberg	NWU
Osnabrück	OSN	Nordbaden	NBD
Aurich	AUR	Sudbaden	SBD
Braunschweig	BRS	Sudwürttemberg	SWU
Oldenburg	OLD	Oberbayern	OBA
Düsseldorf	DUS	Niederbayern	NBA
Köln	KLN	Oberpfalz	OPF
Aachen	AAC	Oberfranken	OFR
Münster	MUN	Mittelfranken	MFR
Detmold	DET	Unterfranken	UFR
Arnsberg	ARN	Schwaben	SCW

FRANCE - Dans les graphiques en couleur les symboles et abréviations sont en bleu.

Les circonscriptions sont désignées par 3 lettres précédées d'un carré □

Nord	NRD	Charente-Maritime	CH-M
Pas-de-Calais	PdC	Deux-Sèvres	SEV
Aisne	AIS	Vienne	VIN
Oise	OIS	Corrèze	CRZ
Somme	SOM	Creuse	CRS
Seine, Seine-et-Oise	SEN	Haute-Vienne	H-VN
Seine-et-Marne	S&M	Dordogne	DRD
Cher	CHE	Gironde	GIR
Eure-et-Loir	E&L	Landes	LDE
Indre	IND	Lot-et-Garonne	L&G
Indre-et-Loire	I&L	Pyrénées Atlantiques	PY-A
Loir-et-Cher	L&C	Ariège	ARG
Loiret	LOI	Aveyron	AVR
Eure	EUR	Haute-Garonne	H-GR
Seine-Maritime	SNM	Gers	GER
Calvados	CLV	Lot	LOT
Manche	MCH	Hauts-Pyrénées	H-PY
Orne	ORN	Tarn	TRN
Côtes-du-Nord	CdN	Tarn-et-Garonne	T&G
Finistère	FIN	Ardennes	ADN
Ille-et-Vilaine	I&V	Aube	AUB
Morbihan	MHN	Marne	MRN
Loire-Atlantique	L-AT	Haute-Marne	H-MR
Maine-et-Loire	M&L	Meurthe-et-Moselle	M&M
Mayenne	MAY	Meuse	MSE
Sarthe	SRT	Moselle	MOS
Vendée	VDE	Vosges	VOG
Charente	CHA	Bas-Rhin	B-RH

FRANCE (suite)

Haut-Rhin	H-RH	Loire	LRE
Doubs	DBS	Rhône	RHO
Jura	JRA	Savoie	SAV
Haute-Saône	H-SO	Haute-Savoie	H-SV
Belfort	BLF	Aude	AUD
Côte d'Or	CdO	Gard	GRD
Nièvre	NVR	Hérault	HRL
Saône-et-Loire	S&L	Lozère	LZR
Yonne	YON	Pyrénées-Orientales	PY-O
Allier	ALL	Alpes de Hte Provence	AL-P
Cantal	CTL	Hautes-Alpes	H-AL
Haute-Loire	H-LR	Alpes Maritimes	AL-M
Fuy-de-Dôme	PdD	Bouches-du-Rhône	BdR
Ain	AIN	Corse	COR
Ardèche	ADC	Var	VAR
Drôme	DRM	Vaucluse	VAU
Isère	ISR		

ITALIE - Dans les graphiques en couleur les symboles et abréviations sont en vert.

Regioni sont désignées par 3 lettres précédées d'un losange plein ◆ pour les graphiques en couleur, d'un losange barré d'un trait horizontal ◊ dans les graphiques en noir.

Piemonte	PIE	Umbria	UMB
Valle d'Aosta	AOS	Lazio	LAZ
Liguria	LIG	Campania	CAM
Lombardia	LOM	Abruzzi	ABR
Trentino-Alto Adige	TRE	Molise	MOL
Veneto	VEN	Puglia	PUG
Friuli Venezia	FRI	Basilicata	BAS
Emilia Romagna	EMI	Calabria	CAL
Marche	MAR	Sicilia	SIC
Toscana	TOS	Sardegna	SRD

LISTE DES CIRCONSCRIPTIONS FIGURANT UNIQUEMENT
DANS L'ANALYSE AGRICOLE

ITALIE - Dans les graphiques en couleur les symboles et abréviations sont en vert.

Zones altimétriques sont désignées par 3 lettres, les 2 premières en majuscules (1) sont propres aux "regioni" suivies d'une minuscule désignant la zone altimétrique : m = montagne, c = colline, p = plaine. Ces 3 lettres sont précédées d'un triangle base en bas ▲

Piemonte montagne	PMm	Lazio montagne	LZm
Piemonte colline	PMc	Lazio colline	LZc
Piemonte plaine	PMp	Lazio plaine	LZp
Valle d'Aosta montagne	AOS	Campania montagne	CPm
Liguria montagne	LGM	Campania colline	CPc
Liguria colline	LGc	Campania plaine	CPp
Lombardia montagne	LBm	Abruzzi montagne	ABm
Lombardia colline	LBC	Abruzzi colline	ABc
Lombardia plaine	LBp	Molise montagne	MOM
Trentino Alto Adige montagne	TRm	Molise colline	MOC
Veneto montagne	VNm	Puglia montagne	PGm
Veneto colline	VNc	Puglia colline	PGc
Veneto plaine	VNp	Puglia plaine	PGp
Friuli montagne	FRm	Basilicata montagne	BAM
Friuli colline	FRc	Basilicata colline	BAC
Friuli plaine	FRp	Basilicata plaine	BAP
Emilia montagne	EMm	Calabria montagne	CLm
Emilia colline	EMc	Calabria colline	CLc
Emilia plaine	EMp	Calabria plaine	CLp
Marche montagne	MRm	Sicilia montagne	SCm
Marche colline	MRC	Sicilia colline	SCc
Toscana montagne	TOm	Sicilia plaine	SCP
Toscana colline	TOc	Sardegna montagne	SDm
Toscana plaine	TOP	Sardegna colline	SDc
Umbria montagne	UMm	Sardegna plaine	SDp
Umbria colline	UMc		

(1) Sauf pour Valle d'Aosta : AOS

LISTE DES CIRCONSCRIPTIONS FIGURANT UNIQUEMENT
DANS L'ANALYSE GENERALE

ITALIE - Dans les graphiques en couleur les symboles et abréviations sont en vert.

Provincie sont désignées par 2 lettres majuscules (1) précédées d'un triangle pointe en bas ▼

PIEMONTE

Torino TO
Vercelli VC
Novara NO
Cuneo CN
Asti AT
Alessandria AL

LOMBARDIA

Varese VA
Como CO
Sondrio SO
Milano MI
Bergamo BG
Brescia BS
Pavia PV
Cremona CR
Mantova MN

TRENTINO-ALTO ADIGE

Bolzano-Bozen BZ
Trento TN

VENETO

Verona VR
Vicenza VI
Belluno BL
Treviso TV
Venezia VE
Padova PD
Rovigo RO

FRIULI-VENEZIA-GIULIA

Pordenone PN
Udine UD
Gorizia GO
Trieste TS

(1) Sauf pour ROMA, Aosta = AOS, Molise = MOL

ITALIE - Provincie (suite)

LIGURIA

Imperia IM
Savona SV
Genova GE
La Spezia SP

MARCHE

Pesaro e urbino PS
Ancona AN
Macerata MC
Ascoli Piceno AP

EMILIA-ROMAGNA

Piacenza PC
Parma PR
Reggio nell'Emilia RE
Modena MO
Bologna BO
Ferrara FE
Ravenna RA
Forli FO

LAZIO

Viterbo VT
Rieti RI
Roma ROMA
Latina LT
Frosinone FR

TOSCANA

Massa-Carrara MS
Lucca LU
Postoia PT
Firenze FI
Livorno LI
Pisa PI
Arezzo AR
Siena SI
Grosseto GR

ABRUZZI

L'Aquila AQ
Teramo TE
Pescara PE
Chieti CH

Molise ⁽¹⁾ MOL

CAMPANIA

Caserta CE
Benevento BN
Napoli NA
Avellino AV
Salerno SA

UMBRIA

Perugia PG
Terni TR

(1) Regroupe Isernia et Campobasso (cf. page 42, point 2).

ITALIE - Province (suite)

PUGLIA

Foggia	FG
Bari	BA
Taranto	TA
Brindisi	BR
Lecce	LE

BASILICATA

Potenza	PZ
Matera	MT

CALABRIA

Cosenza	CS
Catanzaro	CZ
Reggio di Calabria	RC

SICILIA

Trapani	TP
Palermo	PA
Messina	ME
Agrigento	AG
Caltanissetta	CL
Enna	EN
Catania	CT
Ragusa	RG
Siracusa	SR

SARDEGNA

Sassari	SS
Nuori	NU
Cagliari	CA

VALLE D'AOSTA

AOS

SIGNIFICATION des SYMBOLES et ABBREVIATIONS UTILISÉS pour l'ANALYSE AGRICOLE
(48 VARIABLES)

Graphiques 1 à 9

REPARTITION de l'UTILISATION du SOL par CLASSE de SAUtilisée

- t1 1 à 6 : terres labourables
- cer 1 à 6 : terres cultivées en céréales
- sth 1 à 6 : surfaces toujours en herbe
- perm 1 à 6 : cultures permanentes

REPARTITION des CHEFS d'EXPLOITATIONS AVEC ou SANS SUCESSEUR POTENTIEL par CLASSE de SAUtilisée

- Eas 1 à 6 : Exploitations dont le chef est âgé de plus de 50 ans avec successeur potentiel
- Ess 1 à 6 : Exploitations dont le chef est âgé de plus de 50 ans sans successeur potentiel
- Ev 1 à 6 : Exploitations dont le chef a plus de 50 ans avec ou sans successeur potentiel (Ev = Eas+Ess)
- Ej 1 à 6 : Exploitations dont le chef a moins de 50 ans sans qu'intervienne la notion de successeur

CLASSE de SUPERFICIE (SAUtilisée)

- 1 : moins de 2 ha
- 2 : 2 à 5 ha
- 3 : 5 à 10 ha
- 4 : 10 à 20 ha
- 5 : 20 à 50 ha
- 6 : plus de 50 ha

SIGNIFICATION des SYMBOLES et ABREVIATIONS UTILISÉS pour l'ANALYSE ECONOMIQUE GENERALE
(16 et 25 VARIABLES)

Graphiques 10 à 19

POPULATION, MAIN D'OEUVRE

- ▲ { Ptot : population totale
Pat : population active totale
Paa : population active agricole
Pai : population active industrie
Pas : population active services
d : densité de population au km²
- ⊙ S : superficie totale de la région

VARIABLES MONETAIRES

- ◆ { Rt : produit intérieur brut (aux coûts des facteurs)
Ra : valeur ajoutée de l'agriculture
Ri : valeur ajoutée de l'industrie
Rs : valeur ajoutée des services
- { Pr : productivité toutes branches
Pra : productivité agricole
Pri : productivité industrie
Prs : productivité services
Prna : productivité non agricole (Pri + Prs)

REPARTITION du NOMBRE d'EXPLOITATIONS par CLASSE de SUPERFICIE

- ⊙ n₁ à 6 1 : moins de 2 ha 4 : 10 à 20 ha
 2 : 2 à 5 ha 5 : 20 à 50 ha
 3 : 5 à 10 ha 6 : plus de 50 ha

REPARTITION de l'UTILISATION du SOL

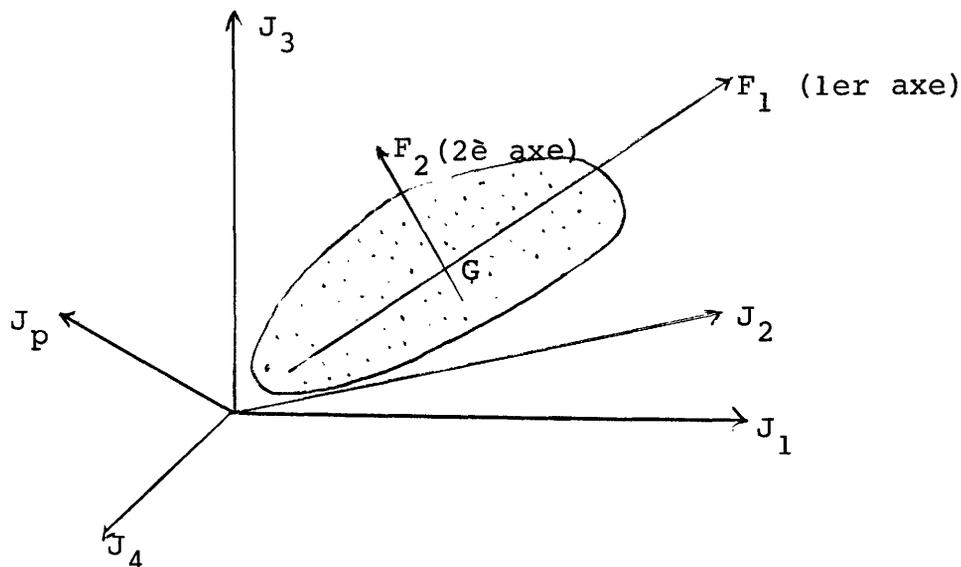
- { t1 : terres labourables
sth : surfaces toujours en herbe
perm : cultures permanentes

CHAPITRE III - PRESENTATION DES RESULTATS

Avant d'indiquer à quels résultats nous sommes parvenus avec l'analyse faisant appel aux seules variables agricoles et avec celle utilisant les variables générales, il s'avère important de présenter sommairement comment s'interprète une analyse factorielle.

Interprétation d'une analyse factorielle

Du tableau de données de départ, de I observations par J variables, il est possible de faire une représentation géométrique, comme on l'a vu dans les principes de l'analyse des correspondances. Plaçons-nous dans l'espace des variables R^p à p dimensions, où figurent les n points d'observations (cf. figure ci-dessous).



Le nuage de points, matérialisé ici dans un plan par une forme ellipsoïdale, a dans l'espace à p dimensions une certaine structure. L'analyse factorielle, tant en composantes principales, que des correspondances, consiste à déterminer les directions privilégiées d'allongement de ce nuage. On cherche successivement selon quels axes se déforme cet ensemble de points. Chacun d'eux étant affecté d'une masse, cela revient à "expliquer", au sens statistique, les dépendances entre les divers points du nuage regroupés autour de leur barycentre G .

La direction d'allongement maximum, déterminée, à l'aide d'une méthode du type des moindres carrés, correspond au premier facteur F_1 . Cet axe étant connu, on procède à l'extraction du second facteur F_2 , deuxième direction selon laquelle s'étire le nuage, puis F_3 , ... Les axes factoriels successifs sont orthogonaux deux à deux. Le nombre de facteurs extraits est généralement de l'ordre de 4 ou 5.

On ne prend en considération le $n^{\text{ième}}$ facteur que dans la mesure où il est jugé significatif ; c'est-à-dire s'il explique une part relativement importante de l'inertie totale du nuage de points. Dans le cas d'un tableau de fréquences le test du chi carré (χ^2) permet de voir à partir de quel moment la part d'inertie expliquée par un facteur a de grandes chances d'être du même ordre de grandeur que ce qui est de nature aléatoire. Dans les autres cas, on a proposé le recours à des méthodes de simulation par génération de tableaux de données en recourant à des tables de nombres au hasard et en extrayant des axes dont les caractéristiques sont comparées à celles des axes issus du tableau initial. L'inconvénient de ce procédé tient à son coût puisqu'il faut une vingtaine de passages si l'on souhaite tester au seuil de 5 % la validité d'un facteur !

En termes mathématiques, l'extraction des facteurs revient à définir une base orthonormée d'un sous-espace vectoriel à q dimensions. Cette base est constituée par les

q vecteurs propres qui correspondent aux q plus grandes valeurs propres de la matrice de départ ayant subi une transformation (1).

Pratiquement, on conduit l'interprétation en se référant aux projections des points sur les plans F_1-F_2 , F_2-F_3 , F_1-F_3 , ... Sur les graphiques ainsi obtenus, on voit apparaître des associations de points qui matérialisent les proximités mathématiques entre observations et variables. Mais quand on raisonne dans le plan F_1-F_2 par exemple, il se peut que deux individus soient proches seulement en apparence. Il faut voir si ces deux individus sont bien représentés dans le plan, ou si l'on a une proximité des seules projections.

Pour déterminer si une observation ou une variable "appartiennent bien" en réalité au plan étudié, on se réfère, à la notion de contribution relative (2). Celle-ci s'analyse comme étant égale au cosinus carré de l'angle formé par le vecteur joignant le point considéré au barycentre et le vecteur axial étudié. Cette grandeur n'est rien d'autre qu'un coefficient de corrélation. Ainsi, une contribution relative forte d'un individu par rapport au premier facteur, permet de dire que ce point est bien représenté sur ce facteur. Il est évident que l'on ne peut pas enregistrer une forte contribution relative de ce même point par rapport au deuxième facteur ou un autre, puisque par définition la valeur absolue d'un coefficient de relation est comprise entre zéro et un.

Mais il est également très intéressant pour l'interprétation des facteurs d'avoir recours à la notion de contribution absolue (2). La contribution absolue d'un point correspond à la fraction de la variance de ce point dans l'ensemble des observations et des variables. Un individu ayant une forte contribution absolue a un grand poids dans l'analyse. S'il intervient dans une proportion importante par rapport à un axe, cela signifie que cet élément joue un rôle dans la définition du facteur.

Notons que la somme des contributions absolues des points pour un facteur équivaut à la valeur propre (2) de ce

- (1) A partir de la matrice de départ, on calcule une matrice de variance-covariance selon la métrique du Chi carré dont on extrait les valeurs propres.
- (2) Pour plus de détails voir Annexe 1.

facteur. En effet, le rapport entre la variance absorbée par un axe et la variance totale est égale à la valeur propre de cet axe. Ainsi, concrètement, une forte valeur propre indique que le facteur concerné explique une forte portion de la variance du nuage. Cela se traduit par un axe sur lequel on a un fort étirement des points, les points les plus significatifs étant ceux dont la variance est forte. Lorsqu'on dit qu'un facteur explique une forte proportion de l'inertie totale (60 % par exemple) ceci veut dire que le rapport entre la valeur propre de ce facteur et la somme des valeurs propres (trace de la matrice) est de 60 %.

Mais, un point peut être caractérisé par une contribution absolue élevée sur un facteur, tandis que sa contribution relative est faible. En effet, un élément qui joue un grand rôle dans l'analyse peut intervenir dans la définition de plusieurs axes, donc avec une forte contribution absolue sur chacun d'eux. Par contre, sa contribution relative ne pourra être élevée que par rapport à un seul facteur. Inversement, il se peut qu'une variable jouant un rôle négligeable dans l'analyse, donc avec une contribution absolue très peu importante, soit malgré tout fortement corrélée à un axe.

Pour terminer, rappelons que la méthode utilisée faisant appel à la distance du chi carré et non à une distance euclidienne, on raisonne en terme de profil. Si une région a un profil identique au profil moyen de l'ensemble des régions pour les $p-1$ variables qui servent à définir les observations, elle apparaîtra caractérisée par la $p^{\text{ième}}$ variable. A condition bien sûr, que cette $p^{\text{ième}}$ variable ait une influence dans l'analyse. Cette région n'est donc pas identifiée au moyen de ses paramètres les plus évidents mais à partir de celui, ou de ceux, qui s'écartent de façon sensible de la norme. Ainsi supposons une région où domine l'agriculture familiale avec système de polyculture élevage classique, mais où la production de fraises, à condition que cette variable soit retenue, est nettement plus importante que dans les autres régions. Si la variable fraises n'a pas une contribution

absolue négligeable, on verra cette région caractérisée par la production de fraises, et non par le critère polyculture-élevage. Dans l'analyse, la variable production de fraises est le critère le plus important pour la région, ceci de façon relative, lorsqu'on se place par rapport à l'ensemble des autres circonscriptions. Ce n'est pas la polyculture, trait dominant à l'intérieur de la région, mais qui ne s'écarte pas du profil moyen de très nombreuses zones, qui identifie cette circonscription.

A - Analyse basée sur les seules variables agricoles

Dans le second Chapitre, on a longuement précisé les variables et les observations à la base de cette analyse. Rappelons seulement que les circonscriptions prises en compte pour l'Italie sont constituées par les zones altimétriques. Les 48 variables, exprimées sous forme d'hectares, concernent la présence ou non d'un successeur, l'âge du chef d'exploitation (jeune ou vieux) et le système de production. Tous ces critères concernent chacune des six classes de surface retenues.

Les caractéristiques générales de l'analyse figurent dans le tableau ci-dessous rassemblant les valeurs propres et le pourcentage d'inertie expliqué par chacun des 5 premiers facteurs, ainsi que le pourcent d'inertie cumulée.

	1er axe	2è axe	3è axe	4è axe	5è axe
valeur propre	0,2769	0,1801	0,098	0,054	0,037
% d'inertie du facteur	35,77	23,27	12,64	7,00	4,85
% inertie cumulée	35,77	59,04	71,68	78,68	83,53

Deux points sont à souligner à la lecture de ce tableau. D'une part, on remarque des valeurs propres d'un niveau élevé, plus particulièrement en ce qui concerne les

deux premiers axes. A titre de comparaison on peut indiquer que dans l'analyse 21 régions françaises (voir annexe 2 - 1ère analyse), la meilleure valeur propre approchait 0,20 alors que dans les analyses communautaires 94 circonscriptions (voir annexe 2 - 2ème analyse), on avait aux environs de 0,16. Les valeurs propres ci-dessus (0,27 pour le premier facteur et 0,18 pour le second) apparaissent d'autant plus élevées que, d'une façon générale, leur niveau a tendance à décroître quand la taille du tableau traité augmente. On peut penser que ces résultats proviennent du choix des variables qui sont très discriminantes. Il est probable que la présentation des critères par classes de surface joue à cet égard un rôle prépondérant.

D'autre part, on constate que les deux premiers facteurs sont assez équilibrés, l'un expliquant plus du tiers de l'inertie totale du nuage, et l'autre moins du quart. A eux deux, ces axes fournissent "l'explication" de près de 60 % de l'inertie totale du nuage de points. A partir du 3è axe, les pourcentages d'inertie diminuent notablement. Il semble que l'on puisse se contenter de 4 ou 5 facteurs dans l'explication du phénomène puisqu'on se situe aux environs de 80 % d'inertie totale expliquée, à ce stade.

La présentation des résultats se fera par l'étude de chacun des facteurs. En vue de dégager leur signification on se réfèrera d'abord aux contributions absolues les plus fortes par rapport à l'axe étudié des variables ainsi que des observations. On complètera ceci par la présentation des points dont la contribution relative élevée signifie leur appartenance au phénomène décrit par le facteur. On a vu que les sous-ensembles définis à l'aide des deux types de contribution peuvent à la limite être adjoints. En fait, il y a des intersections.

1. Premier facteur (voir Graphique n° 1, p.76 bis)

a - Signification : analyse des contributions absolues

Sur cet axe l'analyse des contributions absolues révèle une composition entre les variables concernant :

- d'une part, les exploitations de plus de 50 ha et de 20 à 50 ha. Pour les premières, tous les critères démographiques sont représentés, à l'exception de Ess (sans successeur). Pour les secondes, seuls Ess et Ej (moins de 50 ans) apparaissent. En ce qui concerne le système de production, dans le premier cas on a sth, tl et céréales; dans l'autre cas, on a uniquement sth et tl.
- et, d'autre part, les exploitations de moins de 10 ha. Pour la catégorie des moins de 5 ha se manifeste la variable Ess, tandis que Eas apparaît dans la classe 3 (5 à 10 ha). Quant aux caractères Ej et Ev, ils sont présents dans les classes 2 et 3. Le système de production est représenté par tl en 1, 2 et 3, céréales en 2 et 3 et cultures permanentes en 2 (de 2 à 5 ha).

L'opposition entre les observations a trait :

- aux régions telles que Aisne, Oise, Seine et Marne, Cher, Eure et Loir, Marne, Côte d'Or, Yonne, Schleswig-Holstein, Sardagna (1), Somme, Nièvre, Allier, Sardagna montagne, Sardagna colline, Meuse, Haute-Marne, Ardennes, Eure, Loiret ... Toutes ces circonscriptions sont associées aux variables grandes exploitations.
- aux régions comme Piemonte colline, Campania colline, Piemonte, Veneto, Campania, Abruzzi, Marche-colline, Abruzzi colline, Marche, Emilia, Sicilia, Sicilia colline, Veneto plaine ... Toutes ces circonscriptions correspondent aux critères relatifs aux exploitations de moins de 10 ha.

Ainsi, grossièrement, on peut dire que ce premier facteur associe, d'une part, les zones de grande culture avec terres labourables et céréales, ainsi que les zones de

(1) Lorsque figure pour les circonscriptions italiennes un intitulé tel que celui-ci, Sardagna, cela signifie que la circonscription concernée est au niveau des regioni. Pour la zone altimétrique, on parlera de Sardagna colline, ou montagne, par exemple.

type extensif avec les surfaces toujours en herbe. D'autre part, se trouvent associées les zones de petites exploitations essentiellement méditerranéennes avec un système de production où figurent outre les terres labourables et céréales, les cultures permanentes. Les plus grandes exploitations, où sont représentés à la fois les exploitants de plus de 50 ans et les autres, sont caractérisées par la présence de successeurs potentiels, ce qui est assez naturel. Par contre, les exploitations entre 20 et 50 ha ont un profil plutôt marqué par la coexistence des critères de chefs d'exploitation de moins de 50 ans et l'absence de successeur. A l'opposé, dans les petites exploitations, celles de moins de 5 ha, sont sans successeur alors que celles entre 5 et 10 en ont un.

En grossissant le trait, on voit que le premier facteur différencie les zones à prédominance de grandes exploitations avec successeur, des zones à petites exploitations sans successeur. Bien sûr, cela est vrai si on considère seulement les extrêmes puisque des cas intermédiaires se manifestent également.

b - Les contributions relatives des observations

Le relevé des contributions relatives des variables n'est pas indispensable. Celui des observations (circonscriptions) suffit. Il permet de voir comment se classent, outre les régions déjà citées, certaines circonscriptions dont les contributions absolues sont minimales. Nous indiquons ici l'ensemble des zones assez fortement corrélées au premier axe, même si elles ont été nommées précédemment dans l'étude de la signification du facteur.

Par ordre décroissant, on relève du côté des petites exploitations, les circonscriptions suivantes : Abruzzi colline, Piemonte colline, Veneto, Lombardia colline, Veneto colline, Friuli plaine, Campania plaine, Friuli, Toscana montagne, Piemonte, Emilia montagne, Campania colline, Campania, Veneto plaine, Friuli colline, Molise montagne, Molise, Puglia montagne, Emilia, Abruzzi, Marche, Marche colline, Basilicata montagne, Südbaden, Emilia colline, Sicilia colline, Koblenz-Montabaur, Trier, Toscana plaine, Umbria, Umbria montagne,

Lazio montagne, Calabria colline, Marche montagne, Sicilia colline, Emilia plaine, Umbria colline, Lazio colline, Molise colline. Les contributions relatives varient de 0,75 pour Abruzzi colline, à 0,40 pour Molise colline.

Pour ce qui est des grandes exploitations, on trouve les régions : Allier, Haute-Marne, Eure, Ardennes, Somme, Meurthe-et-Moselle, Meuse, Moselle, Seine-Maritime, Indre, Côte d'Or, Yonne, Nièvre, Loir et Cher, Vienne, Cher, Indre et Loir, Schleswig-Holstein, Stade, Aisne, Loiret, Oise, Aube, Aveyron, Eure et Loir, Alpes de Hte Provence, Marne, Seine et Marne, Saône et Loire, Haute-Garonne, Seine. La corrélation entre ces points observations et le premier axe est plus élevée, variant de 0,84 pour Allier à 0,50 pour la Seine.

2. Deuxième facteur

a - Signification : analyse des contributions absolues

Pour ce qui est des variables, on enregistre un contraste entre :

- d'une part, essentiellement les exploitations de plus de 50 ha où tous les caractères de nature démographique sont représentés, sauf la présence de successeur (variable Eas). Le système de production est présent sous forme des variables Sth, tl, céréales. Il apparaît aussi la variable cultures permanentes dans la catégorie d'exploitations de 2 à 5 ha;
- d'autre part, les exploitations de 10 à 20 ha avec Ej (exploitants de moins de 50 ans) tl et sth, ainsi que les exploitations de 20 à 50 ha avec également Ej, tl et sth.

Le contraste entre observations est relatif aux circonscriptions :

- Sardegna, Puglia, Sicilia, Seine, Marne, Sardegna colline, Puglia plaine, Campania, Lazio, Sicilia colline, Eure et Loir, Aisne, Seine et Marne, Campania colline, Cher, Oise, Lazio colline, Côte d'Or, Puglia colline, Calabria colline, Abruzzi montagne... Ces zones sont associées au premier groupe de variables (plus de 50 ha et 2 à 5 ha).

- Schwaben, Ille-et-Vilaine, Côtes-du-Nord, Manche, Finistère, Mayenne, Loire-Atlantique, Morbihan, Maine et Loire, Niederbayern, Oberpfalz, Vendée, Puy de Dôme, Osnabruck, Nord, Oberfranken, Basses-Pyrénées, Haute-Loire, Münster, Oldenburg, Dordogne... Toutes ces circonscriptions se situent avec les exploitations de 10 à 50 ha.

La partie négative de ce deuxième facteur est très proche de la partie négative du premier axe. Ainsi, les variables E_j, E_v, t_l, s_{th} , céréales concernant la catégorie 6 d'exploitations sont communes aux deux facteurs. Le premier possède en plus pour cette catégorie 6 (plus de 50 ha) le caractère E_{as} , tandis que le second a un caractère E_{ss} . Mais pour le premier facteur en partie négative, on a des éléments de la catégorie 5 (de 20 à 50 ha) tandis que pour le deuxième facteur est présente la variable cultures permanentes pour 2 à 5 ha. Parmi les observations communes à ces deux axes relevons Aisne, Oise, Seine, Seine-et-Marne, Cher, Eure et Loir, Marne, Côte d'Or, Sardegna.

Par contre, la partie positive de ces deux facteurs diffère sensiblement puisque pour le premier se trouvent uniquement des critères concernant les exploitations de moins de 10 ha, alors que pour le second sont des éléments propres à des exploitations de 10 à 50 ha dont le système de production est basé sur les terres labourables et s_{th} , avec l'apparition du caractère E_j (exploitants de moins de 50 ans).

Ce deuxième facteur présente, de façon plus accentuée que le premier, une différenciation des circonscriptions en fonction de leur taille et cela pour un système de production associant terres labourables, surfaces toujours en herbe, et céréales uniquement dans le cas des grandes superficies. Seule la variable culture permanente dans les exploitations de 2 à 5 ha vient perturber ce schéma.

En négligeant ce dernier aspect, on voit une distinction s'opérer entre, d'une part, les régions de grande culture du type du Bassin Parisien, Champagne, ou les zones très exten-

GRAPHIQUE 1: ANALYSE 48 VARIABLES

(Italie: Reg. + Zones altimétriques)

Plan 1-2

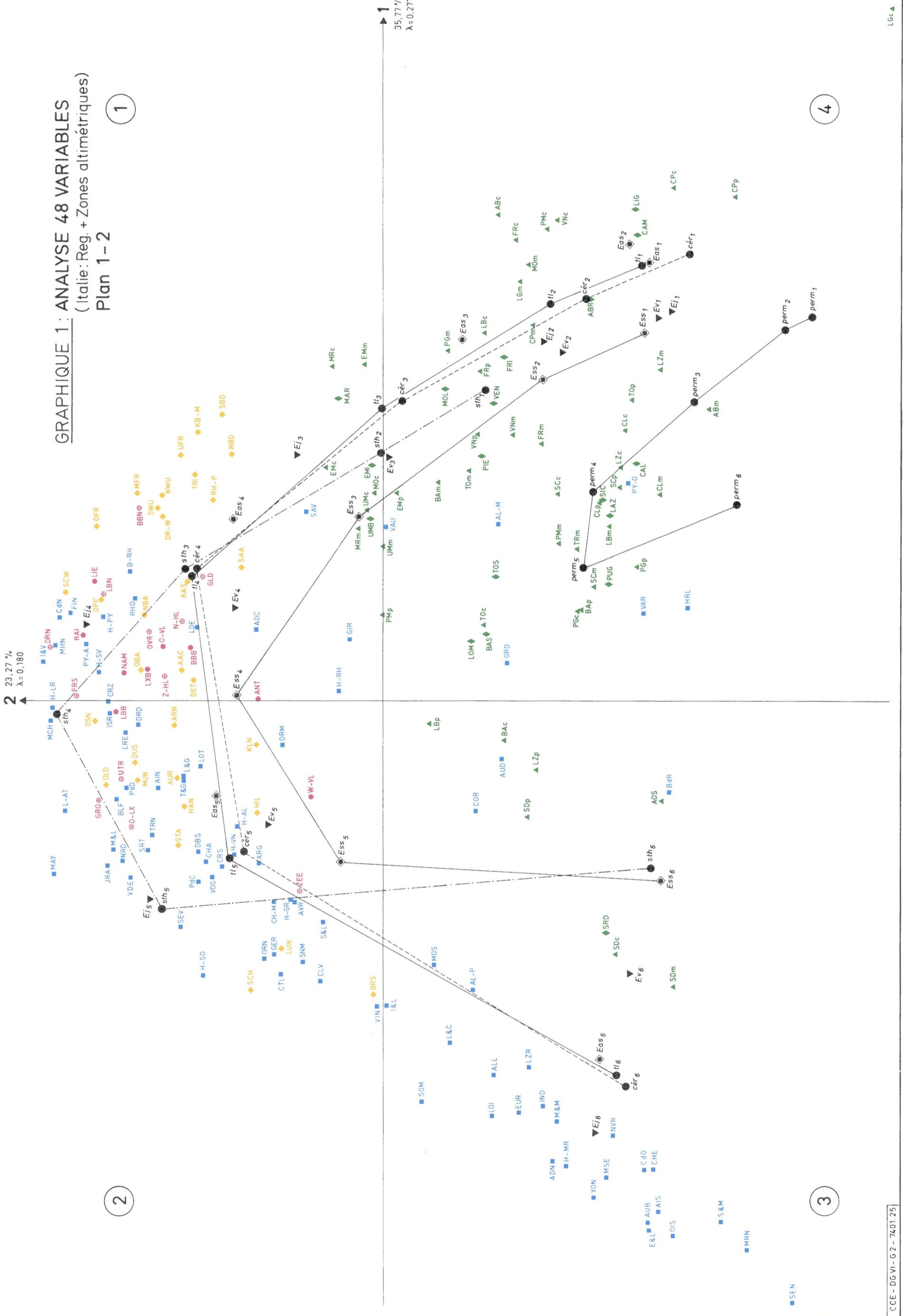
1

2 23,27 %
λ = 0,180

35,77 %
λ = 0,277

2

4



sives comme Sardegnna, Abruzzi montagne et, d'autre part, les zones où prédominent les tailles moyennes avec des terres labourables et des surfaces toujours en herbe. Les plus caractéristiques de ces circonscriptions sont dans l'Ouest de la France ainsi qu'en Allemagne. Par ailleurs, la distinction se fait entre des zones de plus de 50 ha sans successeurs, et des zones caractérisées par la présence d'exploitants jeunes pour les catégories de 10 à 50 ha.

b - Les contributions relatives des observations

Sur la partie négative du deuxième axe, c'est-à-dire avec les grandes exploitations, on trouve par ordre décroissant : Sicilia montagne, Lazio, Calabria montagne, Lazio colline, Puglia colline, Calabria, Puglia, Puglia plaine, Basilicata plaine, Calabria plaine, Calabria colline, Sicilia plaine, Sicilia colline, Basilicata colline, Lazio montagne, Toscana plaine, Toscana. Les contributions relatives s'étagent entre 0,56 pour Sicilia montagne et 0,30 pour Toscana.

Sur la partie positive, avec les zones d'exploitations moyennes se trouvent les circonscriptions : Osnabrück, Isère, Ille-et-Vilaine, Haute-Loire, Pyrénées Atlantique, Dordogne, Morbihan, Arnsberg, Oldenburg, Detmold, Düsseldorf, Loire, Ain, Münster, ~~Niederbayern~~, Corrèze, Oberpfalz, Puy-de-Dôme, Hautes-Pyrénées, Bas-Rhin, Côtes-du-Nord, Finistère, Loire-Atlantique, Lot, Limburg-Nederland, Nord, Zuid-Holland, Schwaben, Mayenne, Stade, Maine et Loire, Oberfranken. La liaison de ces observations avec le deuxième facteur est nettement meilleure que pour les précédentes. Osnabrück a une contribution relative de 0,89 et Oberfranken de 0,50.

Pour compléter l'étude des premier et deuxième facteurs, il est intéressant de se reporter au graphique n° 1 qui donne une représentation du nuage de points dans le plan 1-2.

c - étude du plan 1-2

L'étude du plan 1-2 révèle un phénomène déjà constaté lors des analyses précédentes ; les variables relatives aux classes de surface se répartissent selon une structure

parabolique. Ceci a un caractère général puisque l'on constate la même chose lorsqu'on relie les critères Eas de 1 à 6, Ess de 1 à 6, Ej ... Ev ainsi que les variables décrivant le système de production.

En ce qui concerne les variables démographiques, on note que la parabole des sans successeurs est située vers l'intérieur, tandis que la parabole des exploitants de moins de 50 ans est la plus extérieure. Entre ces deux extrêmes se situent, proches des sans successeurs, les exploitants âgés puis les exploitants avec successeur. Ce schéma est surtout valable pour les classes 4, 5, 6 car on constate souvent au niveau des catégories 1 et 2 un décalage vers l'intérieur du critère chefs d'exploitations de moins de 50 ans. Ceci peut s'expliquer par les caractéristiques propres des petites exploitations.

L'étude des paraboles obtenues en reliant les variables propres au système de production indique une plus grande diversité dans les structures. A l'extérieur se situe la courbe des surfaces toujours en herbe. Tout à fait à l'intérieur, très détachée, se trouve la courbe des cultures permanentes. Cette parabole est très différente des autres puisqu'elle est entièrement comprise dans un quadrant, tandis que les autres recourent les quatre quadrants. Par ailleurs, cette courbe très fermée se situe presque uniquement dans la zone regroupant les circonscriptions méditerranéennes. Ceci explique l'important décalage constaté entre la variable cultures permanentes des exploitations de plus de 50 ha et toutes les autres variables de cette même classe. En effet les cultures permanentes caractérisent surtout les zones méditerranéennes qui sont fortement attirées par les variables relatives aux petites exploitations. Ainsi se constitue une parabole cultures permanentes propre au sous-ensemble zone méditerranéenne. Cette zone ayant pour caractère principal la prédominance de petites exploitations, la variable cultures permanentes sur plus de 50 ha voisine avec l'ensemble des variables propres à la classe moins de 5 ha.

Un autre point mérité d'être souligné. Les courbes terres labourables et céréales sont quasiment identiques. Cela signifie que la distribution du nuage de points sur notre graphique n'aurait pas été modifiée si l'on avait eu recours au seul critère terres labourables, plus large que céréales, puisque ces dernières entrent dans la composition des terres labourables. Ce point est important car dans l'analyse privilégiant les variables économiques générales, on a introduit le système de production à l'aide des surfaces toujours en herbe, des cultures permanentes et des terres labourables. Les superficies céréales n'ont pu être incluses, les données n'étant pas disponibles. La très grande similitude des paraboles terres labourables et céréales nous incite à penser que l'absence des surfaces en céréales, dans l'analyse générale, ne se traduit pas par une perte d'information importante.

Si l'on excepte les variables cultures permanentes, on constate que les deux branches des paraboles correspondent respectivement aux petites exploitations de la zone méditerranéenne avec Liguria, Toscana, Basilicata, Calabria, Lazio, Pyrénées-Orientales, Campania... et aux exploitations de grande culture de Franche-Comté, Champagne, Centre... Au sommet de la parabole se trouvent les provinces du Bénélux, l'Allemagne et les régions de l'Ouest de la France.

3. Troisième facteur (voir Graphique n° 2, p. 80 bis)

a - Signification : analyse des contributions absolues

- Dans l'étude des variables, l'opposition se fait entre :
- d'une part, le système de production céréales sur les exploitations de plus de 50 ha, et les terres labourables dans la catégorie 10-20 ha, 20-50 ha et plus de 50 ha.
 - d'autre part, le système de production surface toujours en herbe dans les classes 10-20 ha, 20-50 ha et plus de 50 ha.

Pour les observations, on voit une distinction entre :

- Marne, Aube, Yonne, circonscriptions associées aux céréales et terres labourables dans les grandes exploitations,
- Sardegna, Sardegna colline, Sardegna montagne, Piemonte montagne, Lombardia montagne, Trentino montagne, Lozère, Abruzzi montagne, Cantal, Orne, Manche, Valle d'Aosta, Veneto montagne, Saône et Loire.

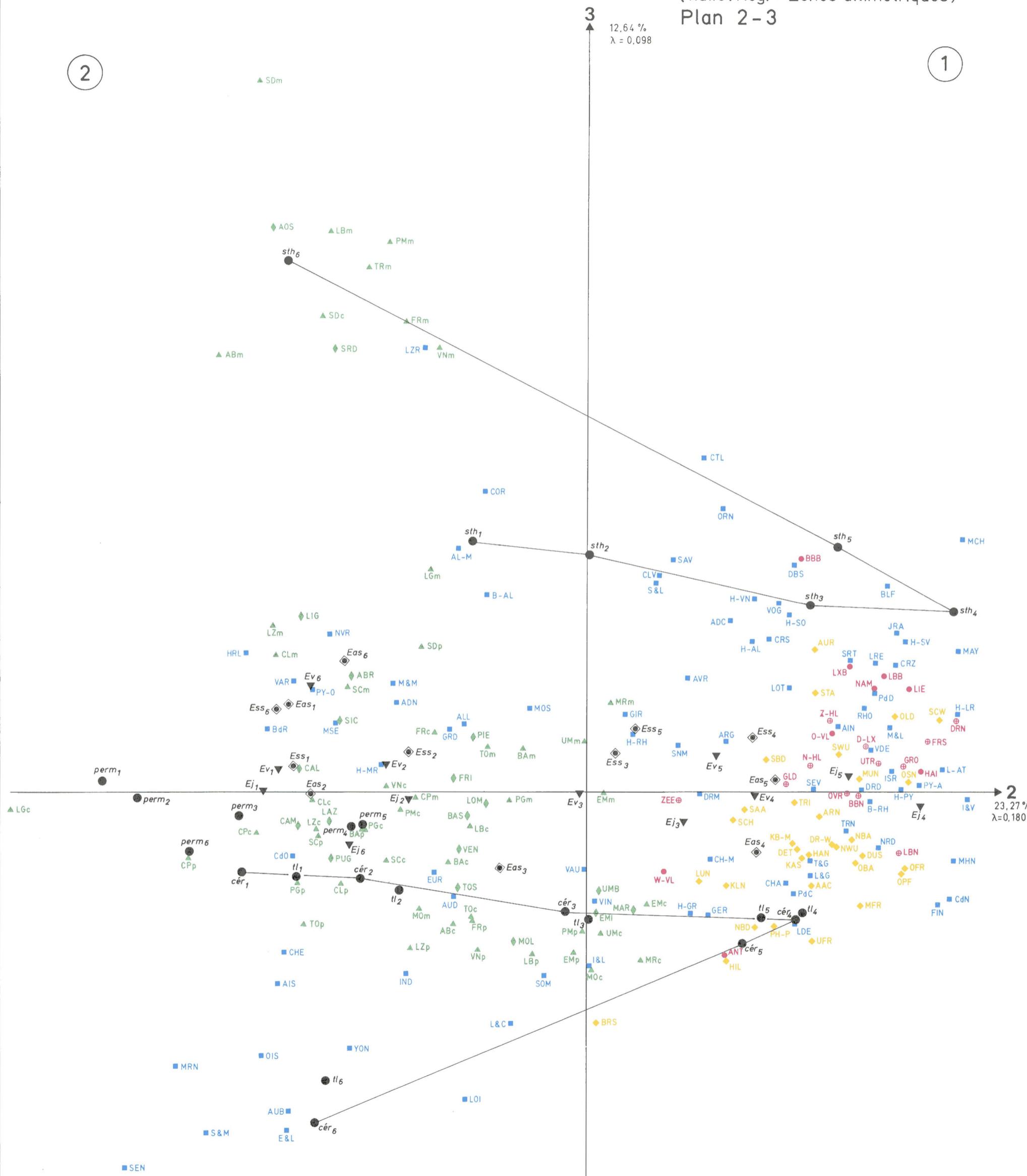
La signification de ce facteur est claire. Il concerne uniquement les grandes exploitations les éclatant selon leur orientation. Dans la partie négative se concentrent les zones de grande culture riches, caractérisées par les variables terres labourables et céréales. Dans la partie positive, prédomine un système extensif basé sur les surfaces toujours en herbe, associant aussi bien les zones de montagnes italiennes ou françaises du Massif Central, que les zones herbagères normandes.

Ainsi, ce troisième axe apparaît comme un dédoublement du premier facteur où l'on retrouvait toutes ces caractéristiques, mais liées ensemble dans une opposition aux petites et moyennes exploitations. Notons néanmoins, à la différence du premier facteur, qu'ici les variables démographiques jouent un rôle négligeable.

b - les contributions relatives

L'étude des contributions relatives des observations sur le 3^e axe permet de voir que les circonscriptions suivantes appartiennent à la catégorie extensive : Sardegna montagne, Valle d'Aosta, Trentino montagne, Piemonte montagne, Sardegna colline, Sardegna, Lozère, Cantal, Lombardia montagne, Saône et Loire, Haute-Vienne, Orne, Veneto montagne, Corse, Hautes-Alpes, Calvados, Vosges, Sardegna plaine, Friuli montagne, Manche, Basses-Alpes, Les corrélations vont de 0,61 pour Sardegna montagne à 0,20 pour les Basses-Alpes.

GRAPHIQUE 2 : ANALYSE 48 VARIABLES
 (Italie : Reg. + Zones altimétriques)
 Plan 2-3



Grâce aux contributions relatives on peut indiquer que les circonscriptions Lombardia plaine, Piemonte plaine, Hildesheim, Braunschweig, Toscana colline, Loiret, Loir et Cher, Lazio appartiennent aux zones baptisées précédemment de "grande culture". Les corrélations s'échelonnent entre 0,49 pour Lombardia plaine et 0,23 pour Lazio plaine. Il est évident que ces régions qui ont des points communs n'appartiennent pas exactement au même système. L'étude du plan 2-3 nous permettra de préciser ce point.

Il convient de remarquer par ailleurs que ce troisième axe donne des précisions non seulement par rapport au premier, mais aussi par rapport au second, puisque ces deux facteurs sont assez proches. La partie négative du deuxième facteur regroupe en effet les variables terres labourables et céréales, ainsi que les surfaces toujours en herbe, l'ensemble de ces critères concernant les grandes surfaces. Mais on a vu aussi la présence des cultures permanentes sur les exploitations de 2 à 5 ha. Le troisième axe opposant les zones de grande culture aux zones extensives permet donc une clarification, et ceci d'autant mieux qu'il n'y a plus aucune interférence avec les cultures spécialisées de petites superficies. C'est à ce titre également que la représentation graphique dans le plan 2-3 présente un intérêt certain.

c - étude du plan 2-3

Dans ce plan qui explique près de 36 % de l'inertie totale de nuage de points, la structure évoquée précédemment se matérialise très clairement :

- en haut à gauche du graphique se trouvent les zones extensives italiennes associées aux surfaces toujours en herbe dans les exploitations supérieures à 50 ha. Ce nuage de points est isolé. Il souligne la particularité des régions qui s'y trouvent.

- en bas à gauche, regroupées autour de terres labourables 6 (tl 6) et céréales 6 (cer 6) apparaissent les départements Seine, Marne, Seine-et-Marne, Oise, Yonne, Aube, Eure-et-Loir, Loiret. Ce nuage symétrique du précédent par rapport au 2^e axe se situe au même emplacement que dans le plan 1-2. Néanmoins, il est plus individualisé.

- plus proches du troisième axe et liées aux variables t1 4 et t1 5 sont les circonscriptions Piemonte plaine, Lombardia plaine, Toscana colline, Lazio plaine, Braunschweig, Hildesheim, Loir et Cher. Dans l'étude des contributions absolues et relatives ce groupe confondu avec le précédent nous avait permis de définir la fraction négative du troisième facteur dite de grande culture. Ici, grâce au graphique du plan 2-3, on a pu distinguer au sein du système grande culture deux sous-systèmes : l'un sur les superficies de plus de 50 ha et propre à la France, et l'autre concernant les exploitations de 20 à 50 ha et plus largement répandu.

Si l'on relie entre eux dans l'ordre de 1 à 6 tous les points relatifs aux céréales, et tous les points relatifs aux surfaces toujours en herbe, on aboutit à deux courbes très semblables à peu près symétriques par rapport à l'axe 2. Notons que si l'on établissait la courbe joignant les variables terres labourables, on obtiendrait une courbe se superposant presque exactement à celle des céréales. Ceci met en évidence l'opposition sur le deuxième facteur entre les sommets des paraboles (partie positive avec sth, terres labourables pour les exploitations entre 10 et 20 ha) et les branches de ces paraboles (partie négative avec sth, céréales et terres labourables sur les exploitations de plus de 50 ha).

La disposition sur le graphique nous permet de distinguer dans les plus grandes exploitations, donc en négatif sur l'axe 2, la variable E_j 6 en-dessous de l'axe, des deux variables E_v 6 et E_{ss} 6, au-dessous de l'axe. Associé aux exploitants jeunes, on trouve un groupe composé de Puglia, Puglia plaine, Sicilia colline, Campania colline, Puglia colline, Lazio, Lazio colline, Puglia colline, Basilicata colline, Basilicata plaine, Calabria plaine, Sicilia plaine et Sicilia colline. Toutes ces circonscriptions bénéficient d'une contribution relative ou absolue assez forte par rapport à la partie négative du deuxième facteur. Accompagnant les exploitants âgés et sans successeur sur plus de 50 ha sont les régions Sicilia montagne, Sicilia et Calabria montagne.

A ces deux groupes situés sur la partie négative du deuxième axe s'opposent essentiellement les régions allemandes, la France de l'Ouest et du Sud-Ouest ainsi que des provinces néerlandaises. Ces zones caractérisées par des superficies entre 10 et 50 ha sont écartelées entre deux pôles. L'un au-dessus du 2^e facteur, est représentatif d'un système plutôt intensif avec la variable terres labourables et les circonscriptions Niederbayern, Oberpfalz, Finistère, Côtes-du-Nord, Morbihan. L'autre, en dessous du deuxième facteur, concerne un système plus extensif avec la variable sth et les circonscriptions Manche, Mayenne, Haute-Savoie. Mais il n'existe pas deux blocs aussi nettement séparés que sur la partie négative de ce deuxième facteur. Beaucoup d'observations sont liées à l'axe lui-même, et donc ne s'en écartent pas, comme Osnabrück, Isère, Munster, Basses-Pyrénées, Dordogne... On peut supposer que ces régions dépendent étroitement des variables démographiques Ej 4 et Ej 5, elles-mêmes très dépendantes du deuxième facteur.

4. Quatrième facteur (Voir Graphique n° 3, p. 84 bis)

a - Signification : analyse des contributions absolues

L'examen des variables montre la présence :

- d'un côté : des surfaces toujours en herbe et terres labourables dans les exploitations entre 2 et 10 ha, ainsi que des surfaces toujours en herbe, des terres labourables et des céréales dans les exploitations de plus de 50 ha,
- de l'autre côté : des cultures permanentes dans l'ensemble des classes de surface, excepté entre 20 et 50 ha.

Pour les observations, on peut opérer une distinction entre :

- les circonscriptions Marche, Lombardia montagne, Piemonte montagne, Marche colline, Abruzzi, Abruzzi montagne, Seine, Piemonte, Veneto montagne associées aux surfaces toujours en herbe et aux terres labourables,
- les circonscriptions Puglia, Sicilia, Puglia plaine, Hérault, Aude, Gard, Puglia colline, Gironde, Pyrénées-Orientales, Sicilia plaine, Vaucluse, Sicilia colline.

L'opposition que l'on constate a trait au système de production plutôt qu'à la taille des exploitations. Il s'agit d'une situation identique à celle observée avec le troisième facteur. Sur ce dernier, la différenciation est entre terres labourables - céréales et surfaces toujours en herbe, alors que ici l'opposition est entre cultures permanentes et autres modes d'occupation du sol. Les cultures permanentes concernent en premier lieu les zones méditerranéennes, tandis que les surfaces toujours en herbe se rencontrent essentiellement dans les zones montagneuses, les caractéristiques terres labourables étant le propre des zones de grande culture.

b - Contributions relatives

Dans la partie négative du quatrième facteur, du côté des surfaces toujours en herbe et des terres labourables, on relève uniquement la région Marche montagne. Sa contribution relative est de 0,24.

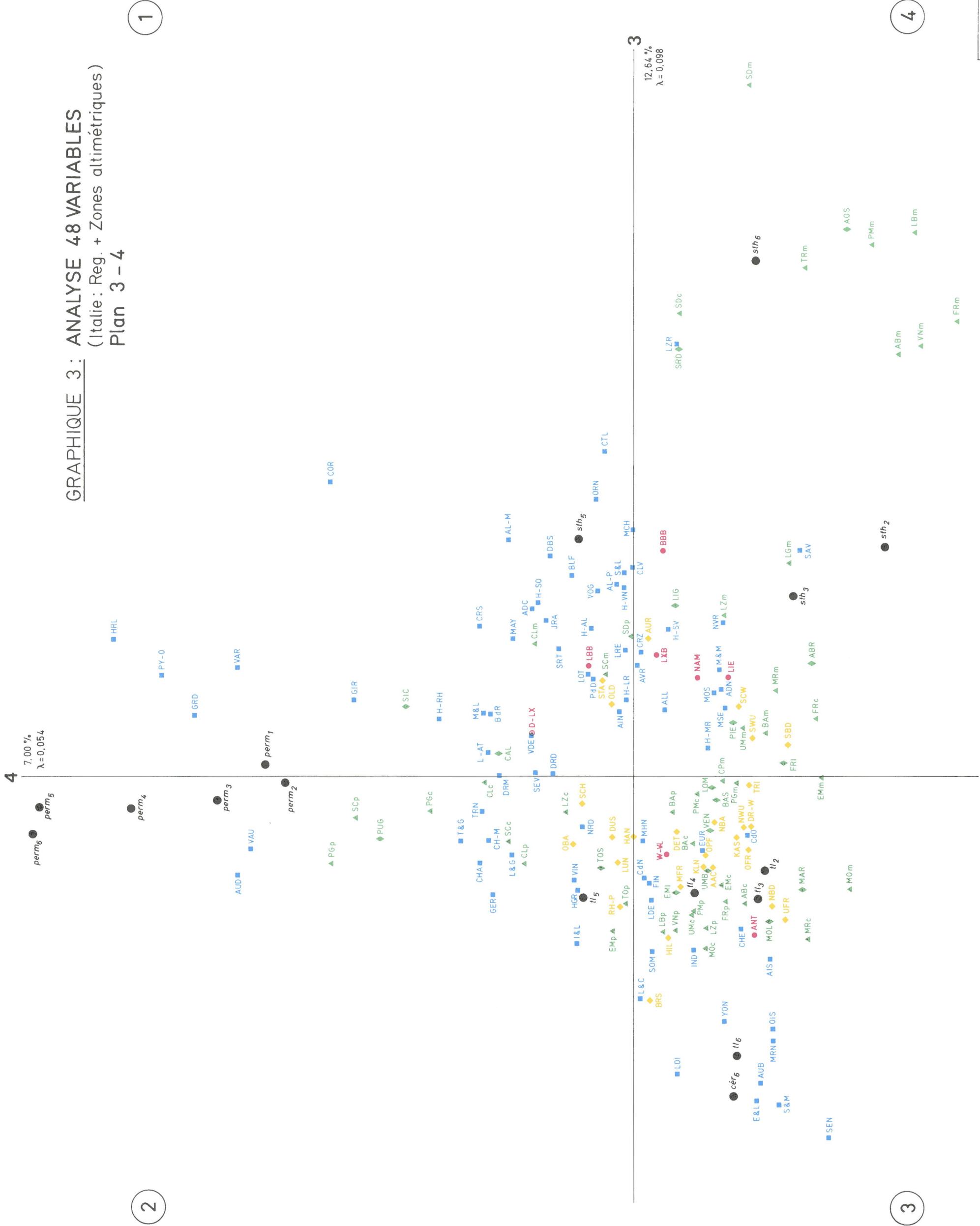
Dans la partie positive, avec les cultures permanentes se trouvent : Aude, Hérault, Gard, Pyrénées-Orientales, Gironde, Var, Puglia plaine, Puglia, Vaucluse, Puglia colline, Sicilia plaine, Corse, Tarn et Garonne, Drôme, Sicilia. Les contributions relatives sont nettement plus élevées, puisque l'on a 0,61 pour Aude et 0,24 pour Sicilia.

L'étude des contributions relatives n'apporte pas de meilleure compréhension à la nature de l'axe 4. L'examen du graphique donnant la représentation du nuage de points dans le plan 3-4 permet par contre de préciser certains aspects.

c - Etude du plan 3-4

Quand on se réfère à un plan n'expliquant qu'une proportion assez faible de l'inertie totale du nuage de points (moins de 20 % pour le plan 3-4), on a une représentation des observations et des variables en forme de sphère autour du centre. Afin d'alléger la représentation graphique, on n'a pas reporté dans le plan 3-4 les observations dont la contribution relative ou absolue par rapport à l'un des deux facteurs était négligeable. De même seules les variables significatives figurent sur le graphique.

GRAPHIQUE 3 : ANALYSE 48 VARIABLES (Italie: Reg. + Zones altimétriques) Plan 3 - 4



Se détachant d'un groupe de points proches du centre, on note plusieurs associations de variables-observations.

En bas à droite, on retrouve le groupe de régions Lozère, Sardegna, Sardegna montagne, Trentino montagne, Valle d'Aosta, Abruzzi montagne, Piemonte montagne, Veneto montagne, Friuli montagne et Lombardia montagne entourant la variable Sth 6.

On a bien là une agriculture de montagne du type alpages où les superficies très grandes sont exploitées de façon très extensive. Ce groupe correspond presque exactement au sous-ensemble noté lors de l'étude du plan 2-3.

Au-dessus de l'axe trois, en partie positive, on observe un rassemblement de nature comparable mais moins individualisé. Il comprend les départements français, Cantal, Orne, Manche, Saône et Loire, Haute-Vienne, Hautes-Alpes, Alpes de Hte Provence et Vosges centrés sur le critère sth 5. L'aspect de culture extensive est toujours présent, mais avec des surfaces moins importantes. Il y a une différence de degré par rapport aux régions de montagne attirées par la variable Sth 6. Ces dernières toutes italiennes sauf Lozère, apparaissent plus défavorisées. Il semble qu'on puisse faire la distinction entre des zones extensives de type alpages et des zones extensives relevant d'un système herbager.

En bas à gauche, symétriques aux zones d'alpage sont les régions de grande culture, déjà vues et qui sont parmi les régions les plus favorisées. On retrouve Marne, Aube, Loiret, Yonne, Eure et Loir, Oise, Seine et Marne, Seine dépendant des variables céréales 6 et terres labourables 6.

En haut, proches du quatrième facteur bien isolées, sont les zones de cultures permanentes avec Hérault, Pyrénées-Orientales, Gard, Aude, Var, Gironde, Vaucluse, Puglia plaine, Puglia colline, Puglia, Sicilia, Sicilia plaine, Drôme, Tarn et Garonne et Corse. On notera que la Corse est à peu près aussi bien représentée sur l'axe 3 du côté des surfaces toujours en herbe (système herbager) que sur l'axe 4 du côté des cultures permanentes.

L'analyse du cinquième facteur ne paraissant pas très intéressante, d'autant plus que cet axe n'explique pas 5 % de l'inertie totale, il est possible de passer à présent à la synthèse des résultats enregistrés.

B - Synthèse des résultats de l'analyse agricole

Pour déterminer aussi précisément que possible les caractéristiques de chacune des observations on a regardé dans quel plan elles étaient le mieux représentées. Etudiant les contributions absolues et relatives de chaque région on a noté par rapport à quel facteur elle était le plus liée. Certaines circonscriptions sont bien représentées sur un facteur et médiocrement sur un ou deux autres. Ou bien elles sont très liées à deux facteurs et accessoirement à un troisième ou à un quatrième. Ainsi le même point peut apparaître à titre principal dans un plan et à titre secondaire dans d'autres plans.

Nous avons ainsi établi toute une série de graphiques n° 4 à 9 (1) correspondant aux plans 1-2, 1-3, 1-4, 2-3, 2-4 et 3-4. Chacun de ces graphes est de lecture aisée car on a fait figurer seulement les points intéressants. Toutes les observations qui ne sont pas dépendantes du système formé par les deux axes du plan étudié sont écartées. De la même manière, on ne trouve dans les graphiques que les variables les plus significatives. De ce fait on aboutit à des associations de points formant des ensembles cohérents et bien individualisés.

Brièvement on peut passer en revue chacun des plans avant de proposer une typologie basée sur les seules variables agricoles.

(1) p. 88, 90, 92, 94, 95 et 96.

1. Les associations significatives de chaque plan

a - Plan 1-2 (Graphique n° 4, p. 83)

Ce plan que l'on a déjà longuement étudié, révèle dans cette nouvelle présentation l'existence de quatre groupes distincts situés chacun dans un quadrant. Dans le premier quadrant (Nord-Est) on a principalement des circonscriptions allemandes et du Bénélux, avec quelques-unes françaises. A ces observations sont associées les variables exploitants de moins de 50 ans sur les exploitations de 5 à 10 ha et de 10 à 20 ha. Quant au système de production, il apparaît sous forme de terres labourables dans la classe 10-20 ha.

Le deuxième quadrant (Nord-Ouest) correspond à un groupe assez semblable : circonscriptions essentiellement françaises et allemandes avec quelques provinces du Bénélux. Ces régions sont associées à la variable exploitants de moins de 50 ans sur les exploitations de 20 à 50 ha. Le système de production est formé de terres labourables et de surfaces toujours en herbe dans la classe 20-50 ha. Par rapport au premier quadrant, ce deuxième groupe se distingue surtout par une taille plus importante, ce qui se traduit aussi au plan des modes d'occupation du sol par l'apparition des surfaces toujours en herbe.

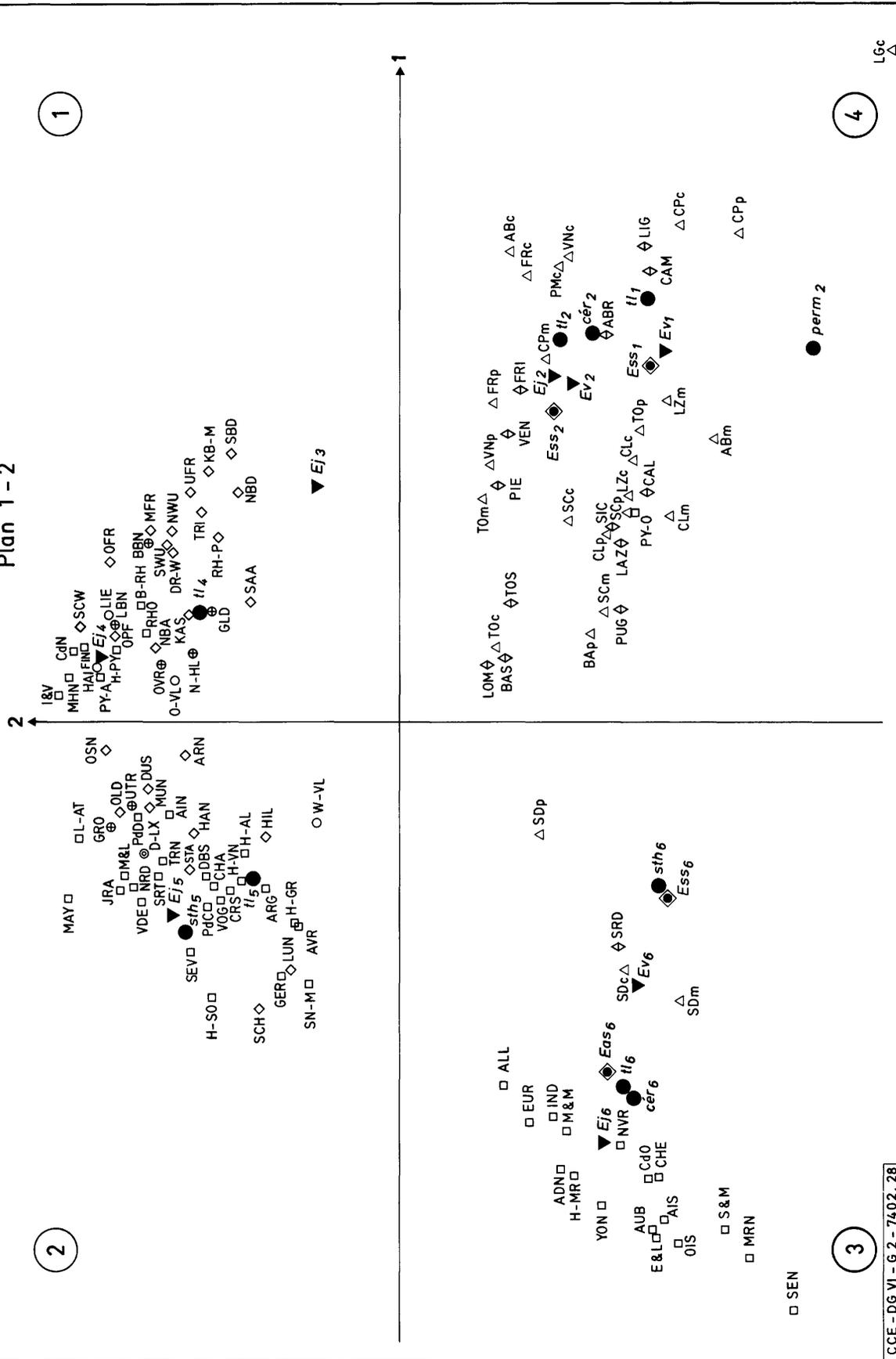
Le troisième quadrant (Sud-Ouest) comprend les départements français de grande culture avec céréales et terres labourables. Ces départements ont des chefs d'exploitation de moins de 50 ans et sont caractérisés par la présence de successeurs. Dans ce même quadrant coexistent les circonscriptions de Sardegna où par opposition aux zones précédentes les exploitants sont âgés et dépourvus de successeurs.

Le quatrième quadrant (Sud-Est) est le plus hétérogène, beaucoup plus que le troisième, même si à l'exception des Pyrénées-Orientales, il ne contient que des circonscriptions italiennes. Il s'agit de régions de petites exploitations dont le système de production repose sur les terres labourables. Les céréales interviennent dans la catégorie de 2 à 5 ha de même que les cultures permanentes. Sur les exploitations de

GRAPHIQUE 4: ANALYSE SIMPLIFIÉE 48 VARIABLES

(Italie: Reg. + Zones altimétriques)

Plan 1 - 2



moins de 2 ha sont des exploitants âgés et sans successeurs, tandis que sur 2 à 5 ha sont à la fois des exploitants jeunes et âgés qui eux aussi sont sans successeur. Pour l'étude plus approfondie de ce groupe, il est intéressant de se référer aux autres plans où certaines régions sont représentées.

b - Plan 1-3 (Graphique n° 5, p.90)

Le premier quadrant comprend les régions Liguria montagne, Alpes-Maritimes, Veneto montagne, Friuli montagne, Abruzzi montagne et Savoie dépendant de la variable surfaces toujours en herbe sur les exploitations de 2 à 5 ha. A l'intérieur de ce quadrant se trouve un sous-groupe composé de Schwaben, Gironde et Sicilia montagne dont le caractère Sth 2 est accessoire. Ces circonscriptions sont mieux représentées dans d'autres plans.

Le deuxième quadrant est de nature semblable avec un premier sous-groupe formé par Calvados, Orne, Vosges, Saône et Loire, Cantal, Haute-Vienne, Aveyron, Alpes de Hte Provence et Hautes-Alpes attirées par la variable surfaces toujours en herbe dans les exploitations de 20 à 50 ha. Lozère, Sardegna et Sardegna colline associées aux surfaces toujours en herbe dans la classe de SAU supérieure à 50 ha, forment un autre sous-groupe. Les autres circonscriptions Creuse, Gironde, Doubs, Jura, Haute-Saône, Puy de Dôme, Sardegna plaine d'une part et Sardegna montagne d'autre part sont également caractérisées par les surfaces toujours en herbe, mais de manière accessoire.

Le troisième quadrant comprend un sous-espace formé de Loir et Cher, Loiret, Somme, Aube, Yonne, Braunschweig caractérisées par les céréales et les terres labourables sur plus de 50 ha. Quant à Indre et Loir, Vienne, Haute-Garonne, Moselle, elles sont associées aux céréales et terres labourables sur les exploitations de 20 à 50 ha. Restent Aisne, Oise, Seine, Seine et Marne, Eure et Loir, Marne ainsi que Gers, Hildesheim et Düsseldorf.

Le quatrième quadrant contient Molise, Emilia, Umbria, Veneto plaine, Friuli plaine, Emilia colline, Emilia plaine, Marche colline, Toscana colline, Umbria colline et Abruzzi colline associées aux céréales et terres labourables sur les exploitations de 5 à 20 ha. Quant à Toscana, Morbihan, Côtes du Nord, Finistère, West-Vlaanderen, Nordwürttemberg, Nordbaden, Oberpfalz, Mittelfranken et Unterfranken leur meilleure représentation est ailleurs.

c - Plan 1-4 (Graphique n° 6, p.92)

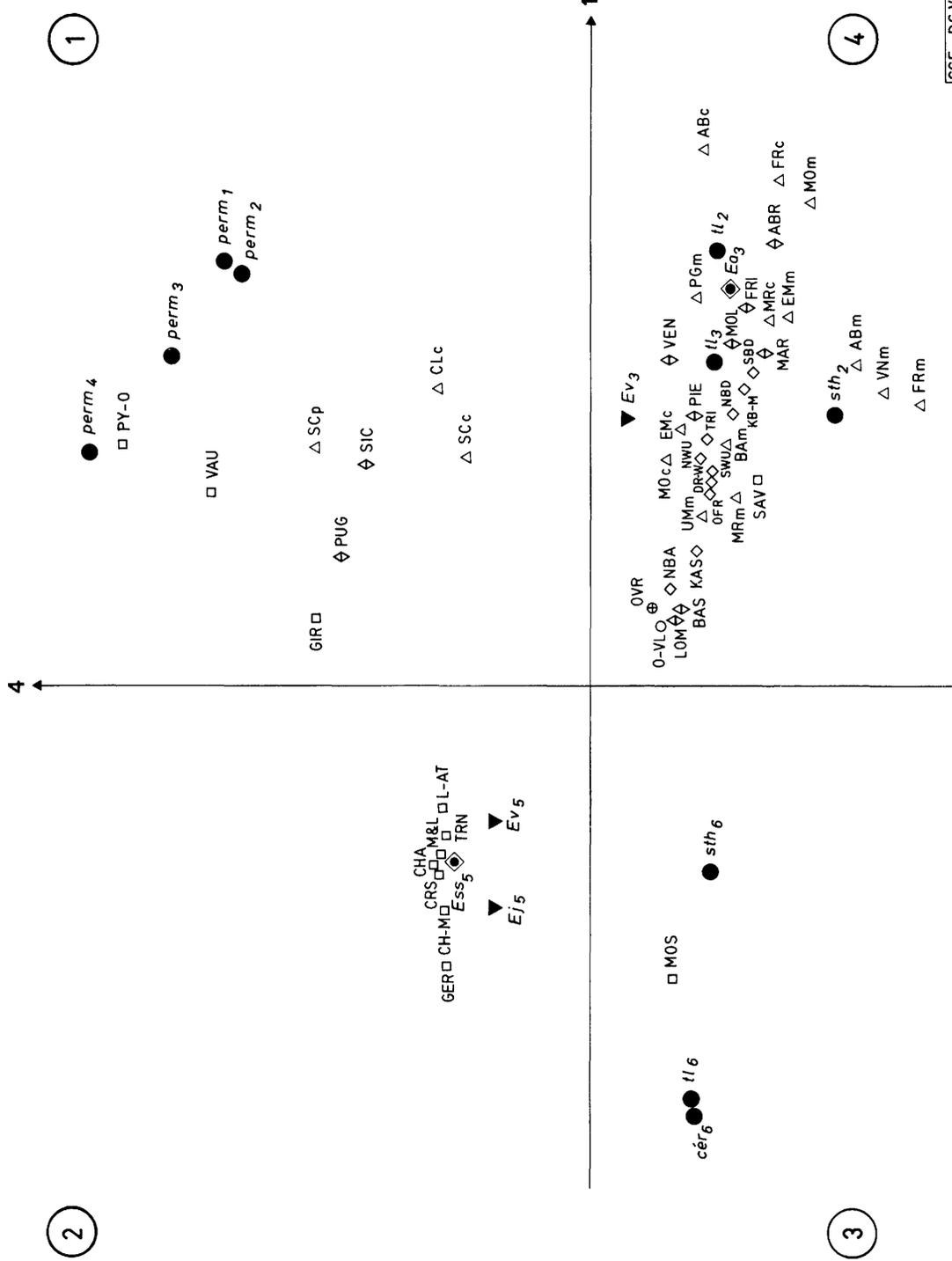
Seuls les départements Vaucluse et Gironde sont représentés à titre principal dans le premier quadrant associés directement aux cultures permanentes dans les exploitations de 5 à 20 ha. Toutes les autres circonscriptions Puglia, Sicilia, Sicilia plaine, Sicilia colline, Calabrie colline et Pyrénées Orientales, ainsi que les cultures permanentes sur les exploitations de moins de 5 ha ne sont dans le plan 1-4 qu'à titre secondaire.

Le deuxième quadrant contient seulement Charente-Maritime et Gers à titre principal. Loire-Atlantique, Maine et Loire, Charente, Creuse et Tarn sont mieux représentées dans d'autres plans. Ces régions sont caractérisées par l'absence de successeur sur les exploitations de 20 à 50 ha, ainsi que la coexistence d'exploitants jeunes et âgés dans cette même classe de surface.

Le troisième quadrant a pour seule composante la Moselle attirée à la fois par les surfaces toujours en herbe, les céréales et les terres labourables dans les exploitations supérieures à 50 ha.

Le dernier quadrant est hétérogène avec beaucoup de circonscriptions et des variables diversifiées. On retiendra que les principales circonscriptions sont italiennes avec Friuli colline, Emilia montagne, Marche montagne, Marche colline, Umbria montagne, Molise montagne, Molise colline, Puglia montagne, Basilicata montagne, Marche et Molise. Ces régions sont avec

GRAPHIQUE 6 : ANALYSE SIMPLIFIÉE 48 VARIABLES
 (Italie : Reg. + Zones altimétriques)
Plan 1 - 4



des chefs d'exploitations âgés et pourvus de successeur et ceci dans la catégorie de 10 à 20 ha. Quant au système de production, il fait référence aux terres labourables ainsi qu'aux surfaces toujours en herbe sur des superficies restreintes.

d - Plan 2-3 (Graphique n° 7, p. 94)

Le premier quadrant comprend Doubs, Puy de Dôme, Haute-Savoie, Ardèche, Loire, Hautes-Alpes, Belfort, Haute-Loire, Brabant belge, Limburg belge, Luxembourg belge, Namur, Drenthe Zuid-Holland, Manche, Corrèze, Haute-Vienne, Lot, Vosges, Aurich caractérisées par les surfaces toujours en herbe dans les exploitations de 5 à 50 ha. Jura, Haute-Saône, Creuse, Aveyron, Schwaben ont une meilleure représentation soit dans le plan 1-2, soit dans 1-3.

Dans le deuxième quadrant on trouve Valle d'Aosta, Trentino montagne, Toscana colline, Abruzzi montagne, Sardegna montagne, Sardegna plaine et Sardegna attirées par les surfaces toujours en herbe dans des exploitations de plus de 50 ha.

Le troisième quadrant contient un grand nombre de régions céréalières mieux représentées dans le plan 1-2. Seules, Lombardia plaine, Lazio plaine et Basilicata colline apparaissent mais liées, et ceci de façon relativement lâche, aux variables terres labourables et céréales sur les exploitations de 2 à 5 ha.

Le dernier quadrant inclut Köln, Aachen, Detmold, Arnsberg, Oberbayern, Hildesheim, Düsseldorf, Antwerpen, Côtes-du-Nord, Finistère, Morbihan et Landes avec des variables céréales et terres labourables sur les exploitations de 20 à 50 ha.

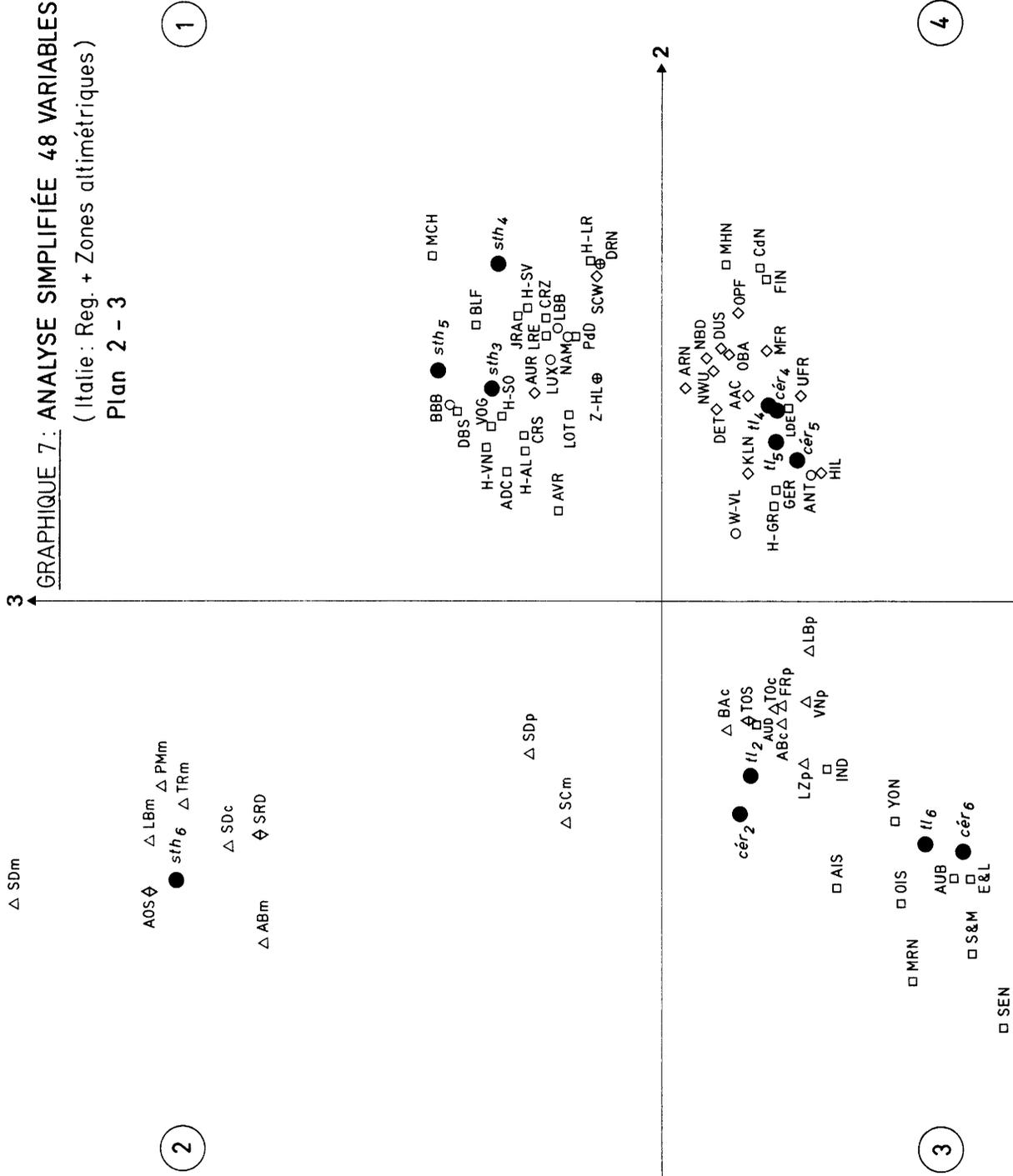
e - Plan 2-4 (Graphique n° 8, p. 95)

Le premier quadrant contient Loire-Atlantique, Maine et Loire, Lot et Garonne, Dordogne, Tarn et Garonne, Tarn, Drôme, Isère, Friesland, et Oberbayern dépendant des variables exploitants âgés et jeunes sur des surfaces de 20 à 50 ha. Quant à Haut-Rhin, Charente, Charente-Maritime, Creuse et Gers, ces circonscriptions mieux représentées dans un autre plan.

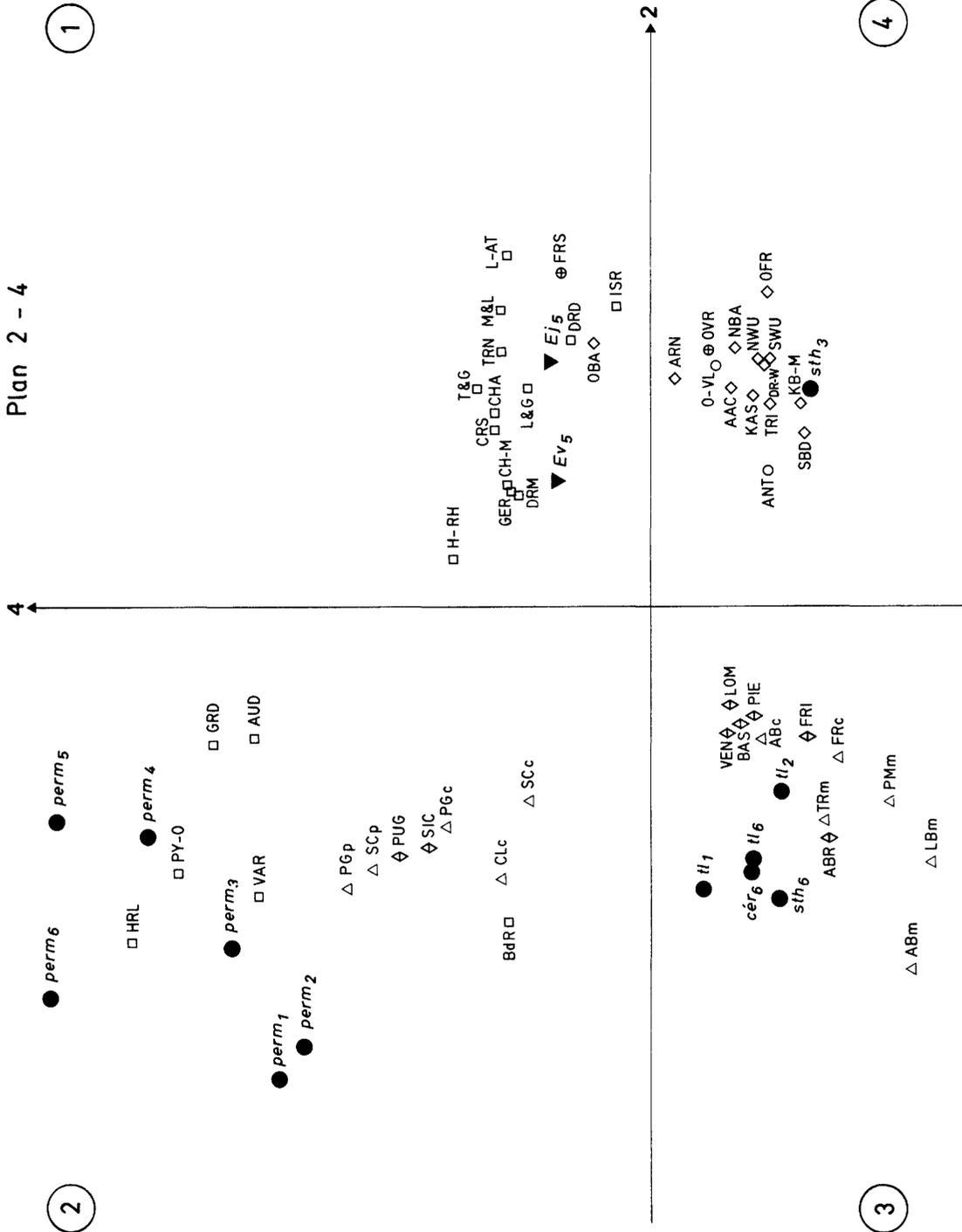
GRAPHIQUE 7 : ANALYSE SIMPLIFIÉE 48 VARIABLES

(Italie : Reg. + Zones altimétriques)

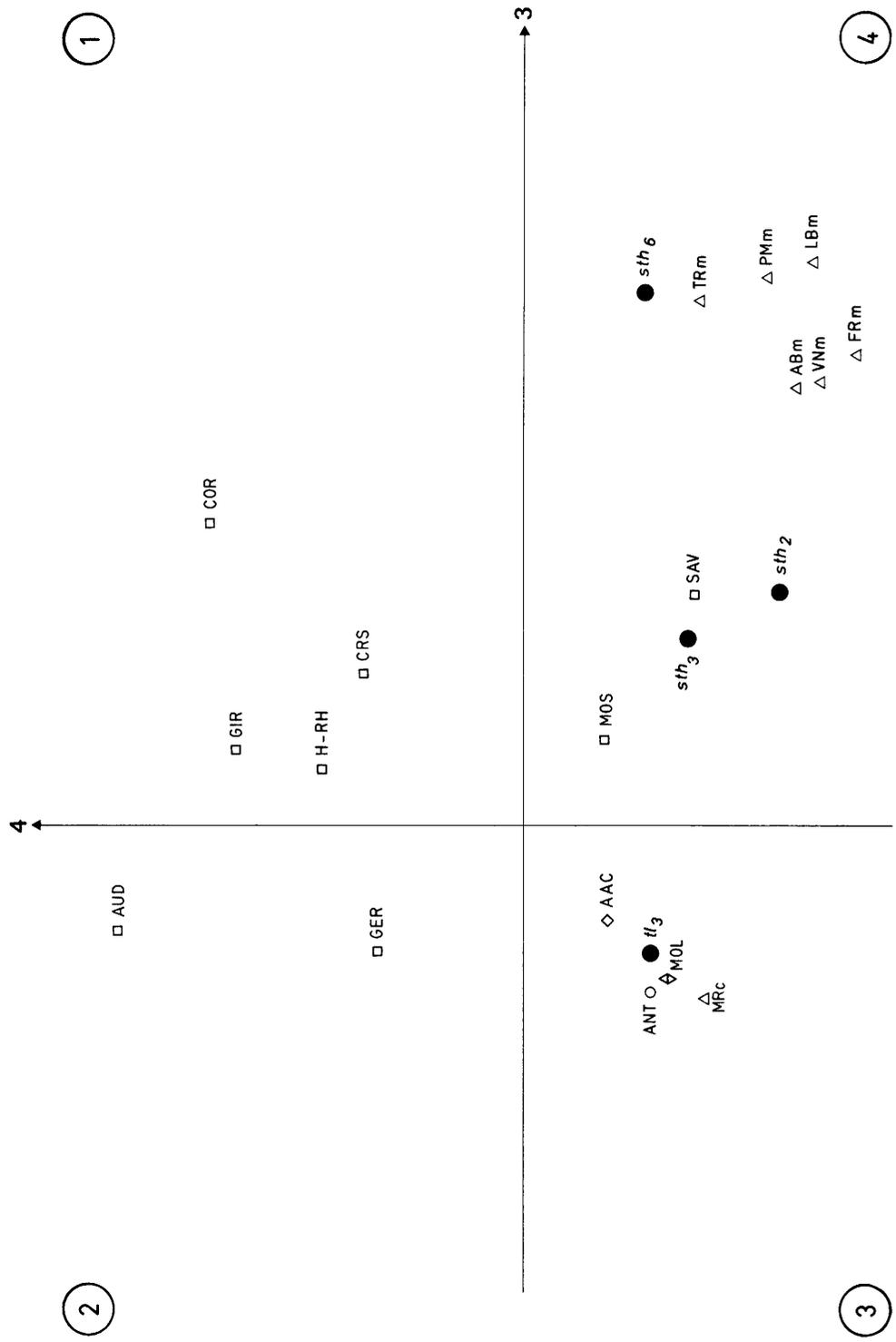
Plan 2 - 3



GRAPHIQUE 8: ANALYSE SIMPLIFIÉE 48 VARIABLES
 (Italie: Reg. + Zones altimétriques)
 Plan 2 - 4



GRAPHIQUE 9: ANALYSE SIMPLIFIÉE 48 VARIABLES
(Italie: Reg. + Zones altimétriques)
Plan 3 - 4



Dans le deuxième quadrant, Aude, Gard, Hérault, Pyrénées-Orientales, Bouches-du-Rhône, Var, Puglia colline, Puglia plaine, Puglia, Sicilia plaine et Sicilia sont associées aux cultures permanentes. Calabria colline et Sicilia colline apparaissent plutôt dans le plan 1-2.

Dans le troisième quadrant ne figure que Lombardia qui doit plutôt dépendre des variables terres labourables sur moins de 5 ha.

Enfin le dernier quadrant contient Oost-Vlaanderen, Overijssel et Aachen proches des surfaces toujours en herbe dans les exploitations entre 5 et 10 ha.

f - Plan 3-4 (Graphiques n° 9, p. 96)

Ce plan contient très peu de points. Dans le premier quadrant se trouvent Gironde, Corse et Haut-Rhin mais sans que l'on puisse associer précisément des variables significatives. L'explication de ces présences ne peut pas être trouvée simultanément par rapport aux deux axes.

Les points qui figurent dans les quadrants 2 (Aude, Gers) et 3 (Molise, Antwerpen, Aachen, Marche colline avec terres labourables sur 5 à 10 ha) se situent dans ce plan à titre secondaire.

Dans le quatrième quadrant se situent Piemonte montagne, Lombardia montagne, Friuli montagne avec la variable surfaces toujours en herbe sur les exploitations de plus de 50 ha.

2 Essai d'une typologie agricole

Dans le tableau récapitulatif détaillé ci-après (p.99) nous avons regroupé tous les éléments permettant de caractériser les circonscriptions : taille des exploitations, variables démographiques et système de production. Dans ce tableau, on a donné un rôle prépondérant à la superficie.

La classification ainsi présentée est assez souple et fait apparaître certaines régions en plusieurs endroits. En effet, il existe des recoupements. Dans certains plans, on a vu des associations du type exploitations de 20 à 50 ha où se trouvent des chefs d'exploitation jeunes ainsi que des chefs

d'exploitation âgés. Mais certaines des régions ainsi regroupées apparaissent ailleurs, avec les exploitations de 20 à 50 ha où les exploitants sont jeunes (ce qui n'a rien de contradictoire avec l'appartenance au sous-ensemble précédent) et avec un système de production à base de surfaces toujours en herbe et de terres labourables.

Rappelons pour la bonne interprétation de ce tableau qu'une région n'est pas caractérisée obligatoirement par son système de production dominant, ou par la surface occupée par la classe de SAU la plus importante. Une circonscription est définie par la variable ou le groupe de variables qui a par rapport aux autres circonscriptions le plus d'importance. C'est-à-dire que par exemple, la Savoie a une superficie moyenne supérieure à 5 ha (moyenne en 1967 de 7,5 ha), mais c'est la classe moins de 5 ha qui est relativement (en fonction de l'ensemble des circonscriptions) la caractéristique la plus marquante. C'est le cas aussi des régions situées dans la catégorie moins de 5 ha, certaines d'entre elles ayant une moyenne inférieure à 5 ha (exemple Alpes Maritimes, 3,0 ha).

Il importe de distinguer dans l'interprétation des résultats la notion de variable caractéristique de la notion de moyenne. De même une région peut apparaître sous une rubrique ne correspondant pas à la classe de surface la plus importante dans la région si au niveau communautaire, cette caractéristique est banale. Elle apparaîtra sous la rubrique, où relativement à l'ensemble, elle est la mieux pourvue (1).

(1) En ce qui concerne la nature des résultats proprement dits, il faut souligner la présence certaine d'une erreur dans le dépouillement de l'enquête communautaire pour les circonscriptions Oberbayern et Mittelfranken. En effet, le contrôle des données de base fait apparaître un rapport de 1 à 20 dans ces deux circonscriptions entre l'importance des terres labourables et l'importance des surfaces toujours en herbe, cela dans la classe de SAU inférieure à 2 ha. Dans la classe 20-50 ha, ce rapport est de 1 à 10. A priori, il semble étonnant que l'on ait dans ces régions si peu de surfaces toujours en herbe.

Par ailleurs, il apparaît dans Oberbayern et Mittelfranken une surface en céréales supérieure aux terres labourables, dont pourtant les céréales constituent une sous-rubrique. Il convient donc de s'interroger sur la place respective de ces deux circonscriptions au moins dans l'analyse. N'ayant pas de données complémentaires, il ne nous a pas été possible de rectifier les chiffres de base et donc d'attribuer ces régions à une catégorie différente.

TABLEAU RECAPITULATIF DETAILLE DE L'ANALYSE AGRICOLE

par la structure des exploitations		Circoscriptions caractérisées :		Circoscriptions concernées	
Classe de S.A.U.	Système de production basé sur :	par la démographie des chefs d'exploitations			
moins de 5 ha	terres labourables (tl) + céréales (cer)	Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels		LOM, BAS, LB-P, LZ-P, BA-C	
	surfaces toujours en herbe (sth)	Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels + Ev : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels		TOS, LAZ, CAM, CAL, PM-C, LG-C, LB-C, VN-C, FR-P, MR-C, TO-m, TO-P, CP-m, CP-C, CP-P, BA-C, CL-m, SC-m	
	cultures permanentes (perm)			PIE, LIG, VEN, ABR	
5 à 20 ha	terres labourables (tl)	Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels		SAV, AL-M, LG-m, VN-m, FR-m, MR-m, LZ-m, AB-m	
	terres labourables (tl) + céréales (cer)			AUD, GRD, HRI, BAR, VAR, LZ-C, PG-C, PG-P, SC-P	
	cultures permanentes (perm)	Ej : Exploitations dont le chef a moins de 50 ans sans successeurs potentiels		FT-O, FUG, SIC, CL-C, SC-C	
10 à 20 ha	surfaces toujours en herbe (sth)	Eas : Exploitations dont le chef a plus de 50 ans avec successeurs potentiels + Ev : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels		CAN, FIN, IGI, MHN, PV-A, H-PV, B-RH, RHO, LIE, HAI, O-VL, OVR, GLD, N-HL, BBN, LBN, SAA, RH-P, NBD, SBD, OFR, DR-W, KAS, KB-W, TRI, NMU, OPR, MPR, UPR, SCN	
	terres labourables (tl) + surfaces toujours en herbe (sth)			EMI, UMB, MOL, VN-P, FR-P, EM-C, EM-P, TO-C, UM-C, MO-C	
	céréales (cer) + terres labourables (tl)			VAU, GIR	
20 à 50 ha	surfaces toujours en herbe (sth)			FRI, MAR, FR-C, EM-m, MR-m, UM-m, MO-m, PG-m, BA-m, BA-P	
	terres labourables (tl) + surfaces toujours en herbe (sth)			MCH, CZZ, H-VN, LOT, VOG, DBS, BLF, H-LR, PAd, ADC, LRE, H-SV, H-AL, BBB, LBB, LXB, MAM, DRN, Z-HL	
	céréales (cer) + terres labourables (tl)			I&L, GAN, FIN, MHN, VIN, LDE, H-GR, MOS, ANT, HLL, BRS, DUS, KLN, AAC, DFT, ARN, OBA	
plus de 50 ha	surfaces toujours en herbe (sth)			CVL, ORN, H-VN, AVR, VOG, S&L, CTL, H-AL, AUR	
	terres labourables (tl) + surfaces toujours en herbe (sth)			NRD, P&C, SNM, L-AT, M&L, MAY, SPT, VDE, CHA, CH-M, SEV, CRS, H-VN, ARG, AVR, TRN, JRA, H-SO, P&D, AIN, H-AL, D-LX, UTR, ZEE, W-VL, GRO, SCH, HAN, LUN, STA, OSN, OLD, DUS, MUN, ARN	
	céréales (cer) + terres labourables (tl)	Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels + Ev : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels		L-AT, M&L, DRD, L&G, TRN, T&G, H-RH, DRM, ISR, PRS, OBA	
plus de 50 ha	terres labourables (tl) + surfaces toujours en herbe (sth)	Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels + Ev : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels		L-AT, M&L, CHA, CH-M, CRS, GEE, TRN	
	surfaces toujours en herbe (sth)			AIS, OIS, SEN, S&M, CHE, C&C, NVR, E&L, IND, EUR, MRN, AUB, YOM, H-MR	
	cultures permanentes (perm) + surfaces toujours en herbe (sth)			L&C, LOI, SOM, AUB, YON	
				NVR, ADN, ALL, MSE, M&M	
				LZR, AL-P, ACS, SRD, TR-m, TO-C, AB-m, SD-C, SD-m, SD-P	
				COR	

Le signe - figurant dans les colonnes variables indique que les circoscriptions concernées ne sont pas caractérisées par cette catégorie de variables.

Afin de rendre la typologie plus opérationnelle nous avons repris le tableau précédent présenté de manière différente (cf. tableau synthétique de l'analyse agricole p. 101). Au lieu de donner un rôle prépondérant à la superficie, on a privilégié le système de production. On a distingué les systèmes à base de terres labourables et céréales (terres labourables seules, terres labourables et céréales, céréales seules), de surfaces toujours en herbe, de cultures permanentes. A ces systèmes se superposent les combinaisons entre grandes catégories (essentiellement terres labourables + surfaces toujours en herbe). Pour chacun des systèmes, on a tenu compte des possibilités pour une circonscription d'être caractérisée par de très petites surfaces (moins de 5 ha), de petites surfaces (5-20 ha), des surfaces moyennes (20 à 50 ha) ou de grande surfaces (plus de 50 ha).

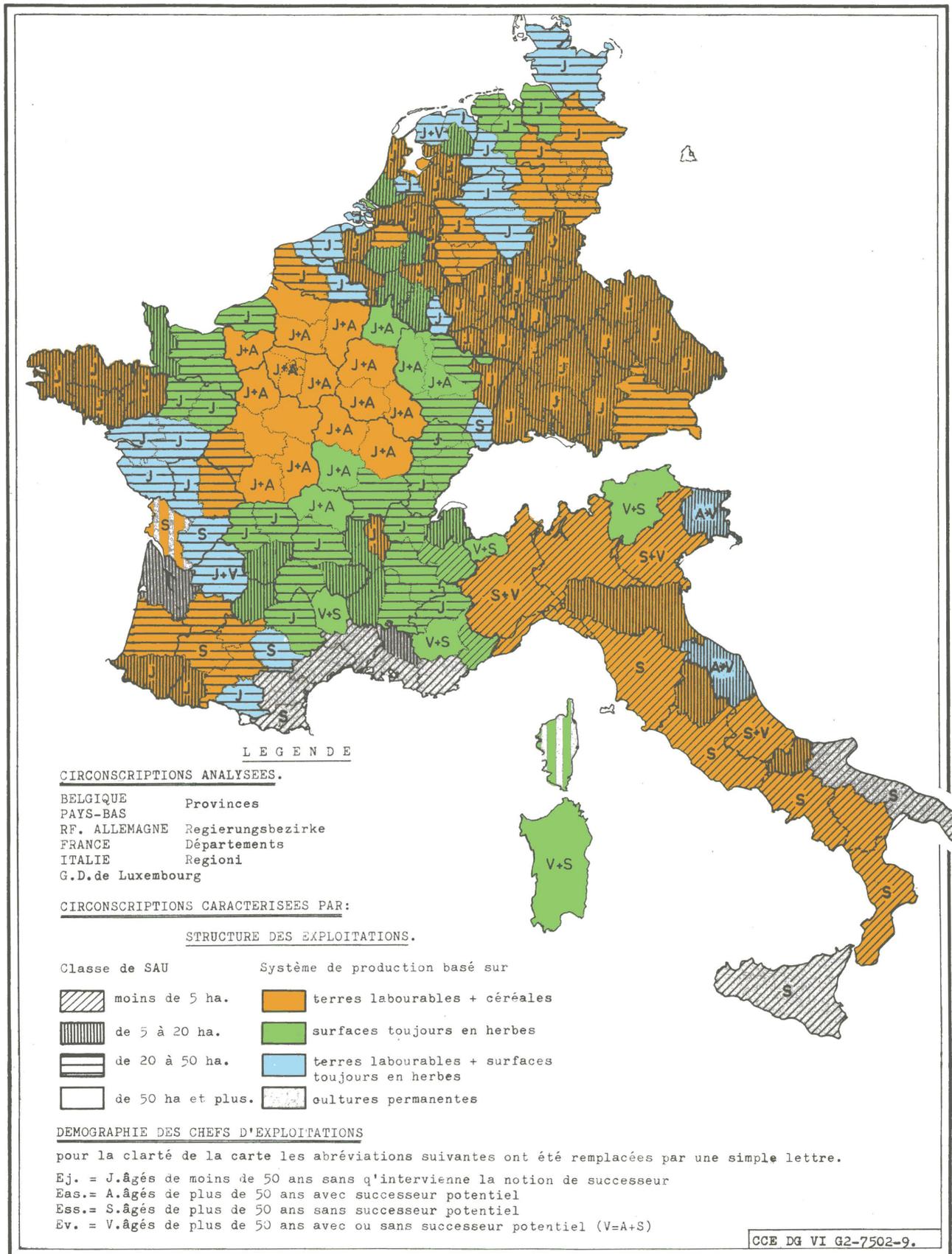
Ceci a été traduit sous forme de carte rendant compte des caractéristiques des régions du point de vue système et taille des exploitations. L'aspect démographique, lorsqu'il se manifeste, apparaît pour une circonscription à l'aide d'un système de lettres signifiant la prédominance relative d'exploitants de moins de 50 ans, ou de plus de 50 ans ainsi que la présence ou non d'un successeur pour ces derniers (cf. carte relative à l'analyse agricole. p. 102).

TABLEAU SYNTHETIQUE DE L'ANALYSE AGRICOLE

Système de production basé sur terres labourables (tl) et terres labourables + céréales (tl + cer)		Circonscriptions concernées	
moins de 5 ha	Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels + Ev : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels	LOM, BAS, LB-S, LZ-D, BA-C (TOS, LAZ, CAM, OAL, PM-C, LG-C, LB-C, VN-C, FR-p, MR-c, TO-m, CP-m, CP-c, CP-p, BA-c, CL-m, CL-D, SC-m)	
5 - 20 ha	Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels	(PIE, LIG, VEN, ABR)	
20 - 50 ha	Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels + Ess : Exploitations dont le chef a plus de 50 ans avec successeurs potentiels	EMI, UMB, MOL, VN-S, FR-D, EM-C, EM-p, TO-c, UM-c, MO-c (OBN, FIN, L&V, WHN, PY-A, H-PY, B-RH, RHO, LIE, HAI, NEA, O-VI, OVR, GLD, M-HL, BBN, LBN, SAA, RH-P, NBU, SBD, SWU, OFR, DR-W, KAS, KB-M, TRI, NMU, OFF, MFR*, UFR, SCH) L&L, LDE, VIN, H-GR, L&G, T&G, ANT, HIL, KLN, BRS, AAC, DET, ORA*	
plus de 50 ha	Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels + Ess : Exploitations dont le chef a plus de 50 ans avec successeurs potentiels	(P&C, HAN, LUM, DUS) (GER, CH-M + cultures permanentes) (AIS, OIS, SEN, S&M, CHE, E&L, EUR, IND, MRN, H-MR, C&O, YOM, AUB) L&G, LOI, SOM	
Système de production basé sur terres labourables (tl) + surfaces toujours en herbe (sth)			
10 - 20 ha	Eas : Exploitations dont le chef a plus de 50 ans avec successeurs potentiels + Ev : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels	(PRI, MAR, PR-C, EM-m, MR-m, UM-m, MO-m, PG-m, BA-m, BA-p)	
20 - 50 ha	Ess : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels + Ev : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels	(NRD, L-AT, M&L, VDE, SEV, ARG, D-LX, GRO, UFR, ZEE, SCH, OSU, MUN, ARN, W-VL) (CHA, TRN, H-RH) DRD, FRB	
Système de production basé sur surfaces toujours en herbe (sth)			
moins de 5 ha		SAV, AL-M, LG-m, VM-m, PR-m, MR-m, LZ-m, AB-m	
5 - 20 ha		MCH, CRZ, LOT, BLF, H-LR, ADC, LHE, H-SV, BBB, LBB, LXB, NAM, DRN, Z-HL	
20 - 50 ha	Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels + Ej : Exploitations dont le chef a moins de 50 ans avec ou sans successeurs potentiels + Ess : Exploitations dont le chef a plus de 50 ans avec successeurs potentiels	CLV, ORN, CTL, VOG, DBS, S&L, MOS, DRM, ISR, AUR MAY, SRT, H-SO, SMM, P&D, AVR, JRA, AIN, H-VN, CRS, H-AL, STA, OLD)	
plus de 50 ha	Ev : Exploitations dont le chef a plus de 50 ans avec ou sans successeurs potentiels + Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels	(NVR, ADM, ALL, MSE, M&M) (LZR, AL-P, AOS, SRD, TR-m, TO-c, AB-m, SD-c, SD-m, SD-p)	
Système de production basé sur cultures permanentes (perm)			
moins de 5 ha	Ess : Exploitations dont le chef a plus de 50 ans sans successeurs potentiels	AUD, GRD, HRL, B&R, VAR, LZ-C, PG-c, PG-D, SC-p PY-O, PUG, SIC, CL-c, SC-c)	
5 - 20 ha		VAU, GIR	
Système de production basé sur cultures permanentes (perm) + surfaces toujours en herbe (sth)			
		COO	

* Rappelons que les données de base concernant Oberbayern et Mittelfranken sont sujettes à caution.
Le signe - figurant dans les colonnes variables indique que les circonscriptions concernées ne sont pas caractérisées par cette catégorie de variables.

ANALYSE 48 VARIABLES AGRICOLE



C - Analyse économique générale

Rappelons d'abord que cette analyse est basée sur 25 variables dont 16 économiques générales (cf. p. 55 et 56).

La situation générale de chaque région est décrite à l'aide d'un ensemble de données relatif :

- à la population : population totale, population active totale, et par secteur, surface et densité,
- au revenu : Produit Intérieur Brut et sa ventilation par activité,
- au revenu par tête et par actif indiquant la productivité du travail par secteur.

L'agriculture de chaque région est saisie à partir :

- de la structure des exploitations : nombre par classe de SAU,
- de la structure du système productif : répartition du sol en terres labourables, surfaces toujours en herbe et cultures permanentes.

Rappelons aussi que les circonscriptions prises en compte pour l'Italie sont les "province" ici au nombre de 93 (Isernia et Campobasso n'ont pas été distinguées à l'intérieur de Molise).

Les caractéristiques générales de l'analyse sont rassemblées dans le tableau ci-dessous où figurent les valeurs propres et le pourcentage d'inertie expliqué par chacun des cinq premiers facteurs, ainsi que le pourcent d'inertie cumulée.

	1er axe	2è axe	3è axe	4è axe	5è axe
valeur propre	0,2121	0,1679	0,0531	0,0467	0,0223
% inertie du facteur	38,86	30,76	9,72	8,56	4,08
% inertie cumulée	38,86	69,62	79,34	87,90	91,98

De ce tableau, il ressort par rapport à l'analyse précédente (basée uniquement sur les variables agricoles) que les valeurs propres sont moins élevées, mais que le pourcentage d'inertie expliqué par chacun des quatre premiers facteurs est plus important.

Bien que satisfaisantes pour l'interprétation, les valeurs propres sont nettement inférieures à celles obtenues lors du traitement avec 48 variables agricoles. Ceci est net pour le premier facteur (0,21 contre 0,27) et le troisième (0,05 contre 0,10). Néanmoins ces valeurs propres sont plus importantes que celles fournies par les analyses portant sur 21 régions françaises ou 94 circonscriptions européennes (Annexe 2).

Les différences par rapport aux résultats constatés avec les données uniquement agricoles tiennent à la nature des variables. On a vu que la présentation des critères démographiques par classe de S.A.U. permet une bonne discrimination des individus. A priori, on peut être étonné par le fait que les valeurs propres ici enregistrées soient supérieures à ce que l'on avait constaté lors d'analyses portant sur des données comparables au niveau de 94 circonscriptions. En effet, ayant un plus grand nombre d'observations, on devrait avoir une diminution des valeurs propres, or c'est l'inverse. Il est probable que l'on peut attribuer cela au fait que les données économiques caractérisant de grandes régions sont assez "plates", en raison des compensations qui se produisent au sein d'une zone hétérogène. Mais si cette dernière est décomposée en éléments de base, ceux-ci étant fort dissemblables, les mêmes données sont plus discriminantes.

Les rapprochements entre l'analyse 48 variables agricoles et l'analyse 25 variables montrent de nettes différences dans l'explication apportée par les premiers axes. Ici le premier plan (1-2) représente 69,6 % d'inertie totale contre 59,0% dans l'autre traitement. D'autre part, les deux premiers facteurs sont très équilibrés, plus que précédemment (38,9 % - 30,8 % contre 35,8 % - 23,3 %). L'écart vient donc surtout du deuxième axe qui a plus de poids dans l'analyse 25 variables.

Nous procéderons facteur par facteur de la même manière que dans le point A (cf. p. 71 à 86).

1. Premier facteur (voir Graphique n° 10, p. 108 bis)

a - Signification : analyse des contributions absolues

L'étude des contributions absolues met en évidence deux sortes de variables :

- du côté négatif, la variable nombre d'exploitations de moins de 2 ha a une contribution absolue très forte. Ce critère est suivi du nombre d'exploitations entre 2 et 5 ha, et des surfaces occupées par les cultures permanentes,
- du côté positif se trouvent les variables Produit Intérieur Brut, valeur ajoutée dans l'industrie et valeur ajoutée dans les services.

Pour ce qui est des observations les plus caractéristiques, on a :

- avec les petites exploitations : Sicilia, Puglia, Calabria, Bari, Lecce Salerno, Veneto, Messina, Frosinone, Avellino, Catania, Basilicata, Agrigento, Trapani, Potenza, Chieti, Caserta, Benevento, Sardegna, Molise ;
- avec le Produit Intérieur Brut et la valeur ajoutée dans les secteurs non agricoles : Seine, Düsseldorf, Arnsberg, Darmstadt-Wiesbaden, Milano, Köln, Oberbayern, Nordwürttemberg, Zuid-Holland, Schleswig-Holstein, Noord-Holland, Nord, Münster, Lüneburg, Hannover, Brabant belge, Nordbaden, Stade, Rhône, Mittelfranken, Antwerpen, Detmold.

Ainsi ce premier facteur oppose de façon claire les zones méditerranéennes où domine l'activité agricole sous forme de cultures permanentes pratiquées sur de petites exploitations, aux zones continentales industrialisées et avec des services. Ces zones sont constituées par un certain nombre de provinces du Bénélux, de Regierungsbezirke, de départements français caractéristiques (Seine, Nord, Rhône) ainsi que de Milano. Ce schéma est conforme aux analyses déjà effectuées tant sur la France avec 21 régions de programme, qu'au niveau communautaire avec 64 circonscriptions (1). Il s'agit là d'un phénomène intéressant puisqu'il se retrouve à tous les niveaux

(1) cf. annexe 2 - 1ère analyse pour la France et la 2è analyse pour 94 régions - les graphiques 1 & 6.

de l'analyse, même si la forme peut varier selon que l'on a à faire à 21 régions françaises ou 94 circonscriptions communautaires. Cette distinction tranchée entre zones où dominent d'une part l'industrie et les services et zones agricoles défavorisées d'autre part, est importante sur le plan de l'action, puisqu'elle indique déjà clairement où doivent porter les efforts dans une perspective d'atténuation des disparités régionales.

b - Contributions relatives des observations

Par ordre décroissant sur la partie de l'axe 1 indiquant la présence de petites exploitations spécialisées, on relève : Catanzaro, Cosenza, Avellino, Benevento, Agrigento, Palermo, Chieti, Viterbo, Calabria, Abruzzi, Trapani, Perugia, Potenza, Salerno, Alessandria, Sicilia, Catania, Caserta, Umbria, Puglia, Udine, Lecce, Basilicata, Reggio di Calabria, Rieti, Bari, Asti, Treviso, Latina, Messina, Molise, Brindisi, Cuneo, Perugia, Taranto. Ces contributions relatives sont fortes, puisqu'elles vont de 0,87 pour Catanzaro à 0,60 pour Taranto.

En ce qui concerne la partie de l'axe indiquant l'industrialisation et la présence de services, on relève, toujours par ordre décroissant, Hannover, Lüneburg, Hainaut, Nord, Aachen, Detmold, Münster, Stade, Rhône, Köln, West-Vlaanderen, Hildesheim, Nordbaden, Saarland, Seine, Zuid-Holland, Oberbayern, Nordwürttemberg, Milano, Schleswig-Holstein, Antwerpen. Ces contributions sont encore plus importantes que celles propres à la partie agricole, ayant un des coefficients de corrélation variant de 0,94 pour Hannover à 0,61 pour Antwerpen.

2. Deuxième facteur

a - Signification : analyse des contributions absolues

Sur cet axe, l'analyse des contributions absolues met en évidence deux sortes de variables :

- sur la partie négative : les exploitations inférieures à 2 ha que l'on trouvait déjà en négatif sur le premier axe, suivies du Produit Intérieur Brut, de la valeur ajoutée dans les services de la valeur ajoutée dans l'industrie (ces trois variables figurant en positif sur le premier facteur), ainsi que de la

population totale, et des exploitations de 2 à 5 ha ;

- sur la partie positive : les surfaces toujours en herbe, les terres labourables, le produit par actif dans l'industrie, le produit par actif dans le secteur non agricole, le nombre d'exploitations entre 20 et 50 ha, le produit par actif dans les services, le produit par actif (au niveau global), la surface totale, le produit par actif agricole, le nombre d'exploitations entre 10 et 20 ha et le nombre d'exploitations de plus de 50 ha.

Quant aux observations, on relève :

- sur la partie négative : Seine, Campania, Düsseldorf, Lombardia, Puglia, Sicilia, Milano, Veneto, Roma, Calabria, Napoli, Darmstadt-Wiesbaden, Liguria, Arnsberg, Piemonte, Bari, Nordwürttemberg, Sudwürttemberg, Zuid-Holland, Köln, Brabant belge, Toscana.
- sur la partie positive : Lozère, Cantal, Orne, Mayenne, Aveyron, Sardegna, Gers, Creuse, Deux Sèvres, Allier, Nièvre, Vendée, Indre, Manche, Saône-et-Loire, Meuse, Haute-Marne, Haute-Saône, Haute-Loire, Cher, Vienne, Sassari, Eure et Loir, Yonne, Alpes de Hte Provence, Ardennes, Nuoro, Lot, Dordogne, Calvados, Haute-Vienne, Sarthe, Corrèze, Maine et Loire, Aisne, Hautes-Alpes, Eure, Oise, Charente, Côte d'Or, Tarn, Ariège, Jura, Aube, Tarn et Garonne, Lot et Garonne, Puy de Dôme.

Ce deuxième facteur oppose donc un ensemble de régions agricoles avec des superficies moyennes et grandes orientées soit vers les surfaces toujours en herbe soit vers les terres labourables, à des régions peuplées certaines industrialisées, d'autres avec beaucoup de tertiaire, ou agricoles avec de petites exploitations. La signification de cet axe est plus difficile à expliciter que celle du premier, car la partie négative combine des régions où dominant les activités secondaires (Seine, Düsseldorf, Arnsberg, Milano, Nordwürttemberg, Köln) avec des zones très agricoles (Puglia, Calabria, Veneto). Le point commun de ces circonscriptions réside dans un peuplement dense, ce qui les oppose aux autres zones, à grande majorité françaises, qui sont plus clairsemées. Il est

remarquable d'ailleurs que les seules observations figurant avec les départements français et qui concernent d'autres pays sont Sardegnna ainsi que Sassari, Nuoro et Cagliari (ces trois provinces constituent la Sardegnna). Ceci confirme donc le caractère extensif de la partie positive de ce deuxième facteur.

b - Contributions relatives

Par ordre décroissant sur la partie de l'axe indiquant les zones peuplées, on relève Liguria avec une contribution de 0,63 puis Lombardia, Roma, Lazio, Napoli, Torino, Veneto, Campania, Milano, Brabant belge, Seine, Darmstadt-Wiesbaden, Düsseldorf, Zuid-Holland, Genova, Bari, Firenze, Köln, Antwerpen avec une corrélation de 0,20.

Du côté des zones extensives, on trouve des contributions beaucoup plus fortes puisque Ain vient en tête avec 0,88. Viennent ensuite : Creuse, Aisne, Aveyron, Haute-Loire, Haute-Saône, Corrèze, Haute-Vienne, Haute-Marne, Eure, Deux-Sèvres, Meuse, Jura, Ardennes, Gironde, Nièvre, Ariège, Ardennes, Mayenne, Sarthe, Allier, Drente, Tarn et Garonne, Puy de Dôme, Maine et Loire, Basses-Alpes, Côte d'Or, Charentes-Maritimes, Saône et Loire, Drôme, Luxembourg belge, Aurich, Gers, Orne, Osnabrück, Cher, Calvados, Loiret, Cher, Landes, Vienne et Doubs. Cette dernière circonscription est liée au facteur avec une corrélation de 0,57, ce qui reste important, surtout par rapport aux contributions dans la fraction négative de l'axe.

c - Plan 1-2

Le plan 1-2 qui représente plus des deux-tiers de l'inertie totale du nuage de points permet la visualisation des phénomènes décrits. On constatera la forme triangulaire du nuage de points avec le regroupement dans toute la partie gauche des circonscriptions italiennes. Très peu de régions non italiennes (essentiellement Pyrénées-Orientales, Namur, Aude et Trier) sont incluses dans cette partie du graphique.

Le sommet du triangle est de nature différente avec les zones toujours en herbe relevant des alpages telles que Sardegna, Gers, Lot, Creuse, Lozère, Cantal.

Par contre, quelques circonscriptions italiennes se situent dans la partie droite et en bas du graphique. Il s'agit des régions peuplées (Roma, Torino, Genova) ou industrialisées (Milano). Ce quadrant IV contient surtout les régions allemandes avec quelques provinces du Bénélux et très peu de départements français. Ces derniers se trouvent plutôt dans la premier quadrant.

3. Troisième facteur (Voir Graphique n° 11, p. 112 bis)

a - Signification : analyse des contributions absolues

La partie négative de cet axe contient les variables : surfaces toujours en herbe, les terres labourables, et les exploitations entre 5 et 50 ha, donc uniquement des variables agricoles.

La partie positive fait référence aux seuls critères de productivité : productivité par travailleur pour les services, l'agriculture, l'ensemble des activités, les activités non agricoles et l'industrie. La densité de population intervient mais beaucoup moins nettement que ces critères.

Quant aux régions associées aux variables agricoles, il s'agit de : Sardegna, Nuoro, Piemonte, Schleswig-Holstein, Marche, Sassari, Schwaben, Lombardia, Niederbayern, Manche, Sicilia, Cagliari, Saône et Loire, Basilicata.

Du côté des variables concernant la productivité on note : Trieste, Belfort, Gorizia, Livorno, La Spezia, Imperia, Pyrénées-Orientales, Savona, Massa-Carrara, Alpes-Maritimes, Hautes-Alpes, Vaucluse, Pistoia, Var, Corse, Gard. Une telle association paraît à première vue surprenante, puisqu'il s'agit de régions qui ne sont pas particulièrement riches sur le plan industriel. Ainsi, Corse, Hautes-Alpes

sont des régions défavorisées. Aussi avons-nous noté dans le tableau ci-dessous les valeurs prises par les variables concernant la productivité pour un certain nombre de circonscriptions.

Valeur en U.C. de la productivité (1) par travailleur (1970)

	totale	agricole	non agricole	industrielle	services
Belfort.....	7177	3238	7360	7093	7764
Gard.....	7148	5032	7601	8183	7142
Alpes-Maritimes	7136	3967	7355	8363	6788
Corse.....	6484	3958	7342	8473	6773
Var.....	6990	3967	7390	8303	6800
Gorizia.....	5809	4643	5892	5286	6634
Trieste.....	6679	6862	6677	5873	7170
Imperia.....	5260	6026	5022	4905	5100
Savona.....	5770	3189	6066	4676	7632
La Spezia.....	5488	3911	5594	3941	7385

(1) Produit intérieur brut et valeur ajoutée par actif.

Pour la productivité totale les valeurs les plus fortes se rencontrent dans Seine = 9431 ; Seine-et-Marne = 9328 ; Köln = 8554 ; Düsseldorf = 8499 ; Lüneburg = 8349 ; Stade = 8353 ; Zuid-Holland = 8201. A l'inverse Avellino a le montant le plus faible avec 2227. Ainsi les chiffres observés dans les circonscriptions retenues sont élevés, surtout pour les circonscriptions françaises.

Pour la productivité agricole, les valeurs les plus fortes se trouvent dans Varese = 8690 ; Zuid-Holland = 7148 ; Trieste = 6862 ; Namur = 6825 ; Milano = 6755 ; Noord-Holland = 6418. La plus faible concerne Potenza = 1318. La productivité des circonscriptions analysées est donc élevée et plus particulièrement dans les provinces italiennes.

Pour la productivité de l'ensemble des activités non agricoles, on trouve en tête Seine = 9547 ; Seine et Marne = 9494 ; Stade = 9099 ; Köln = 8742 ; Lüneburg = 8713 ; Düsseldorf = 8656. Le minimum se rencontre avec Avellino = 2406. Les chiffres observés dans les régions que l'on a notées ci-dessus apparaissent élevés, surtout pour les départements français.

Les secteurs de l'industrie et des services, dont le total constitue le secteur non agricole, pourraient donner lieu à des réflexions identiques.

Ainsi les régions caractérisées par les variables "productivité" ne sont pas celles que l'on trouve en tête d'un classement par rang décroissant. Les régions les plus favorisées de ce point de vue sont plutôt associées aux critères de valeur ajoutée dans les branches non agricoles. A l'inverse les plus fortes productivités par actif agricole ne sont pas forcément le fait de zones où prédomine l'activité agricole. Tel est le cas de Trieste où les actifs agricoles représentent moins de 1 % du total.

Le troisième facteur s'interprète en termes d'opposition entre des régions qui ne font pas partie des zones industrialisées telles qu'elles apparaissent en négatif sur le deuxième axe, mais où la productivité du travail est relativement élevée, et des régions de type essentiellement agricole avec des exploitations moyennes et grandes, tournées vers les terres labourables et les surfaces toujours en herbe.

b - Les contributions relatives

Par ordre décroissant du côté négatif, on note Sardegnna dont la liaison avec le facteur est peu importante (0,39) suivie de Niederbayern, Schwaben, Südwürttemberg, Cagliari, Schleswig-Holstein, Piemonte, Saône et Loire, et Nuoro dont la contribution relative est de 0,20.

Du côté des zones à forte productivité, on relève d'abord Livorno dont la contribution est de 0,85. Viennent ensuite Gorizia, Savona, Massa-Carrara, Pistoia, Pisa, Pescara, Trieste, Belfort, La Spezia, Terni, Var, Venezia, Pyrénées-Orientales, Alpes-Maritimes, Varese, Imperia, Ravenna, Vaucluse, Pordenone, Como, Haut-Rhin, Genova, Forli, Siracusa, Cremona, Rovigo, Arezzo, Vercelli, Lucca, Hautes-Alpes, Ragusa, Gard, Grand Dûché du Luxembourg, Reggio nell'Emilia et enfin Groningen dont le coefficient de corrélation est de 0,30.

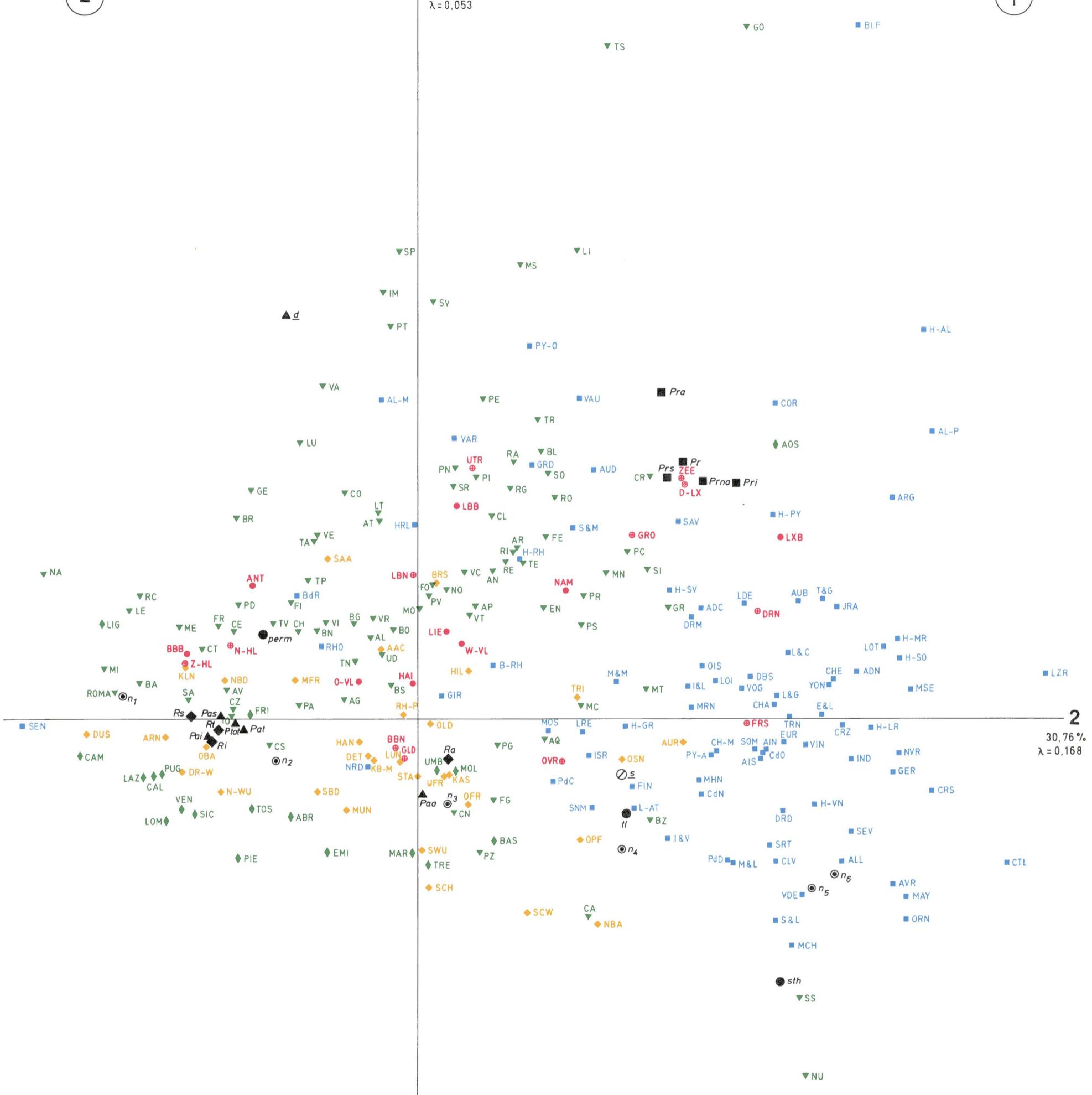
GRAPHIQUE 11: ANALYSE 25 VARIABLES
 (Italie: Reg. + Prov.)
 Plan 2 - 3

2

1

3

9,72 %
 $\lambda = 0,053$



2
 30,76 %
 $\lambda = 0,168$

3

4

GRAPHIQUE 12: ANALYSE 25 VARIABLES

(Italie: Reg. + Prov.)

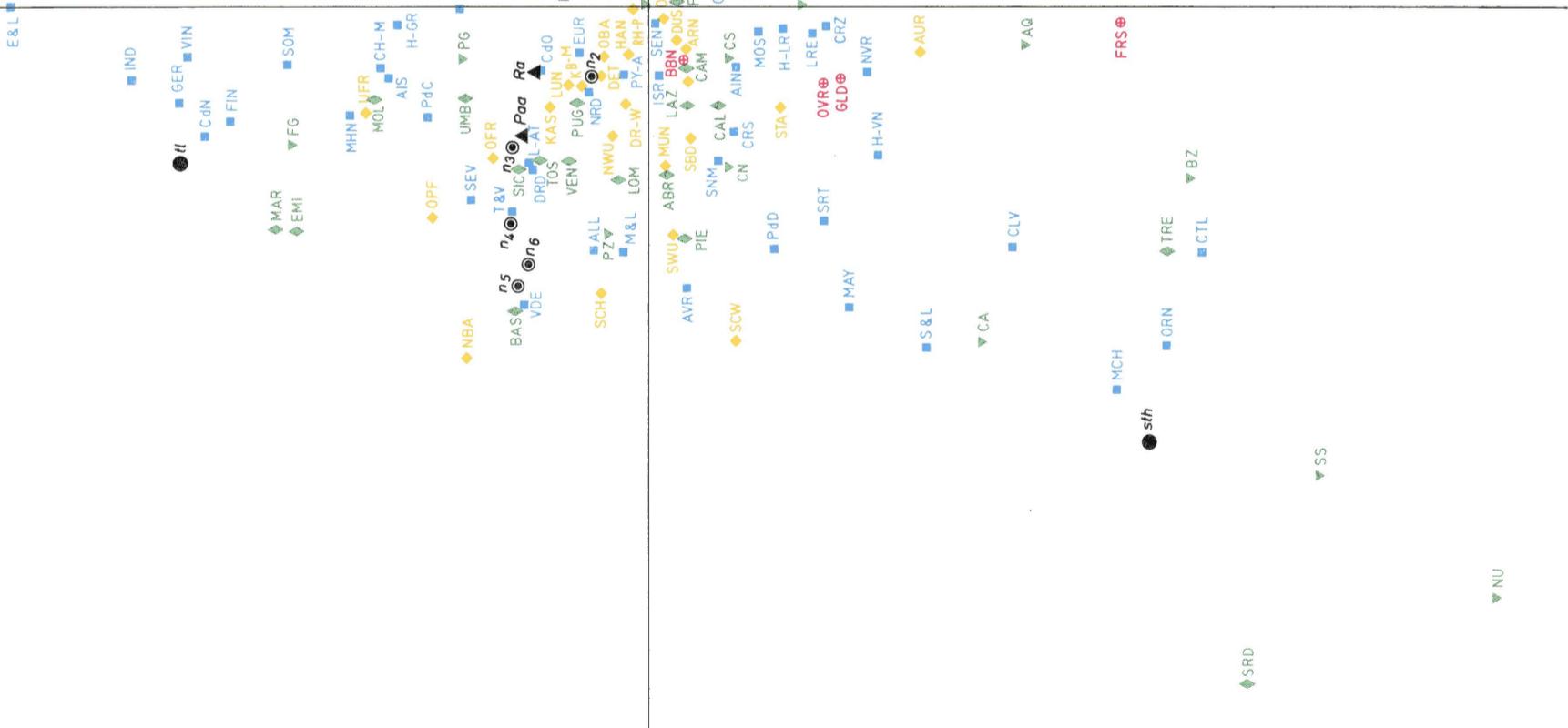
Plan 3 - 4

1

4

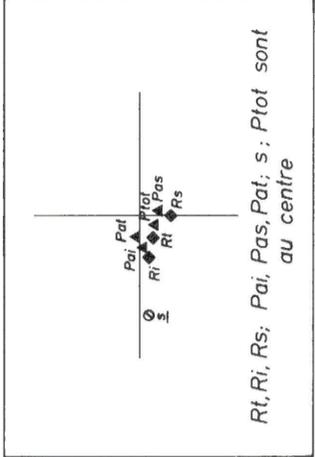
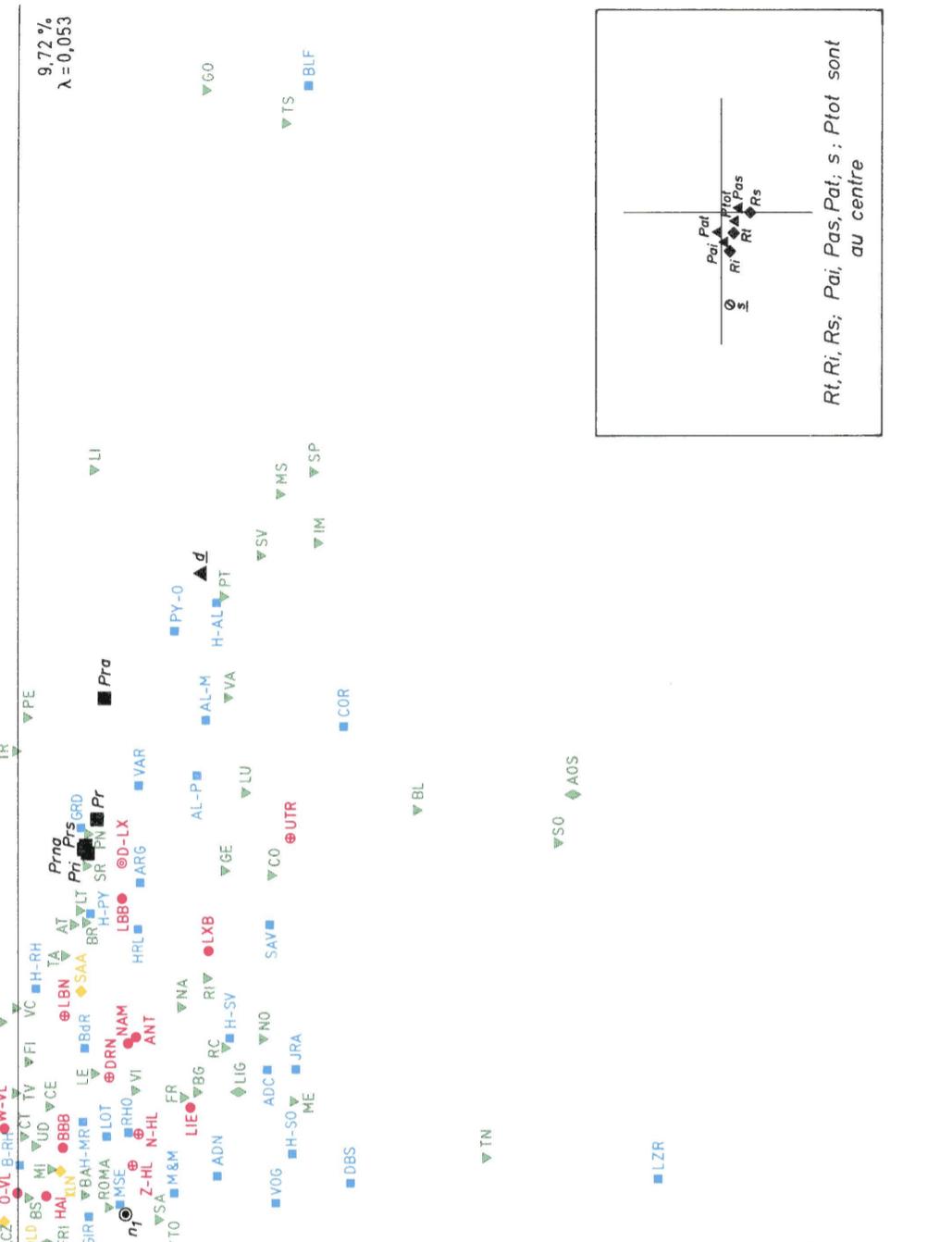
8.56%
 $\lambda = 0.046$

2



3

9.72%
 $\lambda = 0.053$



3

4

c - Plan 2-3

La représentation du nuage de points dans le plan 2-3 fait apparaître nettement dans le premier quadrant le regroupement des variables de productivité, ainsi que les régions qui sont attirées par celles-ci. On notera d'ailleurs que les zones les mieux caractérisées sont rejetées vers le haut et à proximité du troisième facteur.

L'opposition entre régions françaises de type extensif et zones peuplées d'Italie et d'Allemagne se manifeste sur le deuxième axe assez clairement, avec la prépondérance des départements français dans la partie droite du graphique. Seule, la Seine est rejetée dans la partie gauche. Nord et Rhône bien que dans la fraction négative du deuxième axe sont plus proches du centre.

4. Quatrième facteur (Graphique n° 12, p. 112 ter)

a - Signification : analyse des contributions absolues

Dans la partie négative de cet axe apparaissent les variables surfaces toujours en herbe ainsi que le nombre d'exploitations de moins de 2 ha. En fait, cette dernière variable intervient beaucoup moins que la première dans la définition du facteur.

Les variables situées dans la partie positive sont surface occupée par les terres labourables et nombre d'exploitations entre 2 et 5 ha et de 5 à 10 ha. Là aussi les critères nombre d'exploitations ont une contribution absolue très inférieure à celle des terres labourables.

Pour ce qui est des observations, en négatif, on relève Sardegna, Nuoro, Sassari, Trentino, Manche, Lozère, Bolzano, Trento, Friesland, Sandrio, Calvados.

En positif, on trouve : Emilia, Eure et Loir, Loiret, Marne, Marche, Côtes-du-Nord, Finistère, Vienne, Gers, Yonne, Foggia, Loir et Cher, Aube, Somme, Indre et Loir, Unterfranken, Cher.

Il s'agit ici d'un axe purement agricole opposant les surfaces en herbe des régions de type herbager ou avec des alpages, aux terres labourables des régions céréalières (Eure et Loir, Loiret...) ou des régions plus intensives (Côtes-du-Nord, Finistère...)

b - Contributions relatives

Du côté des surfaces toujours en herbe, on trouve Bolzano-Bozen dont la contribution relative s'élève à 0,53. Puis viennent Trentino, Nuoro, Sassari, Friesland, Sondrio, Trento, Gelderland, Sardegna, Valle d'Aosta, l'Aquila, Lozère, Orne, Aurich, Belluno, Calvados, Bergamo, Doubs. Cette dernière circonscription est corrélée à 0,20 au 4^e axe.

Du côté des surfaces occupées par les terres labourables, on note d'abord Emilia dont la contribution relative est de 0,46. Ensuite on trouve Loiret, Finistère, Unterfranken, Bologna, Côtes-du-Nord, Mantova, Marne, Eure-et-Loir, Marche, Indre et Loir, Ferrara, Foggia, Loir-et-Cher, Vienne, Ancona, Pavia, Pesaro e Urbino, Yonne, Modena, Macerata, Aube, Gers, Como, Somme, Piacenza, Rovigo, Perugia. Cette dernière circonscription a une contribution relative égale à 0,20.

c - Plan 3-4

Dans ce plan, toutes les variables concernant le PIB, la valeur ajoutée dans les secteurs non agricoles et les actifs dans les activités non agricoles sont très mal représentées. Elles se trouvent au centre, ce qui indique bien la nature essentiellement agricole de ce plan, bien qu'interviennent les critères de productivité liés au troisième axe.

On avait noté que la partie négative du quatrième facteur associait les surfaces toujours en herbe aux exploitations de moins de 2 ha. Or sur le graphique, ces deux points sont très nettement séparés montrant sans ambiguïté qu'il ne s'agit nullement de surfaces toujours en herbe sur les petites exploitations. De la même manière, la variable surface occupée

par les terres labourables est bien distincte du nombre d'exploitations de 5 à 10 ha et de 10 à 20 ha.

Seuls sont intéressants à considérer les points qui s'éloignent du noyau central. Il s'agit de ceux qui viennent d'être énumérés.

D - Synthèse des résultats de l'analyse générale

Pour avoir une vue aussi claire et synthétique que possible, nous procéderons de la même manière que pour l'analyse en termes uniquement agricoles. Toutefois nous compléterons ceci par une analyse tenant compte des seules variables économiques générales.

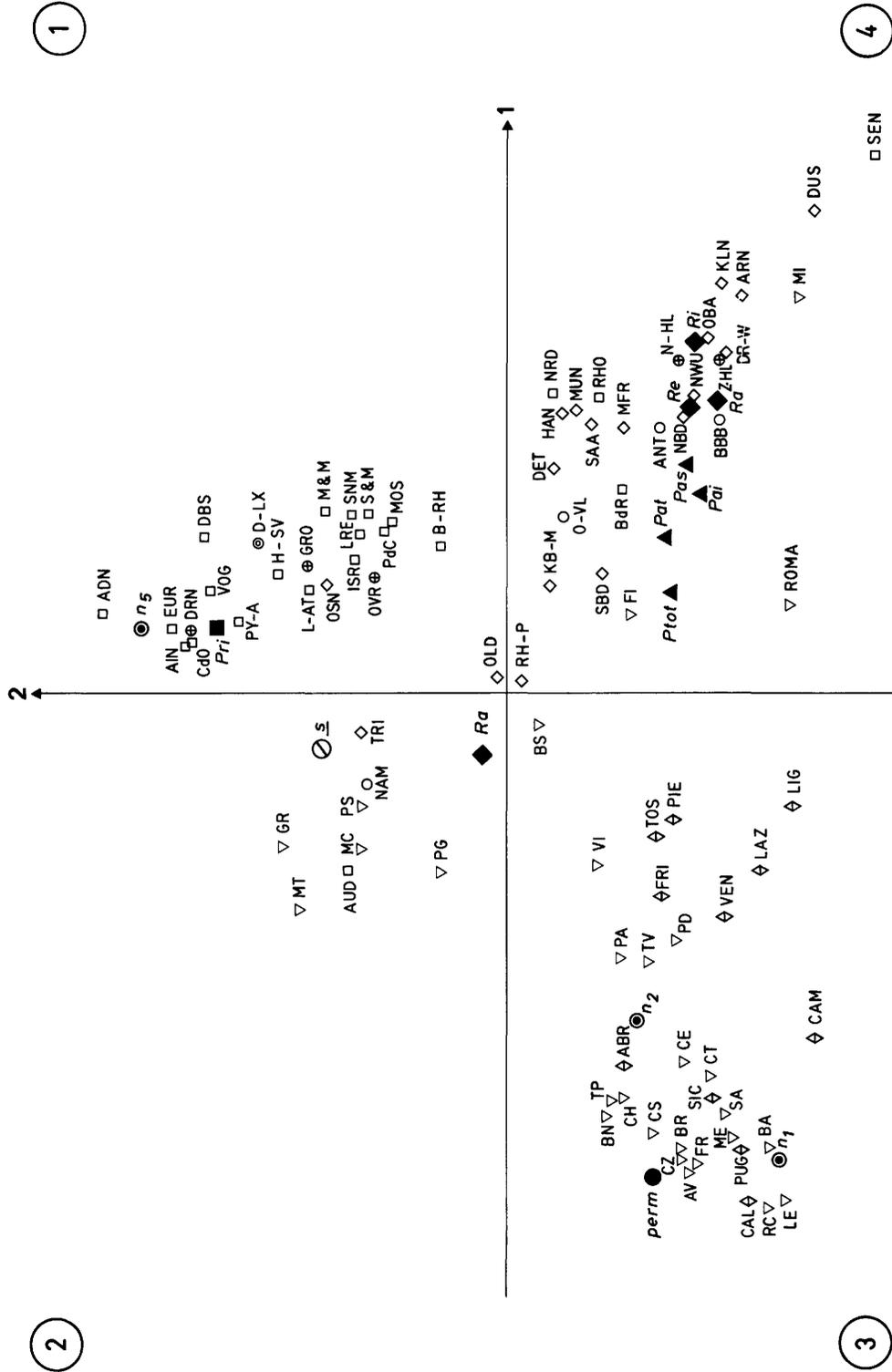
1. Les associations significatives de chaque plan de l'analyse économique générale

a - Plan 1-2 (Graphique n°13, p.116)

Les observations du premier quadrant pour lesquelles la représentation dans le plan 1-2 est la meilleure sont : Pas-de-Calais, Eure, Seine-Maritime, Loire-Atlantique, Pyrénées-Atlantiques, Meurthe-et-Moselle, Moselle, Bas-Rhin, Ain, Isère, Loire, Haute-Savoie, Overijssel, Osnabrück et Oldenburg. Les autres circonscriptions Seine et Marne, Doubs, Vosges, Ardennes, Côte d'Or, Grand Dûché du Luxembourg, Groningen et Drenthe sont mieux représentées sur d'autres plans essentiellement 2-3 et 3-4 (voir graphiques n° 16 et 17). Aucune variable n'a de représentation à titre principal dans ce quadrant. On notera seulement la présence de la productivité dans l'industrie, ainsi que du nombre d'exploitations de 20 à 50 ha qui ne figurent là qu'à titre accessoire.

Le deuxième quadrant contient les observations Trier, Grosseto, Macerata et Matera. Namur, Perugia, Pesaro e Urbino, Aude sont plutôt dans d'autres plans. La variable caractérisant de manière marquante ce quadrant est la surface des circonscriptions. La valeur ajoutée par l'agriculture est présente de manière secondaire.

GRAPHIQUE 13 : ANALYSE SIMPLIFIÉE 25 VARIABLES
(Italie: Reg. + Prov.)
Plan 1-2



Le troisième quadrant est composé uniquement de circonscriptions italiennes attirées par la variable cultures permanentes. Il s'agit surtout de Cuneo, Brescia, Treviso, Frosinone, Padova, Chieti, Caserta, Benevento, Avellino, Salerno, Bari, Brindisi, Lecce, Matera, Cosenza, Catanzaro, Vicenza, Palermo, Reggio di Calabria, Messina, Catania, Piemonte, Lombardia, Veneto, Campania, Friuli, Toscana, Lazio, Puglia, Sicilia, Calabria. Quant à Trapani et Liguria elles sont mieux représentées ailleurs.

Le quatrième quadrant contient les zones à forte activité non agricole avec les actifs de l'industrie et des services, les actifs totaux, la valeur ajoutée dans l'industrie et dans les services, le produit intérieur brut. Il s'agit aussi de régions peuplées puisqu'on trouve la variable population. Les observations présentes sont : Nord, Seine, Rhône, Bouches du Rhône, Antwerpen, Brabant belge, Oost-Vlaanderen, Noord-Holland, Zuid-Holland, Hannover, Düsseldorf, Köln, Detmold, Arnsberg, Darmstadt-Wiesbaden, Koblenz-Montabaur, Rheinhessen-Pfalz, Nordwürttemberg, Nordbaden, Südbaden, Oberbayern, Mittelfranken, Torino, Milano et Roma. Saarland, Münster et Firenze figurent là de façon secondaire.

b - Plan 1-3 (Graphique n° 14, p. 118)

Le premier quadrant caractérisé par la variable densité, qui apparaît aussi en 2-3, contient les observations : Seine et Marne, Alpes-Maritimes, Haut-Rhin, Bouches-du-Rhône, Hainaut, Limburg (belge), West-Vlaanderen, Utrecht, Limburg (Nederland), Saarland, Braunschweig, Aachen et Trieste. Les circonscriptions Antwerpen, Liège, Grand Dûché du Luxembourg et Firenze sont mieux représentées ailleurs.

Aucune variable n'a une représentation privilégiée dans le deuxième quadrant. Seules les superficies occupées par les cultures permanentes s'y trouvent de manière secondaire. Les circonscriptions sont Aude, Hérault, Pyrénées-Orientales, Asti, Alessandria, Brescia, Rovigo, Pavia, Pordenone, Verona, Udine, Imperia, Vercelli, La Spezia, Lucca, Terni, Pistoia,

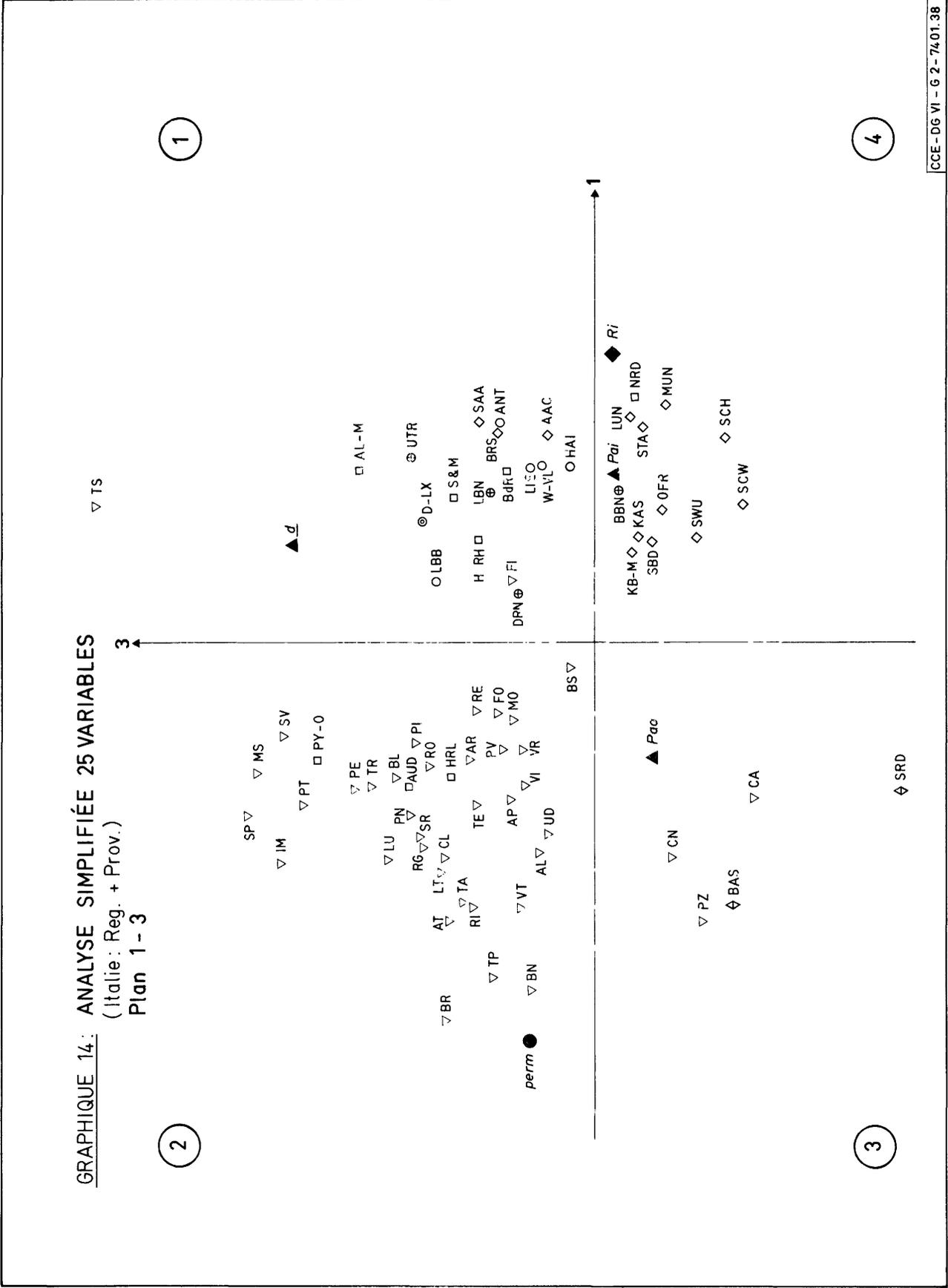
GRAPHIQUE 14: ANALYSE SIMPLIFIÉE 25 VARIABLES
(Italie: Reg. + Prov.)
Plan 1 - 3

2

1

3

4



Ascoli, Pisa, Viterbo, Arezzo, Rieti, Latina, Reggio Emilia, Teramo, Forli, Pescara, Massa-Carrara, Trapani, Taranto, Brindisi, Caltanissetta, Ragusa, Siracusa. A titre annexe, sont Belluno, Savona, Benevento, Modena, Vicenza.

Dans le troisième quadrant avec les actifs agricoles figurent Potenza, Cagliari, Cuneo, et Basilicata. Sardegnna est présente de façon accessoire.

Le quatrième quadrant contient Noord-Brabant, Schleswig-Holstein, Münster, Kassel, Koblenz-Montabaur, Südwürttemberg, Schwaben. Les circonscriptions Lüneburg, Stade, Südbaden et Oberfranken figurent aussi de manière secondaire. Les variables actifs de l'industrie et valeur ajoutée par l'industrie sont là aussi de manière secondaire.

c - Plan 1-4 (Graphique n° 15, p.120)

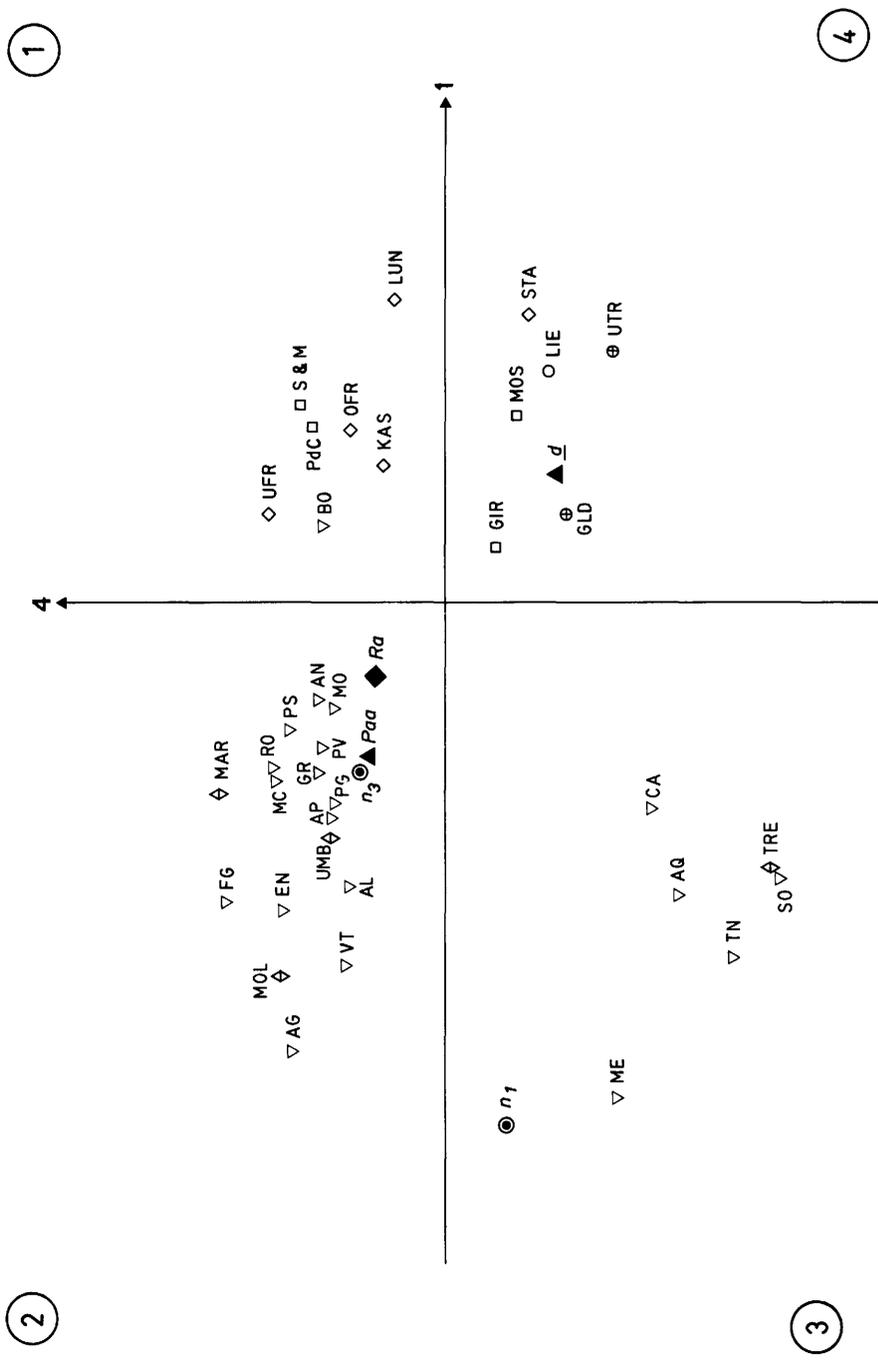
Le premier quadrant contient Pas de Calais, Lüneburg, Kassel et Unterfranken ainsi que Seine et Marne, Bologna mieux représentées ailleurs. Aucune variable n'apparaît nettement dans cette partie du graphique, ce qui signifie que l'on a intérêt à voir où se situent ces observations par rapport à un autre système d'axe afin de les classer.

Dans le deuxième quadrant, apparaissent les actifs agricoles et la valeur ajoutée de l'agriculture. A titre secondaire, on voit également le nombre d'exploitations entre 5 et 10 ha. Associées à ces critères sont Pavia, Modena, Perugia, Macerata, Ascoli, Viterbo, Molise, Foggia, Agrigento, Enna, Marche, Umbria. En plus, figurent Alessandria, Pesaro e Urbino, Grossetto.

Le troisième quadrant met en évidence Sondrio, Trento, l'Aquila, Cagliari, Trentino ainsi que Messina à titre secondaire. La variable exploitations de moins de 2 ha, qui est là, a une représentation meilleure dans un autre plan.

Le dernier quadrant est composé de Gironde, Liège, Gelderland, Stade, Oberfranken ainsi que de Moselle et Utrecht. La seule variable qui apparaisse, de façon assez secondaire, est la densité.

GRAPHIQUE 15 : ANALYSE SIMPLIFIÉE 25 VARIABLES
(Italie : Reg. + Prov.)
Plan 1 - 4



d - Plan 2-3 (Graphique n°16, p. 122)

Les variables caractéristiques du premier quadrant concernent la productivité par travailleur dans : l'agriculture, l'industrie, les services, le secteur non agricole et dans l'ensemble des activités. Y sont associées les observations : Ariège, Hautes-Pyrénées, Lot, Haute-Marne, Belfort, Drôme, Savoie, Haute-Savoie, Gard, Corse, Alpes de Hte Provence, Hautes-Alpes, Vaucluse, Grand-Duché de Luxembourg, Groningen, Namur, Drenthe, Zeeland, Limburg belge, Gorizia, Trieste, Piacenza, Livorno, Siena, Cremona. A titre secondaire, se trouvent : Seine et Marne, Landes, Tarn et Garonne, Haut-Rhin, Jura, Aude, Mantova, Parma, Reggio Emilia, Ravenna, Valle d'Aosta.

Le deuxième quadrant contient : Brescia, Venezia, Genova, Firenze et Napoli attirées par le critère densité.

Dans le troisième quadrant on trouve les regioni : Piemonte ainsi que Lombardia et Toscana mieux représentées en 1-2. Aucune variable n'apparaît dans cette partie du graphique.

Le quatrième quadrant contient les variables concernant la superficie totale des circonscriptions ainsi que les exploitations moyennes et grandes caractérisées par les surfaces toujours en herbe. Les observations que l'on relève sont : Ille-et-Vilaine, Maine-et-Loire, Mayenne, Sarthe, Vendée, Deux-Sèvres, Creuse, Dordogne, Aveyron, Saône et Loire, Allier, Puy de Dôme, Aurich, Niederbayern. A titre secondaire, figurent Manche, Orne, Loire-Atlantique, Oberpfalz, Sardegna.

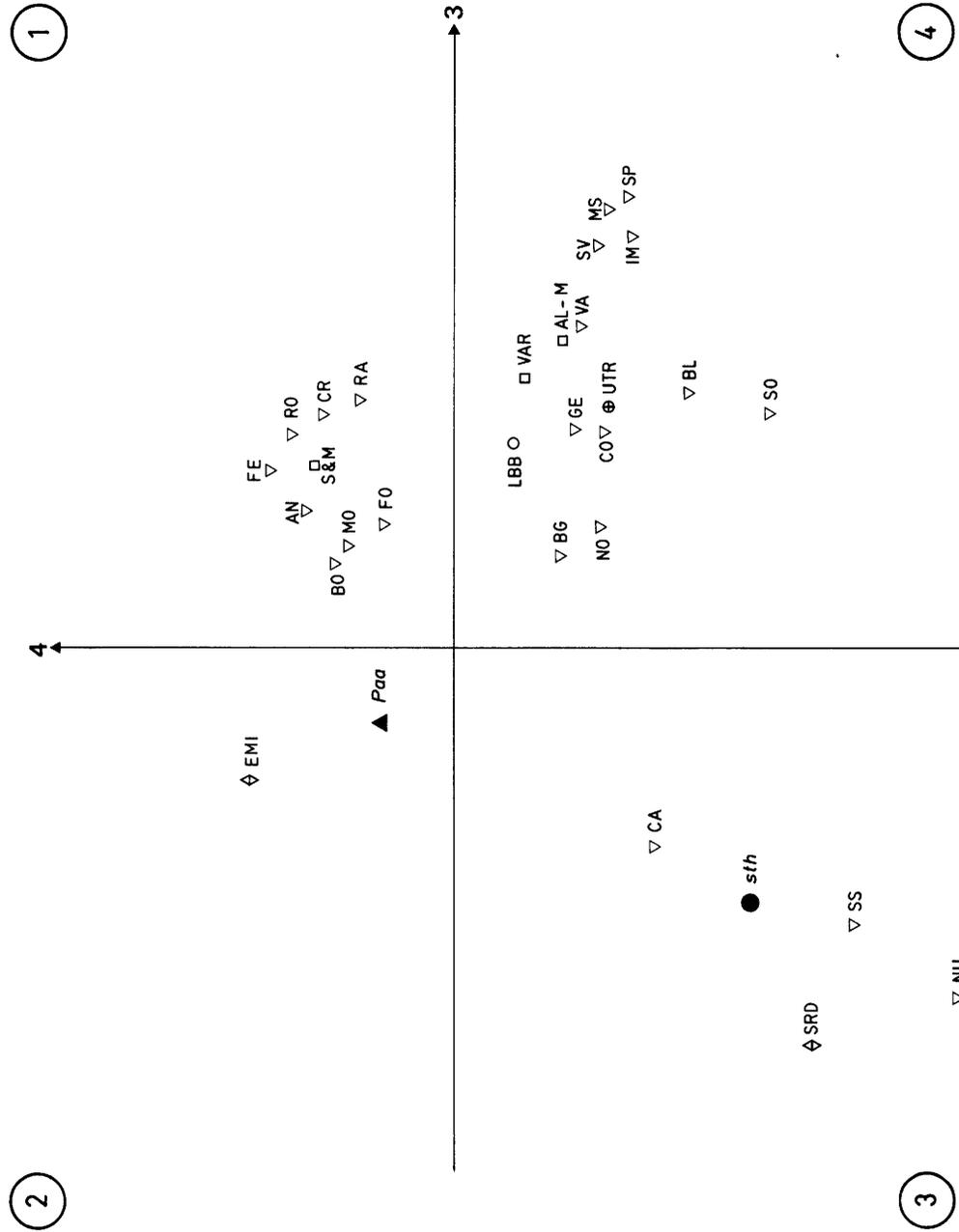
e - Plan 2-4 (Graphique n° 17, p. 123)

Les variables propres à ce quadrant sont d'une part les surfaces en terres labourables, et d'autre part les exploitations entre 10-20 ha, 20-50 ha et plus de 50 ha, et cela de manière secondaire. Les circonscriptions associées sont Aisne, Oise, Somme, Cher, Eure et Loir, Indre, Indre et Loir, Loiret, Loir et Cher, Côtes du Nord, Finistère, Morbihan, Charentes, Charentes-Maritimes, Vienne, Lande, Lot et Garonne, Haute-Garonne, Gers, Tam, Tarn et Garonne, Aube, Marne, Côte d'Or, Oberpfalz, Mantova, Piacenza, Parma, Pesaro e Urbino, Macerata, Cremona. Les circonscriptions Ille-et-Vilaine, Deux-

GRAPHIQUE 16: ANALYSE SIMPLIF. 25 VARIABLES
(Italie : Reg. + Prov.)
Plan 2 - 3



GRAPHIQUE 18: ANALYSE SIMPLIFIÉE 25 VARIABLES
(Italie: Reg. + Prov.)
Plan 3 - 4



Sèvres, Dordogne, Drôme, Ferrara et Grossetto sont également présentes de manière accessoire.

Nul point ne figure dans le deuxième quadrant. Le troisième contient uniquement Liguria avec les exploitations de moins de 2 ha. Quant au dernier, il contient : Calvados, Manche, Orne, Corrèze, Haute-Vienne, Ardennes, Vosges, Doubs, Jura, Haute-Saône, Nièvre, Cantal, Haute-Loire, Ain, Ardèche, Lozère, Meuse, Valle d'Aosta, Bolzano, Sassari, Nuoro, Friesland. A titre accessoire, on trouve aussi Mayenne, Sarthe, Saône et Loire, Puy de Dôme, Savoie, Haute Savoie, Luxembourg belge. La variable caractéristique de ce quadrant est la surface toujours en herbe.

f - Plan 3-4 (Graphique n° 18, p.124)

Par commodité, on peut regrouper les deux premiers quadrants où se trouvent la variable actifs agricoles ainsi que la valeur ajoutée dans l'agriculture et les surfaces en terres labourables, ces deux derniers critères ayant une meilleure représentation dans un autre plan. Pour ce qui est des observations, on relève Emilia, Bergamo, Rovigo, Modena, Bologna, Ferrara, Ravenna, Ancona. A titre accessoire, on trouve aussi Seine et Marne, Cremona et Forli.

Le troisième quadrant contient Cagliari et Sardegnna avec les surfaces toujours en herbe. De façon secondaire, les province Sassari et Nuoro figurent également.

Enfin , le dernier quadrant concerne Var, Novara, Varese, Como, Sondrio, Belluno, Savona. Les circonscriptions Limburg belge, Utrecht, Genova, La Spezia, Massa-Carrara et Imperia sont là à titre accessoire. Les variables productivité par travailleur dans l'agriculture et densité ont une meilleure représentation dans d'autres plans.

2. L'analyse économique générale dont on a supprimé les données concernant les structures agricoles et le système de production.

Afin de déterminer le rôle des variables agricoles et des variables générales dans l'analyse précédente, on a effectué un traitement des données en supprimant :

- les variables sur la structure agricole, c'est-à-dire le nombre d'exploitations par classe de SAU,
- les variables relatives au système de production, c'est-à-dire les surfaces toujours en herbe, les terres labourables et les cultures permanentes.

L'analyse se base ainsi sur 16 variables économiques générales exclusivement au lieu de 25 variables économiques générales et agricoles dans l'analyse précédente (voir p.103).

On a fait figurer les caractéristiques générales de l'analyse dans un tableau où se trouvent valeur propre et inertie :

	1er axe	2è axe	3è axe	4è axe
valeur propre	0,1996	0,0318	0,0143	0,083
% inertie du facteur	74,76	11,93	5,36	3,11
% inertie cumulée	74,76	86,69	92,05	95,16

Ce tableau montre que dans ce type d'analyse le premier facteur avec près de trois quarts de l'inertie totale, joue un rôle prépondérant. Le premier plan (1-2), donne une excellente représentation du nuage de points avec 86,7 % de l'inertie totale soit l'équivalent des quatre premiers facteurs dans l'analyse comprenant 25 variables. On note que la valeur propre du premier facteur, très proche de 0,20 est élevée, se situant presque au niveau constaté lors du traitement avec les 25 variables (0,21). Par contre, avec 0,03 la deuxième valeur propre est faible. La représentation du nuage de points est donc de type unidimensionnel puisque toute l'explication statistique se situe sur un axe.

Ces résultats sont à rapprocher de ceux constatés lors du traitement effectué sur 94 régions avec 14 variables (1). Rappelons que dans cette analyse, où les critères retenus étaient identiques (actifs, valeur ajoutée et productivité, pour le secteur agricole et le secteur non agricole, population,

(1) Cf. Annexe 2 - 2è analyse - graphique 5 concernant 94 régions communautaires.

surface et densité) le premier axe n'expliquait pas la moitié de l'inertie totale avec une valeur propre de 0,087. Le premier plan correspondait à 76 % de la variance. Ces chiffres montrent à l'évidence que l'on a obtenu avec un découpage géographique plus fin des éléments d'explication de meilleure qualité.

Pour saisir l'intérêt de cette analyse, examinons la signification des deux premiers facteurs.

a - Etude du premier facteur (Graphique n° 19, p. 128 bis)

Les contributions absolues les plus élevées du côté négatif concernent le Produit Intérieur Brut et la valeur ajoutée dans l'industrie et dans les services. En opposition figurent toutes les variables de productivité (services, secteur non agricole, industrie, ensemble des activités, agriculture).

Associées au PIB et à la valeur ajoutée des secteurs autres que l'agriculture se trouvent les observations Seine, Düsseldorf, Lombardia, Darmstadt-Wiesbaden, Nordwürttemberg, Arnsberg, Milano, Oberbayern, Piemonte, Lazio, Schleswig-Holstein, Köln, Roma, Zuid-Holland, Veneto, Emilia, Münster, Campania, Brabant belge.

Avec les critères de productivité coexistent les circonscriptions Lozère, Alpes de Hte Provence, Hautes-Alpes, Corse, Ariège, Lot, Belfort, Gers, Creuse, Valle d'Aosta, Tarn et Garonne, Aude, Haute-Marne.

Ainsi ce premier facteur oppose les régions où l'activité non agricole prédomine aux zones plutôt agricoles où, sans être nécessairement très élevée, la productivité est nettement au-dessus de la moyenne tant pour l'industrie, les services que l'agriculture (ceci peut se constater par lecture des données de base). Cet axe résulte d'un télescopage entre la partie négative du premier facteur de l'analyse 25 variables et la partie positive du troisième axe de cette même analyse.

Pour ce qui est des contributions relatives des circonscriptions associées à la partie négative, on relève Darmstadt-Wiesbaden dont la corrélation est de 0,96. Ensuite, on relève Oberbayern, Lombardia, Münster, Schleswig-Holstein,

Dusseldorf, Milano, Arnsberg, Seine, Hannover, Nord, Nordwürttemberg, Köln, Mittelfranken, Nordbaden, Südbaden, Zuid-Holland, Piemonte, Lüneburg, Brabant belge, Koblenz-Montabaur, Rhône, Liguria. Cette dernière observation est liée à l'axe par un coefficient de corrélation de 0,55.

En ce qui concerne la partie positive vient en tête Siena avec une contribution relative s'élevant à 0,98. Par ordre décroissant, on relève Arezzo, Rieti, Aude, Charente, Loir et Cher, Eure et Loir, Lozère, Haute Saône, Haute-Marne, Indre, Indre et Loir, Vienne, Haute-Loire, Cantal, Nièvre, Jura, Mayenne, Grossetto, Matera, Hautes-Alpes, Meuse, Ardennes, Loiret, Ardèche, Pyrénées-Orientales, Aube, Lot, Lot et Garonne, Charente-Maritime, Gard, Yonne, Sarthe, Calvados, Cher, Piacenza, Parma, Pesaro e Urbina, Enna, Alessandria, Terni, Viterbo, l'Aquila, Ain, Gers, Ariège, Drenthe Basses-Alpes, Creuse, Tarn et Garonne, Corrèze, Drôme, Savoie, Valle d'Aosta. Valle d'Aosta a une contribution relative de 0,90.

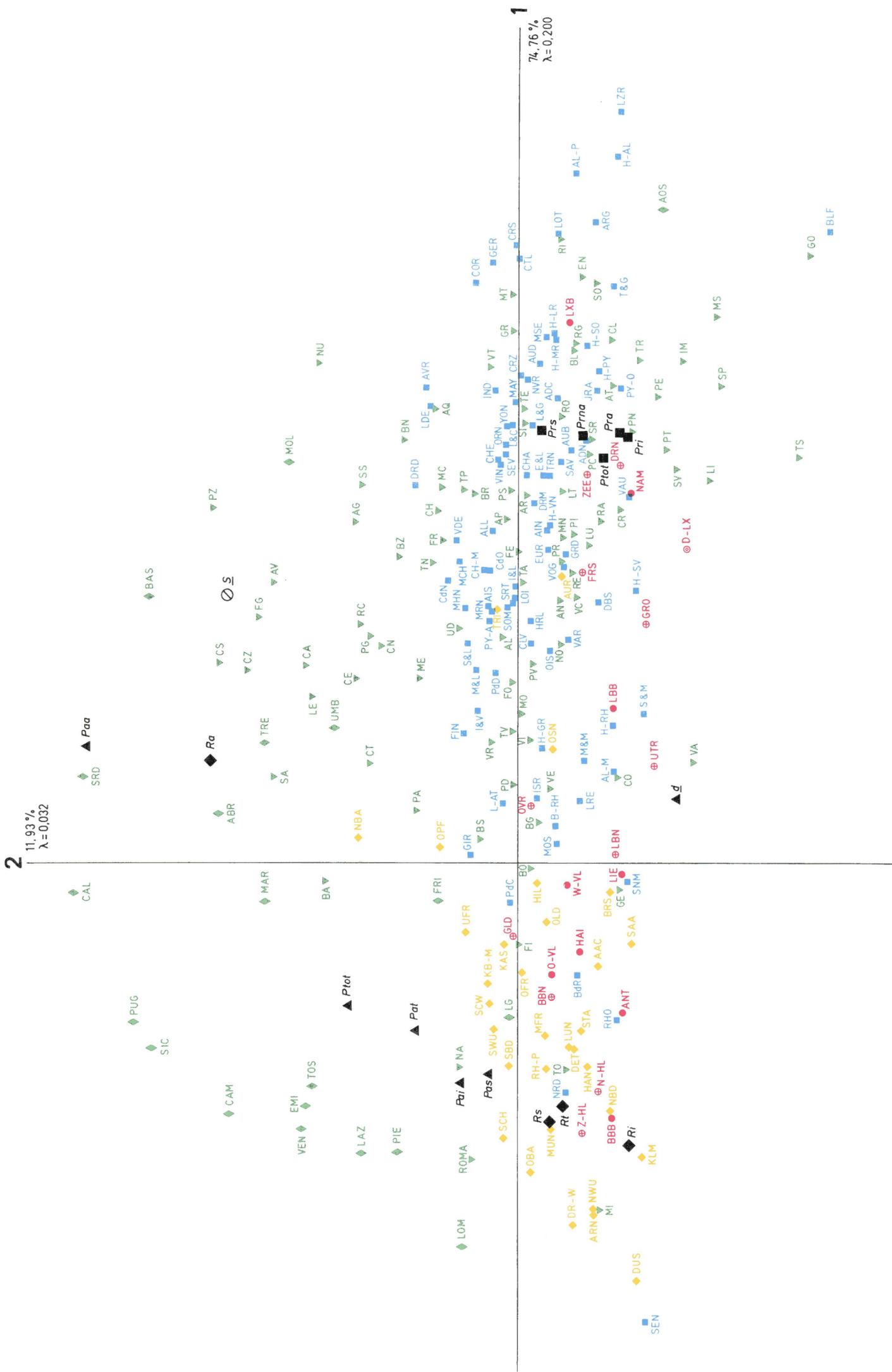
b - Etude du deuxième facteur (Graphique n° 19, p. 128 bis)

En négatif, les contributions absolues les plus fortes sont relatives aux variables valeur ajoutée dans l'industrie et productivité du travail dans l'industrie. Y sont associées les observations Seine, Düsseldorf, Trieste, Belfort, Köln.

En positif, on trouve les variables surface des circonscriptions, population, valeur ajoutée dans l'agriculture et actifs agricoles. Les observations dont les contributions absolues sont les plus élevées sont Sicilia, Puglia, Sardegna, Campania, Emilia, Veneto, Toscana, Abruzzi, Basilicata, Lazio, Trentino, Piemonte, Cosenza, Foggia.

On voit donc très clairement une opposition entre zones industrialisées où la productivité par actif dans l'industrie est forte, et des zones, uniquement italiennes, de nature agricole et avec une forte population.

GRAPHIQUE 19: ANALYSE 16 VARIABLES
 (Italie : Reg. + Prov.)
 Plan 1 - 2



2

1

3

4

Compte tenu de la très forte liaison de nombre de régions avec le premier axe, les contributions relatives par rapport à ce deuxième axe sont beaucoup moins importantes. En négatif, on relève Liège avec 0,74 suivie de Braunschweig, Limburg Nederland Saarland, Seine et Marne, Utrecht, Alpes-Maritimes, Bas-Rhin, Hildesheim, Genova, Haut Rhin et Loire dont la contribution relative est de 0,30.

En positif, en dehors des régions italiennes énumérées à propos de l'étude des contributions absolues on note Salerno avec 0,68 suivie de Bari, Cosenza, Niederbayern, Catanzaro, Foggia, Cagliari, Potenza, Catania, Palermo, Avellino, et Oberpfalz. Oberpfalz a un coefficient de corrélation de 0,30.

Vu la très faible valeur propre du troisième axe (0,014) son étude n'apparaît pas justifiée .

E - Essai d'une typologie sur la base de l'analyse économique générale

Les résultats avec 16 variables économiques générales conduisant à des constatations semblables à celles mises en évidence lors de l'analyse avec les 25 variables, on se réfèrera uniquement à cette dernière pour notre essai de typologie.

Pour établir notre classification, nous nous baserons principalement sur l'opposition entre zones à dominante agricole et zones à dominante PIB, industries et services telle qu'elle apparaît sur le premier axe. On appréciera si une circonscription est plutôt favorisée grâce à son PIB et à ses activités non agricoles, ou au contraire en se fiant à la liaison entre cette circonscription et le premier facteur. Une contribution relative forte de signe négatif signifie que l'observation est essentiellement agricole. Une contribution relative forte de signe positif, au contraire, nous permet d'attribuer à la circonscription l'autre caractère (PIB, acti-

vités non agricoles). Si une observation a une contribution relative faible et cela quel que soit son signe, nous examinerons le niveau de sa contribution absolue. S'il est assez élevé, cela nous indiquera l'appartenance au facteur.

Une contribution relative faible et une contribution absolue faible également ne nous permettent pas de trancher. On a donc une zone pour laquelle on ne peut pas indiquer si la région concernée est caractérisée ou non par le premier axe. Dans ce cas il faut se référer aux autres facteurs. Notons que la position d'une région est relative, puisqu'elle dépend du corpus sur lequel on travaille. L'introduction de nouvelles régions plus riches, ou plus pauvres que la moyenne de la Communauté à Six modifierait la répartition obtenue. Il faut voir aussi qu'il n'y a pas de limite très précise dans la distinction à laquelle on aboutit, et que les choix que nous sommes amenés à faire par nécessité contiennent une part d'arbitraire. Ainsi, on a mis comme seuil minimal pour dire si un point était lié ou non au facteur, une contribution relative égale à 0,30.

Grossièrement, on peut dire que les quadrants Q1 et Q4 du plan 1-2, Q1 et Q4 du plan 1-3, Q1 et Q4 du plan 1-4 contiennent des régions à forte activité non agricole.

Les quadrants Q2 et Q3 du plan 1-2, Q2 et Q3 du plan 1-3, Q2 et Q3 du plan 1-4 contiennent surtout des régions à dominante agricole. Les régions qui se trouvent dans les plans 2-3, 2-4 et 3-4 n'appartiennent ni à l'une ni à l'autre de ces deux catégories.

En reprenant dans le détail les circonscriptions figurant dans les plans 1-2, 1-3, 1-4 on peut corriger le schéma que l'on vient de voir.

Plan 1-2 (Graphique n° 13, p. 116)

Les régions à dominante activités non agricoles de Q1 sont : Pas de Calais, Seine-Maritime, Meurthe et Moselle, Moselle, Bas-Rhin, Isère, Loire, Oldenburg, Seine et Marne,

Overijssel et Osnabrück ont une liaison lâche avec le premier axe (corrélation de 0,25 à 0,20). Quant à Eure, Loire-Atlantique, Pyrénées-Atlantique, Ain, leur liaison avec le premier facteur est encore moins précise. Toutes les autres circonscriptions, dont la représentation dans le plan 1-2 n'est pas la meilleure (Drenthe) Groningen, Grand Dûché du Luxembourg, Ardennes, Côte d'Or, Vosges, Doubs et Seine et Marne qui est proche du seuil) sont difficiles à situer a priori.

Dans Q2, seules Perugia (mieux représentée dans 1-4) et Matera sont caractéristiques des zones agricoles. Macerata est à la limite, tandis que Trier, Grossetto, Pesaro e Urbino, Namur, Aude ne sont pas tranchées.

Dans Q3, toutes les observations sont bien caractérisées par la partie négative de l'axe, relative à l'agriculture, sauf Brescia qui est en dehors.

Dans Q4, toutes les circonscriptions appartiennent à la catégorie des zones développées sauf Torino, Firenze et Roma non caractérisées.

Plan 1-3 (Graphique n° 14, p.118)

Parmi les régions dont la meilleure représentation est dans le quadrant Q1 de 1-3 on a Haut-Rhin, Bouches du Rhône, Hainaut, Utrecht, Limburg, Nederland, West-Vlaanderen, Saarland, Braunschweig et Aachen parmi les zones industrialisées tandis qu'Alpes-Maritimes (contribution relative de 0,29) et Seine et Marne (0,25) sont très proches. Seules Limburg belge et Trieste sont sans appartenance. Liège et Antwerpen sont à activités non agricoles dominantes, le Grand Dûché de Luxembourg et Firenze n'étant pas caractérisées.

Dans Q2, les zones fortement agricoles sont : Pordenone, Vicenza, Udine, Asti, Alessandria, Pavia, Verona, Trapani, Lucca et Benevento. Les circonscriptions : Aude, Hérault, Brescia, Reggio nell'Emilia, La Spezia, Massa-Carrara, Belluno, Savona, Modena ne sont pas caractérisées. Les obser-

vations Rovigo, Forli et Imperia ont une contribution entre 0,20 et 0,24 soit au-dessous du seuil que nous avons fixé.

En Q3, toutes les observations relèvent de la zone défavorisée avec Potenza, Cuneo, Basilicata et Sardegna. Cagliari se situe en-dessous de la limite, bien que proche (0,24).

En Q4, toutes les circonscriptions sont à activités non agricoles dominantes avec Nord, Noord-Brabant, Schleswig-Holstein, Münster, Kassel, Koblenz-Montabaur, Lüneburg, Südbaden, Oberfranken. Les régions Südwürttemberg et Schwaben ne figurent pas dans la liste malgré les contributions relatives de 0,29 et 0,25.

Plan 1-4 (Graphique n° 15, p. 120)

Dans Q1, on voit que Pas de Calais, Lüneburg, Kassel, Unterfranken, qui ont là leur meilleure représentation, sont dans la catégorie des régions à activités non agricoles dominantes. Par contre, Seine et Marne et Bologna ne figurent dans aucune catégorie, bien qu'assez proches du seuil.

En Q2, Pavia, Perugia, Ascoli, Viterbo, Molise, Foggia, Agrigento, Enna, Umbria, Alessandria sont dans les zones défavorisées. Les observations Modena, Macerata, Grossetto, Pesaro ne sont pas suffisamment liées à la partie négative du premier facteur pour que l'on puisse trancher.

Dans Q3, Trento, l'Aquila, Trentino, Messina sont peu développées, alors que Sondrio et Cagliari sont proches de la zone limite.

Enfin pour Q4, on relève Liège, Stade, Oberfranken, Moselle et Utrecht dans la zone des régions développées, tandis que Gironde, Gelderland sont en marge.

L'étude détaillée des autres plans met en évidence l'appartenance du Haut-Rhin (2-3 ; Q1) dans les régions à activités non agricoles dominantes, de Piemonte et Toscana dans les régions agricoles (2-3 ; Q3) tandis que Lombardia se situe avec les zones à activités non agricoles dominantes (2-3 ; Q3).

La circonscription Sardegna (2-3 ; Q4 et 3-4 ; Q3) est dans la catégorie défavorisée, alors que c'est le contraire pour Utrecht (3-4 ; Q4).

Sur la base de l'ensemble de ces éléments et dans un souci de simplification, nous sommes en mesure de proposer la typologie suivante :

Présentation détaillée des résultats de l'analyse générale

I. Régions à activités non agricoles dominantes

a - circonscriptions étroitement liées au premier axe (positivement)

Les zones concernées sont présentées par rang décroissant de liaison au premier facteur et en tenant compte des contributions absolues. Entre parenthèses, on fait figurer les caractères agricoles lorsque ceux-ci se manifestent.

Hannover	Schleswig-Holstein
Lüneburg	Antwerpen
Hainaut	Bas-Rhin (n 5)
Nord	Limburg Nederland
Aachen	Brabant beige
Detmold	Moselle (n 5)
Munster	Bouches du Rhône
Stade	Oldenburg (n 5)
Rhône	Mittelfranken
Koln	Kassel
West-Vlaanderen	Loire
Arnsberg	Seine-Maritime (n 5)
Braunschweig	Oberfranken
Brabant Nederland	Utrecht
Düsseldorf	Meurthe-et-Moselle (n 5)
Darmstadt-Wiesbaden	Pas de Calais (n 5)
Liège	Isère (n 5)
Noord-Holland	Koblenz-Montabaur
Oost-Vlaanderen	Haut-Rhin
Hildesheim (n 5)	Südbaden
Nordbaden	Lombardia (c. perm.)
Saarland	Schwaben (Sth)
Seine	Alpes-Maritimes
Zuid-Holland	Südwürttemberg
Oberbayern	Seine et Marne (n 5)
Nordwürttemberg	Torino (n 1)
Milano	

On notera que les régions développées ne sont pas caractérisées par rapport aux critères agricoles.

b - régions faiblement liées au premier axe

Sous cette rubrique se trouvent les régions peu liées et moyennement liées à la partie positive du premier facteur, mais qui sont dans les quadrants où se trouvent les circonscriptions définies ci-dessus.

1) Plan 1-2

Dans le quadrant Q1 avec les variables productivité des travailleurs dans l'industrie et nombre d'exploitations entre 20 et 50 ha (n 5) on trouve :

Eure (n 5)	Ain (Sth)
Loire-Atlantique (Sth)	Haute-Savoie (Sth)
Pyrénées Atlantiques	Duché Luxembourg
Haute Garonne (t1)	Groningen
Ardennes (Sth)	Drenthe
Vosges (Sth)	Overijssel
Doubs (Sth)	Osnabrück
Côte d'Or (t1)	

Dans le quadrant Q4 où sont les actifs non agricoles et les actifs totaux, la valeur ajoutée dans le secteur non agricole et le Produit Intérieur Brut, ainsi que la population figurent Roma et Firenze. Cette dernière circonscription est mieux représentée en 2-3 Q2 avec la densité.

2) Plan 1-3

Il n'y a que dans le premier quadrant (variable densité) où apparaissent des circonscriptions autres que celles caractérisées dans le plan 1-2. Il s'agit de Limburg belge, Trieste et à titre secondaire Genova.

3) Plan 1-4

En Q1, où il n'y a pas de variable, se trouvent Unterfranken (Sth) et Bologna (t1). Cette dernière circonscription est mieux représentée dans le plan 3-4 Q1 avec les actifs agricoles et la valeur ajoutée agricole.

En Q4 avec la variable densité on a Gironde (n 5) et Gelderland.

II. Régions à activité agricole dominante

a - circonscriptions liées négativement au premier axe (forte liaison)

Le classement est fait par contributions relatives décroissantes et en tenant compte des contributions absolues. Entre parenthèses, se trouvent les caractéristiques agricoles lorsque celles-ci apparaissent.

Catanzaro (perm)	Taranto (perm)
Cosenza (perm)	Foggia (n 3)
Avellino (perm)	Enna (n 3)
Agrigento (n 3)	Padova (perm)
Benevento	Friuli (perm)
Palermo (perm)	Trento (n 1)
Chieti (perm)	Campania (perm)
Viterbo (perm)	Siracusa (perm)
Calabria (perm)	Caltanissetta (perm)
Abruzzi (perm)	Ragusa (perm)
Trapani (perm)	Vicenza (perm)
Frosinone (perm)	l'Aquila (n 1)
Potenza	Veneto (perm)
Salerno (perm)	Matera (tl)
Alessandria (perm)	Pordenone (perm)
Sicilia (perm)	Verona (perm)
Catania (perm)	Lucca (perm)
Caserta (perm)	Ascoli (perm)
Umbria (n 3)	Teramo (perm)
Puglia (perm)	Pavia (perm)
Udine (perm)	Trentino (n 1)
Lecce (perm)	Arezzo (perm)
Basilicata	Toscana (perm)
Reggio Calabria (perm)	Lazio (perm)
Rieti (perm)	Marche (n 3)
Bari (perm)	Cagliari (n 1, sth)
Asti (perm)	Piemonte (perm)
Treviso (perm)	Sondrio (n 1)
Latina (perm)	Nuoro (sth)
Messina (perm)	Sardegna (sth)
Molise (n 3)	
Brindisi (perm)	
Cuneo (perm)	
Perugia (tl)	

b - circonscriptions faiblement liées au premier axe

Il s'agit des régions proches de celles énumérées précédemment mais dont la liaison avec le premier facteur est lâche.

1) Plan 1-2

En Q2 avec la variable valeur ajoutée dans l'agriculture, on note : Macerata (tl), Grosseto (tl), Pesaro e Urbino (tl) et Siena. Bien que dépendant presque uniquement du deuxième facteur la région Trier peut être adjointe à cette liste.

2) Plan 1-3

En Q2 avec les cultures permanentes se trouvent : Aude, Hérault, Pyrénées-Orientales, Brescia, Belluno, Rovigo, La Spezia, Reggio nell'Emilia, Forli, Pistoia, Pisa, Terni, Pescara, Liguria, Imperia, Vercelli et Massa-Carrara.

3) Plan 1-4

En Q2 avec les actifs agricoles et la valeur ajoutée dans l'agriculture, on relève Modena (perm.).

En Q3 avec les exploitations de moins de 2 ha (n 1) on a Novara.

4) Plan 3-4

Le quadrant Q1 avec les caractères actifs agricoles et valeur ajoutée dans l'agriculture est de même nature que Q2 de 1-4. Parmi les circonscriptions de ce quadrant, on notera que celles qui ne sont pas déjà apparues. Il s'agit de : Bergamo (tl), Ferrara (tl), Ravenna (perm.), Ancona (perm). A cette série on peut adjoindre Emilia en Q2.

5) Plan 2-3

Le quadrant Q2 contient les circonscriptions Venezia et Napoli caractérisées par la variable densité.

III. Régions que l'on ne peut pas classer par rapport au premier axe

a - circonscriptions caractérisées par leur productivité

1) Plan 2-3

Dans le premier quadrant de ce plan se trouvent toutes les variables caractéristiques de la productivité du travail pour l'ensemble des activités. Pour ce qui est des circonscriptions qui ont là leur meilleure représentation, on relève :

Ariège	Corse (Sth)
Lot	Vaucluse
Hautes-Pyrénées	Luxembourg belge (Sth)
Haute-Marne	Zeeland (tl)
Belfort	Cremona (tl)
Drôme (tl)	Gorizia (tl)
Savoie (Sth)	Piacenza (tl)
Alpes de Hte Provence (sth)	Livorno (n 5)
Hautes-Alpes (sth)	Namur
Gard (perm.)	

2) Plan 3-4

Le quatrième quadrant contient à titre accessoire les variables concernant la productivité dans l'agriculture et la densité. Les circonscriptions intéressées sont Var, Varese, Como (n 1), Savona (perm.)

b - circonscriptions non caractérisées précédemment et de nature agricole

1) Terres labourables et grandes exploitations

Aisne	Landes
Oise	Lot-et-Garonne
Somme	Gers
Cher	Tarn
Eure-et-Loir	Tarn-et-Garonne
Indre	Aube
Indre-et-Loir	Marne
Loir-et-Cher	Jura
Loiret	Haute-Saône
Côtes-du-Nord	Yonne
Finistère	Mantova
Morbihan	Parma
Vienne	

2) Terres labourables et surfaces toujours en herbe

Ille-et-Vilaine ; Deux-Sèvres ; Dordogne ; Oberpfalz.

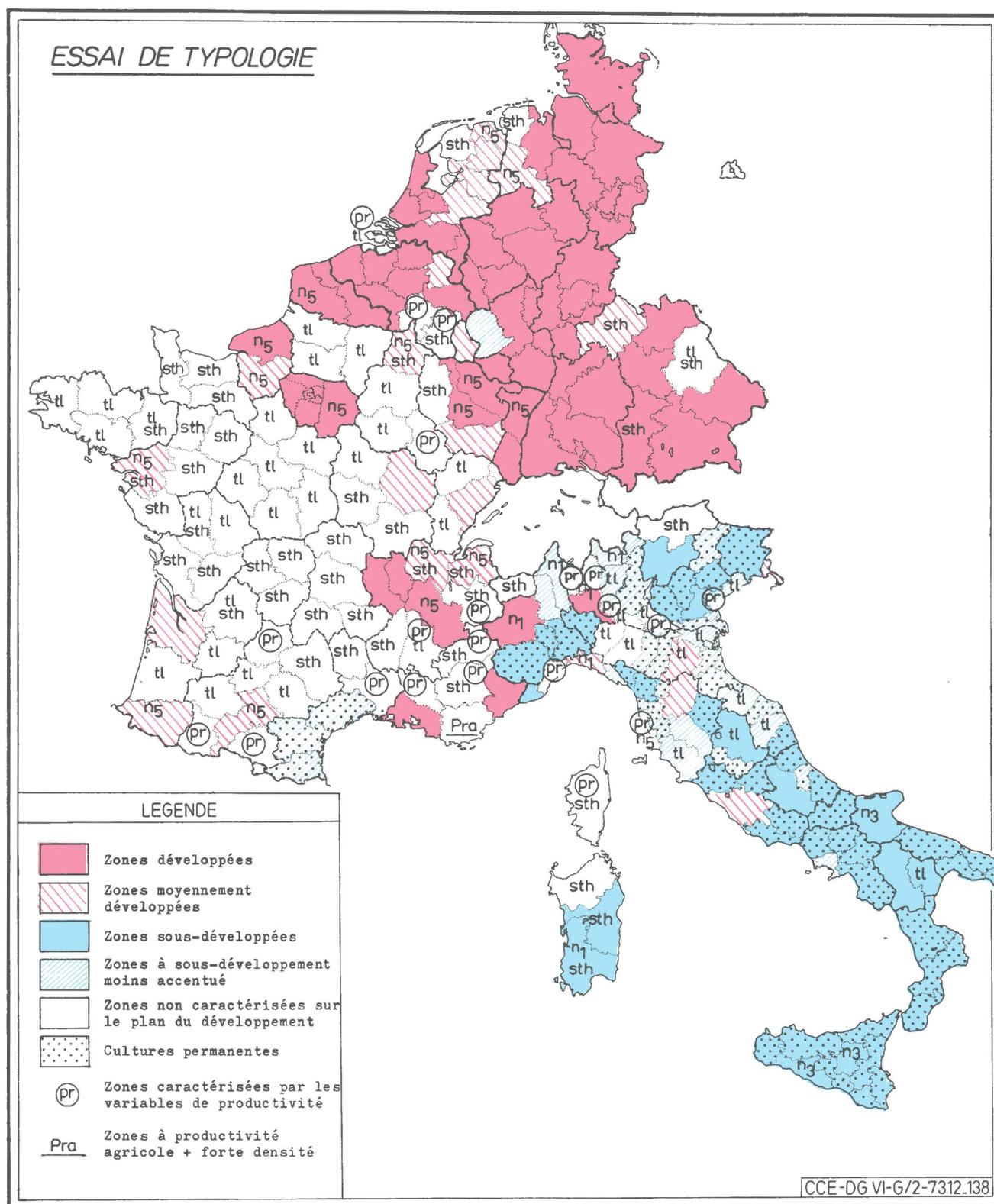
3) Surfaces toujours en herbe

Calvados	Nièvre
Manche	Saône-et-Loire
Orne	Allier
Maine-et-Loire	Cantal
Mayenne	Haute-Loire
Sarthe	Puy-de-Dôme
Vendée	Ardèche
Charente	Lozère
Charente-Maritime	Friesland
Corrèze	Aurich
Creuse	Niederbayern
Haute-Vienne	Bolzano
Aveyron	Sassari
Meuse	Aosta

4) Circonscriptions sans caractéristiques précises

Dans cette catégorie figure seulement Trier dont la seule caractéristique qui apparaît en recourant au 5^e facteur, est le nombre d'exploitations entre 10 et 20 ha.

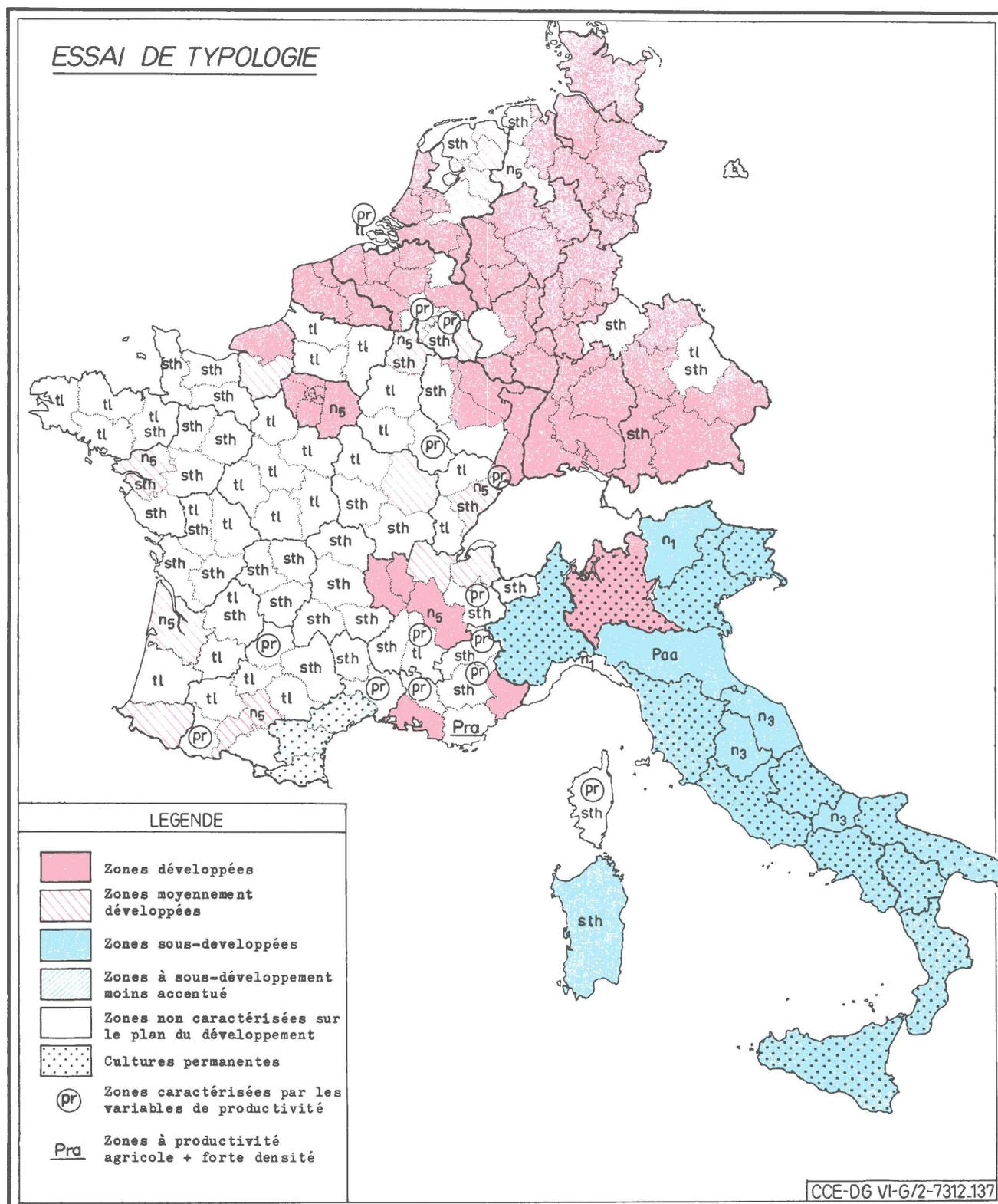
ANALYSE ECONOMIQUE GENERALE 25 VARIABLES



Circonscriptions analysées : FRANCE, départements. R.F.d'ALLEMAGNE, Regierungsbezirke. ITALIE, provincia. BELGIQUE et PAYS-BAS, provinces. GRAND-DUCHE de LUXEMBOURG.

N.B.- Les abréviations non reprises dans la légende figurent page 65 et page 66 .

ANALYSE ECONOMIQUE GENERALE 25 VARIABLES



Circonscriptions analysées : FRANCE, départements. R.F.d'ALLEMAGNE, Regierungsbezirke. ITALIE, regioni. BELGIQUE et PAYS-BAS, provinces. GRAND-DUCHE de LUXEMBOURG.

N.B.- Les abréviations non reprises dans la légende figurent page 65 et page 66 .

CHAPITRE IV. RECAPITULATION

Afin d'élaborer une typologie des régions de la Communauté Européenne des six, on a utilisé l'analyse factorielle des correspondances. Le recours à cette technique a l'avantage d'autoriser la prise en compte simultanée d'un grand nombre de critères et cela sans nécessiter des hypothèses à priori tant sur les caractéristiques mathématiques des variables (normalité, continuité) que sur leur poids respectif dans l'analyse.

Cette méthode a pour résultat une représentation simplifiée de l'ensemble de départ. Avant traitement, les circonscriptions sont définies dans un espace de dimension importante, dimension égale au nombre de critères décrivant chacune des régions. Après traitement, on a affaire à un espace de dimension réduite (4 pour ce qui nous concerne). Le référentiel d'arrivée est constitué d'axes statistiquement indépendants se définissant comme une combinaison linéaire des variables de départ.

L'interprétation se fait en repérant par rapport à un système d'axes du référentiel d'arrivée, les associations de points circonscriptions (ou de points régions) et de points variables c.à.d les caractéristiques. Qu'une circonscription soit caractérisée par une variable donnée (proximité graphique entre ces deux points traduisant une proximité mathématique) signifie que par rapport aux autres régions, celle-ci a ce caractère plus accentué. Ce n'est pas forcément la caractéristique la plus apparente qui se manifeste, mais l'aspect distinctif le plus marqué, relativement à l'ensemble des données de base.

La définition de zones homogènes d'après la situation de l'agriculture et compte-tenu du contexte économique général a conduit à deux sortes d'analyses :

- L'une, plus spécifiquement agricole, concerne la France subdivisée en 89 départements. Les 21 provinces du Bénélux, la R.F. d'Allemagne découpée en 32 *Regierungsbezirke* (les villes-états n'étant pas considérées en tant que telles) et les régions italiennes subdivisées en zones altimétriques (plaine, montagne, colline) soit 51 circonscriptions.

- L'autre, plus générale, utilise le même découpage pour la France, le Bénélux et la R.F. d'Allemagne, mais elle a trait aux 93 provinces italiennes, sans qu'on puisse établir une correspondance entre zones administratives et zones altimétriques. Le seul niveau de concordance possible se situe pour les *regioni*, au nombre de 20.

a - Pour décrire les régions de la Communauté sur le plan agricole on a retenu pour chaque classe de surface (classe 2 ha; 2-5 ha; 5-10 ha; 10-20 ha; 20-50 ha et plus de 50 ha). Les principaux éléments du système de production (surface des terres labourables dont céréales, surfaces toujours en herbe, cultures permanentes).

Par ailleurs, on a fait intervenir les surfaces occupées par les exploitants :

- de moins de 50 ans (variable dite "jeune" : E_j)
- de plus de 50 ans ayant un successeur potentiel (E_{as})
- de plus de 50 ans dépourvus de successeur potentiel (E_{ss})
- de plus de 50 ans avec ou sans successeurs potentiels (variable dite "vieux" : E_v; E_v = E_{as} + E_{ss}).

Cet ensemble de critères permet d'élaborer une typologie agricole dont les grandes lignes sont :

- la prédominance de structures favorables dans le Bassin Parisien, les régions de programme Picardie, Champagne et Centre. Ces zones sont caractérisées par l'importance des surfaces occupées par les exploitations de plus de 50 ha avec système céréalier et présence de chefs d'exploitation de moins de 50 ans ayant des successeurs.

- L'existence globalement en Italie de structures peu favorables, liées à un système basé sur les cultures permanentes. Sur ces catégories d'exploitations sont surtout présents des exploitants âgés et dépourvus de successeurs.

- La présence d'exploitations de taille moyenne liées à un système de polyculture-élevage et avec des chefs d'exploitations de moins de 50 ans en R.F. d'Allemagne, dans le Bénélux, et dans la France de l'Ouest et du Sud-Ouest.

- L'existence de zones particulières telles que Sardegna, Lozère, Alpes de Hte Provence, Valle d'Aosta, Trentino Alto Adige, où l'on a de grandes exploitations de type extensif avec prédominance d'exploitants âgés sans successeur.

b - Pour décrire les régions de la Communauté sur un plan économique général on a retenu des variables telles que la répartition de la population active et du Produit Intérieur Brut en 3 secteurs. La productivité du travail dans chacun de ces secteurs, ainsi que la population totale et la densité. Ceci a été complété par la prise en compte simplifiée du système de production des exploitations ainsi que de leur structure.

Les résultats de cette analyse permettent de dégager grossièrement 3 grandes zones :

- la première globalement défavorisée quant à l'importance des activités secondaires et tertiaires : l'Italie. Font exception les provinces de Milano et Turino, et dans une moindre mesure Roma, Trieste, Genova, Firenze, et Bologna.

- la seconde où dominent les activités industrielles ainsi que les services recouvrant grossièrement l'Europe Lotharingienne avec R.F. Allemagne, le Bénélux, la France du Nord, de l'Est, de la région parisienne, de la région lyonnaise, et

les Bouches du Rhône. Dans l'Europe septentrionale sont exclues un certain nombre de régions dont Unterfranken, Osnabruck, Groningen, Drenthe, Overijssel, Gelderland, Luxembourg (belge et Grand Duché) Limburg (belge) ... dont la situation sur le plan des activités secondaires et tertiaires n'est pas comparable.

- La dernière enfin qui a pour caractéristique d'apparaître ni favorisée, ni défavorisée quant à l'industrie et aux services et qui correspond à la France, déduction faite des départements précédemment cités. La dominante qui se manifeste est l'agriculture.

c - L'analyse agricole et l'analyse économique générale peuvent être combinées en distinguant dans chacune des grandes catégories de l'analyse générale les diverses situations agricoles (1). Pour ce qui est de l'Italie, on raisonne au niveau des "province". Pour cela on retient :

- comme système de production celui qui apparaît dans l'analyse générale pour le découpage par province. Quant une circonscription de base n'est pas caractérisée on se réfère aux caractéristiques de la région à laquelle elle appartient.
- comme caractéristiques démographiques (âge des chefs d'exploitation, présence ou non de successeur) les données de l'analyse agricole en supposant que toutes les provinces rattachées à une même région ont un profil identique à celui de la région.
- comme taille d'exploitation le résultat qui apparaît au niveau des provinces ou à défaut celui des régions soit de l'analyse globale, soit de l'analyse agricole.

La classification est établie en cherchant à ranger à

(1) Rappelons que certains résultats relatifs à l'analyse sont douteux, notamment en ce qui concerne les systèmes de production de Oberbayern et Mittelfranken en raison d'erreurs soit au niveau du dépouillement, soit au niveau de la publication des résultats de l'Enquête communautaire sur les structures agricoles.

l'intérieur de chaque grande catégorie générale les circonscriptions en fonction d'une situation agricole allant du favorable au défavorable. Dans un système de production donné on classe les régions d'après le critère taille d'exploitations : grandes (plus de 50 ha), moyenne (20 à 50), petites (5-20) et très petites (moins de 5 ha).

Les résultats de l'essai de typologie présentés dans les pages suivantes (p. 145 à 149) sont repris sous forme de cartes et de tableau encartés à la fin de ce volume.

A - Circonscriptions caractérisées par les activités secondaires et tertiaires

a - Zones à système de production basé sur les terres labourables et les céréales, à exploitations

- grandes avec exploitants de moins de 50 ans et exploitants de plus de 50 ans pourvus de successeurs : Seine (y compris Seine-et-Oise), Seine et Marne.
- moyennes et exploitants de moins de 50 ans : Hannover, Lüneburg, Düsseldorf, Pas de Calais.
- moyennes sans caractéristique démographique : Hildesheim, Aachen, Detmold, Köln, Braunschweig, Oberbayern, Antwerpen
- petites et exploitants de moins de 50 ans : Hainaut, Rhône, Noord-Brabant, Darmstadt-Wiesbaden, Liège, Noord-Holland, Oost-Vlaanderen, Nordbaden, Saarland, Nordwürttemberg, Bas-Rhin, Limburg-Nederland, Rheinhessen-Pfalz, Mittelfranken, Kassel, Oberfranken, Koblenz-Montabaur, Südbaden, Schwaben, Südwürttemberg.

b - Zones à système de production basé sur les terres labourables et les surfaces toujours en herbe et d'exploitations moyennes

- Régions avec des exploitants de moins de 50 ans : Nord, Utrecht, West-Vlaanderen, Schleswig-Holstein, Arnsberg, Münster.
- Régions sans caractéristique démographique : Haut-Rhin.

c - Zones à système de production basé sur les surfaces toujours en herbe, à exploitations :

- grandes avec exploitants de moins de 50 ans et exploitants de plus de 50 ans pourvus de successeurs : Meurthe et Moselle.

- moyennes avec exploitants de moins de 50 ans : Seine-Maritime, Stade, Oldenburg.
- moyennes sans caractéristique démographique : Moselle, Isère.
- petites sans caractéristique démographique : Zuid-Holland, Brabant belge, Loire.
- très petites sans caractéristique démographique : Alpes-Maritimes.

d - Zones à système de production basé sur les cultures permanentes avec de très petites exploitations

- avec des exploitants de plus de 50 ans sans successeurs : Torino.
- sans caractéristique démographique : Milano, Bouches-du-Rhône.

B - Circonscriptions moyennement marquées par les critères d'industrialisation et d'activités tertiaires

a - Zones à système basé sur les terres labourables et les céréales, à exploitations

- grandes et exploitants de moins de 50 ans et exploitants de plus de 50 ans pourvus de successeurs : Eure, Côte-d'Or.
- moyennes sans caractéristique démographique : Haute-Garonne.
- petites et exploitants de moins de 50 ans : Pyrénées Atlantique, Overijssel, Gelderland, Unterfranken.

b - Zones à système basé sur les terres labourables et les surfaces toujours en herbe à exploitations moyennes et exploitants de moins de 50 ans : Ain, Loire-Atlantique, Grand-Duché de Luxembourg, Groningen, Osnabrück.

c - Zones à système basé sur les surfaces toujours en herbe, à exploitations :

- grandes avec des exploitants de moins de 50 ans et exploitants de plus de 50 ans pourvus de successeurs : Ardennes.
- moyennes sans caractéristique démographique : Vosges, Doubs.
- petites sans caractéristique démographique : Haute-Savoie, Drenthe, Limburg belge.

d - Zones à système basé sur les cultures permanentes, à exploitations :

- petites sans caractéristique démographique : Trieste, Gironde.

- très petites avec exploitants de plus de 50 ans et exploitants de plus de 50 ans sans successeurs : Genova.
- très petites avec exploitants de plus de 50 ans sans successeurs : Firenze, Roma.

C - Circonscriptions non caractérisées par les activités secondaires et tertiaires

a - Zones à système basé sur les terres labourables et céréales, à exploitations :

- grandes avec exploitants de moins de 50 ans et exploitants de plus de 50 ans pourvus de successeurs : Aisne, Oise, Eure-et-Loir, Indre, Aube, Marne, Yonne, Haute-Marne.
- grandes sans caractéristique démographique : Loir et Cher, Loiret, Somme.
- moyennes avec exploitants de plus de 50 ans sans successeurs : Gers, Charente-Maritime (caractérisée d'autre part par les cultures permanentes).
- moyennes sans caractéristique démographique : Indre et Loire, Landes, Vienne, Lot et Garonne, Tarn et Garonne, Parma.
- petites sans caractéristique démographique : Piacenza, Cremona, Mantova.
- petites avec exploitants de moins de 50 ans : Hautes-Pyrénées, Côtes-du-Nord, Finistère, Ille-et-Vilaine, Morbihan, Niederbayern, Oberpfalz.
- petites avec exploitants de plus de 50 ans et exploitants de plus de 50 ans pourvus de successeurs : Gorizia.

b - Zones à système basé sur les terres labourables et les surfaces toujours en herbe à exploitations :

- moyennes avec exploitants de moins de 50 ans : Ariège, Zeeland, Deux-Sèvres, Vendée, Maine et Loire.
- moyennes avec exploitants de plus de 50 ans dépourvus de successeurs : Tarn, Charente.
- moyennes avec exploitants de moins de 50 ans et exploitants de plus de 50 ans : Dordogne, Friesland.

c - Zones à système basé sur les surfaces toujours en herbe, à exploitations :

- grandes avec exploitants de moins de 50 ans et exploitants de plus de 50 ans pourvus de successeurs : Meuse, Nièvre Allier.

- grandes avec exploitants de plus de 50 ans sans successeurs : Alpes de Haute Provence, Lozère, Valle d'Aosta, Bolzano, Sassari.
- moyennes avec exploitants de moins de 50 ans : Mayenne, Sarthe, Creuse, Haute-Vienne, Aveyron, Puy-de-Dôme, Jura, Haute-Saône, Hautes-Alpes.
- moyennes sans caractéristique démographique : Drôme, Calvados, Orne, Saône-et-Loire, Cantal, Aurich, Corse (ce département est caractérisé d'autre part par les cultures permanentes).
- petites sans caractéristique démographique : Lot, Manche, Belfort, Corrèze, Haute-Loire, Ardèche, Namur, Luxembourg belge.
- très petites sans caractéristique démographique : Savoie.

d - Zones à système basé sur les cultures permanentes, à exploitations :

- petites sans caractéristique démographique : Vaucluse.
- très petites avec exploitants de plus de 50 ans et exploitants de 50 ans dépourvus de successeurs : Savona.
- très petites avec exploitants de plus de 50 ans sans successeurs : Livorno.
- très petites sans caractéristique démographique : Gard, Var, Varese, Como.

D - Circonscriptions défavorisées quant aux caractères d'industrialisation et d'activités tertiaires

1 - Circonscriptions moyennement défavorisées

a - Zones à système basé sur les terres labourables et céréales, à exploitations :

- petites avec exploitants de plus de 50 ans et exploitants de plus de 50 ans dépourvus de successeurs : Macerata, Pesaro e Urbino.
- petites sans caractéristique démographique : Ferrara, Bologna.
- très petites avec exploitants de plus de 50 ans sans successeurs : Grosseto, Siena.
- très petites sans caractéristique démographique : Bergamo.

b - Zones à système basé sur les cultures permanentes, à exploitations :

- petites avec exploitants de plus de 50 ans et exploitants de plus de 50 ans sans successeurs : Ancona.

- petites sans caractéristique démographique : Forli, Modena, Ravenna, Reggio Emilia, Terni.
- très petites avec exploitants de plus de 50 ans et exploitants de plus de 50 ans dépourvus de successeurs : Pescara, La Spezia, Belluno, Rovigo, Imperia, Vercelli et Novara.
- très petites avec exploitants de plus de 50 ans sans successeurs : Pyrénées-Orientales, Pistoia, Pisa, Massa-Carrara.
- très petites sans caractéristique démographique : Aude, Hérault, Brescia.

2 - Circonscriptions très défavorisées

a - Zones à système basé sur les terres labourables et les céréales, à exploitations :

- petites sans caractéristique démographique : Molise, Perugia.
- très petites avec exploitants de plus de 50 ans et exploitants de plus de 50 ans sans successeurs : Venezia.
- très petites avec exploitants de plus de 50 ans sans successeurs : Napoli.
- très petites sans caractéristique démographique : Potenza, Matera.

b - Zones à système basé sur les surfaces toujours en herbe à exploitations grandes avec des exploitants de plus de 50 ans et exploitants de plus de 50 ans dépourvus de successeurs : Trento, Cagliari, Nuoro.

c - Zones à système basé sur les cultures permanentes, à exploitations :

- petites sans caractéristique démographique : Foggia, Agrigento, Enna.
- petites avec exploitants de plus de 50 ans et exploitants de plus de 50 ans avec successeurs : Udine, Pordenone, Ascoli.
- très petites avec exploitants de plus de 50 ans et exploitants de plus de 50 ans dépourvus de successeurs : Alessandria, Asti, Cuneo, Treviso, Padova, Vicenza, Verona, Chieti, l'Aquila, Teramo.
- très petites avec exploitants de plus de 50 ans dépourvus de successeurs : Lucca, Arezzo, Benevento, Viterbo, Prosinone, Rieti, Latina, Avellino, Salerno, Caserta, Catanzaro, Cosenza, Reggio di Calabria, Lecce, Bari, Brindisi, Taranto, Palermo, Trapani, Catania, Messina, Siracusa, Caltanissetta, Ragusa.
- très petites sans caractéristique démographique : Sondrio, Pavia.

Notons que si les grandes lignes de la typologie à laquelle nous aboutissons est valable certaines réserves doivent être faites dans le détail (1).

Les premières tiennent aux hypothèses retenues pour définir le système agricole ainsi que les caractéristiques de structure et d'ordre démographique pour les provinces italiennes. En effet, au moins pour les aspects démographiques on a eu recours chaque fois aux résultats obtenus lors de l'analyse agricole au niveau des régions, d'où l'existence d'approximations. Celles-ci sont moins prononcées pour la structure, et le système.

Les deuxièmes tiennent à ce que l'on a fait figurer à la suite les résultats concernant chaque système de production. Or, il y a quelque arbitraire à admettre que les petites exploitations à système terres labourables-céréales se situent mieux, quant à leurs résultats économiques, que les petites exploitations de cultures permanentes. Il faut donc signaler qu'il y a des recoupements entre systèmes de production, les derniers résultats de l'un n'étant pas nécessairement inférieurs aux résultats du suivant.

Les dernières, les moins importantes sont relatives à la classification retenue à l'intérieur d'un système de production. La présence d'exploitants âgés de moins de 50 ans peut être retenue comme un handicap sur les petites exploitations car bloquant l'évolution des structures, ou au contraire comme facteur favorable dans des exploitations suffisantes, car signe de dynamisme. Il en est de même de la présence de successeurs.

Plus généralement, les limites de ce travail tiennent aux données traitées, qu'il s'agisse de leur nature même ou de la qualité des résultats disponibles. Ainsi, au plan

(1) Notons qu'à la classification précédente il faut rajouter Trier que l'on n'a pas pu faire entrer dans aucune des 3 grandes catégories quant au caractère industrialisation et services. Cette circonscription est orientée vers les terres labourables et céréales avec de petites exploitations et des exploitants de moins de 50 ans.

de l'analyse générale il faut noter le caractère statique de celle-ci, car on connaît seulement la situation à un moment donné sans avoir de critère indiquant l'évolution en cours. Sur le plan agricole, on peut déplorer l'absence de données monétaires qui ont conduit à caractériser les systèmes de production à l'aide de seuls critères physiques (surface occupée par grande catégorie de culture).

A N N E X E S

ANNEXE I - PRESENTATION MATHEMATIQUE SIMPLIFIEE DE L'ANALYSE
DES DONNEES

Quand on cherche à rapprocher des entités telles que les régions, dont le contenu est à la fois physique, humain et économique, l'idéal est de caractériser chacune d'elles à l'aide du plus grand nombre possible de variables, ce qui permet d'avoir une description assez correcte de chaque entité. Malheureusement, l'esprit humain étant dans l'impossibilité de considérer simultanément un grand nombre d'éléments, il est obligé, pour se représenter l'ensemble des entités, de procéder à une simplification.

Si l'on raisonne sur un nombre important d'unités (par exemple les 234 circonscriptions définies au niveau de la C.E.E.) les rapprochements entre unités ne sont possibles, quand on opère avec les moyens habituels, qu'après avoir effectué une simplification a priori des caractères définissant ces entités. C'est-à-dire que l'on cherche à caractériser chaque région à l'aide de quelques variables synthétisant le maximum d'informations. Ces variables synthétiques peuvent être, dans le cas qui nous intéresse, le revenu des agriculteurs d'une région rapporté au revenu des autres catégories socio-professionnelles de la région, ou bien le revenu régional des agriculteurs comparé au revenu communautaire moyen des agriculteurs. De même, on peut utiliser le critère superficie des exploitations, densité de population, etc...

Le développement de nouvelles méthodes statistiques, lié aux progrès dans la conception et l'utilisation des ordinateurs, permet d'envisager le traitement global d'un ensemble

important de données. La simplification de l'ensemble intervient a posteriori, grâce à l'utilisation de procédures mathématiques plus ou moins complexes. Ces méthodes aboutissent à des rapprochements entre unités qui sont aussi "objectifs" que possible, car on a réalisé les calculs sur toutes les variables qu'on pouvait supposer intervenir dans le phénomène étudié.

Les deux méthodes qui de notre point de vue, semblent les plus intéressantes sont l'analyse factorielle, sous sa forme la plus élaborée (analyse des correspondances), et la classification automatique. Nous présenterons ces méthodes en posant les principes qui les guident, sans chercher à démontrer les résultats. La bibliographie figurant en fin d'exposé permettra au lecteur intéressé de se reporter aux publications à caractère mathématique traitant ces sujets.

I - L'ANALYSE FACTORIELLE DES CORRESPONDANCES

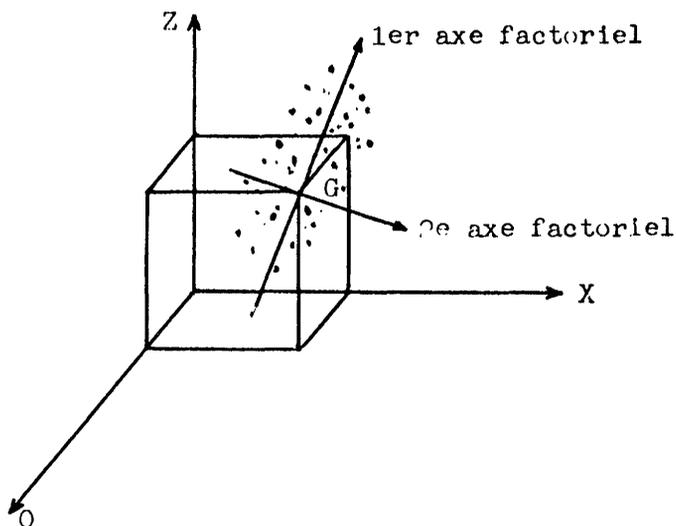
A - Présentation générale

1) L'analyse factorielle

L'analyse factorielle permet de prendre en compte un très grand nombre de critères, sans distinguer entre variables explicatives et variables expliquées, ou entre variables indépendantes et dépendantes. Tous les critères sont expliqués à l'aide d'un certain nombre de facteurs. Ceux-ci sont des variables hypothétiques déterminées par calcul et dont la signification économique n'est pas toujours aisément déchiffrable. L'intérêt de la méthode tient à ce que l'on peut expliquer chacun des critères retenus a priori au moyen d'une relation linéaire le liant aux facteurs. Les facteurs sont statistiquement indépendants.

En raisonnant en termes géométriques, la grande masse d'informations relative à un problème complexe peut s'interpréter comme un nuage de points dans un espace de grandes dimensions. L'analyse factorielle donne de nouveaux axes, les facteurs, permettant de reconstituer avec le minimum de coordonnées, la position de nuage de points. Pour cela, on projette le nuage sur des sous-espaces choisis de manière à ce que la déformation soit la plus faible possible, et ce, afin de réduire au maximum la perte d'informations. L'ajustement se fait en déterminant successivement certaines directions d'allongement du nuage, auxquelles on associe ensuite un facteur (cf. figure).

Représentation du nuage de points dans un espace à 3 dimensions



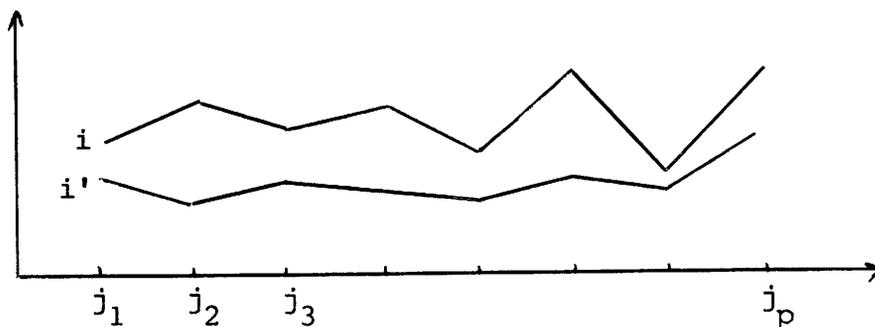
Schématiquement, on peut dire qu'il y a deux types d'analyse factorielle : l'analyse en composantes principales, plus classique, et l'analyse de correspondance. La première est valable uniquement pour des variables prenant des valeurs numériques continues et normales. D'autre part, l'interprétation des résultats n'est pas sans poser des problèmes théoriques et pratiques. Mieux vaut se référer à l'autre méthode due à J. P. BENZECRI, Professeur à l'Institut de Statistique de

l'Université de Paris. Bien que de création récente, elle a fait l'objet de nombreuses applications tant en linguistique, biologie et plus récemment en économie.

2) L'analyse des correspondances

Le principal avantage de cette méthode est de nécessiter aucune hypothèse a priori concernant l'ensemble de données que l'on cherche à représenter de façon simplifiée. L'interprétation des résultats est plus simple et meilleure qu'avec l'analyse aux composantes principales.

Au lieu de raisonner sur les valeurs absolues des caractères définissant les individus étudiés, on raisonne en terme de fréquences ce qui permet de définir le profil des individus (les régions en l'occurrence). Pour définir l'identité entre individus on utilise une notion de distance. Soient deux individus i et i' décrits par les valeurs $p(i, j)$ et $p(i', j)$ $j \in (1, n)$ et où $p(i, j)$ est la valeur de la variable j dans l'observation i . Ces deux individus i et i' sont considérés comme identiques si $p(i, j) = \lambda p(i', j)$ pour toutes les valeurs de j . Graphiquement cette similitude entre points se traduit par des profils identiques si l'on porte les points définissant les individus sur une échelle semi-logarithmique.



La notion de distance entre l'ensemble des individus se traduit par un tableau des distances, ou matrice. Connaissant ces distances, le problème consiste à trouver la figure géométrique qui respecte au mieux toutes ces distances.

Répartition du nombre d'exploitation par classes de SAUtilisée en 1967

Régions	nombre d'exploitations par classe (en 10)										total en 10 k (i)
	< 1	1-2	2-5	5-10	10-20	20-50	50-100	> 100			
1 Nord	1	2	3	4	5	6	7	8			5 311
2 Picardie	191	327	627	754	1 616	1 576	187	32			3 510
3 Région Parisienne	126	192	327	288	564	1 232	501	278			1 389
4 Centre	144	106	157	123	129	291	227	211			8 882
5 Haute Normandie	698	643	945	1 043	1 418	2 554	1 158	422			3 713
6 Basse Normandie	60	283	628	599	745	966	350	82			8 406
7 Bretagne	92	525	1 540	1 740	2 141	1 937	375	55			16 268
8 Pays de Loire	257	1 392	2 808	3 260	5 549	2 585	56	1			14 840
9 Poitou Charentes	970	879	1 647	2 284	4 015	4 638	374	34			9 356
10 Limousin	560	587	1 067	1 238	2 205	3 059	566	74			5 340
11 Aquitaine	23	151	652	1 223	1 765	1 277	220	30			12 912
12 Midi Pyrénées	604	856	2 022	3 297	3 891	1 996	212	34			14 099
13 Champagne	628	677	1 498	2 623	4 306	3 688	554	125			3 928
14 Lorraine	399	305	387	280	340	1 075	838	302			4 697
15 Alsace	269	370	763	716	758	1 149	538	133			3 653
16 Franche Comté	328	445	795	798	856	390	34	7			3 320
17 Bourgogne	77	136	294	747	832	133	188	13			6 676
18 Auvergne	420	409	801	982	1 153	1 703	952	258			8 144
19 Rhône-Alpes	298	296	873	1 581	2 427	2 070	502	96			15 179
20 Languedoc	761	1 164	2 892	3 906	4 132	2 045	245	35			11 029
21 Provence Côte d'Azur	2 631	1 755	2 178	1 812	1 365	897	276	115			8 272
	1 693	1 245	1 829	1 394	1 122	745	162	80			
k' (j) =	11 267	12 759	24 739	30 758	41 332	37 208	8 516	2 419			168 993 = K

De façon symétrique, on peut concevoir l'analyse du profil des caractères des individus.

On peut reprendre plus en détail les divers points essentiels à la compréhension de l'analyse des correspondances.

B - La correspondance

Raisonnons à partir d'un exemple simple, soit les 21 régions de programme françaises décrites dans un tableau rectangulaire variables X régions, ou matrice des données. Il s'appelle également tableau de correspondance, car ce tableau définit les correspondances qui existent entre les éléments de l'ensemble I et les éléments de l'ensemble J, qu'il s'agisse de mesures ou de fréquences.

L'ensemble des caractères retenus pour décrire les régions sera appelé conventionnellement ensemble des variables et noté J. Si l'on prend comme variables la répartition du nombre d'exploitations selon 8 classes de surface (1 ha ; 1 à 2 ; 2 à 5 ; 5 à 10 ; 10 à 20 ; 20 à 50 ; 50 à 100 et 100 ha selon les résultats de l'enquête communautaire de 1967) j variera de 1 à 8. L'ensemble des régions correspond aux observations. Il est noté I avec i variant de 1 à 21.

Chaque point observation sera représenté dans l'espace R^J des variables par un système à 8 coordonnées qui sont les valeurs des 8 variables qui lui correspondent. De même, un point variable sera représenté dans l'espace R^I des observations par un système à 21 coordonnées qui sont les valeurs prises pour chaque variable par les 21 régions.

La matrice des données est reproduite dans le tableau n° 1. A l'intersection de la ligne i (par exemple ligne 20, le Languedoc) et la colonne j (par exemple la colonne 4, le nombre d'exploitations entre 5 et 10 ha) se trouve l'effectif de la case, le nombre k (i, j) qui est ici 1812.

Pour que la comparaison entre la répartition du nombre d'exploitations par classes de surface pour deux régions soit intéressante, mieux vaut confronter la part de chaque classe dans l'ensemble, que les nombres absolus.

On posera :

$$K = \sum_{i,j} k(i, j) \text{ l'effectif total du corpus} \quad K = 168\ 993$$

$$k'(j) = \text{somme des effectifs de la colonne } j \quad k'(4) = 30\ 752$$

$$k(i) = \text{somme des effectifs de la ligne } i \quad k(20) = 11\ 029$$

On pourra définir une estimation de probabilité $p(i, j)$ qui est la fréquence avec laquelle apparaît l'association (i, j) dans la population étudiée.

$$p(i, j) = \frac{k(i, j)}{K} \quad p(20, 4) = \frac{1\ 812}{168\ 993} = 0,0107223$$

$$p(i) = \frac{k(i)}{K} = \sum_j p(i, j) \quad p(20) = \frac{11\ 029}{168\ 993} = 0,0652630$$

$p(i)$ désigne l'importance relative de la région i .

$$p'(j) = \frac{k'(j)}{K} = \sum_i p(i, j) \quad p'(4) = \frac{30\ 752}{168\ 993} = 0,181920$$

$p'(j)$ désigne l'importance relative de la classe d'exploitation j .

Notons que $p(i, j)$ est assimilé à une probabilité conditionnelle, $p(i)$ et $p'(j)$ s'interprètent en termes de probabilités marginales.

On montre que la correspondance qui existe entre éléments des lignes et éléments des colonnes définit sur I une notion de proximité. Si i et i' s'associent de façon semblable aux éléments de J , ils sont proches. De même, pour j et j' si l'on raisonne par rapport aux éléments de I .

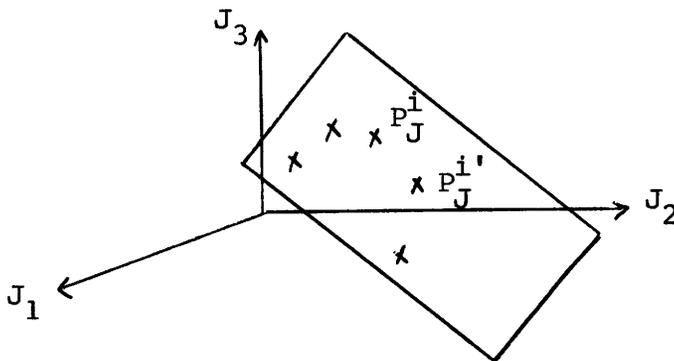
L'analyse des correspondances permet de décrire des proximités pouvant exister entre les lignes et les colonnes du tableau des correspondances (proximités entre formes, indépendamment des niveaux et des tailles), compte tenu des poids différents de ces lignes et de ces colonnes.

Bien que la méthode soit d'inspiration probabiliste, il faut ici remarquer que l'analyse des correspondances ne s'applique pas seulement à des tableaux de probabilités ou de fréquences. Elle donne des résultats intéressants avec des tableaux booléens (cas où l'on porte l'appréciation oui ou non codée 1 ou 0). Son emploi est également possible lorsqu'on a des variables discrètes telles que les notations de 1 à 5 par exemple, ou avec des tableaux en valeurs continues.

C - La distance

Géométriquement, il est possible de représenter dans l'espace R^J un nuage de I points, chacun d'entre eux ayant pour coordonnées la quantité $\left[\frac{p(i, j)}{p(i)} \right]_{j \in J}$ affectée de la masse $p(i)$.

Le vecteur dont les J composantes sont la quantité précédente est noté P_J^i .



Les I points p_J^i se trouvent tous dans un hyperplan (sous-espace à J-1 dimensions), étant donné que leurs coordonnées vérifient la relation

$$\sum_j \frac{p(i,j)}{p(i)} = 1 \text{ pour } i = 1, 2, \dots, I.$$

On dira que dans cet espace les deux modalités de l'observation I sont proches lorsque le passage de i à i' n'a que peu d'influence sur la répartition de l'ensemble des modalités j de la variable J. Mais on peut préciser cette notion de proximité par l'introduction d'une distance $d(i, i')$ entre les deux modalités de l'observation I.

On pourrait utiliser la formule classique :

$$d^2(i, i') = \sum_{j \in J} \left[\frac{p(i,j)}{p(i)} - \frac{p(i',j)}{p(i')} \right]^2$$

Deux points sont proches dans cet espace si les profils des lignes représentés par ces points très peu différents. Ceci revient à dire que l'on compare les proportions d'exploitation qui sont dans chaque classe de surface.

Cette manière de procéder permet de supprimer la taille de la modalité i ou i' dans la mesure des proximités, mais on introduit ainsi un inconvénient. En effet, dans la comparaison terme à terme des J éléments des profils des lignes i et i', on donne aux j éléments le même poids.

Si les effectifs mesurés par $p(j_0)$ de la colonne j_0 sont très peu importants dans le calcul de la distance $d^2(i, i')$ on aura le terme :

$$\left[\frac{p(i, j_0)}{p(i)} - \frac{p(i', j_0)}{p(i')} \right]^2$$

qui jouera un rôle négligeable dans la détermination des proximités, étant donné que son poids minime par rapport aux autres termes.

Si nous considérons la structure des exploitations agricoles de deux régions très différenciées (Bretagne et Champagne) $i = 7$ et $i' = 13$, on a :

$$d^2(7, 13) = \left[\frac{257}{16 \cdot 268} - \frac{399}{3 \cdot 928} \right]^2 + \left[\frac{1 \ 392}{16 \cdot 268} - \frac{305}{3 \cdot 928} \right]^2 + \left[\frac{2 \ 808}{16 \cdot 268} - \frac{387}{3 \cdot 928} \right]^2 \\ + \left[\frac{3 \ 620}{16 \cdot 268} - \frac{280}{3 \cdot 928} \right]^2 + \left[\frac{5 \ 549}{16 \cdot 268} - \frac{340}{3 \cdot 928} \right]^2 + \left[\frac{2 \ 585}{16 \cdot 268} - \frac{1 \ 075}{3 \cdot 928} \right]^2 \\ + \left[\frac{56}{16 \cdot 268} - \frac{658}{3 \cdot 928} \right]^2 + \left[\frac{1}{16 \cdot 268} - \frac{302}{3 \cdot 928} \right]^2$$

$$d^2(7, 13) = \frac{10^4}{639} (30,046 + 256 + 2 \ 781 \ 585 + 93 \ 399 + 264 \ 561 \\ + 53 \ 790 + 179 \ 898 + 24 \ 098) = \frac{1}{6,39} \times 34 \ 276 \ 330 = \\ = 5 \ 364 \ 057$$

Mais si l'on introduit le terme $\frac{1}{p(j)}$ cela permet d'avoir une expression pondérée qui atténue les trop importantes disparités. On a donc la distance :

$$d^2(i, i') = \sum_{j \in J} \frac{1}{p(j)} \left[\frac{p(i, j)}{p(i)} - \frac{p(i', j)}{p(i')} \right]^2$$

Si l'on reprend le calcul précédent avec la nouvelle distance :

$$d^2(i, i') = \frac{10^4}{639} \left(\frac{30 \ 046}{11 \ 267} + \frac{256}{12 \ 759} + \frac{2 \ 781 \ 585}{24 \ 739} + \frac{93 \ 399}{30 \ 752} + \frac{264 \ 561}{41 \ 332} \right. \\ \left. + \frac{53 \ 790}{37 \ 208} + \frac{179 \ 898}{8 \ 516} + \frac{24 \ 098}{2 \ 419} \right)$$

$$\text{soit } d^2(i, i') = \frac{1}{6,39} (266 + 2 + 1 \ 124 + 303 + 640 + 144 + \\ 2 \ 112 + 996) = \frac{1}{6,39} \times 5 \ 587 = 874,3348982$$

On voit qu'avec la nouvelle distance le poids de la colonne 2 par rapport à la colonne 3, particulièrement faible dans le calcul classique de la distance, est revalorisé.

Ainsi l'effet de taille n'est pas éliminé. Il subsiste, jouant un rôle semblable à celui d'un "poids" lorsqu'on élabore un indice synthétique.

La distance que l'on calcule pour comparer la structure des exploitations de deux régions, peut être utilisée pour comparer le profil de deux classes de surface et l'on a :

$$d^2(j, j') = \sum_{i \in I} \left[\frac{1}{p(i)} \frac{p(j, i)}{p(j)} - \frac{p(j', i)}{p(j')} \right]^2$$

- 1) Il apparaît que la symétrie entre régions et classes de surface est parfaite.
- 2) Deux régions sont proches si les probabilités conditionnelles des diverses modalités de la structure sont semblables (c'est-à-dire si les régions s'associent de la même façon aux mêmes classes de surface).
- 3) Si deux points sont proches dans l'espace, c'est-à-dire si $p(i, j_1) = p(i, j_2) \forall_{i \in I}$, il est possible de remplacer j_1 et j_2 par une troisième colonne j_0 telle que :

$$\begin{aligned} \forall_{i \in I} \quad p(j_0) &= p(j_1) + p(j_2) \\ p(i, j_0) &= p(i, j_1) = p(i, j_2) \end{aligned}$$

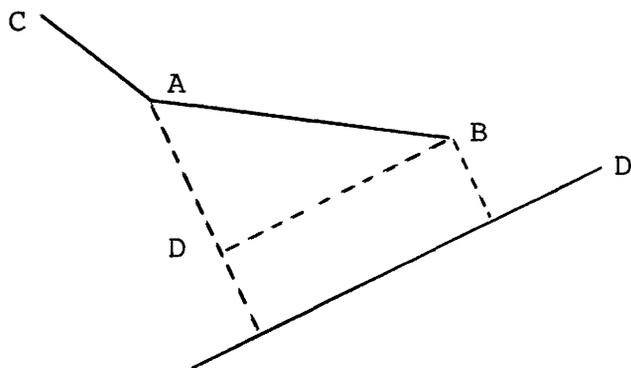
Autrement formulé, ce principe, dit d'équivalence distributionnelle, signifie que si les deux points P_J^i et $P_J^{i'}$ sont considérés comme un seul point affecté de la somme des masses de i et de i' , les distances entre les éléments de J ne sont pas changées. Cette propriété qui est très intéressante explique la stabilité des résultats.

D - Résolution du problème

On a défini un nuage de points R^J muni de la forme quadratique. Chaque point est affecté d'une masse qui est la probabilité propre de l'évènement (soit p_j). D'autre part, les J colonnes sont J vecteurs, ou J points de l'espace vectoriel R^I . Ainsi on peut assimiler le tableau de mesure à un solide, ce qui permet de raisonner par analogie avec la mécanique classique.

Le problème consiste à projeter les nuages de points observations (les régions), et les nuages de points variables (nombre d'exploitations par classe de surface), dans un sous-espace. La projection doit s'effectuer de manière à ce que le nuage de points se déforme le moins possible. Pour cela, on cherche le sous-espace dans lequel la distance entre deux points diffère aussi peu que possible de la distance précédemment définie.

Quand on considère trois points A, B, C par rapport au sous-espace à une dimension (la droite D), tel que les distances entre les projections soient le plus proche possible des distances retenues, on a à maximiser les distances projetées du type BD^2 .



En effet, d'après le théorème de Pythagore, la distance entre les 2 points A et B de l'espace R^J est $AB^2 = BD^2 + DA^2$ où DA est un vecteur de l'espace à $p-1$ dimensions orthogonal à D.

Au lieu de chercher le sous-espace D qui maximise la distance projetée BD^2 on peut, ce qui revient au même, chercher le sous-espace qui minimise AD^2 . Pour le nuage de points variables, le but sera de trouver le sous-espace donnant les longueurs des projections maximum en moyenne quadratique, de façon à ce que le nuage soit le moins déformé possible.

On montre que la solution consiste à trouver les valeurs et les vecteurs propres d'une matrice S' qui est la matrice de variance-covariance des variables, en prenant comme coordonnées des observations les quantités

$$\frac{p(i, j)}{p(i) \sqrt{p(j)}}$$

En fait, il suffit de rechercher les vecteurs propres et les valeurs propres de la matrice S de dimension (M, M) définie par :

$$S = [S_{jk}]$$

où le terme général de S_{ik} s'écrit :

$$S_{jk} = \sum_i \frac{p(i, j) p(i, k)}{p(i) \sqrt{p(j) p(k)}}$$

Sur notre exemple les éléments de S sont :

$$s_{11} = \frac{1}{11 \ 267} \left[\frac{191^2}{5 \ 311} + \frac{126^2}{3 \ 510} + \frac{144^2}{1 \ 389} + \dots + \frac{693^2}{8 \ 272} \right]$$

$$s_{12} = \frac{1}{\sqrt{11 \ 267 \times 12 \ 759}} \left[\frac{191 \times 327}{5 \ 311} + \frac{126 \times 192}{3 \ 510} + \frac{144 \times 106}{1 \ 389} + \dots + \frac{693 \times 1245}{8 \ 272} \right]$$

Par diagonalisation de la matrice, on obtient un certain nombre de valeurs propres $\lambda_1 \dots \lambda_r \dots \lambda_5$ (puisque sont extraits 5 axes factoriels dans l'exemple traité) avec $1 > \lambda_1 > \dots > \lambda_r > \lambda_5 > 0$. A chacune d'elle est associée un vecteur propre normalisé Z_r .

Si l'on reprend la matrice $S = [S_{jk}]$ on constate que cette matrice carrée est égale au produit d'une matrice rectangulaire X par son transposée t_X avec

$$X = [X_{ij}] \text{ et } X_{ij} = \frac{p(i, j)}{\sqrt{p(i) p(j)}}$$

$$S = X t_X \text{ les dimensions de } X \text{ étant } (M, N)$$

$$\text{d'où } (M, M) = (M, N) (N, M)$$

Au lieu de raisonner par rapport à la matrice de variance-covariance des variables, on aurait pu raisonner par rapport à celle des observations. Ceci nous aurait conduit à rechercher les valeurs propres d'une matrice R symétrique d'ordre N

$$R = [k_{ik}] = \sum_j \frac{p(i, j) p(k, j)}{\sqrt{p(i) p(k)} \times p(j)}$$

$$\text{Il apparaît que } R = t_X X$$

$$(N, N) = (N, M) (M, N)$$

On démontre que les deux matrices R et S ont les mêmes valeurs propres, avec le même ordre de multiplicité. Par ailleurs on passe des vecteurs propres de l'une, aux vecteurs propres de l'autre par une application linéaire simple (on a les mêmes vecteurs propres à $\frac{1}{\sqrt{\lambda}}$ près). Ces propriétés sont particulièrement importantes, car elles montrent l'existence d'une symétrie entre espace des observations et espace des variables appelée aussi dualité.

Plus concrètement à chaque λ_r est associé le vecteur propre Z_r ainsi qu'un couple de fonctions (G_r, F_r) . La première (G_r) porte sur l'ensemble des observations, G désignant les coordonnées des observations sur les axes factoriels. La deuxième (F_r) , porte sur l'ensemble des variables, F désignant les coordonnées des variables sur les axes.

Pour chaque variable j la projection sur le r ième axe factoriel Fr s'écrit :

$$Fr(j) = \sqrt{\lambda_r} \quad Z_r(j) \times \frac{1}{p(j)^{1/2}}$$

Pour chaque observation i la projection sur le r ième axe est :

$$Gr(i) = \frac{1}{\sqrt{\lambda_r}} \sum_j p(j,i) Fr(j)$$

$$Fr(j) = \frac{1}{\sqrt{\lambda_r}} \sum_i Gr(i) p(i,j)$$

Ce système montre bien que toute observation est le centre de gravité, au facteur multiplicatif près $\frac{1}{\sqrt{\lambda_r}}$, des variables qui la caractérisent. Ainsi une région est le centre de gravité des effectifs définissant les huit classes de surface des exploitations.

Par symétrie, une classe de surface est au centre de gravité, toujours au facteur multiplicatif près $\frac{1}{\sqrt{\lambda_r}}$, des 21 régions. Une classe de surface est donc proche des régions dans lesquelles elle représente une part importante des effectifs, ceci de façon relative. La région i est d'autant plus près de la variable j que celle-ci intervient avec beaucoup d'intensité dans le profil de cette région.

E - Présentation des résultats

Pratiquement la sortie des résultats apparaît sous une forme semblable à celle du tableau n° 2. On trouve à la suite les résultats représentant les coordonnées des variables sur les axes factoriels $Fr(j)$ et les coordonnées des observations sur les axes $Gr(i)$. En tête des sorties se trouvent les valeurs propres de chaque axe, ainsi que le pourcentage d'inertie expliqué par chaque axe (cette part d'inertie ou part de la variance est le rapport de la valeur propre correspondant à un axe sur la somme des valeurs propres de la matrice S).

Tableau 2 - Coordonnées des variables sur les axes

	axe 1	axe 2	axe 3	axe 4	axe 5
Valeur propre	0,106	0,093	0,019	0,005	0,001
% d'inertie	46,24 %	40,63 %	8,34 %	2,36 %	1,85 %
F (n ₁)	- 0,07	0,88	0,20	- 0,05	- 0,84
F (n ₂)	0,07	0,40	- 0,02	0,02	0,73
F (n ₃)	0,15	0,16	- 0,10	0,06	0,72
F (n ₄)	0,22	- 0,04	- 0,11	0,06	- 0,30
F (n ₅)	0,19	- 0,21	- 0,01	- 0,05	- 0,38
F (n ₆)	- 0,21	- 0,24	0,19	0,01	- 0,37
F (n ₇)	- 0,92	- 0,06	- 0,17	0,15	- 1,08
F (n ₈)	- 1,41	0,14	- 0,44	- 0,40	1,29
Coordonnées des observations sur les axes					
G (NRD)	0,02	- 0,22	0,17	- 0,02	0,36
G (BNM)	0,09	- 0,18	- 0,07	0,10	0,79
G (PIC)	- 0,78	- 0,16	- 0,05	- 0,13	1,09
G (PAR)	- 1,10	0,25	- 0,43	- 0,50	1,36
G (LAN)	0,09	0,83	0,08	- 0,02	- 0,33
G (PRO)	0,13	0,71	0,03	0,01	0,29

La région Nord désignée par NRD s'exprime par rapport au premier facteur sous la forme suivante :

$$G_1(\text{NRD}) = \frac{1}{\sqrt{0,106}} \times \frac{1}{5 \ 311} [191 F_1(n_1) + 327 F_1(n_2) + \dots + 32 F_1(n_8)]$$

$$G_1(\text{NRD}) = \frac{1}{\sqrt{0,106}} \times \frac{1}{5 \ 311} [191 \times -0,07 + 327 \times 0,07 + \dots + 32 \times -1,41]$$

Quant à la classe d'exploitation de moins de 1 ha (n_1) elle s'exprime toujours par rapport au premier axe sous la forme :

$$F_1(n_1) = \frac{1}{\sqrt{0,106}} \times \frac{1}{11\ 267} [191 G_1(\text{NRD}) + 126 G_1(\text{BNM}) + \dots + 1693 G_1(\text{PRO})]$$

$$F_1(n_1) = \frac{1}{\sqrt{0,106}} \times \frac{1}{11\ 267} [191 \times 0,02 + 126 \times 0,09 + \dots + 1693 \times 0,13]$$

Une représentation graphique des points observations et des points variables a l'avantage d'être plus explicite. Pour cela il suffit de reporter les projections des i et j sur les axes pris deux à deux (axes 1-2, 1-3, 2-3, ...) sur un graphique. On peut avoir séparément le graphique des régions, et celui des classes de surface. Mais, puisque l'étude des proximités des éléments de l'un des deux ensembles permet de déduire les proximités entre éléments de l'autre ensemble, il est possible d'avoir une représentation simultanée, sur le même graphique, des deux ensembles.

Dans l'analyse des régions en fonction du nombre d'exploitations réparties en 8 classes de surface, il a été extrait 5 facteurs, ceux-ci représentent la quasi totalité de l'information puisque la variance totale expliquée sur les 5 axes factoriels s'élève à 99,4 %. A eux seuls les trois premiers axes suffisent à expliquer 95,2 % de l'inertie (les deux premiers représentant déjà 86,87 %).

Dans l'analyse des correspondances on extrait uniquement un certain nombre de valeurs propres, ce nombre étant tel que :

$(T - \lambda_1 - \lambda_2 \dots - \lambda_r)$ a une probabilité supérieure à un certain seuil d'être dépassée par un Chi-carré à $(I-1)$ $(J-1)$ degrés de libertés, avec T la trace de la matrice et r le rang auquel on s'arrête. Quant à K sa valeur est

$$K = \sum_{i,j} k(i, j).$$

Ainsi au lieu d'avoir toutes les valeurs propres de la matrice S, on ne dispose que de 5 valeurs propres donnant une approximation $p'(i, j)$ de $p(i, j)$.

$$p'(i, j) = p(i) p(j) [1 + \lambda_1^{-1/2} F_1(j) G_1(i) + \dots + \lambda_5^{-1/2} F_5(j) G_5(i)]$$

Dans cette expression chaque terme est proportionnel à la racine carrée de la valeur propre qui lui correspond car $F(j)$ et $G(i)$ ayant pour variance λ sont d'ordre $\lambda^{1/2}$. Ainsi plus la valeur propre est élevée et plus le terme est grand, d'où une meilleure approximation de $p(i, j)$.

Le graphique n° 1 (1) donne la représentation simultanée de l'ensemble des régions et de l'ensemble des variables par rapport aux axes 1 et 2.

Une proximité entre deux points observations, tels que LAN (Languedoc) et PRO (Provence-Côte d'Azur) s'interprète en terme de comportement semblable vis-à-vis des variables. Si deux points variables sont proches, cela signifie qu'il y a corrélation entre ces variables, et la corrélation est d'autant plus forte que ces points sont éloignés de l'origine.

Une proximité entre points variables et points observations, telle la proximité existant entre LAN, PRO et n_1 (exploitations de moins de 1 ha) signifie que les deux régions Languedoc et Provence-Côte d'Azur ont un profil caractérisé par l'importance des très petites exploitations. C'est dans ces régions que l'on constate la plus forte proportion d'exploitations inférieures à 1 ha. Inversement, la variable n_1 a un profil d'où ressortent les régions LAN et PRO. Notons que n_2 (exploitations entre 1 et 2 ha) se situe entre ces deux régions et l'Alsace.

(1) Cf. Annexe 2 - Graphiques 1 (1ère série) Analyse 47 variables, 21 régions de programme françaises.

Parmi les autres proximités marquantes, il faut remarquer l'association entre les régions PAR (région parisienne) et CHP (Champagne) avec n_8 (exploitations > 100 ha et n_7 (50 à 100) ainsi qu'avec PIC (Picardie).

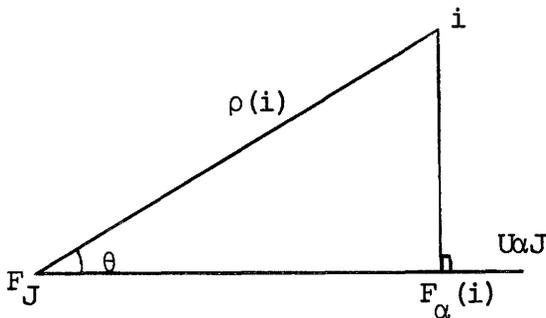
F - Notion de contribution absolue et de contribution relative

Soit F_j le centre de gravité de l'ensemble des points j et un point i appartenant à I . On a sous forme vectorielle :

$$F_J^i = \{ p(i, j) \} j \in J$$

Si dans le référentiel J dont le cardinal est A , A étant le nombre de facteurs que l'on peut interpréter, nombre $\leq J-1$, on prend un facteur u_α et la projection orthogonale de i sur ce facteur, on a :

$\rho(i)$ qui désigne la distance au sens de chicarré de i à F_J



$$\rho^2(i) = \sum_{j \in J} \left[\frac{p(i, j)}{p(i)} - p(j) \right]^2 \frac{1}{p(j)}$$

Avec $p(i)$ qui est le poids de i (rappelons que $p(i) = \sum_{j \in J} p(i, j)$), on a $\rho^2(i) \cdot p(i)$ qui constitue une variance dont l'expression générale $\sum_{i \in I} (x_i - m)^2 \frac{1}{n}$ se réduit à $\sum_{i \in I} x_i^2 p(i)$ quand $m = 0$ et $\frac{1}{n} = p(i)$.

$$\sum_{i \in I} \rho^2(i) \cdot p(i) = \sum_{\alpha \in A} \lambda_{\alpha} = \sum_{K \in J} A(K, K)$$

Ainsi $\rho^2(i) \cdot p(i)$ constitue la contribution absolue de i à la somme des valeurs propres.

La projection de i sur le facteur U_{α} est $F_{\alpha}(i)$ et l'on a :

$$\sum F_{\alpha}^2(i) p(i) = \lambda_{\alpha}$$

II - LA CLASSIFICATION AUTOMATIQUE

La classification automatique qui recouvre un ensemble de méthodes et d'algorithmes de classification, appelée aussi analyse arborescente est d'emploi courant en taxinomie, et maintenant en écologie. Il ne semble pas qu'en économie ces techniques soient beaucoup utilisées (1).

(1) Il n'existe à notre connaissance qu'un seul travail important en économie effectué à l'aide de la classification automatique. Il s'agit de "La nomenclature des activités du secteur travail mécanique du bois (BENIER, DEHEDIN, GIUDICELLI, PIETRI), Laboratoire de Statistique Mathématique - ISUP 1969. Ce travail fait pour l'INSEE a par ailleurs fait l'objet de divers prolongements.

Etant donné une collection d'objets décrits à l'aide d'un certain nombre de critères tant quantitatifs que qualitatifs, la classification automatique permet de répartir ces objets en groupes homogènes. On s'assure de l'homogénéité des classes obtenues en vérifiant que les individus à l'intérieur d'un groupe se ressemblent plus entre eux, qu'à des individus appartenant à des groupes différents.

On examinera d'abord les principes généraux de la classification automatique puis on exposera brièvement les bases de l'analyse des nuées dynamiques de DIDAY qui paraît fort intéressante pour notre propos.

A - Principes généraux

En l'état actuel de la statistique mathématique il n'existe pas à proprement parler de théorie de la classification. De ce fait il n'est pas question pour nous d'envisager de façon approfondie les problèmes mathématiques posés. Nous tenterons de voir simplement les principes généraux à la base des diverses méthodes en ayant en vue leur application au problème de typologie d'unités économiques de taille et de nature différentes.

1° - La ressemblance

La plupart des techniques de classification reposent sur la définition d'indices de similarité qui donnent une mesure de la ressemblance entre observations ou entre variables. Dans un but de simplification considérons uniquement la ressemblance entre observations.

Suivant DUCIMETIERE (7 p.7) considérons un individu $x \in E$ (soit une région de la Communauté) pour lequel on dispose

de mesures d'un paramètre (nombre d'exploitations correspondant à une classe de surface par exemple). Soit A l'ensemble des modalités que peut prendre le paramètre retenu. Pour ce même paramètre on suppose défini l'écart entre deux modalités a_i et a_j .

Mathématiquement l'indice de similarité est une fonction définie sur $A \times A$ vérifiant les conditions

$$\begin{cases} e(a_i, a_j) = e(a_j, a_i) \\ e(a_i, a_i) = 0 \end{cases}$$

Au lieu d'écart on peut raisonner en termes de similarité avec

$$\begin{cases} s(a_i, a_j) = s(a_j, a_i) \\ s(a_i, a_i) = v \end{cases} \quad v \text{ ayant dans ce cas la valeur maximale.}$$

On ne pose pas de règle stricte quand à la nature de e (resp. s) Il s'agit simplement d'une fonction numérique non négative définie sur $A \times A$. En fait, il est possible de définir sur A une vraie distance qui ne soit fonction que de e (resp. s).

Considérons à présent deux observations de E x et y ainsi que les diverses valeurs prises par les variables les décrivant. L'écart entre les deux individus x et y est une fonction numérique non négative définie sur $E_p \times E_p$ (E_p sous-ensemble de E de $A_1 \times A_2 \dots \times A_p$ où A_i constitue l'ensemble des modalités de la ième variable). Cette fonction vérifie les conditions

$$\begin{cases} D(x, y) = D(y, x) \\ \forall i \text{ les deux fonctions } e(x_i, y_i) \text{ et la restriction de } D(x, y) \text{ à l'ensemble } A \times A \text{ sont simultanément croissantes ou décroissantes. Les mêmes conditions peuvent être traduites en termes de similarité.} \end{cases}$$

En général on utilise la notion d'écart $D(x, y)$ lorsque $D(x, y)$ définit une distance sur E. Autrement on parle en termes de similarité.

2° - Notion de lien et classification ascendante

En reprenant la présentation de VOLLE et al [10] on peut voir, la ressemblance étant définie, comment on procède à partir de la notion de lien pour constituer des classes d'individus.

On peut définir entre A ensemble des i , et E ensemble des j la notion de lien.

$$\text{Lien } (A, E) = \sum_{i,j} \frac{[p(i,j) - p(i)p(j)]^2}{p(i)p(j)}$$

Ce lien s'exprime aussi sous la forme de la variance totale du nuage des X_i

$$\text{Lien } (A, E) = \sum_i p(i) \| X_i - G \|^2 \quad \text{où } G \text{ représente le barycentre des } X_i.$$

La classification automatique vise à isoler les regroupements de X_i tels que l'information apportée en supplément par la connaissance de $p(i,j)$ par rapport à $p(i)$ et $p(j)$ soit le moins possible affectée par ces regroupements.

On montre que tout regroupement de X_i , ce qui correspond à une agglutination de rubriques de A , a pour effet une diminution du lien entre ensemble des observations et ensemble des variables. Pour Q , une partition de A on a :

$$p(q,j) = \sum_{i \in Q} p(i,j)$$

$$p(q) = \sum_{i \in Q} p(i)$$

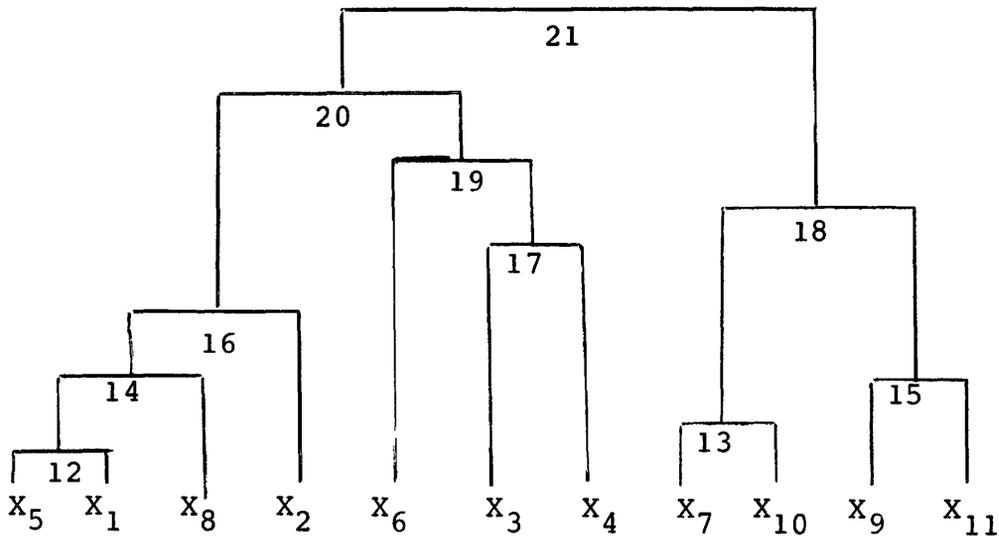
L'information contenue dans $p(q,i)$ est inférieure à celle contenue dans $p(i,j)$. On va donc chercher les partitions Q telles que le passage de A à Q se fasse avec le moins de perte possible d'information, cela revient à chercher la plus petite diminution du lien possible.

Concrètement, on va chercher à partir de la partition Q de A une nouvelle partition Q' obtenue par l'agrégation de deux rubriques seulement de Q , toujours en s'arrangeant pour avoir la diminution minimale du lien. On procède au regroupement des X_i couple de points par couple de points, d'où le nom de méthode binaire. Lorsque deux points sont regroupés, on les remplace par leur barycentre affecté de la somme de leurs masses, et on continue jusqu'à ce qu'il ne reste plus qu'un seul point confondu avec G de masse 1.

Les résultats obtenus par classification automatique sont en général présentés sous la forme d'un arbre ou dendrogramme obtenu directement sur le calculateur. Ci-dessous figure un exemple théorique d'arborescence.

Au niveau de jonction de deux points se trouve le "noeud", pour lequel est calculé le diamètre. Sur le graphique l'ordonnée d'un noeud est égale à son diamètre. Souvent chaque noeud a un numéro d'ordre correspondant à l'ordre croissant des diamètres.

Exemple d'arbre



Plus le diamètre du premier noeud qui correspond à un couple de points est faible (noeud bas), et plus ces points sont proches. Ainsi X_5 et X_1 sont plus proches (diamètre du noeud 12) que X_3 et X_4 (diamètre 17).

3°) - Les limites de la classification automatique ascendante

Pratiquement la classification automatique est d'un usage délicat car se posent de nombreux problèmes. En effet, il s'agit de choisir une distance ou un indice de similarité aussi satisfaisant que possible. Il existe un grand nombre d'algorithmes basés sur des distances différentes. DUCIMETIERE [7] en a dressé une liste critique, sinon exhaustive. Pour résoudre un problème déterminé, principalement en ce qui ^{nous} concerne, quelle

est la meilleure distance ? Notons que si l'on veut rapprocher les résultats de la classification automatique de ceux obtenus avec l'analyse des correspondances on a intérêt à choisir le même type de distance (celle du Chi-Carré).

Par ailleurs, l'analyse arborescente suppose la définition d'un seuil S de ressemblance. En deçà de ce seuil les observations sont agrégées au sein d'une classe, alors qu'au delà on estime qu'il s'agit d'une classe différente. Or, du moins dans les procédés classiques, la définition de ce seuil est plus ou moins arbitraire. Il existe à présent des algorithmes ayant un caractère plus objectif, tels ceux de ROUX [8] et [9] .

Mais le plus grave inconvénient de l'analyse arborescente résulte de son caractère binaire. On construit une classification pas à pas, sans que l'on puisse revenir en arrière pour modifier une branche déjà adoptée. En effet, cette méthode donne un arbre dans lequel chaque agrégation binaire est optimale, mais on n'est pas assuré que l'arbre fournisse des partitions optimales à tous les niveaux (cf. BENZECRI [5]).

Pour remédier à cela on a tenté (REGNER et DE LA VEGA de la Maison des Sciences de l'Homme) de mettre au point une procédure dans laquelle on essaye toutes les affectations possibles pour tous les individus, au lieu de les considérer deux à deux. Malheureusement, ceci suppose, en l'état actuel des choses, des ordinateurs d'une capacité de calcul considérable, dès que les observations sont en nombre important. ROUX [8] et [9] a plus simplement mis au point un algorithme d'échange qui permet d'adopter en cours de construction de l'arbre, un changement d'affectation pour un individu. Bien que moins rigoureux, sur le plan théorique, cet algorithme donne des résultats intéressants.

B - La méthode des nuées dynamiques de Diday [6]

Dans cette méthode, de nature non hiérarchique on part d'un certain nombre de points soit tirés au hasard, soit choisis parmi l'ensemble des individus que l'on cherche à classifier. Ces points constituent l'amorce d'un groupe, les autres individus venant les rejoindre en fonction de leur ressemblance avec l'un ou l'autre. Ainsi apparaissent des noyaux à l'intérieur desquels il y a une forte affinité entre un individu et un groupe d'individus. Ensuite on procède à des rapprochements entre groupes.

Cette méthode a un caractère dynamique car à chaque opération d'adjonction d'un point, ou d'un groupe, il y a possibilité de remise en cause. Ainsi les centres peuvent se déplacer après qu'il y ait eu affectation d'un nouvel individu. Chaque fois qu'un nouveau point est agrégé à une classe, la moyenne de cette classe change, prenant en compte ce nouveau point. Les centres peuvent changer également lorsque tous les individus sont affectés à un groupe.

Les remises en cause des noyaux sont arrêtées au bout d'un certain nombre d'itérations fixé à l'avance, ou lorsque il y a stabilisation du processus. La méthode évite la mise en mémoire d'informations sur les couples d'individus, ce qui réduit considérablement les temps de passage sur ordinateur.

Nous présenterons assez sommairement les bases mathématiques de la méthode des nuées dynamiques (1).

(1) Pour cette présentation nous nous sommes inspirés de MORIN (A.M.) Classification automatique - reconnaissance des formes - nuées dynamiques. Note interne - Station d'Economie Rurale I.N.R.A. Rennes - juin 1973.

Dans l'ensemble fini E d'individus que l'on cherche à classer on définit un sous-ensemble L_k . L_k est inclus dans

$$\{L = (A_1 \dots A_k) / A_i \subset A\}$$

A représente aussi bien E que \mathbb{R}^n ou un autre espace

les A_i sont ce que l'on appelle les noyaux
les éléments composant les A_i sont les étalons

On notera $P_k(E)$ l'ensemble des partitions ordonnées de E en un nombre $n < k$ de parties.

$$D'autre part on a $V_k = L_k \times P_k$$$

Etant donné un critère w , application de $V_k \rightarrow \mathbb{R}^+$ il s'agit de trouver :

- la partition de E qui optimise w
- la partition de E qui optimise u parmi toutes les partitions en k classes
- parmi les partitions en k classes, celles où chaque classe a le noyau le plus représentatif.

Pour arriver à ce résultat on se donne deux fonctions f et g que l'on applique alternativement à partir de $L_0 \in U_k$ estimé, ou pris au hasard.

L'application f associe à chaque noyau A_i la partie P_i constituée de l'ensemble des éléments de E tels que la ressemblance entre P_i et A_i est plus grande que la ressemblance entre P_i et les autres noyaux définis par U. Ceci suppose évidemment que l'on ait choisi un indice de similarité, en se référant par exemple à la distance euclidienne, ou toute autre. Notons que si l'on utilise l'analyse des nuées dynamiques couplée avec l'analyse des correspondances, on a intérêt à choisir la distance du Chi Carré.

$$\text{On a } f : L_k \rightarrow P_k \text{ définie par } f(L) = P$$

L'application $g : \mathbb{P}_k \rightarrow \mathbb{L}_k$ est définie par $g(P) = L = (A_1 \dots A_k)$

Cette fonction a pour effet d'associer à une partition donnée les noyaux les plus proches de cette partition.

Remarquons enfin que l'on utilise une fonction R

$$R : E \times T \times \mathbb{P}_k \rightarrow \mathbb{R}^+ \quad \text{où } T = [1, k] \cap \mathbb{N}$$

Cette application a pour but d'agrégier plus fortement les noyaux en ramenant les étalons vers le centre de leur classe. D'autre part, elle sert à écarter les noyaux entre eux. D'où le nom de fonction d'agrégation - écartement donné à R.

La méthode des nuées dynamiques permet de définir des formes faibles et des formes fortes que l'on visualise au moyen d'une hiérarchie. Celle-ci se manifeste sous forme d'un tableau apparaissant dans les sorties d'ordinateur.

BIBLIOGRAPHIE SOMMAIRE

A) ANALYSE DES CORRESPONDANCES

- [1] BENZECRI (J.P.) Leçons sur l'analyse factorielle et la reconnaissance des formes. Cours 3è cycle IEUP - 2 tomes.
- [2] CORDIER (Brigitte) L'analyse des correspondances - Thèse 3è cycle - Rennes 1965.
- [3] LEBART (L) - FENELON (J.P.) Statistique et informatique appliquées - Dunod 1971.
- [4] LENCO (M) - HAMROUNI Essai d'établissement d'une typologie des exploitations agricoles à partir des méthodes d'analyse factorielle. Ministère de l'Agriculture - S.C.E.E.S. 1970 - 44 p.

B) CLASSIFICATION AUTOMATIQUE

- [5] BENZECRI (J.P.) Problèmes et méthodes de la taxinomie. Revue de Statistique Appliquée 1970. Vol XVIII-n°4, p.73-96.
- [6] DIDAY (Edwin) Nouvelles méthodes et nouveaux concepts en classification automatique et reconnaissance des formes. Thèse d'Etat - Paris VI. Décembre 1972 (Ième thèse).
- [7] DUCIMETIERE (P.) Les méthodes de la classification numérique Revue de Statistique Appliquée, 1970, vol XVIII, n°4, p.5-25.
- [8] ROUX (Maurice) Un algorithme pour construire une hiérarchie particulière - Thèse 3è cycle, Paris 1968.

- [9] ROUX (Maurice) - Deux algorithmes récents en classification
autotique, Revue de Statistique Appliquée,
1970, vol XVIII, n°4, p.35-39.
- [10] VOLLE et Al L'analyse arborescente - Annales de
l'I.N.S.E.E. n°4, mai septembre 1970, p.87-99.
-

ANNEXE 2

RESULTATS DES ANALYSES PRELIMINAIRES

1ère ANALYSE

FRANCE - 21 REGIONS DE PROGRAMME

I - PRESENTATION DES RESULTATS POUR LA FRANCE

Une application de l'analyse des correspondances a été effectuée au niveau des 21 régions programme de la France. Cet essai a été mené à l'aide de deux séries de données. L'une fait appel essentiellement à des éléments exprimés en unités monétaires, tandis que l'autre contient uniquement des données physiques. Ainsi chaque région a été caractérisée à l'aide d'un jeu de données exprimées en valeur, complétées par quelques éléments non monétaires, et à l'aide d'un jeu de données ayant un contenu strictement physique. On a tenté de voir si les résultats obtenus avec les deux analyses se superposaient.

L'intérêt d'une telle démarche tient à la difficulté de disposer au niveau européen et par région de données en valeur ayant un contenu homogène. A supposer même que l'on puisse connaître pour chacune des unités retenues le montant du revenu par catégorie socio-professionnelle, se poserait encore le problème de la comparabilité des chiffres. On sait que le choix d'une unité monétaire commune entraîne des distorsions sensibles. Il apparaît donc souhaitable de savoir si l'on peut élaborer une typologie à l'aide d'éléments à caractère physique que l'on sait être parfaitement comparables d'une région à l'autre, et qui sont faciles à rassembler. Tel est le but de l'analyse entreprise pour la France.

La nécessité d'avoir deux séries relatives à la même période, nous a conduit à centrer l'analyse sur 1962-1963. A cette date, on dispose en effet, de la décontraction des comptes nationaux pour chaque région **et** par activité économique, des résultats du recensement général de la population, et de l'enquête au dixième sur la structure des exploitations agricoles.

1° - Analyse avec des données en termes monétaires (47 variables)

Les catégories de variables retenues sont :

- la répartition du nombre d'exploitations selon la taille (mesurée d'après la superficie agricole utilisée) ; 7 classes ont été retenues (de moins d'1 ha à plus de 100 ha). On a également noté la Superficie Agricole Utilisée Totale de la région.
- la valeur de la production finale agricole.

- La composition de la production finale en distinguant 11 types de productions : céréales, betteraves sucrières et pommes de terre, légumes, fruits, vin, gros bovins, ovins et caprins, veaux, porcs, lait, oeufs et volailles (composition exprimée en pour mille du total).
- La valeur ajoutée globale pour l'agriculture, pour l'industrie (branches 02 à 12) et les services (branches 13 à 19).
- La population active agricole de l'agriculture, de l'industrie et des services.
- La masse des salaires nets versés à tous les salariés.
- Le total des salariés pour l'ensemble des branches.
- Le nombre de personnes-année-travail agricoles (P.a.t.) en distinguant les P.a.t familiales et salariées (Patf) des P.a.t totales.
- Les taux de remplacement des chefs d'exploitation (en pour mille) évalués par BRUN en notant le taux minimum et le taux maximum ; la proportion de terres libérées chaque année par absence de successeur potentiel.
- La valeur des terres, des bâtiments d'exploitation et du cheptel vif. Le montant des amortissements en bâtiments et les amortissements en matériel.
- Les coefficients d'inputs évalués en pour mille de la production finale, tels que les a évalués J. L. BRANGEON : coefficient de travail, terre, consommations intermédiaires, amortissements, intérêts du capital.
- Le nombre moyen d'heures d'utilisation par an et par Tracteur Année Disponible (T.A.D.) ; le nombre total de tracteurs utilisés dans l'année et le nombre total de T.A.D.

On aboutit ainsi à caractériser chaque région à l'aide de 47 variables les unes étant uniquement de nature physique et propres à l'agriculture (structure des exploitations et tracteurs), les autres définissant l'agriculture à partir d'éléments en valeur (la production finale et sa structure, la valeur ajoutée, le résultat net d'exploitation, le capital et les amortissements, les coefficients d'inputs). Restent enfin des caractères économiques généraux, les uns physiques (nombre de salariés, répartition de la population active), les autres en valeur (salaires nets et valeur ajoutée de l'industrie, ainsi que des services).

L'ensemble de ces données traitées, a permis d'obtenir les résultats récapitulés ci-dessous :

	axe 1	axe 2	axe 3	axe 4	axe 5
% de la variance expliquée	68,69	11,31	9,24	3,34	1,90
valeur propre	0,1991	0,0328	0,0268	0,0097	0,0055

Il apparaît que les trois premiers axes permettent d'expliquer l'inertie totale du nuage de points dans la proportion de 90 % environ. On peut donc considérer qu'il suffit de raisonner par rapport aux trois premiers facteurs. La non prise en considération des autres axes n'est pas susceptible d'engendrer une grave erreur dans l'interprétation des résultats. Il faut noter la grande importance jouée par le premier axe avec environ 68,7 % de la variance et une valeur propre proche de 0,20. On peut dès à présent affirmer le rôle prépondérant de ce facteur.

Le graphique n° 2 concernant le plan 1-2 permet de formuler un certain nombre d'observations relatives à l'ensemble des observations, l'ensemble des variables et à l'ensemble observations-variables.

- ensemble des observations

Par rapport au premier axe, on note très déportée vers une extrémité la région parisienne (PAR) alors qu'à l'autre extrémité figure un groupe important de régions qui sont : l'Auvergne (AUV), le Limousin (LIM), Midi-Pyrénées (MPY), Basse-Normandie (BNM), Pays de Loire (LOI) Poitou-Charentes (PCH), Bretagne (BRE). Par rapport au deuxième axe, trois régions sont nettement individualisées : Languedoc (LAN), Provence (PRO) et Alsace (ALS), toutes les autres se situant de la même façon par rapport à ce facteur.

- ensemble des variables

Par rapport au premier axe se trouvent, d'une part, les variables: masse des salaires versés, valeur ajoutée par l'industrie, population active de l'industrie, nombre de salariés, valeur ajoutée des services. D'autre part, apparaissent les variables telles que importance de la production de veaux, valeur du cheptel, exploitations inférieures à 100 ha, population active agricole et Personnes-Années-Travail (PAT). Par rapport au deuxième axe les extrêmes sont d'un côté veaux, gros bovins, lait, valeur du cheptel, porc, exploitations de 50 à 100 ha, céréales et exploitations de 20 à 50 ha, et de l'autre côté exploitations inférieures à 5 ha, légumes, fruits et vin.

- ensemble variables-observations

Sur le premier axe, on constate une association assez nette entre variables économiques caractérisant le niveau d'activité non agricole (valeur ajoutée des branches non agricoles, nombre de salariés, total des salaires) et des régions telles que PAR, NRD, LOR, HNM, RHO. A l'opposé sont associées les régions AUV, LIM, MPY, BNM, LOI, PCH, BRE aux variables concernant essentiellement l'activité agricole (valeur des bâtiments, PAT, valeur ajoutée de l'agriculture, aviculture, porcs...)

La signification du premier facteur apparaît assez clairement c'est un facteur de développement économique. En effet, les régions riches parce qu'industrialisées, PAR, NRD, LOR ... sont proches des variables économiques non agricoles (salariés, masse des salaires, valeur ajoutée et population active non agricole ...), alors que les régions défavorisées sont proches des variables uniquement agricoles.

Pour le deuxième facteur, spécifiquement agricole, il est plus difficile de lui attribuer une signification claire. Il semblerait qu'on puisse le définir comme un axe de différenciation des systèmes de production par spécialisation intensive. En effet, des régions telles que LAN, PRO, ALS sont proches des variables relatives aux exploitations de faible surface (inférieures à 5 ha) fortement spécialisées (vins,

fruits, légumes) et très intensives. A l'inverse, les autres régions sont liées à des variables plus générales caractérisant les exploitations de polyculture-élevage de taille moyenne.

Dans le graphique 3 se trouve la représentation dans le plan 2-3. Cela permet de voir la situation du nuage de points en fonction du 3ème axe et donc sa signification. A une extrémité, se situent les variables propres aux grandes exploitations (50ha et au-dessus) à système spécialisé (plantes sarclées, céréales) mais de nature extensive. Ces variables sont à proximité des régions Champagne (CHP) et Picardie (PIC). A l'autre extrémité, on trouve les régions BRE, AQU, LIM, RHO ... où domine la main d'oeuvre (variables PAT) et les exploitations de taille moyenne (10 à 50 ha). Ce troisième facteur assez voisin du second, dans sa signification, pourrait être considéré comme un facteur de différenciation des systèmes de production par spécialisation extensive. Cette interprétation est corroborée par le fait qu'à proximité de PIC et CHP ne se trouve pas PAR, région dans laquelle existent en proportion notable des cultures intensives (légumes).

2° - Analyse avec des données uniquement physiques

Les données purement physiques relatives à la seule activité agricole sont nombreuses puisqu'au nombre de 36. Elles ont trait à la structure des exploitations, au système de production et à la population. On a retenu :

- nombre d'exploitations par classe de SAU utilisée (on a retenu 8 classes : moins de 1 ha, 1 à 2, 2 à 5, 5 à 10, 10 à 20, 20 à 50, 50 à 100, et plus de 100 ha, au lieu de 7 précédemment, les classes 1 à 2 ha et 2 à 5 ha ayant été fusionnées).
- surface totale des exploitations occupée par l'ensemble des exploitations de chaque classe
- surface agricole utilisée moyenne

- Répartition de la SAUtilisée selon le mode de faire-valoir (en pour 1000 faire-valoir direct, fermage, métayage, direction par un salarié)
- Répartition de la SAUtilisée en pour 1000 selon l'occupation du sol (céréales, fourrages, autres terres labourées, STH, cultures permanentes et maraîchères)
- Nombre de têtes de bétail (brebis, truies, équidés, caprins, vaches adultes)
- Répartition en pour 1000 des chefs d'exploitation par classe d'âge (< 35 ans ; 35-50 ; 50-65 ; > 65 ans)
- Population active agricole

Les variables physiques permettant de caractériser l'activité économique générale sont :

- Population active de l'industrie et des services
- Nombre total de salariés

En plus de ces variables, on a inclus dans l'analyse la surface totale de la région, la population totale et la densité de population au km carré.

Les résultats obtenus avec la deuxième série de données sont les suivants :

	axe 1	axe 2	axe 3	axe 4	axe 5
% de la variance expliquée	69,065	15,157	6,681	2,101	0
valeur propre	0,1663	0,0365	0,0161	0,0051	-

La variance totale exprimée sur les trois premiers facteurs dépasse 90 % ce qui est légèrement supérieur à l'analyse précédente. Le premier axe avec 69 % de l'inertie du nuage de points a une valeur explicative aussi forte que dans le premier cas. Toutefois, sa valeur propre est moins élevée, ce qui signifie que les résultats sur cet axe sont moins contrastés que précédemment.

Le graphique n° 4 relatif à la projection des points dans le plan 1-2 permet de voir que les observations et les variables se répartissent quasiment de la même façon par rapport au premier facteur. Il existe toujours l'opposition entre PAR, NRD, LOR, HNM, RHO et ALS PRO associées aux variables économiques générales (total salariés, population active non agricole, population totale, densité) et les régions à dominante agricole. Ces dernières regroupant presque toutes les autres régions autour du noyau AUV, LIM, PCH, MPY, BNM, avec uniquement à proximité des variables agricoles.

La disposition du nuage de points en fonction du deuxième axe conduit à l'opposition entre d'une part CHP, PIC, grandes exploitations (plus de 50 ha), céréales, CEN, BOU et d'autre part, BRE associée aux exploitations entre 5 et 20 ha, système porcin ainsi que AQU, LOI, BNM ... Ce deuxième facteur qui est celui d'une différenciation des systèmes de production par spécialisation extensive correspond exactement au troisième axe trouvé dans l'analyse avec des données économiques.

L'examen des graphiques 5 et 6 relatifs aux plans 1-3 et 2-3 nous conduit à constater que le troisième axe de cette analyse oppose petites exploitations très spécialisées, LAN et PRO à la grande masse des régions pratiquant la polyculture-élevage. On retrouve donc là notre facteur de différenciation par spécialisation intensive.

A une intervention d'axe près (facteurs 2 et 3) les résultats de l'analyse avec des éléments en valeur concordent avec ceux de l'analyse en termes non monétaires. Ainsi il semble possible de tenter d'élaborer une typologie des régions au niveau européen, en utilisant seulement des données physiques.

3° - Résultats complémentaires et propositions de typologie pour la France

a) Résultats complémentaires

Il a été effectué un assez grand nombre de passages sur machine afin de déterminer si possible le rôle d'un certain nombre de variables, (On a procédé à une vingtaine de traitements). Nous présenterons ici quelques résultats intéressants ou susceptibles de faciliter l'élaboration de la typologie.

En vue de tester la signification du premier facteur, nous avons fait un passage en partant des 47 variables de l'analyse avec des données en valeur, dont on avait ôté les variables économiques générales (valeur ajoutée et population active non agricole, masse des salaires, total des salariés). Le graphique n° 7 présente les résultats obtenus dans le plan 1-2. Ceux-ci sont à rapprocher des résultats donnés par l'analyse menée avec la série de données non monétaires dont on avait supprimé les variables caractérisant l'activité générale (graphique n°8).

Les structures observées dans les graphiques 7 et 8 sont très semblables, mais la nature du premier facteur est bien différente. Si PAR se retrouve bien isolée, les régions qui lui sont proches sont PIC et CHP accompagnées des variables propres aux exploitations de grande surface à système céréalier. A l'opposé sont situées les régions RHO, AQU et BRE avec les surfaces moyennes. Il est assez frappant de constater que NRD voisine avec PCH, LOI et BNM alors que dans les analyses comportant les variables économiques générales, cette région était près de PAR. Il en va de même pour RHO. Ainsi, il est aisé de voir a contrario que les variables économiques générales définissent bien un facteur de développement économique.

Il nous a paru intéressant de voir si les résultats obtenus avec 41 variables non monétaires pouvaient être retrouvés avec un nombre moins important de données. En effet, l'étude approfondie des graphes indique une proximité constante pour de nombreuses variables.

Ceci est particulièrement sensible pour les variables concernant le nombre d'exploitations par classe de surface, et la surface occupée par les exploitations de cette classe (cas de n_1 et s_1 , n_2 et s_2 , n_3 et s_3 ... n_8 et s_8). De même les variables population active de l'industrie, population active des services et nombre de salariés sont toujours groupés. En vertu du principe d'équivalence distributionnelle, on pourrait regrouper les données proches et les remplacer par leur somme. Toutefois, cela n'est pensable que pour les données agrégeables.

Dans le graphique n° 9, nous avons reporté l'ensemble des variables physiques, hormis les données n_1 à n_8 (nombre d'exploitations par classes de surface) et la population active de l'industrie et la population active des services. Notons que le principe d'équivalence distributionnelle aurait permis de réduire à 2 le nombre de classes d'âge des chefs d'exploitation puisque seule la classe moins de 35 ans est isolée. De même pour la classe s_2 et s_3 (surface des exploitations de 1 à 2 ha et de 2 à 5 ha). Ceci donne un nuage de 29 variables qui pourrait être réduit à 25. Or, la structure de ce nuage est identique à celle obtenue avec la série complète de données physiques. Ceci montre qu'il est possible après testage des données, de travailler avec un nombre relativement restreint de variables, ce qui facilite les passages sur machine et l'interprétation des résultats.

b) Propositions de typologie des régions françaises

Selon les facteurs que l'on souhaite privilégier, et le degré de finesse que l'on désire, on peut proposer trois typologies, la première basée sur les facteurs développement économique et différenciation intensive, la deuxième à partir du développement et de la différenciation extensive, et la troisième en combinant les 3 facteurs.

- Typologie basée sur le développement économique et la différenciation intensive . Assez à part se trouve PAR, région très particulière caractérisée par une forte densité et un grand nombre de salariés.

Puis on a des régions industrialisées NRD, LOR, HNM, RHO, ALS. PRO qui est également industrialisée se détache en raison de son agriculture comportant une forte proportion de cultures maraîchères et légumières. Dans une certaine mesure, ALS se trouve dans cette situation.

Dans une zone intermédiaire figurent 3 régions : PIC, FRC et CHP. PIC et CHP sont des régions où l'agriculture bénéficie de structures particulièrement favorables, alors que l'on a vu sur les graphiques 7 et 8 (les deux séries données traitées sans les variables économiques générales) que FRC, tout comme NRD se trouvait dans le peloton des régions défavorisées.

La quatrième catégorie regroupe BOU, CEN, et AQU, régions difficiles à caractériser ; BOU et CEN étant très semblables et ayant des structures agricoles plus favorables qu'AQU ainsi que le montrent les graphiques 7 et 8.

Enfin restent les régions faiblement industrialisées et où l'agriculture ne bénéficie pas de conditions très favorables. Il s'agit de LOI, BNM, AUV, PCH, LIM, MPY et BRE.

Le Languedoc, LAN, est bien à part, région peu industrialisée mais à système très intensif avec des productions spéciales.

- Typologie basée sur le développement économique et la différenciation extensive.

Par rapport à ces deux axes, la région parisienne PAR se détache toujours nettement.

Ensuite les régions industrialisées NRD, LOR, HNM, RHO, ALS comprennent également PRO qui est englobée dans ce groupe. Deux petits groupes se manifestent avec d'une part CHP, PIC et d'autre part CEN, BOU. Dans cette zone se situe également FRC qui est isolée. Il s'agit de régions situées dans une sorte de zone tampon entre les régions industrialisées et les régions pauvres.

Avec les données en valeur, LAN apparaît nettement isolé des groupes précédents et du groupe suivant, alors qu'avec les variables non monétaires LAN rejoint l'ensemble régions peu industrialisées de polyculture-élevage.

Cet ensemble polyculture élevage comprend outre LAN selon le type de données, PCH, BNM, LOI, AQU, AUV, LIM, MPY.

La Bretagne BRE se détache de ce groupe.

En définitive, l'utilisation des facteurs développement et différenciation intensive permet une typologie plus satisfaisante, du moins sur le plan de l'esprit, que lorsqu'on se base sur l'autre facteur de différenciation extensive.

- Typologie basée sur les 3 facteurs

En combinant les deux analyses précédentes, on aboutit à proposer :

PAR région caractérisée par l'activité tertiaire

NRD, LOR, HNM, RHO, ALS régions où domine l'activité industrielle avec une agriculture moyennement ou faiblement favorisée mais sans particularités.

PRO dont le degré de développement est semblable à celui du groupe précédent mais où l'agriculture est orientée vers des cultures spéciales effectuées sur de faibles surfaces.

CHP, PIC régions où l'agriculture bénéficie de conditions très favorables grâce à des structures peu contraignantes qui autorisent la grande culture.

CEN, BOU régions du même type que CHP et PIC mais un degré en dessous (surfaces moins importantes).

FRC région assez proche de CEN et BOU sur le plan agricole mais plus industrialisée.

LAN faiblement industrialisé et qui a un système de production très marqué par les petites exploitations et les cultures spéciales (vigne, fruits).

AQU mieux placée que LAN quant au développement économique avec quelques particularités agricoles qui le mettent sur le même plan de ce point de vue que BRE et RHO.

LOI, BNM, AUV, PCH, LIM, MPY régions sans industrie et dont le système de production est traditionnel. Ce groupe assez homogène pourrait certainement être scindé si on le désirait.

BRE occupe une place particulière qui est presque symétrique de celle de PAR : région la plus mal placée dans une analyse où n'entrent que les variables agricoles (PAR est en tête avec PIC et CHP), région la moins bien située par rapport au facteur de développement économique.

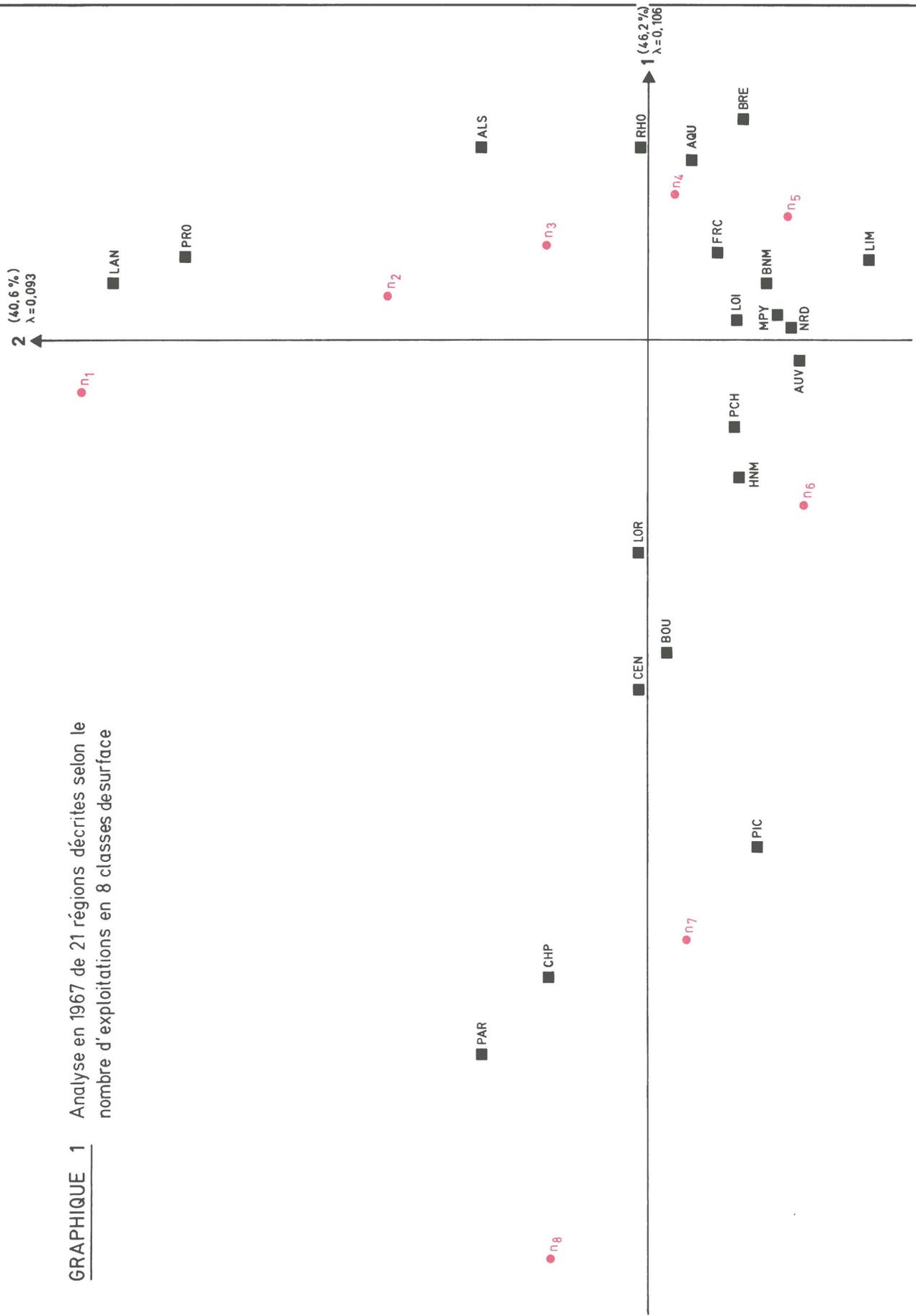
On aboutit ainsi à classer les 21 régions de programme en 10 groupes ce qui est trop pour avoir un outil opérationnel. Mais il est possible de restreindre le nombre de classes en proposant, région parisienne mise à part :

1er groupe : régions riches en raison de leur développement industriel : NRD, LOR, HNM, RHO, ALS, PRO.

2e groupe : régions riches en raison de leur agriculture, ou sans problème agricole : CHP, PIC, CEN, BOU, FRC.

3e groupe : régions pauvres tant sur le plan industriel que sur le plan agricole : PCH, BNM, LOI, AQU, AUV, LIM, MPY, BRE, LAN.

GRAPHIQUE 1 Analyse en 1967 de 21 régions décrites selon le nombre d'exploitations en 8 classes de surface

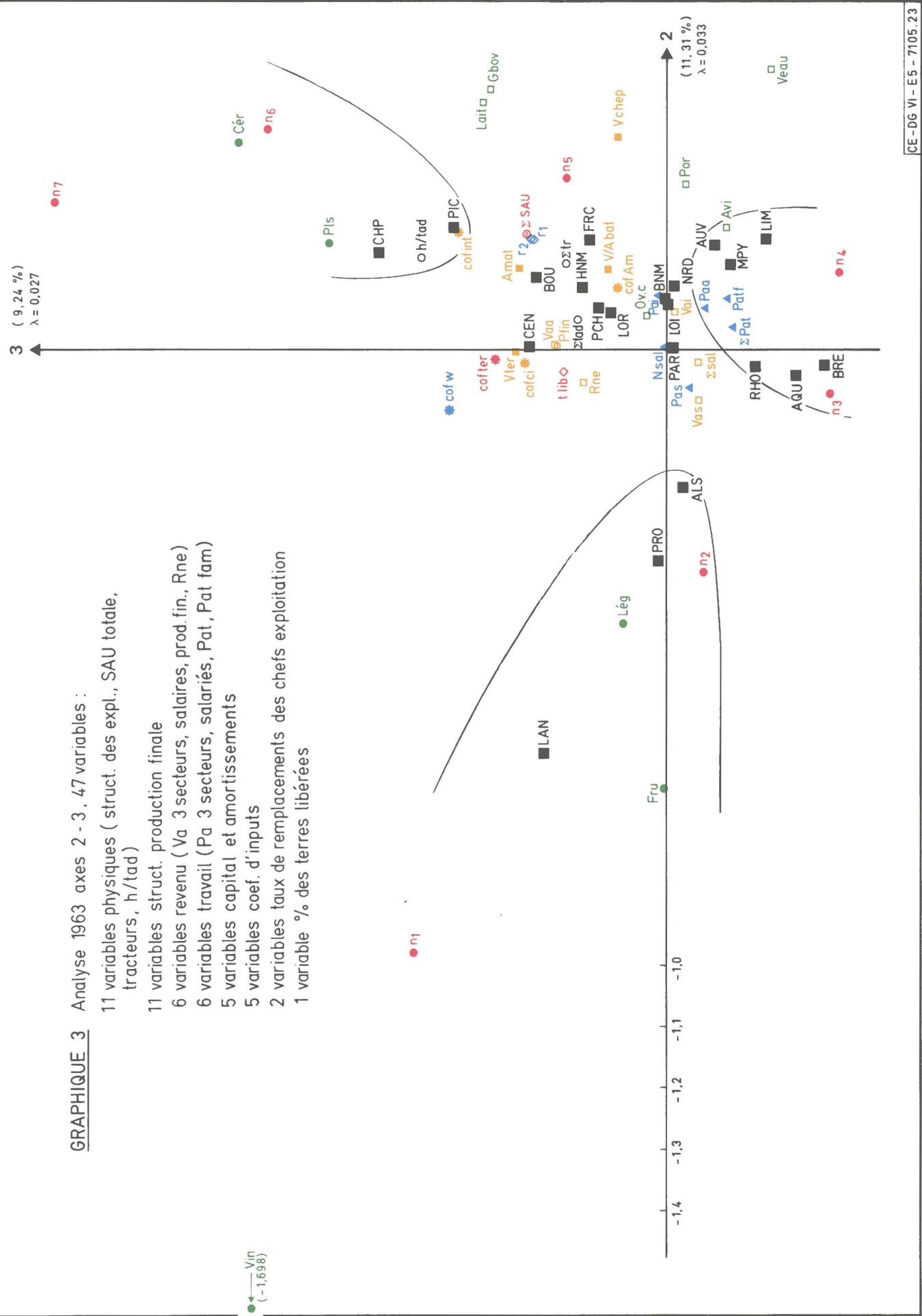


GRAPHIQUE 3

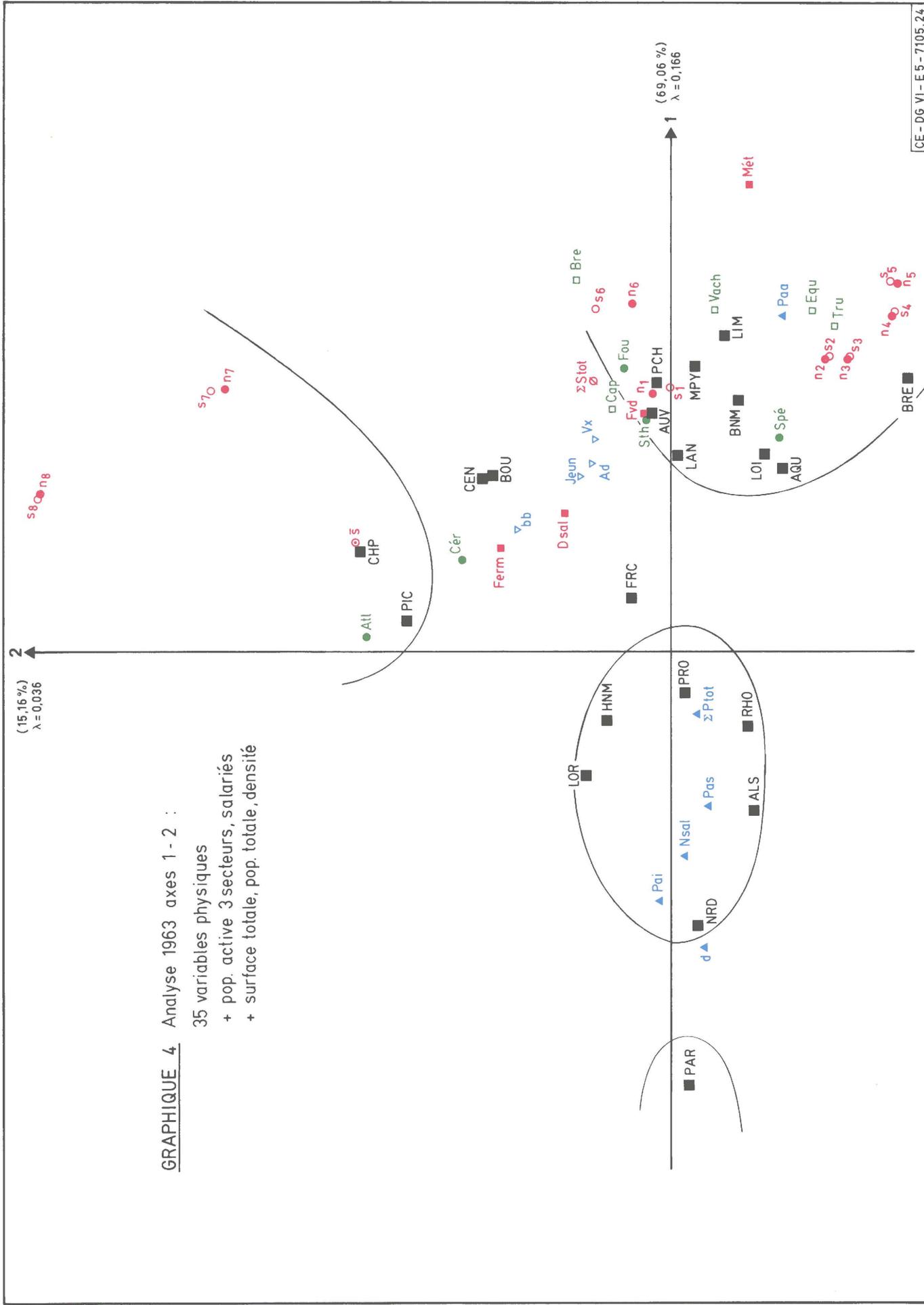
Analyse 1963 axes 2 - 3, 47 variables :

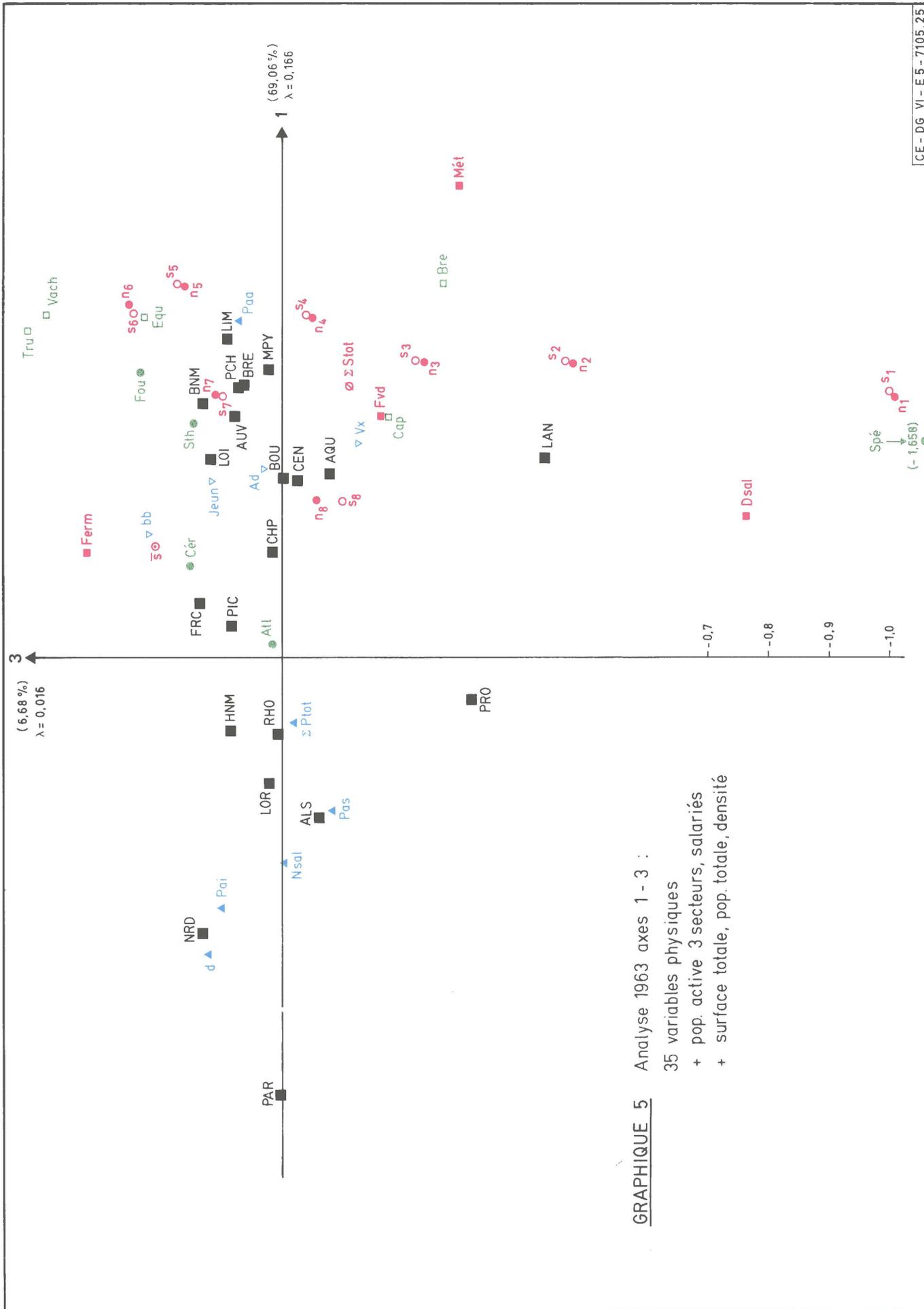
- 11 variables physiques (struct. des expl., SAU totale, tracteurs, h/tad)
- 11 variables struct. production finale
- 6 variables revenu (Va 3 secteurs, salaires, prod. fin., Rne)
- 6 variables travail (Pa 3 secteurs, salariés, Pat, Pat fam)
- 5 variables capital et amortissements
- 5 variables coef. d'inputs
- 2 variables taux de remplacements des chefs exploitation
- 1 variable % des terres libérées

← Vin
(-1,698)

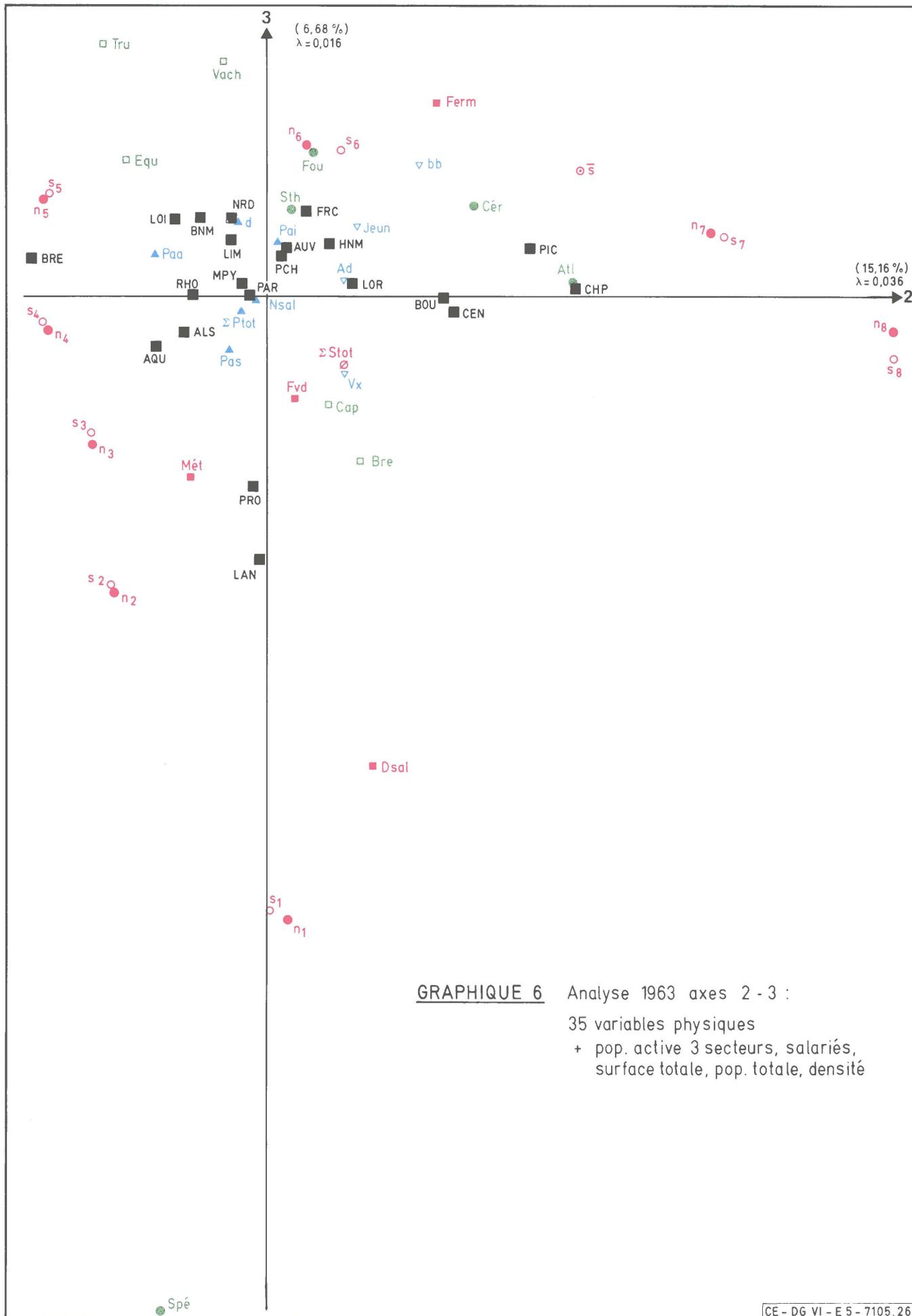


GRAPHIQUE 4 Analyse 1963 axes 1 - 2 :
 35 variables physiques
 + pop. active 3 secteurs, salariés
 + surface totale, pop. totale, densité

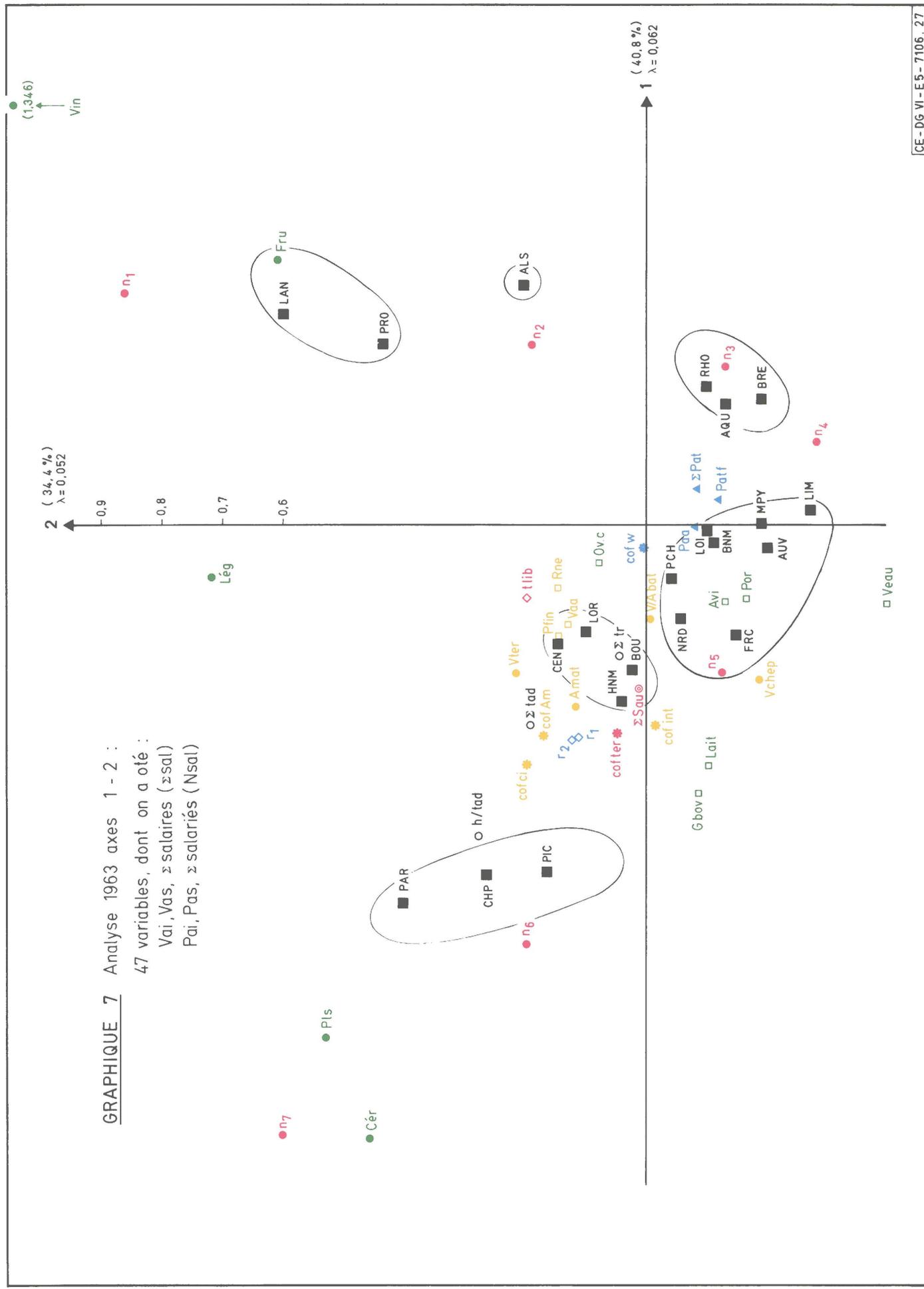




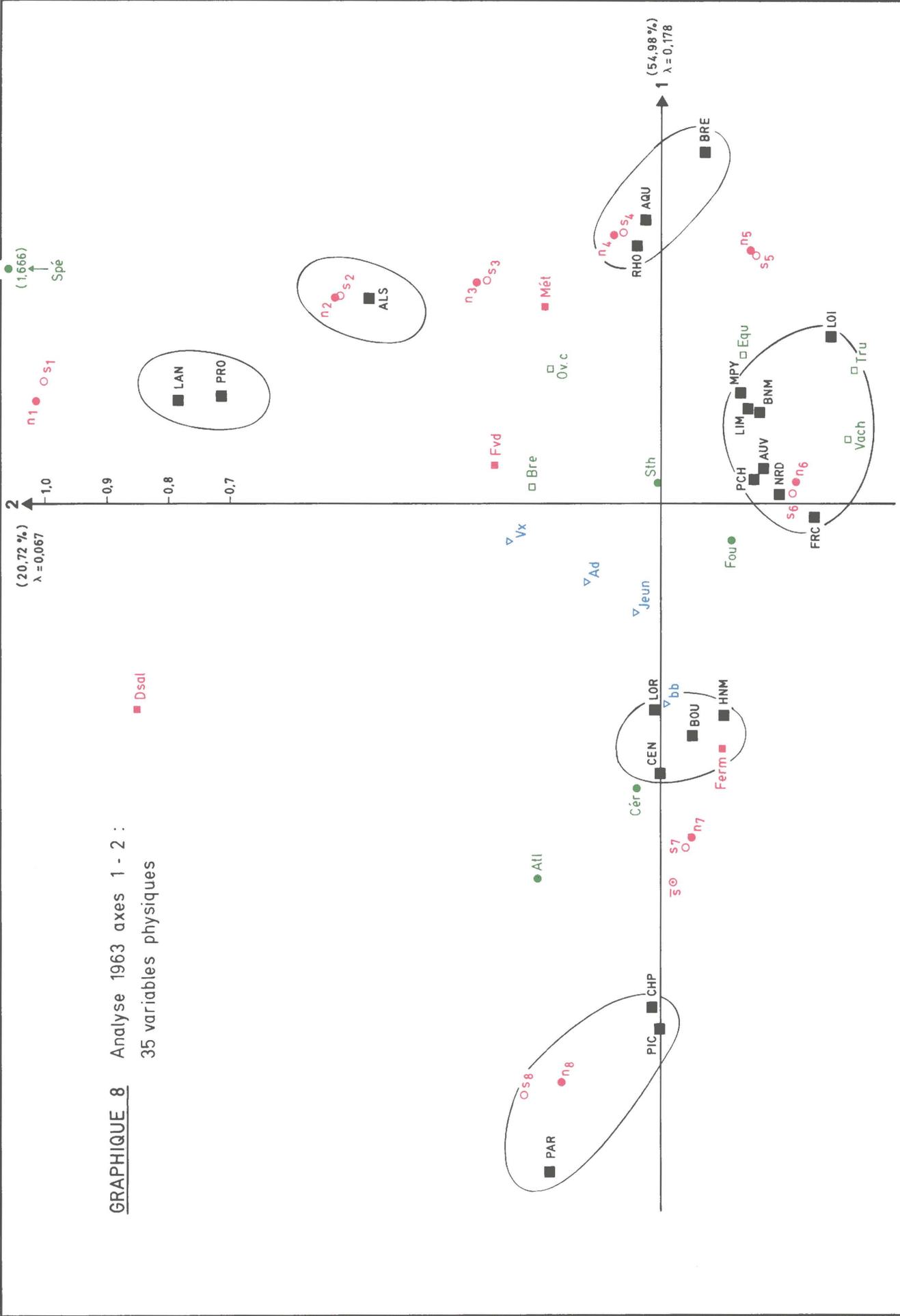
GRAPHIQUE 5 Analyse 1963 axes 1 - 3 :
 35 variables physiques
 + pop. active 3 secteurs, salariés
 + surface totale, pop. totale, densité



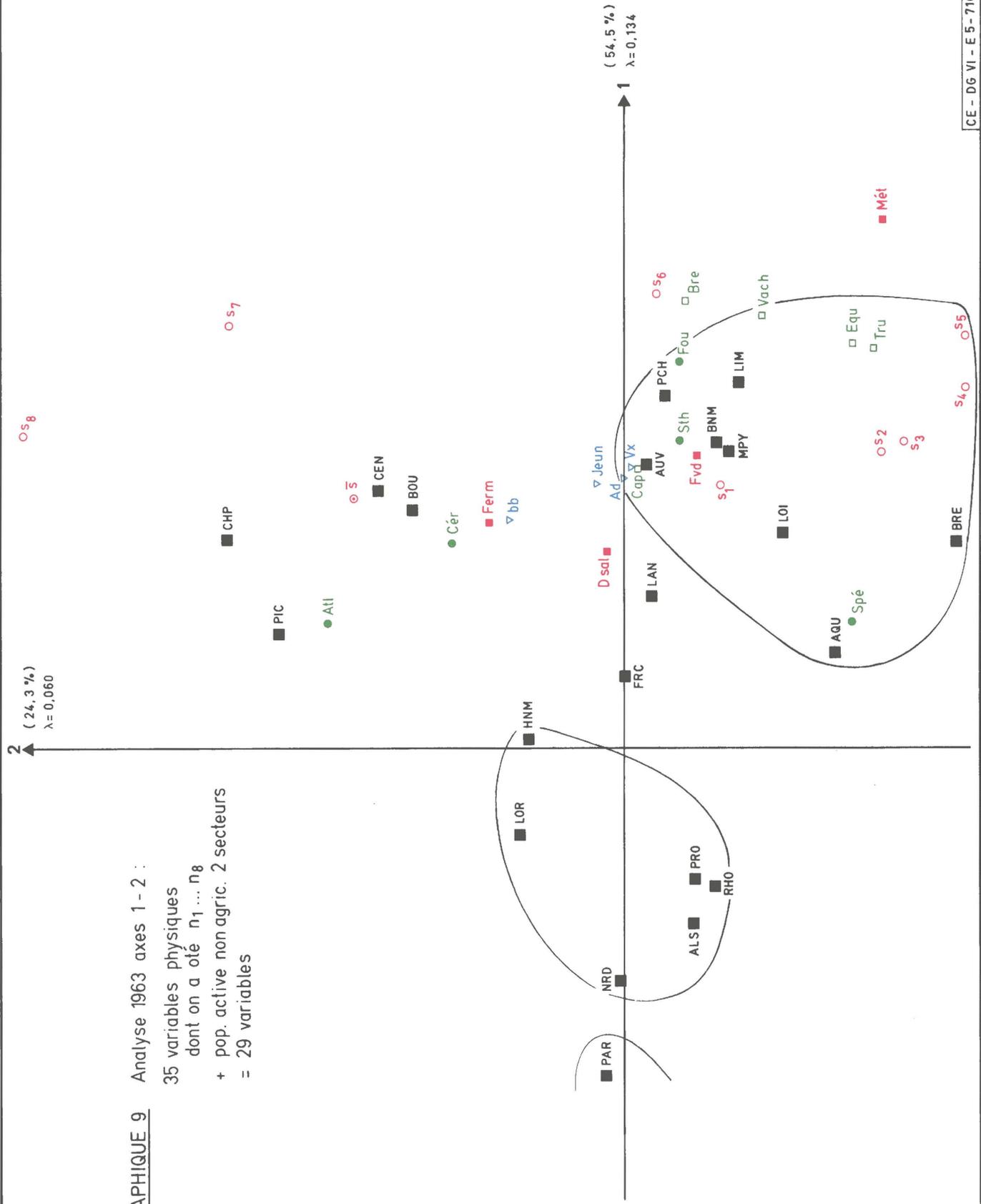
GRAPHIQUE 7 Analyse 1963 axes 1 - 2 :
 47 variables, dont on a oté :
 Vai, Vas, z salaires (zsal)
 Pai, Pas, z salariés (Nsal)



GRAPHIQUE 8 Analyse 1963 axes 1 - 2 :
35 variables physiques



GRAPHIQUE 9 Analyse 1963 axes 1 - 2 :
 35 variables physiques
 dont on a été $n_1 \dots n_8$
 + pop. active non agric. 2 secteurs
 = 29 variables



SYMBLES ET ABREVIATIONS EMPLOYES DANS LES GRAPHIQUES

■	REGIONS :
ALS	: Alsace
AQU	: Aquitaine
AUV	: Auvergne
BNM	: Basse-Normandie
BOU	: Bourgogne
BRE	: Bretagne
CEN	: Centre
CHP	: Champagne
FRC	: Franche-Comté
HNM	: Haute-Normandie
LAN	: Languedoc
LIM	: Limousin
LOI	: Pays de la Loire
LOR	: Lorraine
MPY	: Midi-Pyrénées
NRD	: Nord
PAR	: Région Parisienne
PCH	: Poitou-Charente
PIC	: Picardie
PRD	: Provence-Côte d'Azur-Corse
RHO	: Rhône-Alpes

Surface et nombre d'exploitations

Structure des exploitations

- n_1 à n_8 Répartition du nombre d'exploit. par classe de SAU utilisée
- Graphiques 1, 2, 3 :
 - n_1 : < 1ha; n_2 : 1 à 5ha; n_3 : 5 à 10ha; n_4 : 10 à 20ha;
 - n_5 : 20 à 50ha; n_6 : 50 à 100ha; n_7 : > 100ha.

Graphiques 4 à 9 :

- s_1 à s_8 Surface totale des exploitations occupée par l'ensemble des exploitations de chaque classe (mêmes classes que pour la répartition du nombre d'exploitations).
- \bar{s} Surface agricole utilisée moyenne
- ΣSAU Surface agricole utilisée totale de la région
- $\Sigma Stot$ Surface totale

Répartition de la SAU utilisée selon le mode de faire-valoir

- F_{vd} : Faire-valoir direct
- F_{ferm} : Ferme
- F_{sal} : direction par un salarié
- ◇ Terres libérées
- ◇ t_{lib} : proportion de terres libérées chaque année par absence de successeur potentiel.

Composition de la production finale (exprimée en pour mille du total) ou répartition de la SAU utilisée (en pour mille selon l'occupation du sol)

- Production végétale
 - Cér Céréales
 - Pls plantes sarclées
 - Lég légumes
 - Fru fruits
 - Vin viticulture
 - Fou cultures fourragères
 - Atl autres terres labourables
 - Sth surfaces toujours en herbe
 - Spé cultures spéciales
- Production animale
 - Gbov gros bovins
 - Ov.c ovins, caprins
 - Lait lait
 - Bre brebis
 - Équ équidés
 - Vach vaches adultes
 - Veau veaux
 - Por porcs
 - Avi aviculture
 - Iru truies
 - Cap caprins

Population et main d'œuvre

- ▽ Répartition des chefs d'exploitation par classe d'âge (en pour mille)
 - bb < 35 ans
 - Jeun 35 à 50 ans
 - Vx > 65 ans
- ◇ Remplacement des chefs d'exploitations (en pour mille)
 - r_1 taux minimum
 - r_2 taux maximum

Population main d'œuvre

- Paa population active agricole
- Pai population active industrie
- Pas population active services
- Nsal total des salariés pour l'ensemble des branches
- Patf nombre de personnes-année-travail familiales agricoles
- ΣPat nombre de personnes-année-travail agricoles totales (familiales et sal.)
- $\Sigma Ptot$ population totale
- \bar{d} densité de la population au km²

Coefficients d'inputs évalués en pour mille de la production finale

- Cof w travail
- Cof ci consommation intermédiaire
- Cof Am amortissements
- Cof int intérêts
- Cof ter terres
- Traction
- h/tad nombre moyen d'heures d'utilisation par an et par Tracteur-Année-Disponible
- Σtr nombre total des tracteurs utilisés dans l'année
- Σtao nombre total de Tracteurs-Année-Disponible

Variables monétaires

- Capital et amortissements
- Vter valeur terre
- Vbât valeur bâtiments
- Vchep valeur cheptel vif
- Amat amortissements matériel
- Abât amortissements bâtiments
- Revenu
- Pfin valeur de la production finale agricole
- Vaa valeur ajoutée globale de l'agriculture
- Vai valeur ajoutée globale de l'industrie
- Vas valeur ajoutée globale des services
- Σsal masse des salaires nets versés à tous les salariés
- Rne résultat net d'exploitation

2ème ANALYSE

COMMUNAUTE EUROPEENNE (SIX ANCIENS ETATS MEMBRES) - 94 REGIONS

II - ANALYSE AU NIVEAU COMMUNAUTAIRE (Six anciens Etats membres)

On précisera successivement le champ des variables et des observations, les résultats obtenus avec l'ensemble des variables, puis avec un nombre plus restreint. On proposera ensuite une classification des régions, et enfin on indiquera dans quelle mesure, il est possible d'aboutir aux mêmes résultats avec moins de données.

A - LES VARIABLES RETENUES ET LES OBSERVATIONS

Pour caractériser de façon aussi complète que possible, l'activité économique tant agricole que générale de chacune des régions, nous avons retenu au départ un jeu de 64 variables. Dans le choix de celles-ci, nous avons dû tenir compte des statistiques disponibles au niveau communautaire, étant entendu que l'on doit avoir des informations dont le contenu est homogène pour l'ensemble des régions. Inversement, les statistiques disponibles sont une contrainte dans le choix de la circonscription géographique. En effet, plus on descend bas dans le découpage des pays, et moins on a de renseignements détaillés et homogènes, soit que les résultats aient été obtenus par sondage, soit, s'il s'agit de recensements généraux, que les résultats ne soient pas publiés. On peut dire approximativement que le détail des résultats est inversement proportionnel à la finesse du découpage des régions.

1 - Les observations (les régions de la C.E.E.)

Les premiers essais furent effectués en utilisant pour la R.F. d'Allemagne le découpage en 11 Länder, puis devant le caractère insuffisamment précis de cette division, on eut recours au nouveau découpage en 35 Regierungsbezirke. Pour les autres pays, on se référa aux 21 régions de programme françaises, aux 20 regioni italiennes, 9 provinces belges, 11 provinces néerlandaises et le Grand Duché de Luxembourg. Il apparut que les "villes-état" : HAMBURG, BREMEN et BERLIN avaient des caractéristiques très particulières, puisque la place qu'y tient l'agriculture est très réduite. Notons que pour la région

parisienne, le problème est un peu différent car si la place de l'agriculture est faible par rapport à l'activité de cette région, elle n'est pas négligeable par rapport à l'agriculture française et européenne. Il a donc été décidé d'exclure HAMBURG, BREMEN et BERLIN de l'analyse. Ainsi, le nombre des observations s'élève à 94 (voir liste annexe 1).

En ce qui concerne le niveau de la circonscription géographique, il convient de souligner les très grandes disparités de taille entre régions italiennes ou régions françaises et les autres régions. Ainsi la région de programme Midi-Pyrénées a une superficie totale de 43 700 km² alors que les Pays-Bas divisés en 11 provinces ont une surface de 33 620 km² et la Belgique divisée en 9 provinces de 30 500 km² ! Pour avoir un découpage à peu près identique par pays, il conviendrait donc de travailler à un niveau beaucoup plus fin pour l'Italie et pour la France. Pour ce dernier pays, il faudrait utiliser le cadre du département. Il importe de voir que si la typologie reposait sur des critères avantageant l'activité non agricole, le découpage utilisé serait moins discutable car au point de vue population ou revenu, il y a une certaine équivalence entre Regierungsbezirke et regioni par exemple.

2 - Les variables

On peut distinguer, par commodité, des variables donnant une idée de l'activité économique générale, et des variables proprement agricoles, ayant un caractère physique (voir liste complète des variables ; annexe 2).

Parmi les variables générales, se trouvent essentiellement des données d'ordre démographique : population totale, surface totale et densité et population active. Cette dernière a été appréhendée par secteur : population active agricole, industrielle et tertiaire. Les chômeurs ont été également notés. A ces variables, on a joint les données monétaires élaborées par la Direction Générale de la politique régionale de la Commission des Communautés Européennes, à partir d'une décontraction des agrégats globaux estimés en U.C. Elles concernent le Produit Intérieur Brut (global, agricole et non-agricole), ainsi que la PIB par tête et la productivité du travail (globale, agricole et non agricole).

Pour l'agriculture, on a fait appel essentiellement aux résultats de l'enquête communautaire sur la structure des exploitations. Cette enquête ayant eu lieu en 1966 ou 1967 selon les pays, on a essayé de faire correspondre les autres données à la date de l'enquête communautaire.

Grossièrement, on peut classer en deux types les variables agricoles retenues. Les unes relatives aux personnes occupées en agriculture (chef d'exploitation et sa famille) ; les autres sont relatives au système de production et au mode de faire valoir.

a) Parmi les variables caractérisant les actifs agricoles se trouvent :

- la structure par classe d'âge des chefs d'exploitations (moins de 35 ans, 35-45, 45-50, 50-55, 55-65 et plus de 65 ans),
- la présence ou non d'un successeur potentiel,
- le nombre d'exploitations et la surface qu'elles occupent selon que le chef a une activité à l'extérieur nulle, de plus de la moitié de son temps de travail, de la moitié ou moins de son temps de travail,
- l'importance de la main d'oeuvre familiale régulièrement occupée sur l'exploitation et de celle qui n'est pas régulièrement occupée, ainsi que le nombre de personnes, parmi la main d'oeuvre familiale régulièrement occupée, dont le temps de travail est supérieur à 2 400 heures par an.

b) Parmi les variables caractérisant le système de production agricole se trouvent :

- le nombre d'exploitations par classes de surface (exploitations sans sol, de moins de 1 ha, 1 à 2, 2 à 5, 5 à 10, 10 à 20, 20 à 50, 50 à 100 et plus de 100 ha),
- la répartition de la surface agricole utilisée en cultures permanentes, céréales, pommes de terre, betteraves sucrières, surfaces toujours en herbe, plantes industrielles, fleurs et plantes ornementales, cultures sous verre, cultures de fourrages verts des terres arables. A ces catégories, ont été jointes les surfaces boisées faisant partie des exploitations agricoles
- l'importance du bétail appréciée à l'aide du nombre total d'unités de bétail ainsi que le nombre d'unités de bétail d'équins, bovins, ovins-caprins, porcins et volailles,
- l'équipement exprimé à l'aide du nombre d'exploitations n'utilisant pas de moyens de traction, utilisant de la traction exclusivement mécanique, utilisant de la traction exclusivement animale et du nombre de tracteurs appartenant à l'exploitation.

c) Par ailleurs, on a tenu compte du mode de faire-valoir grâce à la surface en faire-valoir direct, en fermage et en métayage.

Ainsi, les variables sélectionnées donnent une vue sur la situation générale de chaque région (produit intérieur brut, population active, population totale). Quant à l'agriculture, un ensemble de critères fournit une description assez complète tant du système de production que des relations avec le milieu économique (importance de l'activité extérieure). D'autres éléments donnent enfin une vue dynamique de la situation (présence ou non de successeurs potentiels, âge des chefs d'exploitation). En l'état actuel des données existantes, il est difficile d'aller plus loin.

Notons que le nombre de paramètres définissant l'agriculture d'une façon technique, est beaucoup plus important que ceux caractérisant l'environnement économique général (49 dans le 1er cas contre 15 dans le second).

On a fait figurer à l'annexe 2 les codes et sigles utilisés dans les graphiques pour caractériser les 94 régions et les 64 variables.

B - RESULTATS DE L'ANALYSE AVEC 64 VARIABLES

Dans la première analyse comportant 64 variables, 10 facteurs ont été extraits (1), mais pratiquement seuls les quatre premiers ont été utilisés. A titre indicatif, nous fournissons les pourcentages d'inertie et valeurs propres des six premiers.

	1	2	3	4	5	6
% d'inertie	35,12	13,80	9,76	8,26	6,90	5,38
valeur propre	0,1619	0,0636	0,045	0,038	0,0318	0,0248

L'inertie cumulée sur les quatre premiers axes représente plus des deux-tiers de l'inertie totale (66,94 %). Avec cinq axes, on a près des trois-quarts. Les pourcentages "d'explication" sont moins élevés que pour l'analyse relative à la France, mais a priori cela est normal car on est passé d'un tableau 21 x 47 à un tableau 94 x 64.

(1) Pour la France ayant un tableau de taille moyenne (21 x 47), il avait été possible, en utilisant un programme mis au point par M. KERBAOL, de traiter le tableau sur une machine petite (IBM, 1130). Devant l'importance du tableau des données européennes, il a fallu utiliser une machine de plus grosse capacité (C.I.I. 10 070 à 64 K). Le traitement a pu être effectué à l'aide d'un programme mis au point au Laboratoire de l'Institut de Statistiques de l'Université de Paris par l'équipe du professeur J. P. BENZECRI.

L'étude du **nuage** de points est menée en cherchant la signification des facteurs. Ceux-ci sont définis mathématiquement à partir des variables et des observations qui leur sont liées de façon plus ou moins étroite (1). Dans chaque cas, nous raisonnons par rapport aux plans obtenus respectivement avec les axes 1, 2, 3 et 4. Les points qui sont les plus significatifs sont énumérés, mais leur liste n'est pas limitative. C'est l'association de ces points (points-régions et points-paramètres), ainsi que les oppositions entre groupes de points, qui permet de saisir la signification économique d'un facteur. On peut être amené pour des raisons de commodité à désigner un axe d'une façon simple, mais on aura présent à l'esprit qu'il s'agit uniquement d'une simplification.

Rappelons que la méthode utilisée (distance du chi carré et non distance euclidienne) permet de raisonner en terme de profil. Si une région a un profil semblable au profil moyen de l'ensemble des régions pour 63 des 64 variables, elle apparaîtra caractérisée par la 64ème variable, à condition évidemment que celle-ci ait une influence. Cette région n'est donc pas identifiée au moyen de ses paramètres les plus évidents, mais à partir de celui qui s'écarte de façon sensible de la norme. Ainsi, une région où dominerait l'agriculture familiale avec système de polyculture élevage classique, mais où la classe d'âge des exploitants de 35 à 45 ans représenterait une proportion nettement plus élevée^{que} dans les autres régions, serait identifiée non pas à l'aide de son système de production, mais par la caractéristique : nombre de chefs d'exploitation entre 35 et 45 ans.

Notons enfin qu'un point très lié à un facteur ne saurait avoir une bonne représentation sur un quelconque des autres axes. Si l'on désire connaître la position de ce point par rapport aux autres axes, il importe d'éliminer l'influence du facteur primordial.

(1) Pour la définition des facteurs, on s'est basé sur les contributions relatives des variables et des observations aux facteurs. Celles-ci, qui peuvent s'analyser comme un coefficient de corrélation, permettent de voir si un point est bien représenté par rapport aux axes factoriels étudiés. cf. CAZES (Pierre) : Application de l'analyse des données au traitement de problèmes géologiques - Thèse de 3ème cycle, ISUP, 1970, p. 32-33

1er facteur

Si l'on considère le plan 1-2 (voir graphique 1), celui-ci définit 4 quadrants numérotés I, II, III, IV. Ceux-ci permettent de situer les points. Le premier axe^{sg} définit par une opposition entre des régions telles que CAL (II), LAZ (I), SIC (I), PUG (I), CAM (I), ABR (II), LAN (II), etc... et SCHL (sur l'axe), STA (III), OVER (IV), SCHW (sur l'axe), NRD (IV), BRE (IV), HNM (III) etc....

En ce qui concerne les variables, les premières régions sont associées aux petites exploitations (moins de 10 ha avec prédominance des exploitations de 1 à 5 ha), aux cultures permanentes (fruits, vignes). Se trouvent également les exploitations dans lesquelles la main d'oeuvre familiale n'est pas occupée régulièrement, et les exploitations avec traction uniquement animale, ou sans traction. Les variables nombre d'exploitants à temps complet ainsi que le nombre d'exploitants dont l'activité extérieure est supérieure à la moitié du temps et la surface qu'ils occupent, sont associées à ces régions. Enfin, on trouve là les exploitants de plus de 35 ans et ceux qui sont sans successeur potentiel. Les autres régions SCHL, STA sont proches des variables total des unités de bétail, unités de bétail bovin, et dans une moindre mesure, unités de bétail porcin.

On peut dire que le premier facteur différencie ;

- d'une part ce que l'on pourrait qualifier d'agriculture méridionale où il existe une forte proportion d'actifs agricoles qui n'arrivent pas à s'employer à temps plein ni dans une agriculture très spécialisée, ni dans les autres activités,
- d'autre part, ce qu'il est possible de nommer agriculture septentrionale où domine le système de polyculture-élevage à base de bovins.

Cet axe a une forte valeur discriminante, ce qui se marque par une valeur propre de 0,16 relativement élevée.

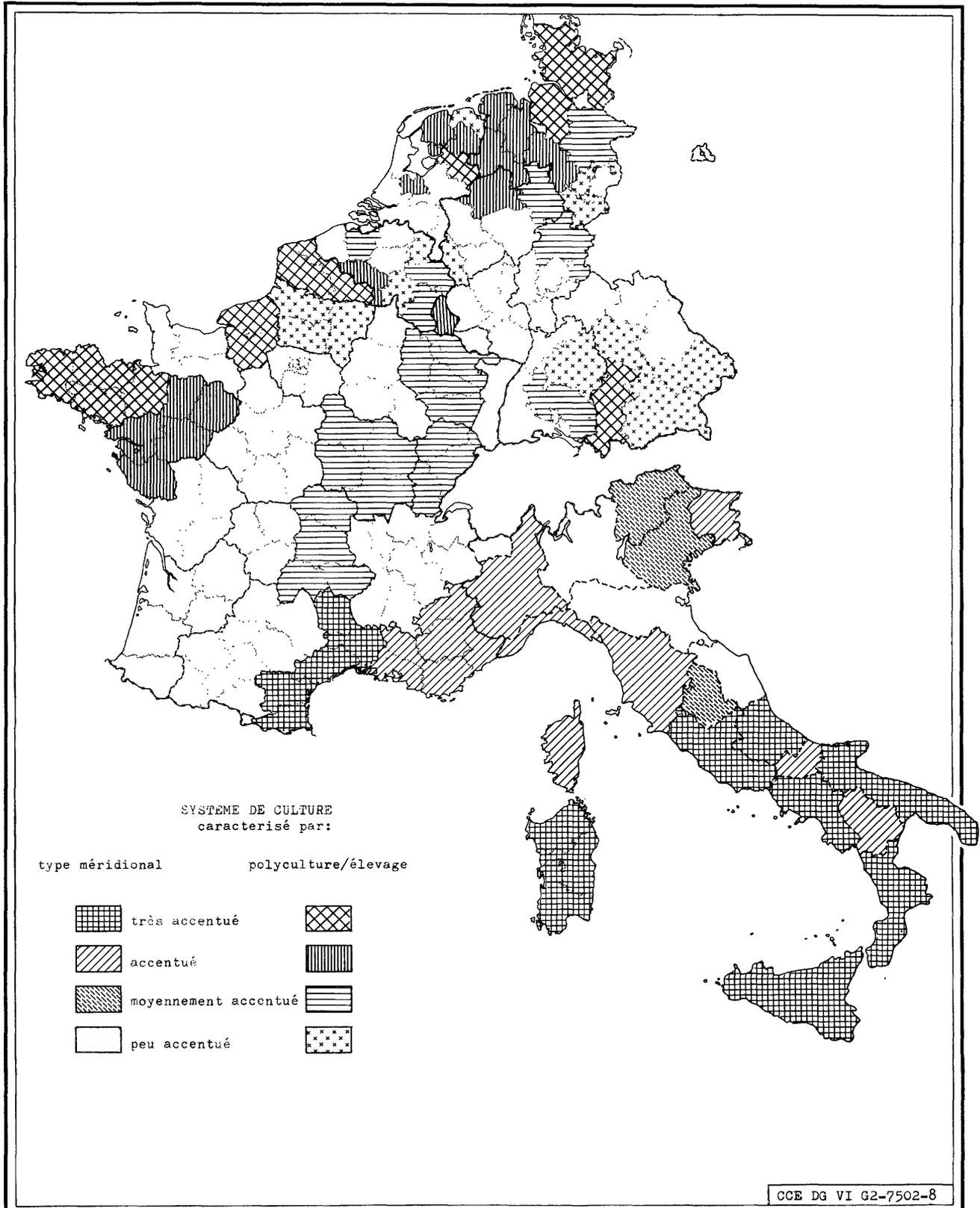
2ème facteur

Il permet de dissocier :

- le groupe d'observations, NBAD (I), NWU (IV), DÜSS (IV), DAR.W (IV), SAAR (IV), R.H. PF (I) qui est caractérisé par la proximité des variables actifs totaux, actifs de l'industrie, actifs du tertiaire, population totale, densité, ainsi que des variables aviculture et porc.

REPARTITION DES REGIONS SUR LE 1er AXE

ANALYSE 64 VARIABLES



- du groupe AUV (III), LIM (III), SARD (II), BOU (III), FRC (III) MPY (sur l'axe) proche des paramètres, surface toujours en herbe, surface occupée par les exploitants à temps complet, exploitations entre 20 et 100 ha, et unités de bétail ovin-caprin.

Ce deuxième axe met en évidence d'un côté les régions urbanisées où l'industrie est importante et dont l'agriculture est orientée vers des productions animales intensives, et de l'autre côté des régions peu industrialisées dont l'agriculture est de type animal extensif. Cet axe peut être considéré comme reflétant le niveau de développement des régions qui y sont représentées de façon correcte. Toutefois, l'examen du graphe concernant le plan 1-2 de l'analyse des 64 variables (graphique 1) montre que la plupart des régions liées au premier facteur ne sont pas concernées par le second. Ainsi la plupart des régions méditerranéennes ne peuvent pas faire l'objet d'un classement en fonction de leur degré de développement. Ceci explique que l'on trouve ensemble deux régions telles BASIL et PRO (II) dont le niveau d'activité économique est fort différent.

3ème facteur

Son étude sur le plan 1-3 (voir graphique 2) permet de différencier :

- les observations PAR (IV), CHP (IV) CEN (IV) KOLN (IV), BS (IV) etc.... qui sont caractérisées par la présence des surfaces en céréales, en plantes industrielles, en betteraves sucrières, exploitations de plus de 50 ha, les surfaces en fermage, ainsi que par le revenu global, le revenu non agricole, les actifs totaux (I) et ceux de l'industrie et des services,
- les régions GELD (III), BRE (III) etc... associées aux exploitations de 5 à 10 ha et les surfaces de cultures fourragères.

Ce troisième facteur met en contraste les régions urbanisées et industrialisées et une agriculture de type grande culture avec système céréalier, par rapport aux régions où dominant les petites exploitations fourragères. Notons que l'on a des régions où se trouvent toutes les caractéristiques précédemment énoncées (système céréalier, grandes surfaces, fermage, haut niveau d'industrialisation) telles PAR, KOLN tandis que d'autres n'auront pour critère qu'un certain nombre de variables. C'est ainsi que CHP et CEN correspondent bien à des zones dont l'agriculture est de type grande culture avec céréales et grandes exploitations en fermage, mais où l'activité industrielle est faible.

4ème facteur

Le quatrième axe, d'interprétation aisée, met en évidence dans le plan 1-4 (**Graphique 4**) :

- d'une part, des observations comme OPF (III), AQU (II), NBAY (III), OFR (III), UMB (II), OBAY (III) etc... dont le caractère principal est la présence de superficies boisées dans les exploitations,
- d'autre part, les observations LIEG (IV), LUX.B (IV) AUR (IV), BRAB.B (IV) etc.. associées aux équins.

Les résultats de l'analyse portant sur ces 64 variables, nous permettent déjà d'ébaucher une typologie des régions en distinguant 8 groupes de régions :

- régions méditerranéennes à cultures spécialisées et excès de population agricole,
- régions septentrionales à spéculations animales, à base de bovins,
- régions industrialisées à spéculation animale intensive (aviculture, porc),
- régions peu industrialisées et à spéculation animale extensive (grandes superficies en pâturages, production ovine),
- régions de grande culture céréalière et où l'activité non agricole est importante,
- régions d'exploitations moyennes fourragères,
- régions caractérisées par la présence de surfaces boisées dans les exploitations,
- régions où l'on trouve des équins.

En fait, certains rapprochements sont possibles et certaines distinctions de peu d'intérêt. Mais on constatera que les facteurs précédemment définis ont un contenu essentiellement agricole. Ceci s'explique par le fait que les variables définissant le système de production ont une variance très forte, par rapport aux variables économiques générales. Ainsi la surface en cultures permanentes a un profil très particulier caractérisant de façon spécifique un certain nombre d'observations, C'est une variable forte, alors que le PIB par tête a un profil relativement régulier et une variance faible. Cette variable n'apparaît donc pas dans la définition d'un facteur.

Pour que la typologie envisagée tienne compte du contexte économique général de chaque région, l'agriculture ne devant pas être envisagée isolément, nous avons effectué une analyse portant uniquement sur les variables économiques générales (14 seulement ont été retenues sur les 15).

C - RESULTATS DE L'ANALYSE DES 14 VARIABLES ECONOMIQUES GENERALES

Les 14 variables retenues sont : les actifs totaux, ceux de l'agriculture, de l'industrie, des services, et les chômeurs; la population totale, la surface totale et la densité ; le revenu total, agricole et non agricole ; le PIB par tête, la productivité globale, agricole et non agricole. (Rappelons qu'il s'agit de la productivité du travail obtenue en divisant la valeur ajoutée par le nombre d'actifs).

Les caractéristiques des cinq premiers facteurs sont :

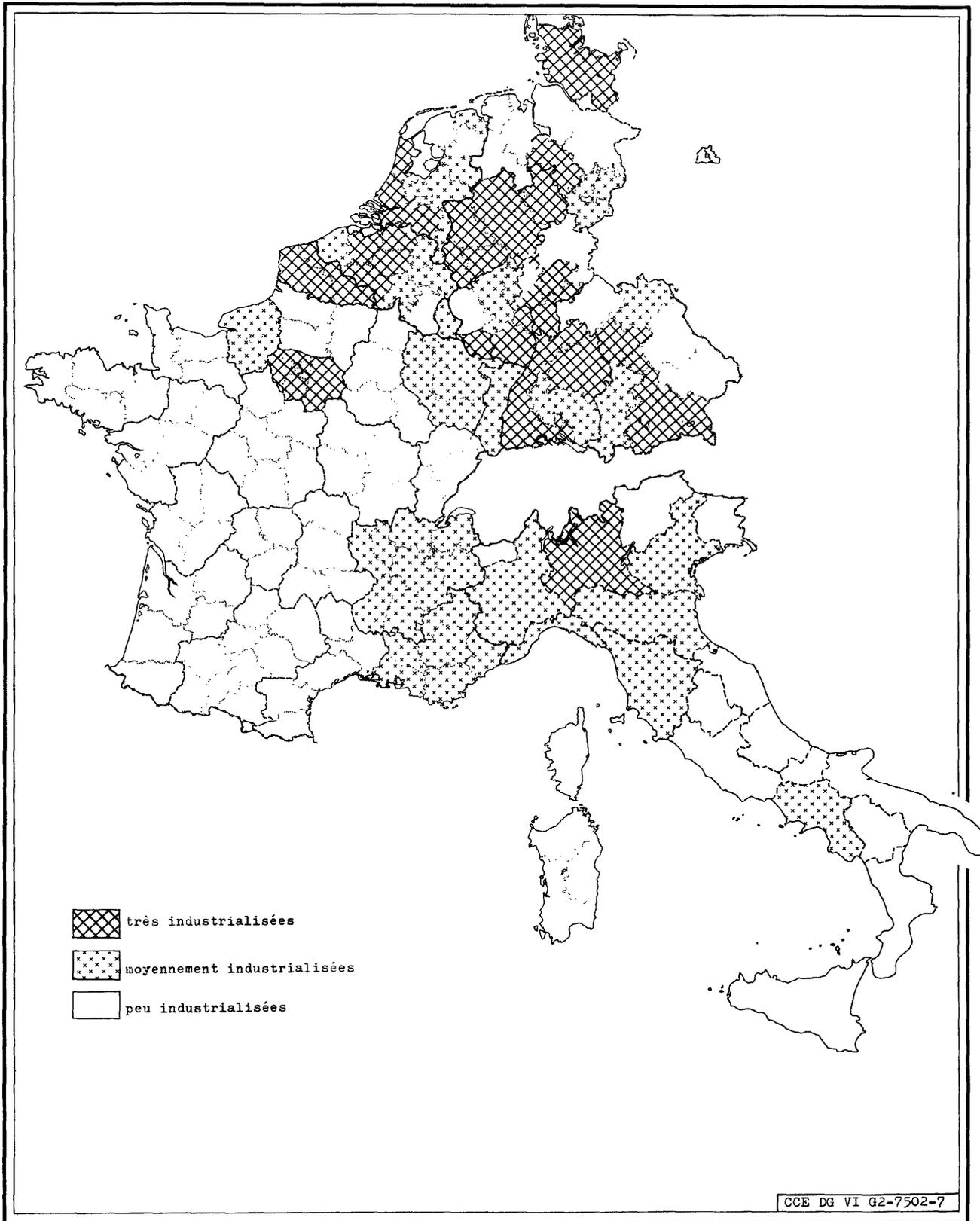
	1	2	3	4	5
% de la variance	48,28	27,51	11,09	7,57	2,89
valeur propre	0,087	0,049	0,020	0,013	0,005

Les deux premiers axes (voir graphique 5) représentant plus de 75 % de l'inertie totale du nuage de points, il n'apparaît pas intéressant d'utiliser les autres facteurs, d'autant plus que les valeurs propres obtenues sont peu élevées. Ceci montre bien la faible variance des variables économiques générales, par rapport aux variables agricoles.

Le premier facteur est particulièrement intéressant car il oppose des régions comme DÜSS, DAR.W., KOLN, ARNS, MUN, PAR, etc... caractérisées par les actifs de l'industrie, actifs totaux, actifs des services, revenu non agricole et revenu total à BNM, BASIL, PCH, UMB, LIM, TREN etc... proches des variables superficie totale, revenu agricole et actifs agricoles. Il est clair que cet axe est un axe d'industrialisation qui sépare les régions bénéficiant d'un niveau de développement élevé, des régions essentiellement agricoles peu peuplées et où l'industrie a une place négligeable.

Le deuxième facteur permet de distinguer HILD, LIEG, LIMB.B, GRON, BS, LUX, ALS, etc.. à proximité desquelles sont les variables concernant la productivité globale, agricole, le revenu par tête et la productivité non agricole, de TOSC, VEN, PIEM, EMI SIC, etc... proches du nombre d'actifs agricoles ainsi que de la population totale et des actifs totaux. Il faut toutefois avoir présent à l'esprit que les données concernant le PIB ainsi que le produit de l'agriculture et le produit des autres activités sont très difficiles à obtenir du niveau régional.

DEGRE D'INDUSTRIALISATION DES REGIONS REPARTITIONS SUR LE 1er AXE
ANALYSE 14 VARIABLES



Il est probable que ces chiffres, ainsi que ceux de la productivité, sont entachés d'une erreur assez importante. Ils fournissent uniquement un ordre de grandeur et n'ont de sens que pour les comparaisons.

Par rapport aux résultats précédemment obtenus dans l'analyse avec 64 variables, plusieurs constatations s'imposent :

- les régions caractérisées dans la première analyse, sur le 2ème axe, par la présence des actifs totaux, de la population totale et de l'aviculture, sont toutes situées sur l'axe d'industrialisation avec les régions développées. La seule exception est constituée par UFR (quadrant(IV), graphique 1, 1ère analyse et II, graphique 5, 2ème analyse). Cette observation dans le premier cas se situe avec les régions telles DAR.W, DUSS, NBAD alors que dans le 2ème cas, elle est opposée à ces mêmes régions se trouvant dans la zone où dominent surface totale, revenu agricole et actifs agricoles. Dans une moindre mesure, EMI serait dans la même situation.

- Inversement, les régions du 2ème axe (analyse 64 variables) proches des grandes exploitations avec une surface toujours en herbe importante, et des unités de bétail ovins-caprins se retrouvent dans la zone non industrialisée de l'axe 1 (analyse 14 variables). La seule exception vient de RHO qui ne peut être classée ni parmi les circonscriptions industrialisées, ni parmi celles essentiellement agricoles.

D - ESSAI DE SIMPLIFICATION DANS LE TABLEAU DES DONNEES

Les 64 variables définissant chaque région sont difficiles à rassembler lorsqu'on choisit les circonscriptions géographiques plus fines que celles utilisées. Nous avons donc tenté en supprimant certains paramètres de voir si la structure des graphes était modifiée ou non.

Un premier essai a porté sur la suppression du nombre d'exploitations par classe de surface ainsi que sur les variables définissant la productivité du travail et le revenu par tête. Cet analyse avec 51 variables donne des résultats absolument identiques à ceux obtenus avec 64 variables. Etant donné la très faible variance des chiffres concernant la productivité, celle-ci n'intervenant dans la définition d'aucun facteur, il n'était pas étonnant que leur suppression soit sans effet.

Par contre, il est plus étonnant, a priori, que la suppression du nombre d'exploitations par classe de surface soit sans effet. En fait, il semble bien que ceci soit dû à l'existence d'une relation très étroite entre taille des exploitations et système de production. Comme on dispose d'un jeu très complet de paramètres décrivant le système de production, la disparition du nombre d'exploitations par classe de S.A.U. n'a pas d'influence.

Poursuivant l'essai de simplification, on a oté un très grand nombre d'autres variables, aboutissant ainsi à un tableau 94 x 23. Les paramètres restant sont : Actifs totaux, densité, actifs de l'industrie, actifs des services, nombre de personnes qui dans la main d'oeuvre familiale travaillent plus de 2400 heures par an, surface en faire vabir direct, en fermage, en métayage, cultures permanentes, céréales, pommes de terre, betteraves, bois, surfaces toujours en herbe, plantes industrielles, cultures fourragères, total des unités de bétail, équins, bovins, ovins, porcs et volailles.

Les caractéristiques de l'analyse 23 variables sont :

	1	2	3	4	5
% inertie	33,04	15,25	10,79	9,38	8,00
valeur propre	0,1495	0,069	0,0488	0,0424	0,0362

La comparaison des résultats (graphique 6 et 7) avec l'analyse 64 variables (graphique 1 et 2) montre que les facteurs 1 et 2 ont une signification identique. Par contre, il apparait des modifications sur l'axe 3. Il semble que ce nouvel axe 3 soit la résultante du 3^e et du 4^{ème} facteur de l'analyse avec 64 variables. En effet, il oppose les superficies boisées et la main d'oeuvre familiale travaillant plus de 2 400 heures, aux surfaces en fermage, plantes industrielles et céréales.

Ainsi, il est possible de procéder à une analyse générale du même type que celle entreprise avec 64 variables en disposant d'un nombre plus restreint de paramètres. Le fait que l'on puisse obtenir des résultats comparables en raisonnant avec des données en termes monétaires et avec des données uniquement physiques, était déjà apparu dans l'analyse pour la France. Par ailleurs, l'étude du graphe obtenu avec le deuxième et le troisième facteurs de l'analyse avec 64 variables mettait en évidence le parallélisme entre les paramètres "actifs" et les paramètres "revenus", d'où la possibilité de suppression des variables revenus.

En nous aidant des analyses 64 variables et 14 variables, nous sommes en mesure de faire un essai de classification des régions européennes compte tenu à la fois du type d'agriculture existant et du niveau d'activité économique générale. Nous avons vu qu'il eût été possible de raisonner avec 23 variables, mais nous avons préféré utiliser l'ensemble des données disponibles.

E - ESSAIS DE CLASSIFICATION

Dans un premier stade, on a distingué les régions d'après leur niveau d'industrialisation, fort, faible ou moyen et d'après leur système de production agricole. Dans un deuxième stade, on a procédé à la fois à un regroupement des régions définies en fonction de leur système et à un éclatement dans certains cas en tenant compte du niveau de productivité.

1 - Classification d'après le degré d'industrialisation et le système de production

Certaines définitions de groupes pourront sembler se contredire, une observation apparaissant sous deux rubriques différentes. En fait, une région peut parfois être définie simultanément par plusieurs facteurs. En général, l'un prime et des caractères que l'on pourrait qualifier de secondaires, peuvent se manifester sur d'autres axes.

a - Régions industrialisées dont l'agriculture est de type

- céréaliier : Région Parisienne (PAR), Köln (KOLN), Düsseldorf (DUSS), Aachen (AC), Arnsberg (ARNS)
- polyculture élevage (système bovin dominant et porc) ; Schleswig-Holstein (SCHL) Nord (NRD), Münster (MUN), Hainaut (HAIN), Nord Brabant (N.BRAB), Hannover (HAN) Detmold (DETM), Flandre Orientale (O.VLA), Mittelfranken (MFR), Nord Holland (N.HOL), Zuid Holland (Z.HOL).

- spéculation animale intensive (aviculture et porc) : Nordbaden (NBAD), Nord Württemberg (NWUR), Düsseldorf (DUSS), Darmstadt-Wiesbaden (DAR.W), Saarland (SAAR), Rheinhesse-Pfalz (R.H.PF), Lombardia (LOM), Köln (KOLN), Aachen (AC), Detmold (DETM), Arnberg (ARNS), Brabant belge (BRAB.B), Région Parisienne (PAR), Hainaut (HAIN), Anvers (ANT).
- forestier : Oberbayern (OBAY), Arnberg (ARNS), Südbaden (SBAD).

b - régions peu industrialisées dont l'agriculture est de type

- céréaliier : Champagne (CHP), Centre (CEN), Picardie (PIC), Bourgogne (BOU), Zeeland (ZEEL).
- spéculation animale extensive (paturages, moutons) : Auvergne (AUV), Limousin (LIM), Sardagna (SARD), Bourgogne (BOU), Franche Comté (FRC), Midi-Pyrénées (MPY), Basse-Normandie (BNM), Poitou-Charante (PCH), Pays de Loire (LOI), Val d'Aosta (AOS).
- polyculture élevage : Stade (STAD), Bretagne (BRE), Drenthe (DREN), Oldenburg (OLD), Friesland (FRIES), Pays de Loire (LOI), Aurich (AUR), Osnabrück (OSN), Luxembourg belge (LUX.B), Kassel (KAS), Franche-Comté (FRC), Auvergne (AUV), Lüneburg (LUN).
- spéculation animale intensive : Unterfranken (UFR),
- méridional (cultures permanentes, petites exploitations, excès de population active) : Calabria (CAL), Sicilia (SIC), Puglia (PUG), Abruzzi (ABR), Languedoc (LAN), Basilicata (BASIL), Molise (MOL), Friuli (FRIU), Umbria (UMB), Trentino (TREN).
- forestier : Oberpfalz (OPF), Niederbayern (NBAY), Umbria (UMB), Aquitaine (AQU), Midi-Pyrénées (MPY), TRIER (TRI), Trentino (TREN), Lüneburg (LUN), Marche (MAR).

c - Régions moyennement industrialisées dont l'agriculture est de type

- céréaliier : Alsace . L'Alsace qui se situe avec le même système de production que CEN, CHP, ou PIC, mais à un niveau différent d'industrialisation. Ce n'est pas une région de grande culture comme les précédentes, mais elle possède beaucoup de céréales (c'est la sixième région française pour ce qui concerne la part relative de la S.A.U. en céréales).

- spéculation animale extensive ; Lorraine (LOR), Rhône-Alpes (RHO).
- polyculture-élevage : Overijssel (OVER), Schwaben (SCHW), Haute-Normandie (HNM), Luxembourg Grand Duché (LUX), Utrecht (UTR), Südwürttemberg (SWUR), Liège (LIEG), Lorraine (LOR), Flandre Occidentale (W.VLA), Limbourg (LIMB) Hildesheim (HILD), Limbourg belge (LIMB.B), Namur (NAM), Groningen (GRON).
- spéculation animale intensive : Limbourg (LIMB), Hildesheim (HILD), Emilia (EMI), Braunschweig (BS), Veneto (VEN), Koblenz-Montabaur (KOB.M.), Liège (LIEG), Namur (NAM).
- méridional : Provence-Côte d'Azur (PRO), Lazio (LAZ), Campania (CAM) Piemonte (PIEM), Liguria (LIG).
- forestier ; Südwürttemberg (SWUR), Oberfranken (OFR), Toscana (TOSC).

2 - Classification faisant intervenir le niveau de productivité

Il est aisé de constater que certaines régions figurent sous plusieurs rubriques soit parce qu'elles possèdent à la fois des surfaces boisées importantes par exemple et des bovins, soit parce que les rubriques ont un caractère complémentaire. Ainsi FRC est caractérisée à la fois par l'importance de ses surfaces toujours en herbe (axe 2 - analyse 64 variables) et par la prédominance de bovins (axe 1, analyse 64 variables). Certains regroupements sont de ce fait possibles. Par ailleurs, on peut distinguer parmi les régions à système identique et à même niveau de développement économique général, celles dont la productivité du travail est plus ou moins élevée. Il suffit pour cela de se référer au 2ème axe de l'analyse 14 variables. Toutefois, il faut noter que pour toutes les circonscriptions très fortement liées au 1er axe dans cette analyse, le 2ème axe ne peut pas être d'une grande utilité. Ainsi, toutes les régions fortement caractérisées par leur développement ou leur sous-développement ne peuvent pas être définies par rapport à un autre facteur. L'axe de productivité ne peut donc être utilisé que pour le cas intermédiaire d'une industrialisation moyenne. Dans le cas des régions développées, la productivité est obligatoirement élevée, et inversement en situation de sous-industrialisation il y a généralement une faible productivité du travail.

a - Régions industrialisées à agriculture de type

- 1 - céréaliier dominant : PAR
- 2 - céréaliier + spéculations animales : KOLN, DUSS, AC, ARNS
- 3 - spéculations animales (polyculture élevage et système intensif) SCHL, NBAD, NRD, MUN, NWUR, HAIN, N.BRAB, HAN, DAR.W, SAAR, R.H.PF, LOM, DETM, O.VLA, BRAB.B, MFR, ANT, N.HOL, Z.HOL.
- 4 - forestier : OBAY, SBAD.

On peut classer l'ensemble de ces circonscriptions dans la catégorie des zones favorisées. Ceci ne signifie pas qu'il n'y ait aucun problème agricole dans ces régions, mais que s'il y a des difficultés celles-ci sont plus faciles à résoudre en raison de l'industrialisation qui permet aux agriculteurs de trouver des emplois tout en restant sur leur exploitation ou en la quittant.

b - Régions peu industrialisées à agriculture de type

- 1 - céréaliier dominant : CHP, CEN, PIC, ZEEL
- 2 - spéculatiion animale extensive : AUV, FRC, BNM, LOI, LIM ainsi que BOU qui possède en plus des céréales ; SARD, PCH, AOS avec des ovins et MPY qui en outre a des surfaces boisées.
- 3 - polyculture-élevage : STAD, BRE, DREN, OLD , FRIES, AUR, OSN, LUX.B, KAS, LÜN, UFR et TRI avec des bois.
- 4 - Méridional : CAL, SIC, PUG, ABR, LAN, BASI, MOL, FRIU, avec UMB et TREN possédant des bois.
- 5 - forestier : OPF, NBAY, AQU, MAR.

Parmi ces régions peu industrialisées, toutes ne sont pas en difficulté. Celles où domine le système céréaliier (PIC, ZEEL, CHP et CEN) doivent être classées à part car malgré le petit nombre d'activités industrielles, elles ne souffrent d'aucun maux, par rapport à l'ensemble des régions agricoles européennes.

c - Régions moyennement industrialisés

Pour ces régions, il est intéressant de faire intervenir le critère de productivité générale et agricole qui permet de distinguer celles où existent des problèmes :

- 1 - système céréaliier : ALS à forte productivité.
- 2 - système animal intensif : KOB.M à productivité moyenne et EMI à faible productivité.
- 3 - système polyculture-élevage : OVER, SCHW, HNM, LUX, UTR, LOR, SWUR, LIEG, W.VLA, LIMB.B, NAM, GRON. Parmi ces régions : OVER, HNM, LUX, UTR, LOR,

LIEG, W.VLA, LIMB, HILD, LIMB.B, BS, NAM, GRON sont à forte productivité; SCHW, SWUR sont à productivité moyenne.

4 - système animal extensif : RHO avec faible productivité.

5 - méridional : PRO, LAZ, CAM, TOSC avec bois, PIEM, LIG, VEN avec productions animales intensives. Parmi ces régions : LIG est à forte productivité, LAZ, CAM, TOSC, PIEM et VEN ont une faible productivité. PRO a une productivité moyenne.

Dans ces régions à niveau d'activité économique moyen, se détache un groupe de régions dont l'agriculture ne paraît pas poser de problèmes majeurs, telles que ALS, HNM, LOR et les régions de Belgique (W.VLA, LIMB.B) et GRON, FRIES, GELD des Pays-Bas. Parmi les régions italiennes EMI, VEN (spéculations animales) LAZ, CAM, PIEM (cultures permanentes), TOSC (cultures permanentes et bois) PIEM (cultures permanentes) sont en difficulté. De même RHO pour la France.

En conclusion, on pourrait considérer comme régions "problèmes" celles qui sont peu développées (industrialisées) en classant par ordre décroissant celles à système méridional (10), forestier (4), animal extensif (8), polyculture-élevage (11). Les zones à système céréalier sont exclues. Au total on a donc 33 régions.

A ce groupe, on peut rajouter les régions moyennement développées (industrialisées) et à faible productivité en classant toujours par ordre décroissant : système méridional (5), animal extensif (1) et animal intensif (1), soit 7 régions.

Dans un tableau ci-joint, on a représenté l'ensemble des régions en fonction du système de production (résultats de l'analyse 64 variables) et du degré d'industrialisation ainsi que le niveau de productivité. Les régions "problèmes" sont entourées d'un trait.

Niveau d'industrialisation	système de production	céréaliier	animal intensif	Polyculture-élevage		animal extensif	avec activité forestière importante	type méridional avec cultures permanentes
				dominant	avec animal intensif			
FORT		PAR AC KOLN DUSS ARNS	NBAD NWDR DAR.W SAAR R.H.PF LOM BRAB.B ANT	NRD MUN HAIN N.BRAB HAN DETM O.VLA MFR Z.HOL N.HOL			OBAY SBAD	
MOYEN	(productivité élevée			SCHL OVER HNM LUX UFR LOR W.VLA LIMB.B GRON FRIES GELD SCHW SWUR	LIEG LIMB NAM HILD			LIG
		(productivité moyenne	BS AIS	KOB.M.			OFR	PRO
FAIBLE	(productivité faible		EMI.					LAZ CAM PIEM TOSC + bois VEN + animal intensif
		PIC CHP CEN ZEEL		DREN AUR OSN LUX.B KAS LUN OLD STA UFR BRE CAL			OFF NEAY AGU MAR	ABR LAN BASIL MOL PRIU CAL SIC PUG UMB) + bois TREN)

LISTE ET ABREVIATIONS DES 94 REGIONS

● FRANCE

ALS	Alsace
AQU	Aquitaine
AUV	Auvergne
BNM	Basse-Normandie
BOU	Bourgogne
BRE	Bretagne
CEN	Centre
CHP	Champagne
FRC	Franche-Comté
HNM	Haute-Normandie
LAN	Languedoc
LIM	Limousin
LOI	Pays de la Loire
LOR	Lorraine
MPY	Midi-Pyrénées
NRD	Nord
PAR	Région parisienne
PCH	Poitou-Charentes
PIC	Picardie
PRO	Provence-Côte d'Azur-Corse
RHO	Rhône-Alpes

○ BELGIQUE/BELGIË

ANT	Antwerpen
BRAB. B.	Brabant
HAIN	Hainaut
LIËG	Liège
LIMB. B.	Limburg
LUX. B.	Luxembourg
NAM	Namur
O. VLA	Oost-Vlaanderen
W. VLA	West-Vlaanderen

□ NEDERLAND

DREN	Drenthe
FRIES	Friesland
GELD	Gelderland
GRON	Groningen
LIMB	Limburg
N. BRAB	Noord-Brabant
N. HOL	Noord-Holland
OVER	Overijssel
UTR	Utrecht
ZEEL	Zeeland
Z. HOL	Zuid-Holland

○ LUXEMBOURG

LUX	Luxembourg
-----	------------

▼ B. R. DEUTSCHLAND

AC	Aachen
ARNS	Arnsberg
AUR	Aurich
BS	Braunschweig
DAR. W.	Darmstadt-Wiesbaden
DETM	Detmold
DÜSS	Düsseldorf
HAN	Hannover
HILD	Hildesheim
KAS	Kassel
KOB. M.	Koblenz-Montabaur
KÖLN	Köln
LÜN	Lüneburg
MFR	Mittelfranken
MÜN	Münster
NBAD	Nordbaden
NBAY	Niederbayern
NWUR	Nordwürttemberg
OBAY	Oberbayern
OFR	Oberfranken
OLD	Oldenburg
OPF	Oberpfalz
OSN	Osnabrück
R. H. PF	Rheinessen-Pfalz
SAAR	Saarland
SCHL	Schleswig-Holstein
SCHW	Schwaben
SBAD	Südbaden
STA	Stade
SWUR	Südwestfalen-Lippe
TRI	Trier
UFR	Unterfranken

■ ITALIA

ABR	Abruzzi
AOS	Val d'Aosta
BASIL	Basilicata
CAL	Calabria
CAM	Campania
EMI	Emilia-Romagna
FRIU	Friuli-Venezia Giulia
LAZ	Lazio
LIG	Liguria
LOM	Lombardia
MAR	Marche
MOL	Molise
PIEM	Piemonte
PUG	Puglia
SARD	Sardegna
SIC	Sicilia
TOSC	Toscana
TREN	Trentino-Alto Adige
UMB	Umbria
VEN	Veneto

● STRUCTURE DES EXPLOITATIONS

● Répartition du nombre d'exploitations par classe de SAU utilisée (en valeur absolue)

n_0 à n_8 n_0 : sans terre; n_1 : < 1ha; n_2 : 1 à 2ha; n_3 : 2 à 5ha;
 n_4 : 5 à 10ha; n_5 : 10 à 20ha; n_6 : 20 à 50ha;
 n_7 : 50 à 100ha; n_8 : > 100ha.

○ Superficie des exploitations dont le chef d'exploitation (en ha SAU) :

Stc travaille à temps complet;
 SEXT + à une activité extérieure de plus de 50 % de son temps de travail;
 Sext - à une activité extérieure de 50 % ou moins de son temps de travail.

■ Répartition de la SAU utilisée selon le mode de faire-valoir (en ha)

Fvd faire-valoir direct
 Ferm fermage
 Mét métayage

⊖ S superficie totale (agricole et autres) de la région (en ha)

◇ Traction (en valeur absolue)

Irs nombre d'exploitations n'utilisant pas de moyens de traction
 Ira nombre d'exploitations utilisant de la traction exclusivement animale
 Irm nombre d'exploitations utilisant de la traction exclusivement mécanique
 Itr nombre de tracteurs appartenant à l'exploitation

● PRODUCTION VEGETALE

▲ Répartition de l'utilisation du sol (en ha)

perm cultures permanentes
 cér céréales
 pot pommes de terre
 bet betteraves sucrières
 sv cultures sous verre
 boi superficie boisée de l'exploitation
 sth surfaces toujours en herbe (prairies, pâturages, alpages)
 ind plantes industrielles
 fl fleurs et plantes ornementales
 fou cultures fourragères des terres arables

□ PRODUCTION ANIMALE

▽ Répartition exprimée en UB (Unité-Bétail)

Ubt total des unités de bétail
 Eq équins
 Bov bovins
 Ovc ovins-caprins
 Por porcins
 Avi aviculture

VARIABLES MONETAIRES

◆ Revenu (PIB aux coûts des facteurs en mio U.C.)

Rt revenu total
 Rna revenu non agricole
 Ra revenu agricole
 r revenu par tête (PIB/tête en U.C.)

□ Productivité du travail (en U.C. par actif)

Pr productivité toutes branches
 Prna productivité non agricole
 Pra productivité agricole

POPULATION, MAIN-D'OEUVRE, CHEF D'EXPLOITATION

▲ Population, main-d'oeuvre

Ptot population totale (en 1.000)
 Pat population active totale (en 1.000)
 Paa population active agricole (en 1.000)
 Pai population active industrie (en 1.000)
 Pas population active services (en 1.000)
 Chô chômeurs (en 1.000)
 d densité de la population au km²

▲ Nombre de personnes qui, dans la main-d'oeuvre familiale (en valeur absolue) :

fot travaillent plus de 2.400 heures/an;
 for sont occupées régulièrement;
 fir sont occupées non régulièrement.

▽ Répartition du nombre des chefs d'exploitation par classe d'âge (en valeur absolue)

E_1 à E_6
 E_1 : < 35 ans; E_2 : 35 à 45 ans; E_3 : 45 à 50 ans;
 E_4 : 50 à 55 ans; E_5 : 55 à 65 ans; E_6 : > 65 ans.

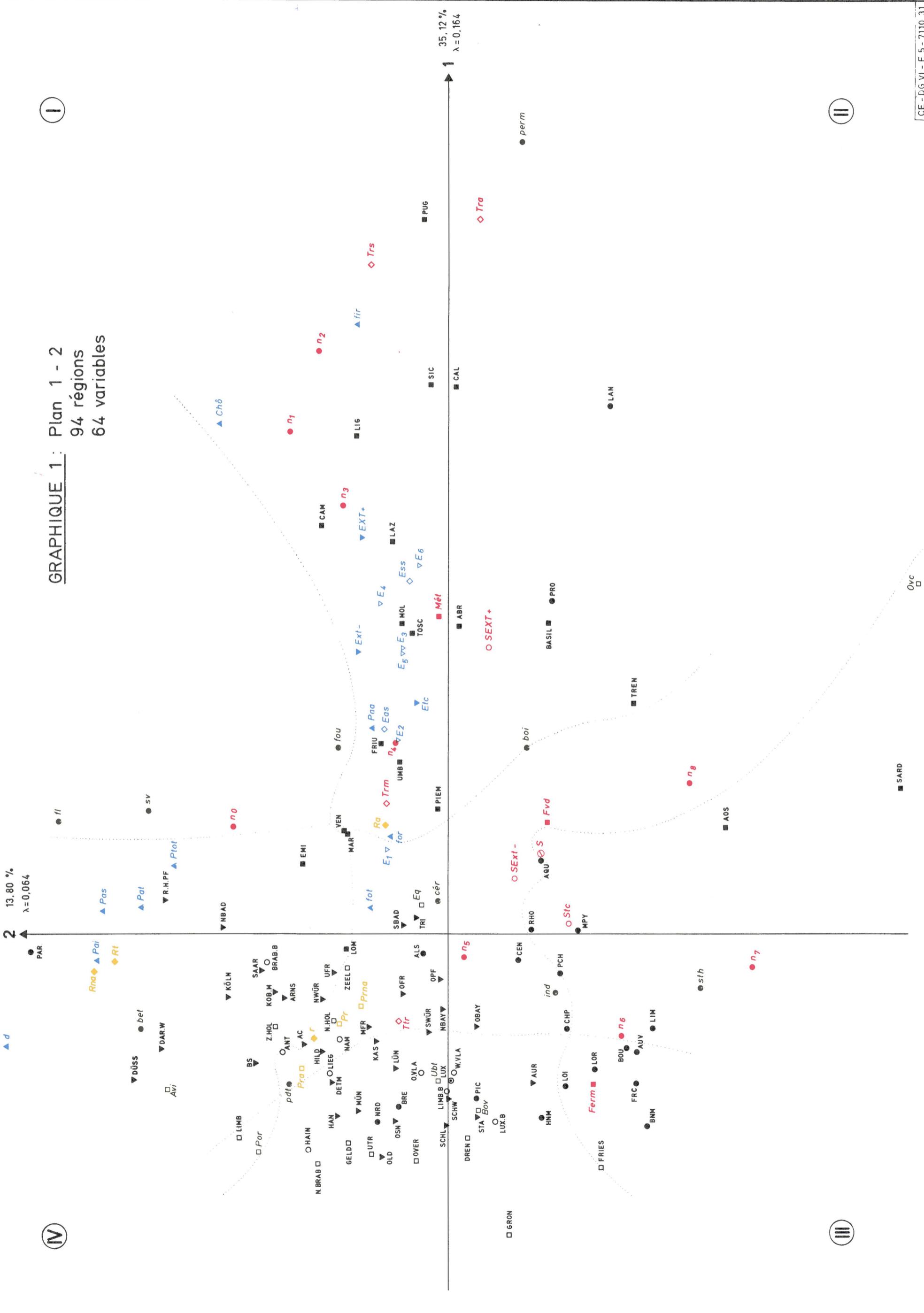
▲ Nombre de chefs d'exploitation qui (en valeur absolue) :

Etc travaillent à temps complet;
 EXT + ont une activité extérieure de plus de 50 % de leur temps de travail;
 Ext - ont une activité extérieure de 50 % ou moins de leur temps de travail.

◇ Nombre de chefs d'exploitation de plus de 50 ans (en valeur absolue)

Ess sans successeur potentiel
 Eas avec successeur potentiel

GRAPHIQUE 1: Plan 1 - 2
 94 régions
 64 variables



I

II

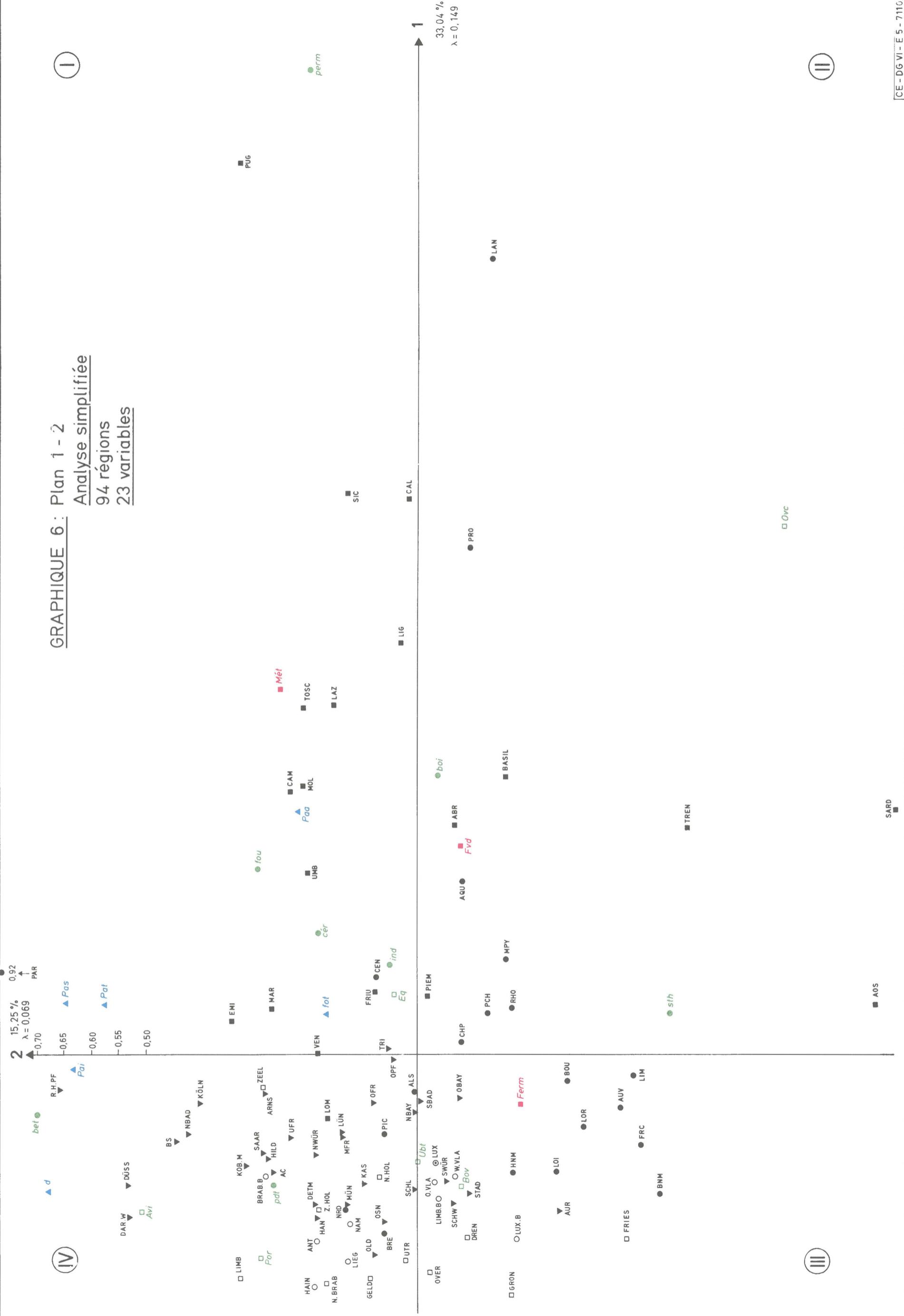
IV

III

GRAPHIQUE 5: Plan 1 - 2
 94 régions
 14 variables économiques génér.



GRAPHIQUE 6: Plan 1 - 2
 Analyse simplifiée
 94 régions
 23 variables



I

II

IV

III

TABLE DES MATIERES

Pages

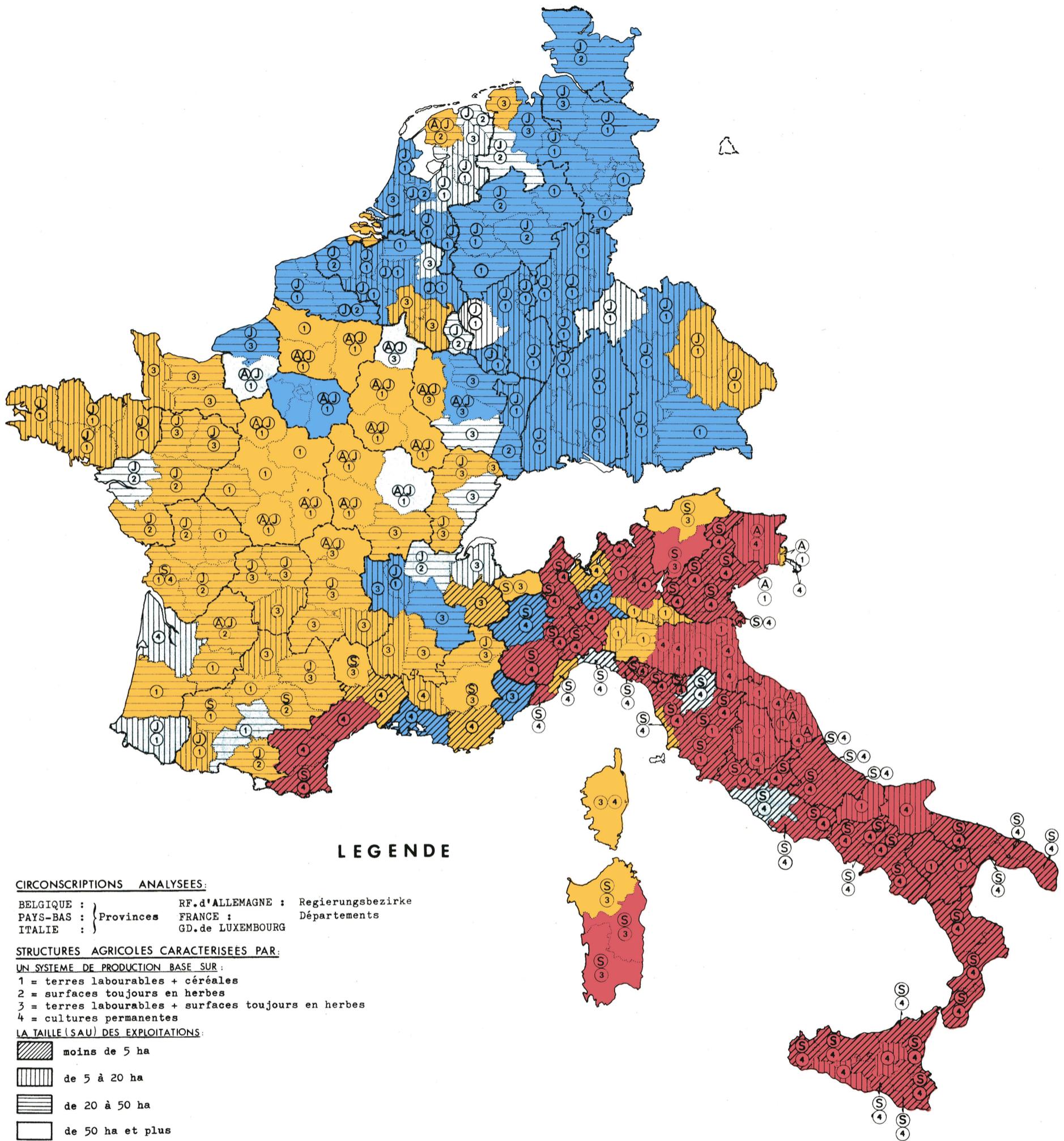
Avant-propos	3
Sommaire	5
<u>INTRODUCTION - LES OBJECTIFS DE L'ETUDE</u>	7
1. Objectif régional agricole	7
2. Objectif de développement régional	9
<u>CHAPITRE I - PROBLEMES METHODOLOGIQUES</u>	10
A. <u>Approche classique</u>	11
B. <u>L'analyse des données</u>	14
§ 1. L'utilisation du concept de distance par Stone	14
§ 2. Principes généraux de l'analyse des données	17
C. <u>Les bases de l'analyse des données</u>	19
§ 1. La classification automatique	20
§ 2. L'analyse factorielle	23
§ 3. Les principes de base de l'analyse des correspondances	26
<u>CHAPITRE II - DEFINITION DE L'ESPACE DES OBSERVATIONS ET DE L'ESPACE DES VARIABLES</u>	33
A. <u>Enseignement des analyses préliminaires</u>	34
§ 1. Nature des variables	34
§ 2. Nombre et type des variables	37
B. <u>Données de base pour l'analyse définitive</u>	41
§ 1. L'espace des observations	41
§ 2. L'espace des variables	44
<u>Annexe au Chapitre II :</u>	
Liste des symboles et abréviations utilisés dans les graphiques	57

TABLE DES MATIERES (suite)

	<u>Pages</u>
<u>CHAPITRE III - PRESENTATION DES RESULTATS</u>	67
Interprétation d'une analyse factorielle	67
A. <u>Analyse basée sur les seules variables agricoles</u>	71
1. Premier facteur	73
2. Deuxième facteur	75
3. Troisième facteur	79
4. Quatrième facteur	83
B. <u>Synthèse des résultats de l'analyse agricole</u>	86
1. Les associations significatives de chaque plan	87
2. Essai d'une typologie agricole	97
C. <u>Analyse économique générale</u>	103
1. Premier facteur	105
2. Deuxième facteur	106
3. Troisième facteur	109
4. Quatrième facteur	113
D. <u>Synthèse des résultats de l'analyse générale</u>	115
1. Les associations significatives de chaque plan de l'analyse économique générale	115
2. L'analyse économique générale dont on a supprimé les données concernant les structures agricoles et le système de production	125
E. <u>Essai d'une typologie sur la base de l'analyse économique générale</u>	129
<u>CHAPITRE IV - RECAPITULATION</u>	139

TABLE DES MATIERES (suite)

	<u>Pages</u>
<u>ANNEXES</u>	151
<u>Annexe 1 - Présentation mathématique simplifiée de l'analyse des données</u>	153
I. L'analyse factorielle des correspondances	154
II. La classification automatique	172
Bibliographie	182
<u>Annexe 2 - Résultats des analyses préliminaires</u>	185
I. Présentation des résultats pour la France - 21 régions	189
II. Présentation des résultats pour la C.E. (Six anciens Etats membres) - 94 régions	203



LEGENDE

CIRCONSCRIPTIONS ANALYSEES:

BELGIQUE : } RF.d'ALLEMAGNE : Regierungsbezirke
 PAYS-BAS : } Provinces FRANCE : Départements
 ITALIE : } GD.de LUXEMBOURG

STRUCTURES AGRICOLES CARACTERISEES PAR:

UN SYSTEME DE PRODUCTION BASE SUR:

- 1 = terres labourables + céréales
- 2 = surfaces toujours en herbes
- 3 = terres labourables + surfaces toujours en herbes
- 4 = cultures permanentes

LA TAILLE (SAU) DES EXPLOITATIONS:

- moins de 5 ha
- de 5 à 20 ha
- de 20 à 50 ha
- de 50 ha et plus

LA DEMOGRAPHIE DES CHEFS D'EXPLOITATIONS:

- J = âgés de moins de 50 ans
- A = âgés de plus de 50 ans avec successeurs potentiels
- S = âgés de plus de 50 ans sans successeur potentiel
- V = âgés de plus de 50 ans

DEVELOPEMENT ECONOMIQUE GENERAL:

- fortement
 - moyennement
 - non caractérisées par des activités secondaires et tertiaires
 - moyennement défavorisées
 - très défavorisées
- } caractérisées par des activités secondaires et tertiaires
- } sur le plan de l'industrialisation et des activités tertiaires

ESSAI DE TYPOLOGIE DES CIRCONSCRIPTIONS DE LA COMMUNAUTE A SIX

Structure agricole caractérisée par :			Développement économique général					
Système de production basé sur :	Taille des exploitations	Démographie des chefs d'exploitation*	Zones caractérisées par: activités secondaires et tertiaires		Zones non caractérisées par: activités secondaires et tertiaires	Zones défavorisées sur le plan de l'industrialisation et des activités tertiaires		
			fortement	moyennement		moyennement défavorisées	très défavorisées	
Terres labourables et Céréales	Grande	plus de 50 ans avec successeurs et moins de 50 ans	Seine (y compris Seine-et-Oise), Seine-et-Marne.	Eure, Côte-d'Or.	Aisne, Oise, Cher, Eure-et-Loir, Indre, Aube, Marne, Yonne, Haute-Marne.			
		non caractérisée			Somme, Loir-et-Cher, Loiret.			
	Moyenne	moins de 50 ans	Hannover, Lüneburg, Düsseldorf, Pas-de-Calais.					
		plus de 50 ans sans successeurs			Gers, (2) Charente-Maritime**			
		non caractérisée	Hildesheim, Aachen, Detmold, Köln, Braunschweig, Oberbayern, Antwerpen.	Haute-Garonne	Indre-et-Loire, Landes, Vienne, Lot-et-Garonne, Tarn-et-Garonne, Parma. (2)			
	Petite	moins de 50 ans	Hainaut, Rhône, Noord-Brabant, Darmstadt-Wiesbaden, Liège, Noord-Holland, Oost-Vlaanderen, Nordbaden, Saarland, Nordwürttemberg, Bas-Rhin, Limburg (Nederland), Rheinhessen-Pfalz, Mittelfranken, Kassel, Oberfranken, Koblenz-Montabaur, Südbaden, Schwaben, Südwürttemberg.	Pyrénées Atlantiques, Nord-Hollande, Gelderland, Unterfranken.	Hautes-Pyrénées (2) Côtes-du-Nord, Finistère, Ille-et-Vilaine, Morbihan, Niederbayern, Oberpfalz.			
		plus de 50 ans avec successeurs			Gorizia (2)	Macerata (1) Pesaro e Urbino.		
		non caractérisée			Piacenza, Cremona, Mantova. (2)	Ferrara, Bologna. (1)	Molise, Perugia. (1)	
		plus de 50 ans avec successeurs					Venezia (1)	
		plus de 50 ans sans successeurs					Grosseto, Siena. (1)	Napoli (1)
		non caractérisée					Bergamo. (1)	Potenza, Matera.
	Terres labourables et surfaces toujours en herbe	Moyenne	moins de 50 ans	Nord, Utrecht, West-Vlaanderen, Schleswig-Holstein, Arnberg, Münster.	Loire-Atlantique, Ain, Grand Duché de Luxembourg, Groningen, Osnabrück.	Ariège, Zeeland, Deux-Sèvres, Vendée, Maine-et-Loire.		
moins de 50 ans et plus de 50 ans (avec ou sans successeurs)						Dordogne, Friesland.		
plus de 50 ans sans successeurs						Tarn, Charente.		
non caractérisée				Haut-Rhin.				
Surfaces toujours en herbe	Grande	moins de 50 ans et plus de 50 ans avec successeurs	Meurthe-et-Moselle.	Ardennes.	Meuse, Nièvre, Allier. (2)			
		plus de 50 ans sans successeurs			Alpes de Haute (2) Provence, Lozère, Valle d'Aosta, Bolzano, Sassari.		Trento, Cagliari, (1) Nuoro.	
	Moyenne	moins de 50 ans	Seine-Maritime, Stade, Oldenburg.			Mayenne, Sarthe, (2) Creuse, Haute-Vienne, Aveyron, Puy-de-Dôme, Jura, Haute-Saône, Hautes-Alpes.		
		non caractérisée	Moselle, Isère.	Vosges, Doubs.	Drôme, Calvados, (2) Orne, Saône-et-Loire, Cantal, Aurich, Corse **.			
	Petite	non caractérisée	Zuid-Holland, Brabant belge (3) Loire.	Haute-Savoie (3) Drenthe, Limbourg belge.	Lot, Manche, (2) Belfort, Corrèze, Haute-Loire, Ardèche, Namur, Luxembourg belge.			
	Très petite	non caractérisée	Alpes-Maritimes (3)		Savoie. (2)			
	Cultures permanentes	Petite	plus de 50 ans avec successeurs				Ancona. (1)	Fordonone, (1) Udine, Ascoli.
			non caractérisée		Trieste, (3) Gironde.	Vaucluse. (2)	Forli, Modena, (1) Ravenna, Reggio Emilia, Terni.	Foggia, (1) Agrigento, Enna.
Très petite		plus de 50 ans sans successeurs	Torino (3)	Genova, (3) Firenze, Roma.	Savona, Livorno. (2)	Pescara, La Spezia, (1) Belluno, Rovigo, Imperia, Vercelli Novara, Pyrénées-Orientales, Pistoia, Pisa, Massa-Carrara.	Chieti, l'Aquila, (1) Teramo, Alessandria Asti, Cuneo, Treviso, Padova, Vicenza, Verona, Lucca, Arezzo, Benvenuto, Viterbo, Frosinone, Rieti, Latina, Avellino, Salerno, Caserta, Catanzaro, Cosenza, Reggio Calabria, Lecce, Bari, Brindisi, Taranto, Palermo, Trapani, Catania, Messina, Siracusa, Ragusa, Caltanissetta.	
		non caractérisée	Milano, (3) Bouches-du-Rhône.		Gard, Var, (2) Varese, Como.	Aude, Hérault, (1) Brescia.	Sondrio, Pavia. (1)	

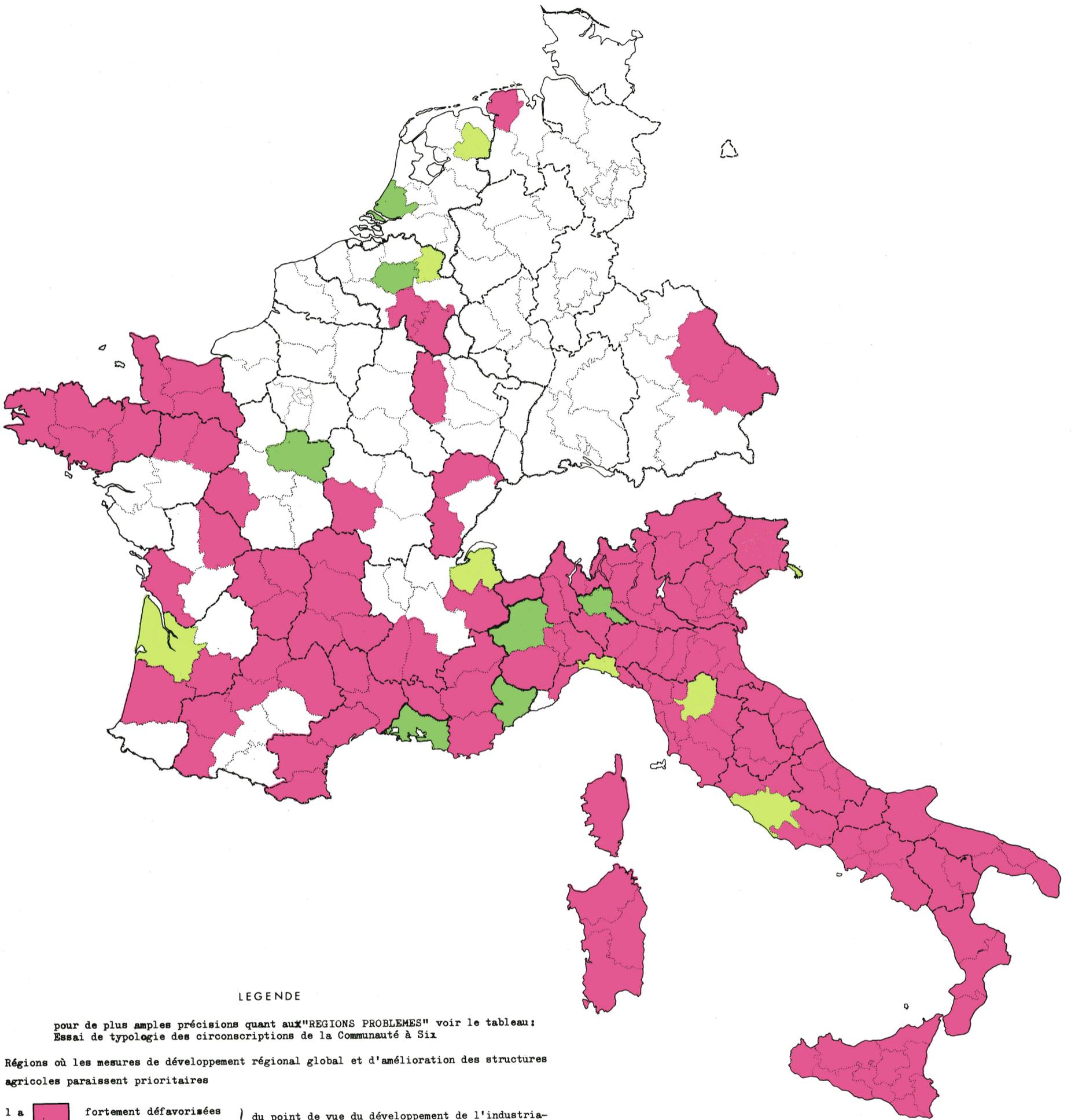
* Démographie des chefs d'exploitations cf. p. 46, p. 65 et p. 142 § a

** Caractérisées d'autre part par les cultures permanentes.

REGIONS "PROBLEMES"

1. Zones fortement ou moyennement défavorisées d'un point de vue de l'industrialisation et des activités tertiaires, le plus souvent avec des exploitations agricoles petites ou très petites ou, dans quelques cas, des grandes exploitations à système extensif, avec des chefs d'exploitation de plus de 50 ans dont la succession n'est, dans la majorité des cas, pas assurée ou incertaine. Ces régions constituent le groupe où les mesures de développement régional global et d'amélioration des structures agricoles paraissent prioritaires.
2. Zones non caractérisées par un développement des secteurs secondaires et tertiaires et où les systèmes de production agricole peu intensifs sont pratiqués dans des exploitations de taille le plus souvent moyenne ou petite et qui comprennent notamment des régions de montagne posant des problèmes de développement régional global moins urgents que le premier groupe, mais des problèmes graves du point de vue des structures agricoles dont la solution est conditionnée notamment par des efforts de diversification des activités agricoles.
3. Zones caractérisées par des exploitations agricoles petites et très petites, mais voisinant des structures économiques générales fortement ou moyennement développées. Ces zones ont besoin avant tout de pouvoir faire appel à des mesures favorisant l'agrandissement des exploitations agricoles et la conversion ou la cessation d'activité.

Essai de typologie des circonscriptions de la Communauté à Six
 "REGIONS PROBLEMES"



LEGENDE

pour de plus amples précisions quant aux "REGIONS PROBLEMES" voir le tableau:
 Essai de typologie des circonscriptions de la Communauté à Six

1. Régions où les mesures de développement régional global et d'amélioration des structures agricoles paraissent prioritaires

- 1 a fortement défavorisées
 - 1 b moyennement défavorisées
- } du point de vue du développement de l'industrialisation et des activités tertiaires

2. Régions posant des problèmes de développement régional global moins urgents que le premier groupe, mais des problèmes graves du point de vue des structures agricoles dont la solution est conditionnée notamment par des efforts de diversification des activités agricoles



3. Régions qui ont besoin avant tout de mesures favorisant l'agrandissement des exploitations agricoles et la conversion ou la cessation d'activité

- 3 a caractérisées moyennement
 - 3 b caractérisées fortement
- } par des activités secondaires et tertiaires

Informations internes sur L'AGRICULTURE

		Date	Langues
N° 1	Le boisement des terres marginales	juin 1964	F ⁽¹⁾ D ⁽¹⁾
N° 2	Répercussions à court terme d'un alignement du prix des céréales dans la CEE en ce qui concerne l'évolution de la production de viande de porc, d'œufs et de viande de volaille	juillet 1964	F ⁽¹⁾ D ⁽¹⁾
N° 3	Le marché de poissons frais en république fédérale d'Allemagne et aux Pays-Bas et les facteurs qui interviennent dans la formation du prix du hareng frais	mars 1965	F ⁽¹⁾ D ⁽¹⁾
N° 4	Organisation de la production et de la commercialisation du poulet de chair dans les pays de la CEE	mai 1965	F ⁽¹⁾ D ⁽¹⁾
N° 5	Problèmes de la stabilisation du marché du beurre à l'aide de mesures de l'Etat dans les pays de la CEE	juillet 1965	F D
N° 6	Méthode d'échantillonnage appliquée en vue de l'établissement de la statistique belge de la main-d'œuvre agricole	août 1965	F ⁽¹⁾ D ⁽²⁾
N° 7	Comparaison entre les « trends » actuels de production et de consommation et ceux prévus dans l'étude des perspectives « 1970 » 1. Produits laitiers 2. Viande bovine 3. Céréales	juin 1966	F ⁽¹⁾ D
N° 8	Mesures et problèmes relatifs à la suppression du morcellement de la propriété rurale dans les Etats membres de la CEE	novembre 1965	F ⁽¹⁾ D
N° 9	La limitation de l'offre des produits agricoles au moyen des mesures administratives	janvier 1966	F D
N° 10	Le marché des produits d'œufs dans la CEE	avril 1966	F ⁽¹⁾ D ⁽¹⁾
N° 11	Incidence du développement de l'intégration verticale et horizontale sur les structures de production agricole – Contributions monographiques	avril 1966	F ⁽¹⁾ D
N° 12	Problèmes méthodologiques posés par l'établissement de comparaisons en matière de productivité et de revenu entre exploitations agricoles dans les pays membres de la CEE	août 1966	F ⁽¹⁾ D
N° 13	Les conditions de productivité et la situation des revenus d'exploitations agricoles familiales dans les Etats membres de la CEE	août 1966	F D
N° 14	Situation et tendances des marchés mondiaux des principaux produits agricoles – « bovins – viande bovine »	août 1966	F D
N° 15	Situation et tendances des marchés mondiaux des principaux produits agricoles – « sucre »	février 1967	F D ⁽¹⁾
N° 16	Détermination des erreurs lors des recensements du bétail au moyen de sondages	mars 1967	F ⁽¹⁾ D ⁽³⁾

(1) Epuisé.

(2) La version allemande est parue sous le n° 4/1963 de la série « Informations statistiques » de l'Office statistique des Communautés européennes.

(3) La version allemande est parue sous le n° 2/1966 de la série « Informations statistiques » de l'Office statistique des Communautés européennes.

		Date	Langues
N° 17	Les abattoirs dans la CEE I. Analyse de la situation	juin 1967	F D
N° 18	Les abattoirs dans la CEE II. Contribution à l'analyse des principales conditions de fonctionnement	octobre 1967	F D
N° 19	Situation et tendances des marchés mondiaux des principaux produits agricoles – « produits laitiers »	octobre 1967	F D ⁽¹⁾
N° 20	Les tendances d'évolution des structures des exploitations agricoles – Causes et motifs d'abandon et de restructuration	décembre 1967	F D
N° 21	Accès à l'exploitation agricole	décembre 1967	F D
N° 22	L'agrumiculture dans les pays du bassin méditerranéen – Production, commerce, débouchés	décembre 1967	F D
N° 23	La production de produits animaux dans des entreprises à grande capacité de la CEE – Partie I	février 1968	F D
N° 24	Situation et tendances des marchés mondiaux des principaux produits agricoles – « céréales »	mars 1968	F D
N° 25	Possibilités d'un service de nouvelles de marchés pour les produits horticoles non-comestibles dans la CEE	avril 1968	F D
N° 26	Données objectives concernant la composition des carcasses de porcs en vue de l'élaboration de coefficients de valeur	mai 1968	F D
N° 27	Régime fiscal des exploitations agricoles et imposition de l'exploitant agricole dans les pays de la CEE	juin 1968	F D
N° 28	Les établissements de stockage de céréales dans la CEE – Partie I	septembre 1968	F D
N° 29	Les établissements de stockage de céréales dans la CEE – Partie II	septembre 1968	F D
N° 30	Incidence du rapport des prix de l'huile de graines et de l'huile d'olive sur la consommation de ces huiles	septembre 1968	F D
N° 31	Points de départ pour une politique agricole internationale	octobre 1968	F D
N° 32	Volume et degré de l'emploi dans la pêche maritime	octobre 1968	F D
N° 33	Concepts et méthodes de comparaison du revenu de la population agricole avec celui d'autres groupes de professions comparables	octobre 1968	F D
N° 34	Structure et évolution de l'industrie de transformation du lait dans la CEE	novembre 1968	F D
N° 35	Possibilités d'introduire un système de gradation pour le blé et l'orge produits dans la CEE	décembre 1968	F D
N° 36	L'utilisation du sucre dans l'alimentation des animaux – Aspects physiologiques, technologiques et économiques	décembre 1968	F D

(¹) Epuisé.

		Date	Langues
N° 37	La production de produits animaux dans des entreprises à grande capacité de la CEE – Partie II	février 1969	F D
N° 38	Examen des possibilités de simplification et d'accélération de certaines opérations administratives de remembrement	mars 1969	F D
N° 39	Evolution régionale de la population active agricole – I : Synthèse	mars 1969	F D
N° 40	Evolution régionale de la population active agricole – II : R.F. d'Allemagne	mars 1969	F D
N° 41	Evolution régionale de la population active agricole – III : Bénélux	avril 1969	F D
N° 42	Evolution régionale de la population active agricole – IV : France	mai 1969	F
N° 43	Evolution régionale de la population active agricole – V : Italie	mai 1969	F D
N° 44	Evolution de la productivité de l'agriculture dans la CEE	juin 1969	F D
N° 45	Situation socio-économique et perspectives de développement d'une région agricole déshéritée et à déficiences structurelles – Etude méthodologique de trois localités siciliennes de montagne	juin 1969	F I
N° 46	La consommation du vin et les facteurs qui la déterminent I. R.F. d'Allemagne	juin 1969	F D
N° 47	La formation de prix du hareng frais dans la Communauté économique européenne	août 1969	F D
N° 48	Prévisions agricoles – I : Méthodes, techniques et modèles	septembre 1969	F D
N° 49	L'industrie de conservation et de transformation de fruits et légumes dans la CEE	octobre 1969	F D
N° 50	Le lin textile dans la CEE	novembre 1969	F D
N° 51	Conditions de commercialisation et de formation des prix des vins de consommation courante au niveau de la première vente – Synthèse, R.F. d'Allemagne, G.D. de Luxembourg	décembre 1969	F D
N° 52	Conditions de commercialisation et de formation des prix des vins de consommation courante au niveau de la première vente – France, Italie	décembre 1969	F D
N° 53	Incidences économiques de certains types d'investissements structurels en agriculture – Remembrement, irrigation	décembre 1969	F
N° 54	Les équipements pour la commercialisation des fruits et légumes frais dans la CEE – Synthèse, Belgique et G.D. de Luxembourg, Pays-Bas, France	janvier 1970	F

		Date	Langues
N° 55	Les équipements pour la commercialisation des fruits et légumes frais dans la CEE – R.F. d'Allemagne, Italie	janvier 1970	F D
N° 56	Agriculture et politique agricole de quelques pays de l'Europe occidentale I. Autriche	mars 1970	F D
N° 57	Agriculture et politique agricole de quelques pays de l'Europe occidentale II. Danemark	avril 1970	F D
N° 58	Agriculture et politique agricole de quelques pays de l'Europe occidentale III. Norvège	avril 1970	F D
N° 59	Constatation des cours des vins de table à la production I. France et R.F. d'Allemagne	mai 1970	F D
N° 60	Orientation de la production communautaire de viande bovine	juin 1970	F
N° 61	Evolution et prévisions de la population active agricole	septembre 1970	F D
N° 62	Enseignements à tirer en agriculture d'expérience des «Revolving funds»	octobre 1970	F D
N° 63	Prévisions agricoles II. Possibilités d'utilisations de certains modèles, méthodes et techniques dans la Communauté	octobre 1970	F D
N° 64	Agriculture et politique agricole de quelques pays de l'Europe occidentale IV. Suède	novembre 1970	F D
N° 65	Les besoins en cadres dans les activités agricoles et connexes à l'agriculture	décembre 1970	F D
N° 66	Agriculture et politique agricole de quelques pays de l'Europe occidentale V. Royaume-Uni	décembre 1970	F D
N° 67	Agriculture et politique agricole de quelques pays de l'Europe occidentale VI. Suisse	décembre 1970	F D
N° 68	Formes de coopération dans le secteur de la pêche I. Synthèse, R.F. d'Allemagne, Italie	décembre 1970	F D
N° 69	Formes de coopération dans le secteur de la pêche II. France, Belgique, Pays-Bas	décembre 1970	F D
N° 70	Comparaison entre le soutien accordé à l'agriculture aux Etats-Unis et dans la Communauté	janvier 1971	F D
N° 71	Agriculture et politique agricole de quelques pays de l'Europe occidentale VII. Portugal	février 1971	F D
N° 72	Possibilités et conditions de développement des systèmes de production agricole extensifs dans la CEE	avril 1971	F D
N° 73	Agriculture et politique agricole de quelques pays de l'Europe occidentale VIII. Irlande	mai 1971	D

		Date	Langues
N° 74	Recherche sur les additifs pouvant être utilisés comme révélateurs pour la matière grasse butyrique – Partie I	mai 1971	F ⁽¹⁾
N° 75	Constatation de cours des vins de table II. Italie, G.D. de Luxembourg	mai 1971	F D
N° 76	Enquête auprès des consommateurs sur les qualités de riz consommées dans la Communauté	juin 1971	F D I
N° 77	Surfaces agricoles pouvant être mobilisées pour une réforme de structure	août 1971	F D
N° 78	Problèmes des huileries d'olive Contribution à l'étude de leur rationalisation	octobre 1971	F I
N° 79	Gestion économique des bateaux pour la pêche à la sardine – Recherche des conditions optimales – Italie, Côte Méditerranéenne française I. Synthèse	décembre 1971	F I
N° 80	Gestion économique des bateaux pour la pêche à la sardine – Recherche des conditions optimales – Italie, Côte Méditerranéenne française II. Résultats des enquêtes dans les zones de pêche	décembre 1971	F I
N° 81	Le marché foncier et les baux ruraux – Effets des mesures de réforme des structures agricoles I. Italie	janvier 1972	F D
N° 82	Le marché foncier et les baux ruraux – Effets des mesures de réforme des structures agricoles II. R.F. d'Allemagne, France	janvier 1972	F D
N° 83	Dispositions fiscales en matière de coopération et de fusion d'exploitations agricoles I. Belgique, France, G.D. de Luxembourg	février 1972	F
N° 84	Dispositions fiscales en matière de coopération et de fusion d'exploitations agricoles II. R.F. d'Allemagne	février 1972	D
N° 85	Dispositions fiscales en matière de coopération et de fusion d'exploitations agricoles III. Pays-Bas	février 1972	N
N° 86	Agriculture et politique agricole de quelques pays de l'Europe occidentale IX. Finlande	avril 1972	F D
N° 87	Recherche sur les incidences du poids du tubercule sur la floraison du dahlia	mai 1972	F D
N° 88	Le marché foncier et les baux ruraux – Effets des mesures de réforme des structures agricoles III. Pays-Bas	juin 1972	F D
N° 89	Agriculture et politique agricole de quelques pays de l'Europe occidentale X. Aperçu synoptique	septembre 1972	D

(¹) Etude adressée uniquement sur demande.

		Date	Langues
N° 90	La spéculation ovine	Septembre 1972	F
N° 91	Méthodes pour la détermination du taux d'humidité du tabac	Octobre 1972	F
N° 92	Recherches sur les révélateurs pouvant être additionnés au lait écrémé en poudre	Octobre 1972	F ⁽¹⁾ D ⁽¹⁾
N° 93	Nouvelles formes de collaboration dans le domaine de la production agricole - I : Italie	Novembre 1972	F I
N° 94	Nouvelles formes de collaboration dans le domaine de la production agricole - II : Benelux	Décembre 1972	F N
N° 95	Nouvelles formes de collaboration dans le domaine de la production agricole - III : R.F. d'Allemagne	Décembre 1972	F D
N° 96	Recherche sur les additifs pouvant être utilisés comme révélateurs pour la matière grasse butyrique - Partie II	Janvier 1973	F ⁽¹⁾ D ⁽¹⁾
N° 97	Modèles d'analyse d'entreprises de polyculture-élevage bovin - I : Caractéristiques et possibilités d'utilisation	Janvier 1973	F D
N° 98	Dispositions fiscales en matière de coopération et de fusion d'exploitations agricoles - IV : Italie	Janvier 1973	F I
N° 99	La spéculation ovine II. France, Belgique	Février 1973	F
N° 100	Agriculture de montagne dans la région alpine de la Communauté I. Bases et suggestions d'une politique de développement	Février 1973	F D I
N° 101	Coûts de construction de bâtiments d'exploitation agricole - Étables pour vaches laitières, veaux et jeunes bovins à l'engrais	Mars 1973	F en prép. D
N° 102	Crédits à l'agriculture I. Belgique, France, G.D. de Luxembourg	Mars 1973	F D
N° 103	La spéculation ovine III. R.F. d'Allemagne, Pays-Bas	Avril 1973	F
N° 104	Crédits à l'agriculture II. R.F. d'Allemagne	Avril 1973	D
N° 105	Agriculture de montagne dans la région alpine de la Communauté II. France	Mai 1973	F D
N° 106	Intégration verticale et contrats en agriculture I. R.F. d'Allemagne	Juin 1973	F D
N° 107	Agriculture de montagne dans la région alpine de la Communauté III. R.F. d'Allemagne	Juin 1973	F D

(¹) Etude adressée uniquement sur demande.

		Date	Langues
N° 108	Projections de la production et de la consommation de produits agricoles – « 1977 » I. Royaume-Uni	Août 1973	F D E en prép.
N° 109	Projections de la production et de la consommation de produits agricoles – « 1977 » II. Danemark, Irlande	Août 1973	F D E en prép.
N° 110	Nouvelles formes de collaboration dans le domaine de la production agricole IV. Synthèse	Septembre 1973	F D
N° 111	Modèles d'analyse d'entreprises de polyculture-élevage bovin II. Données technico-économiques de base Circonscription Nord-Picardie et région limoneuse du Limbourg belge	Septembre 1973	F
N° 112	La consommation du vin et les facteurs qui la déterminent II. Belgique	Septembre 1973	F N
N° 113	Crédits à l'agriculture III. Italie	Octobre 1973	F I
N° 114	Dispositions législatives et administratives concernant les résidus dans le lait, les produits laitiers et les aliments pour le cheptel laitier	Octobre 1973	F D
N° 115	Analyse du marché du porcelet dans l'optique d'une stabilisation du mar- ché du porc	Octobre 1973	F en prép. D
N° 116	Besoins de détente en tant que facteurs pour le développement régional et agricole	Novembre 1973	F
N° 117	Projections de la production et de la consommation de produits agricoles – « 1977 » III. Italie	Décembre 1973	F D en prép.
N° 118	Nouvelles formes de collaboration dans le domaine de la production agricole V. France	Décembre 1973	F
N° 119	Intégration verticale et contrats en agriculture II. Italie	Décembre 1973	F E I
N° 120	Projections de la production et de la consommation de produits agricoles – « 1977 » IV. R.F. d'Allemagne	Janvier 1974	F en prép. D
N° 121	Production laitière dans les exploitations ne disposant pas de ressources fourragères propres suffisantes	Janvier 1974	F D en prép. N
N° 122	Le rôle des ports de la Communauté pour le trafic de céréales et de farines I. Synthèse pour les principaux ports français et italiens	Février 1974	F
N° 123	Le rôle des ports de la Communauté pour le trafic de céréales et de farines II. Monographies pour les principaux ports français de la Manche	Février 1974	F
N° 124	Le rôle des ports de la Communauté pour le trafic de céréales et de farines III. Monographies pour les principaux ports français de l'Atlantique	Février 1974	F

		Date	Langues
N° 125	Le rôle des ports de la Communauté pour le trafic de céréales et de farines IV. Monographies pour les principaux ports français de la Méditerranée	Février 1974	F
N° 126	Le rôle des ports de la Communauté pour le trafic de céréales et de farines V. Monographies pour les principaux ports italiens de la côte Ouest	Février 1974	F
N° 127	Le rôle des ports de la Communauté pour le trafic de céréales et de farines VI. Monographies pour les principaux ports italiens de la côte Est	Février 1974	F
N° 128	Projections de la production et de la consommation de produits agricoles - « 1977 » V. Pays-Bas	Mars 1974	F D
N° 129	Projections de la production et de la consommation de produits agricoles - « 1977 » VI. Résultats pour la Communauté européenne	Avril 1974	F D
N° 130	Utilisation de produits de remplacement dans l'alimentation animale	Mai 1974	F E en prép.
N° 131	Recherche sur les additifs pouvant être utilisés comme révélateurs pour la matière grasse butyrique - Partie III	Juin 1974	F
N° 132	La consommation du vin et les facteurs qui la déterminent III. Pays-Bas	Juin 1974	F en prép. N
N° 133	Les produits dérivés de la pomme de terre	Août 1974	F
N° 134	Projections de la production et de la consommation de produits agricoles - « 1977 » VII. Belgique, Grand-Duché de Luxembourg	Septembre 1974	F D en prép.
N° 135	La pêche artisanale en Méditerranée - Situation et revenus	Octobre 1974	F I en prép.
N° 136	La production et la commercialisation de parties de volaille	Octobre 1974	F D en prép.
N° 137	Conséquences écologiques de l'application des techniques modernes de production en agriculture	Novembre 1974	F D en prép.
N° 138	Essai d'appréciation des conditions d'application et des résultats d'une politique de réforme en agriculture dans des régions agricoles difficiles I. Morvan	Décembre 1974	F
N° 139	Analyse régionale des structures socio-économiques agricoles - Essai de typologie régionale pour la Communauté des Six Partie I : Rapport	Janvier 1975	F

