REVUE DE STATISTIQUE APPLIQUÉE

Y. LE FOLL
B. BURTSCHY

Représentations optimales des matrices imports-exports

Revue de statistique appliquée, tome 31, nº 3 (1983), p. 57-72
<http://www.numdam.org/item?id=RSA_1983__31_3_57_0>

© Société française de statistique, 1983, tous droits réservés.

L'accès aux archives de la revue « Revue de statistique appliquée » (http://www.sfds.asso.fr/publicat/rsa.htm) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

REPRESENTATIONS OPTIMALES DES MATRICES IMPORTS-EXPORTS

Y. LE FOLL et B. BURTSCHY

Ecole Nationale Supérieure des Télécommunications
46, rue Barrault 75634 Paris Cedex 13

RESUME

Pour analyser les échanges commerciaux de 53 pays (ou groupes de pays), il y a lieu de choisir une méthode de représentation graphique des matrices d'importations-exportations (éventuellement ventilées par produits).

Après avoir discuté ce choix, on démontre que la représentation simutanée de l'Analyse des Correspondances multiples est graphiquement optimale. Les échanges intérieurs des pays risquant de perturber arbitrairement cette représentation, un algorithme itératif est proposé pour rendre cette perturbation minimale.

L'optimalité des projections factorielles est en fait relative à la façon dont on pondère les distances de la représentation. C'est ainsi que pour les échanges de 1976, on analyse deux systèmes de projection : l'un pour identifier des facteurs généraux d'échanges (système usuel), l'autre pour apprécier les spécificités d'un exportateur particulier (système local).

REPRESENTATIONS OPTIMALES
DES MATRICES IMPORTS-EXPORTS

INTRODUCTION

Les experts du Commerce Extérieur (cf. [13]) étudient avec beaucoup de profit les flux d'échanges commerciaux entre nations. Habituellement représentés par une matrice C à double entrée (pays importateurs en ligne, pays exportateurs en colonne), ceux-ci peuvent être ventilés par produits, par périodes... etc. De nombreuses études ont montré l'intérêt qu'il y a à traiter ces tableaux par analyse multidimensionnelle, tant pour DECRIRE (cf. [3, 5, 23] et [30]) que pour PREVOIR les principaux flux d'échanges (cf. [17] et [30] notamment).

Cependant même au plan descriptif, certains aspects méthodologiques demandent encore à être précisés. Ainsi en ce qui concerne la représentation des matrices imports-exports (ventilées ou non), plusieurs approches sont envisageables : les techniques suggérées par la théorie des graphes, les méthodes taxonomiques, les représentations factorielles. Comment choisir entre ces possibilités et notamment entre l'Analyse des Correspondances et l'Analyse Sphérique (cf. [14]) ? Les éléments "diagonaux" de la matrice (c'est-à-dire les échanges des pays avec eux-mêmes) ne perturbent-ils pas arbitrairement la représentation ? Est-il possible

Mots clés : **Représentation optimale, Matrice d'échanges, Analyse des correspondances, Analyse pondérée, Analyse de données incomplètes.**

d'adapter cette représentation pour analyser la position d'un importateur ou d'un exportateur particulier ?

Il ne s'agit pas de donner des réponses exclusives à ces questions, mais de justifier les solutions retenues en termes de REPRESENTATIONS OPTIMALES. Après avoir introduit quelques notations (§.1) et avoir discuté le choix du modèle de représentation (§.2), on montre l'optimalité graphique de l'Analyse Factorielle des Correspondances (AFC) pour la visualisation de telles données relationnelles (§.3). L'étude de l'influence des éléments "diagonaux" sur la représentation conduit à une technique itérative d'estimation toujours convergente (§.4). Quant à la vision "locale" des échanges, elle ne remet pas en cause la représentation d'ensemble mais ses plans de projections (§.5). Comme application, on présente une analyse globale du commerce international de 1976 et une analyse locale de la situation des exportations françaises dans le contexte précédent (§.6).

§.1 NOTATIONS

Etant donné un ensemble $I = \{1, \ldots, i, \ldots n\}$ de pays importateurs et un ensemble $E = \{1, \ldots, e, \ldots p\}$de pays exportateurs, on note C la matrice (n lignes et p colonnes) d'imports-exports : C_{ie} est l'importation du pays i venant de l'exportateur e. Par cumul, on obtient le total T_i importé par i, le total T_e exporté par e et le total général T.

$$T_i = \sum_e C_{ie} \quad T_e = \sum_i C_{ie} \quad \text{et} \quad T = \sum_i T_i \tag{1}$$

Mais les échanges peuvent être ventilés en fonction d'un ensemble $L = \{1, \ldots, \ell, \ldots q\}$ de produits. On note alors $C_{i\ell e}$ l'importantion de i en produit ℓ venant de l'exporteur e. C_{ie} possède la propriété additive suivante :

$$\forall\, i \in I \quad \text{et} \quad e \in E \quad C_{ie} = \sum_{\ell} C_{i\ell e} \tag{2}$$

$\{C_{i\ell e} | i \in I,\ \ell \in L,\ e \in E\}$ constitue une hypertable de contingence, à laquelle il est d'usage d'associer un tableau de Burt B (cf [22] p. 125). Si J note la somme ensembliste $I + L + E$, alors B est une matrice symétrique de dimension $n + q + n$, se décomposant en 3×3 blocs :

$$B = \begin{bmatrix} B_{II} & B_{IL} & B_{IE} \\ B_{LI} & B_{LL} & B_{LE} \\ B_{EI} & B_{EL} & B_{EE} \end{bmatrix}$$

Les blocs diagonaux sont diagonaux : par exemple

$$\forall j \in I,\ \forall j' \in I \quad B_{jj'} = \delta_{jj'} T_j \quad \text{pour le bloc } B_{II}$$

où $\delta_{jj'}$ est le symbole de Kronecker

avec

$$\forall j \in L \quad T_j = \sum_i \sum_e C_{ije} \quad \text{pour le bloc } B_{LL}$$

Les blocs non-diagonaux sont des tables de contingence binaires :

$$\forall j \in I, \quad \forall j' \in L \quad B_{jj'} = \sum_e C_{jj'e} \quad \text{pour le bloc } B_{IL}$$

Ces notations pourraient être étendues à un nombre quelconque de ventilations. Pour simplifier les écritures, on s'est limité à trois ensembles I, L et E. Mais il est simple d'imaginer le cas où J est la somme de m ensembles.

§.2 CHOIX D'UNE METHODE DE REPRESENTATION

La méthode recherchée ici doit permettre d'appréhender sur des graphiques les principaux flux d'échanges entre les pays tels qu'ils apparaissent dans la matrice C par exemple. Essayons de passer en revue les approches les plus "naturelles".

A la matrice C, il est possible d'associer un graphe valué, ayant les éléments de I et E comme sommets et les éléments de C comme valeurs des arcs. Ce graphe étant complet ou presque, il s'agit d'en donner une vision simplifiée. La méthode "des seuils" (suppression des arcs inférieurs à un certain seuil cf [28]) est trop sensible aux différences de taille $\{T_i | i \in I\}$ et $\{T_e | e \in E\}$ des pays, pour être utilisée ici. La méthode des "hiérarchies centrifuges" (cf [12]) offre une possibilité de représentation automatique des graphes non valués : il faut donc savoir choisir les arc "pertinents" du graphe associé à C, ce qui ramène au problème précédent.

Une autre approche consiste à définir un indice de proximité, soit entre pays importateurs soit entre pays exportateurs (cf [5]) ; il est alors possible de construire des arbres hiérarchiques ou des images euclidiennes restituant approximativement ces proximités. Les solutions les plus intéressantes sont associées à la notation d'information mutuelle et à la notion de distance distributionnelle (cf [1]). Les pays y sont classés en fonction de leurs profils d'exportation ou de leurs profils d'importation.

Parmi ces méthodes, l'AFC et l'analyse sphérique (cf [14]) sont particulièrement intéressantes : car elles offrent une représentation simultanée de I, E et éventuellement L. Cela est dû au fait qu'en échangeant les rôles de I et de E, les représentations obtenues sont liées par des relations de transition. En AFC, ces relations sont pseudo-barycentriques (cf [22]), ce qui place un importeur au barycentre (au facteur $(\lambda_a)^{-1/2}$ près) de ses fournisseurs et vice-versa. Par exemple

$$F_a^i \sqrt{\lambda_a} = \sum_e \frac{C_{ie}}{T_i} F_a^e \quad \forall i \in I$$

où F_a^i et F_a^e sont les coordonnées factorielles de rang a des points i et e respectivement.

En analyse sphérique, la formule correspondante (l'autre s'en déduisant aisément) comporte un effet de taille en $\sqrt{T_e}$.

$$F_a^i \sqrt{\lambda_a} = \sum_e \sqrt{T_e}\, F_a^e \sqrt{\frac{C_{ie}}{T_i}} \quad \forall i \in I$$

Si l'analyse sphérique ne peut être retenue ici, vues les grandes fluctuations des tailles T_i et T_e, ses développements originaux suggèrent des aménagements en AFC (cf § 4 et § 5). D'autres modèles factoriels (cf [3,11] et [15]) ont été proposés, mais en mettant l'accent sur la symétrie ou non de C, ils répondent à d'autres objectifs.

Il semble donc que l'AFC soit particulièrement adaptée pour représenter les matrices imports-exports. Non seulement elle élimine bien les différences de taille, mais sa représentation simultanée est optimale, comme on va le voir.

Remarque

D'intéressants modèles statistiques (cf [9] et [10]) ont été proposés pour l'analyse des tableaux d'échanges ou des tables de contingence. Mais ils ne correspondent pas au problème posé ici.

§.3 OPTIMALITE GRAPHIQUE DE L'AFC

Il s'agit d'interpréter (dans le cas d'une matrice imports - exports) et de généraliser la propriété extrêmale suivante de l'AFC du tableau C :

Propriété 1

Le premier facteur normé $\{a_1^i \mid i \in I\}$ et $\{b_1^e \mid e \in E\}$ de l'AFC du tableau C rend extremum (minimum) le résidu :

$$R = \sum_i \sum_e C_{ie}(a^i - b^e)^2$$

sous la contrainte de normalisation suivante :

$$\sum_i T_i(a^i)^2 + \sum_e T_e(b^e)^2 = 1$$

et si la valeur propre associée à ce facteur est notée λ_1^2,

$$1 - \lambda_1 = \sum_i \sum_e C_{ie}(a_1^i - b_1^e)^2$$

Cette propriété très classique (cf [7]) montre que la représentation simultanée de l'AFC de C tend à rapprocher sur son premier facteur les importateurs i et les exportateurs e ayant les échanges les plus importants :

$$C_{ie} \text{ important} \Rightarrow |a_1^i - b_1^e| \text{ faible.}$$

Cette propriété peut être étendue au cas de tables C à trois entrées (ou plus) et de plusieurs facteurs, ce qui permet de caractériser la représentation en facteurs normés à 1 associée à l'analyse des correspondances multiples (c'est-à-dire l'AFC de B)

Propriété 2

Si $\{b_a^j \mid j \in J,\ a = 1, k\}$ sont k vecteurs orthonormés au sens suivant :

$$\sum_j 3T_j b_a^j b_{a'}^j = \delta_{aa'} \quad \forall a = 1, k\,;\ a' = 1, k \tag{4}$$

Les trois conditions suivantes sont équivalentes :

(i) ils sont les k premiers facteurs de norme 1 de l'AFC de B

(ii) ils rendent minimum : $R_k = \sum_{a=1}^{k} \sum_j \sum_{j'} B_{jj'}(b_a^j - b_a^{j'})^2$

(iii) ils rendent minimum la quantité :

$$S_k = \sum_{a=1}^{k} \sum_i \sum_l \sum_e C_{ile}\, \Delta^2(b_a^i, b_a^l, b_a^e)$$

où $\Delta^2(b_a^i, b_a^l, b_a^e)$ est l'écart quadratique moyen associé

Preuve

- Pour montrer que (i) $\Rightarrow$ (ii), il suffit d'expliciter R_k

$$R_k = \sum_{a=1} \sum_j \sum_{j'} B_{jj'}[(b_a^j)^2 + (b_a^{j'})^2 - 2\, b_a^j\, b_a^{j'}]$$

Or $$(4) \Rightarrow \sum_j \sum_{j'} B_{jj'}\,(b_a^j)^2 = \sum_j 3\,T_j\,(b_a^j)^2 = 1$$

de plus (i) $\Rightarrow$ Si λ_a^2 est la valeur propre associée au facteur de rang a

$$\sum_{j'} B_{jj'}\, b_a^{j'} = 3\,\lambda_a\, T_j\, b_a^j$$

d'où $$R_k = \sum_{a=1}^{k} \left[1 + 1 - 2\lambda_a \sum_j 3T_j\,(b_a^j)^2\right]$$

$$(4) \Rightarrow R_k = 2\,[k - \lambda_1 - \lambda_2 - \ldots - \lambda_k]$$

Comme les valeurs propres associées aux k premiers facteurs sont les k plus grandes, il est clair que R_k est minimal.

• Réciproquement, montrer que (ii) $\Rightarrow$ (i) est une simple extension de la propriété 1 avec substitution de B à C ce qui ne pose formellement aucune difficulté. Sous les conditions de normalisation (4), rendre minimum R_k équivaut à rendre maximum la quantité :

$$I_k = \sum_{a=1} \sum_j \sum_{j'} B_{jj'} \, b_a^j \, b_a^{j'}$$

où l'on reconnaît une inertie projetée sur un sous-espace de dimension k. Ce problème classique conduit à identifier les k premiers facteurs de norme 1 de l'AFC de B (cf. [2]).

• Il reste à montrer l'équivalence de (ii) et (iii). L'écart quadratique moyen du m-uplet $(x_1, \ldots, x_m)$ a pour expressions :

$$\Delta^2 (x_1, \ldots, x_m) = \frac{1}{m} \sum_i \left[x_i - \sum_{i'} \frac{x_{i'}}{m} \right]^2$$

$$= \frac{1}{2m^2} \sum_i \sum_{i'} (x_i - x_{i'})^2 .$$

$$\Rightarrow \Delta^2 (b_a^i, b_a^e, b_a^1) = \frac{1}{9} [(b_a^i - b_a^e)^2 + (b_a^i - b_a^1)^2 + (b_a^e - b_a^1)^2]$$

$$\Rightarrow S_k = \frac{1}{9} \sum_{a=1}^{k} \left[\sum_i \sum_e C_{ie} \, (b_a^i - b_a^e)^2 + \sum_i \sum_1 C_{i1} (b_a^i - b_a^1)^2 + \sum_e \sum_1 C_{e1} \, (b_a^e - b_a^1)^2 \right]$$

$$\Longleftrightarrow S_k = \frac{1}{18} \sum_a \sum_j \sum_{j'} B_{jj'} \, (b_a^j - b_a^{j'})^2 = \frac{1}{18} \, R_k$$

Plus généralement, si la table C est à m entrées, il est clair que

$$S_k = \frac{1}{2m^2} \, R_k \quad \text{ce qui rend : (ii)} \Longleftrightarrow \text{(iii)}$$ ■

Remarques

a) En permutant l'ordre des sommations dans S_k, on obtient :

$$9 S_k = \sum_i \sum_1 \sum_e C_{i1e} \, [{}_k d^2 (i, 1) + {}_k d^2 (i, e) + {}_k d^2 (1, e)]$$

où ${}_k d (i, e)$ note la distance euclidienne classique des points i et e en projection sur l'ensemble des k premiers facteurs considérés.

La condition (iii) signifie donc que les points i, 1 et e sont d'autant plus proches sur l'ensemble des k premiers facteurs que l'échange C_{i1e} est important.

b) La propriété 2 montre que le positionnement d'éléments, tels que (i, 1, e), ou (i, 1) ou (i, e) ou (1, e) sur les facteurs de l'AFC de B a un sens : il suffit de les imaginer au centre de gravité des éléments qui les composent.

c) Le tableau de BURT B ne dépendant que des tables de contingence binaires, la représentation considérée élimine les interactions d'odre $\geqslant 2$ (cf. [10]). Cette difficulté est classiquement résolue (cf. [1]) en rangeant les C_{ile} dans un tableau à deux entrées (I et L x E par exemple) et en traitant celle-ci comme telle.

d) Alors qu'en analyse des proximités il est naturel de construire les représentations graphiques avec des facteurs de norme λ_a, c'est avec des facteurs normés à 1, qu'il convient de procéder en AFC, pour disposer des propriétés 1 et 2. On montre que cette représentation s'obtient aussi par simple projection à partir de la distance "atomique" du χ^2 (cf. [24]) ; d'ailleurs les propriétés 1 et 2 servent alors de critère de sélection des axes factoriels.

§.4 ESTIMATION OPTIMALE DES ELEMENTS "DIAGONAUX"

La matrice C comporte des éléments "diagonaux" (c'est-à-dire à l'intersection des lignes et des colonnes associées à un même pays), dont les valeurs sont purement conventionnelles. Comme le notent D. DOMENGES et M. VOLLE à propos des tableaux d'échanges inter-industriels (cf. [14]), ces valeurs peuvent être estimées sur la base d'une nomenclature fine des zones d'échanges. Par agrégation des zones suivant la nomenclature des (groupes de) pays composant I et E, on retrouve bien les montants d'échanges inter-pays. En revanche les montants d'échanges intra-pays, c'est-à-dire les valeurs des éléments "diagonaux", dépendent beaucoup de la finesse de la nomenclature de départ.

Précisément un découpage arbitrairement fin conduit à des éléments "diagonaux" variant entre 0 et ∞. Etudions les représentations optimales (au sens de §.3) associées à ces deux cas limites. Celles-ci peuvent être construites (cf. [26]) en normant les facteurs de l'analyse factorielle sur le tableau des distances suivantes :

$$d^2(j\,,j') \doteq \frac{T}{T_j} + \frac{T}{T_{j'}} - \frac{2T\,B_{jj'}}{T_j\,T_{j'}} \qquad \forall (j\,,j') \in J \times J$$

Si i et $\mathring{e}$ sont associés à un même pays, $C_{i\mathring{e}}$ note la valeur d'un élément "diagonal" courant :

$$C_{i\mathring{e}} \doteq 0 \;\Rightarrow B_{i\mathring{e}} \doteq 0 \;\Rightarrow d^2(i\,,\mathring{e}) \text{ importante}$$

$$C_{i\mathring{e}} \to \infty \quad \text{et} \quad B_{i\mathring{e}} \sim T_i \sim T_{\mathring{e}} \Rightarrow d^2(i\,,\mathring{e}) \to 0\,.$$

En pratique, ces relations peuvent fixer les premiers axes factoriels de l'analyse, du moins si T_i ou $T_{\mathring{e}}$ sont importants. Une solution tendant à réduire l'influence de ces éléments "diagonaux" semble préférable. De la décomposition classique de l'inertie cumulée des axes factoriels d'une AFC du tableau C :

$$\sum_a \lambda_a = \sum_i \sum_e \frac{1}{T_i\,T_e}\left(C_{ie} - \frac{T_i\,T_e}{T}\right)^2$$

il résulte que si les éléments "diagonaux" vérifient :

$$C_{i\mathring{e}} = T_i\,T_{\mathring{e}}/T \quad (i = \mathring{e} = 1\,,n) \tag{5}$$

ces éléments ont une contribution nulle (c'est-à-dire sont neutres). Si tel est le cas, on tend à mettre en relief les seuls échanges extérieurs sur les axes factoriels et donc à renforcer la stabilité de leur analyse (au sens de [14].

Il existe une méthode de résolution permettant de satisfaire la relation (5) pour les n éléments "diagonaux" simultanément (cf. [3]).

A la première itération on pose : $C_{i\mathring{e}}^{(1)} = 0 \quad (i = \mathring{e} = 1, n)$ $\Rightarrow$

$$T_i^{(1)} = \sum_{\substack{e \\ e \neq i}} C_{ie} \qquad T_{\mathring{e}}^{(1)} = \sum_{\substack{i \\ i \neq \mathring{e}}} C_{ie} \qquad \text{et} \qquad T^{(1)} = \sum_i T_i^{(1)}$$

Pour passer de l'itération k à l'itération k + 1, on a la recurrence :

$$C_{i\mathring{e}}^{(k+1)} = \frac{T_i^{(k)}\, T_{\mathring{e}}^{(k)}}{T^{(k)} - T_i^{(k)}\, T_{\mathring{e}}^{(k)}} \quad (i = \mathring{e} = 1, n)$$

Cet algorithme a déjà été étudié dans un cadre plus général (cf. [25]). On montre qu'il est CONVERGENT lorsque

$$T^{(k)} > T_i^{(k)}\, T_{\mathring{e}}^{(k)} - \sqrt{T_i^{(k)}\, T_{\mathring{e}}^{(k)}} \qquad \begin{array}{l} k = 1, \ldots \\ i = \mathring{e} = 1, n \end{array}$$

Ce qui est formellement satisfait. En pratique si l'on teste les valeurs de

$$| C_{i\mathring{e}}^{(k+1)} - C_{i\mathring{e}}^{(k)} |$$

par rapport à un seuil approprié, la convergence est atteinte en moins de 5 itérations.

Remarques

a) Cette méthode pourrait être généralisée pour un tableau ventilé tel que C_{ile}. En fait il est plus naturel de l'appliquer telle quelle à chaque tableau C_l avant tout calcul de B. Par ailleurs il convient de réestimer les éléments "diagonaux" chaque fois que la composition de I ou de E est modifiée.

b) D'autres méthodes d'estimation ont été testées (cf. [3]). Parmi elles, la plus séduisante à notre avis, cherche à rendre homogène la matrice d'imports-exports, en se référant aux échanges intérieurs de même "nature" que les échanges extérieurs. Cette définition, qui suscite quelques difficultés de saisie, a conduit à des représentations factorielles proches de celles obtenues ici, même si les valeurs estimées différaient sensiblement.

§.5 VUES PARTICULIERES DE LA REPRESENTATION

La représentation précédente est optimale au sens de la reconstruction des distances $\{d^2(j, j') | (j, j') \in J \times J\}$ définies au §4, du moins lorsque celles-ci sont pondérées par les réels :

$$\{T_j T_{j'} \mid (j, j') \in J \times J\}$$

Les distances $\{{}_k d^2(j, j') \mid (j, j') \in J \times J\}$ entre les projections de j et j' sur l'espace des k premiers axes factoriels de l'AFC de B rendent en effet minimum :

$$I_r = \sum_j \sum_{j'} T_j T_{j'} \, [d^2(j, j') - {}_k d^2(j, j')]$$

qui est l'inertie résiduelle du nuage $\mathcal{N}(J)$.

Il est clair que cette optimalité met particulièrement en relief le barycentre 0 des points du nuage $\mathcal{N}(J)$ respectivement affectés des masses $\{2T \, . \, T_j \mid j \in J\}$ puisque :

$$I_r = \sum_j 2TT_j \, [d^2(j, 0) - {}_k d^2(j, 0)]$$

Les vues particulières, qu'il s'agit de construire pour "voir" $\mathcal{N}(J)$ d'un de ses points j_0 (par exemple le point exportations françaises), cherchent plutôt à rendre minimum :

$$I_0 = \sum_j 2T \, . \, T_j \, [d^2(j, j_0) - {}_k d^2(j, j_0)]$$

Dans ce cas, il suffit de *rapporter la matrice d'inertie au point j_0 et non au point 0*. Cela assure une bonne cohérence entre ces facteurs "locaux" et les facteurs "globaux" usuels (cf. [24]). Mais alors que ceux-ci rendent compte de l'écart entre le tableau B et le tableau produit des marges au sens de la distance de χ^2 associée (cf. [14]), les facteurs "locaux" restituent la distance entre le tableau B et le tableau local : $B^0 = \{B^0_{jj'} \mid j \in J\}$ défini par :

$$B^0_{jj'} = T_j B_{j_0 j'} / T_{j_0} \quad \forall (j, j') \in J \times J$$

En pratique il est fréquent (cf. §6.2 par exemple) que le premier facteur "local" $j_0 l_1$ soit très proche de $j_0 0$ et que les facteurs suivants diffèrent très peu des facteurs "globaux", mais décalés d'un rang.

Remarque

Cette méthode est un cas particulier très simple de l'**analyse locale** proposée par L. LEBART (cf [21]). **L'analyse factorielle privilégiante** développée par D. LAFAYE de MICHEAUX offre d'autre possibilités d'extension au cas de plusieurs points privilégiés, en nombre fini ou infini (cf [20]).

§.6 APPLICATION AU COMMERCE INTERNATIONAL DE 1976

Cette application s'appuie sur les échanges commerciaux enregistrés en 1976 dans 53 pays (ou groupes de pays). La liste des pays, leurs codes alphabétiques et leurs volumes relatifs d'exportation font l'objet de la table 1. Les éléments "diagonaux" de cette matrice (53 x 53) ont été estimés suivant la méthode du §.4, la convergence ayant été atteinte après 4 itérations. Il s'agit ici de montrer l'intérêt et la complémentarité des approches globales et locales proposées au §.3 et §.5.

TABLE 1

CODES DES 63 (GROUPES DE) PAYS ET POIDS DES EXPORTATIONS EN %

BEL = Belgique 36
DAN = Danemark. 10
FRA = France 62
ITA = Italie. 40
IRL = Irlande 3
HOL = Pays-Bas. 43
RFA = Allemagne Ouest 117
GB = Grande Bretagne. 51
CAN = Canada. 42
USA = Etats Unis 139
JAP = Japon 72
AUT = Autriche. 9
AUS = Australie. 14
ESP = Espagne. 9
FIN = Finlance 7
GRE = Grèce. 3
NOR = Norvège 8
NZE = Nouvelle Zélande 3
SUE = Suède 20
SUI = Suisse 16
YOY = Yougoslavie 5
OCD = Autres OCDE 4
AFS = Afrique du Sud. 5
ISR = Israël. 2
ALG = Algérie 5
ARA = Arabie Saoudite. 38
EMI = Emirats 9
INO = Indonésie. 9
IRN = Iran 22
IRK = Irak 9
KOW = Koweït 9
LIB = Lybie 9
NIG = Nigéria 10
VEN = Vénézuéla. 7
PET = Autres pétroliers. 3
ARG = Argentine. 4
BRE = Brésil 10
COR = Corée. 8
NKO = Hong Kong 9
IND = Inde 6
MAL = Malaisie 6
MEX = Mexique. 3
PHI = Philippines 3
SIN = Singapour. 7
TAI = Taïwan 9
THA = Thaïlande. . . . . .
PVD = Autres pays en voie de développement 54
URS = URSS. 14
POL = Pologne. 5
ROU = Roumanie. 5
TCH = Tchécoslovaquie. 3
CHI = Chine 6
SOC = Autres pays socialistes 6

6.1. Analyse globale de la représentation

Les huits premiers axes factoriels de cette AFC classique sont caractérisés par les valeurs propres et pourcentages suivants :

Axe n°	Valeur propre	Pourcentage d'inertie	Pourcentage cumulé
1	0.3122	25.1 %	25.1 %
2	0.1315	10.6 %	35.6 %
3	0.1218	9.8 %	45.4 %
4	0.0878	7.1 %	52.4 %
5	0.0768	6.2 %	58.6 %
6	0.0683	5.5 %	64.1 %
7	0.0595	4.8 %	68.9 %
8	0.0464	3.7 %	72.6 %

Les principes de dépouillement de la méthode étant bien connus (cf [1, 6] ou [22]), on se limitera ici à souligner, en liaison avec la propriété 1 du §.3, le caractère suggestif de la représentation simultanée des importateurs et des exportateurs (codes pays suivis des signes — et + respectivement). Les figures 1 et 2 permettent de suivre les premières interprétations suivantes des axes factoriels.

Axe n°	Interprétation proposée pour le facteur associé
1	Attraction relative vers le bloc Européen ou vers le bloc Américano-Japonais.
2	Distance intercontinentale entre Amérique et Asie
3	Dépendance des pays vis-à-vis des USA
4	Dépendance des pays vis-à-vis du Japon et des pays pétroliers
5	Distance continentale entre l'Europe de l'Ouest et l'Europe de l'Est ou du Nord
6	Distance politique entre les pays de l'Est et les pays Anglo-nordiques
7	Dépendance des pays vis-à-vis de l'URSS

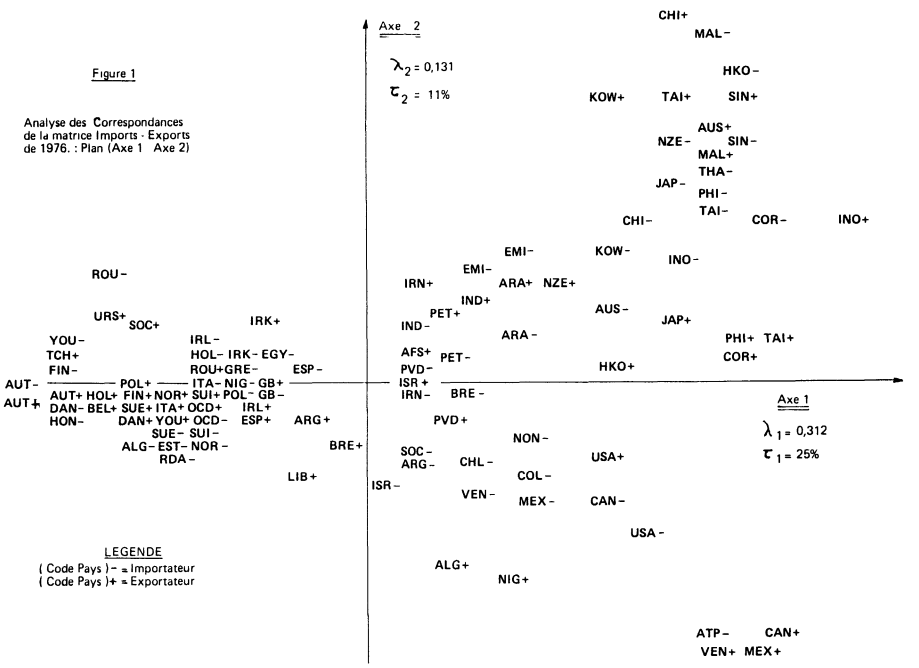

Figure 1. – Analyse des Correspondances de la matrice Imports - Exports de 1976 : Plan (Axe 1 – Axe 2).

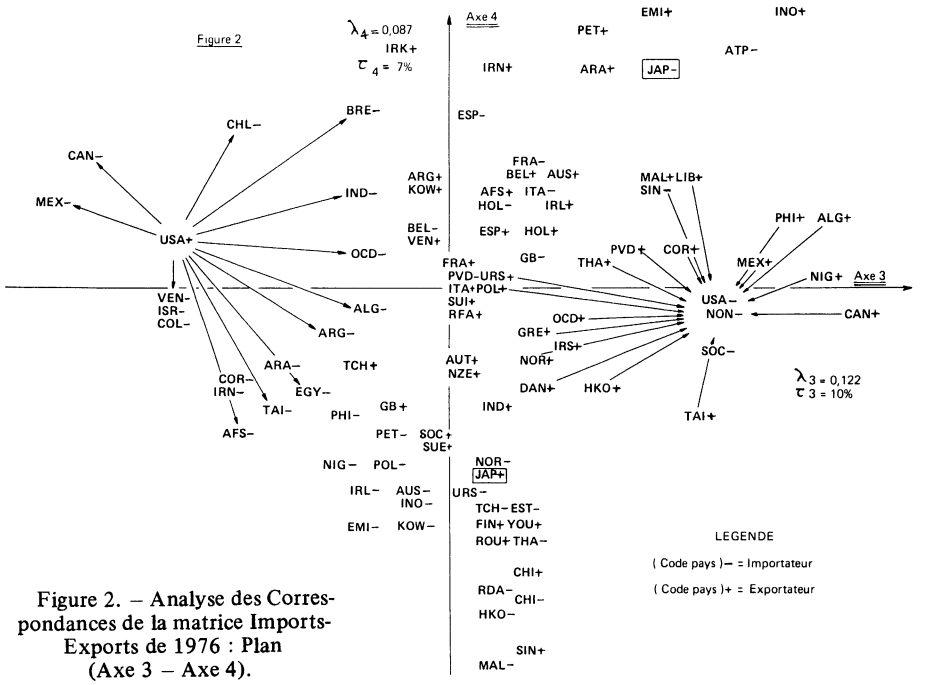

Figure 2. – Analyse des Correspondances de la matrice Imports-Exports de 1976 : Plan (Axe 3 – Axe 4).

A la lumière de ces interprétations, il apparaît que les échanges extérieurs des nations résultent de trois types de facteurs :

– des facteurs d'**opportunité** (les axes 1 et 6 par exemple), qui ont des origines diverses et multiples (économiques, politiques, historiques. . .) ;
– des facteurs de **proximité géographique** (cf les axes 2 et 5) dont les effets ne sont pas toujours prépondérants ;
– des facteurs de **dépendance** vis à vis de certains pays leader (USA, Japon, Pays Pétroliers, URSS), que les économistes (cf [13] par exemple) connaissent bien.

Pour les spécialistes de la question, ces **opportunités** et ces **dépendances** sont synonymes de faits commerciaux et de produits précis. Faute d'éléments supplémentaires à projeter ou à faire participer à l'analyse, on notera seulement que *le rang des facteurs* extraits permet d'ordonner les principales opportunités, contraintes ou difficultés que les nations rencontrent dans leurs échanges avec l'extérieur : Après s'être déterminés pour l'un ou l'autre bloc sur l'axe 1, les pays semblent s'associer en fonction de leurs proximités géopolitiques (cf l'axe 2 ou les axes 5 et 6) et, à un niveau équivalent, en raison de leurs dépendances économiques (cf les axes 3 et 4 ou l'axe 7). Ceux-ci, qu'il n'est pas possible d'étudier en détail ici, sont d'autant plus significatifs que la méthode d'estimation des valeurs "diagonales" ne favorisait en rien leur apparition. Pour répondre à des questions plus précises sur les exportations françaises par exemple (la qualité de leur représentation sur ces 7 axes n'étant que de 60 %), il convient de construire et d'analyser des vues particulières de cette représentation.

6.2. Analyse de la représentation vue de FRA+

Il s'agit d'interpréter le système d'axes locaux développés au §.5 et issu du point exportations Fançaises (FRA+).

On note que son premier facteur passe juste par le barycentre 0 des points importations et exportations. La figure 3 montre, comment les points importations sont attirés par le point FRA+, effet particulièrement sensible sur l'axe 1′ (24 % d'inertie). Cet axe permet d'apprécier les *spécificités des marchés extérieurs de la France* en 1976. Les importateurs sont en effet d'autant plus proches de FRA+ qu'ils achètent plus systématiquement "français". De plus il ne s'agit pas de situations concurrentielles, les autres exportateurs étant sur cet axe éloignés de FRA +. L'analyse détaillée de la projection des divers pays sur cette dimension conduit aux deux observations suivantes :

• Les marchés extérieurs de la France concernent des pays qui apparaissent proches, soit historiquement (Algérie, pays en voie de développement, Pologne. . .), soit géographiquement (Belgique, Suisse, Italie, RFA, Espagne), soit par nécessité (pays pétroliers. . . etc).

• A quelques exceptions près (la Chine par exemple), la France ne semble pas rechercher de partenaires lointains offrant une complémentarité économique ou des débouchés commerciaux, comme le font des pays tels que le Japon ou les USA (cf fig. 2).

Les axes principaux suivants ressemblent beaucoup à ceux de l'analyse globale du §.6.1 (mais de rang r − 1), comme on peut le voir en comparant l'axe 4′ de la figure 3 à l'axe 3 de la figure 2. Leurs interprétations sont donc faciles à imaginer. Ces axes permettent de préciser comment se situe la Fance par rapport

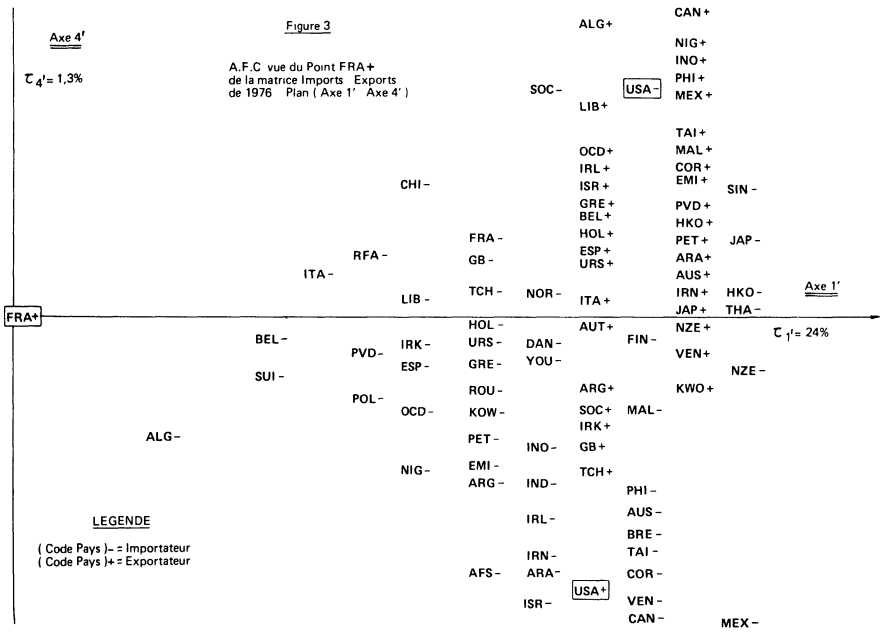

Figure 3. – A.F.C. vue du Point FRA+ de la matrice imports-Exports de 1976 : Plan (Axe 1′ – Axe 4′)

aux opportunités, aux difficultés ou aux contraintes générales d'échanges. Ce n'est pas le lieu de développer ici les réflexions et les stratégies commerciales que cette représentation suggèrent aux spécialistes du commerce extérieur.

Remarque

Géométriquement, la ressemblance entre facteurs locaux et facteurs globaux signifie seulement que l'axe 1′ est quasi-orthogonal aux axes 1,2,... etc. Mais la situation serait sans doute très différente pour des axes vus de USA+ ou de JAP+.

§.7 CONCLUSION

Pour l'économiste, les analyses précédentes mériteraient d'être détaillées par produits (cf propriété 2), car elles semblent révélatrices des lois et des possibilités des marchés extérieurs.

Pour le statisticien, cette application montre l'intérêt et l'efficiacité de la méthodologie mise au point. Ses développements formels peuvent être aisément étendus à d'autres données. Ainsi l'estimation itérative des termes "diagonaux" s'applique encore à toute matrice d'échanges ou de flux. L'optimalité graphique de l'AFC concerne le traitement de toute table de contingence, qu'elle soit binaire ou multiple. Elle indique que les représentations simultanées associées visualisent au mieux les intensités relationnelles impliquées par de telles données. Cette optimalité peut en fait être adaptée à des besoins de description très variés, soit par pondération des distances de la représentation (cf §.5, [21] et [24]), soit par référence à d'autres critères (cf [7, 8, 20] et [29]).

Les distances factorielles, qui en résultent, peuvent conduire à d'intéressantes classifications hiérarchiques (cf [5]) utiles pour affiner certaines interprétations (cf [1, 6, 18] et [19]).

BIBLIOGRAPHIE

[1] J.P. BENZECRI et COLL. (1973). – *L'Analyse des données* (2 tomes) Dunod, Paris.

[2] J.P. BENZECRI (1977). – Sur l'analyse des Tableaux binaires associés à une correspondance multiple. *Cahiers de l'Analyse des données*, vol. II, n° 1.

[3] B. BURTSCHY (1981). – *Analyse des matrices d'échanges*, thèse de 3e cycle ; Université P. et M. Curie, Paris.

[4] B. BURTSCHY et C. NORA (1978). – *The problem of missing data, using the properties of factor analysis of correspondence*, Proceedings of 4th symposium on Pattern recognition, p. 276-279 Kyoto.

[5] B. BURTSCHY et Y. LE FOLL (1981). – *Graphical representation of input-output matrice*, Proceedings 14 st Annual meeting of the American Statistical Association, Business and Economic Section, Detroit.

[6] F. CAILLEZ et J.P. PAGES (1976). – *Introduction à l'analyse des données*, SMASH, Paris.

[7] P. CAZES (1977). – *Traitement de variables qualitatives par l'analyse des correspondances et l'analyse canonique*, Séminaires IRIA. Classification Automatique et perception par ordinateur. p. 41-61.

[8] P. CAZES (1981). – L'analyse de certains tableaux rectangulaires décomposés en blocs : Généralisation des propriétés rencontrées dans l'étude des correspondances multiples *Cahiers de l'Analyse des données*, vol. V, n° 2 et 4 et vol. VI, n° 1 et 2. Dunod, Paris.

[9] H. CAUSSINUS (1976). – Quelques points de vue sur l'analyse des tableaux d'échanges. *Annales de l'INSEE,* n° 22-23, p. 83-100.

[10] D. COLLOMBIER (1980). – *Recherche sur l'analyse des tables de contingence*, thèse d'Etat, Université Paul Sabatier, Toulouse.

[11] A.G. CONSTANTINE et J.C. GOWER (1978). – Graphical representation of asymetric matrices, *Applied Statistics*, vol. 27, 3, p. 297-304.

[12] M. DELARCHE (1978). – Problèmes liés à la hiérarchisation et au dessin automatique de graphes, *Rapport de recherche IMAG,* n° 151, Grenoble.

[13] P.H. DERYCKE et B. BOBE (1975). – Projection des échanges extérieurs et balances des paiements, *Economica.*

[14] D. DOMENGES et M. VOLLE (1979). – Analyse factorielle sphérique : une exploration. *Annales de l'INSEE*, n° 35, p. 3-83.

[15] B. ESCOFIER (1979). – Stabilité et approximation en analyse factorielle, thèse d'Etat, Université P. et M. Curie, Paris.

[16] Y. ESCOUFIER et A. GRORUD (1979). – *Analyse factorielle des matrices carrées non symétriques*, in : Data Analysis and Informatics (E. Diday *et al.* Ed.) p. 263-276, North Holland, Amsterdam.

[17] T. FOUCART (1981). – Suites de tableaux et de sous-tableaux, *Revue de Statistique Appliquée,* n° 2, p. 31-42.

[18] E.W. HOLLMAN (1972). – The relation between hierarchical and euclidean models for psycological distances, *Psychometrika*, vol. 37, n° 4, p. 472-486.

[19] M. JAMBU (1976). – Sur l'interprétation mutuelle d'une classification hiérarchique et d'une analyse des correspondances, *Revue de Statistique Appliquée*, n° 2, p. 45-73.

[20] LAFAYE DE MICHEAUX (1978). – *Approximation d'analyses canoniques non linéaires de variables aléatoires et analyses factorielles privilégiantes,* thèse de Docteur-Ingénieur, Université de Nice.

[21] L. LEBART et N. TABARD (1973). – *Recherche sur la description automatique des données socio-économiques,* Rapport CORDES-CREDOC, Paris.

[22] L. LEBART, A. MORINEAU et N. TABARD (1977). – *Technique de la description statistique*, Dunod, Paris.

[23] Y. LE FOLL (1972). – *Analyse des échanges économiques internationaux entre 1958 et 1968*, Actes du colloque de statistique et de géographie, p. 129-152, Besançon.

[24] Y. LE FOLL (1981). – Pondération des distances en analyse factorielle, *statistique et Analyse des Données*, vol. 7, n° 1, p. 13-31.

[25] C. NORA-CHOUTEAU (1975). – *Une méthode de reconstitution et d'analyse de données incomplètes,* thèse d'Etat, Université P. et M. Curie, Paris.

[26] J.P. ORFEUIL (1974). – Structure métrique des espaces probabilisés : application à l'étude des contingences multiples, *cahier du Centre de Morphologie Mathématique de l'Ecole des Mines de Paris.*

[27] C.R. RAO (1974). – The use and interpretation of principal component analysis in applied research, *Sankhia,* Ser. A, 26, p. 329-355.

[28] B. ROY (1969). – *Algèbre moderne et théorie des graphes*, Dunod, Paris.

[29] G. SAPORTA (1980). – *About some remarquable properties of generalized canonical analysis*, European Meeting of the Psychometric Soviety, Groningen.

[30] E. STEMMELEN (1977). – "Tableaux d'échanges : description et prévision", *Cahier du B.U.R.O.*, n° 28, Paris.