

VIVE LA GÉOGRAPHIE RÉGIONALE !

DÉFENSE ET ILLUSTRATION D'UNE DISCIPLINE-MÈRE RECADRÉE PAR LES MÉTHODES MODERNES DE L'ANALYSE SYNTHÉTIQUE

Jean-Max PALIERNE

Conseil Scientifique de la Région
des Pays de la Loire

"Sur des pensées nouveaux faisons des vers antiques"
A. CHÉNIER

Résumé : Dans ce texte, l'auteur montre que, pour rester la science fondamentale qu'elle fut, la Géographie Régionale doit recourir aujourd'hui aux méthodes modernes de l'Analyse Factorielle (assistée par ordinateur). En prenant pour exemple les États-Unis d'Amérique (et la France à un moindre titre), l'A. expose sa manière personnelle d'exploiter les traitements d'une mathématique efficace, féconde et non répulsive.

Abstract : In this paper, the author shows that, in order to stay the fundamental science it used to be, regional geography must nowadays use modern methods of Factorial Analysis (computer assisted) Taking for example the United States of America (and France, in lesser degree), the A. displays his personal way of using the process of efficient, prosperous and non repulsive mathematics.

Mots-clés : Géosstructures - Composantes Principales - Polygones d'inertie - Transfacteur prédictif

Key-words : Geostructures - Principal Components - Polygon of inertia - Predictive transfactor

Grâces soient rendues à ceux qui ont toujours défendu la Géographie Régionale, car elle seule est en mesure d'assurer aux activités géographiques une place singulière dans la sphère des activités scientifiques majeures. Elle seule en effet, contrairement aux "spécialités" associées (irremplaçables néanmoins), relève des disciplines-mères, c'est-à-dire celles-là mêmes auxquelles appartiennent en propre un objet d'étude global et finalisé, un champ d'investigation spécifique et un ensemble de modes opératoires originaux. Plus l'aventure humaine se construit, plus elle est porteuse des nécessités d'un aménagement raisonné de l'espace -de la simple "niche" locale à la planète tout entière- y compris et surtout dans ses perspectives environnementales, desquelles dépend l'avenir de ladite humanité. Une telle "construction" de l'aire de vie (écoumène ou non) implique nécessairement une **approche synthétique** du Réel ; des infrastructures morphogéologiques aux superstructures géopolitiques. À cette diversité des connaissances, dimension historique incluse, il faut une discipline pratiquant une authentique analyse de la globalité. Aujourd'hui, pour tendre à ce but, il suffit de réadapter une méthodologie "régionale" qui a fait ses preuves, en l'appuyant sur la logistique moderne d'investigation et de traitement de l'information. Ce texte, qui n'est ni une pétition de principe ni une profession de foi, espère en apporter une démonstration qui devrait convenir aux géothéoriciens, chercheurs et enseignants, comme aux géopraticiens, prévisionnistes et aménageurs.

I - CHANGER POUR MAINTENIR : L'ANALYSE SYNTHÉTIQUE MODERNE

1. Un paradoxe fondateur

Analyse synthétique ... On croirait avoir affaire ici à une antinomie ou à un oxymoron ; si ce n'est à une aporie. En fait, cette bizarrerie tient à la nature même de la géographie (régionale notamment), laquelle, en prétendant traiter par la synthèse ce que d'autres disciplines avaient toujours confié à l'analyse, n'a cessé, depuis qu'elle est apparue en tant qu'activité définie, de soulever des querelles épistémologiques, souvent orchestrées d'ailleurs par l'inflation sémantique. Ainsi en va-t-il du "fait géographique" : est-il scientifique ou non ? De la "région" : est-elle une réalité effective ou une entité d'"effectuation" dont les différentes "branches" (physique, économique, humaine, etc.), en fondant leurs particularismes ensemble, peuvent rendre compte de façon satisfaisante ? Et ainsi de suite... Car cette manière-là d'approcher le Réel intrigue nécessairement : quel est donc cet "espace" dont la nature, l'échelle et les formes varient à volonté, et qui sert de champ, à la fois, à la recherche et à l'"application" ? Et surtout, comment justifier l'accumulation "encyclopédique" de faits que les critiques réputent hétéroclites, et dont la seule compatibilité relèverait, selon les mêmes, de la compilation complaisante (sinon de l'alibi pseudoscientifique), plutôt que d'une synthèse parfaite... "pointue" et "performante" ?

Certes, il y a de quoi s'interroger à propos de l'amalgame des données hétérogènes et hétérométriques que la "Régionale" se propose d'utiliser de manière concomitante et congruente. Qu'y-a-t-il en effet de commun et, plus encore, de commensurable, entre un degré centigrade, une strate géologique, un taux de boisement, un indice de productivité industrielle, un flux d'importation et un vote conservateur, pour ne prendre que quelques unes des centaines de variables, quantitatives ou qualitatives, que l'étude régionale peut envisager de regrouper en une seule et même étude, afin d'en faire, à travers leurs relations synergiques ou leurs rapports conflictuels, les éléments qualificatifs et constitutifs de ses fondements, de ses structures et de ses réseaux ? Et pourtant... Pourtant, à l'époque de sa splendeur, la *Géographie Universelle*, production éminente de cette fameuse école française qui pointait les "caps" et traçait les routes à suivre au sein de la communauté internationale, la vieille "G.U." des Baulig, Demangeon, de Martonne et consorts, dressait, à travers le monde entier, des "portraits" territoriaux si ressemblants, qu'actuellement encore les Américains, pour mieux comprendre leur pays kaléidoscopique, doivent plus à ces vénérables découpages qu'aux modèles, tout ensemble, sommaires et tarabiscotés des *Nine Census divisions of Mainland* fédéraux (suivre -Planche 3, "Géostructures des individus", plan A/C- les traces des "régions" *Mountain*, en rouge, et *Pacific*, en bleu).

Toutefois, il est avéré que ces études du passé, pour brillantes et convaincantes qu'elles furent, ne sauraient être reprises telles quelles aujourd'hui. L'"intuition géographique" (le "flair"...selon certains!), le temps disponible, les "techniques" empiriques et manuelles, tout cela est irrémédiablement caduc. Les informations contemporaines et leur "problématique" sont trop nouvelles, trop imprévisibles, trop abondantes, et trop pesantes aussi, pour se satisfaire d'approximations. Il faut impérativement (et pas seulement en Régionale...!) s'adresser à l'arsenal moderne des moyens d'investigation et de traitement, lesquels savent rendre pertinente l'analyse synthétique, en redonnant tout son poids et tout son prix à une activité dont le caractère intrinsèque (transdisciplinarité) et les principes méthodologiques (globalisation) sont, à coup sûr, plus que jamais d'actualité.

2. Les nombres et les mots

Qu'ici, les "profanes" ou les novices ne s'émeuvent pas trop devant le vocabulaire des nouvelles techniques et ce qu'il recouvre : tout individu de bonne volonté, fût-il dépourvu de connaissances mathématiques, peut, en moins de cinquante heures, acquérir la maîtrise des méthodes qui vont de la fruste *Variance* à la très fine *Analyse Canonique Discriminante*, en passant par tout mode ou échelon de l'exploitation raisonnée de ces nombres que l'on traduit ensuite en mots d'usage courant. Dix ans de pratique avec les étudiants de Licence et de Maîtrise me l'ont

surabondamment montré. Au demeurant, les moyens actuels d'utilisation des données sont nombreux et d'accès facile ; et ils possèdent une simplicité de maniement et une flexibilité d'exploitation qui les rend aussi solides que subtils et dociles. Pour l'essentiel, ils puisent à quatre sources que l'on peut regrouper deux à deux : d'une part, et pour ce qui concerne la base "intellectuelle" du travail, il y a la **STATISTIQUE** (de l'unidimensionnelle à la pluridimensionnelle), et le **CALCUL MATRICIEL**, tiré de l'algèbre linéaire (qui s'exprime dans le multidimensionnel des hyperespaces); d'autre part, et pour ce qui concerne le travail pratique, il y a des **LOGICIELS**, de plus en plus "conviviaux", et des **ORDINATEURS** de plus en plus puissants ("Micros" notamment). Dans ces conditions, exploiter les données, qu'elles soient quantitatives ou qualitatives, procure les plus grandes satisfactions, d'autant, on va le voir, qu'en s'appuyant sur les propriétés remarquables de la mathématique des *Ensembles*, la géographie dispose d'un corps de protocoles opératoires incontestables, de très haute tenue et aptes à traiter, par combinaison, des données de toute nature, de tout ordre et de toute grandeur.

Débarassé des itérations hasardeuses et des collationnements fastidieux, le géographe a la possibilité désormais de donner sa pleine mesure ; car ce ne sont ni le mathématicien ni l'informaticien qui décideront du choix des phénomènes à faire traiter par les machines, non plus que des conclusions à tirer des traitements : cela, seule la **longue formation géographique complète** sera en mesure d'y préparer. Et, comme la géographie régionale a décidément beaucoup de chance, elle est toute désignée pour utiliser à plein cette propriété fondamentale de la méthodologie ensembliste, selon laquelle, deux ensembles étant combinés, *tout élément de l'un dépend à tout moment de tous les éléments de l'autre*. Outre que cela définit et régit les analyses dites factorielles, cela correspond exactement aussi à l'activité géographique qui traite toujours, simultanément et conjointement, de deux ensembles : celui des **INDIVIDUS** (forêt, bassin industriel, réseau hydrographique, massif montagneux, ville, port, commune rurale, département, continent, et toute autre unité qui plaira); et celui des **DESCRIPTEURS** (surface, densité démographique, hauteur de marnage, indice de productivité, suffrages exprimés, et tout autre paramètre ou variable quantitative ou qualitative que l'on voudra). Mais, comme "l'intelligence" artificielle n'est qu'une naïve ou vaniteuse métaphore langagière, comme les hypothèses mathématiques sont très strictes, et comme les machines programmées sont capables de traiter n'importe quoi, il ne faut pas s'attendre que les ordinateurs fassent des miracles : ils calculent, un point c'est tout. Pour vérifier tout cela, il n'est que de se reporter à un exemple pris en vraie grandeur, sinon en "temps réel".

II - UN EXEMPLE REMARQUABLE : LES ÉTAT-UNIS DE LA DÉCENNIE 60

1. Problématique, protocoles et procédure d'étude

Remarquable, cet exemple l'est d'abord parce qu'il intéresse un énorme ensemble régional, complexe autant que composite. Il l'est aussi parce que la période considérée a été un moment crucial dans l'évolution de cet ensemble : politique de la *Nouvelle Frontière*, de la *Grande Société*, *Droits Civiques*, *Black Power*, assassinats retentissants (J.F. Kennedy, M.L. King), etc. Remarquable, il l'est encore parce que, y ayant facilement accès, tous et chacun peuvent, *a posteriori*, contrôler l'efficacité de son traitement, et éprouver la validité des conclusions le concernant. Enfin, en le confrontant à une étude similaire de deux géographes anglo-saxons (Cole et King, *Quantitative Geography*, Wiley éd., pp. 287-304), on pourra comparer les techniques respectives d'exploitation des résultats (dimensions prédictives comprises), la nôtre étant marquée par des constructions particulières, et définies par un vocabulaire spécifique (mots en *italiques grasses* avec "guillemets", ci-après, à leur première mention). Pour d'évidentes raisons de place, on n'entrera point dans le détail des calculs matriciels ; on se bornera à mentionner, seulement et sommairement, l'orientation du traitement des éléments numériques.

Avant tout, il y a les données, regroupées en **DEUX ENSEMBLES** : celui des *individus* constitué, ici, par les cinquante états "américains", rangés en lignes, par ordre alphabétique, et notés de 01 (Alabama) à 50 (Wyoming); et celui des *descripteurs* qui compte vingt-cinq variables (certaines ayant été transformées en paramètres), classées de A à Y, et rangées en colonnes (v.

Pl.1). Il est nécessaire, en effet, de prendre deux individus (au moins) pour un descripteur. Les "bordereaux de perforation" (dans lesquels sont reportées les données à "saisir" sur bande magnétique) comportant un nombre limité de colonnes (80 en l'occurrence), on attribue celles-ci -par 1,2,3 ou 4- aux descripteurs selon leur "encombrement" (voir ci-après). On veille à éliminer les décimales (grâce à un multiplicateur et à l'arrondi) et les valeurs négatives (en changeant d'origine, e.g. : pour une variable "étendue" de -20 à 60, on compte de 1 à 81); aussi bien faut-il retenir le moins possible de valeurs nulles, et ne pas abuser des pourcentages (idéalement moins d'1 sur 2 données); trop nombreux, en effet, ils biaisent l'analyse, en créant des sous-espaces parasites, et risquent de scotomiser des phénomènes majeurs (v. Pl.4). Dans le cas présent, les colonnes 1 et 2 étant réservées aux états-individus, les descripteurs commencent à la colonnes 3 ; dès lors, le descripteur A (altitude) occupe les colonnes 3,4,5 ; B (températures) est en 6,7,8 ; H (solde naturel de la population), en 24,25 ; I (migrations internes), en 26,27,28,29 ; etc. A titre indicatif, voici le "*profil numérique*" de l'Arkansas (O4), de A à Y :

04-58-167-125-13-300-210-235-19-29-165-45-22-20-288-1760-90-49-300-330-245-1200-95-431-566-430

Pour les 25 descripteurs, il y a ainsi 1250 nombres (soit 3400 chiffres), relevés selon 14 modes métrologiques différents (v.Pl.1). Les algorithmes vont les DÉCOMPOSER (Analyse), et les RECOMPOSER (Synthèse) en 25 "facteurs", de telle manière que les descripteurs participent plus ou moins intensément à chaque facteur. Le facteur, dès lors, est une sorte de nouveau descripteur densifié et simplifié. D'où il résulte que certains facteurs sont très riches de sens (les tout premiers), alors que d'autres sont extrêmement pauvres ou sans signification (les derniers des quatre cinquièmes restants -v. ci-après en fin de II,1). À partir de là, la question est de savoir si les nombres choisis, peuvent, à eux seuls, décrire la situation régionale des États-Unis, extraire les problèmes qu'elle pose, et "sortir" un pronostic pour la période suivante (décennale ou non). Par parenthèse, le lecteur notera le caractère rudimentaire des données initiales: par exemple -pour une amplitude de surface en km. carrés (S) de 1 à 220- un seul nombre définit l'altitude du Rhode Island (S= env. 3000), aussi bien que celle du Texas (S= env. 700 000).

Voilà pourquoi l'analyse est faite en A.C.P. (Analyse en Composantes Principales). C'est la plus robuste de toutes les analyses factorielles dont elle est la "mère" ; et quand on dispose des meilleurs logiciels, il faut l'utiliser en priorité absolue. L'un d'entre eux, *Principal Components*, dû à S.A.S. (*Statistical Analysis System*), est très puissant et des plus solides. C'est lui qui a été mis en oeuvre dans la présente étude. Les MILLIERS DE CALCULS produits (moments centrés jusqu'à l'ordre 4, corrélations, valeurs propres, inerties de variance, etc.) n'ont demandé que TRENTE DEUX CENTIÈMES DE SECONDE à un gros ordinateur (Orsay), alors que les seules corrélations interdescripteurs (trois cents résultats), par calculatrice programmée, demandent TRENTE HEURES effectives de calcul intensif et continu (vérifications non comprises). Quant à l'analyse complète, à supposer que l'on sache la conduire (expansions de Laplace), et à supposer que l'on résiste à l'épuisement physique et nerveux qu'elle entraînerait, il y faudrait des centaines d'heures (au moins)...au regard de cette fraction de seconde.

Comme le but de l'A.C.P. est de mettre en rapport toutes les données entre elles, le problème se pose de les rendre compatibles. Car, même en système ensembliste, on ne peut multiplier, à l'état brut, des dollars par des kilomètres carrés, des salles de bains par des infirmières, ou des bulletins de vote par des tonnes de céréales...et remultiplier le tout par des Peaux-Rouges ou des wagons de chemin de fer. C'est pourquoi l'on procède à une sorte de "*neutralisation métrique*", dite réduction en termes de Statistique (ou encore, selon les auteurs, "dédimensionnement" ou "standardisation"), et qui consiste à ajuster tous les nombres à une même unité de compte, soit le quotient de l'écart à la moyenne par l'écart-type. Par exemple, en donnée initiale, la moyenne altimétrique de l'Arkansas (04A) étant de 58, et celle des états de 158, pour un écart-type de 162, la réduction fait venir -0.617, soit : $(58-158)/162$. Voici encore quelques données réduites, pour montrer la faible amplitude du système, dans tous ses cas de figures:

32A=-0.419; 04B=+1.6; 32B=-0.53; 04Q=-1.6; 32Q=+1.77; 04C=+1.02; 32C=+0.25.

Au terme des traitements, tout est réordonné sur des "axes-facteurs" qui condensent l'information en valeurs fluctuant de -1.00 à +1.00. Il y a donc, comme dit plus haut, autant d'axes-facteurs que de descripteurs, soit, ici, 25 ; mais, du fait de la condensation des informations, seuls les premiers sont significatifs (exploitables). L'analyse numérique est, par conséquent, une *simplification* et un *éclaircissement* : précisément, tout ce que l'on attend d'une *synthèse*. La masse des "influences", qu'exercent en effet les données les unes sur les autres, est une sorte de grande variation globale et solidaire qui est dite, pour cette raison, variance générale (inertie). Puisque les données ont été "travaillées" en "espace réduit" (*i.e.* corrélations), et les vecteurs normés à 1, la somme de la variance est égale à 25 (ou 100%). Cela explique que, en vue de l'interprétation, seuls sont retenus les facteurs rendant compte ou "saturant" d'au moins 1/25 la variance générale (soit 4%). Dans l'analyse en cours, seuls les cinq (5) premiers facteurs sont utilisables : ils saturent la variance à 80% (soit 20/25 -v.Pl.1) ; les vingt autres ne représentant à eux tous que 5/25 de variance, soit 20% ; le facteur XIV ne valant que 1% ; le XX: 0,1 ; le XXIV et le XXV: 0.

Le travail du "comptable" achevé, celui du géographe -qui avait débuté par la collecte et la sélection des données- reprend avec la lecture des résultats qui doivent être interprétés dans les deux "espaces" analytiques, à commencer par celui des caractères.

2. Les "géstructures" des descripteurs : conservatismes et novations

Pour exploiter les résultats, on étudie les plans factoriels (euclidiens, orthonormés par des axes centraux de +1.00 à -1.00, v. Pl.2 et 3), obtenus par combinaison de deux facteurs, tels que I et II, I et III, jusqu'à IV et V, ce qui donne dix (10) plans pour les cinq facteurs retenus, soit : ((5x5)-5/2). Contrairement à ce que l'on affirme trop souvent, il n'y a pas que le premier plan qui ait de l'intérêt ; l'exploitation, toutefois, dépasse rarement cinq à sept plans, non pris, d'ailleurs, dans l'ordre "hiérarchique" que donneraient les cumulations successives de variance expliquée. Ainsi, dans notre cas d'espèce, le plan I-IV, qui explique, ou sature, 43% de la variance générale, est sans intérêt, alors que IV-V, qui n'en sature que 10, est fort intéressant, notamment en ce qu'il dévoile la nature du facteur cinq (V). Il en va de même pour les plans formés par les individus (moins complexes cependant). Dans la mise en plans, les facteurs ayant déjà absorbé les "déformations" (vectorielles), on ne retiendra pas l'objection d'illusion d'optique (de Bourouche et Saporta, *Que Sais-je?* N°1854, p.40) due censément à l'hyperespace, et qui voudrait que l'on traitât restrictivement les projections de points.

À partir du tableau des coefficients de corrélation entre descripteurs et axes-facteurs (Pl.1), on projette, dans les plans, les points donnés par ces coefficients qui leur servent de coordonnées (Pl. 2 et 3). D'ordinaire, on cherche à regrouper ces points en "nuages" que l'on essaie d'interpréter. La méthode présentée ici est assez différente, qui cherche à définir des structures d'interrelation ("géstructures"). Partant du fait que les descripteurs ont un lien commun (celui de l'analyse) et des liens particuliers (ceux de leur appartenance à tel secteur propre : géographie physique, démographie, etc.), je forme ce que j'appelle des "polygones d'inertie", en commençant par le "polygone global" (ou polygone de variation générale), par jonction des points les plus éloignés du centre de gravité où se trouve une zone de très faible corrélation, que j'appelle "puits d'inactivation". À l'intérieur de ce polygone, je dresse les "polygones catégoriels" (jusqu'aux triangles), réunissant les points des communautés ou des affinités d'appartenance. Tous ces polygones définissent des "aires significatives", qui jouent un rôle important dans les structures. De manière plus étroite, on peut distinguer encore, tantôt des "synergies" (conjonctions), qui vont par paires, triplettes ou plurinucléaires, lesquels forment des "pôles directeurs" lorsqu'ils sont en "tête" d'axe (proches de +1) ; tantôt des "isolats" (un seul point), liés à des situations de particularisme, de faible dynamisme ou de crise. À quoi s'ajoutent des "antagonismes" (oppositions), et des "neutres" (quadratures). Pour faciliter le repérage des points, de +1 à -1, je nomme, métaphoriquement, Nord-Sud (N.S), l'axe vertical, et Est-Ouest (E.W), l'horizontal. Les "quadrants" sont alors désignés N.E (++) , S.W (--), N.W (+-), S.E (-+). L'expérience montre que le quadrant N.E est porteur de points dynamiques, anticipant l'évolution de l'ensemble analysé : j'y place ce que j'appelle un "transfacteur" (prédictif).

PLANCHE 1

LES ÉTATS-UNIS D'AMÉRIQUE
(DÉCENNIE 60)

NOMENCLATURE ET MÉTROLOGIE DES DESCRIPTEURS

A. Altitude moyenne (mètres x 10)- B. Températures moy. an. (degrés centigrades x 10)- C. Précipitations tot. an. (cm)- D. Densité démograph. (personnes/km carré)- E. Pop. âge moyen (35-55 ans °/∞ pop. tot.)- F. Taux natal. (°/∞∞ pop. tot.)- G. Mort. inf. (°/∞∞ pop. tot.)- H. Sold démogr. natur. (°/∞ pop. tot.)- I. Solde migr. inter. (°/∞ pop. tot. +220)- J. Pop. rur. (°/∞ pop. tot.)- K. Pop. urb. (°/∞ pop. tot.)- L. Pop. noire (°/∞ pop. tot.)- M. Peaux-Rouges (°/∞ pop. non blanche)- N. Travailleurs indus. lourdes (°/∞ pop. tot.)- O. Revenu moyen (\$ p. pers.)- P. Durée de la scol. (mois)- Q. Confort résident. (°/∞ des habit.)- R. Téléphones (°/∞ pers.)- S. Voitures partic. (°/∞ pers.)- T. Budget fédéral (\$)- U. Cabinets jurid. (°/∞∞∞ pers)- V. Médecins (°/∞∞∞ pers.)- W. Suffrages pr. Nixon (nbre)- X. Suffrages pr. Johnson (nbre)- Y. Suffrages gouv. répub. (nbre.)

La variance générale vaut 25, soit 100%

La variance expliquée vaut 20, soit 80% répartis comme suit entre les cinq premiers facteurs
I= 37- II= 20- III= 13- IV= 6- V= 4

Dans les tableaux ci-dessous, le zéro a été supprimé, les valeurs ont été arrondies à deux chiffres, les négatives étant soulignées

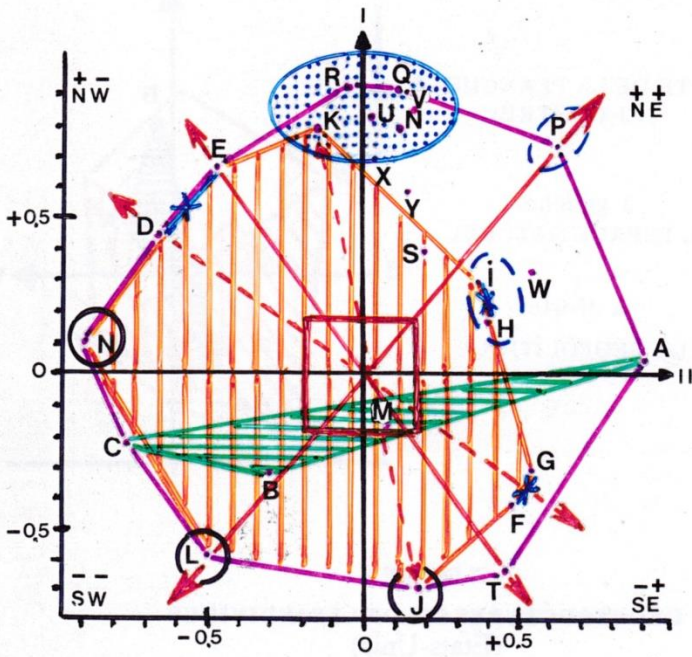
CORRÉLATIONS DESCRIPTEURS-FACTEURS

	I	II	III	IV	V
A	06	92	05	16	14
B	<u>32</u>	<u>30</u>	62	52	05
C	<u>22</u>	<u>77</u>	25	<u>35</u>	<u>16</u>
D	45	<u>65</u>	10	<u>36</u>	<u>05</u>
E	66	<u>50</u>	<u>40</u>	02	<u>05</u>
F	<u>40</u>	50	57	<u>45</u>	<u>00</u>
G	<u>30</u>	55	09	<u>20</u>	<u>25</u>
H	17	40	73	20	<u>05</u>
I	30	37	75	17	<u>20</u>
J	<u>70</u>	17	<u>70</u>	30	28
K	80	<u>15</u>	50	24	15
L	<u>58</u>	<u>51</u>	50	25	06
M	<u>15</u>	05	22	31	01
N	09	<u>91</u>	<u>20</u>	02	<u>30</u>
O	81	11	14	<u>15</u>	<u>02</u>
P	74	64	15	<u>15</u>	02
Q	93	12	08	09	03
R	92	<u>06</u>	05	<u>04</u>	02
S	40	20	<u>40</u>	60	00
T	<u>63</u>	47	16	<u>30</u>	06
U	81	02	08	05	05
V	81	15	10	05	00
W	28	50	<u>50</u>	30	<u>50</u>
X	70	05	<u>18</u>	<u>50</u>	<u>40</u>
Y	52	15	<u>40</u>	<u>22</u>	25

CORRÉLATIONS INDIVIDUS-FACTEURS

		A	B	C			A	B	C		
01	Alabama	<u>80</u>	<u>30</u>	24	Al.	26	Montana	05	46	<u>19</u>	Ma.
02	Alaska	<u>29</u>	58	34	Ak.	27	Nebraska	05	25	<u>27</u>	Nk.
03	Arizona	12	37	44	Az.	28	Nevada	40	79	60	Nv.
04	Arkansas	<u>75</u>	<u>20</u>	<u>10</u>	As.	29	New Hampshire	15	<u>05</u>	<u>10</u>	Nh.
05	California	65	19	28	Ca.	30	New Jersey	60	<u>41</u>	15	Nj.
06	Colorado	38	57	08	Co.	31	New Mexico	<u>18</u>	60	27	Nm.
07	Connecticut	66	<u>29</u>	05	Ct.	32	New York	75	<u>38</u>	05	Ny.
08	Delaware	<u>33</u>	20	23	Del.	33	North Carolina	<u>65</u>	<u>25</u>	08	Nc.
09	Florida	17	<u>20</u>	48	Fa.	34	North Dakota	<u>38</u>	<u>38</u>	<u>45</u>	Nd.
10	Georgia	<u>61</u>	<u>26</u>	35	Ga.	35	Ohio	30	<u>15</u>	<u>09</u>	Oh.
11	Hawai	11	22	50	Hw.	36	Oklahoma	05	12	<u>18</u>	Ok.
12	Idaho	<u>11</u>	51	<u>28</u>	Id.	37	Oregon	35	15	<u>19</u>	Or.
13	Illinois	40	<u>15</u>	<u>08</u>	Il.	38	Pennsylvania	35	<u>35</u>	<u>28</u>	Pn.
14	Indiana	05	<u>15</u>	<u>02</u>	In.	39	Rode Island	42	<u>41</u>	06	Ri.
15	Iowa	10	<u>02</u>	<u>35</u>	Io.	40	South Carolina	<u>86</u>	<u>30</u>	20	Sc.
16	Kansas	20	12	<u>29</u>	Ks.	41	South Dakota	<u>40</u>	32	<u>42</u>	Sd.
17	Kentucky	<u>45</u>	<u>04</u>	<u>27</u>	Ky.	42	Tennessee	<u>47</u>	<u>14</u>	<u>12</u>	Tn.
18	Louisiana	<u>53</u>	<u>38</u>	40	La.	43	Texas	<u>20</u>	03	25	Tx.
19	Maine	<u>06</u>	<u>15</u>	<u>20</u>	Mn.	44	Utah	18	53	16	Ut.
20	Maryland	40	<u>27</u>	21	My.	45	Vermont	<u>05</u>	<u>05</u>	<u>28</u>	Vt.
21	Massachussets	63	<u>36</u>	10	Ms.	46	Virginia	<u>31</u>	<u>09</u>	15	Va.
22	Michigan	25	<u>10</u>	<u>04</u>	Mg.	47	Washington	40	14	<u>12</u>	Wn.
23	Minnesota	19	03	<u>25</u>	Me.	48	West Virginia	<u>50</u>	<u>20</u>	<u>30</u>	Wv.
24	Mississippi	<u>91</u>	<u>33</u>	30	Mp.	49	Wisconsin	20	<u>08</u>	<u>18</u>	Wi.
25	Missouri	10	<u>27</u>	<u>03</u>	Mi.	50	Wyoming	06	64	<u>15</u>	Wy.

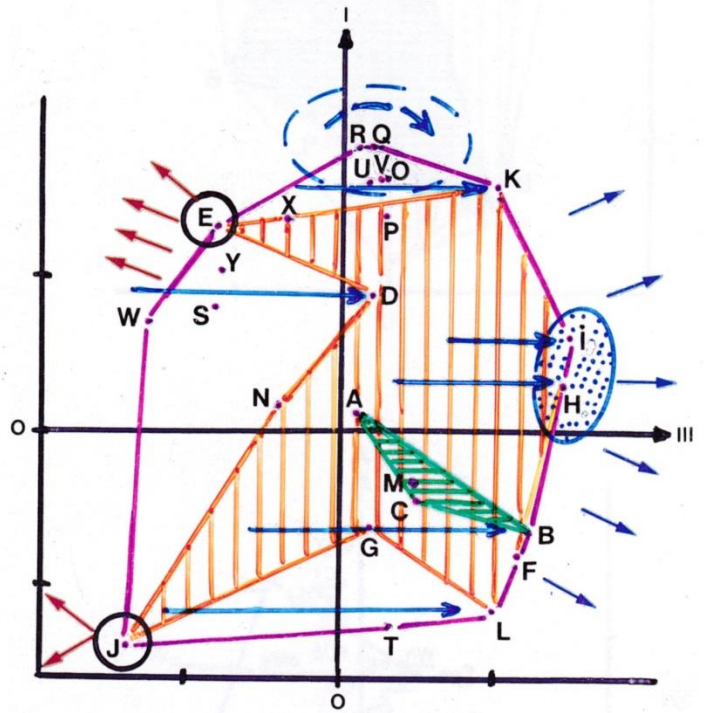
ÉTUDE RÉGIONALE
DES ÉTATS-UNIS
GÉOSTRUCTURES
SELON LES DESCRIPTEURS



Ci-dessus : BILAN DE L'UNION (I-II)

Ci-contre : MOUVEMENTS NOVATEURS (I-III)

Ci-dessous : L'AVENIR ANNONCÉ (II-III)



CLEF

Nature



Démographie



Politique



Pôle directeur



Isolat



Puits d'inactivation



Front



Traîne



Synergie



Antagonisme



Neutres : D-P, E-P, P-F, etc. (en I-II)

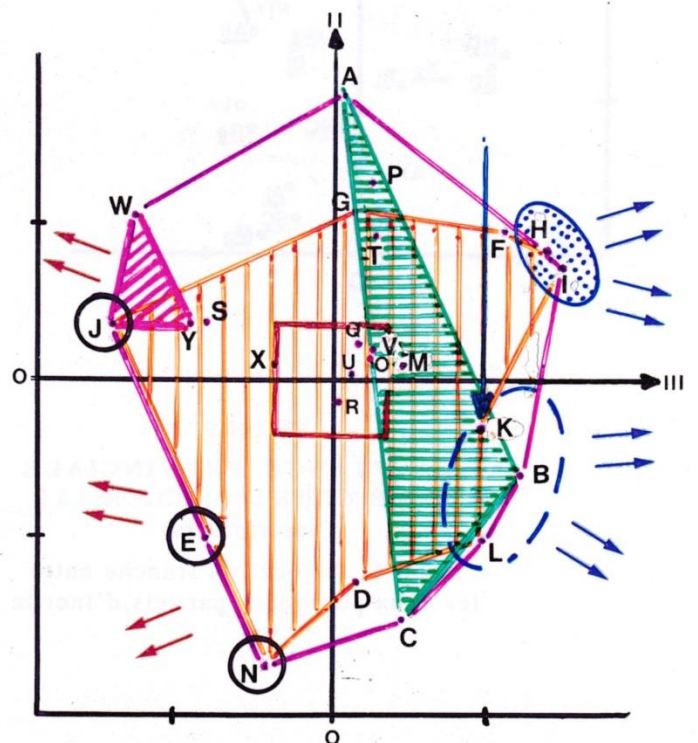
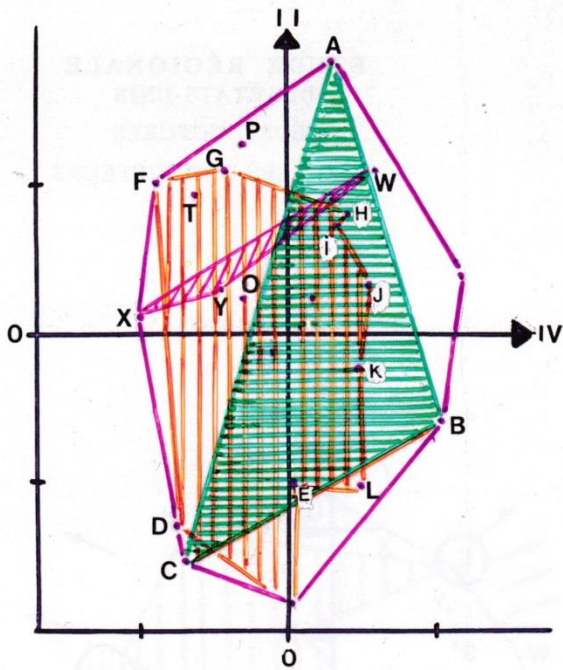


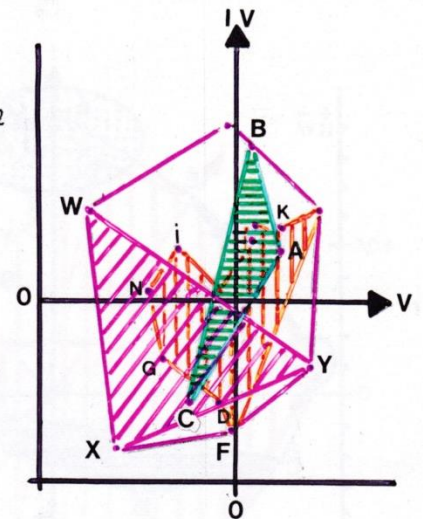
PLANCHE 3



SUITE DE LA PLANCHE 2
(CI-CONTRE)

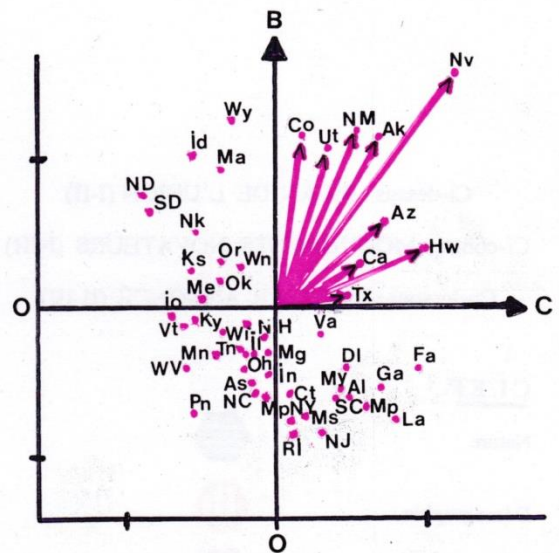
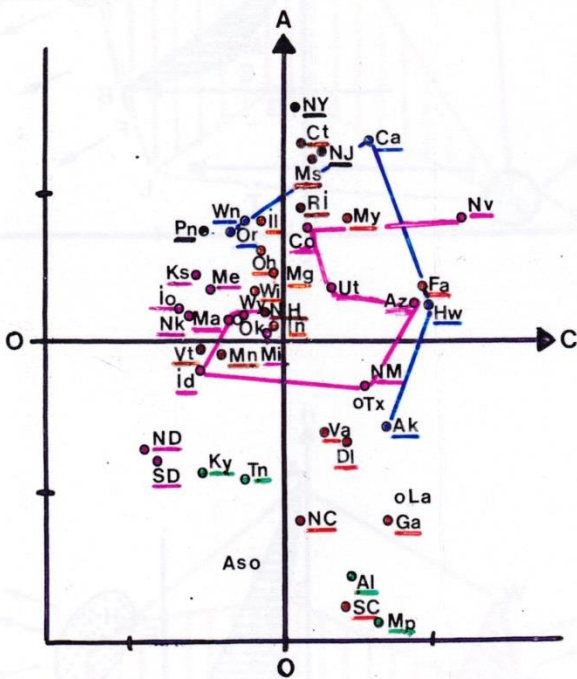
à gauche:
L'ESPACE NATUREL

à droite:
LA GÉOPOLITIQUE



CI-DESSOUS

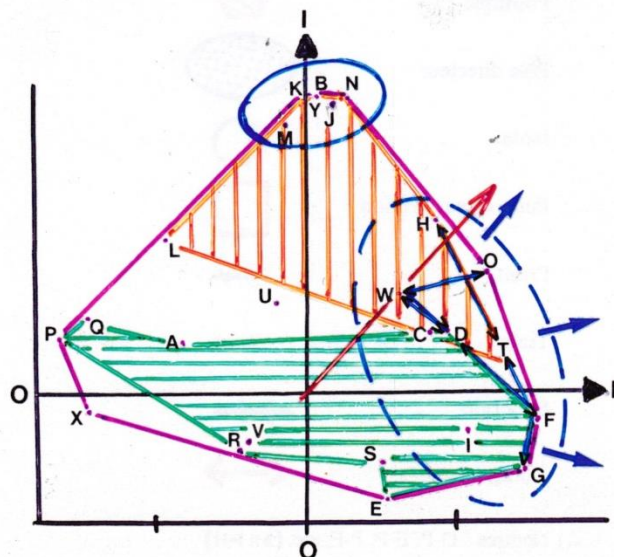
GÉOSTRUCTURES SELON LES INDIVIDUS
(États-Unis)



CI-CONTRE

LA FRANCE PROVINCIALE
GÉOSTRUCTURE DES FACTEURS I & II
(descripteurs)

Noter la disjonction franche entre
les deux polygones partiels d'inertie



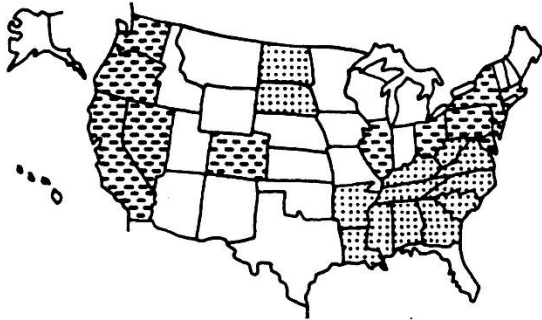
C'est sur tout cela que se fonde l'analyse de la "géostructure", laquelle est constamment éclairée, dans le commentaire, par un double va-et-vient : entre elle et le tableau des données initiales, d'un côté ; et entre elle et le tableau des corrélations interdescripteurs, de l'autre (v.Pl.5.) Il reste à dire que les "géostructures" n'ont pas toutes la même signification. La première est statique et dresse une sorte de bilan des grandes tendances de la région étudiée, au moment de l'étude : c'est donc une "structure d'état" (ou de constat). Les suivantes sont dynamiques et approfondissent ou affinent le bilan. Prospectives, et parfois prédictives, ce sont des "structures d'évolution". Quant aux dernières, qui décèlent les ultimes facteurs, et, parfois, dégagent des faits "enfouis" (cf. France ci-après, en partie III), elles ferment l'analyse et sont donc des "structures d'occlusion". Les principes étant posés et les géographes connaissant leur métier, on ne s'appesantira pas sur les commentaires ; seules seront mentionnées les lignes de force des "géostructures". Cinq plans vont être examinés, qui vont dévoiler "l'état de l'Union", dans la décennie 60 ; ses pesanteurs socio-économiques et socio-ethniques fortes ; et ses pulsions socio-économico-territoriales. De la richesse établie à la course au soleil, en passant par les antagonismes culturels et politico-économiques, tout est inscrit, ici, en clair ou en filigrane, jusqu'à une certaine forme d'anticipation de l'évolution historique, ce qui n'est pas le moins surprenant des résultats obtenus par A.C.P. À cet égard, la présente analyse, réalisée "après coup" (1984), fera dire aux sceptiques que l'histoire est plus aisée à raconter une fois qu'elle a été écrite. Je répondrai simplement ceci : en janvier 1988, j'ai analysé, selon les mêmes procédures et devant une centaine d'étudiants, la "géostructure" météorologique de la France, laquelle montrait de façon éclatante, pour l'été, l'occurrence d'une sécheresse au moins aussi grave que celle de 1976. J'écris sous leur contrôle, certains étant, aujourd'hui, enseignants d'Université.

La première "géostructure" des E.U. (axes I-II) est conforme à la règle : elle traduit l'"existant", avec un "pôle directeur" dense, formé par la *richesse établie* ou l'aisance (urbaines); les professions libérales lui donnent sa tonalité socio-économique (K,O,Q,R,U,V, en synergie) qui fait de l'axe I le support d'un facteur d'"establishment". En sont exclus, les ouvriers (N), les ruraux (J) et les Noirs (L), tous confinés dans leur isolat respectif, lié, pour le dernier, aux températures chaudes (B) et aux précipitations du Sud (C), qu'explicite le tableau des données originelles. On notera que S, automobiles, déjà banalisées aux E.U., n'est pas dans le pôle. Pour ce qui est des "Indiens" (M), ils sont dans le "puits", leur délaissement traduisant bien l'insignifiance de leur rôle, mis sous séquestre dans les *Réserves*. On ne citera que pour mémoire les oppositions d'évidence, tels que citadins/ruraux (K/J), montagnes/industries (A/N), population d'âge moyen/taux de natalité-mortalité infantile (E/F-G). En revanche, l'antagonisme population d'âge moyen aisée (E)/dépenses fédérales (T) souligne la forte hostilité à la politique sociale des Démocrates. Voilà l'exemple d'une observation à fouiller dans les structures suivantes. Aussi bien, l'antagonisme Noirs (L)/études longues (P) pourrait-il être préoccupant pour l'avenir. La place originale de P sur le "transfacteur", en quadrant N.E, laisse prévoir que ce descripteur est en train de devenir un élément essentiel aux E.U., d'autant que l'on note, à peu de distance, les descripteurs du dynamisme démographique (H et I), proches, sur l'axe II, des descripteurs des mouvements de la population (F et G). Ces mouvements mettent d'ailleurs en valeur l'axe II, qui, par le triangle A-B-C, livre clairement la nature *spatio-physique* du facteur II.

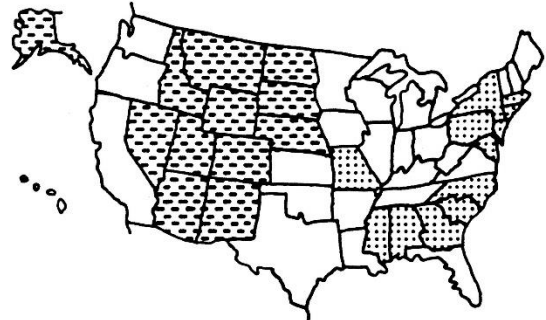
La deuxième structure (I-III) est fort intéressante car elle explique comment l'"existant", vu ci-dessus, se déforme sous l'influence du facteur III, indéniablement démographique en raison d'un "pôle directeur" en formation (H,I). Du coup, l'ancien pôle bascule vers le quadrant N.E et attire P, comme il fallait s'y attendre. L'"immobilité" de E est symptomatique du conservatisme des populations aisées d'âge moyen ; aussi bien, le renforcement de l'isolement des *farmers*, sur la "traîne" de la structure, est-il significatif de la permanence des tendances "déprimées" d'une partie non négligeable de la population. Inversement, la translation de B et de L vers le "front" marque un progrès spectaculaire, mais elle ne doit pas faire illusion : ce sont les états du Sud à fort pourcentage de Noirs qui bénéficient d'une expansion, et non leurs habitants noirs eux-mêmes. Cette translation est révélatrice, par ailleurs, des richesses cachées de l'analyse : elle invite en effet à reprendre les données originelles pour les approfondir. Dès lors, hors analyse, on peut découvrir qu'une partie de cette expansion est liée au développement du système de la "New plantation".

PLANCHE 4

ÉTATS-UNIS
 CARTES FACTORIELLES DES INDIVIDUS
 (v. les corrélations, Pl. I)



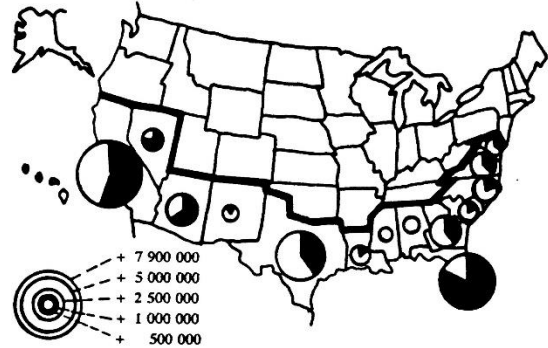
Facteur A
 Tireté: corrélations de 0.30 à 1.0
 Pointillé: corrélations de - 0.30 à -1.0
 (étendue 166: de 0.75 à -0.91)



Facteur B
 Tireté: corrélations de 0.25 à 1.0
 Pointillé: corrélations de - 0.25 à -1.0
 (étendue 120: de 0.79 à -0.41)

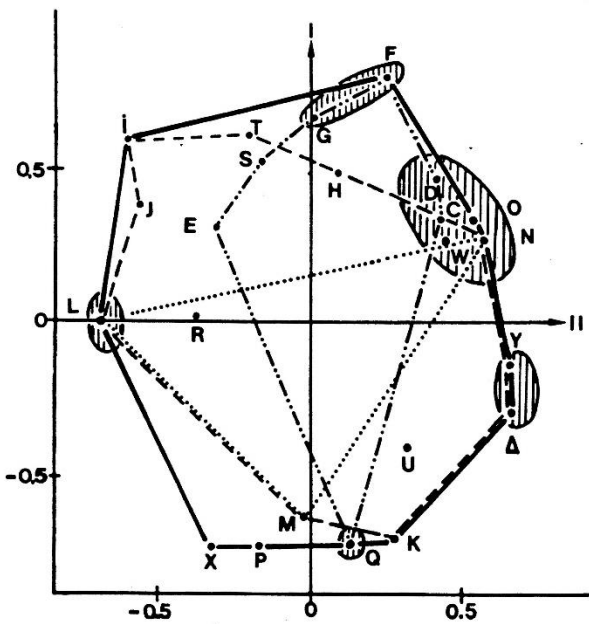


Facteur C
 Tireté: corrélations de 0.20 à 1.0
 Pointillé: corrélations de - 0.2 à -1.0
 (étendue 105: de +0.60 à -0.45)

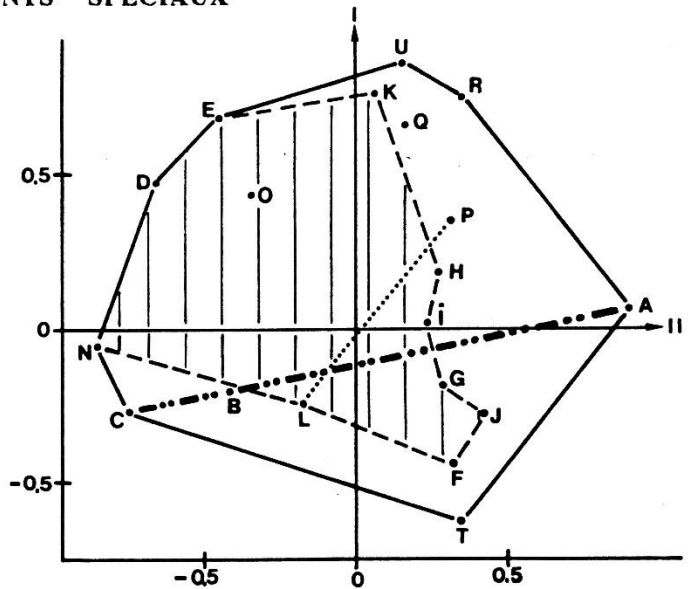


Sun Belt
 inspiré de G. DOREL "Historiens et Géographes"
 N°312, p.507, Paris 1987
 en NOIR dans les cercles: Solde Migratoire

TRAITEMENTS SPÉCIAUX



Plan factoriel (I-II) des Descripteurs
 de la France Provinciale (traitée en % généralisés)
 Structure modifiée et trop simplifiée: inversion injustifiée
 les facteurs, exagération des évidences inutiles (triangle L-M-N)



Plan factoriel (I-II) des Descripteurs
 des États-Unis, après rotation Varimax
 Structure très appauvrie: suppression du pôle directeur,
 réduction linéaire du triangle A.B.C. etc. Seul avantage éventuel:
 personnalisation franche de f II

La troisième structure (II-III) est également très parlante en ce qu'elle montre que l'avenir ne dépend plus seulement de la richesse ancienne : effondrement du "pôle" premier dans le "puits d'inactivation". Le front de l'expansion se précise, avec l'ébauche d'un nouveau "pôle" dont les éléments fondateurs sont d'ordre démographique dynamique (F,H,I) et spatial (B,C,L), renforcés par l'affirmation des villes du Sud : translation de K dans le "noyau" B,L. Le facteur II est donc bien confirmé dans son rôle spatio-physique. L'isolement des personnes d'âge moyen (E), la conjonction forte des ruraux (J) et du parti républicain (triangle J,W,Y) sur la "traîne" ("queue" du facteur III), accusent le durcissement anti-démocrate (X, à proximité du "puits") et la permanence des tendances conservatrices. On a bien vu, depuis, ce qu'il en a été...notamment à la fin de 1994 (Congrès majoritairement républicain). Avec cette géostructure, nous sommes au coeur du débat, puisque le rôle de l'espace et celui des mouvements de la population semblent préparer l'"Amérique" à venir.

Les quatrième et cinquième structures (resp. II-IV et IV-V) annoncent la fin de l'analyse par épuisement du sens : rétraction du polygone global. Néanmoins, celle-là permet de comprendre le rôle éminent de l'espace aux E.U., puisque le facteur IV est à nouveau d'ordre physique (climatique, tandis que II est plutôt topographique); celle-ci, de son côté, révèle la nature politique du facteur V dont le triangle d'inertie occupe à lui seul la moitié d'un "polygone global", très réduit.

Richesse traditionnelle, rôle de l'espace, mouvements de population, tendances socio-politiques claires, l'essentiel de ce qui faisait les E.U. des années 60 (et qui les fait encore) se trouve donc révélé en cinq figures, à partir de seulement vingt-cinq types de données très simples. Transposons cet exemple, banal plus encore que classique, dans le domaine de la recherche (ou dans celui de l'application pédagogique ou pratique), et nous voyons s'ouvrir d'infinies perspectives ; d'autant que les algorithmes permettent de très nombreux développements : analyses partielles (v. plus loin, France en III), rotations simplificatrices (v. *Pl.4*), etc.

3. Les "géostructures" des individus : "nouvelle frontière" et Sun-Belt

Ici, on ira très vite, car plans et cartes factoriels, parlent d'eux-mêmes. De plus, ils "recourent" les enseignements des plans-descripteurs (v. *Pl.3*). Ainsi, le lecteur verra-t-il, s'il le dresse, que le plan A-B (deux premiers facteurs) décèle spontanément le découpage régional des E.U., tel qu'on le connaît de manière traditionnelle : Nouvelle Angleterre, Mégalopolis, Grandes Plaines, Sud ("profond" et "jeune"), etc.; et il le comparera aux divisions "régionales" administratives du *Census* ... Pour ce qui est des plans factoriels A-C et B-C (*Pl.3*), ils établissent, sans ambiguïté, comment, dès les années 60, la géographie régionale traditionnelle des E-U est en train d'éclater en une spectaculaire redistribution de l'espace, qui, somme toute, illustre le thème de cette "Nouvelle Frontière" que Kennedy ne voyait peut-être pas sous cette forme, mais qu'il annonçait cependant comme devant être la grande "affaire" à venir pour son pays. Les cartes factorielles (*Pl.4*), de leur côté, sont du plus haut intérêt : A, en dressant un utile complément au plan factoriel I-II, par opposition des régions pauvres aux régions d'ancienne ou de nouvelle richesse (Côte Ouest); B, en montrant que l'analyse peut "révéler" des faits non supposés par la sélection des descripteurs : on voit clairement se différencier régions de vieille occupation (anciennes "Colonies" notamment) et bloc des nouvelles terres acquises lors de la "conquête de l'Ouest" -le facteur B est donc de nature spatio-historique franche ; et C, plus remarquablement encore, en "dévoilant", bien avant que le phénomène n'ait été reconnu, l'émergence de ce que l'on a appelé depuis la *Sun Belt* (v. l'excellente étude de G. Dorel, cité à la *Pl.4*) : c'est le rôle prédictif des Composantes Principales. Et pour se convaincre de l'excellence de ces méthodes, le lecteur peut s'essayer à dresser et à commenter les plans factoriels de l'exemple qui lui est proposé ci-après.

PLANCHE 5

LA FRANCE PROVINCIALE

A	1
B	08 1
C	26 18 1
D	04 10 43 1
E	10 25 13 33 1
F	12 09 32 66 03 1
G	27 18 18 02 47 50 1
H	06 47 15 39 16 37 11 1
I	07 24 01 47 05 56 28 67 1
J	14 96 13 22 25 01 18 51 09 1
K	08 99 18 04 26 16 21 42 32 93 1
L	45 38 00 18 21 23 48 25 01 44 40 1
M	04 95 11 08 22 26 18 35 40 88 97 29 1
N	06 98 17 15 25 02 12 54 15 96 96 33 92 1
O	06 39 20 12 24 31 57 50 33 45 33 02 30 47 1
P	65 08 27 11 37 44 54 14 25 09 13 63 10 01 33 1
Q	50 11 16 04 51 44 63 01 15 11 15 52 13 06 32 81 1
R	17 10 11 34 32 17 13 29 28 12 07 13 05 14 08 27 24 1
S	46 17 17 01 37 15 20 00 22 13 21 29 17 14 09 29 29 01 1
T	41 18 31 14 09 32 49 22 07 11 14 59 23 23 35 64 50 25 17 1
U	12 21 24 23 02 22 05 41 18 22 20 19 24 28 32 15 29 09 07 03 1
V	28 19 41 03 20 10 00 04 20 14 21 29 18 17 01 22 28 06 15 31 28 1
W	19 22 22 41 06 36 02 52 46 32 19 37 07 23 33 07 05 17 08 14 04 03 1
X	07 06 24 53 09 68 49 32 44 14 02 16 11 14 46 44 39 27 21 32 15 03 31 1
Y	11 98 19 14 27 07 20 46 22 96 97 41 91 96 37 11 12 08 17 11 18 24 25 08 1
A	B C D E F G H I J K L M N O P Q R S T U V W X Y

TABLEAU DES CORRÉLATIONS ENTRE DESCRIPTEURS

Zéro supprimé. Arrondi à 2 chiffres. Valeurs négatives soulignées

Variance totale à 82% de saturation, avec pour chaque facteur

I= 28%. II= 21%. III= 12.5%. IV= 9%. V= 6.5%. VI= 5%

Classement des descripteurs par importance décroissante

B.K.N.Y.M.J.S.A.L.Q.I.P.D.H.F.T.U.R.E.O.G.X.W.C.V.

A.C.P. *Principal Components*Durée de l'analyse= 51/100 seconde
pour 2200 données en 5430 chiffres

CORRÉLATIONS DESCRIPTEURS-FACTEURS

	I	II	III	IV	V	VI
A	15	41	53	32	36	43
B	98	04	18	05	06	06
C	20	41	08	52	23	21
D	20	43	55	52	14	08
E	35	26	19	66	25	17
F	06	76	36	24	16	01
G	25	72	15	33	08	19
H	56	43	46	13	26	12
I	13	52	71	10	21	10
J	96	08	02	07	09	05
K	96	04	22	03	06	04
L	50	48	48	02	25	31
M	91	07	34	11	02	09
N	96	13	13	08	01	06
O	40	60	05	42	03	22
P	18	83	32	02	14	06
Q	24	74	39	16	27	26
R	16	25	25	32	61	35
S	23	26	14	46	25	70
T	10	68	42	03	40	01
U	29	12	16	51	59	19
V	15	21	47	47	19	01
W	32	31	56	03	25	19
X	06	73	29	04	13	12
Y	97	02	13	01	11	05

NOMENCLATURE DES DESCRIPTEURS

A: Surface totale- B: Population tot.- C: Distance au centre géographique- D: Température (moy. an)- E: Précipitations (som.)- F: Ensoleillement (heures/an)- G: Energie du relief- H: Pop. (solde nat.)- I: Migrations internes- J: Personnes de plus de 65 a.- K: Jeunes (0-19 a.)- L: Pop. active (I.re)- M: Pop.act. (II.re)- N: Pop. act. (III.re)- O: Chambres d'hôtel- P: S.A.U. (ha)- Q: Terres lab. (ha)- R: Surf. tjrs en herbe (ha)- S: Forêts (ha)- T : Pop. immigrée- U: Logements tout confort- V: Ménages à 2 voitures- W: Places de camping- X: Équipement médical (méd. + lits)- Y: Demandeurs d'emploi (Pas de pourcentages)

GÉOSTRUCTURES PARTIELLES

Corrélations avec les facteurs I et II nouveaux
(mêmes conventions que dans le Tableau Général ci-dessus)

	I	II		I	II
H	51	04	C	35	57
I	20	05	D	20	80
J	98	03	E	54	67
K	99	07	F	65	52
L	44	78	G	83	15
M	95	13	Q	88	21
N	99	10	R	12	60
T	14	87	S	28	35
Y	99	02			

III - UNE ANALYSE EN KIT : LA FRANCE "PROVINCIALE"

Par France provinciale, il faut entendre la France métropolitaine privée des départements de sa région 10 (75, 77, 78, 91, 92, 93, 94, 95) et des D.O.M., trop "typés" les uns et les autres ; soit quatre-vingt huit (88) individus pour vingt-cinq (25) variables. En effet, dans la situation du début des années 80, on a voulu savoir ce qu'était devenu le "Désert français" de F. Gravier. Le résultat est concluant. On trouvera à la *Pl.5* les éléments à monter aux fins d'analyse, selon le mode de représentation des *Planches 2* et *3*, en complément du plan I-II déjà effectué (*Pl.3*). Avant tout commentaire, on relèvera que, *dès 1984* (date des calculs exploités en "Licence" à Nantes), et donc bien avant l'explosion du C.I.P., la simple lecture des résultats de l'A.C.P. révélait, sans ambiguïté aucune, l'effrayante *montée du chômage des jeunes*, puisque, dans le classement des 25 descripteurs (de toute nature), selon leur participation à la variance générale, se trouvaient aux cinq premières places : la population brute, les jeunes de 0 à 19 ans, le secteur tertiaire, le chômage et le secteur secondaire... La première structure (plan I-II), par le déséquilibre remarquable de son "*polygone global*" (dû à l'absence de la région Île-de-France), indique bien la gravité des problèmes dans la France provinciale, où s'opposent deux aires : celle de la France "grise" (zones industrielles traditionnelles en crise) du "*pôle directeur*", et celle de la France "ocre", étendue entre "*transfacteur*" N.E. et "*tête*" du facteur II, où se regroupent, sous le soleil et dans les zones de loisirs, les migrants et les populations en accroissement, pour former un embryon de *Sun Belt* à la française.

À partir des structures partielles de la *Pl.5* (que l'ordinateur calcule, à la demande, sur descripteurs sélectifs), on peut voir que le "*polygone global*" est immédiatement dépendant des descripteurs démographiques (quasi identité de formes), et que les descripteurs physiques pèsent aussi d'un poids très lourd. L'étude des plans généraux suivants est, de son côté, très significative : I-III, montre le rôle perturbateur du facteur I1II (démographie dynamique); II-III accuse l'opposition entre une "France fermée" ("*tête*" de II) et une "France ouverte" ("*queue*" de II); une France des labours et une France des herbages ; une France de l'hôtellerie (montagne) et une France du camping (soleil), etc. Le plan II-IV, par l'expansion extraordinaire des faits physiques, est indicatif du poids que ces faits peuvent prendre dans l'avenir : comparer à ce que devient, dans cette "*géostucture*", l'ancien "*pôle*". Cette analyse, exceptionnellement riche, mérite que l'on forme aussi, pour y déceler de subtils détails, les plans I-IV, IV-V et IV-VI qui sont de remarquables condenseurs-simplificateurs, propres à confirmer les commentaires précédents. Il n'est pas jusqu'au dernier plan, V-VI, qui n'ait de l'intérêt : avec un peu d'attention en effet, le géographe y découvrira comment, au travers de descripteurs non choisis à cette fin, l'analyse, en "désenfouissant" des faits implicites, permet de restituer la France des bocages et celle des openfields, entre "*tête*" et "*queue*" du facteur VI.

Voilà ce que peuvent apporter les méthodes modernes à l'analyse synthétique. L'efficacité qu'elles insufflent à la pratique géographique ne saurait être mise en rapport avec les préjugés qu'inspire, trop souvent, une mathématique tenue pour inapprochable. C'est pourquoi, en donnant bien volontiers acte à Alain Chauvet de son "combat" pour la Régionale, je formerai le voeu que ne soient pas perdus les efforts pionniers de ceux qui, comme M. Phipps ou J-C. Wieber, ont ouvert la voie à une action rénovatrice, sans laquelle il faudrait craindre que la géographie future ne s'enlise dans des exercices mineurs, marginaux, ou mimétiques de ceux des vraies disciplines scientifiques.

Nantes, Janvier 1995

Erratum : *Pl.2*, dans le "Pôle directeur", lire O au lieu de N