LES CAHIERS DE L'ANALYSE DES DONNÉES

M. LESTY
P. BUAT-MÉNARD

La synthèse géométrique des corrélations multidimensionnelles

Les cahiers de l'analyse des données, tome 7, nº 3 (1982), p. 355-370

<http://www.numdam.org/item?id=CAD_1982__7_3_355_0>

© Les cahiers de l'analyse des données, Dunod, 1982, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » 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/

Les Cahiers de l'Analyse des Données
Vol. VII - 1982 - n° 3 - p. 355-370

LA SYNTHÈSE GÉOMÉTRIQUE DES CORRÉLATIONS MULTIDIMENSIONNELLES [SYNTH. GÉOM. CORR.]

par M. Lesty (1)
et P. Buat-Ménard (2)

Il existe, dans l'atmosphère et la mer, une quantité considérable de matière particulaire en suspension qui a une grande influence sur l'équilibre écologique de notre planète. Connaître l'*origine* des éléments chimiques dans l'air marin et leur influence sur la géochimie de l'océan prend de plus en plus d'importance en raison de la contribution croissante des apports d'aérosols d'origine industrielle.

Lorsqu'on recueille des aérosols au-dessus de l'océan, la difficulté ne réside pas dans les techniques de mesures, maintenant très précises, ni dans le manque de mesure (celles-ci sont même surabondantes), mais dans l'interprétation des données.

En effet, les conditions météorologiques varient sans cesse, les masses d'air se succèdent, le bateau se déplace et les sources de particules sont variées et variables dans le temps. La composition de ces sources est mal connue, l'aérosol peut modifier chimiquement et physiquement pendant son voyage, sa dissémination dépend de sa taille et de sa forme, de la météorologie etc.

Lorsqu'on est confronté à un processus multiparamétrique, la première étape consiste à rassembler les données accessibles se rapportant au phénomène à expliquer. Elles se présentent d'ordinaire sous la forme d'un tableau numérique à plusieurs variables mesurées en parallèle sur plusieurs échantillons.

Ces paramètres ne sont pas tous indépendants. Les intéractions sont, en général, si complexes qu'une description mathématique serait inextricable. On se trouve donc réduit à la méthode suivante :

1 - Regrouper les variables qui ont des variations similaires (par exemple, au moyen d'un tableau de corrélation).

2 - Trouver les intéractions entre ces groupes.

3 - Représenter le tout sous forme intuitive et rapidement accessible.

Dans ce but, un nouvel outil informatique a été mis au point au C.F.R. à Gif-sur-Yvette, qui permet la représentation globale immédiate de l'information interprétable des grands tableaux de données multivariables.

---

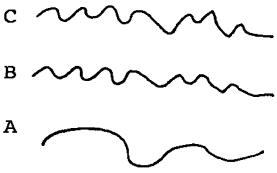
(1) *Docteur d'état*

(2) *Docteur d'état*
Centre des Faibles Radioactivités - C.N.R.S. - Gif-sur Yvette.

On propose ici une application de cet outil au problème de l'origine des particules dans l'atmosphère, à partir de données récoltées dans l'Atlantique tropical Nord-Est qui avaient déjà été traitées selon les méthodes d'analyse classiques (Buat-Ménard 1979). Nous verrons que le traitement informatique confirme les interprétations qui avaient été données et apporte des informations supplémentaires.

1 *Description de la méthode*

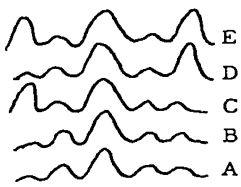
Dans les sciences naturelles, où les conditions de mesure sont rarement identiques, un calcul de moyenne ne présente de sens physique que si le contexte est parfaitement précisé. Fermer les yeux sur la variabilité fine des mesures qualifiée d' "aléatoire" irait à l'encontre des tentatives d'amélioration de la précision des appareils.



Si, par exemple, on lisse les courbes ci-contre, les trois variables A, B, C semblent appartenir à un groupe unique. L'examen des micro-variations montre cependant que B et C forment un groupe à part. Dans le cas où ces variables ont été mesurées indépendamment au moyen d'appareils différents dans leur principe, la similitude des micro-variations constitue une "macro-information". Les variations relatives de B et C ne sont pas dues à une erreur de mesure.

Pour toutes ces raisons, nous nous placerons d'un point de vue *non- statistique*, en particulier dans l'utilisation du "coefficient de corrélation".

a - *Visualisation globale des variables par des points dans l'espace.*

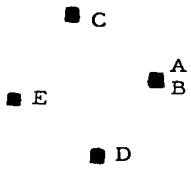


Considérons les courbes ci-contre :

- B est semblable à A.

- Quantitativement, il y a autant de différence entre C et A qu'entre D et A ; mais qualitativement, ce n'est pas la même différence.

- De même, il y a autant de différence entre C et E qu'entre D et E.



Représentons les différences entre les variations par une "distance" d'autant plus longue que la différence est grande. Nous pouvons visualiser les relations entre ces courbes de la manière suivante :

- A et B : points confondus.

- C et D à égale distance de A, mais dans deux directions distinctes.

- E à égale distance de C et D.

(Il y a une infinité de manières d'être à égale distance de A (cercle), mais si l'on connaît la position de A et C, et la distan- r_{CD} , la position de D est connue par triangulation).

Le problème de la représentation graphique automatique sera donc de caractériser la différence entre deux variables par un nombre possédant les propriétés d'une distance.

Soit $\bar{x}$ et $\bar{y}$ les moyennes des mesures de deux variables x et y. Les écarts à la moyenne sont :

$$X_1 = x_1 - \bar{x}\ ,\ X_2 = x_2 - \bar{x}\ ,\ X_3 = x_3 - \bar{x} \quad \text{etc.}$$

$$Y_1 = y_1 - \bar{y}\ ,\ Y_2 = y_2 - \bar{y}\ ,\ Y_3 = y_3 - \bar{y} \quad \text{etc.}$$

Les ouvrages de statistique caractérisent la "corrélation" entre x et y par la formule suivante (définie à partir des notions statistiques de "variance" et de "covariance") :

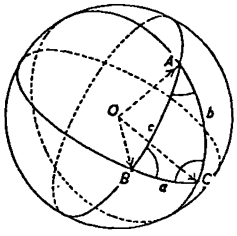
$$\text{Coeff. de corrélation } r_{xy} = \frac{X_1\ Y_1 + X_2\ Y_2 + X_3\ Y_3 + \ldots}{\sqrt{X_1^2 + X_2^2 + X_3^2 + \ldots}\sqrt{Y_1^2 + Y_2^2 + Y_3^2 + \ldots}}$$

D'un point de vue *non-statistique* cependant, le numérateur de cette formule est un produit scalaire ; et le dénominateur est le produit des longueurs de deux vecteurs $\vec{X}$ et $\vec{Y}$, respectivement de coordonnées X_1 , X_2 ,X_3 ,..., et Y_1 ,Y_2 ,Y_3 ,...

Pour la géométrie analytique, en effet, cette formule est le COSINUS de l'angle entre les deux vecteurs $\vec{X}$ et $\vec{Y}$.

Il en résulte (quel que soit par ailleurs le nombre n de coordonnées de X et Y) que l'arc-cosinus de r_{xy} est un nombre qui possède toutes les propriétés d'un ANGLE : en particulier, c'est un nombre *limité* (entre $-\pi$ et $+\pi$), donc toujours représentable graphiquement dans les limites d'une feuille de papier.

Nous disposons donc d'une DISTANCE angulaire qui va nous permettre de REPRESENTER graphiquement sur une *sphère* unique les différences entre les variations de paramètres mesurés simultanément.



Connaissant, par exemple, la valeur du coeff. de corrélation r_{BC} , nous représentons les variables B et C par deux points séparés sur la sphère par l'angle $\alpha = \text{arc-cos}(r_{BC})$

Connaissant r_{AB} et r_{AC} ,nous pourrons, par triangulation, positionner n'importe quelle autre variable A, et, de proche en proche, toutes les variables disponibles.

(Les problèmes techniques qui se posent parfois lorsque le nombre n de mesures est supérieur à 3 sont décrits ailleurs (Lesty, 1981)).

Si $r_{BC} = 1$, l'angle est nul : B et C sont colinéaires.

Si B et C sont anticorrélés ($r_{BC} = -1$), l'angle = 180°.

Note :

Du point de vue statistique, on s'intéresse essentiellement aux relations linéaires (c'est-à-dire aux angles nuls). Lorsque la relation n'est pas linéaire, on calcule la probabilité pour que l'on puisse considérer cet écart entre les vecteurs comme "aléatoire", et admettre que la relation est linéaire. D'où l'usage de "niveaux de probabilité d'erreur" en dessous desquels on considère la corrélation comme "non-significative" (en fait, c'est seulement la conclusion de linéarité qui n'est plus possible : un cosinus est toujours significatif, même s'il est nul).

Du point de vue physique, au contraire, on s'intéresse essentiellement aux corrélations différentes de l'unité (c'est-à-dire aux angles non-nuls). L'organisation des différences de variation entre paramètres (fondée sur les propriétés du triangle sphérique) fait toute la matière de la physique.

En pratique, le premier point sur la sphère sera placé n'importe où ; ensuite la variable la moins corrélée à la première ; puis , par triangulation, la variable la moins corrélée à la fois aux deux premières. A partir de ce triangle de base, les autres paramètres se situent de proche en proche. La sphère est représentée en perspective . Les points seront des cubes dont la taille permettra de distinguer ceux qui sont au premier plan (gros) de ceux qui sont à l'arrière plan. Pour alléger la figure, on ne dessine pas le pourtour de la sphère.

b - *Analyse de la qualité des relations observées.*

Placer les variables d'après les corrélations ne suffit pas pour interpréter les données car la corrélation est une notion relative : Par rapport à une étoile, le mouvement d'un stylo, par exemple , est corrélé à ceux du bras, de la terre, du papier et du soleil, car l'ensemble est entraîné dans le mouvement du système solaire. Dans le repaire de la terre, au contraire on n'observe plus la corrélation stylo, papier mais la corrélation stylo, main demeure forte. Connaissant dans le repère de l'étoile, les corrélations deux à deux des variables disponibles, l'observateur peut calculer les corrélations qu'il aurait trouvées s'il avait mesuré les mouvements dans un autre repère z, par exemple la terre ("corrélation partielle par rapport à z"). Ces corrélations partielles sont données par la formule :

$$r_{xy/z} = \frac{r_{xy} - r_{xz}\, r_{yz}}{\sqrt{(1-r_{xz}^2)}\sqrt{(1-r_{yz}^2)}}$$

qui n'est autre que la formule fondamentale des triangles sphériques.

Ayant placé les points sur la sphère (cf. § précédent), nous conviendrons de relier les variables x et y par un *trait* lorsque les corrélations partielles $r_{xy/z}$ de même signe que la corrélation totale, sont toutes supérieures en valeur absolue à une certaine valeur prise comme SEUIL, *quel que soit* z parmi les variables disponibles.

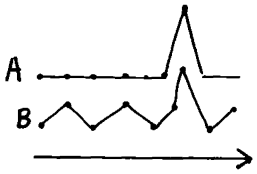
STYLO ——— bras(soleil) ——— MAIN

Exemple : Stylo et main sont liés, car, même dans le repère du bras où la corrélation partielle est la plus faible, celle-ci est supérieure au seuil. Le long du trait sont mentionnés le repère dans lequel la corrélation partielle est la plus faible et celui dans lequel la corrélation partielle est la plus forte . Ce dernier, inscrit entre parenthèses, est donc le paramètre qui a le

moins d'influence sur ce lien (et même, parfois, varie en quadrature avec les 2 variables reliées).

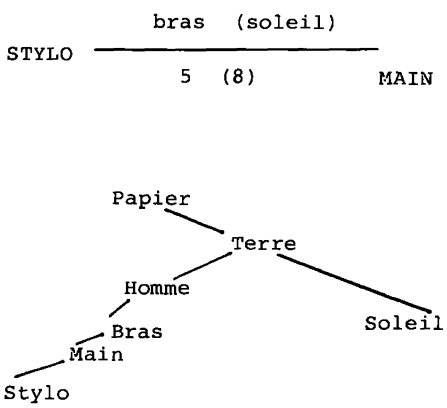
Il n'y aura pas de lien entre stylo et bras : car dans l'exercice de l'écriture la corrélation partielle stylo, bras par rapport à la main devient négative (la valeur du seuil a donc moins d'importance qu'il paraitraît d'abord : il est choisi automatiquement en fonction des données afin que le schéma ne soit pas un réseau inextricable de traits).

Cette évaluation de la "solidité" des relations, par un tri méthodique des corrélations partielles permet un tracé schématique de l'organisation des variations selon les articulations naturelles.



Avant le tracé du trait, un dernier contrôle est cependant nécessaire : Les variables A et B, ci-contre, sont fortement corrélées à cause du seul échantillon n° 7 . On calculera donc la corrélation r_{AB} successivement en l'absence de chacun des n échantillons. Si une seule de ces n valeurs est inférieure au seuil, on considérera la liaison comme "douteuse" et l'on ne reliera pas A et B.

En résumé, le programme informatique comporte deux phases :



1- Une phase d'*analyse* de chaque relation, au cours de laquelle on cherche *la plus faible* des corrélations partielles : le bras, *la plus forte* : le (soleil), l'échantillon qui a *le plus d'influence* : par ex. le n° 5, et *le moins d'influence* : le (n° 8). ("Le secret de toute la méthode est là : en toute chose repérer soigneusement ce qui est le plus absolu" Descartes).

2 - Une phase de *synthèse* rendue possible par le tracé, sur une seule figure globale, de l'agencement de l'ensemble des relations simples.

("Toute la science humaine consiste en la vision distincte de la façon dont ces natures simples concourent ensemble à la composition des autres choses" Descartes).

La figure est aussi complète qu'elle peut l'être, en fonction des données disponibles (elle serait évidemment plus complète si l'on connaissait aussi les mouvements des phalanges des doigts etc.).

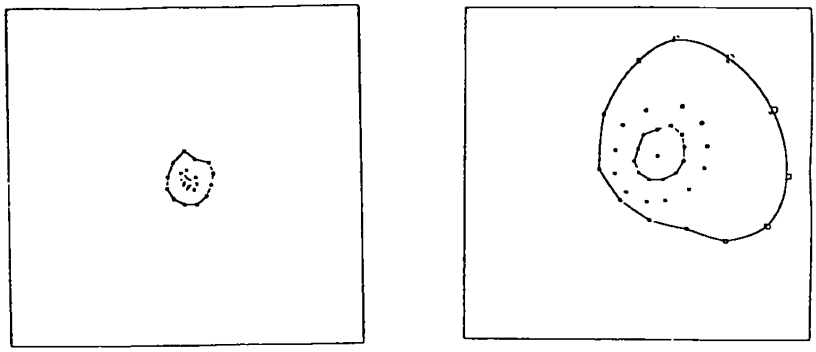
Note :

Les liaisons de *corrélation* seront représentées par un *arc* de grand cercle à la surface de la sphère ; les liaisons d'*anticorrélation* par une *droite* pointillée traversant la sphère.

c - Dispositif d'accomodation.

Il reste un problème technique : Il peut arriver que toutes les variables disponibles soient fortement corrélées auquel cas les points représentatifs sont tous concentrés en un amas difficilement distinct

(exemple ci-dessous à gauche). Pour pallier à ce genre d'inconvénient, le programme est doté d'un dispositif qui dilate automatiquement la figure afin d'occuper de façon plus rationnelle la surface disponible sur la sphère (cette technique est basée sur le calcul des corrélations des corrélations : voir Lesty, 1981). On obtient la figure de droite :



CORRELATIONS ⇨ CODISTAN

Inversement la même technique conduira automatiquement à la contraction de la figure s'il faut faire place à des données indépendantes.

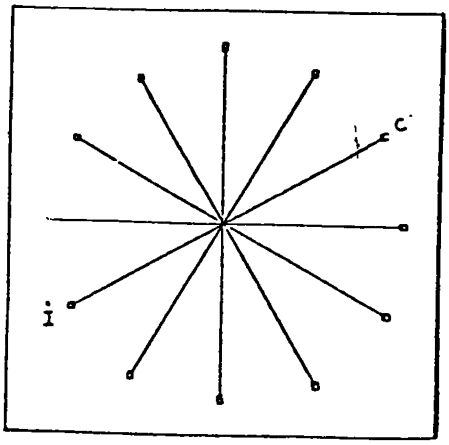
Exemple : Sur la figure de gauche est représenté un ensemble de paramètres présentant une organisation cyclique. A droite a été tracée l'organisation d'un autre ensemble de variables, cycliques aussi, quoique indépendantes du précédent. Rassemblons ces deux ensembles de données, ajoutons-y un autre ensemble de mesures sans rapport avec elles ; le tout formant un seul tableau numérique (ces trois ensembles de variables sont mesurés sur un nombre identique d'échantillons). Le schéma de la page 361 représente ce tableau numérique: le programme a bien séparé les trois ensembles de données indépendantes, en conservant la forme des figures, mais en réduisant la taille de l'une d'elle (pour plus de détails sur ces exemples : voir Lesty, 1981).

d - Commentaire sur l'utilisation de la méthode.

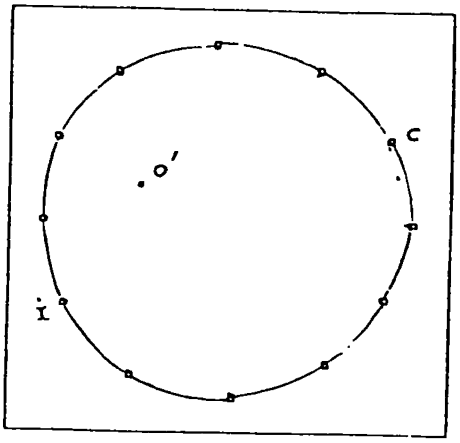
- C'est une méthode non statistique qui n'exige pas d'hypothèse préalable sur la distribution statistique des mesures, et permet la visualisation purement descriptive des intéractions. On peut l'assimiler à une "photographie" de l'évolution dynamique des données disponibles, pour un lieu et une période donnés.

On simplifie toujours la difficulté en augmentant le nombre de variables en rapport avec le problème (une figure est d'autant plus claire qu'elle est plus complète : par exemple, les cycles de la page précédente n'apparaissent que si l'on dispose de suffisamment de points pour former les cercles).

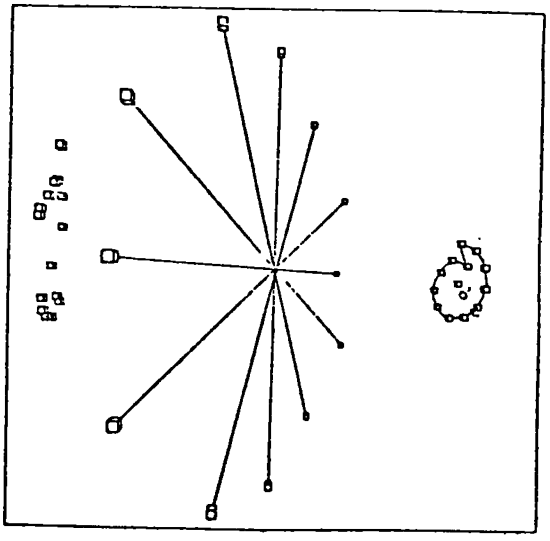
- C'est une méthode fiable : l'introduction d'un ou plusieurs paramètres mal mesurés ou sans rapport avec le problème ne fausse pas l'interprétation des paramètres correctement mesurés (on ne peut, bien sûr, tirer des données plus d'informations qu'elles n'en contiennent). Rappelons encore que la corrélation ne dépend pas des erreurs systématiques.



Projection sur Oy



Projection sur Ox



Données quelconques

Projection sur Oy

Projection sur Ox

- C'est une méthode purement graphique : elle nous épargne la consultation fastidieuse de tableaux numériques, sans sacrifier de l'information, ce qui la rend accessible au non-mathématicien . La densité d'information réunie sur chaque schéma permet l'intercomparaison rapide des figures.

- Une analyse plus fine des variations découle de la prise en compte simultanée de toutes les informations disponibles . Il est possible, mieux que par l'analyse spectrale, d'identifier les diverses composantes des variations des concentrations des éléments, qui, chacunes, peuvent caractériser les origines des particules ou polluants. Les déphasages sont mis en évidence.

e - Comparaison avec l'Analyse Factorielle.

La méthode décrite ici présente quatre principales nouveautés par rapport à l'Analyse Factorielle :

1) Positionnement des points sur une surface globale unique. Au contraire, dans une analyse factorielle, chaque variable est assimilée à "un point dans un espace à n dimension" (n information par variable) on ne peut visualiser que des coupes à 2 dimensions de cet espace ("plans factoriels") il faut donc tracer plusieurs figures, et contrôler les proximités par des "tableaux de contributions".

2) Evaluation de la "solidité" des relations par un tri méthodique des corrélations partielles.

3) Contrôle méthodique de l'influence de chaque échantillon.

4) Dispositif automatique d'accomodation pour optimiser la clarté graphique.

Par ailleurs les analyses factorielles sont parfois gênées par un "effet de taille" qui peut diminuer l'intérêt explicatif du premier plan factoriel. De plus, l'introduction d'une variable sans rapport avec le problème peut déplacer les plans factoriels et quelquefois empêcher l'interprétation.

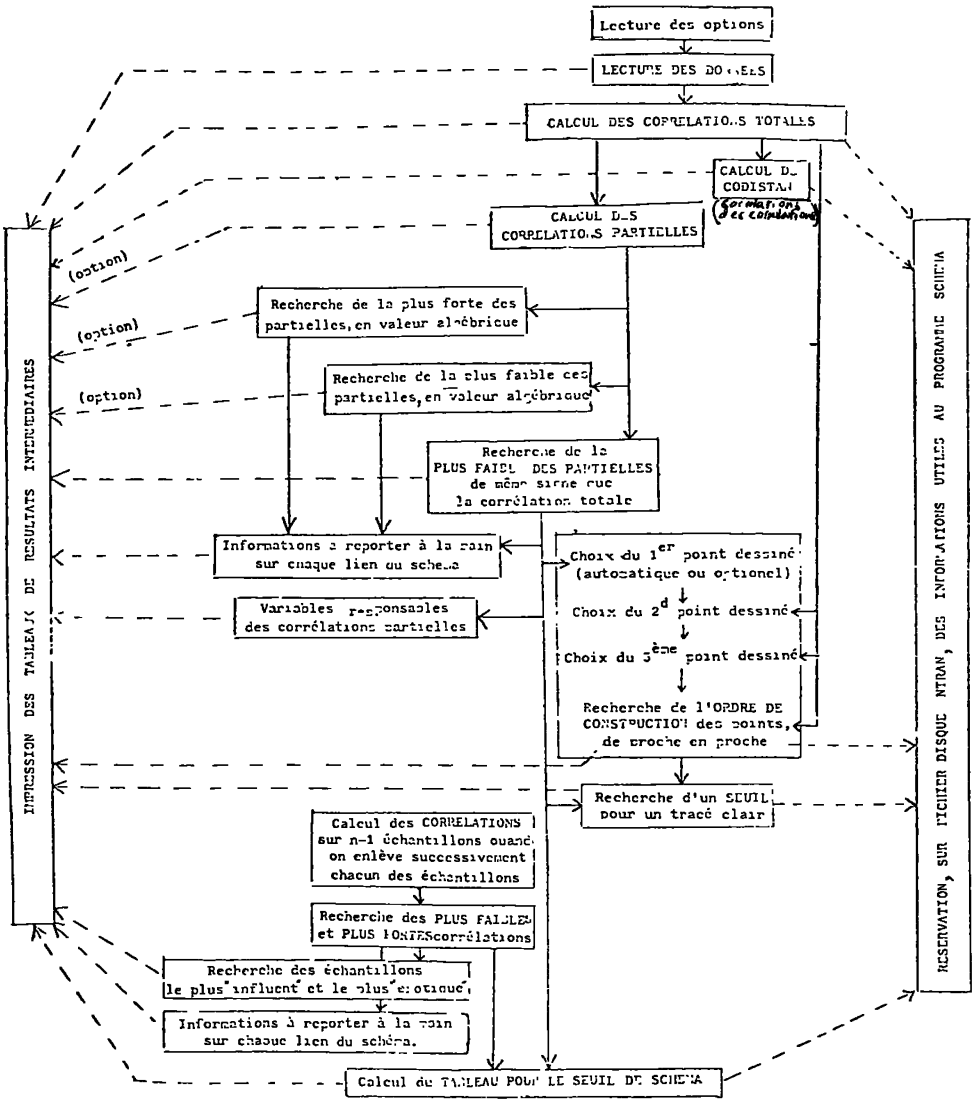
f - Organigramme.

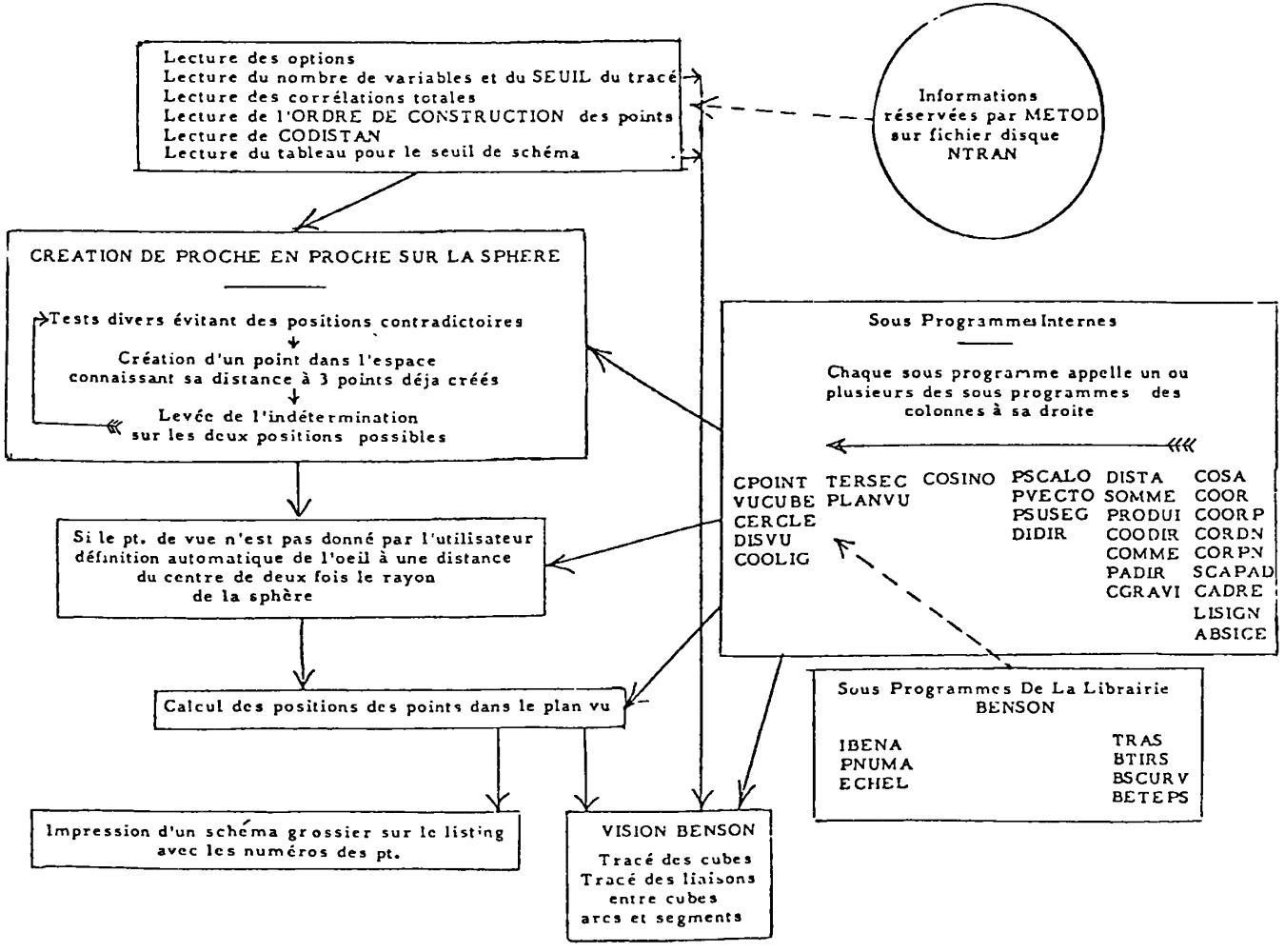
Pour des raisons d'économie de temps de calcul et de taille de mémoire, le programme a été segmenté en deux sous-programmes :METHOD et SCHEMA dont les organigrammes simplifiés sont donnés ci-après. Un programme principal : CORICO, appelle l'un ou l'autre des deux sous-programmes selon les besoins.

2 *Origine des éléments métalliques dans l'aérosol de l'Atlantique tropical Nord-Est*

On présente ici l'analyse de prélèvements de particules par filtration d'air sur filtre Whatman 41, à environ 10 m. au-dessus de la surface de la mer, au cours des campagnes MIDLANTE, RIFT et GATE, d'avril à juillet 1974. La description des techniques de prélèvements et d'analyse chimique, ainsi qu'une première interprétation de la géochimie de ces prélèvements a été publiée par Buat-Ménard (1979). Nous avons traité ces données au moyen du programme informatique décrit plus haut, après y avoir joint les paramètres physiques, météorologiques ou géographiques susceptibles de nous aider à mieux caractériser les sources possibles d'aérosols.

Sous-programme METOD




Lecture des options
Lecture du nombre de variables et du SEUIL du tracé
Lecture des corrélations totales
Lecture de l'ORDRE DE CONSTRUCTION des points
Lecture de CODISTAN
Lecture du tableau pour le seuil de schéma
Informations réservées par METOD sur fichier disque NTRAN
CREATION DE PROCHE EN PROCHE SUR LA SPHERE
Tests divers évitant des positions contradictoires
Création d'un point dans l'espace connaissant sa distance à 3 points déja créés
Levée de l'indétermination sur les deux positions possibles
Sous Programmes Internes
Chaque sous programme appelle un ou plusieurs des sous programmes des colonnes à sa droite
CPOINT VUCUBE CERCLE DISVU COOLIG
TERSEC PLANVU
COSINO
PSCALO PVECTO PSUSEG DIDIR
DISTA SOMME PRODUI COODIR COMME PADIR CGRAVI
COSA COOR COORP CORDN CORPN SCAPAD CADRE LISIGN ABSICE
Si le pt. de vue n'est pas donné par l'utilisateur définition automatique de l'oeil à une distance du centre de deux fois le rayon de la sphère
Calcul des positions des points dans le plan vu
Sous Programmes De La Librairie BENSON
IBENA PNUMA ECHEL
TRAS BTIRS BSCURV BETEPS
Impression d'un schéma grossier sur le listing avec les numéros des pt.
VISION BENSON
Tracé des cubes
Tracé des liaisons entre cubes
arcs et segments

a - Campagne GATE

Située à la latitude de Dakar, cette campagne est caractérisée par d'importants apports d'aérosols sahariens par les alizés.

La figure A correspond aux données de cette camapagne ; on y observe un certain nombre de paramètres isolés, soit parce que leur comportement géochimique est complexe, soit, peut-être, en raison de difficultés analytiques (contaminations...).

Il apparaît ensuite trois groupes d'explication aisée en raison de "traceurs" géochimiques, bien identifiés par ailleurs, qu'ils contiennent :

1) Al et Sc sont des traceurs des argiles d'origine terrigène (ici, l'érosion des sols sahariens). Fe, Mn, Co, Sm, Cr, Ce, Eu, Cs, La, Th sont rattachés à ce groupe. On remarque que Ba et V dont on connaît pourtant une composante terrigène sont ici à l'écart de ce groupe ce qui manifeste une autre composante pour ces éléments.

1) En pleine mer, Na est le traceur des sels d'origine marine (embruns). Cl et Mg sont liés à ce groupe.

3) On sait maintenant que la microcouche superficielle de l'océan est fortement enrichie en matière organique et métaux lourds. Buat-Ménard (1979) a montré que Zn et Sb dans l'aérosol à 10 m. de la surface pouvait servir de traceur de l'éjection par éclatement de bulles de microgouttelettes de cette microcouche. Ni, Cu_{total} et Pb_{total} sont ici rattachés à ce groupe alors que $Cu_{soluble}$ et $Pb_{soluble}$ sont isolés.

Le groupe bien différencié Rn, SO_4 , Se, Ca_T , Ca_S , Br est, en en revanche, une information nouvelle révélée par ce schéma. Son interprétation appelle quelques commentaires car les éléments qui le composent ont, en principe, des origines diverses :

- Rn est un gaz qui émane, pour sa plus grande partie, de l'écorce terrestre. Sa période de 3.8 jours en fait, en principe, un traceur des masses d'air d'origine continentale récente. (La position du radon près de la différence de température entre la mer et l'air nous rappelle l'existence d'une autre composante des variations du Rn, laquelle est d'ailleurs plus apparente lors de la campagne Rift (Lesty, 1981)).

- SO_4 possède une composante marine. Il peut également stagner dans la haute troposphère. (Voir en particulier Bonsang, 1980).

- Se a été identifié par Buat-Ménard et Arnold (1978) comme un traceur possible de la source volcanique.

- Ca peut avoir une origine marine ou terrigène.

- Br peut avoir une origine marine, liée , en particulier , à des mécanismes photosynthétiques (Rancher, 1978).

Pour expliquer ce regroupement, on peut invoquer un mécanisme commun de transport : ces éléments seraient amenés par bouffées depuis la haute troposphère. La plupart étant volatils (SO_4 , Se , Br) ou gazeux (Rn), leur histoire serait différente de celle des éléments terrigènes, bien que ceux-ci transitent aussi par la haute troposphère. Toutefois, il faudrait aussi expliquer pourquoi Ca qui n'est pas volatil, est lié à Se et Br. On pourrait remarquer aussi que Rn, Se, SO_4 sont en général associés aux phénomènes volcaniques et que Br dégaze

plus facilement des magmas en fusion riches en Ca. (Unni et Schilling, 1978). Ce groupe serait donc lié à l'aérosol volcanique . Il faudrait néanmoins expliquer pourquoi il affecte cette structure cyclique.

b - Données réunies des campagnes MIDLANTE, RIFT et GATE

Nous n'avons conservé sur la fig. B que les 33 variables qui ont pu être mesurées pour les trois campagnes. On retrouve, de façon distincte, les groupes terrigènes (SC...), marin (Na...) et de la microcouche superficielle (Zn ,Sb). Par contre, le groupe cyclique discuté plus haut se réduit au lien Se, SO_4 avec le souvenir de Rn mentionné sur ce lien. Le calcium est maintenant isolé, tandis que la composante marine de SO_4 est soulignée (lien avec Cl).

L'information nouvelle de ce schéma est d'abord la liaison de Rn avec les phases de la lune : un examen plus approfondi montre que Rn est plus exactement lié avec le coefficient de marée (variable que nous n'avions pas pensé à intégrer au schéma). On peut donc supposer que la production de Rn est en partie liée aux marées terrestres (qui aideraient au cheminement du gaz dans les failles de l'écorce ou bien aux marées atmosphériques facilitant peut-être le pompage du gaz . Notons qu'une périodicité de 28 jours dans les concentrations du radon a été observée dans l'Océan Indien entre 40° et 70° Sud par Lambert et al. (1970).

La seconde information nouvelle est le lien du mercure avec la latitude : Lorsque le navire remonte vers le Nord, on entre dans le régime atmosphérique d'Ouest. L'examen des cartes météorologiques montre que les pics de mercure correspondent tous au passage d'un front. On serait donc en présence d'une pollution d'origine nord-américaine encore perceptible aux abords des Açores.

Les positions des points sont calculées, fig. B, d'après les corrélations totales (du point de vue du navire). Considérons, fig. C , les *corrélations partielles* par rapport au Sc (variable représentative des argiles terrigènes) ; c'est-à-dire, faisons abstraction de la composante terrigène pour mettre en évidence les composantes secondaires des variations des éléments AL, Th, Fe, Cr, Ce, Mn, Co. Il apparaît une composante marine non seulement pour le fer, comme l'avait signalé Buat-Ménard (1979), mais également pour Cr, Ca et peut-être Ce. La composante marine de Fe, Cr et Ca a été réobservée depuis, en Mer de Chine (Lesty, 1981).

c - Traitement comparé par l'Analyse Factorielle des Correspondances

Arbib (1981) a repris les données qui correspondent à la figure B et les a traitées par l'Analyse Factorielle des Correspondances (cf. Benzécri, 1973). Les informations obtenues sont en gros similaires à celles qui sont décrites au § précédent pour la fig. B, mais elles ont nécessité l'examen de six figures (projection sur six plans factoriels) tirées du "tableau de notes d'intensité", de six figures tirées des "tableaux de contingence", et de dix tableaux numériques (tableaux des valeurs propres et tableaux des contributions).

Il apparaît donc que, pour les données océano-météorologiques du type discuté ici, les méthodes d'analyse factorielle sont d'un emploi beaucoup plus lourd que la méthode présentée au § 1, où l'information est rassemblée sur une seule figure, et où la "solidité" des relations est contrôlée et soulignée par des liaisons tracées par la machine.

Dans l'analyse factorielle des correspondances, les données, présentées au départ sous forme d'un tableau à deux entrées, ont dû être recodées et découpées en 5 classes d'effectifs approximativement égaux : c'est-à-dire que chaque mesure, qui comporte initialement plusieurs

chiffres significatifs, a été remplacée par un caractère logique pouvant prendre 5 modalités. Si les limites des classes sont mal choisies, la variabilité fine des mesures peut être effacée. C'est un inconvénient dans une étude géochimique où le souci premier est la précision des mesures à quelques décimales près (certaines substances, même à l'état de trace, ont un effet non négligeable sur l'environnement). C'est ainsi que l'analyse des correspondances n'a pas fait clairement apparaître la petite composante marine de Fe, Ca, Cr qui se superpose à la composante terrigène prépondérante de ces éléments (voir § IIb).

Conclusion

Lors d'une étude manuelle, le trop grand nombre de mesures conduit à une sélection toujours un peu arbitraire des paramètres dont l'analyse est jugée prioritaire. Les schémas informatiques présentés ici permettent d'emblée, au contraire, une vue globale des paramètres disponibles. Ils ont attiré notre attention sur des phénomènes géochimiques importants, qui, une fois découverts, se révèlent très apparents sur les données, comme le lien du radon avec le coefficient de marée, ou l'association du mercure avec le passage des fronts . D'autres phénomènes, de nature plus complexe, comme le groupe Rn, SO_4 ,Se, Ca, Br ne pouvaient être mis en lumière que par ce type de figure. Enfin, la possibilité de séparer les diverses composantes a mis en évidence la composante marine du chrome associée à celle du fer.

NOTE :

Le programme COCORICO figure dans la bibliothèque mathématique du centre de calcul de P.S.I. (Université d'Orsay. Il est accessible de tout autre ordinateur par le système TRANSPAC. un mode d'emploi est disponible.

Remerciements

Nous remercions R. Chesselet qui a suscité et encouragé ce travail, J.P. Benzécri pour ses précieux conseils, et C. Lambert pour ses critiques et suggestions.

BIBLIOGRAPHIE

ARBIB L. : Sur l'étude de composants chimiques d'aérosols marins et de l'influence de l'environnement écologique sur leurs variations . Thèse de 3° cycle, université P. et M. Curie, (PARIS VI), 1981.

BENZECRI J.P.: L'Analyse des Données - Tomes I & II - DUNOD, Paris (1973)

BONSANG G. : Cycle atmosphérique du soufre d'origine marine. Thèse d'état, université de Picardie, 1980.

BUAT-MENARD P. and ARNOLD M. : The heavy metals chemistry of atmospheric particulate matter emitted by mount Etna volcano. Geophys. Res. lett. 5, 245-248, 1978.

BUAT-MENARD P.:Influence de la retombée atmosphérique sur la chimie des métaux traces dans la matière en suspension dans l'Atlantique Nord . Thèse d'état, université PARIS VII, 1979.

LAMBERT G., POLIAN G., et TAUPIN D. : Existence of periodicity in radon concentration at lower latitude between 40° and 70° south.. Journ. Geophys. Res., 75, 2341-2345, 1970.

LESTY M. : Analyse informatique de données géochimiques sur les échanges air-mer. Thèse d'état, université PARIS VII, 1981.

RANCHER J. : Contribution à l'étude du cycle du chlore, du brome et de l'iode dans l'atmosphère au-dessus de l'océan. Thèse de spécialité , université PARIS VII, 170 p., 1978.

UNNI C.K. and SCHILING J.G. : Cl and Br degassing by volcanism along the Reykjanes Ridge and Iceland. Nature vol. 272, 2 march 1978.

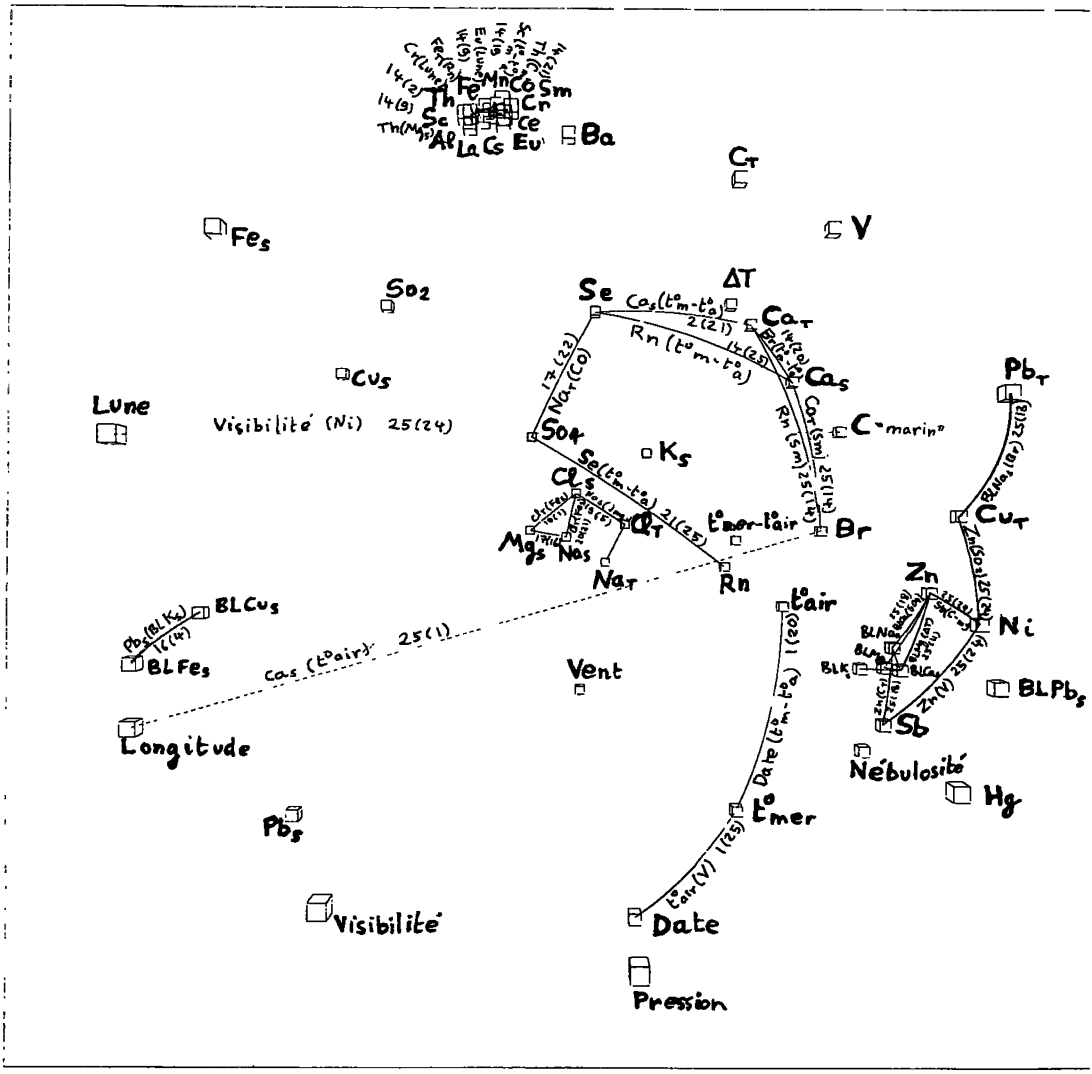


Figure A : CAMPAGNE GATE

Les positions des points sont calculées d'après les corrélations totales. Le tri méthodique des corrélations partielles et de l'influence de chaque échantillon a mis en évidence les liens les plus "solides". Certaines variables, relativement éloignées s'assemblent (ex.: le groupe en forme de boucle Rn, SO_4 ,Se, Ca_T ,Ca_S ,Br). D'autres quoique proches ne sont pas liées (ex. : K_s n'est pas liée au groupe des sels marins (Na...)).

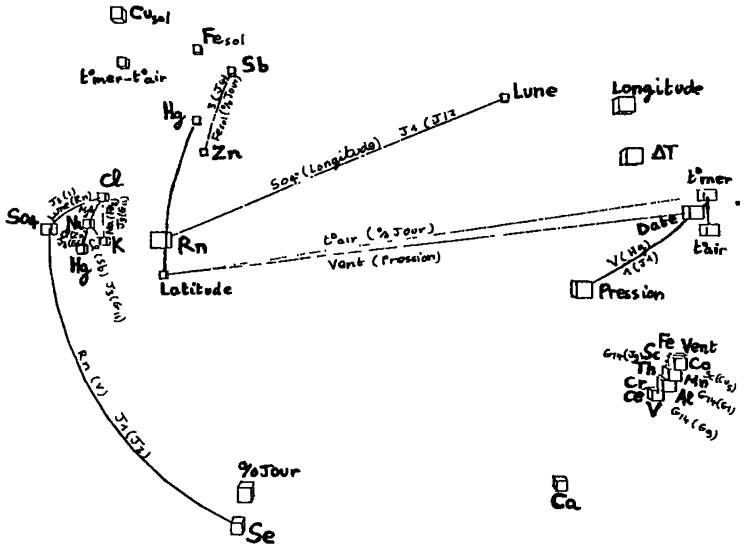


FIGURE B MIDLANTE + RIFT + GATE

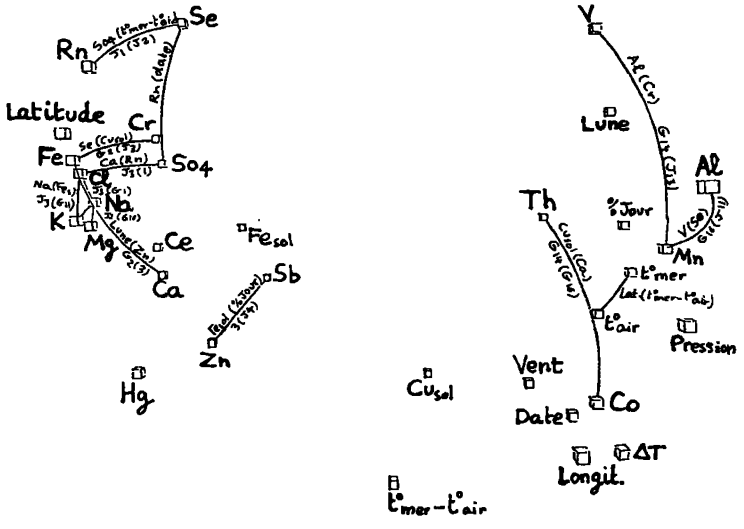
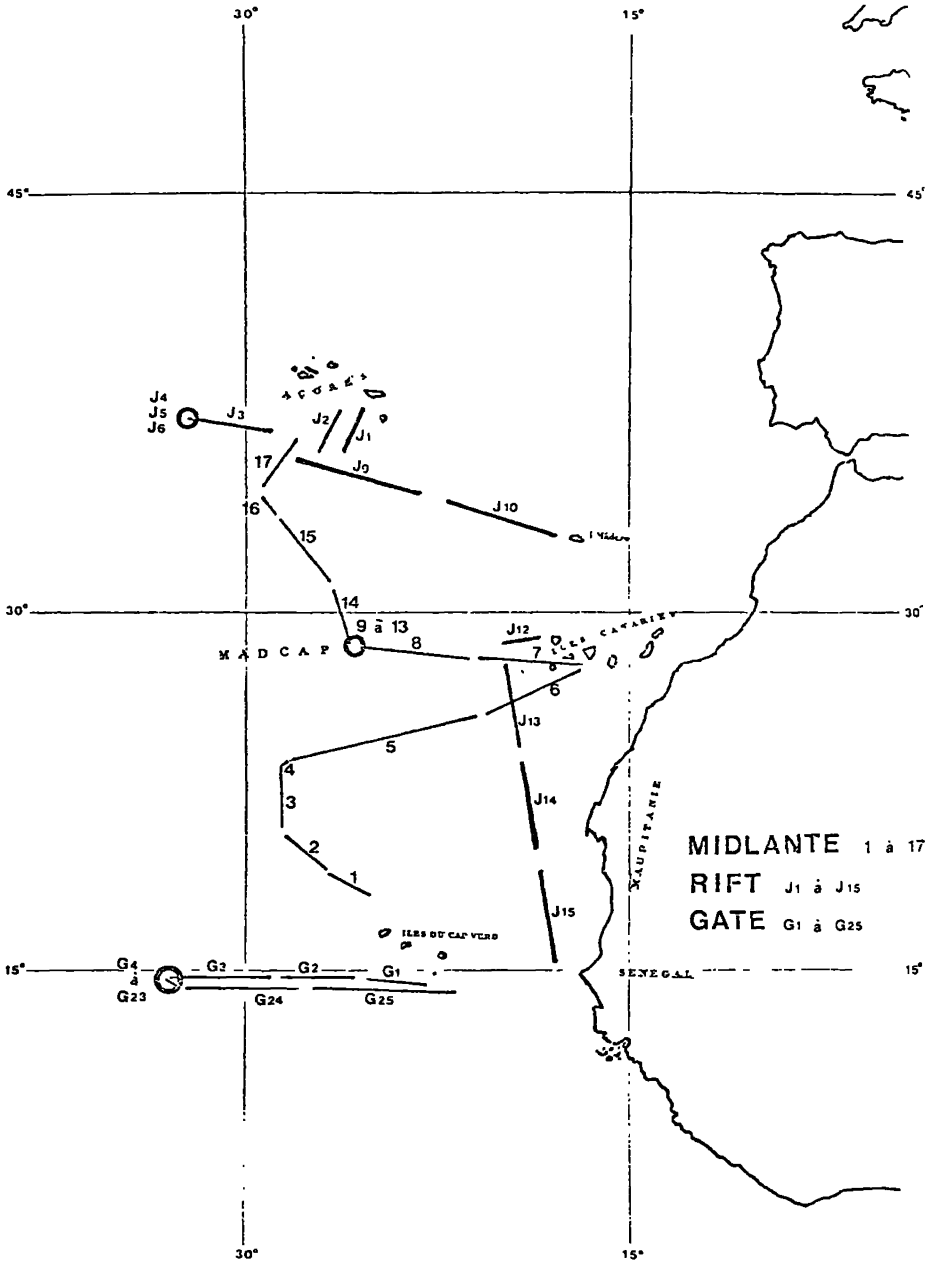


Figure C : MIDLANTE++RIFT+GATE

Influence terrigène retirée (corr. partielles/Sc)

FIGURE D

Localisation des prélèvements d'aérosols
MIDLANTE (avril-mai 1974)
RIFT (mai-juin 1974)
GATE (juin-juillet 1974)