

REVUE DE STATISTIQUE APPLIQUÉE

J. PAGES

L. LE CALVEZ

Analyse statistique d'une formation limoneuse en vue de préciser son origine

Revue de statistique appliquée, tome 28, n° 2 (1980), p. 45-60

http://www.numdam.org/item?id=RSA_1980__28_2_45_0

© Société française de statistique, 1980, 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/>

ANALYSE STATISTIQUE D'UNE FORMATION LIMONEUSE EN VUE DE PRÉCISER SON ORIGINE

J. PAGES * et L. LE CALVEZ **

1. — INTRODUCTION

Le Massif Armoricaïn est constitué de roches très variées. Par contre, les sols qui les surmontent sont, du point de vue granulométrique, assez uniformes. Ils renferment relativement peu d'argile et de sables, mais, beaucoup de limons. (Particules fines comprises entre 2 et 50 microns).

Ce contraste entre la diversité des roches et l'homogénéité des sols a induit de nombreux travaux sur l'origine des limons ; proviennent-ils :

- des roches sous-jacentes ?
- d'un ou de plusieurs apports éoliens (Loess) ?
- ont-ils une origine mixte (mélange) ?

L'étude au microscope met en évidence, dans les sols limoneux de Bretagne centrale, à la fois des minéraux provenant des roches locales et des minéraux absents de ces roches.

Dans cette région, les limons ont donc une origine mixte. Les analyses chimiques confirment cette double origine.

Pour compléter cette étude qualitative, il convient d'identifier les constituants initiaux du mélange et d'évaluer leurs proportions relatives dans chaque échantillon.

Sur le terrain, il existe, naturellement, des variations entre les proportions des constituants du mélange. Celles-ci entraînent des variations dans les compositions granulométriques des sols.

Une analyse statistique de ces variations devrait permettre de déterminer la composition des constituants initiaux, à condition que les sols n'aient pas subi de transformations postérieures au mélange trop importantes — la présence de minéraux très fragiles (Hornblende) permet de penser que cette condition est réalisée.

*
* *

Pour cette étude, trois séquences, situées sur les trois principales roches de la région — grès, granite, schistes — ont été retenues. Dans chaque séquence, plusieurs échantillons ont été prélevés et analysés, du point de vue de leur composition

(*) Statistique et Informatique ENSAR
(**) Sciences du Sol INRA ENSAR.

granulométrique⁽¹⁾. En outre, on dispose de la composition chimique de quelques échantillons. L'analyse statistique de ces données chimiques étant conduite selon les mêmes principes que celle des données granulométriques, nous nous référons dans le texte essentiellement à ces dernières, qui sont les plus fréquentes car beaucoup moins coûteuses à obtenir.

2. — LE MODÈLE DE MÉLANGE

2.1. — Notations

Soit :

- X_{ij} , la mesure de la variable j pour le prélèvement i .
- $\bar{I}$, le nombre de prélèvements $i = 1, \bar{I}$
- $\bar{J}$, le nombre de classes granulométriques $j = 1, \bar{J}$
- X_{iJ} , l'ensemble des mesures pour le prélèvement i :

$$X_{iJ} = \{X_{ij} \mid j = 1, \bar{J}\}$$

- X_{IJ} , l'ensemble des mesures pour tous les prélèvements :

$$X_{IJ} = \{X_{ij} \mid i = 1, \bar{I} ; j = 1, \bar{J}\}$$

Ces notations sont récapitulées sur la figure 1 :

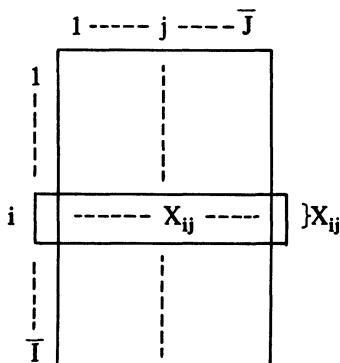


FIG. 1. — Récapitulation des notations

(1) Un sol est constitué de particules de différentes tailles. Classiquement, on distingue 8 classes de tailles.

Argile :	particules comprises entre	0 et	2 microns
Limon fin :	" "	2 et	20 microns
Limon grossier :	" "	20 et	50 microns
Sable 1 :	" "	50 et	100 microns
Sable 2 :	" "	100 et	200 microns
Sable 3 :	" "	200 et	500 microns
Sable 4 :	" "	500 et	1 000 microns
Sable 5 :	" "	1 000 et	2 000 microns

La composition granulométrique d'un échantillon indique la masse totale des particules de chaque classe ; elle s'exprime généralement en pourcentage.

2.2. — Le modèle de mélange

Si la formation limoneuse est issue du mélange entre deux matériaux, alors, la composition granulométrique des prélèvements que l'on y fait doit présenter une certaine structure. Nous appelons cette structure "modèle de mélange".

Ce modèle est composé de deux propositions :

M1 : il existe deux constituants initiaux Y et Z ; nous noterons leur composition granulométrique respectivement

$$Y_j = \{Y_j | j = 1, \bar{J}\} \text{ et } Z_j = \{Z_j | j = 1, \bar{J}\} \quad \text{où } Y_j \text{ (resp } Z_j)$$

est la teneur en éléments appartenant à la classe j pour le constituant initial noté Y (resp. Z).

M2 : chaque prélèvement i observé est un mélange des constituants Y et Z ; soit, en appelant P_i la proportion du constituant Y :

$$\forall i = 1, \bar{I} ; j = 1, \bar{J} : X_{ij} = P_i Y_j + (1 - P_i) Z_j$$

En réalité, la proposition M2 ne sera jamais vérifiée exactement pour deux raisons :

- X_{ij} est une mesure sujette à erreur ;
- il s'écoule nécessairement un certain temps entre la réalisation du mélange et notre observation. Aussi, les matériaux issus du mélange vont subir une transformation pédologique, et c'est le résultat du mélange *et* de la transformation post-mélange que l'on observe.

Ceci conduit à une nouvelle formulation de la proposition M2 :

$$\forall i = 1, \bar{I} ; j = 1, \bar{J} : X_{ij} = P_i Y_j + (1 - P_i) Z_j + A_{ij}$$

où A_{ij} est une variable aléatoire qui représente à la fois les erreurs de mesure et la transformation post-mélange.

Dans ce modèle, X_{ij} est une variable aléatoire, P_i , Y_j , Z_j sont des paramètres. Pour estimer ces paramètres, quelle que soit la méthode utilisée, il est nécessaire de lui adjoindre une proposition supplémentaire :

$$M3 : \quad \forall i = 1, \bar{I} ; j = 1, \bar{J} : E[A_{ij}] = 0$$

Sans cette hypothèse, il sera impossible de distinguer les influences respectives sur X_{ij} de P_i , Y_j , Z_j d'une part, de A_{ij} d'autre part. Sauf cas très particulier, il est certain que M3 n'est jamais strictement vérifiée, mais il est vraisemblable que les situations naturelles ne s'en écartent pas trop. (Nous citerons les arguments mis en jeu par cette affirmation après la présentation de la méthode que nous proposons, ce qui permettra de mieux en apprécier la portée (§ 343)).

Il convient de remarquer que la méthodologie que nous allons développer à partir d'un exemple, est de portée assez générale. Les modèles de mélange ont déjà fait l'objet de plusieurs travaux. Généralement, les composants initiaux sont connus (e.g. P. CAZES et P.Y. TURPIN [1]), et l'estimation des pourcentages s'effectue au moyen d'une régression sous contraintes. Les situations où les composants sont inconnus sont plus rares, mais existent : à propos des margarines, H. MARTENS et al. [4] ont proposé une méthode très proche de la nôtre.

2.3. — Le problème posé en terme de modèle de mélange

Soit :

- $\hat{P}_i, \hat{Y}_j, \hat{Z}_j$ des estimations de P_i, Y_j, Z_j .
- $\hat{X}_{ij} = \hat{P}_i \hat{Y}_j + (1 - \hat{P}_i) \hat{Z}_j$ l'estimation de X_{ij} .
- $\hat{P}_I = \{\hat{P}_i | i = 1, \bar{I}\}$
- $\hat{Y}_J = \{\hat{Y}_j | j = 1, \bar{J}\}$
- $\hat{Z}_J = \{\hat{Z}_j | j = 1, \bar{J}\}$
- $\hat{X}_{I,J} = \{\hat{X}_{ij} | i = 1, \bar{I} ; j = 1, \bar{J}\}$

Les données observées seront compatibles avec le modèle de mélange si l'on peut trouver $\hat{P}_I, \hat{Y}_J, \hat{Z}_J$ tels que $\hat{X}_{IJ}$ ne soit pas trop différent de X_{IJ} . Pour cela, il est nécessaire de définir un critère pour mesurer l'écart entre $\hat{X}_{IJ}$ et X_{IJ} . Le plus naturel est le critère des moindres carrés, soit :

$$Q = \sum_i \sum_j (\hat{X}_{ij} - X_{ij})^2$$

Le problème pédologique peut donc être formulé ainsi : trouver une estimation de P_i, Y_j, Z_j qui minimise Q ; si la valeur calculée de Q est grande, on refusera l'hypothèse selon laquelle la formation limoneuse étudiée est issue du mélange de deux matériaux ; si la valeur calculée de Q est petite, on acceptera l'hypothèse et simultanément, le problème de l'estimation des paramètres sera résolu.

Une difficulté se présente pour préciser ce que l'on appelle une valeur de Q grande ou petite. Il est possible d'envisager un test statistique, mais il convient plutôt ici de profiter de la nature particulière du problème posé. En effet, si l'hypothèse de mélange est juste, alors, les composants des matériaux de départ ne peuvent être quelconques : l'un est nécessairement un produit de l'altération du substrat, et l'autre un apport éolien. Or, nous disposons d'idées a priori précises sur ces compositions. Le test de l'hypothèse de mélange consistera donc à vérifier que les estimations de Y_j et Z_j qui minimisent Q sont cohérentes aux yeux du pédologue. Nous avons ici un critère de validation externe (par rapport aux données) qui est très puissant en ce sens qu'il est très improbable que nous acceptions l'hypothèse de mélange si elle est fausse.

De toutes façons, pour le pédologue, le problème du test de l'hypothèse de mélange est relativement secondaire. En effet, nous avons signalé qu'il existe des arguments qualitatifs (présence de tel ou tel minéral par exemple) incitant à penser que telle formation limoneuse est issue d'un mélange de deux matériaux. C'est l'estimation des paramètres qui est le point essentiel.

3. — ESTIMATION DES PARAMÈTRES DU MODÈLE DE MÉLANGE

3.1. — Remarque préliminaire

Nous sommes en présence d'un système de $\bar{I} \times \bar{J}$ équations du type :

$$X_{ij} = \hat{P}_i \hat{Y}_j + (1 - \hat{P}_i) \hat{Z}_j + \hat{A}_{ij} = \hat{X}_{ij} + \hat{A}_{ij}$$

Le problème consiste à estimer les $\hat{P}_i$, $\hat{Y}_j$ et $\hat{Z}_j$ en minimisant $Q = \sum_{ij} \hat{A}_{ij}^2$

Les paramètres à estimer doivent vérifier certaines contraintes :

$$C1 : 0 \leq \hat{P}_i \leq 1 \quad i = 1, \bar{I}$$

$$C2 : 0 \leq \hat{Y}_j \leq 100 \quad j = 1, \bar{J}$$

$$C3 : 0 \leq \hat{Z}_j \leq 100 \quad j = 1, \bar{J}$$

$$C4 : \sum_j \hat{Y}_j = 100$$

$$C5 : \sum_j \hat{Z}_j = 100$$

Ce problème peut s'interpréter comme une régression non linéaire sous contraintes, avec $\bar{I} + 2\bar{J}$ paramètres et $\bar{I} \times \bar{J}$ équations. De ce point de vue, sa résolution est extrêmement complexe. En fait, nous proposons plus loin une solution très naturelle et relativement simple.

3.2. — Interprétation géométrique

On peut considérer chaque X_{ij} comme un élément de l'espace R^J . Supposons trouvées les estimations $\hat{P}_i$, $\hat{Y}_j$ et $\hat{Z}_j$. On remarquera que $\hat{Y}_j$ et $\hat{Z}_j$ appartiennent à R^J et que $\hat{X}_{ij} = \hat{P}_i \hat{Y}_j + (1 - \hat{P}_i) \hat{Z}_j$ appartient au segment qui relie $\hat{Y}_j$ et $\hat{Z}_j$.

En outre Q est la somme des carrés des distances (au sens de la distance euclidienne usuelle) entre $\hat{X}_{ij}$ et X_{ij} .

$$Q = \sum_i \sum_j (\hat{X}_{ij} - X_{ij})^2 = \sum_i d^2(\hat{X}_{ij}, X_{ij})$$

Ainsi, $\hat{X}_{ij}$ peut s'interpréter comme la projection de X_{ij} sur le segment reliant $\hat{Y}_j$ et $\hat{Z}_j$. (S'il n'en était pas ainsi, Q ne serait pas minimisé).

Ces résultats sont récapitulés dans la figure 2.

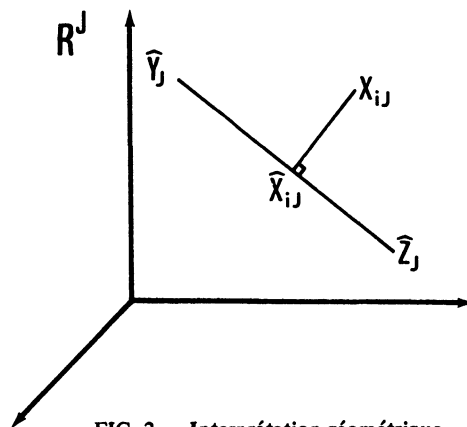


FIG. 2. — Interprétation géométrique du modèle de mélange

Nous sommes ainsi conduits à une nouvelle formulation du problème d'estimation : rechercher les deux points $\hat{Y}_J$ et $\hat{Z}_J$ de R^J tels que l'inertie de l'ensemble des points X_{iJ} par rapport à l'axe reliant $\hat{Y}_J$ et $\hat{Z}_J$ soit minimale.

D'après cette interprétation, il apparaît que, si $\hat{Y}_J$ et $\hat{Z}_J$ sont trouvés, il existe d'autres points $\hat{Y}'_J$ et $\hat{Z}'_J$ qui conduisent à la même valeur du critère Q : ces points sont ceux de la droite passant par $\hat{Y}_J$ et $\hat{Z}_J$. Il existe donc une double infinité de solutions.

En réalité, tous les points de cette droite ne conviennent pas : on ne prendra en compte que ceux qui satisfont les contraintes. On remarquera que les contraintes C4 et C5 sont automatiquement vérifiées. En effet, tous les points X_{iJ} vérifient $\sum_j X_{ij} = 100$. L'ensemble du problème se situe donc dans l'hyperplan de R^J contenant les points dont la somme des composantes vaut 100. (Figure 3).

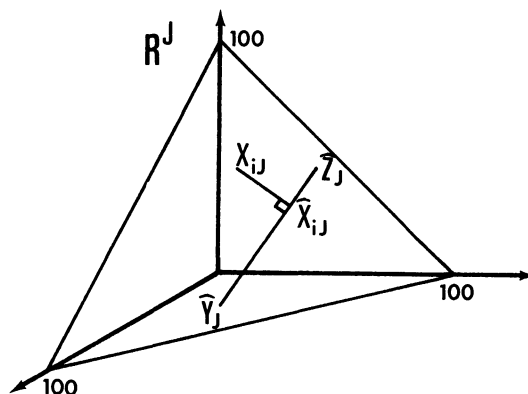


FIG. 3. — Interprétation géométrique du modèle de mélange mettant en évidence les contraintes C4 et C5

3.3. — Méthode proposée

La méthode que nous proposons comporte 3 phases :

Phase 1 : détermination de la droite contenant les $\hat{Y}_J$ et les $\hat{Z}_J$.

Phase 2 : détermination des segments de cette droite qui contiennent les $\hat{Y}_J$ et les $\hat{Z}_J$ satisfaisant les contraintes.

Phase 3 : détermination de $\hat{P}_I$ compte tenu de la multiplicité des $\hat{Y}_J$ et $\hat{Z}_J$.

Phase 1 : détermination de la droite contenant les $\hat{Y}_J$ et les $\hat{Z}_J$.

Le problème consiste à rechercher dans R^J , l'axe affine Δ tel que l'inertie par rapport à Δ du nuage N_I des points X_{iJ} soit minimum. D'après le théorème de Huyghens, cet axe passe par le centre de gravité de N_I . Il conviendra donc de réaliser une analyse en composantes principales (A.C.P.) non normée du tableau X_{iJ} : l'axe Δ est le premier axe fourni par cette analyse.

A ce niveau, se pose le problème de l'identification de l'axe Δ avec une opposition substrat-apport. Sous réserve de confirmation lorsque l'on disposera des estimations, il est possible d'accepter l'identification d'après l'examen des corrélations entre les variables initiales et Δ . (Les oppositions de variables mises en évidence par Δ correspondent-elles aux idées a priori que l'on a sur la différence entre le substrat et l'apport ?).

De ce point de vue, le pourcentage d'inertie associé à Δ n'est pas un indicateur de la validité de l'hypothèse de mélange, mais plutôt un indicateur de l'influence de l'origine des prélèvements sur leur composition actuelle.

Phase 2 : détermination des $\hat{Y}_J$ et $\hat{Z}_J$.

Cette phase est décrite par le schéma de la figure 4, qui représente l'axe Δ fourni par l'analyse en composantes principales.

On détermine d'abord le segment S_c , ensemble des points de Δ vérifiant les contraintes C2 et C3 : on obtient ainsi les points $\hat{Y}_{0J}$ et $\hat{Z}_{0J}$.

En projetant le nuage N_I sur Δ , on détermine le segment S_e , plus petit segment contenant les projections des X_{iJ} .

Soit $\hat{X}_{yJ}$ et $\hat{X}_{zJ}$ les extrémités de S_e .

Pour que la contrainte C1 soit vérifiée, $\hat{Y}_J$ et $\hat{Z}_J$ doivent être extérieurs à S_e .

Les trois contraintes seront donc vérifiées dans l'ensemble complémentaire de S_e dans S_c ; cet ensemble comprend deux segments : l'un, S_y , contient les valeurs possibles de $\hat{Y}_J$, l'autre S_z , contient les valeurs possibles de $\hat{Z}_J$.

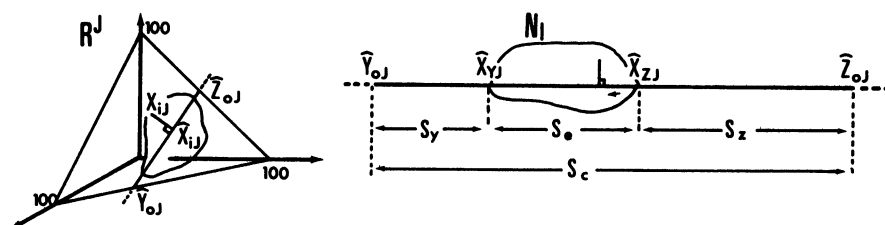


FIG. 4. — Détermination des ensembles des valeurs de $\hat{Y}_J(S_y)$ et de $\hat{Z}_J(S_z)$

Phase 3 détermination de $\hat{P}_I$.

La valeur de $\hat{P}_I$ dépend de celles de $\hat{Y}_J$ et $\hat{Z}_J$. Comme ces dernières ne sont pas uniques, on pourra proposer plusieurs valeurs de $\hat{P}_I$.

Toutefois, les différentes estimations $\hat{P}_I$ ont en commun l'ordre de leurs composantes. En effet, pour chaque $\hat{P}_I$, les valeurs $\hat{P}_i$ respectent l'ordre des $\hat{X}_{iJ}$ sur l'axe factoriel. La méthode proposée fournit de manière unique un classement des échantillons selon leur plus ou moins grande proportion en matériel d'apport.

Si le classement ne suffit pas, on pourra calculer les valeurs extrêmes de $\hat{P}_I$. En effet, pour chaque individu i , la valeur de $\hat{P}_i$ (proportion du matériel Y) la plus grande sera obtenue en choisissant comme valeur de $\hat{Y}_J$ la valeur la plus proche de l'échantillon ($\hat{X}_{yJ}$) et comme valeur de $\hat{Z}_J$ la plus éloignée ($\hat{Z}_{0J}$). Réciproquement, la valeur de $\hat{P}_i$ la plus petite sera obtenue en choisissant pour $\hat{Y}_J$ et $\hat{Z}_J$ le couple ($\hat{Y}_{0J}$, $\hat{X}_{zJ}$). Ce procédé est rigoureux, mais risque de conduire à des intervalles étendus de façon exagérément pessimiste ; pratiquement, on pourra se contenter de l'ordre de grandeur de $\hat{P}_I$, calculé à partir du couple :

$$((\hat{Y}_{0J} + \hat{X}_{yJ})/2 ; (\hat{Z}_{0J} + \hat{X}_{zJ})/2) .$$

3.4. — Discussion sur la méthode

3.4.1. — Les estimations des composants initiaux

Nous avons indiqué que les $\hat{Z}_J$ et $\hat{Y}_J$ n'étaient pas uniques : tous les points d'un segment de droite convenaient. Ainsi, pratiquement nous calculerons $\hat{Z}_{0J}$ et $\hat{X}_{zJ}$. Il convient de remarquer que ces bornes ne constituent pas un intervalle de confiance au sens usuel du terme. Chaque borne est elle-même une estimation ponctuelle, et notre modèle ne contient aucune hypothèse probabiliste.

La longueur des segments $(\hat{Z}_{0J}, \hat{X}_{zJ})$ et $(\hat{X}_{yJ}, \hat{Y}_{0J})$ dépend de la présence ou non dans l'échantillon de prélèvements qui sont proches des limites des contraintes $0 < X_{ij} < 100$. Ceci est illustré figure 5 à partir d'un exemple très simplifié où seules deux classes interviennent : la situation de la figure 5a conduit à des intervalles de longueur beaucoup plus petites que ceux de la figure 5b, et ceci indépendamment de la qualité de l'ajustement.

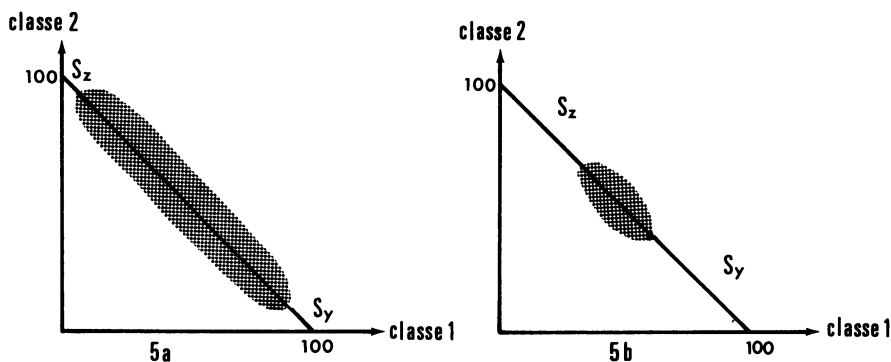


FIG. 5. — Influence de l'échantillonnage sur la longueur des intervalles contenant les estimations de l'apport et du substrat. S_y et S_z

Pratiquement, cette condition implique, dans l'hypothèse de mélange que :

a) Il existe dans l'échantillon des prélèvements très proches des composants initiaux.

b) Les composants initiaux sont tels qu'ils saturent certaines contraintes (certaines de leur composantes sont nulles par exemple).

Il est possible d'agir sur ces deux conditions :

a) en choisissant l'échantillon de telle façon que certains prélèvements soient très proches de l'apport et d'autres très proches du substrat. Ceci est possible par de simples observations sur le terrain ;

b) en modifiant le système de contraintes. En effet, nous avons imposé jusqu'ici aux différents pourcentages d'être compris entre 0 et 100. Or, on dispose généralement d'idées a priori sur le produit de la dégradation de la roche mère, laquelle est connue.

Ce dernier point revêt une grande importance pratique car un système de contraintes trop général risque de conduire à des estimations très vagues. Toutefois, il convient de n'ajouter des contraintes qu'avec prudence.

Pratiquement, on pourra appliquer l'algorithme suivant : calcul de $\hat{Y}_j$ et $\hat{Z}_j$ sans contrainte supplémentaire

faire

examen de $\hat{Y}_j$ et $\hat{Z}_j$

tant que (il existe une teneur "irréaliste")

calcul de $\hat{Y}_j$ et $\hat{Z}_j$ avec une contrainte supplémentaire à propos de la teneur la plus irréaliste.

refaire

3.4.2. — Robustesse

Lors de la présentation du modèle, nous avons indiqué que l'hypothèse $H : E(A_{ij}) = 0 \quad i = 1, I ; j = 1, J$ n'est jamais strictement vérifiée mais que les données réelles ne s'en écartent probablement pas trop.

En ce qui concerne les données granulométriques, l'écart le plus important à l'hypothèse H est dû au lessivage de l'argile. Or, on pourra considérer que ce dernier n'influe pas trop sur la qualité des estimations si :

a) Le lessivage porte sur de faibles quantités.

b) Le lessivage agit "indépendamment" de la teneur en apport en ce sens qu'il est possible de rencontrer des échantillons riches en substrat aussi bien à de grandes qu'à de petites profondeurs. Cette condition est réalisable du fait de l'érosion lorsque l'on étudie un terrain présentant des pentes.

c) Le substrat et l'apport ne diffèrent pas essentiellement du point de vue de l'argile.

3.4.3. — Les données chimiques

La présentation qui vient d'être effectuée ne considère que les données granulométriques. Or, pour certains prélèvements, nous disposons en plus d'analyses chimiques. (Ces analyses sont coûteuses, et n'ont été réalisées que sur une dizaine de prélèvements).

Naturellement, toute la présentation précédente reste applicable pour les données chimiques. A partir d'une séquence de prélèvements, on pourra donc réaliser 2 traitements statistiques ; celui qui concerne les données chimiques devra être surtout considéré comme un complément de l'autre puisqu'il ne s'appuie que sur un nombre restreint de prélèvements ; mais ce complément joue un rôle important dans la validation des résultats puisqu'il conduit à des estimations de P_i qu'il est facile de comparer avec celles calculées à partir des données granulométriques.

Si les deux estimations de P_i sont voisines⁽²⁾, alors on peut envisager l'estima-

(2) On a vu que l'information contenue dans P_i consistait surtout dans l'ordre des P_i . On pourra donc mesurer la proximité entre les deux estimations à l'aide d'un coefficient de corrélation de rang.

tion des compositions chimiques des constituants initiaux. Trois démarches sont possibles :

D1 : estimation à partir des seules données chimiques.

D2 : estimation à partir des données chimiques et granulométriques qui concernent les prélèvements pour lesquels on dispose des deux types de données.

D3 : estimation restreinte à Y_j et Z_j à partir des données chimiques. P_1 est alors considéré comme connu et on utilise les valeurs $\hat{P}_1$ issues des données granulométriques.

Pour comparer ces trois démarches, nous nous référerons à l'ensemble des données (granulométrie + chimie) tel qu'il est présenté figure 6.

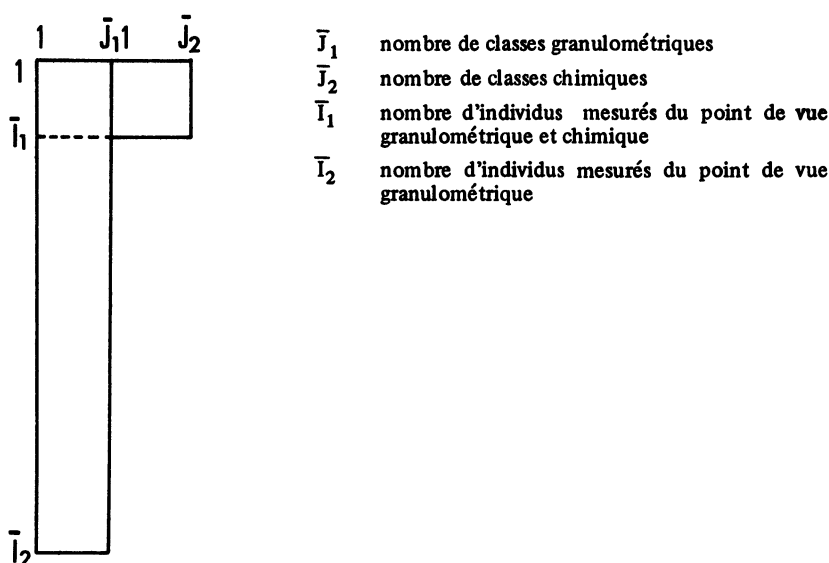


FIG. 6. – Représentation des données granulométriques et chimiques

La démarche D1 reprend exactement ce que nous avons vu pour la granulométrie, appliquée cette fois à la chimie, c'est-à-dire à un tableau de taille $\bar{I}_1 \times \bar{J}_2$.

La démarche D2 recouvre une A.C.P. réalisée sur les individus possédant à la fois des données granulométriques et chimiques, c'est-à-dire sur un tableau de taille : $\bar{I}_1 \times (\bar{J}_1 + \bar{J}_2)$. Toute la méthodologie précédente s'applique à condition de remplacer les contraintes C4 et C5 par

$$\sum_{j=1}^{\bar{J}_1} Y_j = \sum_{j=1}^{\bar{J}_1} \hat{Z}_j = \sum_{j=\bar{J}_1+1}^{\bar{J}_1+\bar{J}_2} Y_j = \sum_{j=\bar{J}_1+1}^{\bar{J}_1+\bar{J}_2} \hat{Z}_j = 100$$

La démarche D3 comprend deux étapes :

a) A.C.P. sur les données granulométriques (à partir du tableau de taille $\bar{I}_2 \times \bar{J}_1$). Cette A.C.P. nous fournit $\hat{P}_1$.

b) On ré-écrit le modèle de mélange pour la chimie en supposant P_I connu (P_I est fourni par la granulométrie) :

$$\forall i = 1, \bar{I}_1 \quad ; \quad j = 1, \bar{J}_2 \quad X_{ij} = P_i Y_j + (1 - P_i) Z_j + A_{ij} = P_i (Y_j - Z_j) + Z_j + A_{ij}$$

On reconnaît là un modèle linéaire, dont on pourra estimer les paramètres au moyen de $\bar{J}_2$ régressions simples. Les paramètres à estimer, Y_j et Z_j , étant des pourcentages, doivent vérifier les contraintes habituelles de positivité et de sommation à 100. En fait, des considérations empiriques fondées sur la nature particulière des données :

$$- \forall i = 1, \bar{I}_1 \quad ; \quad j = 1, \bar{J}_2 : 0 \leq X_{ij} < 100$$

- $\forall i = 1, \bar{I}_1 : 0 \leq P_i \leq 1$; les P_i "remplissant" à peu près $(0, 1)$ (cf. § 342, condition a))

laissent supposer que, si le modèle linéaire convient, ces contraintes seront pratiquement vérifiées automatiquement.

Nous préférons la démarche D3 : elle conduit à une valeur unique de $\hat{P}_I$ compatible avec les données granulométriques et chimiques et qui présente le plus fort rapport — nombre de données/nombre de paramètres estimés — dans l'estimation des compositions chimiques des constituants initiaux.

Si toutefois les estimations des compositions chimiques par D3 ne satisfont pas les contraintes, et si l'on ne dispose pas des programmes de régression adaptés, la démarche D2 convient parfaitement.

3.4.4. — La mise en œuvre

Elle est très simple et s'effectue en 2 temps. Tout d'abord, on réalise une A.C.P. non normée (à l'aide d'un programme standard) sur le tableau X_{IJ} . Les résultats sont, d'une part imprimés, et d'autre part, stockés en fichiers permanents. Dans un deuxième temps, ces fichiers sont introduits en entrée dans un programme spécifique qui édite $\hat{Y}_{0J}$, $\hat{X}_{YJ}$, $\hat{X}_{ZJ}$, $\hat{Z}_{0J}$, les différentes valeurs de $\hat{P}_I$ et $\hat{X}_{IJ}$.

4. — EXEMPLE D'APPLICATION

La présentation qui suit est très sommaire : elle vise à montrer comment les considérations précédentes peuvent s'appliquer à des données naturelles et non à exploiter complètement un ensemble de données. Ainsi, bon nombre d'idées, soulevées à propos de cette étude seront seulement citées et non argumentées. Dans le même esprit, nous présenterons l'analyse des données granulométriques d'une unique séquence, en ne mentionnant que dans la conclusion les analyses se fondant sur les données chimiques et/ou sur les autres séquences. Le lecteur désireux d'approfondir nos conclusions pédologiques se reportera avec profit à la thèse de L. LE CALVEZ (cf. [3]).

4.1. — Les données

Les échantillons proviennent d'un sol limoneux à limono-sableux, qui se situe le long d'un versant, au-dessus d'une arène granitique sableuse, au Sourn, près de Pontivy (Morbihan).

Les points de prélèvements —profils— sont régulièrement espacés le long de la pente. A chaque point de prélèvement, on recueille des échantillons de sol à différentes profondeurs ; 66 échantillons ont été recueillis et analysés du point de vue granulométrique, 10 d'entre eux ont donné lieu à une analyse chimique.

4.2. — Les résultats

Phase 1 : détermination de la droite contenant les $\hat{Y}_J$ et les $\hat{Z}_J$. Cette phase s'appuie sur une analyse en composantes principales non normée du tableau X_{IJ} . Les pourcentages d'inertie des différents axes sont les suivants :

- 1^{er} axe : 79,8 %
- 2^e axe : 10,3 %
- 3^e axe : 5,9 %
- 4^e axe : 3,1 %

Il se dégage un axe prépondérant qui explique 80 % des variations des mesures.

Les corrélations des variables à cet axe sont de :

- Argile (A) : 0,310
- Limon fin (L.F.) : 0,766
- Limon grossier (L.G.) : 0,981
- Sable très fin (S1) : -0,750
- Sable fin (S2) : -0,940
- Sable moyen (S3) : -0,927
- Sable grossier (S4) : -0,961
- Sable très grossier (S5) : -0,861

A l'exception de l'argile, toutes les variables sont fortement corrélées à l'axe. Celui-ci oppose les limons, qui présentent des corrélations positives, aux sables, dont les corrélations sont négatives. Il sépare donc les échantillons dont la composition est voisine de celle des arènes (sables), des échantillons dont la composition apparaît a priori proche de celle des loess. Aussi, cet axe peut s'interpréter en terme de mélange. Il est donc possible, à partir de cet axe, de procéder à l'estimation de Y_J et Z_J .

Phase 2 : détermination de $\hat{Y}_J$ et $\hat{Z}_J$

Cette phase s'appuie sur un programme utilisant en entrées les résultats de l'A.C.P. . Les estimations fournies sont les suivantes.

Matériel Y

		A	LF	LG	S1	S2	S3	S4	S5
$\hat{X}_{YJ}$	borne du nuage	16,4	31,4	43,6	3,9	0,85	1,1	1,4	1,2
$\hat{Y}_{0J}$	saturation d'une contrainte	16,8	32,4	46,2	3,6	0,3	0	0,4	0,2

Matériel Z

		A	LF	LG	S1	S2	S3	S4	S5
$\hat{X}_{zJ}$	borne du nuage	12,5	20,8	15,2	7,3	6,4	13,3	12,6	11,9
$\hat{Z}_{0J}$	saturation d'une contrainte	10,3	15	0	9	9,3	13,8	18,6	17,6

Les écarts entre les deux estimations ponctuelles de Y_J sont relativement étroits. La composante qui s'annule est S3, mais S2, S4, S5 sont également voisins de zéro. Cette composition granulométrique, du double point de vue de l'aspect sigmoïde de sa courbe cumulative, (fig. 7) et de l'absence de fractions sableuses moyennes et grossières, est caractéristique d'un limon éolien.

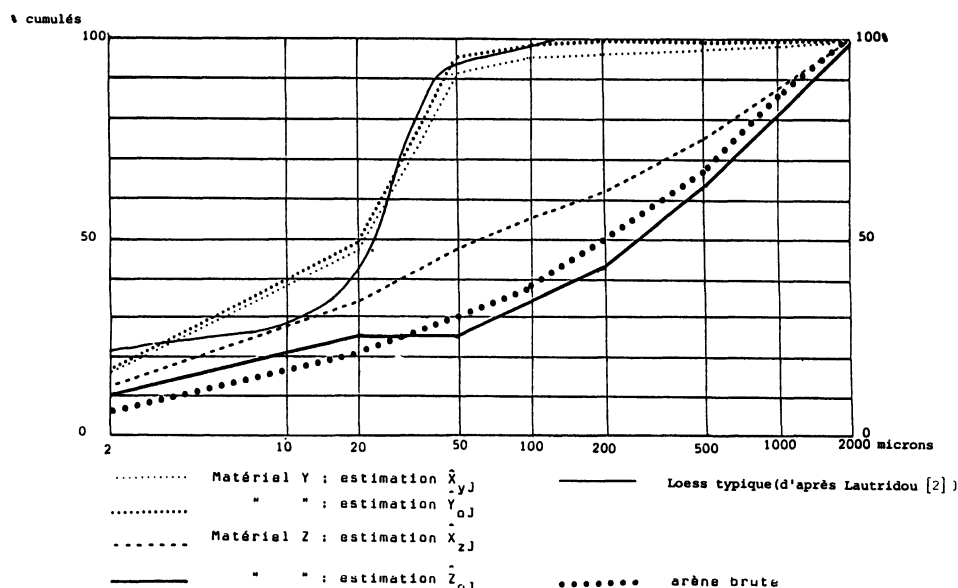


FIG. 7. – Courbes granulométriques cumulatives de différents matériaux

A côté des courbes estimées, une courbe de l'arène et une courbe de loess (Apport éolien) typique ont été représentées

Les écarts entre les deux estimations $\hat{X}_{zJ}$ et $\hat{Z}_{0J}$ de Z_J sont assez élevés. Les courbes cumulatives correspondantes ont un aspect concave, comparable à celui des arènes sous-jacentes. Or, il est improbable que cette arène en s'altérant donne un matériel qui ne contienne pas de limon grossier (cas de $\hat{Z}_{0J}$) et même, qu'elle fournisse moins de limon grossier qu'elle n'en renferme (soit 12 % en moyenne). Nous pouvons donc admettre que ce matériel Z ne contient pas moins

de 12 % de limon grossier, et, calculer une nouvelle estimation ($\hat{Z}'_{0J}$) qui tienne compte de cette contrainte. Soit :

	A	LF	LG	S1	S2	S3	S4	S5
$\hat{Z}'_{0J}$ saturation d'une nouvelle contrainte	12	19,6	12	7,7	7,0	14,7	13,9	13,1

La fourchette entre $\hat{X}_{jJ}$ et $\hat{Z}'_{0J}$ est relativement étroite.

Phase 3 : détermination de $\hat{P}_1$

Les deux composants du mélange étant estimés, il devient possible de donner leur proposition relative dans chaque échantillon. Ces proportions sont fournies par le même programme qui calcule les estimations de Y_j et Z_j . Notons que l'échantillon le plus riche en substrat en contient entre 89 et 100 %, le plus riche en apport en renferme entre 92 et 100 %. Les diverses proportions ont été reportées sur la séquence (fig. 8). Il se dégage une allure en enveloppes, ouvertes vers l'aval, les plus forts pourcentages d'apports se situant au cœur de ces enveloppes. Cette distribution s'explique aisément par l'érosion et s'accorde avec l'existence d'un apport éolien.

4.3. — Conclusion

Dans cette séquence du Sourn, 80 % des variations granulométriques peuvent être expliquées par le mélange entre deux matériaux. La connaissance de leur composition permet d'identifier leur origine :

- l'un est un apport éolien ;
- l'autre est issu de l'arène sous-jacente.

Par ailleurs, le traitement des données chimiques de cette séquence, de façon autonome d'abord, puis couplé avec les données granulométriques, confirme ce résultat, et permet de mettre en évidence que, le matériel Z dérive de l'arène par altération, et non par une simple fragmentation mécanique des grains⁽³⁾.

De plus, l'analyse de deux séquences situées légèrement plus au nord, l'une sur quartzite, l'autre sur schiste, a conduit à des résultats similaires. L'apport, mis en évidence, apparaît légèrement plus grossier, ce qui est en plein accord avec l'origine septentrionale généralement avancée.

5. — CONCLUSION

Au cours de ce travail, l'analyse en composantes principales (A.C.P.) est apparue sous l'aspect peu habituel d'une méthode d'estimation des paramètres

(3) Il n'y a pas lieu de présenter ici les résultats complets, qui se trouvent dans LE CALVEZ 1979 [3]. A propos des données chimiques, nous signalerons simplement que dans la démarche D3, les régressions simples ont conduit à des estimations vérifiant les contraintes.

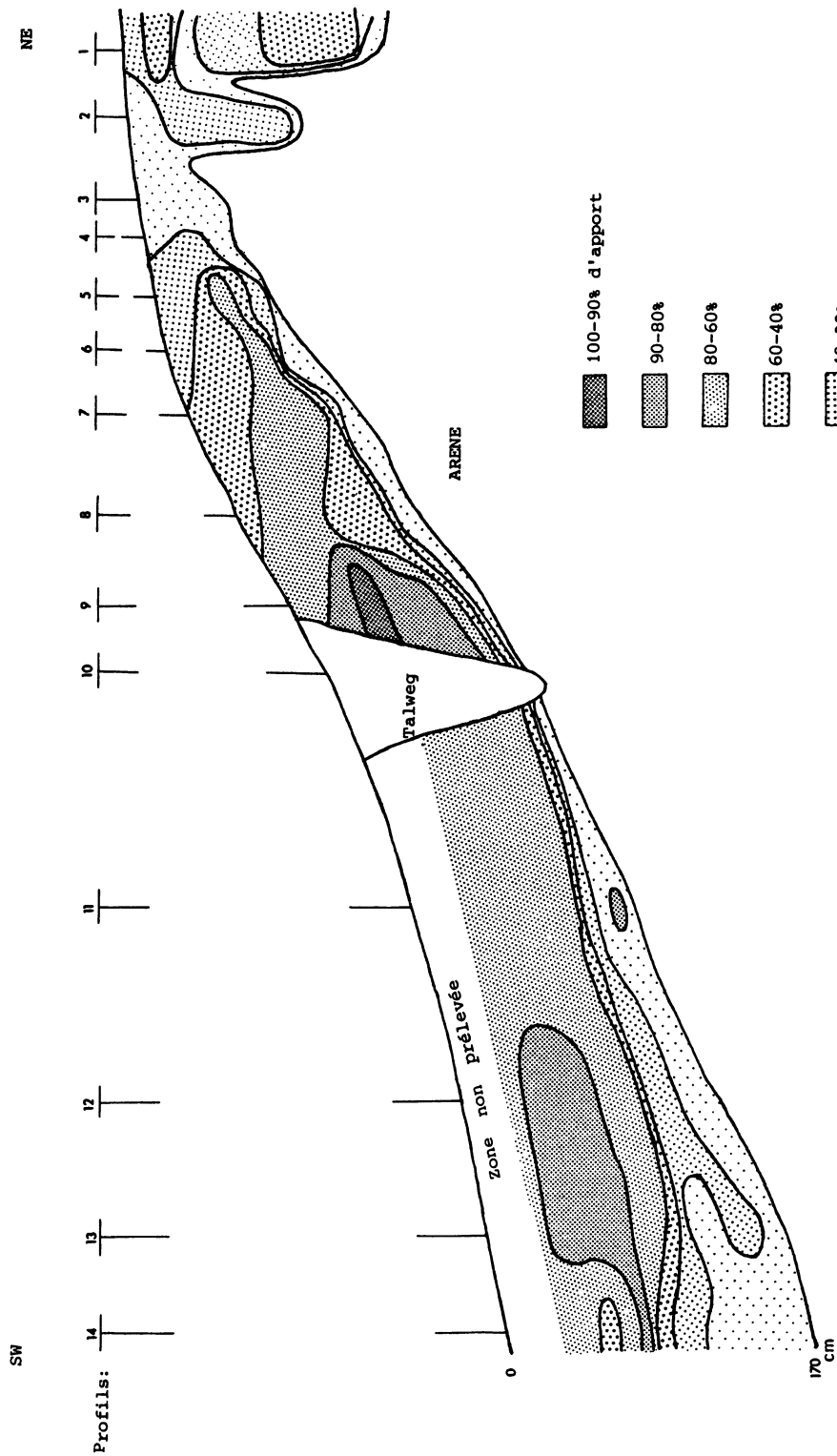


FIG. 8. — Distribution de l'apport le long du versant

Cette distribution est reconstituée à partir des estimations calculées en 66 points de prélèvements. Elle est en accord avec l'action combinée de deux effets : apport éolien et érosion

d'un modèle construit a priori : c'est là que réside, à nos yeux, le principal intérêt méthodologique de cette étude.

Mais il convient de remarquer que, indépendamment du modèle de mélange, l'A.C.P. peut être appliquée directement à nos données : elle possède alors son objectif traditionnel qui est de discriminer au mieux les individus. Le premier axe s'interprète comme mettant en évidence deux tendances que l'on oppose selon les mêmes critères qui opposent les deux matériaux initiaux. Cette interprétation est correcte, même si elle ne fait pas apparaître toutes ses implications. De ce point de vue, il est intéressant d'examiner les autres axes issus de l'A.C.P. ; c'est ce que nous avons fait, et qui nous a conduit, pour toutes les analyses, à identifier le deuxième axe avec les remaniements pédologiques postérieurs au mélange. Il est donc possible de faire apparaître quantitativement et séparément les phénomènes mélange et remaniements post mélange.

REFERENCES

- [1] CAZES P., TURPIN P.Y. (1961). — Régression sous contraintes : application à l'estimation de la courbe granulométrique d'un aérosol. *Revue de Statistique Appliquée*, Vol. XIX, n° 4, p. 23-44.
- [2] LAUTRIDOU J.P. (1979). — Granulométrie des sédiments éoliens périglaciaires — Comparaison avec la fraction fine obtenue par gélifraction expérimentale. 7^e Réunion Annuelle des Sciences de la Terre. Lyon 1979. (Livre en dépôt à la Société Géologique de France. p. 280).
- [3] LE CALVEZ L. (1979). — Genèse des formations limoneuses de Bretagne Centrale : essai de Modélisation. Thèse de 3^e cycle. U.E.R. des Sciences Biologiques. Rennes. (213 p.)
- [4] MARTENS H., PAULSEN F, SPOTVOLL E., VOLDEN R. (1979). — Regression on disjoint factor analysis models. Secondes journées internationales Analyse des données et Informatique. IRIA. Versailles 1979.