

<b>Zeitschrift:</b>	Bulletin de la Société Vaudoise des Sciences Naturelles
<b>Herausgeber:</b>	Société Vaudoise des Sciences Naturelles
<b>Band:</b>	84 (1996-1997)
<b>Heft:</b>	1
<b>Artikel:</b>	Eléments simples d'identification d'anomalies géochimiques par statistique : aspects miniers et environnementaux
<b>Autor:</b>	Jaboyedoff, Michel / Bauchau, Christian / Maignan, Michel
<b>DOI:</b>	<a href="https://doi.org/10.5169/seals-287991">https://doi.org/10.5169/seals-287991</a>

### Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

### Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

### Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

**Download PDF:** 12.01.2026

**ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>**

# Eléments simples d'identification d'anomalies géochimiques par statistique: aspects miniers et environnementaux

*par*

*Michel JABOYEDOFF<sup>1</sup>, Christian BAUCHAU<sup>1</sup>, Michel MAIGNAN<sup>1</sup>*

## TABLE DES MATIERES

Abstract, Résumé .....	74
1. Introduction .....	74
1.1. Représentation des données .....	76
1.2. Des données aux distributions .....	76
1.3. Choix de classes des histogrammes: « règle empirique » .....	79
2. Interprétation en termes de populations sous-jacentes (inférence statistique) .....	79
2.1. Qu'est-ce qu'une anomalie géochimique? .....	79
2.2. Ensemble de données statistiques .....	81
2.3. Le problème du support .....	81
2.4. Lois normale et log-normale: distributions usuelles .....	84
2.5. Utilisation du <i>probability plot</i> .....	87
2.6. Seuil et retour aux cartes .....	90
2.7. Discriminer .....	91
3. Autres outils .....	91
3.1. Evaluation quantitative des anomalies .....	91
3.2. Le variogramme .....	92
4. Exemples .....	94
4.1. Exemple minier .....	94
4.2 Un exemple de variogramme .....	97
4.3. Exemple de traitement de pollutions .....	99
5. Mise en œuvre informatique .....	103
5.1. Statistiques spatiales .....	103
5.2. Représentation cartographique .....	104
5.3. Corrélations spatiales, estimations et simulations .....	104
6. Conclusion .....	104
Remerciements .....	106
Bibliographie .....	106

<sup>1</sup>Institut de minéralogie, Université de Lausanne, BFSH2, CH-1015 Lausanne

*Abstract.*—JABOYEDOFF M., BAUCHAU C. et MAIGNAN M., 1996. Simple statistical elements to identify geochemical anomalies: mineral exploration and environmental aspects. *Bull. Soc. vaud. Sc. nat.* 84.1: 73-108.

The present paper discusses some useful criteria to identify natural or anthropogeneous geochemical anomalies. Such an identification can be done with simple univariate statistical tools. First, simple statistical parameters are reminded. In order to determine the applicability of the normal and log-normal distributions, the signification of those distributions is discussed. The relation between the grades' histogram shape and the type of the anomaly and sampling support is emphasized. Histograms are considered to be a good approximation of the true distributions, thus allowing the analyst to try an identification of one or more populations with the probability plots. If an anomalous distribution is detected, it is possible to define criteria to delineate zones where the concentrations are abnormal. This approach is illustrated by examples. Its efficiency is essentially due to its rapidity and well structured processes.

*Key words:* Geochemical anomalies, mineral exploration, pollutions, statistics, geostatistics, geochemical threshold.

*Résumé.*—JABOYEDOFF M., BAUCHAU C. et MAIGNAN M., 1996. Eléments simples d'identification d'anomalies géochimiques par statistique: aspects miniers et environnementaux. *Bull. Soc. vaud. Sc. nat.* 84.1: 73-108.

On présente un certain nombre de critères permettant l'identification d'anomalies géochimiques tant naturelles que d'origine anthropogène, à l'aide d'outils statistiques univariés simples. On rappelle la détermination des paramètres statistiques de base, puis la signification des lois normale et log-normale afin de cerner leur champ d'application. On souligne les relations qui peuvent exister entre l'histogramme des teneurs et le type d'anomalie étudiée, la maille et la forme du support des prélèvements. On étudie les histogrammes qui sont supposés être une approximation des distributions, dans le but de savoir s'il y a plusieurs populations à l'aide de «probability plots». Si l'on obtient une population anomale, on peut définir des critères pour délimiter des zones générées par des processus de concentration anormaux. Au travers d'exemples, on montre la problématique liée à cette approche et son efficacité, essentiellement due à sa rapidité et à la démarche structurée. Une brève description de logiciels pertinents est également effectuée.

*Mots clefs:* Anomalies géochimiques, prospection minière, pollutions, statistiques, géostatistiques, seuil géochimique.

## 1. INTRODUCTION

«Comment identifier une anomalie géochimique?» est une question certainement réactualisée. En effet, si elle reste essentiellement posée aux géologues prospecteurs, elle devient aussi d'actualité pour le géologue qui s'occupe d'environnement et plus particulièrement de pollutions. Pour les premiers, la détermination de l'importance, de la valeur, d'une anomalie géochimique est primordiale. La géochimie de sols précède la phase, coûteuse, des forages et doit en fournir les premières cibles, si possible de haut potentiel. C'est pourquoi il importe de disposer de critères de sélection pratiques et fiables afin de choisir les anomalies les plus prometteuses, en évitant ainsi un gaspillage de mètres de forage et donc des dépenses inutiles. Il en va de même pour l'éologue de pollutions, bien qu'ici les engagements de dépenses soient en général sensiblement moindres, si l'on ne fore pas.

Nous nous sommes donc interrogés depuis de nombreuses années sur la définition des anomalies géochimiques, par le biais notamment d'un enseigne-

ment de la prospection minière par simulation. Ces simulations furent développées par R. Woodtli à l'Université de Lausanne, puis avec la collaboration de M. Vannier à l'Ecole des Mines de Paris et par la suite avec celle des auteurs (VANNIER et WOODTLI 1992, BAUCHAU et JABOYEDOFF 1990, BAUCHAU *et al.* 1993). Elles ont permis d'expérimenter, par simulations, des méthodes et des stratégies de détection d'anomalies géochimiques et d'en dégager les points essentiels. D'autre part le traitements des données de pollutions fait maintenant partie de cette démarche générale (MAIGNAN and MAIGNAN in press, POURCHET *et al.* in press, KANEVSKI *et al.* in press).

L'objet pratique du présent article est d'indiquer des méthodes simples permettant la détection d'anomalies géochimiques, c'est-à-dire de valeurs élevées qui peuvent représenter une population anomale. Du point de vue de la géologie minière, on débouche ensuite sur des indices et dans le cadre des sciences de l'environnement sur des pollutions, par exemple. Dans ce dernier cas, on suppose que les teneurs anomalies appartiennent à des populations engendrées par des processus de concentration anormaux ou par une source anormale. Il faut voir là le contraire de la notion de seuil défini *a priori*. Les teneurs anomalies peuvent n'être alors qu'une expression exceptionnelle due à la variabilité d'un processus normal. Les concentrations exceptionnelles -par rapport à un fond géochimique (*background*)- liées à des minéralisations, ou bien encore les pollutions, se caractérisent justement par la présence d'une ou de plusieurs populations anomalies.

D'autre part, dans le cas des pollutions, la recherche de la population formée par le fond géochimique est importante, car elle indique la distribution des teneurs à atteindre lors de la réhabilitation complète d'un site pollué.

Dans la pratique, on est souvent confronté à un grand nombre de données géochimiques ou de pollutions distribuées dans l'espace, en général simplement dans un plan. Il faut donc, pour commencer, définir quelles teneurs sont anomalies, soit un seuil. Si le seuil est fixé *a priori*, et cela peut être le cas lorsqu'il existe des réglementations en termes de pollution, une réponse partielle est donnée, sauf en ce qui concerne la taille minimale de l'anomalie (tonnage, volume, surface) pour la considérer comme «dangereuse». Même dans ce cas, du fait que les normes OSOL, par exemple, définissent des teneurs limites, mais pas le support volumétrique, des investigations complémentaires sont recommandables. Au contraire, si l'on cherche à mettre en évidence des processus exceptionnels, l'emploi des statistiques univariées permet souvent de dégager rapidement des populations et des seuils au-dessus desquels les populations anomalies sont dominantes. Il faut analyser la signification des populations du point de vue de leur répartition spatiale, et il n'est pas toujours simple de lier une population statistique donnée à un processus physico-chimique particulier. Les processus de diffusion ou autres ayant engendré une anomalie n'étant généralement pas connus et mesurés *a priori*, il y a lieu de compenser l'absence d'un modèle déterministe par une approche statistique. De plus, la complexité spatiale d'un processus de diffusion, par exemple, justifie une démarche statistique en complément d'un éventuel modèle déterministe. Nous exposons ici quelques principes et méthodes, ainsi que les problèmes liés à ces méthodes. Des exemples sont donnés pour souligner certains problèmes. La démarche adoptée dans cet article suit intentionnellement une approche «manuelle», c'est-à-dire faisant plus appel à la réflexion qu'à la mécanique informatique.

### 1.1. Représentation des données

Le document de base est une carte des teneurs (cf. par exemple figures 15, p. 94 et 21, p. 99), car elle permet de voir immédiatement s'il y a des zones anomalies (à teneurs élevées) et localisées. L'analyse proprement dite des données s'effectue à partir d'histogrammes en échelles linéaires ou logarithmiques et de statistiques simples (DAVIS 1986, JOURNEL 1989, MARSAL 1987, RUEGG 1989, SAPORTA 1990, VENTSEL 1987) telles que les estimations «non-biaisées»  $m$  de la moyenne  $\mu$  d'une variable aléatoire  $X$  et  $S^2$  de sa variance (moment d'ordre 2) et de l'écart type  $S$  sont:

$$\text{Moyenne } m = \frac{1}{N_{tot}} \sum_{i=1}^{i=N_{tot}} x_i \quad (1)$$

$$\text{Variance } S^2 = \frac{1}{(N_{tot} - 1)} \sum_{i=1}^{i=N_{tot}} (x_i - m)^2 \quad (2)$$

où  $N_{tot}$  est le nombre total d'échantillons. Plus généralement, la moyenne  $m$  est l'estimation de l'espérance mathématique notée  $E[X]$ . De même la variance  $S^2$  est l'estimation de l'espérance mathématique du moment d'ordre 2 centré  $E[(X-m)^2]$ . On rencontre fréquemment les autres moments statistiques qui, de façon générale, sont notés:

$$\text{Moment d'ordre } k = m_k = \frac{1}{N_{tot}} \sum_{i=1}^{i=N_{tot}} (x_i - m)^k \quad (3)$$

Le coefficient d'aplatissement, appelé aussi kurtosis à une constante près, et le coefficient d'asymétrie (skewness) sont définis comme:

$$\begin{aligned} \text{Coefficient d'asymétrie} &= \frac{m_3}{S^3}; \\ \text{coefficient d'aplatissement } \frac{m_4}{S^4} - 3 &= \text{kurtosis} - 3 \end{aligned} \quad (4)$$

Le coefficient d'asymétrie est positif si la distribution s'étale plus du côté des valeurs élevées et inversement; il est nul pour des distributions symétriques. On soustrait 3 dans le calcul de l'aplatissement pour obtenir zéro dans le cas d'une distribution normale. Si l'aplatissement est positif, c'est que la distribution est plus aplatie qu'une gaussienne.

### 1.2. Des données aux distributions

L'analyse attentive d'un histogramme en regard des paramètres décrits ci-dessus permet souvent de mettre en évidence les tendances et les éventuels sous-ensembles. Ces ensembles peuvent être modélisés par des distributions. On note généralement la fonction de distribution ( $\approx$  l'histogramme)  $f(x)$  et la fonction de répartition ( $\approx$  l'histogramme cumulé)  $F(x)$  où  $x$  représente les valeurs prises par une variable aléatoire  $X$ . Les histogrammes sont des estimations des fonctions de distribution et de répartition. La distribution est généralement une densité de probabilité, c'est pour cela qu'une probabilité doit être notée

$f(x)\Delta x$ . Remarquons que la fonction de distribution est la dérivée de la fonction de répartition  $F(x)$  (fig. 1):

$$F'(x) = f(x) \text{ ou encore } F(x) = \int f(x)dx \quad (5)$$

$$\text{d'autre part: } F(x)_{x \rightarrow \infty} = 1 \text{ ou } 100\% \quad (6)$$

$F(x)$  est équivalent à la probabilité  $P(X < x_i)$  que la variable  $X$  soit inférieure à  $x_i$  de sorte que:

$$P(X < x_i) = F(x_i) \approx \frac{N_{X < x_i}}{N_{tot}} \quad (7)$$

où  $N_{X < x_i}$  est le nombre des valeurs inférieures à  $x_i$ . Il faut choisir des classes de valeurs  $\Delta x$  pour pouvoir visualiser l'estimation de la fonction de distribution, l'histogramme, c'est-à-dire la probabilité que la variable  $X$  prenne une valeur comprise entre  $x_i - \Delta x/2$  et  $x_i + \Delta x/2$ :

$$P(x_i - \frac{1}{2}\Delta x \leq X < x_i + \frac{1}{2}\Delta x) \approx f(x_i)\Delta x \quad (8)$$

avec

$$\int_{-\infty}^{+\infty} f(x)dx \text{ évalué par } \sum_{i=1}^{i=N_{tot}} f(x_i)\Delta x = 1 \text{ ou } 100\% \quad (9)$$

en pratique on fait l'approximation:

$$f(x_i)\Delta x \approx \frac{N(x_i - \frac{1}{2}\Delta x, x_i + \frac{1}{2}\Delta x)}{N_{tot}} \quad (10)$$

où  $N_{tot}$  représente le nombre total de données,  $N(x_i - \Delta x/2, x_i + \Delta x/2)$  le nombre de données appartenant à une classe. Ou si l'histogramme est construit en nombres on utilise:

$$N_{tot}f(x_i)\Delta x \approx N(x_i - \frac{1}{2}\Delta x, x_i + \frac{1}{2}\Delta x) \quad (11)$$

Souvent, on examine l'histogramme simple et l'histogramme cumulé, puis celui-ci sur papier de probabilité (*probability plot*). Ce dernier s'apparente à un histogramme cumulé dont l'échelle des fréquences a été modifiée; de ce fait, si l'on traite un ensemble présentant les caractéristiques d'une loi normale (en général on utilise l'hypothèse de normalité ou de log-normalité), les points s'alignent sur une droite, la droite de Henry (fig. 2). En fait, il s'agit de la projection des pourcentages cumulés de la loi normale ou de la loi log-normale sur un axe des valeurs normées qui représente l'échelle des fréquences. L'avantage de cette représentation «cumulée» est qu'elle associe à chaque valeur observée  $x_i$  de la variable  $X$  une valeur de  $F(x_i)$ . Ce n'est pas le cas de l'histogramme simple qui est composé de classes. Une fois l'ajustement d'une loi réalisé, quelle que soit la méthode, un test du  $\chi^2$  est toujours possible, mais n'est pas nécessaire dans la phase exploratoire.

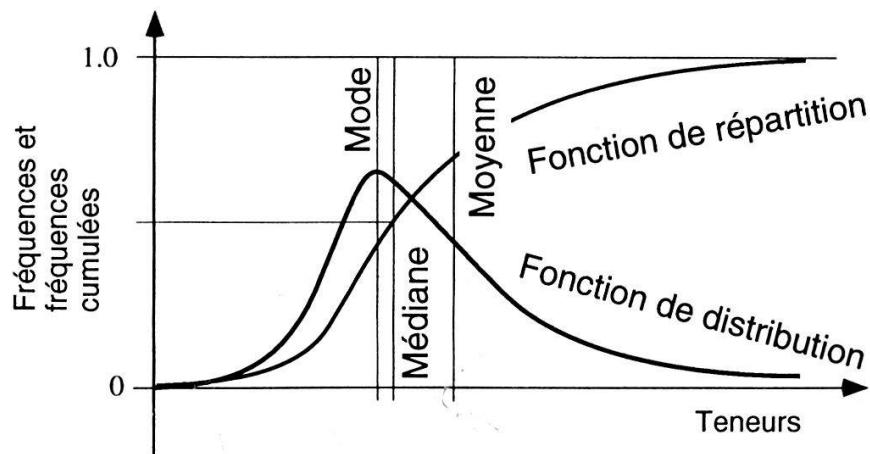
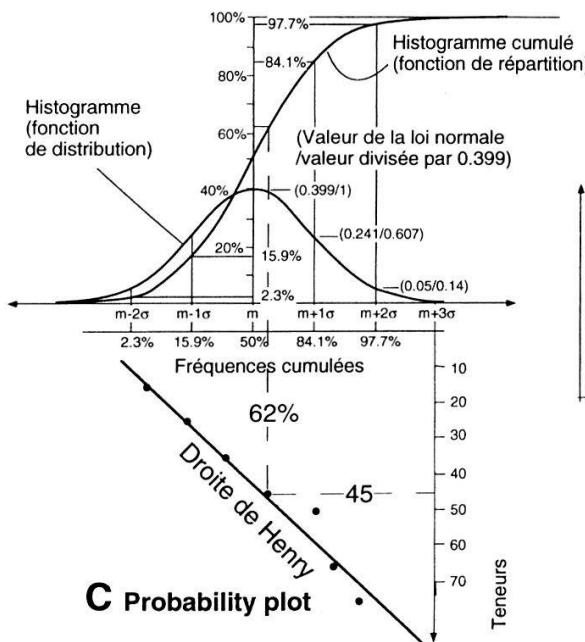
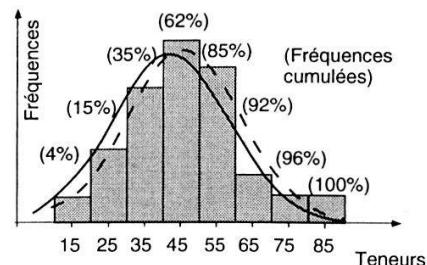


Figure 1.-Dans la pratique, la fonction de répartition est proche de l'histogramme cumulé et la fonction de distribution de l'histogramme. Le schéma présente la terminologie généralement utilisée pour décrire un histogramme. Le mode est la classe la plus représentée et la médiane est la valeur qui divise la distribution en deux parts égales de 50% des observations.

### A Loi normale ( $\sigma=1$ )



### B Histogramme



### C Probability plot

Exemple d'histogramme observé et modélisé:  
 Droite de Henry  
 Moyenne + variance  
 de la population

Figure 2.-On illustre ici la construction d'un «probability plot» de type normal (voir texte). (A) Transformation des valeurs de la fonction de répartition sur l'échelle des valeurs. En B un histogramme proche de la loi normale. Les points représentent cet histogramme sur un «probability plot» (C). Sur l'échelle des fréquences, on reporte la valeur cumulée 62%, comme s'il s'agissait d'une loi normale, et la teneur 45 sur l'échelle des valeurs. Le résultat est une droite si la distribution est normale. Il en va de même si on choisit une échelle logarithmique. On note la légère différence qui existe entre les distributions normales suivant qu'on utilise directement la moyenne et l'écart type de la population ou la droite de Henry. Les premiers sont calculés dans l'espace des fréquences simples et l'autre dans l'espace cumulé. Les écarts types sont comparables, ce sont les moyennes qui diffèrent un peu. Il faut garder cette remarque en mémoire lors des interprétations (d'après PACÉ et CLUZEL 1968).

### 1.3. Choix de classes des histogrammes: règle «empirique»

La mesure la plus simple de la dispersion des valeurs est l'écart-type. Nous proposons ici dans un premier temps d'utiliser la valeur (fig. 3):

$$\Delta x \approx [de \ 0.6 \ à \ 1.2]S \quad (12)$$

où  $\Delta x$  est la largeur des classes de l'histogramme et où  $S$  provient de la formule 2. En effet, si on veut que la classe la mieux fournie comporte approximativement de 20 à 40% des données, on supposera avoir affaire à une distribution de type normal (voir plus loin); sachant que l'accroissement pour  $2S$  est de 68%, la pente de l'histogramme cumulé est d'environ:

$$\frac{.68}{2S} = \frac{de \ .2 \ à \ .4}{\Delta x} \Rightarrow \Delta x \approx [de \ .6 \ à \ 1.2]S \quad (13)$$

S'il y a asymétrie de l'histogramme,  $S$  peut être plus grand que dans le cas symétrique, de sorte que  $\Delta x$  augmente, et cette relation n'est plus nécessairement valable. On peut passer en échelle logarithmique si ce n'est pas déjà fait.

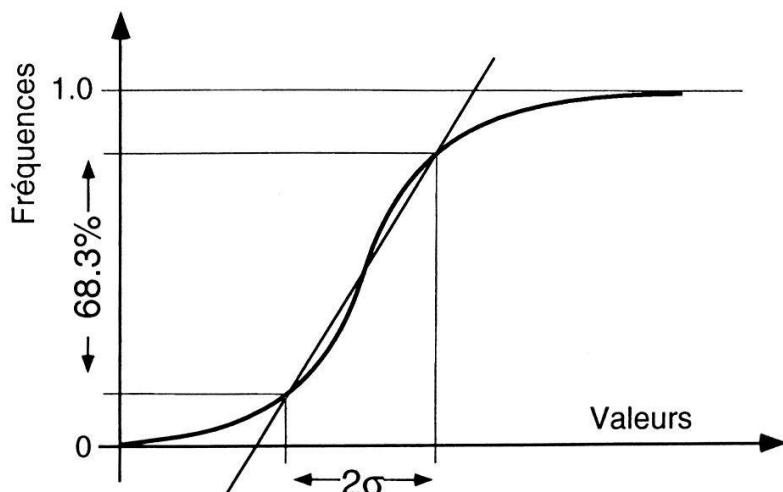


Figure 3.—L'examen de l'histogramme cumulé permet, à l'aide de sa plus forte pente, de choisir la classe la mieux représentée, d'après le principe énoncé dans le texte. Ici le cas gaussien.

## 2. INTERPRÉTATION EN TERMES DE POPULATIONS SOUS-JACENTES (INFÉRENCE STATISTIQUE)

### 2.1. Qu'est-ce qu'une anomalie géochimique?

Une anomalie géochimique est définie comme un ensemble de teneurs élevées, par rapport à la moyenne, spatialement proches les unes des autres ou plus simplement «contourables» d'un point de vue cartographique (fig. 4), voire comme un sous-ensemble connexe.

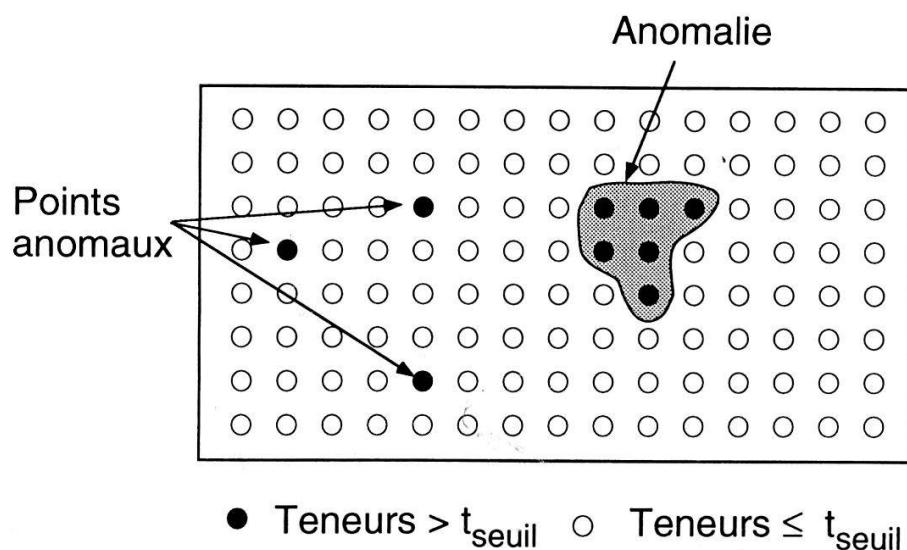


Figure 4.—La notion de «contourable» est importante dans la définition d'anomalies, car un point isolé n'est pas significatif. Il faut qu'une surface soit définie sinon il peut ne s'agir que d'un point aléatoire. Dans le cas d'un point unique, il faut effectuer des prélèvements plus rapprochés pour déterminer s'il s'agit d'une anomalie. La notion d'anomalie est aussi dépendante de l'échelle de travail.

Des valeurs élevées non contiguës ne constituent pas une anomalie, elles peuvent être aléatoirement distribuées dans l'espace. La notion d'anomalie est fortement dépendante de l'échelle à laquelle on travaille. Par exemple, des anomalies d'une taille de 100 m recherchées avec une grille de prélèvements de 1000 m ne se manifesteront que par des points anomaux isolés, sans signification spatiale. Mais l'existence de l'anomalie sera sans doute décelée avec une maille inférieure à 50 m: on peut alors espérer qu'au moins deux prélèvements contigus appartiendront à la même anomalie.

La notion d'anomalie implique nécessairement la notion de *background* (ou fond géochimique), qu'on peut définir comme un ensemble de valeurs oscillant autour d'une valeur moyenne (ou d'une tendance régionale), avec une variabilité constante partout dans l'espace. On ne peut pas «contourer» de zone à l'intérieur de cet ensemble de valeurs sur les données brutes. D'autre part, la notion de *background* peut aussi avoir une signification locale, par exemple en présence d'une dérive régionale: la variance et la moyenne peuvent évoluer. On tente généralement de définir des anomalies par rapport à un *background*: rapport anomalie/*background*.

Il faut nuancer la notion de «contourage», car il se peut que des valeurs élevées soient localisées mais pas contourables parce qu'il existe des valeurs plus faibles entre les points anomaux. Il y a à cela plusieurs raisons. Dans un exemple que nous analyserons au paragraphe 4.1, les teneurs en cuivre des roches présentent de grande variations; en effet, certaines lithologies sont beaucoup plus riches en cuivre que d'autres pour différentes raisons: altérations, minéralogie, etc. Une anomalie n'apparaît alors que lors du lissage des valeurs, opération nécessaire pour dresser une carte de teneurs. Du point de vue des pollutions, on peut rencontrer le même type de problèmes lorsqu'une substance est dispersée en grains. Si on dispose de suffisamment de temps, on peut toujours trier les données suivant d'autres caractéristiques telles que lithologie, autres éléments chimiques, matière organique, argiles, etc., afin de conditionner les investigations statistiques.

## 2.2. Ensemble de données statistiques

Les teneurs géochimiques ne représentent qu'un ensemble limité d'échantillons. On suppose presque toujours que les caractéristiques de cet échantillonnage tendent vers celui de la population globale, c'est-à-dire de la population qu'on obtiendrait si on connaissait complètement, ou mieux, continûment la zone étudiée (ergodicité). De sorte que la distribution (histogramme), sa variance, sa moyenne, ses modes, etc. tendent vers celles de la population globale, en conséquence de la loi des grands nombres.

Un ensemble de données est homogène, ou régulier, si les deux types d'histogrammes par classe et cumulé (échelles linéaires ou logarithmiques), sont continus et que leurs dérivées sont « régulières » de part et d'autres d'un mode unique pour l'histogramme simple. Dans ce cas, l'usage des *probability plot* nous indique si cet ensemble respecte éventuellement une loi normale ou log-normale.

Lorsque les histogrammes ne sont pas « réguliers » (c'est-à-dire qu'une unique distribution ne permet pas d'interpréter l'histogramme), on est amené à diviser l'ensemble des données en plusieurs sous-ensembles. En présence de plusieurs modes ou quand la dérivée ne suit pas une tendance régulière, on peut soupçonner l'existence de plusieurs distributions et faire des hypothèses quant à la normalité ou la log-normalité des distributions.

## 2.3. Le problème du support

Pour interpréter des données spatiales par des statistiques univariées, il faut veiller à ce que la maille de prélèvements soit appropriée, que la zone étudiée soit centrée et que la distribution de l'information soit homogène. Il faut donc considérer les points suivants:

1.-La maille de la grille ou du profil de prélèvements doit être appropriée à la cible recherchée. Dans un premier stade, on se contentera souvent d'une maille légèrement inférieure à la taille de la cible définie d'après les objectifs de teneurs et de tonnage (fig. 5 et 6). Mais cela ne permet pas nécessairement

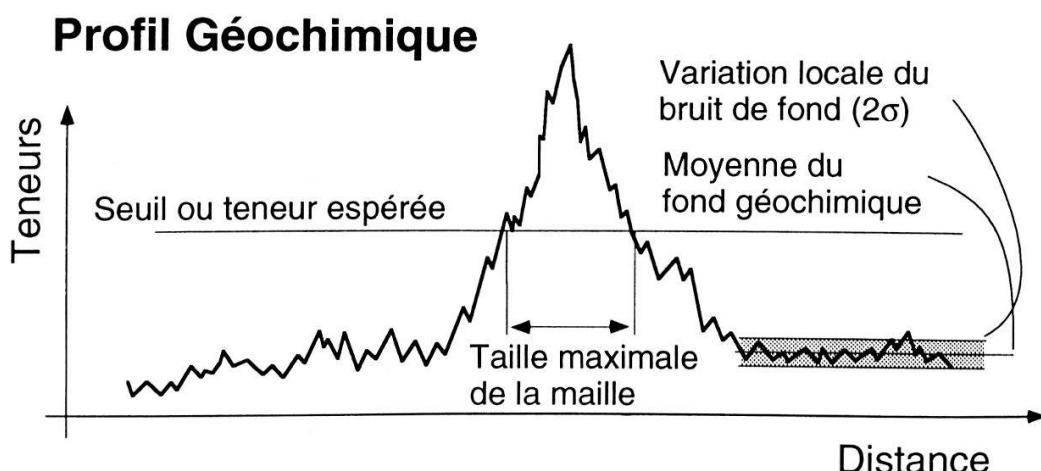


Figure 5.–Les dimensions de la maille de la grille de prélèvements sont définies par la forme, la taille et la teneur de l'anomalie à détecter. Ici le mot grille décrit aussi des profils.

de définir d'anomalies. Puis, une fois un point anomal localisé, on resserre la maille (fig. 6). Cette fois, celle-ci devrait être inférieure à la moitié de la taille de la cible. Le choix d'une grille ou de profils géochimiques doit être adapté à d'éventuelles anisotropies. Il est dispendieux d'avoir une grille carrée si l'objet recherché est allongé. Il est important de consulter une carte géologique, pédologique, hydrologique, d'activité anthropogène, etc. si c'est possible, afin de pouvoir localiser les éventuels sites favorables à la recherche et d'évaluer leurs formes et leurs tailles. On peut aussi utiliser une grille aléatoire et/ou tester l'homogénéité par la dimension fractale afin de définir les zones de basse densité d'information (FLAMM *et al.* 1994, LOVEJOY *et al.* 1986).

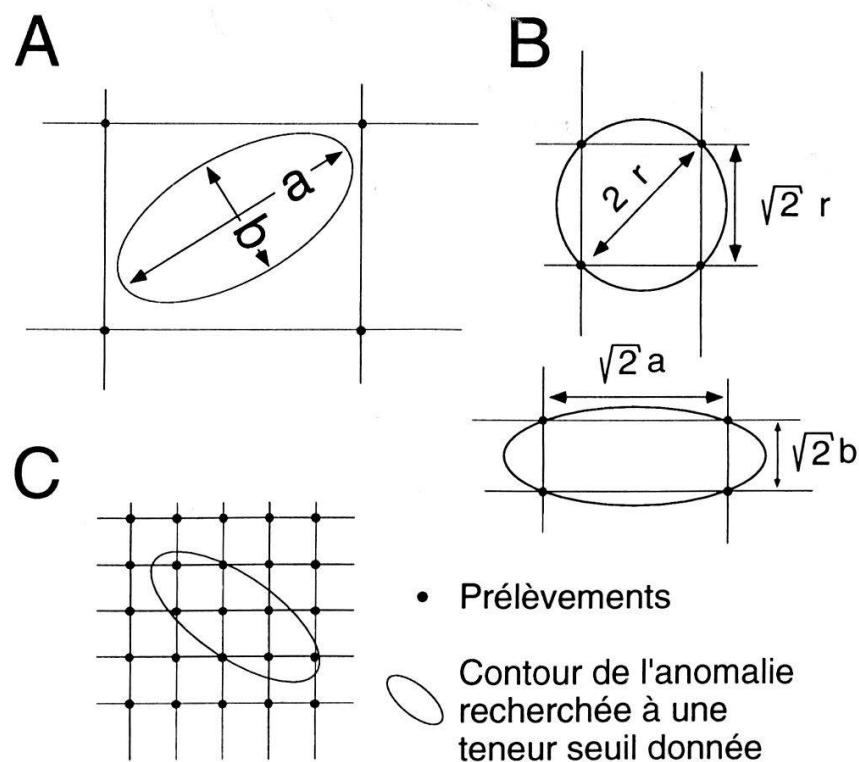


Figure 6.–Exemple idéalisé de choix d'un écartement et d'une orientation des mailles de prélèvement de sols (d'après ELLIOT and FLETCHER 1975). Si l'on recherche une anomalie elliptique (A) la maille est trop lâche et mal orientée. En (B) maille optimale pour une anomalie supposée circulaire de rayon  $r$ . Pour qu'un prélèvement touche à coup sûr l'anomalie, la plus grande distance entre deux points successifs de la maille doit être égale ou inférieure au diamètre du cercle ( $2r$ ). Dans le cas elliptique il suffit de faire une homothétie à partir du cas circulaire. (C) exemple de maille inadéquate dans la première phase de reconnaissance: beaucoup de prélèvement sont inutiles. Par contre, lors de la phase suivante qui sert à définir une anomalie, une telle maille peut se justifier; si l'orientation est connue, on oriente alors la maille parallèlement à l'allongement de l'ellipse.

2.–Pour interpréter des histogrammes, il faut prendre garde à ce que les anomalies ne soient pas en bordure de la zone d'étude ou qu'elles ne soient que partiellement reconnues, car la connaissance d'un *background* de part et d'autre de l'anomalie est importante. Dans le meilleur des cas, il faudrait même que le domaine d'étude soit centré sur l'anomalie et de même forme

que cette dernière, c'est-à-dire dans un cas théorique qu'il soit circulaire ou elliptique, pour ne pas sur-représenter le *background* dans les histogrammes (fig. 7).

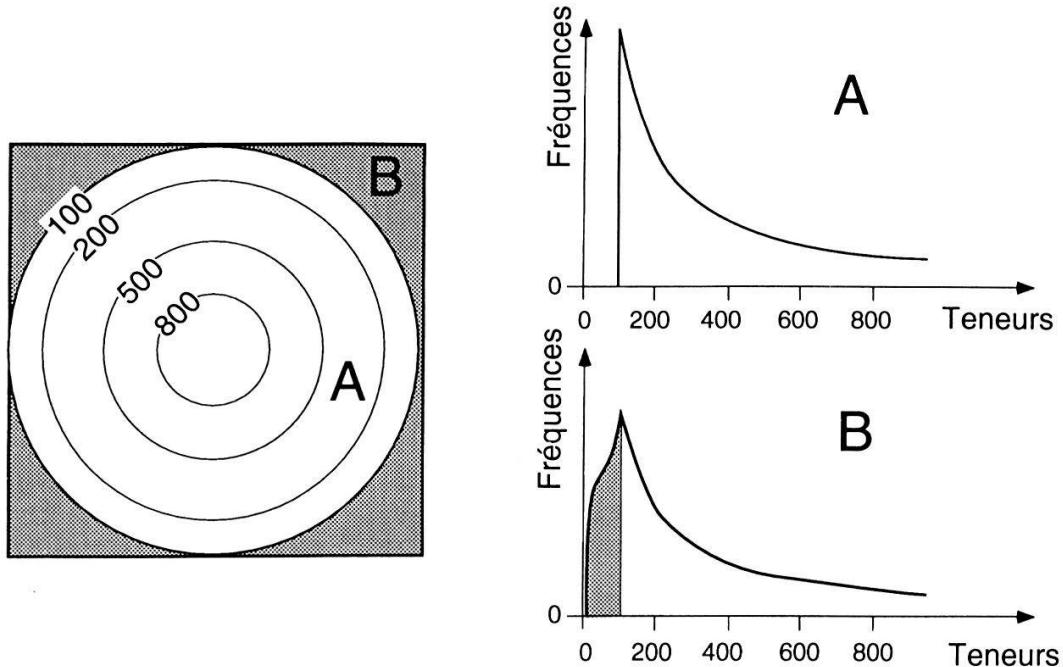


Figure 7.-Effet de la forme du support sur la distribution des teneurs d'une anomalie générée par un processus de diffusion, échantillonnée de façon homogène, en A sur un support circulaire et carré en B. L'effet se voit sur les faibles teneurs (grisé). Une distribution des teneurs produites uniquement par diffusion est inversement proportionnelle à la teneur en échelle linéaire (JABOYEDOFF and MIGNAN in prep.)

3.–Le problème de l'homogénéité est peut-être le plus pervers, car on effectue en général une campagne à maille lâche afin de mettre en évidence des points anomaux, puis on resserre la maille autour de ces points pour vérifier s'il s'agit d'anomalies; de ce fait, la densité d'informations varie en fonction des teneurs (fig. 8). Pour remédier à ce genre de problèmes, on pourrait se limiter aux zones à densité constante; malheureusement les contraintes économiques font que les zones de *background* sont souvent à plus faible densité de prélèvements. On peut aussi supprimer des données afin que la densité soit homogène. Une méthode courante, le *declustering* (description détaillée dans ISSAKS and SRIVASTAVA 1989 et DEUTSCH and JOURNEL 1992), consiste à pondérer les valeurs par l'inverse de la densité. Si on connaît dans un domaine «j» la densité  $\rho_j$  et le nombre  $q_j(x_i)$  de prélèvements, dont les valeurs des teneurs appartiennent à un intervalle centré sur  $x_i$ , on obtient pour m domaines le nombre  $N_c$  de données corrigées contenues entre les valeurs  $x_i - \Delta x/2$  et  $x_i + \Delta x/2$  et en posant  $\rho_h$  la densité désirée on obtient:

$$N_c(x_i + \Delta x \leq X < x_i + \Delta x) = N_c(x_i) \approx \rho_h \sum_{j=1}^{j=m} \frac{q_j(x_i)}{\rho_j} \quad (14)$$

d'où si n est le nombre de classes de valeurs centrées sur les  $x_i$ , le nombre total corrigé vaut:

$$N_{tot} = \sum_{i=1}^{i=n} N_c(x_i) \quad (15)$$

et la probabilité est donnée par:

$$P(x_i - \frac{1}{2}\Delta x \leq X < x_i + \frac{1}{2}\Delta x) = f(x_i)\Delta x \approx \frac{N_c(x_i)}{N_{tot}} \Delta x \quad (16)$$

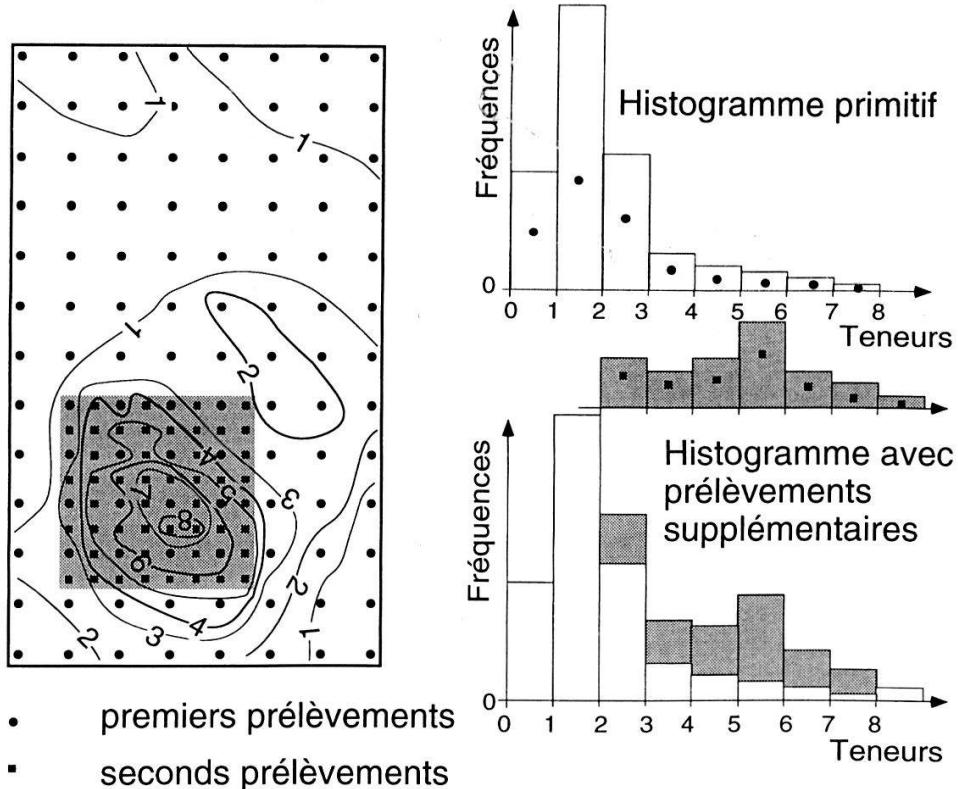


Figure 8.–Exemple synthétique de l'effet pervers dû à la densité inhomogène de prélèvements sur les histogrammes des teneurs. Les prélèvements supplémentaires (carrés) induisent une population artificielle. Si le fond géochimique était représenté on aurait une classe [0-1] très grande: en fait l'histogramme est proche d'un histogramme de diffusion tronqué.

#### 2.4. Lois normale et log-normale: distributions usuelles

Il est commode de comparer les histogrammes d'ensembles de teneurs géochimiques à certaines lois de distribution de référence. Les plus fréquemment utilisées sont les lois normale et log-normale. En effet elles découlent de processus simples et sont souvent proches des observations. Mentionnons le théorème central limite (suivant l'énoncé donné dans VENTSEL 1987), pour comprendre l'intérêt de ces deux distributions.

**Théorème central limite:** Soient  $X_1, X_2, X_3, \dots, X_n$  des variables aléatoires indépendantes de même loi de répartition, d'espérance mathématique  $\mu$  (moyenne) et de variance  $\sigma^2$ ; lorsque  $n$  augmente indéfiniment, la loi de répartition de la somme

$$Y_n = \sum_{k=1}^{k=n} X_k \quad (17)$$

tend vers la loi normale.

De façon intuitive, le théorème central limite est facilement compréhensible au travers d'un exemple (fig. 9). En présence d'un phénomène qui génère une distribution de type tout ou rien, celle-ci a une forme de crêteau. Si, par exemple, la population d'origine ne prend que la valeur  $b$ , alors, après l'accomplissement du processus, la population a une probabilité unique dans un intervalle  $[b-a, b+a]$  et en dehors la probabilité est nulle. La deuxième fois que le phénomène se produira sur la population nouvellement formée, on obtiendra un chapeau de base  $[b-2a, b+2a]$  avec des probabilités variables. Puis, si ce phénomène affecte plusieurs fois de suite la population nouvellement créée, on obtiendra la distribution normale:

$$f(x, \mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (18)$$

où  $x$  est la variable aléatoire,  $\mu$  la moyenne (espérance mathématique),  $\sigma^2$  la variance et  $f(x, \mu, \sigma^2)$  la densité de probabilité. Il existe des formes de ce théorème où la contrainte sur les lois de répartition est relâchée: il a donc une portée encore plus générale. Ce théorème signifie que si plusieurs phénomènes aléatoires indépendants ou faiblement liés s'additionnent, alors plus ils sont nombreux, plus on s'approche de la loi normale. L'exemple classique en est le tir dans une cible.

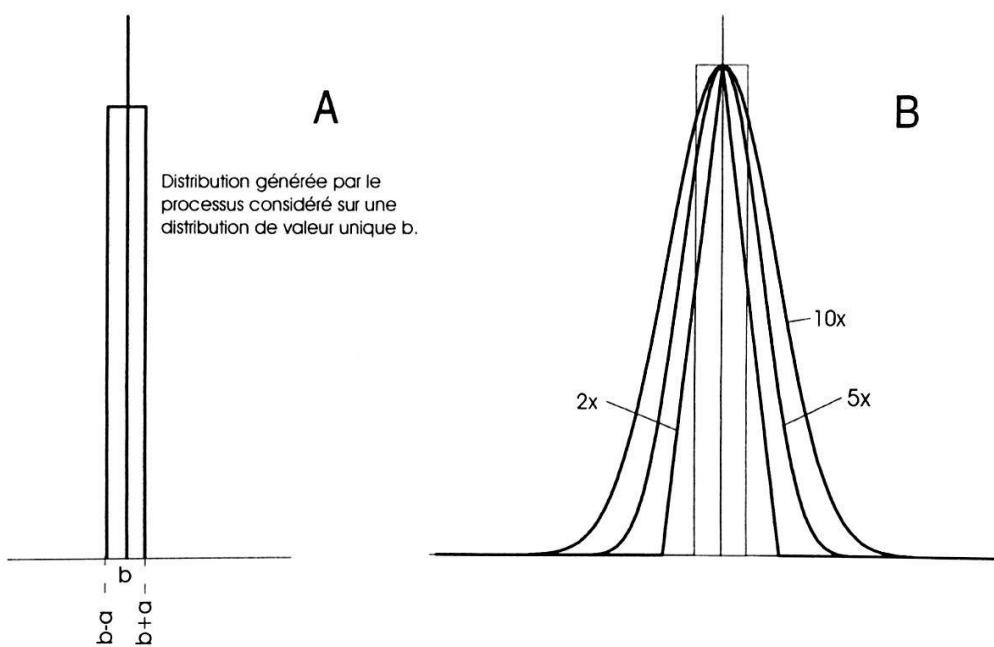


Figure 9.—Illustration graphique du théorème central limite.  $b$  est la valeur unique prise par l'ensemble des données avant que le phénomène n'affecte une fois (A) puis plusieurs fois ( $2x$ ,  $5x$  et  $10x$ ) (B) la distribution. Ceci est formellement égal à des produits de convolution successifs.

En géologie, des phénomènes d'agrégation devraient être de type normal. La loi normale s'applique par exemple à la quantité de silice dans les basaltes ou les rhyolites (SIEGEL 1974). En général, les teneurs en métaux suivent plu-

tôt la loi log-normale, comme par exemple la géochimie des *stream sediments* de zinc et cuivre du Guatemala (LEPELTIER 1969). La loi log-normale est plus fréquente. Elle est la conséquence de phénomènes proportionnels. On remplace simplement dans la loi normale la variable aléatoire X par le logarithme de la variable aléatoire Y, donc:

$$dx = \frac{1}{y} dy \quad (19)$$

Dans un phénomène de type proportionnel, on sait que deux étapes successives sont reliées par un coefficient  $\varepsilon_j$  et on peut alors écrire:

$$(y_j - y_{j-1}) = \varepsilon_j y_{j-1} \text{ ou encore } \frac{(y_j - y_{j-1})}{y_{j-1}} = \varepsilon_j \quad (20)$$

On considère alors les  $\varepsilon_j$  comme des variables aléatoires, et si on répète n fois le processus, on obtient:

$$\sum_{j=1}^{j=n} \frac{(y_j - y_{j-1})}{y_{j-1}} = \sum_{j=1}^{j=n} \varepsilon_j \quad (21)$$

Si n est très grand, on passe au continu  $(y_j - y_{j-1}) \Rightarrow dy$  et il vient:

$$\Rightarrow \int_{y_0}^{y_n} \frac{dy}{y} = \ln y_n - \ln y_0 \quad (22)$$

de sorte que

$$\ln y_n = \ln y_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \dots + \varepsilon_n \quad (23)$$

est une somme de variables aléatoires à laquelle on peut appliquer le théorème central limite. La densité de distribution log-normale est donnée par:

$$f(x, \mu_L, \sigma_L^2) = \frac{1}{\sigma_L \sqrt{2\pi}} e^{-\frac{(x-\mu_L)^2}{2\sigma_L^2}} \quad (24)$$

où x représente les valeurs prises par la variable aléatoire X ( $X = \ln Y$ ),  $\mu_L$  l'espérance mathématique du logarithme et  $\sigma_L^2$  sa variance (fig. 10). Mais on écrit la densité de probabilité log-normale  $l(\ln y, \mu_L, \sigma_L^2)$ :

$$P(Y \leq y_1, Y < y_2) = \int_{y=y_1}^{y=y_2} l(\ln y, \mu_L, \sigma_L^2) \frac{1}{y} dy \text{ car } x = \ln y \quad (25)$$

La relation entre la moyenne arithmétique  $\mu$  et géométrique  $\mu_L$  d'une variable distribuée log-normalement est donné par (AITCHISON and BROWN 1963):

$$\mu = e^{\left(\mu_L + \frac{1}{2}\sigma_L^2\right)} \quad (26)$$

et les variances sont reliées par:

$$\sigma^2 = \mu^2 \left( \left( e^{\sigma_L^2} \right) - 1 \right) \quad (27)$$

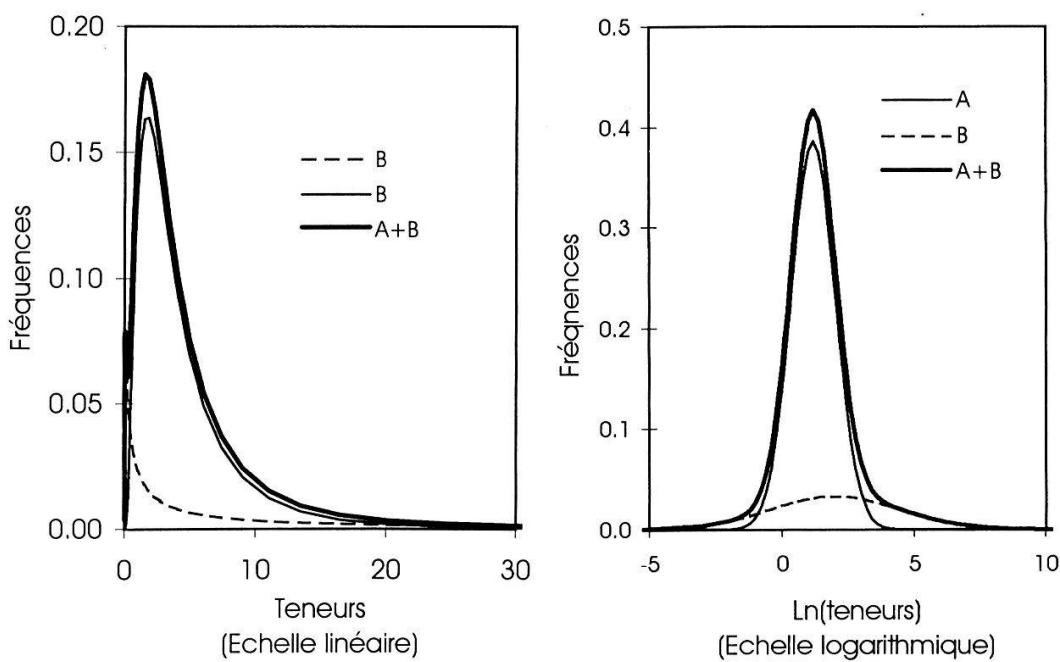


Figure 10.—Somme de deux distributions, A et B log-normales, représentées sur des échelles linéaire et logarithmique. Il faut surtout noter le déplacement du mode de la distribution B entre les deux échelles et le fait que la moyenne et la variance ne sont pas égales via le logarithme (ou l'exponentielle) d'une représentation à l'autre. L'exemple est volontairement particulier pour souligner l'effet du changement d'échelle.

Les réactions chimiques sont régies par des effets proportionnels (loi d'action de masse), de même que les processus de désagrégation (EPSTEIN 1947). Il n'est donc pas surprenant que plusieurs éléments chimiques soient répartis suivant cette loi. Les tailles de particules de sédiments de sols (granulométrie), etc. sont aussi fréquemment distribuées de cette façon.

Notons quelques restrictions à ces lois. Les teneurs infinies n'existent pas, par conséquent la loi log-normale ne s'applique pas aux minéraux à haute teneur, car il y a une limite supérieure (MATHÉRON 1962). Ceux-ci présentent plutôt une asymétrie inverse. Ce n'est pas notre propos. Les teneurs nulles, ce qui est possible pour certains éléments rares, donneraient la valeur moins infini dans le cas log-normal. La loi normale souffre du même problème de la limite supérieure, mais en plus il n'existe pas de teneur négative. Notons que la loi de Poisson peut s'appliquer pour les teneurs faibles, celle-ci décrivant des processus du type oui-non.

En pratique lorsqu'on modélise une distribution donnée par inférence statistique, on utilise les estimateurs  $m$  et  $S^2$  respectivement de la moyenne  $\mu$  et de la variance  $\sigma^2$ .

## 2.5. Utilisation du probability plot

Un histogramme cumulé tracé sur un tel graphique dessine une droite s'il s'agit d'une distribution normale ou log-normale au choix (fig. 2, 11 et 12); le principe étant le même, seule l'échelle des variables change.

Pour une loi normale, la médiane (50%) se confond avec la moyenne. Ainsi, la lecture de la moyenne est aisée. D'autre part, on peut très facilement mesurer la variance sur une droite de Henry. Il suffit de faire la différence

entre les valeurs de la variable à 15,9% et 84,1% pour obtenir la valeur de 2 écarts type:

$$\frac{x_{15.9\%} - x_{84.1\%}}{2} = S \quad (28)$$

On opère de manière inverse si on veut tracer une droite de Henry sur un *probability plot* (figures 11 et 2).

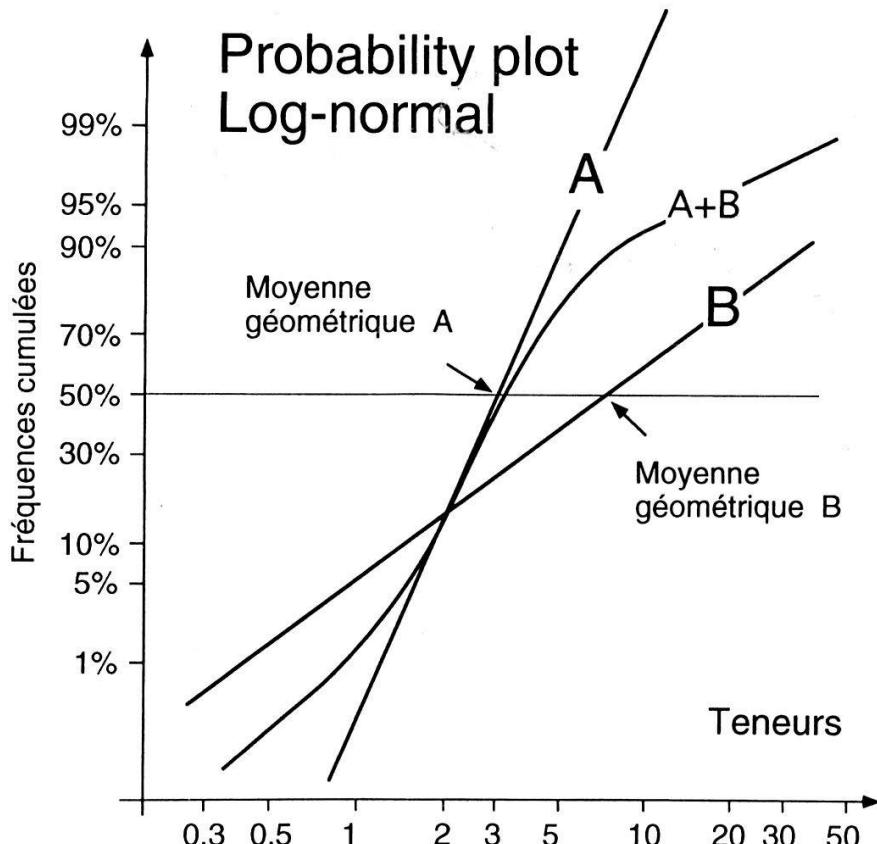


Figure 11.—Probability plot des deux distributions de la figure 10 et de leur somme. Ce cas est intéressant car les deux distributions ont des moyennes proches, mais des variances différentes. De ce fait leurs droites de Henry se recoupent. Par conséquent, la distribution la plus dispersée «domine» dans les valeurs faibles et élevées. C'est le cas ici de la distribution la moins importante (B). Remarquons que deux droites de Henry se recoupent presque toujours, mais que cela se passe souvent dans des domaines de valeurs qui ne nous concernent pas.

Généralement pour analyser des données, on examine simultanément l'histogramme (avec au moins six classes) et le *probability plot* adapté. S'il s'agit d'une population unique et qu'elle est normale ou log-normale, elle trace une droite sur le *probability plot*; il n'y a plus, en prospection minière, qu'à choisir un seuil assez élevé pour sélectionner quelques zones intéressantes. Pour des pollutions il faut déterminer si cette distribution est causée par un processus anthropogène. Cependant, il peut arriver que la distribution apparaisse comme une distribution unique cohérente mais sans être normale ou log-normale. Sur un *probability plot* on observe alors une courbe régulière à laquelle on appliquera un critère de «seuillage» (voir paragraphe 2.6.)

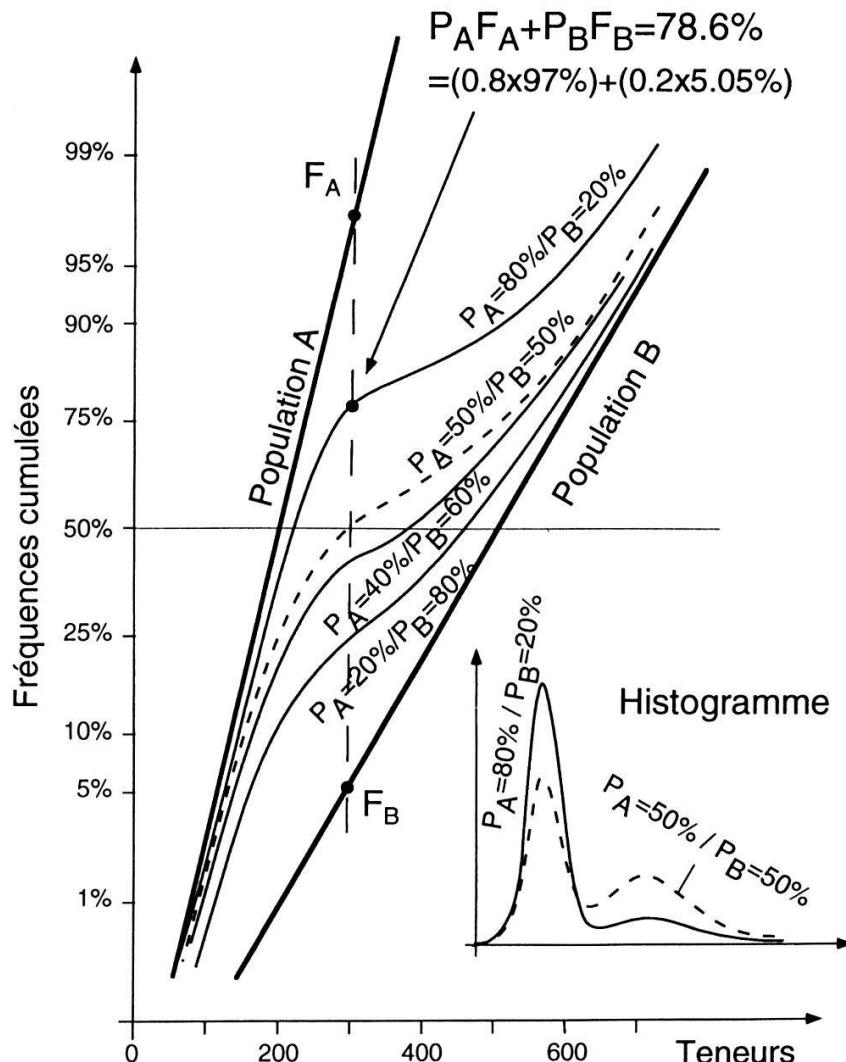


Figure 12.—Probability plot normal représentant la somme de distributions dont on a fait varier le poids. En petit, histogrammes de deux de ces courbes. Les points de plus forte courbure ou les points d'inflexion indiquent à peu près le poids des deux distributions.

On trouvera au paragraphe 4 des applications de la démarche suivante, qui traite de façon théorique l'analyse des *probability plot* en présence de plusieurs distributions. Si on soupçonne plusieurs distributions du même type que l'échelle du *probability plot* utilisé, on peut les mettre en évidence par l'inspection visuelle d'un histogramme qui présente plusieurs modes ou par conjectures. On essaie alors de caractériser les populations par leurs paramètres respectifs ( $S^2$ ,  $m$ ). Les calculs de ces derniers s'effectuent sur les populations supposées obtenues en décomposant l'histogramme en plusieurs «distributions». Ainsi, on peut représenter les droites de Henry de chaque distribution déduite sur le *probability plot* (SINCLAIR 1989, 1991). Pour vérifier la qualité de l'interprétation on construit la courbe cumulée de la somme des distributions en ajoutant les pourcentages cumulés (échelle du *probability plot*)  $F_i$  de chacune d'entre elles et en les multipliant par leur poids respectif  $P_j$ . De sorte que le pourcentage du mélange des distributions A, B, .... est donné par:

$$F_{A+B} = F_A P_A + F_B P_B + \dots \quad (29)$$

En présence de deux distributions, l'examen du *probability plot* des données peut permettre de les mettre directement en évidence. Un point de plus forte courbure ou un point d'infexion du *probability plot*, si une des populations domine, indique à peu près le poids respectif des deux distributions directement sur l'échelle des fréquences (fig. 11) comme on le constate sur l'exemple synthétique de la figure 12 où le point de plus forte courbure (78.6%) est proche des proportions respectives de chaque distribution (80% et 20%). Mais cette analyse doit être faite conjointement avec l'histogramme simple. Généralement la courbe du *probability plot* des données suit une droite proche de celle de la distribution la plus importante en proportion puis elle s'infléchit sous l'influence de la seconde distribution (exemples de cas particuliers aux figures 11 et 23). En s'appuyant sur le *probability plot* et l'histogramme, comme précédemment, on dégage la moyenne et l'écart-type de la distribution dominante, puis il est possible avec l'histogramme d'obtenir la seconde distribution par soustraction. Les paramètres qui permettent la construction d'un histogramme de loi normale sont donnés entre parenthèses à la figure 2. On applique à nouveau la formule 29 et comme précédemment on peut vérifier si les ensembles choisis «miment» bien les observations. Si ce n'est pas le cas on recommence, on révise l'interprétation et on réitère la démarche décrite ci-dessus. S'il y a plus de deux distributions superposées, l'interprétation est délicate et il faut avoir de bonnes raisons de penser qu'elles sont présentes, car on peut à la limite tout simuler avec un grand nombre de distributions.

Pour les données de pollution, la définition d'un *background* est de grande importance. Si le *background* possède une distribution normale ou log-normale, on peut obtenir sa distribution à l'aide d'un *probability plot*. Ceci est particulièrement utile dans le cas de la réhabilitation d'un site pollué, car le but d'une telle action est de retrouver le *background* naturel (FLEISHHAUER and KORTE 1990). FLEISHHAUER and KORTE (1990) indiquent que pour déduire un *background* normal ou log-normal, on peut calculer successivement le coefficient d'asymétrie en enlevant un à un les points de teneur les plus élevées, jusqu'à ce que le coefficient soit nul. Ils proposent aussi de décomposer directement les données en deux populations point par point, en les divisant en deux groupes, puis en reportant chacun d'eux sur un *probability plot* pour voir si ces groupes respectent les lois normales ou log-normales.

Dans cette phase, le traitement informatique est essentiel; il permet en effet de visualiser très vite tous les types de graphiques (les programmes statistiques courants sont par exemple: SPSS®, STATLAB®, S+® et en géostatistique le programme vétéran du domaine public GEOEAS, ENGLUND and SPARKS 1991). Ils permettent surtout de sélectionner les données pour n'en traiter qu'une partie.

## 2.6. Seuil et retour aux cartes

Une fois effectuée l'analyse univariée des données, examinons ce qui se passe spatialement. L'étape suivante est de trouver une limite inférieure (seuil d'anomalie) au-dessus de laquelle les valeurs sont intéressantes ou ont une signification particulière, afin de poursuivre l'objectif fixé.

Dans le cas d'un seul ensemble apparemment homogène de données, on tâche de choisir une limite qui mette en évidence des anomalies comme défini au paragraphe 2.1. En début de campagne, on ne trouve souvent que des points

anomaux isolés. Puis on cherche à établir, par l'acquisition de données supplémentaires à l'aide d'une maille plus serrée (paragraphe 2.3), s'ils appartiennent à des zones à teneurs élevées. Il n'y a pas de règle absolue: souvent, on choisit les 3 à 5% supérieurs de la distribution (moyenne +2S). Il ne faut pas perdre de vue les dimensions de la maille utilisée, et il est important de se rapporter à la carte géologique ou autre, car elle permet de dire si des sites peuvent être intéressants ou non. Une teneur relativement élevée corrélée avec un site favorable (géologique, pédologique, hydro-géologique, etc.) est à retenir même si un seuil relativement haut la faisait disparaître.

Si on a pu mettre en évidence avec ou sans *probability plot* deux ou plusieurs distributions, comme décrit au paragraphe 2.5, la détermination du seuil est en principe plus facile. Soit  $F_{\text{sup}}$  et  $F_{\text{inf}}$  les valeurs des fonctions de répartition (échelle cumulée du *probability plot*), respectivement de la population à hautes teneurs et celles de l'ensemble des populations à plus basses teneurs. On placera le seuil en principe au plus bas, là où la distribution la plus haute en teneurs  $F_{\text{sup}}$  dépasse l'autre en pourcentage c'est-à-dire:

$$F_{\text{sup}} P_{\text{sup}} > F_{\text{inf}} P_{\text{inf}} \quad (30)$$

où  $P_{\text{sup}}$  et  $P_{\text{inf}}$  sont les proportions de chacune des distributions.  $F_{\text{sup}}$  ne respecte pas forcément une loi log-normale. De nouveau, il est préférable alors de contourer ces zones anomalies. Si celles-ci sont trop étendues, on tente une nouvelle interprétation pour dégager un seuil plus élevé. On sera particulièrement attentif à la répartition spatiale de la distribution supérieure, afin d'établir s'il y a des corrélations de sites pédologiquement, géologiquement, hydro-géologiquement, etc., particuliers avec cet ensemble de valeurs.

## 2.7. Discriminer

Les traitements décrits plus haut peuvent être appliqués en réduisant le nombre d'échantillons, c'est-à-dire en imposant des critères de sélection. On pourra, par exemple, ne travailler que sur un environnement particulier ou ne traiter les données qu'au-dessus d'un certain seuil, ou avec la présence d'un seul élément si on en analyse plusieurs à la fois, ou encore effectuer des seuillages croisés entre éléments.

## 3. AUTRES OUTILS

### 3.1. Evaluation quantitative des anomalies

Cette méthode est utile en prospection minière et pour la quantification de l'ampleur de pollutions. Elle est une aide au choix des meilleures cibles (figure 13, BARBIER 1989). Elle consiste à mettre en regard le tonnage par mètre d'approfondissement ou simplement la teneur  $t_i$  (ou tonnage dans le premier mètre de sol) avec la surface. Cela est particulièrement utile pour comparer le potentiel des anomalies; en effet, certaines anomalies peuvent apparaître comme assez vastes, mais d'autres, plus petites, peuvent s'avérer plus importantes et il peut être intéressant de savoir à quelle teneur l'une devient plus importante que l'autre en « quantité de métal ». Une mesure comparable et le tonnage/mètre est donné par:

$$\rho \times t_i \times S(t_i - \Delta t, t_i + \Delta t) \quad (31)$$

où  $\rho$  est la masse volumique du matériel analysé,  $t_i$  la teneur et  $S(t_i - \Delta t, t_i + \Delta t)$  la surface occupée par les valeurs comprise entre  $t_i - \Delta t$  et  $t_i + \Delta t$ . Pour un domaine dont la teneur est supérieure à une teneur donnée, le tonnage par mètre est donné par:

$$T(t_i) = \rho \sum_{j=1}^{j=n} t_i S_j (t_i - \Delta t, t_i + \Delta t) \quad (32)$$

$$\text{avec } n \text{ entier de } \frac{t_{\max} - t_{\min}}{\Delta t} \text{ et } t_i = (t_{\max} - j \times \Delta t) \quad (33)$$

où  $t_{\max}$  et  $t_{\min}$  sont les teneurs maximale et minimale.

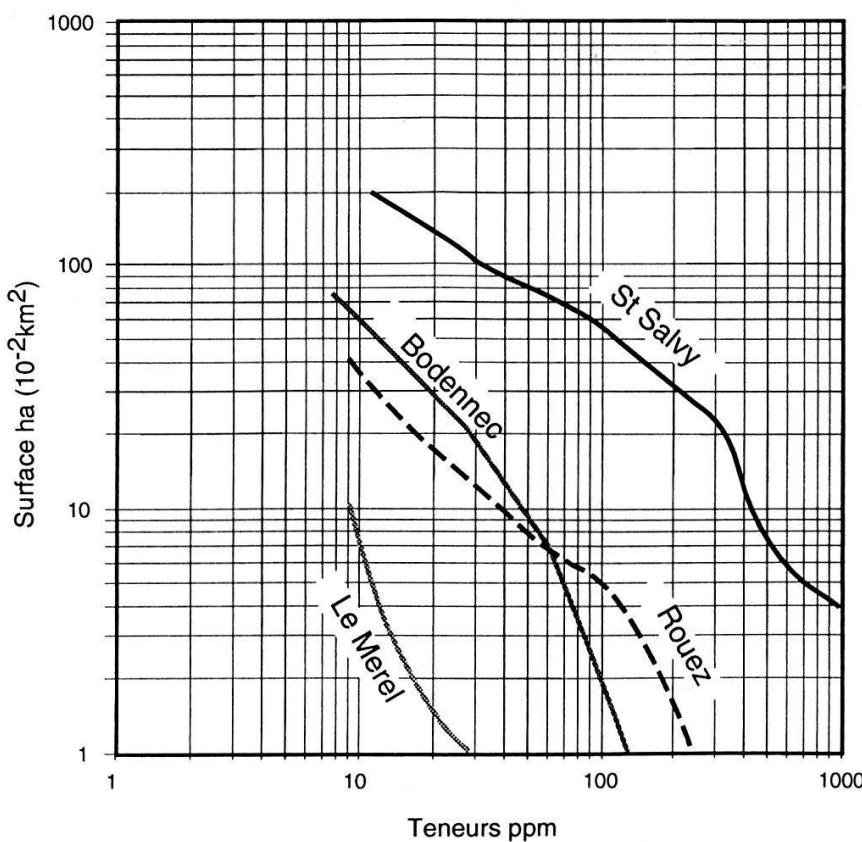


Figure 13.—Courbes de teneurs-surface en Pb pour quatre minéralisations sulfurées en France (d'après BARBIER 1989). Les surfaces sont délimitées par les courbes isano-males. On observe que l'anomalie de St Salvy est toujours plus importante en tonnage que les autres. Par contre Rouez est une anomalie plus grande que Bodennec seulement à partir de 60 ppm.

### 3.2. Le variogramme

Le variogramme est l'outil de base de la géostatistique (JOURNEL 1989, MATHÉRON 1962, LOKOSHA 1991, ISSAKS and SRIVASTAVA 1989, BRUNO et RASPA 1994, PANNATIER 1996). Il permet de mettre en évidence le lien spatial pouvant exister entre des points de prélèvements. On le définit comme:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(d_i) - z(d_i + h)]^2 \quad (34)$$

où  $N(h)$  est le nombre de paires de points se trouvant à une distance  $h$  les uns des autres,  $z(d_i)$  la teneur et  $d_i$  la distance (fig. 14) (ces notations sont standard). De façon formelle le variogramme se note comme la variance de l'accroissement de la valeur  $z(\dots)$  entre deux points séparés d'une distance  $h$ :  $\text{Var}[z(d_i) - z(d_i + h)]$ , ou bien comme l'espérance mathématique des accroissements  $E[(z(d_i) - z(d_i + h))^2]$ .

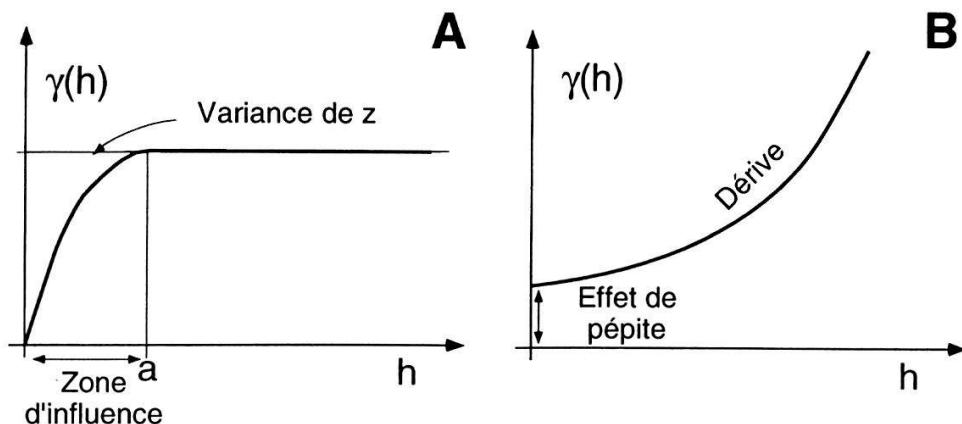


Figure 14.–En A schéma d'un variogramme stationnaire et en B variogramme non stationnaire avec effet de pépite.

De façon pratique, il faut faire des classes de distance et même déterminer des directions de recherche. Le grand intérêt du variogramme est de définir une zone d'influence, c'est-à-dire d'indiquer jusqu'à quelle distance les variables sont corrélées, ou à partir de quelle distance les variations deviennent aléatoires. Plusieurs variogrammes calculés dans différentes directions permettent aussi de mettre en évidence d'éventuelles anisotropies, sous réserve de l'hypothèse de stationnarité. Le variogramme est égal à la variance lorsqu'on est en dehors de la zone d'influence (cas stationnaire). En d'autres termes lorsque l'espace a les mêmes propriétés partout, on obtient un plateau à partir de la zone d'influence. Dans certains cas, le variogramme présente une variation de pente tout en continuant d'augmenter: on est alors en présence d'une dérive régionale (cas non stationnaire). Le comportement à l'origine du variogramme est important mais difficile à obtenir, car il faut des prélèvements rapprochés. Lorsque le variogramme n'est pas nul à l'origine, une teneur supérieure à zéro peut directement jouxter une teneur nulle: c'est l'effet de pépite (en environnement, l'exemple typique est une dispersion de particules polluantes).

Le variogramme est complété par la covariance spatiale et le corrélogramme. La forme du variogramme donne de précieux renseignements sur la maille d'échantillonnage qui doit être en tout cas inférieure à la zone d'influence, sinon les points de mesure n'appartiennent en moyenne plus à la même anomalie. En réduisant la maille de prélèvement on augmente la précision, puisque la variance va diminuer depuis le palier du variogramme jusqu'à l'effet de pépite. Il faut chercher si le variogramme présente plusieurs structures, comme par exemple un plateau intermédiaire. Il faut remarquer que le

variogramme peut montrer, entre autre, la taille de l'anomalie étudiée, ce qui est une information importante. Ceci peut se produire lorsqu'on effectue le variogramme sur une zone riche, car alors le *background* n'est pas suffisamment représenté.

Cet outil a pris un essor considérable dans le traitement des pollutions, dans le but de définir les dimensions des anomalies et la corrélation spatiale des teneurs élevées (MAIGNAN and MAIGNAN in press, ATTEIA *et al.* 1994, FLAMM 1994, WEBSTER and OLIVER 1990).

#### 4. EXAMPLES

#### 4.1. Exemples miniers

Nous avons analysé «à la main» des données géochimiques en cuivre (ppm), qui proviennent d'une prospection géochimique minière systématique en roches de surface dans la région désertique du Hoggar -Algérie du sud- (DJEDDOU 1991). Les prélèvements ont été faits en moyenne tous les 20 m selon des profils plus ou moins continus espacés de 250 ou 500 mètres. La carte a été dressée par l'ajustement de séries de Fourier à deux dimensions (programme utilisé par l'Institut de Géophysique de l'Université de Lausanne) par la méthode des moindres carrés, ce qui explique que les très hautes valeurs aient disparu (fig. 15). Ceci se justifie parce que les différents types de roches assimilent de façon variable les éléments chimiques. Dans le cas présent le lisage des valeurs est donc applicable. Les anomalies ne sont que partiellement centrées et l'information n'est pas parfaitement homogène. On peut donc s'attendre à avoir des anomalies sur-représentées. En fait, en pondérant par 1/2 le nombre de points, l'ensemble de la zone où les profils sont doublés (formules 14 à 16), donc là où le nombre de prélèvements est deux fois supérieur au reste de l'étude, on vérifie que l'histogramme n'est que peu modifié (fig. 16).

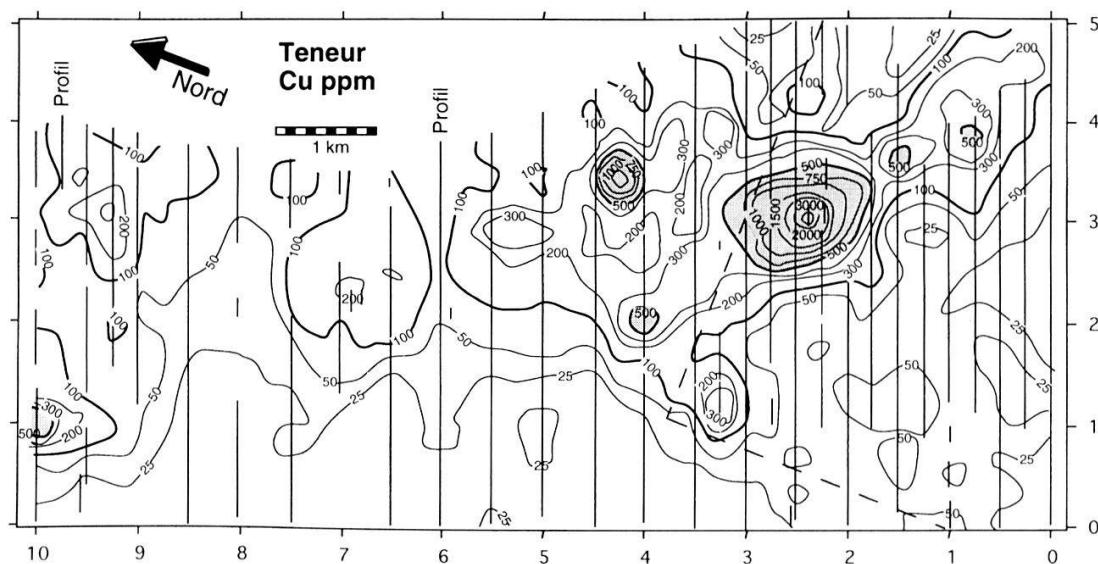


Figure 15.—Carte de géochimie de surface. Les lignes NE-SW indiquent les profils de prélèvements qui sont espacés de 250 m dans la partie sud (d'après DJEDDOU 1991).

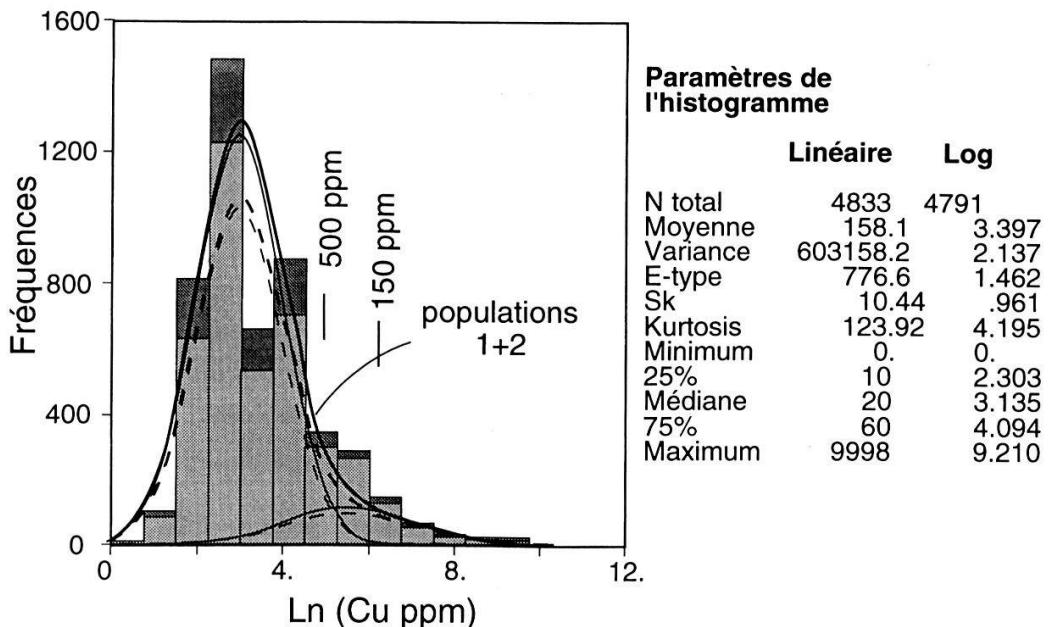


Figure 16.—Histogramme en logarithmes des teneurs Cu ppm. On peut noter que les caractéristiques en échelle linéaire ou logarithmique sont évidemment différentes. Les formules 26 et 27 qui relient les moyennes dans les échelles normale et log-normale, nous donnent les moyennes et les écarts types arithmétiques des deux distributions:  $m_1 \approx 33$ ,  $\sigma_1 \approx 43$  et  $m_2 \approx 755$ ,  $\sigma_2 \approx 2000$  ppm. Ceci caractérise bien le background (le clarke ou teneur moyenne de la croûte terrestre est de Cu=55 ppm) et les teneurs élevées (distribution 2). On rappelle que la moyenne est toujours plus élevée que la médiane dans une distribution log-normale; cela ne contredit pas le seuil choisi. L'histogramme le plus clair est celui pour lequel on a pondéré les données par 1/2, dans la zone délimitée par un traitillé dans la figure 15. On voit que la même interprétation (histogramme en traitillés) est valable au nombre de valeurs près. Remarquons que le nombre de points considérés en échelle logarithmique est moins grand qu'en échelle linéaire car les point de teneurs nulles (invalides) ne sont pas pris en compte.

L'histogramme logarithmique (fig. 16) présente une très nette asymétrie en échelle log-normale vers les hautes teneurs et sa décroissance n'est pas régulière. Nous pouvons supposer qu'il se compose de deux populations log-normales. Le *probability plot* présente un comportement relativement linéaire jusqu'à 80%, que représenterait une première distribution dont le poids est estimé à 0.88 (88% de l'effectif total, fig. 17); c'est à cette valeur environ qu'on observe en effet la plus grande courbure. Lorsqu'une des distributions est dix fois supérieure à l'autre, le point d'inflexion est trop décalé, et de plus nous ne sommes plus en présence d'un cas purement log-normal. A l'aide de l'histogramme, nous avons estimé les moyennes et les variances des distributions. La plus importante a une moyenne de 3.0 ln (Cu ppm) et un écart type de 1. La seconde une moyenne de 5.5 et un écart-type de 1.5 ln (Cu ppm). Avec ces informations, on peut construire le *probability plot* simulé.

Après avoir mis en évidence ou créé deux distributions, il faut trouver un seuil. La distribution 2 domine au-dessus de 150 ppm environ (5 ln (Cu ppm)). Cela représente 10% des données, ce qui est considérable, et occupe une trop grande surface. La distribution étant étalée, il est préférable de considérer l'ensemble des valeurs lorsque l'influence de la première distribution est qua-

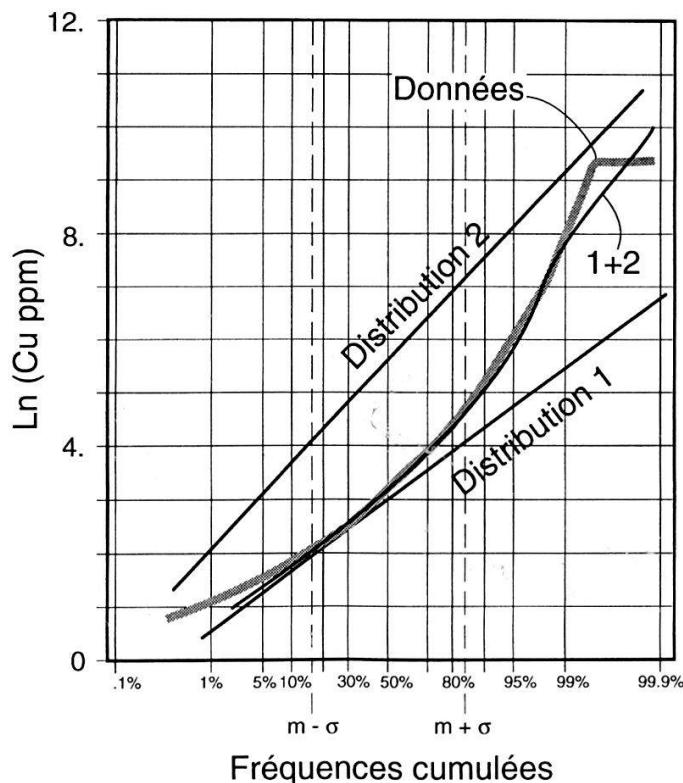


Figure 17.—Probability plot de l'histogramme de la figure 16. On observe une bonne concordance entre la courbe réelle et la simulation. Le poids des deux distributions a été en partie choisi à l'aide de l'histogramme simple, car le point d'inflexion ne correspond à rien sur le probability plot. C'est le point de plus grande courbure qui a été choisi.

siment nulle. Cela se produit autour de 500 ppm ( $6.2 \ln (\text{Cu ppm})$ ), et représente 5% des échantillons. La répartition des zones ainsi définies sur la carte est assez satisfaisante, car elle met en évidence quelques zones anomalies bien individualisées, de 200 m à plus d'un kilomètre d'extension, dimensions tout à fait raisonnables pour l'implantation de sondages à la tarière, par exemple. Naturellement, les données géologiques vont aussi intervenir dans le choix des sites de sondage. Ces zones présentent une grande probabilité de concentrations élevées provenant de processus anormaux (minéralisateurs?). Le premier seuil ne mettait en évidence qu'une grande zone étalée, qui ne permettait pas de choisir des cibles précises pour les sondage.

L'interprétation ci-dessus est statistique, et on se pose la question de sa signification génétique. La première distribution est censée représenter un *background*, c'est-à-dire les valeurs normales des teneurs en cuivre dans cette région. La seconde représenterait un apport dû à des roches plus riches en cuivre et redistribué chimiquement dans les zones proches de ces roches. D'autres explications sont envisageables. Si on suppose que la distribution régionale de base est un peu plus faible et moins variable, on peut inférer que l'asymétrie soulignée par un maximum local sur l'histogramme est plutôt due à l'effet de zones à très haute teneur, qui auraient diffusé dans les zones proches (fig. 7, 18 et 19); et comme la zone de diffusion est limitée, cela expliquerait le maximum local de l'histogramme. Ce type de phénomène génère de telles asymétries, même en échelle logarithmique (JABOYEDOFF and MAIGNAN in prep.). On peut aussi observer qu'un profil au travers d'une anomalie est assez proche de celui d'une diffusion.

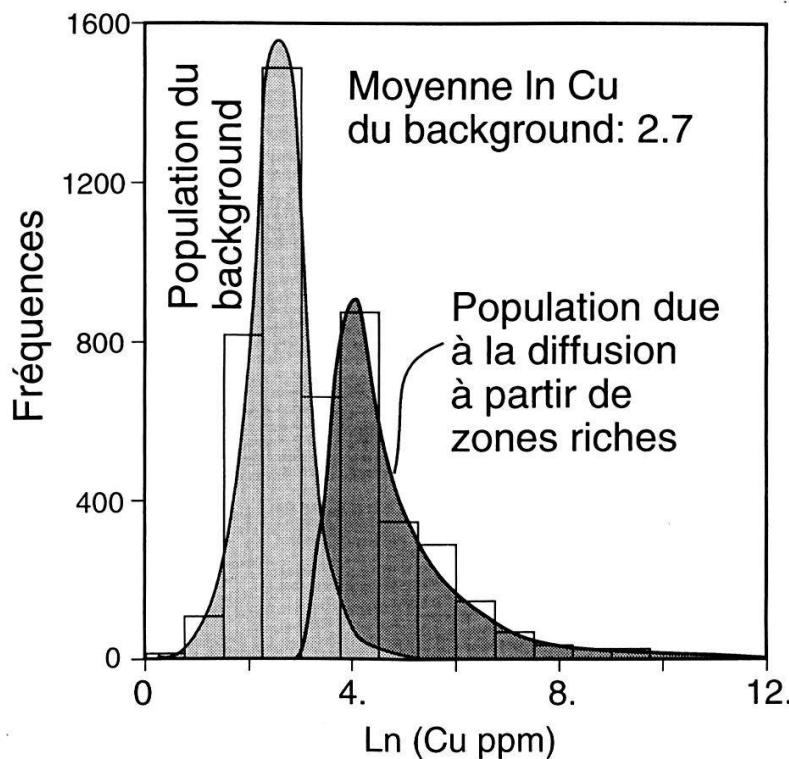


Figure 18.–Exemple d’interprétation de distributions dues à une diffusion et à un background log-normal. La moyenne arithmétique de la distribution du background est d’environ 20 ppm Cu et l’écart type de 10 ppm Cu. C’est peu, mais si on considère que les teneurs au-dessus de 50 ppm sont inexistantes, il s’agirait d’un background sans aucun effet de l’anomalie, ce qui n’est pas le cas de l’interprétation précédente. Le reste proviendrait de l’addition du background et de la diffusion des zones riches, qui représentent environ 30 à 40% des échantillons.

Il importe peu que, dans un premier temps, les populations possèdent ou non une « réalité » génétique, car cette méthode permet de trouver des critères de choix simples à partir des données et non pas à partir de choix préétablis. Par la suite, une identification des populations dans un sens génétique permettra de valider ou non les hypothèses faites à partir des statistiques, et par là de tenter une hiérarchisation des phénomènes. Dans l’exemple précédent il n’y a pas de différences majeures entre les deux interprétations, dont les populations jouent essentiellement le même rôle, bien que les valeurs soient légèrement différentes.

#### 4.2. Un exemple de variogramme

L’exemple qui suit montre les difficultés liées à l’échantillonnage et au calcul pratique d’un variogramme. Nous avons calculé le variogramme (formule 34) pour les teneurs en cuivre (ppm) en géochimie de roches représentées à la figure 20. Il a été effectué parallèlement aux profils avec une tolérance de 5° et un pas de 75 m pour une tolérance de 10 m. D’emblée, on s’aperçoit qu’il n’y a pas de stationnarité. En effet, sur les 2000 premiers mètres de portée du variogramme sa valeur ne fait qu’augmenter. Ceci est en principe dû à une tendance régionale. Ici, les anomalies sont importantes par rapport à la surface de l’étude et elles forment un « îlot » dont le diamètre dépasse localement les

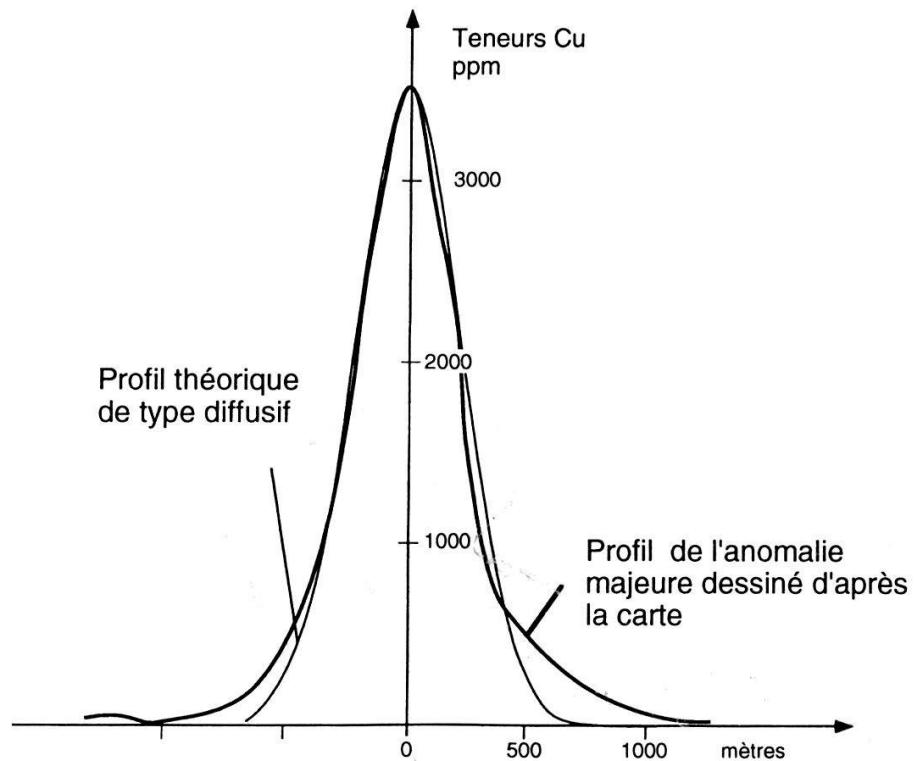


Figure 19.—Profil des teneurs à travers le maximum de l'anomalie principale. La courbe a été établie à partir de la carte (lissage) et parallèlement au profil. Le profil théorique est très proche de celui de l'anomalie. Notons qu'une moyenne mobile sur les teneurs brutes donne un résultat tout à fait similaire.

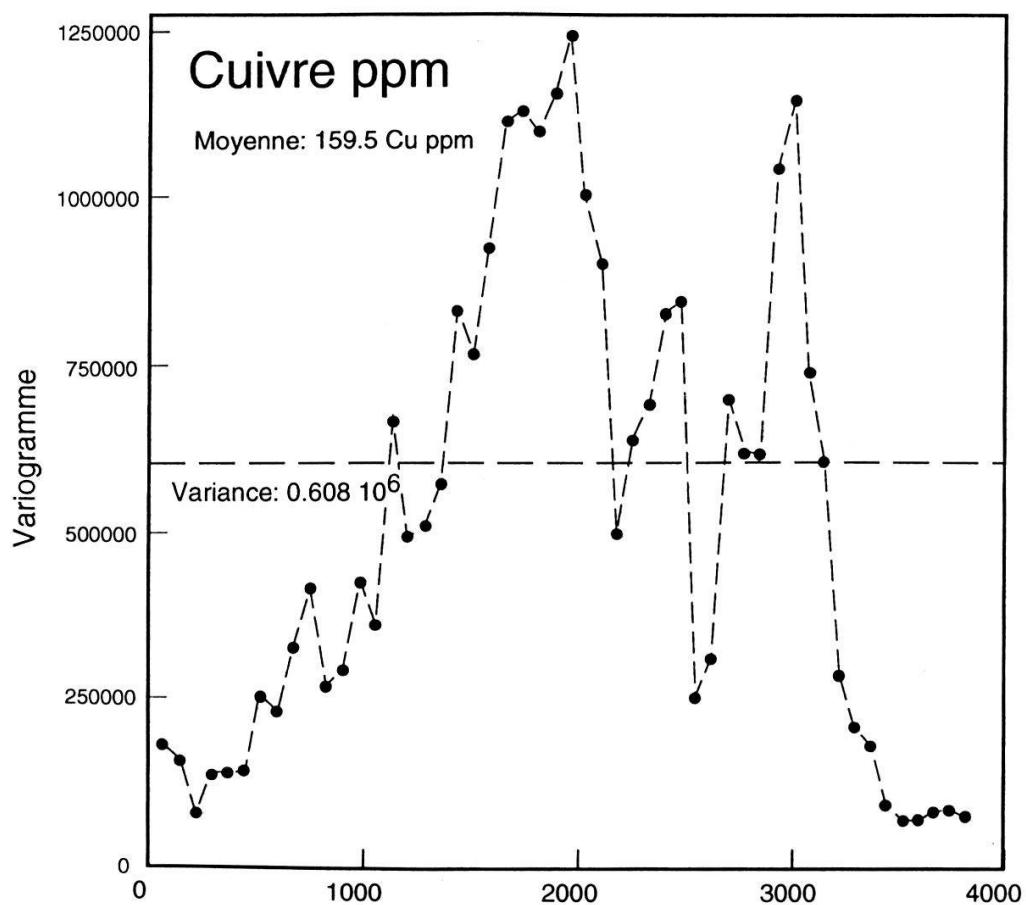


Figure 20.—Variogramme des données de la figure 15. La direction de recherche est celle des profils  $\pm 5^\circ$  et le pas de  $h$  est de 75 m  $\pm 10$ m.

3 km. Les fluctuations de cette dérive sont vraisemblablement dues aux zones dont la teneur est supérieure à 300 ppm Cu, car lorsque  $h$  est égal à la distance séparant deux maxima de ces zones, le variogramme diminue. La diminution rapide au delà de 2000 m suggère un effet de bordure: en effet le centre de l'anomalie principale se trouve à 2 km du bord de l'étude et son importance diminue alors très vite. Les autres variations doivent être liées au même type de problème ou au fait que les profils ne sont pas tous continus, donc qu'il y a un problème de tolérance. A 3500 m environ, on ne perçoit plus que les variations de *background*.

#### 4.3. Exemple de traitement de pollutions

L'exemple proposé provient de données collectées par le laboratoire cantonal d'agronomie de Jussy GE (CÉLARDIN *et al.* 1992, fig. 21). 102 sites ont été analysés (grande culture: 65, prairie: 15, vigne: 9, forêt: 9, arboriculture: 1, culture maraîchère: 1 et réserve naturelle: 1) sur quatre niveaux différents (0-20, 20-40, 40-60, 60-80 cm), mais tous n'ont pas toujours été analysés (102, 101, 95, 73). Les méthodes d'analyse sont décrites dans CÉLARDIN *et al.* (1989). Parmi ces données nous avons choisi les teneurs en zinc.

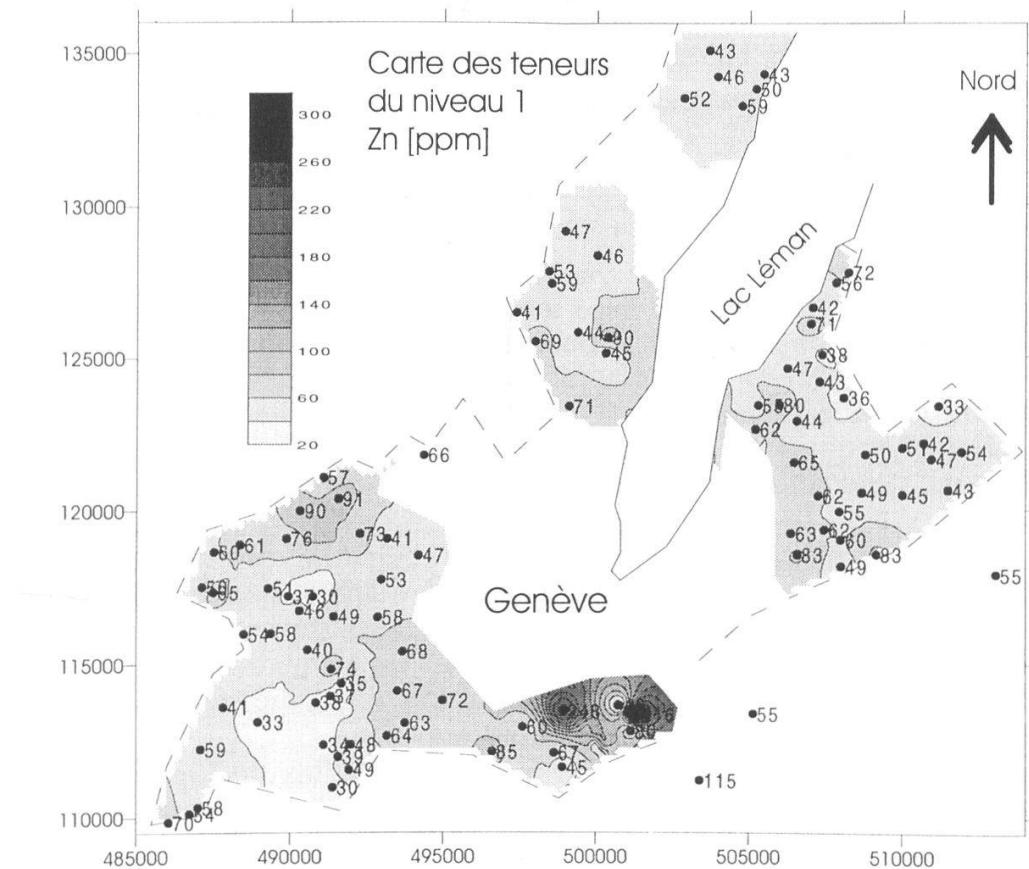


Figure 21.—Carte des teneurs Zn établie pour le niveau 1 par la méthode d'inverse distance d'ordre 2. L'effet de pépite important du variogramme autorise cette méthode pour la représentation. Le rayon de recherche pour l'interpolation est de 2500 m et la grille carrée possède un pas de 250 m et 6 points au maximum sont utilisés. On constate les effets de bord inhérents à toute interpolation. On note les grandes structures du background (voir texte).

Enumérons quelques-unes des propriétés du zinc sans entrer dans les détails du comportement de cet élément dans les sols. C'est un oligo-élément chalcophile (tendance à former des liaisons covalentes) dont la teneur moyenne dans la croûte terrestre est de 70 ppm. L'ordre de grandeur des teneurs dans les sols est identique et le domaine de variation d'environ 10-300 ppm. Les plantes manquent de Zn lorsqu'elles en contiennent moins de 20 ppm et la teneur est toxique lorsqu'elle dépasse 400 ppm; pour les animaux le zinc est toxique s'ils en contiennent plus de 1000 ppm (BOLT and BRUGGENWERT 1976).

La quantité et la mobilité du zinc sont affectées par plusieurs facteurs. C'est essentiellement le pH qui contrôle sa mobilité. D'une façon générale, la solubilité du zinc décroît avec le pH, d'environ 100 fois par unité (ASRARUL HAQUE and SUBRAMANIAN 1982). Dans les sols, le zinc, selon IYENGAR *et al.* (1981), se présente sous différentes formes: adsorbé échangeable ou non ( $\approx 4\%$ ), lié à la matière organique ( $\approx 2\%$ ), associé aux hydroxydes de Mn ( $\approx 2\%$ ), associé aux hydroxydes d'Al et Fe ( $\approx 24\%$ ) et 68% sont résiduels, c'est-à-dire essentiellement insérés ou intimement associés aux minéraux de la fraction argileuse.

Le stock de zinc utilisable par les plantes provient essentiellement de la fraction échangeable et de la matière organique. De faibles pH provoquent la libération du zinc associé aux hydroxydes (Fe, Mn), et favorise sa lixiviation: elle est observée dans les horizons superficiels des sols forestiers acides du canton de Genève alors qu'elle ne l'est pas dans les sols plus alcalins, où le lessivage est moins intense (CÉLARDIN et CHATENOIX 1990).

Une part importante de zinc dans les sols peut avoir une origine anthropique. Par exemple, certains engrains phosphatés peuvent contenir jusqu'à 0.1% de Zn (WEDEPOHL 1969), cet élément peut même être ajouté comme nutriment (TROEH and THOMPSON 1973).

En raison du nombre élevé de variables chimiques qui contrôlent le comportement du zinc dans les sols, bien qu'un petit nombre prédomine, on supposera qu'on peut obtenir pour le *background* une distribution log-normale (paragraphe 2.4): ceci est conforme aux observations faites dans le Jura. De plus le zinc possède un comportement géochimique relativement différent des autres éléments qui se regroupent par affinités (WEBSTER *et al.* 1994). Il semble donc cohérent de supposer que le fond géochimique est de type log-normal.

L'examen des histogrammes normés pour chaque niveau montre que les populations sont assez similaires, avec une légère diminution avec la profondeur. Seuls les niveaux 1 et 2 présentent des teneurs supérieures à 100 ppm. Notons que certaines des données utilisées pour les histogrammes se trouvent en dehors de la carte (fig. 21 et 22), elles ont tout de même été utilisées vu le nombre relativement restreint d'échantillons.

L'examen du *probability plot* du niveau 1, en échelle logarithmique, montre clairement une population qui suit une distribution log-normale (figure 23). Celle-ci peut être attribuée à un *background*; quelques valeurs plus élevées n'en font pas partie (voir paragraphe 2.5). Ainsi on peut, à partir du *probability plot*, déduire les caractéristiques de la distribution du fond géochimique, dont la moyenne  $\ln$  (Zn ppm) vaut 4 et l'écart type 0.25. En échelle linéaire on obtient respectivement 56.3 ppm et 14.3 ppm Zn (tab. 1). On constate qu'il existe une population anomale au-delà de 4.6  $\ln$  (Zn ppm), c'est-à-dire qu'au-dessus de 100 ppm plus des deux tiers des données sont anomalies. Si on ne

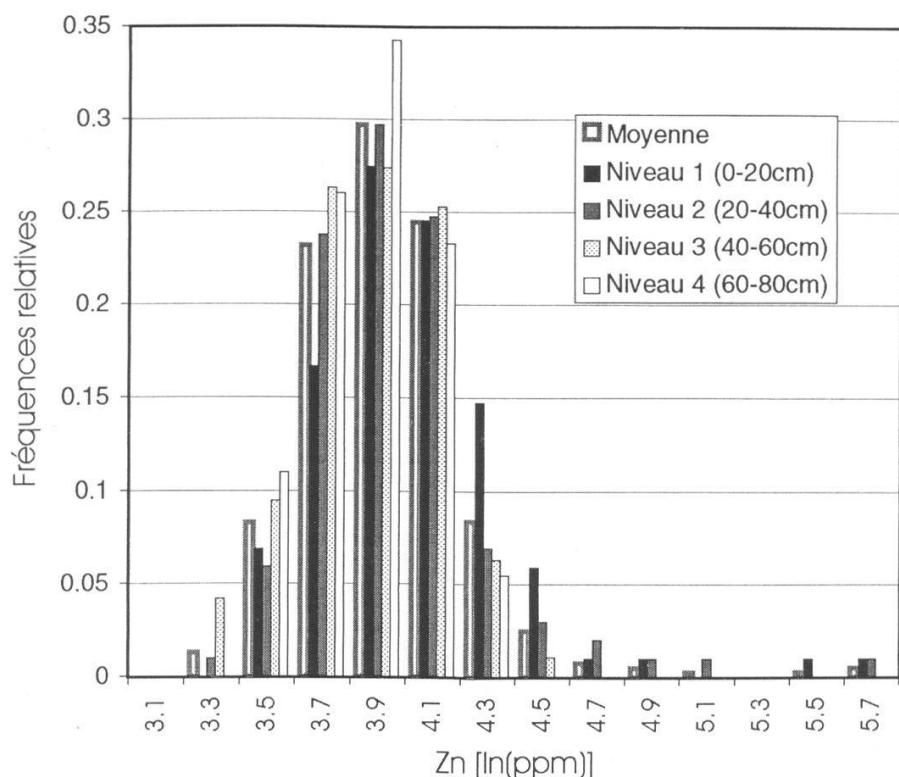


Figure 22.—Histogrammes relatifs des teneurs dans les différents niveaux et l'histogramme moyen. On constate que tous les niveaux possèdent des distributions proches de la distribution log-normale.

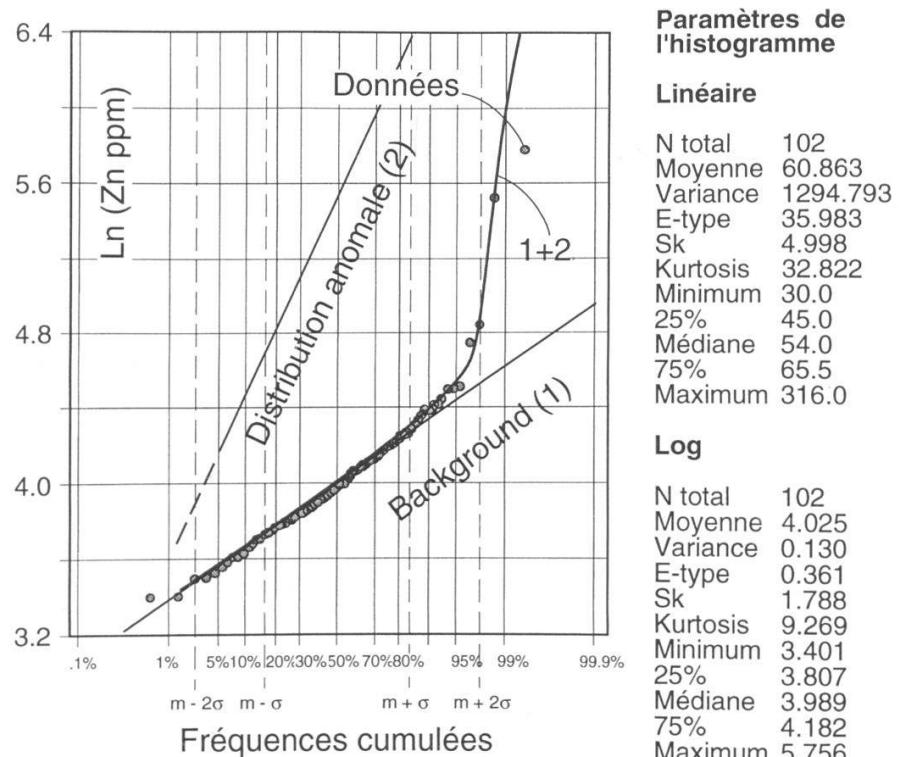


Figure 23.—Probability plot de  $\ln$  (Zn ppm) du niveau 1. La tendance log-normale du background est assez claire. La modélisation est effectuée avec un poids de 97% pour le background (1) et 3% pour la distribution anomale (2).

Tableau 1.—Paramètres statistiques des populations des différents niveaux. Ceux du niveau 1 sont recalculés pour les teneurs inférieures à 100 ppm Zn et comparés à la population déduite du probability plot.

	Moyenne arithmétique	Ecart type arithmétique	Moyenne logarithmique	Ecart type logarithmique	Coefficient d'asymétrie
Niveau 1	60.8	35	4.02	0.36	1.78
Niveau 2	57.3	32.5	3.97	0.26	1.91
Niveau 3	49.1	12.1	3.86	0.26	-0.17
Niveau 4	49.1	10.6	3.87	0.22	-0.03
Niveau 1 <100 ppm	55.0	14.0	3.97	0.26	0.01
Valeur de la population déduite (background niveau 1)	56.33	14.3	4.0	0.25	0

tient pas compte des données élevées on s'aperçoit que les caractéristiques de la distribution résultante sont très proches de la distribution log-normale déduite à l'aide du *probability plot* (tab. 1). Il existe donc des teneurs anomalies, dont on ne connaît pas le type de distribution vu le nombre restreint de données. On peut néanmoins tenter de l'approcher à l'aide d'une distribution log-normale de moyenne 5.5 ln (Zn ppm) et d'écart-type de 0.8 ln (Zn ppm). La simulation obtenue est bonne (fig. 23). Il faut remarquer que pour les faibles valeurs, même si en quantité cela est négligeable, la population anomale devient plus importante que l'autre, ce qui est gênant. Par conséquent on suppose que l'approximation faite ici n'est valable que pour les teneurs suffisamment élevées.

Il existe donc clairement une source autre que le *background* dans la partie superficielle, qui est probablement d'origine anthropogène comme l'ont souligné CÉLARDIN *et al.* (1992). La légère différence qui existe entre les niveaux supérieurs (fig. 22 et tab. 1), même «nettoyés» de leurs teneurs anomalies, et les niveaux inférieurs provient vraisemblablement du fait que les échantillons qui sont issus des cultures sont en plus grand nombre que ceux provenant des forêts qui présentent un enrichissement vers le bas (CÉLARDIN et CHATENOIX 1990, LANDRY et CÉLARDIN 1988). La différence entre les teneurs des bois et des cultures n'est pas suffisante pour provoquer une bimodalité des teneurs, ce qui, vu le nombre restreint d'échantillons, nous a autorisés à tous les traiter ensemble.

Les points anomaux qu'on observe sur la carte des teneurs du niveau 1 (fig. 21) ne permettent pas de définir l'étendue des anomalies. Elles ne doivent pas dépasser un kilomètre de diamètre, et il faudrait donc resserrer la maille à ces endroits. Par contre, un variogramme effectué uniquement sur les valeurs inférieures à 100 ppm montre un léger lien jusqu'à 2500 m avec un fort effet de pépite. C'est bien approximativement le rayon des zones de teneurs légèrement plus élevées au sein du *background*. On ne peut pas juger si les valeurs du *background* sont légèrement enrichies par rapport aux teneurs naturelles. Dans le Jura (ATTEIA *et al.* 1994), la teneur moyenne du zinc est de 78.5 ppm et l'écart type de 38.5 ppm (en logarithme naturel la moyenne vaut 4.3 ln (Zn ppm), l'écart type 0.18 ln (Zn ppm) et le coefficient d'asymétrie 0.28). Nos valeurs ne sont donc pas très élevées. En fait les valeurs élevées du *background* semblent associées à la molasse châtienne et les valeurs basses aux alluvions. L'apport extérieur de zinc, s'il existe, est tout à fait raisonnable au regard des normes OSOL (limite supérieure: 200 ppm). On peut considérer

qu'au-delà de 100 ppm l'essentiel des valeurs sont liées à des processus anormaux, et par conséquent dans le cas d'une réhabilitation de site, qu'il faudrait revenir au-dessous de cette valeur.

Notons que dans le cas de pollutions par appauvrissement des teneurs, ce qui est envisageable pour le zinc qui est un oligo-élément, le problème peut être traité de la même manière mais l'échelle des teneurs est inversée.

## 5. MISE EN ŒUVRE INFORMATIQUE

Au niveau d'une analyse extrêmement simple déjà, des descriptions statistiques et des représentations cartographiques élaborées peuvent être aisément obtenues sur PC.

Tout en étant simples d'utilisation, ces programmes performants sont disponibles et utilisables avec un minimum d'apprentissage. Il faut néanmoins insister sur la nécessité de comprendre les méthodes, afin d'en choisir une adéquatement, et de pouvoir juger les différences de résultats qui en découlent. De plus, certains de ces programmes ne couvrent qu'une partie de la problématique, et il ne faut en aucun cas utiliser des paramètres par défaut, si l'on n'a pas réalisé les étapes précédentes d'analyse.

### 5.1. Statistiques spatiales

Le module S+ $\circ$  spatial statistics permet d'effectuer les investigations de statistiques spatiales au niveau des méthodes déterministes, de la variographie, et de l'estimation.

La répartition spatiale des prélèvements est caractérisée de manière quantitative par:

- la statistique des plus proches voisins point à point: pourcentage de points par rapport à l'ensemble, en fonction de la distance de voisinage,
- la statistique des plus proches voisins entre points de prélèvement et noeuds d'une grille,
- le nombre de points de mesure par unité de surface, ou intensité, à représenter sous la forme d'une carte raster,
- le nombre de mesures à une distance inférieure à  $d$  par rapport à une mesure spécifiée, pondéré par l'intensité (fonction K de Ripley),
- les indices de Moran et de Geary pour mesurer le niveau de significativité de l'autocorrélation.

En ce qui concerne la démarche géostatistique, les quatre démarches de base successives sont:

- l'estimation de variogrammes et corrélogrammes expérimentaux, avec compléments des nuages variographiques,
- l'ajustement des variogrammes par des modèles théoriques,
- l'estimation par calculs BLUE (*Best Linear Unbiased Estimator*) de krigage, avec ou sans dérive; la présence éventuelle d'une dérive étant mise en évidence par median polishing par exemple,
- les simulations, afin de reconstituer la variabilité des observations parmi l'ensemble des résultats interpolés.

Certains types de données ne se prêtent pas à interpolation et doivent être représentés avec une valeur constante sur une zone définie. Parmi ces *lattice data* citons bien entendu les indications de la carte géologique, où il n'est pas question d'interpoler entre granite et calcaire, mais d'étendre la valeur « granite » et la valeur « calcaire » à la zone dont le périmètre a été défini.

Les calculs d'autocorrélation et de régression spatiale sont adéquatement utilisés pour la valorisation de ce type de données.

### 5.2. Représentation cartographique

SURFER<sup>©</sup> permet d'établir des cartes d'interpolation, et de les représenter soit en cartes planes d'isolignes, soit en perspective 3-D. Les cartes d'interpolation sur une maille seront représentées telles quelles, ou sur un fond de carte digitalisé représentant des limites administratives ou des caractéristiques topographiques. On prendra garde de calculer préalablement les variogrammes dont les paramètres (modèles, effet de pépite, zones d'influence, paliers, dérive, anisotropie) sont à indiquer pour effectuer l'interpolation sur une grille par krigage.

Les différentes méthodes déterministes d'interpolation à disposition: *minimum curvature*, *nearest neighbor*, *polynomial regression*, *radial basis functions* (*quadric*, *cubic spline*, etc.), *Shepard's method* (inverse de distance améliorée par moindre carré) et *triangulation* fourniront des cartes différentes de celles établies par krigage, où l'on aura pris soin de fournir une indication correcte du modèle de variogramme.

### 5.3. Corrélations spatiales, estimations et simulations

Le logiciel GSLIB (DEUTSCH and JOURNEL 1992) est utilisable avec l'élaboration d'un court fichier de paramètres, et comporte une grande panoplie de méthodes géostatistiques. Les *outputs* peuvent être visualisés sur des imprimantes habituelles, par exemple à l'aide du programme de *post-processing Upfile*.

Les nombreux programmes GSLIB couvrent les aspects de:

- variographie,
- krigeages: *estimator BLUE (Best Linear Unbiased Estimator)*, avec ou sans dérive, krigage factoriel, cokrigage, krigage d'indicatrices, et diverses variantes,
- simulations: simulations gaussiennes séquentielles, simulation par indicatrices, recuit (*annealing*).

## 6. CONCLUSION

La finalité de cet article est de souligner l'intérêt des statistiques univariées pour l'identification d'anomalies géochimiques, du fait notamment de leur mise en œuvre très rapide à l'aide de programmes informatiques (SPSS<sup>©</sup>, STATLAB<sup>©</sup>, S+<sup>©</sup>, GSLIB et GEOEAS). On peut, grâce aux outils statistiques, dégager des tendances avant d'élaborer des hypothèses sur les processus naturels (fig. 24). Puis une fois ce travail effectué, il faut s'interroger sur la signifi-

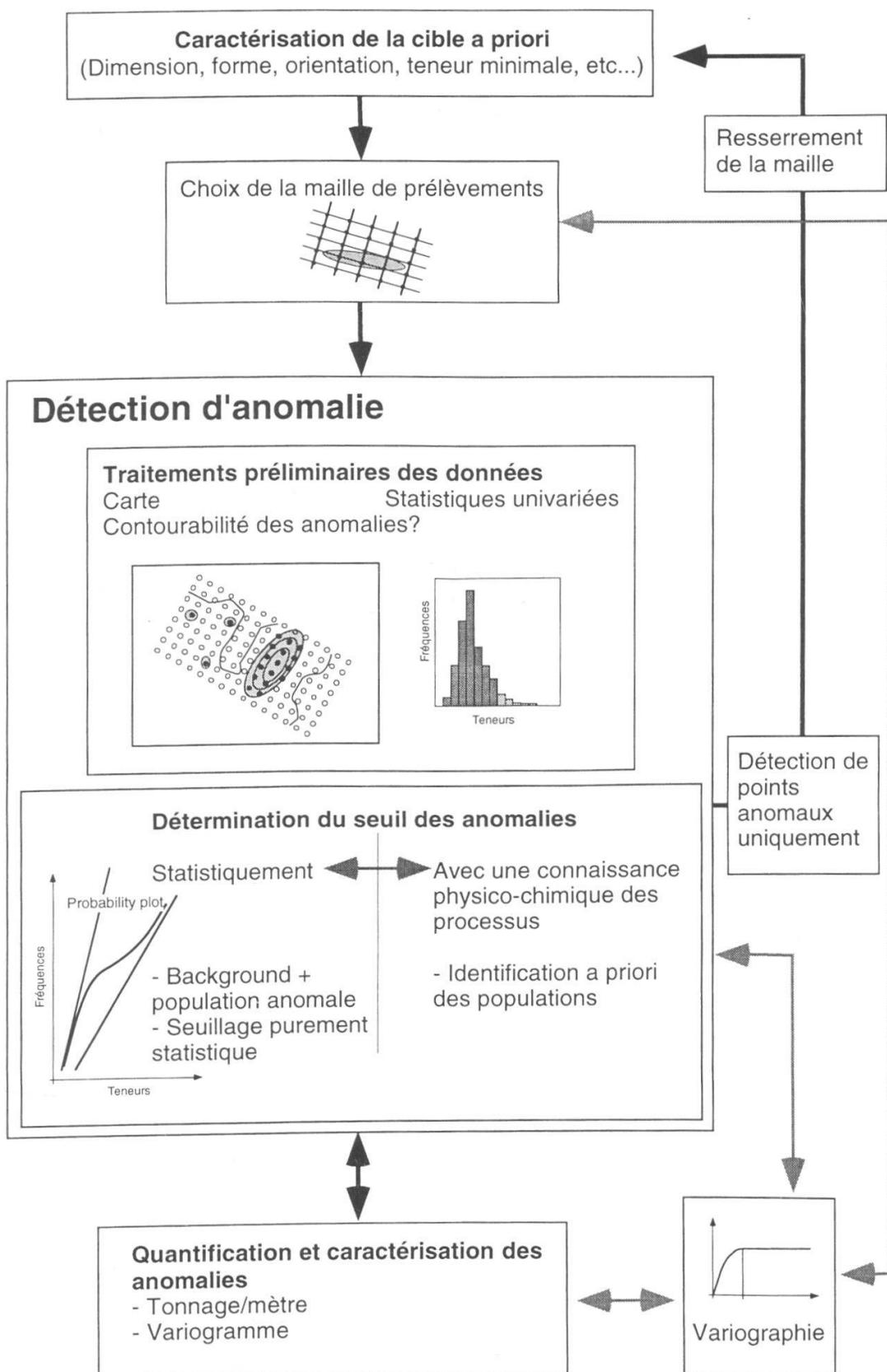


Figure 24.–Organigramme simplifié de la démarche menant à l'identification d'anomalies géochimiques essentiellement à l'aide d'outils statistiques univariés.

cation des populations statistiques en termes de processus physico-chimiques, une telle démarche pouvant induire des modifications dans la définition des populations. La structure induite par une analyse statistique constitue donc une base pour poursuivre d'autres investigations. Par ailleurs, dans les conditions actuelles où les laboratoires automatisés et les bases de données sur *Internet* ou *CD-Rom* nous submergent de données, il faut saisir rapidement les faits importants dans ce foisonnement d'informations: là aussi, les indicateurs univariés résument l'information sont utiles. Certaines techniques multivariées, non abordées ici, telles que l'analyse factorielle, s'avéreront utiles. La vario-graphie est un outil important, elle permet en effet d'évaluer comment et jusqu'à quelle distance les prélèvements sont liés entre eux (zone d'influence), point capital lors de la définition de la maille de prélèvement.

Il est clair que cette panoplie d'outils rapidement et assez facilement applicables est devenu un atout majeur pour analyser efficacement et utilement les données géochimiques, quelle que soit leur origine. A l'heure où beaucoup d'organismes, publics ou privés, voient leurs ressources se restreindre ou être soumises à un contrôle beaucoup plus serré, il est évident que de telles méthodes peuvent représenter une économie substantielle, en temps et en argent, en permettant une sélection mieux assurée des meilleures anomalies, et ceci dans différents contextes de recherches.

#### REMERCIEMENTS

Nous remercions M. Djeddou et l'Institut de géophysique ainsi que M. Célardin et le laboratoire cantonal d'agronomie de Jussy (GE) de nous avoir aimablement mis à disposition quantité de données, ainsi que Messieurs F. Bavaux et R. Della Piazza pour leur révision du manuscrit et leurs suggestions et corrections constructives. Nous sommes reconnaissants envers Mme et M. Calmbach, Mme Magnenat, Messieurs Epard, Flamm, Kanevsky et Lokosha pour les nombreuses discussions et corrections. M. Jaboyedoff exprime sa gratitude envers le FNRS pour lui avoir permis, en complément de ses recherches principales, d'effectuer ce travail dans le cadre de l'enseignement à l'Université de Lausanne.

#### BIBLIOGRAPHIE

- AITCHISON J. and BROWN J.A.C., 1963. The lognormal distribution. Cambridge University Press. 176 p.
- ASRARUL HAQUE M. and SUBRAMANIAN V., 1982. Copper, lead, and zinc pollution of soil environment. *Critical Rev. in environ. Control.* 12: 13-68.
- ATTEIA O., DUBOIS J.-P. and WEBSTER R., 1994. Geostatistical analysis of soil contamination in the Swiss Jura. *Environ. Pollution* 86: 315-327.
- BARBIER J., 1989. L'évaluation quantitative des anomalies en sols résiduels: une aide pour le choix des prospects. *J. Geochem. Explor.* 32: 445-452.
- BAUCHAU C. and JABOYEDOFF M., 1990. L'enseignement de la prospection minière par simulation sur ordinateur PC. *Bull. Soc. vaud. Sc. nat.* 80.1: 49-62.
- BAUCHAU C., JABOYEDOFF M. and VANNIER M., 1993. «CLAIM»: A new personal computer assisted simulation model for teaching mineral exploration techniques. In: KIKHAM R.V., SINCLAIR W.D., THORPE R.I. AND DUKE J.M. (Eds.): Mineral Deposit Modeling: *Geol. Assoc. of Canada, Special Paper* 40: 685-691.
- BOLT G.H. and BRUGGENWERT M.G.M., 1976: Soil chemistry: A. Basic elements. Elsevier, Amsterdam. 281 p.

- BRUNO R. and RASPA G., 1994. La Pratica della Geostatistica lineare. Guerini Studio. Roma. 170 p.
- CÉLARDIN F., CHATENOIX L. et d'ERSU P., 1989. Métaux lourds dans les sols du bassin genevois: état de la question (II). *Arch. Sci. phys. nat. (Genève)* 42: 493-498.
- CÉLARDIN F. et CHATENOIX L., 1990. Répartition en profondeur des métaux lourds dans les sols genevois. *Arch. Sci. phys. nat. (Genève)* 43: 265-272.
- CÉLARDIN F., CHATENOIX L. et MEYLAN G., 1992. Réseau genevois d'observation du sol (GEOS). *Arch. Sci. phys. nat. (Genève)* 45: 171-187.
- DAVIS J.C., 1986. Statistics and data analysis in geology. John Wiley and Sons, New York. 645 p.
- DEUTSCH C.V. and JOURNEL A.G., 1992. GSLIB Geostatistical Software Library and User's guide. Oxford University Press, New York. 340 p.
- DJEDDOU A., 1991. Recherche des minéralisations d'or, de cuivre et de chromite dans la région du «coin» (Hoggar sud-oriental, Algérie) par méthodes géophysiques. Thèse Fac. Sc. Univ. Lausanne. 167 p.
- ELLIOT I.L. and FLETCHER W.K., 1975. Geochemical exploration 1974. Elsevier, Amsterdam. 720 p.
- ENGLUND E. and SPARKS A., 1991: Geo-EAS 1.2.1 User's Guide. US-EPA Report 600/8-91/008. Environ. Protect. Agency-Environ. Monitoring Syst. Lab. Las Vegas, Nevada.
- EPSTEIN B., 1947. The mathematical description of certain breakage mechanisms leading to logarithmico-normal distribution. *J. Franklin Inst.* 244: 471-477.
- FLAMM Ch., 1994. Pollutions de sols: apport des statistiques spatiales et multivariées. Travail de diplôme. Inst. Min. Pétr. Univ. de Lausanne. Inédit. 200 p.
- FLAMM C., KANEVSKY M. and SAVELIEVA E. 1994. Non-regular lag variography and multi-method mapping to determination of origin of heavy metals: case study on Geneva heavy metal survey, Switzerland. Proc. of Int. Amer. Math. Geol. Congr., Mont-Trembant (Quebec).
- FLEISCHHAUER H.L. and KORTE N., 1990. Formulation of cleanup standards for trace elements with probability plots. *Environ. Management* 14: 95-105.
- ISAAKS, E.H. and SRIVASTAVA R. M., 1989. An introduction to applied geostatistics. Oxford University Press, New York. 561 p.
- IYENGAR S.S., MARTENS D.C. and MILLER W.P., 1981. Distribution and plant availability of soil zinc fraction. *Soil Sci. Soc. Am. J.* 45: 735-739.
- JABOYEDOFF M. and MAIGNAN M., in prep. Distribution due to a diffusive anomaly.
- JOURNEL G., 1989: Fundamentals of geostatistics in five lessons. Short course in Geol. 8, Amer. Geophys. Union, Washington. 40 p.
- KANEVSKI M., ARUTYUNIAN R., BLSHOV L., DEMYANOV V., LINGE I., SALIEVA, E. SHERSHAKOV V., HASS T. and MAIGNAN M., in press. Geostatistical Portrayal of the Chernobyl accident. Int. Geostatistical Congr., Wollongong 1996.
- KANEVSKI M., SALIEVA E. and MAIGNAN M., in press. NNRK Neural Network Residual Kriging. *Jap. J. of Geoinformatics*.
- LANDRY J.-C. and CELARDIN F., 1988. Métaux lourds dans les sols du bassin genevois: état de la question. *Arch. Sci. phys.nat. (Genève)* 41: 199-223.
- LEPELTIER C., 1969. A simplified statistical treatment of geochemical data by graphical representation. *Econ. Geol.* 64: 538-550.
- LOVEJOY S., SCHERTZER D. and LADOY P., 1986. Fractal characterization of inhomogeneous geophysical measuring networks. *Nature* 319: 43-44
- LOKOSHA O., 1991. Approche et démonstrations des méthodes géostatistiques d'estimation linéaire. Travail de diplôme Univ. Lausanne. Inédit. 61 p.
- MAIGNAN M. and MAIGNAN M.-F., in press. Heavy metals in soils of Western Switzerland: geogenous origin or pollution? Proc. of 30th World geol. Congr. Beijing.
- MARSAL D., 1987. Statistics for geoscientists. Pergamon Press, Oxford. 176 p.

- MATHÉRON G., 1962. Traité de géostatistique appliquée. Tome 1. Mémoire du BRGM N°14. Ed. Technip, Paris. 333 p.
- OSOL. Ordonnance sur les polluants du sol (Osol), du 9 juin 1986 (RS 814.12). Office fédéral de l'environnement. Berne, Suisse.
- PACÉ P. et CLUZEL R., 1968: Statistique et probabilité. Aide-mémoire. Delagrave. Paris. 156 p.
- PANNATIER Y., 1996. VarioWin: software for spatial data analysis in 2D. Springer-Verlag, New York. 91 p.
- POURCHET M., BATRAYA S.K., MAIGNAN M., PINGLOT J.F., ARISTARAIN A.J., FURADADA G., KOTLYAKOV V.M., MOSLEY-THOMSON E., PREISS N. and YOUNG N.W., in press. Distribution and Fallout of Cs137 and other radionuclides over Antarctica. *J. of Glaciology*.
- RUEGG A., 1989. Probabilités et statistique. Presses polytechniques romandes, Lausanne. 153 p.
- S+ ©, 1996. Spatial Stats User's manual. Mathsoft Inc. USA. 225 p.
- SAPORTA G., 1990. Probabilités, analyse des données et statistique. Ed. Technip, Paris. 493 p.
- SIEGEL F.R., 1974. Applied geochemistry. John Wiley and Sons, New-York. 353 p.
- SINCLAIR A.J., 1989. Statistical interpretation of soil geochemical data. In: ROBERTSON J.C.(Ed.). Exploration geochemistry: design and interpretation of soil surveys. *Rev. in Econ. Geol.* 3: 97-115.
- SINCLAIR A.J., 1991. A fundamental approach to threshold estimation in exploration geochemistry: probability plots revisited. *J. Geochem. Explor.* 41:1-22.
- SPSS©, 1994. Statistical Package for the Social Sciences. User's manual.
- STATLAB©, 1995. Manuel de l'utilisateur. Version 2.1. SLP, France. 400 p.
- SURFER©, 1995. Surfer for Windows, user's guide V6. Contouring and 3D surface mapping. Golden Software Inc. Colorado. 400 p.
- TROEH R. and THOMPSON L-M., 1993. Soils and soil fertility. 6<sup>th</sup> ed. Oxford University Press, New York. 462 p.
- VANNIER M. and WOODLI R., 1992. Teaching and testing strategy in mineral exploration by simulation techniques on PCs. In: MERRIAM D. AND KURZL H. (Eds.): Use of microcomputers in geology (p. 9-23). Plenum Press, New-York. 285 p.
- VENTSEL H., 1987. Théorie des probabilités. Ed. MIR, Moscou. 584 p.
- WEBSTER R. and OLIVER M.A., 1990. Statistical methods in soil and land resource survey. Oxford University Press, Oxford. 316 p.
- WEBSTER R., ATTEIA O. and DUBOIS J.-P., 1994. Coregionalization of trace metals in the soil in the Swiss Jura. *Eur. J. Soil Sci.* 45: 205-218.
- WEDEPOHL K.H., 1969. Handbook of geochemistry. Springer, Berlin. 6 vol.

*Manuscrit reçu le 2 mai 1996*



© Société vaudoise des Sciences naturelles, CH-1005 Lausanne.  
Droits de reproduction réservés.

---

Rédaction:

Pierre Gex, Institut de Géophysique, bâtiment du collège propédeutique, CH-1015 Lausanne

Composition: Société vaudoise des Sciences naturelles, 1005 Lausanne.

Imprimerie: IRL S.A CH-1020 Renens-Lausanne.