M. Sala

B. R. G. M.

DEPARTEMENT DES RESERVES

REMARQUES SUR LA NOTE S E M A intitulée "STRATEGIE DE LA PROSPECTION ALLUVIONNAIRE"

G. MATHERON

Mai 1963

REMARQUES SUR LA NOTE S E M A

intitulée " STRATEGIE DE LA PROSPECTION ALLUVIONNAIRE "

1/- <u>Fondements logiques.</u>

Les développements présentés par la SEMA ne prennent un sens cohérent que si on les rattache, comme à leur fondement logique, à la théorie des processus stochastiques. Cette voie, implicitement empruntée dans la NOTE SEMA, n'est pas mauvaise a priori. On sait que, dans le cas d'un processus stochastique stationnaire, elle conduit à la théorie géostatistique des schémas intrinsèques. C'est même peut être la manière la plus commode d'introduire la théorie intrinsèque, (cf. Note géostatistique n°47). Mais on sait aussi qu'en dehors du cas stationnaire cette formulation probabiliste mène à une impasse, par suite de l'impossibilité de l'inférence statistique.

Précisons. La variable régionalisée $f(M)$, ici égale à $+1$ ou à 0 selon que le point M est classé positif ou non, est définie en tant que processus stochastique par la donnée des lois de probabilités simultanées $F(x_1 .. x_n ; M_1 .. M_n)$ des valeurs $x_i = f(M_i)$ prises par la variable en n points quelconques M_i. Si le processus est stationnaire, une telle loi est invariante par une translation d'ensemble des n points M_i : le phénomène se répète lui-même, indéfiniment, dans l'espace, et l'inférence statistique est possible. Mais, en dehors du cas stationnaire, les lois F dépendent effectivement de la position des points M_i. Tout ensemble de n valeurs expérimentales $f(M_i)$ doit être considéré comme le résultat d'un tirage au sort unique de la variable aléatoire à n composantes $(x_1 .. x_n)$ selon la loi F. Comme on ne peut jamais remonter d'un tirage unique à une loi de probabilité, aucune inférence statistique n'est fondée. La seule réalité possédant un sens physique est l'ensemble des valeurs $f(M_i)$ effectivement prises par la variable au point M_i, c'est-à-dire le phénomène minéralisé lui-même, tel qu'il existe dans la nature. Considérer ces valeurs expérimentales comme une réalisation particulière d'un certain processus stochastique est une hypothèse à la fois arbitraire (puisqu'on ne peut pas remonter de ces valeurs au processus lui-même) et inutile (puisque la théorie non probabiliste des représentations transitives permet d'étudier directement le phénomène lui-même).

Dans le cas présent, où $f(M)$ ne prend que les deux valeurs 0 et 1, la probabilité a priori $P(M)$ pour qu'un point M soit minéralisé doit dépendre effectivement de M dans le cas non stationnaire. Mais, pour inférer la valeur $P(M_i)$ en un point déterminé M_i, on ne dispose que d'une épreuve unique, ayant donné 0 ou 1 selon que M_i était stérile ou minéralisé. Aucune autre épreuve n'est possible dans le cas non stationnaire, et, par conséquent, aucune estimation de $P(M_i)$ - à dire vrai, la notion même de probabilité $P(M_i)$ n'a pas grand sens, à moins de dire $P(M_i) = 1$, si M_i est minéralisée, et $P(M_i) = 0$ si M_i est stérile, point de vue purement formel qui revient à poser $P(M) = f(M)$.

2/- L'HYPOTHESE STATIONNAIRE.

Ce n'est que dans le cas stationnaire, où la probabilité a priori $P(M)$ devient une constante indépendante du point M, que l'inférence statistique a un sens et que le langage probabiliste est utilisable. Il semble bien qu'implicitement la SEMA ait admis l'hypothèse stationnaire : il eut été plus simple, une fois cette hypothèse admise, d'utiliser directement la théorie des schémas géostatistiques intrinsèques, plutôt que d'élaborer une méthode nouvelle, peu cohérente logiquement et mal adaptée au problème étudié.

Malheureusement, cette hypothèse n'est certainement pas admissible. En fait, la notion même d'anomalie implique une hétérogénéité dans l'espace, et contredit le caractère stationnaire. Un processus stationnaire présente des fluctuations stochastiques, mais aucune " anomalie ". Il est bien clair que la probabilité pour qu'un point M soit minéralisé est plus élevée en zone anormale qu'en zone stérile, de sorte que $P(M)$ dépend effectivement de M. Mais comme alors on doit renoncer à donner un sens objectif à $P(M)$, autre que $P(M) = f(M)$, il est illusoire de recourir à un langage probabiliste. La méthode à utiliser est, manifestement, celle des représentations transitives.

On objectera qu'il est peut être possible de décrire ce qui se passe au voisinage d'un point M donné à l'aide d'un processus stationnaire, valable localement, considéré comme tangent en M au processus réel. Cette possibilité n'existe que dans la mesure où les caractéristiques (valeur probable et fonction d'autocorrélation) du processus réel ne se déforment, dans l'espace, que d'une manière lente et régulière. On retombe alors sur la théorie intrinsèque, considérée comme une approximation valable localement. En fait, en ce qui concerne le problème actuel, cette approximation n'est pas utilisable. Le phénomène décisif est ici la transition entre une anomalie (considérée comme un domaine où $f(M) = 1$ presque partout, ou, plus simplement, a une valeur moyenne voisine de 1) et le stérile environnant (ou $f(M) = 0$ ou a une valeur moyenne faible). Dans cette zone de transition, par définition, les caractéristiques locales du phénomène se modifient très rapidement, et le processus stationnaire tangent en un point ne représente qu'un domaine très petit vis-à-vis de la maille kilométrique : aucune inférence " locale " n'est possible.

Dans ces conditions, on se demande ce que représentent les valeurs numériques calculées par la S E M A : en vérité, pas grand chose de plus que le choix arbitraire de la méthode de calcul qui a permis de les obtenir.

3/- L'HYPOTHESE MARKOVIENNE

En plus de l'hypothèse stationnaire, qui eut dû logiquement conduire à la mise en oeuvre des méthodes géostatistiques intrinsèques, la SEMA a formulé une hypothèse markovienne (ce qui n'est pas absurde), et l'a utilisée de manière contradictoire. A une dimension, un processus - représentant, par exemple, l'évolution d'un phénomène dans le temps - est dit markovien si l'avenir ne dépend que du présent, et non du passé, lorsque le présent est connu. Le présent fait écran, et arrête l'influence du passé sur l'avenir. A deux dimensions, la définition est déjà moins immédiate. On peut dire qu'un processus à deux dimensions est markovien, si, lorsque l'on connaît les valeurs prises par la variable sur un contour fermé quelconque, il y a indépendance entre le domaine intérieur et le domaine extérieur. De tels processus markoviens existent effectivement. La théorie géostatistique du krigeage continu en schéma de wijsien isotrope à deux dimensions en constitue sans doute l'exemple le plus simple possible.

Toutefois ce n'est pas en ce sens rigoureux que la SEMA formule son hypothèse markovienne. Elle admet que, le long d'une rivière, (donc à une seule dimension) les prélèvements ne dépendent que de leurs voisins immédiats. L'hypothèse est plausible, du reste, en vertu du caractère linéaire des covariogrammes transitifs construits par M. DEREC pour représenter, plus efficacement, le phénomène étudié. Mais ce caractère markovien sur certaines courbes à une dimension est incompatible avec le caractère markovien dans l'espace, et ne peut même pas être vrai pour n'importe quelle courbe. (par trois points M P Q, on peut faire passer des courbes pour lesquelles P est entre M et Q, et d'autres pour lesquelles M est entre P et Q, ou Q entre M et P : le loi f(xyz) devrait être à la fois de la forme :

$$f(xyz) = f_1(xy)\, f_2(yz) = f_3(yz)\, f_4(zx) = f_5(zy)\, f_6(zx)$$

ce qui n'est possible que s'il y a indépendance stochastique, cas sans intérêt).

Plus concrètement, on peut admettre que le caractère markovien à 1 dimension soit vrai sur des droites. Mais alors, pour trois points M P Q non alignés, M et Q ne peuvent pas être indépendants à P fixé. C'est cependant sous cette forme contradictoire que la SEMA utilise son hypothèse markovienne.

Nous lisons page 8 : " Les fréquences de transition dépendent de la distance qui sépare les points voisins. Elles peuvent être estimées pour les multiples entiers de la maille. Mais il n'est nécessaire de le faire que pour une seule distance, le calcul permettant d'en déduire les probabilités pour tout multiple entier ".

Cette façon de faire est évidemment bien commode, en ce sens qu'elle ne risque pas de conduire à des valeurs que l'expérience démentirait. Malheureusement, elle est fondée sur un modè : de calcul non seulement injustifié, mais mathématiquement incohérent.

On comprend d'autant moins la raison de cette hypothèse markovienne, utilisée de manière incohérente, qu'une fois admise son hypothèse stationnaire, il ne restait plus à la SEMA qu'à appliquer la théorie géostatistique intrinsèque.

4/ <u>CONCLUSION</u>

La construction de la SEMA apparaît bien fragile, et les " résultats " auxquels elle conduit comme quelque peu arbitraires. Sans parler de la manière mathématiquement incohérente dont elle utilise l'hypothèse markovienne, elle admet, comme fondement logique, le caractère stationnaire du processus minéralisateur, caractère qui n'est pas vérifié, même localement. A cet égard, les résultats obtenus par M. DEREC sont décisifs. Il a montré expérimentalement qu'une maille " intelligente "(où les prélèvements sont implantés par des géologues ne connaissant pas la région, et disposant seulement d'une carte géologique comportant le réseau hydrographique) donnait des résultats incomparablement meilleurs qu'une maille systématique ou aléatoire, pour un même nombre de prélèvements. Il en déduit l'existence de points privilégiés, particulièrement représentatifs, de par leur position sur le réseau hydrographique, de l'ensemble d'un bassin. De tels points sont en corrélation particulièrement bonne avec leur voisinage (le bassin qu'ils représentent). La fonction d'autocorrélation n'est pas invariante par translation, mais se déforme, sans doute très rapidement, au voisinage des pôles préférentiels que constituent ces points privilégiés. Il n'est pas réaliste de songer à construire un modèle mathématique tenant compte de cette caractéristique majeure du phénomène, précisément parce qu'on est dans le cas où l'inférence statistique n'est pas fondée. Seul un géologue expérimenté, capable de se représenter d'une manière qualitative et synthétique, les caractères structuraux de la géologie et du réseau hydrographique, peut indiquer approximativement, dans chaque cas, la maille intelligente la meilleure possible, c'est-à-dire celle qui s'adapte au mieux à la réalité d'un phénomène naturel unique.

G. MATHERON.