

ESTIMER ET CHOISIR

- Essai sur la Pratique des Probabilités -

par

G. MATHERON

ÉCOLE NATIONALE SUPÉRIEURE DES MINES DE PARIS

E S T I M E R E T C H O I S I R

- Essai sur la Pratique des Probabilités -

PREMIERE PARTIE

A LA RECHERCHE DE L'OBJECTIVITE

G. MATHERON

—

1978

PREMIERE PARTIE

A la Recherche de l'Objectivité

"Quiconque désire l'absolu, doit prendre la subjectivité -
- l'égoцентриté - par surcroît, et quiconque souhaite l'ob-
jectivité ne peut éviter le problème du relativisme"

H. Weyl

T A B L E D E S M A T I E R E S

P R E M I E R E P A R T I E

<u>I N T R O D U C T I O N .</u>	1
<u>C H A P I T R E I</u> - LE HASARD CHEZ J. MONOD ou : Comment on franchit les limites de l'objectivité.	9
Le Hasard, concept métaphysique.	10
La parabole du Plombier et du Médecin.	12
Les chaînes polypeptidiques.	15
Un paradoxe apparent	19
Le seuil de réalisme ou d'objectivité.	23
<u>C H A P I T R E I I</u> - POURQUOI NOUS NE SOMMES PAS D'ACCORD AVEC LES ETRUSQUES ou : De l'objectivité des modèles probabilistes.	26
Le Problème.	27
Quelques illusions anthropomorphiques.	29
Le critère poppérien de l'objectivité.	31
Les concepts opératoires.	34
La subjectivité	35
Il n'y a pas de probabilité en soi, seulement des modèles probabilistes.	36
Les modèles probabilistes.	38
Le modèle de l'alternative répétée.	39
A la recherche d'un critère d'objectivité.	41
Reconstruction opératoire des concepts probabilistes.	44
L'hypothèse anticipatrice, et le risque d'erreur radicale.	48
Modèles panscopiques et modèles monoscopiques.	50
<u>C H A P I T R E I I</u> (Suite)	
Les critères externes, ou l'objectivité d'une méthodologie.	55
Critères d'objectivité interne, liés au caractère concret de l'espace Ω .	59

INTRODUCTION

Ce livre n'est pas désintéressé. Plus qu'une thèse, je défends une pratique, un ensemble de modèles, méthodes et "tours de main", souvent peu orthodoxes, élaborés¹ au fil des ans pour décrire, étudier, estimer les phénomènes les plus variés : il peut s'agir de cellules cancéreuses ou de forêts ; de la structure d'une roche ou de celle d'un alliage métallique ; d'un gisement de pétrole, ou d'un revêtement de route ; de pollution, ou de météorologie ; de cartographie sous-marine, ou de prospection géophysique etc.... Cette méthodologie et cette pratique tirent leur origine d'un problème tout à fait terre à terre : celui de l'estimation des gisements miniers. C'est ce domaine bien particulier qui a servi de banc d'essai, et a fourni les vérifications les plus nombreuses et les plus décisives.

Le problème des matières premières est, paraît-il, aujourd'hui au coeur des préoccupations de chacun de nous. A l'usage, pourtant, il apparaît vite qu'en dehors d'un très petit nombre de spécialistes de l'industrie minière, peu de personnes supportent d'écouter jusqu'au bout un exposé un peu sérieux sur la notion de réserves récupérables d'un gisement minier. Il est encore moins question de leur montrer les difficultés, multiples et réellement subtiles, qu'il faut élucider une à une, et surmonter effectivement, avant de pouvoir proposer une estimation simplement sensée de ces réserves. La discipline nouvelle qu'il a fallu forger à cette fin - la Géostatistique minière - et qui a maintenant fait ses preuves, surprendrait peut-être, par sa complexité, les statisticiens ou les probabilistes qui auraient le courage d'en prendre connaissance. Mais l'aspect technique de ces problèmes découragerait, je le sais, le lecteur de meilleure volonté. Je n'en parlerai, dans ce qui suit, que par allusions discrètes.

1 - Principalement au Centre de Géostatistique et de Morphologie Mathématique de l'Ecole des Mines, à Fontainebleau.

Quoi qu'il en soit, les phénomènes auxquels nous nous intéresserons présentent deux caractéristiques importantes : d'une part, ils se déploient dans l'espace physique, celui dans lequel nous vivons, et se manifestent, comme nous disons, sous une forme régionalisée. D'autre part, et le plus souvent, il s'agit d'objets ou de phénomènes uniques. Ayant traité et résolu, avec un succès réel, confirmé par la pratique, un grand nombre de cas particuliers, nous soupçonnons, il est vrai, que ces problèmes n'étaient peut-être pas aussi uniques qu'il nous semblait au départ. La méthodologie qui nous a permis de les résoudre, ayant fait ses preuves, relève sans aucun doute d'une certaine forme d'objectivité, que nous appellerons objectivité externe : c'est, tout simplement, la sanction de la pratique. Mais ces objets restent uniques, en tant qu'individus que l'on ne retrouvera jamais ailleurs identiques à eux-mêmes. Il n'existe pas deux gisements, ni deux forêts, ni deux étoiles identiques. Dans quelle mesure une estimation, ou un modèle probabiliste concernant ce gisement-ci, cette forêt-là (et non un gisement ou une forêt en général) possède-t-il une signification objective ? C'est à la recherche des critères de cette objectivité interne qu'est consacré cet essai.

Je ne peux donc, en aucune façon, éluder la fameuse question, si mal formulée : "la probabilité est-elle subjective ou objective ?". En fait, il n'y a pas, il ne saurait y avoir, de probabilité en soi. Il n'y a que des modèles probabilistes. Ou, si l'on préfère, l'"aléatoire" n'est en aucune façon une propriété, univoquement définie, ni même définissable, du phénomène lui-même. Mais, uniquement, une caractéristique du, ou des, modèles que nous choisissons pour le décrire, l'interpréter, et résoudre tel ou tel problème que nous nous posons à son sujet. Et, suivant la nature de ces problèmes, nous pouvons fort bien adopter, tour à tour, des modèles différents, attribuant, suivant le contexte, des probabilités différentes à un événement donné dont l'énoncé, apparemment, reste le même. Le seul problème réel est de savoir si un modèle donné, dans un contexte donné, possède, ou non, une signification objective et, le cas échéant, de faire le tri : je

veux dire, parmi les concepts, énoncés, paramètres etc... figurant dans le modèle, distinguer ceux qui possèdent dans (ce que nous appelons) la réalité une contrepartie objective, observable, mesurable, etc... et les autres : ces derniers, que j'appellerai conventionnels (plutôt que subjectifs) pourront jouer un rôle heuristique utile dans l'élaboration et la mise en oeuvre du modèle. Mais ils devront disparaître de la formulation ultime de nos conclusions et de nos résultats. Car, dans la mesure même où les problèmes que nous cherchons à résoudre sont des problèmes réels, les solutions que nous proposons doivent, elles aussi, sous peine de se révéler illusoire, respecter le principe de réalité.

Il ne s'agit pas, notons-le bien, d'adopter quelque position "conciliatrice" entre ces deux extrêmes que représentent les thèses "objectivistes" et "subjectivistes". Il faut trancher dans le vif. Des bayesiens, je retiens beaucoup : leur critique féroce des postulats implicites des objectivistes, postulats qui, dans bien des cas, ne peuvent être dits ni vrais ni faux, mais seulement vides de sens ; ou bien encore le rôle fondamental qu'ils attribuent, avec raison, à ce qu'ils appellent l'information "a priori", c'est-à-dire, à cette information non numérique, ou qualitative, de nature structurale, qui nous dit, en somme, tout simplement de quoi il s'agit : devant une même suite de 0 et de 1, nous ne choisissons pas le même modèle selon qu'il s'agit des résultats d'une partie de pile ou face, d'une alternance de jours de pluie et de beau temps, ou de l'écriture binaire d'un nombre transcendant. Mais je n'en tire nullement la conclusion que nos modèles probabilistes doivent être abandonnés, sans recours, à l'arbitraire de la subjectivité individuelle. Je propose, au contraire, d'abord de faire le tri, comme je l'ai dit, puis, surtout, de nous mettre à l'école des physiciens, et de procéder à la reconstruction opératoire des concepts de notre modèle, ou du moins de ceux que notre critique préalable aura reconnus aptes à véhiculer une information objective.

Cette position méthodologique se reflète dans la terminologie que j'adopte. Les statisticiens orthodoxes auraient dit :

estimer et prédire. Les bayesiens, ou subjectivistes : évaluer et prévoir. J'ai préféré : estimer et choisir. Ceci mérite un commentaire. J'utilise le verbe "estimer" chaque fois qu'il s'agit d'évaluer une grandeur que nous ne connaissons pas, mais qui n'en existe pas moins indépendamment de nous et de l'état de notre information : l'estimation, donc, se réfère à une grandeur physique, objective, et peu importe que cette grandeur soit associée, dans notre modèle, à une variable aléatoire ou à un "paramètre". De fait, une même réalité - disons, la teneur moyenne d'un gisement minier, sera, suivant le cas, traitée dans le modèle, soit comme un paramètre (objectif), soit comme une variable aléatoire. Mais, dans un cas comme dans l'autre, je parlerai de l'estimation de cette teneur moyenne (et non de sa prédiction, ou de sa prévision). J'utiliserai le verbe "choisir" tantôt dans un sens neutre (ne préjugant pas du sens, objectif ou conventionnel, qu'il conviendra par la suite d'attribuer au paramètre ainsi choisi), tantôt aussi dans un sens limitatif, lorsqu'il s'agira, pour des raisons de commodités, d'attribuer une valeur numérique à un paramètre conventionnel (non objectif), valeur pouvant être choisie de manière plus ou moins arbitraire sans compromettre l'objectivité de la démarche prise dans son ensemble. En somme, on estime des grandeurs (objectives), on choisit des méthodes, des modèles, ou des paramètres conventionnels, et on convient de critères. Pour les bayesiens, il n'y a pas d'estimation, mais seulement des "choix". Ils évaluent des probabilités (subjectives) et en tirent des prévisions (également subjectives) concernant les événements ou les grandeurs inconnues. Les orthodoxes, ou objectivistes, à l'inverse, réservent le terme "estimation" au choix des paramètres de leurs modèles, sans se demander au préalable si ces paramètres possèdent ou non, une contrepartie objective dans la réalité. Par contre, ils prédisent les valeurs (futurs ou inconnues) de variables aléatoires, même si celles-ci représentent des réalités physiques effectives. Leur terminologie véhicule une sorte de platonisme clandestin : le modèle et ses paramètres nous viennent du ciel des idées, et tirent de cette noble origine une réalité substantielle, un statut ontologique bien supérieur à celui que pourraient revendiquer leurs

pâles et terrestres imitations (les variables du modèle, c'est-à-dire les grandeurs physiques de la réalité). C'est ce présupposé substantialiste que dénoncent, avec raison, les bayesiens. Pour nous, c'est à la suite seulement d'un examen critique effectué dans chaque cas particulier, que nous accorderons ce statut objectif à tel ou tel paramètre et, qu'en conséquence, nous parlerons de son estimation.

Ce qui précède dicte le plan de cet essai. Dans la première partie, j'examine les critères, et les limitations, de l'objectivité externe, ou méthodologique, qui est tout simplement la sanction (à la longue) de la pratique, et l'objectivité interne, qui se fonde, en définitive, sur les caractéristiques physiques, concrètes, de l'objet individuel. A ce stade, les exemples que je donne sont empruntés au fond culturel commun : la pluie et le beau temps, la partie de pile ou face, les molécules de protéines. A partir de la seconde partie, il s'agit plus spécifiquement des phénomènes régionalisés dans l'espace, et des modèles "topo-probabilistes" qui les représentent. Le but est alors de dégager des critères d'objectivité interne. Comment choisir un modèle, comment faire le tri entre ce qui est objectif et ce qui ne l'est pas, comment localiser dans cette hiérarchie le point précis où nous introduisons, in praxi, cette hypothèse anticipatrice, lourde d'information objective non contenue dans les données, qui nous permet de résoudre les problèmes que nous nous posons, mais sans nous garantir jamais complètement contre un risque d'erreur radicale ? Telles sont les questions auxquelles j'essaierai de répondre. Dans la troisième partie, je tenterai, mais sur des cas particuliers seulement, de reconstruire en termes opératoires, à la manière des physiciens, quelques-uns des concepts qui sous-tendent nos modèles. Cette reconstruction nous permettra de préciser la nature de la fameuse hypothèse anticipatrice : il s'agira, simplement, d'une approximation banale, mais adoptée, à titre d'hypothèse, avant que sa validité ait pu être réellement contrôlée. D'où le risque, toujours présent, d'erreur radicale. De la même façon, l'"inférence statistique" se dépouillera d'une grande partie de son mystère : pour autant qu'elle vise un paramètre conventionnel, il s'agit en somme d'un faux problème. Pour

autant, par contre, qu'elle vise une réalité physique, il s'agit simplement du calcul numérique approché d'une intégrale d'espace.

La grande victime de cet examen sera peut-être l'espérance conditionnelle, notion mathématique très profonde, qui se révélera pourtant à peu près inapte, dans notre contexte, à recevoir le statut de concept physique. A titre de consolation je proposerai un substitut plus simple, celui d'estimateur disjonctif, et je dirai quelques mots aussi de l'usage, possible et légitime moyennant certaines conditions, de modèles purement heuristiques.

Bien que nos modèles probabilistes permettent d'aborder des problèmes très variés, quantitatifs ou qualitatifs¹, c'est surtout, pour des raisons de simplicité, sur l'estimation que je centrerai l'exposé. Les difficultés que soulève le problème de l'estimation ne sont pas de nature mathématique, ou du moins pas principalement. On les rencontre aux deux niveaux extrêmes : à l'étape initiale de la conceptualisation, c'est-à-dire du choix du modèle, et à l'étape terminale de la mise en oeuvre pratique.

A l'étape initiale, nous voyons surgir des questions inévitables concernant le sens, l'objectivité et la validité de toutes les démarches ultérieures. Etant donné un phénomène unique, parfaitement déterminé même s'il reste largement inconnu, est-il possible, et cela présente-t-il même un sens, de le conceptualiser sous la forme d'un modèle probabiliste, c'est-à-dire, en somme, de se représenter ce phénomène comme l'un, quelconque, tiré au sort dans une famille de phénomènes virtuels considérés comme possibles ? Le choix de ce modèle a-t-il une signification objective, c'est-à-dire est-il susceptible de se prêter à telle ou telle forme de contrôle expérimental, ou doit-il être abandonné sans recours à l'arbitraire de la subjectivité individuelle ? Enfin, à supposer que ces contrôles expérimentaux soient possibles, conduisent-ils à des résultats positifs permettant de con-

1 - Voir, par exemple, ce que nous appelons la Morphologie Mathématique.

clure sans ambiguïté à la validité objective au moins partielle de l'ensemble de la démarche ?

A l'étape terminale, c'est-à-dire au niveau des applications pratiques, on voit surgir des questions analogues, mais sous forme particularisée. A partir de l'information, toujours fragmentaire, souvent de qualité médiocre, dont nous disposons pour estimer, par exemple, un gisement minier, pouvons-nous spécifier un modèle probabiliste adéquat à ces données (donc, en particulier, évaluer de manière sensée les valeurs numériques des paramètres dont dépend ce modèle) et, à supposer que cela soit possible, dans quelle mesure pouvons-nous ensuite extrapoler du connu à l'inconnu, c'est-à-dire déduire du modèle, étalonné sur les seules données connues, des conclusions valables pour les parties non informées du gisement ? A cette question décisive (très kantienne) "Comment l'estimation est-elle possible ?" il semble bien qu'on ne puisse apporter une réponse qu'en acceptant, sous une forme ou une autre, une hypothèse d'homogénéité statistique au moins locale, dont la stationnarité classique constitue une forme extrême, exagérément sévère : le phénomène devrait, en somme, se comporter, là où on ne le connaît pas, d'une manière raisonnablement analogue à ce que l'on peut observer sur les données disponibles au voisinage. Mais, si cette hypothèse est indispensable pour fonder la possibilité de l'estimation, il n'en résulte pas automatiquement qu'elle soit partout et toujours vérifiée. Il est facile de donner des contre-exemples. Sur la figure ci-après, où les points A, B, C...J représentent

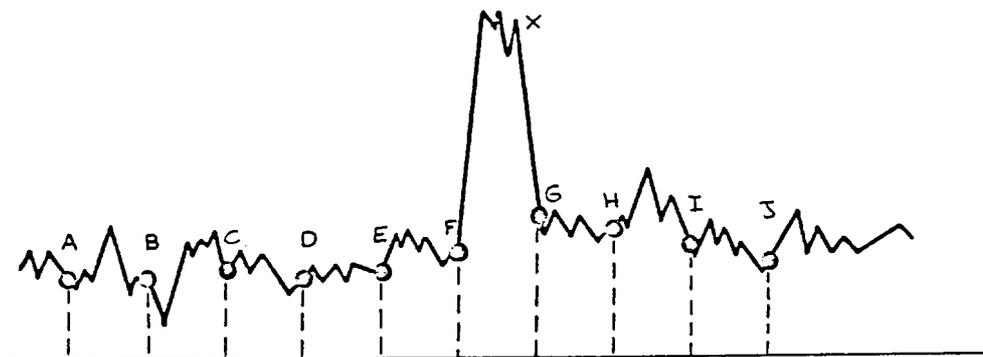


Figure 1

les données disponibles, on peut voir un cas de ce genre : rien ne permet de prévoir l'anomalie X et l'estimation n'est pas possible (quelle que soit la méthode). L'erreur radicale est, ici, inévitable.

Un mot encore, pour terminer cette introduction un peu longue. On ne trouvera dans ce qui suit aucune "conception du monde", explicite ou implicite, mais seulement des conseils méthodologiques à l'intention des praticiens des probabilités. on peut, si l'on veut, distinguer le moteur et le code. L'élément moteur, ou "dialectique", est toujours implicitement présent. C'est lui qui nous incite à progresser, à entreprendre toujours de nouvelles démarches, sans jamais rester en repos. C'est lui aussi qui nous permet de "comprendre" tout le développement antérieur dans une vaste synthèse rétrospective - mais après coup seulement, à la manière de l'"oiseau de Minerve", qui se lève après le coucher du soleil. Il se pourrait bien d'ailleurs qu'il y ait là quelque illusion anthropomorphique. Ce point de vue, de toute façon, convient à un historien ou à un philosophe des sciences, qui cherche à reconstituer les étapes parcourues par l'esprit humain dans son effort pour élaborer et reconstruire le réel, mais pas tellement à un praticien : celui-ci, au cours de son travail, a plutôt besoin d'un code, ou disons d'un fil à plomb pour construire des murs bien droits. Cela ne veut pas dire qu'il adhère à une quelconque philosophie du fil à plomb, mais seulement qu'il aime son métier.

LE HASARD CHEZ J. MONOD

ou

Comment on franchit les limites de l'objectivité

Dans un ouvrage sérieux, écrit par un auteur compétent, sur un sujet scientifique, une expression telle que "ce phénomène est dû au hasard", constitue en principe, simplement, une manière rapide de parler, signifiant en réalité "tout se passe comme si ce phénomène était dû au hasard", ou, dans un langage plus précis : "pour décrire, ou interpréter, ou formaliser ce phénomène, jusqu'à présent, seuls les modèles probabilistes ont donné de bons résultats". Il s'agit alors d'un simple constat de fait. Eventuellement, l'auteur peut ajouter : "et il ne me semble pas que cette situation soit de nature à se modifier dans un avenir prévisible". Il y a cette fois une prise de position personnelle, un choix méthodologique, qui ouvre des possibilités à la recherche, mais en ferme d'autres. On entend bien que l'auteur, qui connaît son sujet pour l'avoir pratiqué pendant des années, souhaite éviter à ses collègues de s'engager dans une impasse et leur épargner un temps précieux. Il y a un risque, car après tout rien ne prouve que demain ou dans 10 ans un autre chercheur ne publiera pas une théorie déterministe expliquant merveilleusement et intégralement le phénomène en question, et nous aurons peut-être raté le coche pour avoir suivi le conseil implicite de notre auteur : c'est là un risque réel, mais normal et inhérent à la pratique du travail scientifique, puisque de toutes les façons nous sommes obligés de faire des choix méthodologiques. Entendu dans ce sens, purement opératoire, le recours au hasard, c'est-à-dire, en réalité, la décision d'utiliser des modèles probabilistes, est parfaitement légitime, et ne nous fait pas sortir du cadre des disciplines positives.

Mais parfois aussi, plus souvent peut-être dans des ouvrages de vulgarisation ou de synthèse philosophique que dans des travaux proprement scientifiques, on glisse à une interprétation tout autre. On suggère, ou même on affirme, que le Hasard (parfois avec une majuscule), en temps que tel, exerce une action décisive sur le cours des événements : il ne s'agit plus d'un choix méthodologique, le Hasard est hypostasié, muni d'attributs positifs, érigé en deus

ex machina. Dans ces conditions, imputer le phénomène au Hasard ou à la Providence constituent deux attitudes équivalentes, étrangères à la méthodologie scientifique.

Le Hasard, concept métaphysique.

L'usage illégitime des concepts scientifiques au-delà des limites à l'intérieur desquelles ils ont un sens opératoire : c'est là le passage subreptice à la métaphysique, danger qui guette même les plus grands. Prenons un ouvrage célèbre, écrit par un biologiste de haut renom, le "Hasard et la Nécessité" de J. Monod. A la dernière page, nous trouvons cette affirmation angoissante : "l'ancienne alliance est rompue ; l'homme sait enfin qu'il est seul dans l'immensité indifférente de l'Univers d'où il a émergé par hasard". Ici, le mot est visiblement employé dans une acception métaphysique (non opératoire). Ailleurs¹, nous saisissons sur le vif le passage d'un statut à l'autre :

"Nous disons que ces altérations [i.e. les altérations génétiques] sont accidentelles, qu'elles ont lieu au hasard². Et, puisqu'elles constituent la seule³ source possible de modifications du texte génétique, seul³ dépositaire à son tour des structures héréditaires de l'organisme, il s'ensuit nécessairement que le hasard est la seule³ source de toute nouveauté, de toute création dans la biosphère. Le hasard pur, le seul hasard, liberté absolue, mais aveugle, à la racine même du prodigieux édifice de l'évolution : cette notion centrale est la seule³ concevable, comme seule compatible avec les faits d'observation et d'expérience".

La première phrase (les altérations génétiques sont dues au hasard) peut encore être entendue dans un sens scientifique (c'est-à-dire opératoire), et signifie dans ce cas : les données expérimentales en notre possession sont compatibles avec tel modèle probabiliste (qu'il aurait d'ailleurs fallu spécifier), et nous ne disposons d'aucun modèle déterministe apte à en rendre compte de

1 - p. 127 2 - souligné par moi 3 - souligné par J. Monod.

façon plus satisfaisante. Dès la seconde phrase, par contre, le hasard est crédité d'un rôle actif, positif : il constituerait la seule source possible de toute nouveauté dans la biosphère. Dans quel sens le hasard peut-il être dit "source" de quelque chose ? Qu'il y ait ici passage subreptice à la métaphysique, il suffit, pour s'en rendre compte, de remplacer le mot "hasard" par sa définition opératoire. Cela donne à peu près : "le fait que, jusqu'à présent, seuls des modèles probabilistes se soient révélés compatibles avec ce que nous savons des altérations génétiques constitue la seule source de toute création et de toute nouveauté dans la biosphère". Le sentiment de malaise que nous éprouvons à la lecture de cet étrange énoncé est lié à l'hétérogénéité radicale du sujet (une constatation méthodologique, relative à notre modeste pratique scientifique et à son échelle limitée) et du rôle actif et universel qui lui est attribué (celui d'un Demiurge cosmologique).

J. Monod avait en partie conscience de cette redoutable confusion des genres. Aussitôt après le passage que nous venons de citer, il ajoute en effet qu'il est "très important de préciser dans quel sens exact le mot de hasard peut et doit être employé, s'agissant des mutations comme source de l'évolution". Il introduit alors une distinction entre deux types d'incertitude : une "incertitude opérationnelle" et une "incertitude essentielle". Je ne suis pas certain qu'il s'agisse d'une distinction pertinente. Car, sous peine de sortir du cadre des disciplines positives, il faudrait définir un critère opératoire permettant de discriminer sans équivoque les deux types d'incertitude. On peut douter que cela soit possible. La base proprement empirique sur laquelle nous fondons notre diagnostic d'incertitude est, en effet, la même dans les deux cas : dans l'état actuel de nos connaissances, nous ne sommes pas à même de prévoir avec certitude tel événement ou tel phénomène. Dans le premier cas, nous admettons qu'une amélioration considérable (concevable, au moins théoriquement, même si nous n'avons guère d'espoir de la voir se réaliser en pratique) de nos connaissances ou de nos moyens d'investigation nous permettrait de lever cette incertitude. Dans le second cas, nous décrétons catégoriquement que cela même est exclu : mais c'est nous qui le décrétons,

non les données empiriques en notre possession, qui restent parfaitement muettes sur ce point. Ce n'est pas une constatation, mais une interprétation surajoutée, une extrapolation à l'infini, et par cela même illégitime, même si elle est basée sur des considérations théoriques sérieuses : car, dans le domaine des disciplines empiriques, nous ne pouvons extrapoler à l'infini une théorie, si bien corroborée soit-elle, sans sortir ipso facto des limites à l'intérieur desquelles cette théorie possède un sens opératoire et a reçu la sanction de l'expérience.

La parabole du Plombier et du Médecin.

Il est remarquable, d'ailleurs, que l'illustre biologiste n'ait en aucune façon songé à définir un critère opératoire permettant de distinguer les deux types d'incertitude. "Le mieux, nous dit-il, est de prendre quelques exemples". Le premier exemple, classique, est celui du jeu de dés ou de la roulette. Il est bien certain que si nous connaissions avec une précision suffisante les conditions initiales (position, vitesse et moment cinétique du dé au moment où il quitte le cornet) et les conditions aux limites (la topographie précise de la pièce où se joue la partie) les lois de la mécanique nous permettraient, en principe, de prévoir avec certitude le numéro destiné à sortir. Or, il se trouve que de très légères variations de ces conditions initiales, très inférieures à celles que nous sommes capables de contrôler expérimentalement, suffisent pour modifier le résultat final (amener le 5 au lieu du 6 etc...). Nous retrouvons ici la condition d'"inséparabilité des conditions initiales" caractéristique, selon J. Ullmo¹, de l'introduction des modèles probabilistes en physique. Il s'agit bien d'une "incertitude opérationnelle, et non essentielle", et c'est bien "pour des raisons purement méthodologiques" que nous avons recours au calcul des probabilités. Ceci est incontestable, et somme toute banal. Nous attendons avec quelque impatience le second exemple, celui qui va nous éclairer sur la nature quelque peu mystérieuse du second type d'incertitude, l'incertitude "essentielle". Hélas, ce qui suit n'est rien d'autre que la "parabole" archi-connue du plombier Dubois travaillant sur une toiture et

1 - Voir, par exemple, "Les concepts physiques", dans "Logique et connaissance scientifique", par J. Piaget et al., Paris, 1967, La Pléiade.

laissant par inadvertance tomber son marteau qui vient fracasser le crâne du Docteur Dupont passant justement par là, appelé de toute urgence auprès d'un de ses malades. Aucun lecteur de ce beau et grand livre qu'est "Le Hasard et la Nécessité" n'a pu, je pense, se défendre d'un sentiment de malaise en rencontrant un argument aussi faible tombé d'une plume aussi illustre. Cette "parabole" est censée définir la notion de "coïncidence absolue" ou, ce qui reviendrait au même, de "hasard essentiel", c'est-à-dire, nous dit Monod, reprenant une idée qui remonte à Cournot, "inhérent à l'indépendance totale des deux séries d'évènements dont la rencontre produit l'accident".

La terminologie employée dénonce d'elle-même l'extrapolation au-delà de toute limite. La coïncidence est absolue, l'indépendance est totale. Ces adjectifs ont-ils un sens, appliqués à des notions approximatives, de portée limitée, comme la coïncidence ou l'indépendance? Il y a d'ailleurs une certaine contradiction à parler de deux séries d'évènements "totalement" indépendantes, et qui pourtant en viennent à se rencontrer. La notion d'"indépendance" utilisée ici n'est même pas définie. On nous dit seulement que ces deux séries d'évènements "n'avaient rien à voir" l'une avec l'autre. Mais, quel que soit le sens de ce langage approximatif, il est clair que c'est à nos yeux seulement que ces deux séries n'avaient "rien à voir". Ce qui signifie, à nouveau, que nous étions incapables de prévoir leur "rencontre" et l'"accident" qui devait en résulter : rien de plus. En quoi cette imprévisibilité diffère-t-elle de celle qui entache une partie de dés? La différence existe bien, mais tient au niveau d'abstraction où nous nous sommes placés pour envisager les deux catégories d'évènements, et non à la "nature" de ces derniers. Dans le premier cas, il s'agissait d'une partie de dés en général, dans le second d'un évènement unique, singulier, entièrement spécifié par l'identité du plombier et du médecin (on aurait pu aussi bien ajouter le lieu et la date) : seule, la dissymétrie des deux points de vue crée une impression de contraste. Or je peux particulariser le premier point de vue, celui qui concernait les parties de dés, et proposer la "parabole" suivante, symétrique de celle du plombier et du médecin : "ce jour-là, à telle heure, M. Duval entre dans le café X, rencontre son vieil ami M. Durand qu'il

n'avait pas vu depuis trois ans, arrose abondamment cet heureux évènement, puis, sur la suggestion du barman, entame avec M. Durand une longue partie de dés, au 26ème coup de laquelle le dé, maladroitement lancé, vient atterrir dans la tasse de café que M. Duparc venait de vider et de poser sur le comptoir, amenant ainsi le chiffre 6. Il s'agit cette fois d'un évènement tout aussi imprévisible, d'un "hasard" tout aussi "essentiel" que dans la fable du plombier et du médecin. La morale de cette histoire triviale est sans doute que n'importe quel évènement singulier survenant dans le monde où nous vivons, dès que nous l'individualisons en spécifiant tous les détails qui l'ont accompagné, y compris les conditions de temps et de lieu, apparait comme le fruit de la rencontre, non pas de deux mais d'un très grand nombre de "séries d'évènements" qui n'avaient pas grand chose "à voir" les uns avec les autres, et relève d'un "hasard" ni plus ni moins "essentiel" que celui qu'invoque J. Monod. Et ceci revient seulement à dire, ce qui est banal, qu'un évènement unique et singulier, en tant que tel, ne peut pas relever d'une connaissance scientifique, la condition de "répétabilité" qui, seule, fonde l'objectivité de nos disciplines empiriques faisant ici défaut par définition.

En sens inverse, si nous dépouillons par la pensée l'évènement unique et singulier de telles ou telles circonstances anecdotiques qui l'accompagnent (et, notamment, des conditions de temps et de lieu) de manière à le réduire au statut d'occurrence particulière d'une classe plus générale d'évènements susceptibles de se répéter, alors, par cette démarche même, nous le constituons en tant qu'objet d'une connaissance scientifique possible. Et, par là même, nous devenons capables d'avancer des modèles, ou théories (déterministes ou probabilistes), susceptibles d'être corroborés ou réfutés selon qu'ils se révéleront ou non "en accord" non seulement avec cet évènement-ci mais aussi avec tous ceux qui relèvent de la même classe (constituée par nous) et que nous aurons eu l'occasion d'observer. Par exemple, nous serions ici conduits à proposer un modèle probabiliste contenant au moins trois évènements, à savoir : A (chute du marteau d'un plombier en un point et un instant donné), B (présence du crâne d'un passant au même point et au même instant)

et AB (conjonction des deux précédents). L'assertion relative à l'indépendance de ces deux événements signifierait simplement que, dans ce modèle, la probabilité de AB est égale au produit des probabilités de A et de B, soit la relation habituelle $P(AB) = P(A) P(B)$. A ce stade, il s'agit seulement d'une caractéristique de notre modèle et non d'un énoncé relatif à la réalité. Pour franchir ce pas et obtenir un énoncé ayant une signification objective, nous devons attribuer à notre modèle d'autres caractéristiques encore (par exemple la stationnarité dans le temps) moyennant lesquelles nous pourrions légitimement identifier (à un degré d'approximation que la statistique mathématique permet en principe d'évaluer) la probabilité $P(A)$ d'un événement A avec la fréquence moyenne (empirique) $v(A)$ de sa réalisation au cours du temps. Alors, mais alors seulement, nous obtenons un énoncé objectif, c'est-à-dire expérimentalement contrôlable. A savoir : la fréquence $v(AB)$ de l'événement produit doit être égale (avec le même degré d'approximation que ci-dessus) au produit $v(A) v(B)$ des fréquences des deux événements en question.

Incidentement, et bien que nous ne disposions pas des données statistiques nécessaires, nous pouvons prévoir que, selon toute vraisemblance, la relation $v(AB) = v(A) v(B)$ ne serait pas vérifiée. Pour la raison simple que les moments où il n'y a pas, en général, de plombiers sur les toits (la nuit, les jours de grand froid ou d'averses torrentielles etc...) sont aussi, en moyenne, ceux où il n'y a que très peu de passants dans les rues. Ainsi, les deux événements qui nous étaient présentés comme l'exemple typique de l'"indépendance totale" révéleraient sans doute, à l'analyse, une corrélation positive, statistiquement significative.

Les chaînes polypeptidiques.

C'est avec un sentiment de gêne que j'avance des critiques aussi pesantes et triviales à l'encontre d'un livre que j'admire par ailleurs. Mais ces critiques ne sont peut-être pas inutiles : si un grand esprit scientifique, comme J. Monod, s'est laissé prendre à un piège épistémologique aussi grossier, c'est qu'apparemment le danger est réel. Pour examiner un exemple moins banal, voyons plutôt

comment notre auteur utilise la notion de hasard dans le domaine qui est le sien, la biologie moléculaire. La même confusion des genres, malheureusement, va se manifester. Ouvrons le livre à la page 110. Il s'agit des protéines, molécules "constituées par la polymérisation linéaire de corps appelés amino-acides". Ces amino-acides, au nombre de 20 au total, se succèdent dans un ordre quelconque le long de cette "chaîne polypeptidique", et le seul modèle compatible avec ce que l'on sait aujourd'hui de cette succession est celui de l'alternative répétée généralisée, à savoir : la fréquence (ou, dans le modèle, la probabilité) avec laquelle un acide donné se manifeste en un point de la chaîne ne dépend ni de la position de ce point dans la chaîne, ni de la nature des acides occupant les positions voisines. En termes probabilistes, il y a (dans ce modèle) à la fois stationnarité le long de la chaîne, et indépendance (stochastique) entre les événements affectant les différents points de cette chaîne. Mais citons plutôt.

"Dire de la séquence des amino-acides dans un polypeptide qu'elle est "au hasard" ne revient nullement, il faut insister là-dessus, à un aveu d'ignorance, mais exprime une constatation de fait : à savoir que, par exemple, la fréquence moyenne avec laquelle tel résidu [= acide] est suivi de tel autre dans les polypeptides est égal au produit des fréquences moyennes de chacun des deux résidus dans les protéines en général". Sous cette forme, l'énoncé est absolument inattaquable. Il ne déborde en aucune manière les limites opératoires du modèle, et ne fait appel qu'à des faits effectivement observés. Malheureusement, il est immédiatement précédé d'un autre énoncé, nullement équivalent, loin de là : "Ces structures (celles des polypeptides) sont "au hasard" en ce sens que, connaissant exactement l'ordre de 199 résidus dans une protéine qui en contient 200, il est impossible de formuler aucune règle, théorique ou empirique, qui permettrait de prévoir la nature du seul résidu non encore identifié par l'analyse".

Bien loin d'être équivalents, ces deux énoncés sont incompatibles et se réfutent l'un l'autre, du moins si nous les considérons comme des énoncés empiriques. Le premier, en effet, affirme qu'un modèle probabiliste bien défini, celui de l'alternative

répétée généralisée, n'a, jusqu'à présent, pas été démenti par l'expérience. Or, pour contrôler un modèle de nature probabiliste, la démarche usuelle consiste à procéder à des tests statistiques : on fait choix (avant, évidemment, d'avoir pris connaissance des données, ou du moins, sans se laisser influencer par elles) d'un évènement ayant (dans le modèle) une probabilité très faible, et on convient de rejeter le modèle si cet évènement est, en fait, réalisé. Il y a, toujours, une certaine équivoque, puisqu'après tout, un évènement peu probable peut, malgré tout, se produire. Mais cette ambiguïté est d'autant plus réduite que la probabilité de l'évènement en question est plus faible. En particulier, si un évènement (défini à l'avance) dont la probabilité, dans le modèle, est inférieure à 10^{-10} ou même à 10^{-100} se réalisait néanmoins, tout le monde s'accorderait à considérer le modèle comme réfuté (dans le cas contraire, il serait tout simplement impossible, faute de critères de contrôle, d'utiliser les modèles probabilistes dans le domaine des sciences empiriques).

Pour servir de test, dans le cas qui nous occupe, choisissons l'évènement suivant, que nous appellerons A pour abrégé : "Parmi les protéines distinctes dont la formule chimique est actuellement connue, il y en a au moins deux qui présentent la même séquence initiale de 199 acides aminés". Evaluons grossièrement la probabilité $P(A)$ de cet évènement dans le modèle "alternative répétée généralisée". J'ignore le nombre exact des protéines distinctes dont la formule chimique est connue aujourd'hui. Mettons, pour être au large, qu'il y en ait cent mille, soit $n = 10^5$, et supposons que les fréquences de chacun des 20 acides aminés soient toutes égales à $1/20$ (en réalité, les différents acides ne sont pas équiprobables, mais un calcul rigoureux, tenant compte des fréquences empiriques observées pour chacun d'eux, conduirait aux mêmes ordres de grandeur). La probabilité pour qu'une molécule donnée commence par une séquence donnée de 199 acides est donc (dans le modèle) $(20)^{-199}$ soit environ $(10)^{-258}$. Il est alors facile de voir que $P(A)$ est au plus égal à $(1/2)n^2 \times (10)^{-258}$, donc, certainement :

$$P(A) \leq 10^{-248}$$

Cette probabilité est inconcevablement petite. Si l'évènement A était réalisé, ce serait l'occasion ou jamais de rejeter un modèle probabiliste.

Or, le second énoncé de Monod - si nous l'interprétons comme une constatation empirique, et non comme le simple résultat d'une déduction théorique effectuée dans le modèle sans avoir reçu de confirmation expérimentale - implique que cet évènement A, extraordinairement improbable, s'est en fait réalisé. En effet, nous ne pouvons affirmer comme un fait expérimentalement prouvé qu'il n'existe aucune règle théorique ou empirique permettant de prévoir le 200^{ème} acide, connaissant les 199 premiers, que si nous avons observé effectivement au moins une fois deux protéines commençant par la même séquence de 199 acides et ne différant l'une de l'autre qu'à partir du 200^{ème} résidu. Or cela implique que A est réalisé et, par cela même, conduit à rejeter le modèle probabiliste.

On peut aller plus loin, et montrer que le modèle probabiliste non seulement n'exclut pas, mais au contraire implique l'existence d'une règle empirique permettant de prévoir le 200^{ème} acide à partir des 199 premiers. Sachant que le nombre de protéines distinctes présentes dans un animal supérieur est de l'ordre du million, et admettant que, dans notre galaxie, un milliard de planètes, au plus, ont pu donner naissance chacune à un milliard d'espèces différentes, nous pouvons évaluer à 10^{24} le nombre des protéines existant ou ayant existé dans notre galaxie (et il s'agit certainement d'une évaluation par excès). Or, la probabilité (dans le modèle) pour que, parmi 10^{24} molécules distinctes, il en existe deux présentant la même séquence initiale de 199 résidus est au plus $(1/2)(10)^{48}/(20)^{199}$, soit environ 10^{-210} : probabilité tellement faible, que l'évènement correspondant doit être considéré comme pratiquement impossible. Autrement dit, et avec un risque d'erreur inférieur à 10^{-210} , nous pouvons affirmer que la donnée de ses 199 premiers résidus permet d'identifier à coup sûr, et sans équivoque, n'importe laquelle des protéines existant ou ayant existé dans notre univers. Cela revient bien à dire qu'il existe une règle (certes, purement empirique) permettant de prévoir sans risque d'erreur non seulement le 200^{ème} résidu, mais également tous les autres, à partir des 199 premiers.

La nature de cette règle est sans mystère. On peut la comparer à l'état-civil. La donnée d'un petit nombre de paramètres (nom, prénoms, date et lieu de naissance) permet, avec un risque d'erreur négligeable, d'identifier chacun des 50 millions de citoyens de notre pays. Si donc, comme on nous en menace, les renseignements biographiques relatifs à chacun d'eux étaient mis en mémoire dans quelque ordinateur, le premier venu (sous réserve d'avoir accès au Grand Ordinateur) serait capable de prévoir à coup sûr la taille, la profession, etc... de n'importe quel citoyen français à partir de son seul état-civil. Qu'on ne se hâte pas de décréter qu'un tel mode opératoire est étranger à la science. Car les classifications systématiques, en botanique ou zoologie, ne procèdent pas de façon fondamentalement différente.

Dans le cas des protéines, la règle permettant de faire cette prévision est actuellement, et restera sans aucun doute à tout jamais, hors de notre portée, puisqu'il est bien certain que nous ne connaissons jamais la formule exacte de 10^{24} protéines. Mais elle n'en existe pas moins (ou alors, le modèle probabiliste de Monod est, par là même, réfuté). Déjà, d'ailleurs, sur la base des quelques centaines ou milliers de protéines dont nous connaissons effectivement la structure, une certaine forme de prévision est possible : si, lors de l'analyse d'une nouvelle espèce, nous trouvons que les 199 premiers résidus coïncident avec ceux d'une molécule déjà connue, nous pouvons parier à coup sûr (si le modèle est bon) qu'il s'agit de la même protéine, et donc prévoir en toute certitude les résidus non encore analysés.

Un paradoxe apparent.

Le lecteur aura peut-être été troublé par ce paradoxe apparent : malgré, ou plutôt justement à cause du caractère "purement aléatoire" du modèle utilisé pour décrire la succession des acides aminés le long de la chaîne de polypeptides, nous sommes à même d'affirmer, avec un degré de certitude rarement atteint ailleurs, (une probabilité d'erreur au plus égale à 10^{-210} !) qu'il existe une règle permettant de prévoir le 200^{ème} résidu connaissant ses 199 prédécesseurs. Il ne s'agit, dira-t-on peut-être, "que d'une

règle empirique", très compliquée, que nous ne saurons pas expliciter autrement que sous la forme d'un inventaire exhaustif, et dont aucune théorie simple ne pourra rendre compte. Mais est-ce là une objection? Personne, et Monod moins que tout autre, n'avait exigé de cette règle qu'elle soit "simple". Outre qu'il n'est pas facile de définir avec précision ce que l'on peut entendre au juste, ici, par simplicité, cette règle, si elle existe, est ce qu'elle est : si elle ne nous paraît pas simple, c'est là un jugement de valeur de notre part, c'est-à-dire une appréciation purement subjective qui n'entache en rien l'objectivité de la règle en question. De quel droit, après tout, exigerions-nous du cosmos qu'il se conforme à nos structures mentales ou à nos préférences esthétiques?

Il n'est d'ailleurs nullement certain que la règle en question ne soit pas "simple", et nous aurions en tout cas bien du mal à prouver expérimentalement qu'elle ne l'est pas. Prenons un exemple. Rangeons les 20 acides aminés dans un ordre quelconque, mais fixe (ce qui est possible de $20! = 2,4 \cdot 10^{18}$ manières différentes), de sorte qu'à chaque acide corresponde un nombre a bien défini compris entre 0 et 19. La règle suivante (appelons-la R) :

$$(R) \quad a_{200} = a_1 + a_2 + \dots + a_{199} \text{ (modulo 20)}$$

peut certainement être dite simple. Cette règle nous semble a priori peu plausible, parce que, dirions-nous, il n'y a "aucune raison" pour que les choses se passent ainsi : j'en conviens volontiers. Mais, pour démontrer expérimentalement que cette règle est fautive, nous devrions essayer successivement les $2,4 \cdot 10^{18}$ versions possibles, et exhiber pour chacune d'elles un contre-exemple expérimental. Or, personne ne l'a fait, et, certainement, ne le fera jamais. A supposer qu'un millième de seconde suffise en moyenne pour trouver chaque contre-exemple, il n'en faudrait pas moins un million de siècles pour achever la vérification. Et, bien sûr, on peut envisager bien d'autres types de règles "simples". On objectera qu'aucune personne sensée, de nos jours, n'accepterait d'envisager la possibilité de règles aussi "absurdes", c'est-à-dire ne présentant aucun rapport intelligible avec le présent contexte biologique, et pour ma part je ne conseillerais pas à un jeune

chercheur de s'engager dans une voie de recherche apparemment aussi stérile. Toutefois, n'oublions pas que c'est pour nous seulement qu'il n'y a pas de rapport intelligible entre cette règle et ce phénomène, et que, d'ailleurs, le bon sens qui nous fait rejeter aujourd'hui, sans examen, une telle hypothèse est de date relativement récente. Pendant des siècles, l'humanité s'est complue à des spéculations arithmologiques bien plus étranges.

Pour rendre plus piquant le paradoxe, notons encore que l'application de la règle R conduirait à des séquences numériques qui ressembleraient à s'y méprendre (sauf pour quelques choix très particuliers et peu nombreux de la séquence initiale, par exemple $a_1 = a_2 = \dots = a_{199}$) à des séquences "purement aléatoires", c'est-à-dire obtenues par des tirages au sort indépendants. Plus précisément, il est facile de démontrer le théorème suivant : Si A_1, A_2, \dots, A_{199} sont 199 variables aléatoires indépendantes et uniformément distribuées sur $\{0, 1, \dots, 19\}$, alors $A_{200} = A_1 + \dots + A_{199}$ (modulo 20) est elle-même une variable aléatoire uniformément distribuée sur les 20 premiers entiers, et, surtout, les 199 variables A_2, A_3, \dots, A_{200} sont encore mutuellement indépendantes. C'est seulement à l'ordre 200, c'est-à-dire à la condition de faire intervenir explicitement la loi de distribution simultanée des 200 variables A_1, \dots, A_{200} , qu'il est possible de mettre en évidence l'effet de dépendance fonctionnelle. Autrement dit, les tests statistiques usuels, même si l'on dispose d'un nombre très élevé (par exemple $n = 10^{24}$) de séquences de 200 valeurs numériques obtenues par ce procédé, concluront régulièrement et inmanquablement à l'absence complète de dépendance.

C'est d'ailleurs sur des procédés très analogues que reposent les techniques de simulation, que connaissent bien tous ceux qui ont eu l'occasion d'utiliser en pratique la théorie des probabilités. Contrairement à ce que l'on pourrait croire, il n'est pas du tout facile de choisir un nombre "au hasard" entre, disons, 1 et 10^6 , et encore moins facile d'effectuer un grand nombre de tels tirages au sort indépendamment les uns des autres. Cette difficulté technique, bien connue des praticiens, devrait nous inciter à plus de circonspection dans le maniement de notions faussement

claires comme celle de "hasard". Il serait bien imprudent, en effet, d'extrapoler à l'échelle cosmique un concept que nous sommes déjà incapables de contrôler de manière réellement opératoire à une échelle aussi modeste que celle de la simulation de quelques milliers de valeurs numériques. La manière dont on s'y prend, en pratique, est tout-à-fait paradoxale (et par là même très instructive). Les procédés qu'utilisent, en général, les praticiens sont, en effet, d'une nature arithmétique simple, et absolument déterministe, tout-à-fait comparable, en cela, à la règle R ci-dessus. Par exemple, on choisit un nombre initial n_1 "quelconque" entre 0 et 10^6 , on l'élève au cube et on retient les 6 chiffres centraux de l'écriture décimale du nombre obtenu : n_2 est le nombre dont l'écriture décimale est donnée par les 6 chiffres ainsi retenus. On peut réitérer l'opération plusieurs milliers de fois, et soumettre les valeurs obtenues aux tests statistiques les plus sévères : on obtient, en général, une bonne confirmation de l'"indépendance" de ces soi-disant variables aléatoires (sauf, de temps en temps, pour certaines valeurs particulières du nombre initial n_1). Et pourtant, le procédé de génération de ces nombres "au hasard" est absolument déterministe. Nous sommes sûrs ici, puisque c'est nous qui l'avons choisie, qu'il existe une "règle théorique" simple permettant de prévoir le millième nombre connaissant le premier. L'apparence "aléatoire" du résultat, qui résiste aux tests les plus sévères, provient, nul n'en doute, du caractère bizarre, hétérogène, des diverses manipulations numériques que nous effectuons. Elever au cube et retenir les 6 décimales centrales sont des opérations arithmétiques qui n'ont "rien à voir" entre elles, du moins pour nous, et cela permet de "comprendre" pourquoi les nombres ainsi fabriqués nous semblent "au hasard". Mais ils ne le sont pas, et des tests plus puissants effectués sur une suite extrêmement longue feraient apparaître le caractère déterministe (par exemple, après au plus un million d'opérations, on retombe nécessairement sur une valeur numérique déjà écrite, et à partir de ce moment la suite devient périodique, et se reproduit elle-même à l'infini). Or, il ne s'agit nullement ici d'une "curiosité" arithmétique, mais d'un procédé couramment utilisé par les praticiens pour fabriquer des nombres "au hasard" : l'existence d'une règle théorique simple,

de nature déterministe, ne contredit nullement, à leurs yeux, le caractère "aléatoire", empiriquement bien corroboré, des nombres qu'ils obtiennent.

Le seuil de réalisme ou d'objectivité.

Ces exemples illustrent bien, je crois, la différence capitale qui existe entre un théorème et un énoncé empirique. Un modèle donné, aussi bien testé et corroboré qu'il ait pu être, contient toujours et nécessairement des théorèmes qui ne correspondent plus à des énoncés empiriques contrôlés, ni même contrôlables au-delà d'une certaine limite. Il existe toujours un seuil de réalisme, au-delà duquel le mathématicien peut, certes, poursuivre joyeusement ses déductions, mais que le physicien doit respecter, sous peine de ne plus obtenir que des énoncés incontrôlés d'abord, puis, peu à peu, incontrôlables, c'est-à-dire dépourvus de signification objective, ou, si l'on veut, "métaphysiques" au sens que prend cet adjectif dans l'usage des sciences positives. Dans le cas de nos protéines, pour tout entier n , la probabilité (dans le modèle) d'une séquence donnée a_1, \dots, a_n de n acides aminés est égale au produit $P(a_1) \dots P(a_n)$ des probabilités individuelles, soit à $(20)^{-n}$ dans le cas où les différents acides sont équiprobables. C'est là un théorème du modèle. Pour qu'il lui corresponde un énoncé "objectif" ou "empirique", c'est-à-dire expérimentalement contrôlable, il faut que, parmi les 20^n séquences possibles, quelques-unes au moins se trouvent répétées un nombre suffisant de fois parmi les 10^{28} séquences¹ qu'il serait (théoriquement!) possible d'observer. Si nous voulons, par exemple, qu'il y ait (dans le modèle) une probabilité supérieure à $1/100$ pour que l'une, au moins, de ces $(20)^n$ séquences se répète au moins 10 fois (et c'est vraiment le minimum que l'on puisse exiger) on voit facilement que n ne peut pas dépasser une valeur de l'ordre de 24 environ. Un énoncé d'ordre supérieur, par exemple d'ordre $n = 200$, comme celui

1 - En admettant que les protéines contiennent 10^4 acides en moyenne, les 10^{24} molécules de l'univers contiennent (environ) $10^4 \times 10^{24} = 10^{28}$ séquences de 200 acides consécutifs.

de Monod est empiriquement vide¹. Si, maintenant, nous considérons non plus l'ensemble des molécules existant dans notre univers, mais seulement celles dont nous connaissons effectivement la formule, au nombre, mettons, de 10000, ce qui donne au plus 10^8 configurations observées, nous devons nous attendre à ce qu'une séquence donnée de longueur n figure, en moyenne, $10^8 \times 20^{-n}$ fois dans notre matériel expérimental. Pour $n = 2$ et 3 , nous aurons plus de 10^5 et 10^4 répétitions, en moyenne, et par suite de très bonnes possibilités de contrôle expérimental. Pour $n = 6$, déjà, ce nombre moyen est de l'ordre de l'unité : à l'ordre 6, il ne sera plus possible que de contrôler quelques séquences (en très petit nombre) celles qui, "par hasard" se seront répétées, disons, au moins dix fois. Au-delà de $n = 7$ ou 8 , le matériel expérimental disponible aujourd'hui ne permet pratiquement plus aucun contrôle (ou, s'il le permet, cela signifie par le fait même que le modèle probabiliste est réfuté).

En résumé : une assertion relative à l'indépendance d'ordre $n = 10$ ne peut pas, si le modèle est bon, être considérée comme expérimentalement établie aujourd'hui. Quant à la notion d'indépendance d'ordre $n \geq 30$ (et a fortiori d'ordre 200) elle est tout simplement dépourvue de signification objective. Comment, dès lors, pourrions-nous extrapoler à l'échelle cosmique, et affirmer catégoriquement, comme s'il s'agissait d'un fait expérimentalement prouvé, ou même simplement d'un énoncé présentant une signification

1 - Attirons, en passant, l'attention des logiciens sur le statut étrange (qui rappelle, en plus complexe, le célèbre paradoxe du menteur) des énoncés obtenus, comme celui de Monod, en transgressant largement le seuil de réalisme d'un modèle probabiliste : ce sont à la fois des théorèmes et des falsificateurs virtuels du modèle. S'ils sont vrais (empiriquement), ils sont faux (théoriquement) [i.e. : s'ils sont vérifiés expérimentalement, ou même simplement si les conditions d'un contrôle expérimental possible se trouvent réunies, le modèle dont ils sont déduits est, par là même, réfuté]. S'ils sont vrais (théoriquement), ils sont incontrôlables (empiriquement) [Si le modèle peut être considéré comme bon, i.e. non démenti par l'expérience, les conditions d'un contrôle expérimental de l'énoncé ne sont pas réunies]. C'est pourquoi ils peuvent être dits objectivement vides : plus précisément, ils ne pourraient avoir de signification objective que si le modèle d'où on les a déduits était réfuté par l'expérience.

objective quelconque, que "le hasard constitue la seule source possible de toute création et de toute nouveauté dans la biosphère"? Il est vraisemblable que J. Monod a conçu sa philosophie, avant tout, comme une machine de guerre contre celle de Teilhard de Chardin¹. C'est ce qui explique leur parenté. Le hasard de Monod est le frère ennemi du point Omega du bon Père : son ennemi, certes, mais fondamentalement son frère ; ils sont bien de la même famille.

1 - Les sentiments de Monod à l'égard de Teilhard de Chardin me semblent du genre "amour déçu, tourné en haine". Althusser a relevé les nombreuses réminiscences teilhardiennes qui émaillent les textes de Monod, et analysé (malheureusement, dans le langage et selon le point de vue un peu rigides d'une Ecole) les rapports ambigus qui existent entre ces deux auteurs. On trouvera une bonne analyse, (dans un langage normal, et selon un point de vue moins particulier) et une bibliographie dans Madeleine Barthélémy-Madaule, "L'idéologie du hasard et de la nécessité", Paris, 1977. Cet auteur, philosophe universitaire, me semble avoir fort bien vu l'essentiel : à savoir que l'illustre biologiste, lorsqu'il employait des mots comme "hasard", ou "probabilité", ne savait tout simplement pas de quoi il parlait. Mais, n'étant pas elle-même probabiliste, elle ne s'est pas crue autorisée à dire les choses aussi brutalement. Le "modèle" de Monod a d'ailleurs été sérieusement contesté sur le fond, (c'est-à-dire du point de vue de la biologie) par les biologistes eux-mêmes. Il ne m'appartient pas, faute de compétence, de prendre position dans ce débat, et je ne peux que renvoyer à l'exposé de J. Ruffié, dans "De la biologie à la culture", Paris, 1976, pp. 205 et suivantes. Le lecteur aura compris que mon seul but, dans les pages qui précèdent, était d'illustrer, sur un exemple frappant, les dangers que comporte l'extrapolation à l'infini à partir de concepts ou de modèles probabilistes.

- CHAPITRE II -

POURQUOI NOUS NE SOMMES PAS D'ACCORD AVEC LES ETRUSQUES

ou

DE L'OBJECTIVITE DES MODELES PROBABILISTES

"Voici en quoi nous ne sommes pas d'accord avec les Etrusques, spécialistes de l'interprétation des foudres. Selon nous, c'est parce qu'il y a collision de nuages que la foudre fait explosion. Selon eux, il n'y a collision que pour que l'explosion se fasse. Comme ils rapportent tout à la divinité, ils sont persuadés, non pas que les foudres annoncent l'avenir parce qu'elles ont été formées, mais qu'elles se forment parce qu'elles doivent annoncer l'avenir".

Sénèque (Questions Naturelles, II, 32)

LE PROBLEME.

Dans les pages qui précèdent, nous avons assisté au franchissement de ce que nous avons appelé le seuil d'objectivité. d'un modèle probabiliste, seuil au-delà duquel la déduction ne fournit plus que des énoncés empiriquement vides. mais existe-t-il réellement un tel seuil en-deça duquel un modèle probabiliste soit susceptible de présenter quelque signification objective ? On sait qu'une école de pensée le conteste. Les "subjectivistes",¹ ou partisans de l'interprétation purement subjective de "la" probabilité, remarquent qu'il est impossible de donner un sens objectif à un énoncé probabiliste concernant un évènement unique et singulier. Si je dis, par exemple, "il y a une chance sur deux pour qu'une forme de vie intelligente existe sur la planète Mars", on ne peut voir dans cet énoncé rien d'autre qu'une opinion personnelle (purement subjective). Car, si un contradicteur m'objecte "mais non, cette probabilité est 7/10 (ou 3/10 etc..)" rien ne pourra jamais nous départager: nous saurons vraisemblablement dans un proche avenir si Mars est ou non "habité", mais cette constatation expérimentale ne confirmera ni ne réfutera aucun de nos énoncés probabilistes, qui resteront ainsi indécidables et indécidés à tout jamais. L'argument est incontestable. mais en résulte-t-il, nécessairement, que "la" probabilité ne puisse jamais, ni dans aucun contexte, représenter rien d'autre qu'une opinion "purement" subjective ? Les objectivistes rétorquent qu'ils ne s'intéressent en aucune façon à la prédiction d'un évènement unique et singulier : il s'agit là, pour eux, d'une question étrangère à la science.

1 - On les appelle aussi "bayesiens" à cause de l'usage intensif qu'ils font du théorème de Bayes (bien classique pourtant) : dénomination impropre, car ce théorème, comme tout théorème, est de nature purement formelle (mathématique) et ne préjuge en rien de l'interprétation subjectiviste ou objectiviste que l'on peut donner, dans les applications, aux probabilités qu'il met en jeu. Les physiciens, objectivistes par vocation, utilisent eux aussi ce théorème (par exemple, dans l'analyse de la "rétrodictio" terme forgé par analogie avec "prédiction" pour désigner la reconstitution du passé à partir du présent).

Pour citer K.R. Popper ¹ "toute controverse relative à la question de savoir s'il y a des événements qui ne se produisent qu'une fois et ne peuvent être répétés ne peut être tranchée par la science : il s'agit d'une controverse métaphysique". Ce qui les intéresse, c'est seulement le comportement statistique des gros effectifs, et les lois parfaitement objectives qui le régissent. Ce qui n'est qu'une probabilité à l'échelle de l'individu émerge, par le jeu de la loi des grands nombres, sous la forme d'une fréquence que nous pouvons atteindre et mesurer. Nos énoncés probabilistes sont donc parfaitement objectifs, puisque nous pouvons les contrôler expérimentalement, sous réserve de répéter un nombre suffisant de fois la même expérience. A quoi les Bayésiens répondent : "C'est vous qui jugez qu'il s'agit de la même expérience. C'est là une opinion subjective que nous ne sommes nullement obligés de partager. De cette opinion que vous avez adoptée, il résulte logiquement qu'après un nombre n assez élevé d'expériences vous devez utiliser la fréquence observée pour évaluer la probabilité (subjective) que vous attribuez à la $(n+1)^{\text{ème}}$ expérience, non encore réalisée. Mais cela ne signifie en aucune façon que cette probabilité ait acquis une signification objective". Et le dialogue de sourds pourra continuer longtemps.

On perçoit bien quelque chose d'outrancier dans l'argument des subjectivistes. Devant les réussites indiscutables, éclatantes, obtenues par les modèles probabilistes dans des domaines aussi divers que la thermodynamique, la mécanique quantique, les assurances, la démographie etc., devant les prévisions précises et régulièrement confirmées par l'expérience auxquelles ils conduisent, il est tout simplement déraisonnable de nier leur valeur objective. Les physiciens, forts de leurs succès, sont convaincus à juste titre de l'objectivité de leurs "probabilités", et opposent une fin de non recevoir catégorique aux empiètements des subjectivistes. Mais, en sens inverse, les arguments de ces

1 - La Logique de la découverte scientifique, p. 43 de l'édition française.

derniers sur l'impossibilité de donner un sens objectif à la probabilité d'un évènement unique sont apparemment irréfutables. On pourrait, comme Salomon, proposer de couper l'enfant en deux, je veux dire abandonner sans recours les phénomènes uniques à l'arbitraire des subjectivistes, et limiter l'objectivité des méthodes probabilistes au seul domaine des phénomènes "réguliers et répétables" semblables à ceux que les physiciens étudient. Mais notre but, dans ces pages, est justement d'examiner si, et dans quelle mesure, un phénomène unique peut faire l'objet d'une estimation ou d'une prévision sur la base d'une information fragmentaire : comme les techniques d'estimation utilisées en pratique sont, le plus souvent, fondées sur des modèles probabilistes, il n'est pas possible d'éluider le débat.

QUELQUES ILLUSIONS ANTHROPOMORPHIQUES.

Le chapitre précédent aura fait justice, j'espère, de l'illusion substantialiste prêtant au Hasard, avec une majuscule, je ne sais quelle responsabilité cosmique. Mais les illusions de ce genre ont la vie dure, car notre nature est ainsi faite que nous recherchons avec avidité des explications profondes et définitives, capables croyons-nous de nous éclairer sur la nature mystérieuse des choses. En physique, les modèles déterministes utilisés à l'échelle macroscopique correspondent, souvent, à la représentation intuitive d'une action qui se transmet de proche en proche et se traduit, mathématiquement, par une ou plusieurs équations aux dérivées partielles. S'agissant de cordes vibrantes, par exemple, il nous paraît clair que le déplacement d'un petit élément situé au point x exerce une action sur l'élément voisin situé en $x + \delta x$, et l'entraîne à son tour. Notre intuition motrice, telle qu'elle a pu se former au cours d'une évolution qui a requis des millions d'années, nous fournit ici des représentations spontanées, de nature archétypiques, qui nous font croire que nous comprenons ce qui se passe réellement. Qu'il y ait là quelque illusion anthropomorphique (ou zoocentrique), c'est ce que la microphysique nous a appris à notre détriment - ou à

notre avantage, car nous avons ainsi assimilé cette leçon essentielle : le modèle, jamais, n'est identique à la réalité. D'innombrables aspects du réel lui échappent toujours, et, inversement, le modèle contient toujours d'innombrables propositions parasites, sans contrepartie aucune dans la réalité. S'il s'agit d'un phénomène unique et d'un modèle probabiliste, c'est-à-dire d'un "espace" (Ω, \mathcal{A}, P) mis en correspondance avec cette réalité unique, le même genre d'illusion nous incite à dire que tout se passe, en somme, comme si l'évènement qui se réalise avait été "tiré au sort" selon la loi P dans l'espace Ω . Mais c'est là un énoncé faussement clair, et les représentations sous-jacentes qui le nourrissent sont tout particulièrement inadéquates. Quel est le mécanisme de ce "choix au hasard" que nous invoquons, quel croupier céleste, quel joueur de dé cosmique agite ici les cornets de fer de la nécessité ? Ce mythe, car c'en est un (au sens péjoratif), du "tirage au sort" est à la fois inutile et gratuit. Gratuit : car, à supposer que quelque demiurge ait procédé une fois pour toutes, en un tirage au sort unique, à l'élection d'un élément ω_0 , nous n'aurions de toute façon aucun espoir de reconstituer jamais, ni cet espace Ω , ni cette probabilité P . Car, s'agissant d'un évènement unique, les seules sources d'information dont nous pourrions disposer seraient contenues dans l'unique élément ω_0 qui aurait été ainsi choisi le premier jour : tout le reste, l'espace Ω entier, et l'océan infini des virtualités qu'il contenait, s'est évanoui, effacé pour toujours par ce tirage unique. Inutile, c'est-à-dire sans valeur explicative, pour la même raison au fond : car les propriétés que nous pouvons observer dans notre univers sont contenues dans cet unique élément ω_0 , et ne dépendent plus de rien d'autre. De sorte que nous pouvons oublier résolument toutes les richesses qui sommeillaient dans les autres éléments ω , ceux qui n'ont pas été choisis. Du reste, cet élément ω_0 , le nôtre, avait sans doute une probabilité nulle d'être tiré, et notre univers était en somme presque impossible : c'est pourtant le seul univers qui nous soit donné, et le seul que nous puissions étudier.

LE CRITERE POPPERIEN DE L'OBJECTIVITE.

Avant d'engager la discussion, il convient de préciser le sens des termes utilisés. Il ne s'agit absolument pas ici d'analyser avec profondeur les rapports du sujet et de l'objet ¹,; mais de donner un critère, aussi simple que possible, qui nous permette de reconnaître à coup sûr si, et dans quelle mesure, un énoncé, un modèle etc... présente un sens objectif, ou doit être considéré comme "purement subjectif". Imaginons le dialogue ² suivant, entre deux personnages A et B :

A : C'est Jupiter qui lance la foudre.

B : Mais non, la foudre se produit lorsque deux nuages porteurs de charges électriques de signes opposés viennent à se rencontrer.

C : Naturellement : c'est là le moyen qu'utilise Jupiter. Lorsqu'il veut lancer la foudre, il fait que ces deux nuages se rencontrent.

ou bien (version d'apparence plus "moderne") :

A : La foudre est due au Hasard.

B : Mais non, etc...

A : Bien sûr, mais c'est justement par hasard que ces deux nuages se rencontrent.

La discussion peut se prolonger indéfiniment. A chaque tentative de réfutation de B, A oppose une objection ad hoc absolument imparable. Il est en effet, parfaitement impossible de prouver que ce n'est pas Jupiter (ou le Hasard etc..) qui lance la foudre. C'est justement à cause de cette impossibilité où nous sommes d'imaginer une expérience, ou une observation etc... dont le résultat éventuel serait susceptible de la réfuter que nous jugeons l'opinion de A "purement subjective", ou dépourvue

1 - Sur l'objectivité, conçue comme le fruit d'une élaboration ou d'une reconstruction du réel par le sujet (épistémique), voir les travaux de J. Piaget, et notamment, Logique et Connaissance Scientifique, déjà citée.

2 - cf. la citation de Sénèque en exergue de ce chapitre.

d'objectivité. De fait, si cette opinion est compatible avec toute chose et avec son contraire, en sens inverse elle ne permet d'en prévoir aucune : ce qui revient à dire qu'elle ne nous apporte aucune information réelle.

En somme, l'objectivité d'un énoncé, d'une opinion etc... est liée à la possibilité d'en contrôler l'exactitude. Selon K.R. Popper ¹, "l'objectivité des énoncés scientifiques réside dans le fait qu'ils peuvent être soumis à des tests intersubjectifs". Il y a lieu, toutefois, de distinguer deux cas assez différents, selon qu'il s'agit d'énoncés singuliers ou universels.

Du type "singulier" relèvent les énoncés du genre "constatation d'un fait". Par exemple : "il pleut (ici, en ce moment)", ou "les troupes alliées ont débarqué le 6 juin 1944 sur les plages de Normandie". J'entends bien qu'il n'existe pas de "données immédiates" ni de "faits bruts", ce que nous désignons ainsi résultant toujours d'une élaboration et d'une construction préalables. Mais, en pratique, (dans la vie de tous les jours, aussi bien que dans la pratique quotidienne du travail scientifique) cette remarque présente peu d'intérêt. Disons simplement que, par hypothèse ou par convention (comme on voudra), nous supposons qu'un certain consensus est réalisé en ce qui concerne le sens des mots et leur adéquation à (ce que nous appelons) la réalité. Deux observateurs présents au même moment et au même endroit se mettront d'accord pour dire s'il pleut (ou non). Deux historiens, ayant accès aux sources, admettront que le débarquement a bien eu lieu à cette date. Ainsi, en ce qui concerne les énoncés singuliers, c'est-à-dire les "constatations" (possibles) concernant un fait (présent, passé ou futur), le critère d'objectivité réside dans le fait qu'une fois réunie toute la documentation nécessaire, un consensus se réalise entre les "gens sensés" concernant la vérité ou la fausseté de l'énoncé en question : il s'agit d'énoncés dé-
cidables, c'est-à-dire dont il serait univoquement possible de

1 - Logique de la découverte scientifique.

décider s'ils sont vrais ou faux, pourvu seulement que l'on dispose des informations voulues. Il se peut qu'un énoncé décidable (comme : "il a plu à Paris le 1er Juillet de l'an 251 avant J.C.") reste, en fait, indécidé, faute d'accès aux sources. Cela ne lui ôte en rien son objectivité. Celle-ci, de même, est indépendante de la question de savoir si, en définitive, l'énoncé se révélera vrai ou faux. Un énoncé objectif peut très bien être faux. Il suffit qu'il soit possible de le déclarer faux.

Viennent ensuite des énoncés du genre universel. On les rencontre, évidemment, dans les sciences dites positives (par exemple "deux corps s'attirent en raison inverse du carré de leur distance") mais, aussi bien, dans la vie quotidienne ("il pleut tous les jours à Londres"). Ils se réfèrent à une classe illimitée, non pas nécessairement infinie, mais d'extension indéfinie : tous les corps existant, ayant existé ou devant exister dans l'univers etc.. Ils affirment que tous les éléments de cette classe vérifient une propriété contrôlable (en principe). On ne peut pas les vérifier (constater qu'ils sont vrais), faute de pouvoir procéder à un inventaire exhaustif de l'univers. Mais on peut les falsifier (montrer qu'ils sont faux) en exhibant des contre-exemples (par exemple, des témoins peuvent nous assurer qu'il n'a pas plu à Londres le 1er Juillet de l'année dernière). Leur objectivité est liée à leur falsifiabilité. En matière scientifique, s'agissant d'un énoncé, d'un modèle, d'une théorie etc... (considérés en tant qu'ils se rapportent à un secteur bien défini du monde que nous appelons réel) nous disons qu'ils ont une signification objective si, et dans la mesure où, il est possible de les soumettre au contrôle d'expériences ou d'observations dont le résultat soit définissable sans équivoque (ce qui veut dire que l'énoncé de ces résultats doit être susceptible de réaliser le consensus des spécialistes). Le plus souvent, un énoncé scientifique est du genre universel, et ne peut par suite faire l'objet d'une vérification (logique) rigoureuse, qui impliquerait la réalisation effective d'une infinité d'observations ou d'expériences. Par contre, il doit toujours être possible de déduire d'un énoncé scientifique général des énoncés plus particuliers (prédiction du résultat d'expériences ou d'observations effectuées dans des

conditions bien définies, susceptibles, eux, d'être confirmés ou infirmés (vérifiés ou falsifiés). S'ils sont confirmés, il n'en résulte pas que l'énoncé général soit vérifié, mais seulement qu'il est "corroboré" (non réfuté). Par contre, si l'un d'eux est infirmé, l'énoncé général est par là même falsifié (réfuté). Autrement dit, l'énoncé général a une signification objective dans la mesure où il est falsifiable, et possède une validité (relative et toujours provisoire) dans la mesure où il a été corroboré, c'est-à-dire a résisté victorieusement à toutes les tentatives de falsification auxquelles il a été soumis jusqu'à présent : et nous lui attribuerons un degré de validité d'autant plus élevé que ces tentatives auront été plus nombreuses et plus sévères. Tel est le critère de falsifiabilité que propose K.R. Popper ¹ comme ligne de démarcation entre énoncés "métaphysiques" et énoncés "empiriques" ou objectifs. C'est ce critère que nous utiliserons.

LES CONCEPTS OPERATOIRES.

La recherche de l'objectivité a depuis longtemps conduit les physiciens (qui peuvent nous servir de guides en la matière) à n'accepter l'usage que des concepts opératoires (au sens de Bridgman), c'est-à-dire, selon J. Ullmo ², les concepts "définis par le procédé régulier et répétable qui permet de les atteindre et de les mesurer". Cela ne signifie pas, ou pas seulement, que leur définition doit être fondée sur des critères permettant d'effectuer des mesures ; mais, plus profondément, que le concept est défini ou constitué par le système même des "relations répétables" qui permettent de le dégager, et par les lois physiques qui résument ce système. Ainsi, la résistance électrique est définie par la loi d'Ohm $V = RI$. Il n'y a donc, dans le concept opératoire, rien de plus (mais aussi bien : rien de moins) qu'un système d'opérations, effectivement réalisables par le physicien, qui se contrôlent et se recoupent mutuellement. Et ce sont justement les invariants que mettent en évidence ces contrôles

1 - Op. cit.

2 - Les concepts physiques, dans Logique et Connaissance scientifique.

et recouvrements mutuels qui constituent les concepts opératoires.

Il en résulte aussitôt que la valeur, et même la signification objective, d'un concept opératoire sont toujours relatives et limitées : relatives à l'échelle et au secteur de la réalité où les opérations qui le constituent ont un sens ; limitées par la précision des mesures qui le définissent. Et il ne s'agit pas seulement de dire que nous ne pouvons jamais avoir qu'une connaissance approximative des "vraies" valeurs qui existeraient par ailleurs, bien que nous ne les connaissions pas. Des concepts parfaitement opératoires à notre échelle, comme la longueur ou la vitesse, s'estompent dans une sorte de flou, et perdent peu à peu toute signification objective au fur et à mesure que nous descendons vers les échelles microscopiques ¹. C'est ici que la distinction nécessaire du modèle et de la réalité prend toute son importance. Car, une fois que les concepts opératoires et les lois physiques qui les sous-tendent ont été rassemblés dans le cadre d'un modèle mathématique, la tentation est grande d'oublier ces limites et, en se fiant aveuglément au formalisme mathématique, de tirer du modèle des déductions qui vont bien au-delà de son domaine de validité objective. Nous avons vu dans le chapitre précédent un exemple de dépassement de ce seuil d'objectivité. L'existence de ce seuil, et la tentation de le franchir, constituent un danger permanent que nous devons tout particulièrement garder présent à l'esprit lorsque nous mettons en oeuvre des modèles probabilistes.

LA SUBJECTIVITE.

La notion de subjectivité, en tant qu'elle désigne les opinions, croyances, sentiments de conviction de tel ou tel individu, nous retiendra moins longtemps. Notons surtout qu'elle ne

1 - La longueur d'une règle est définissable, disons, au dixième de millimètre près, mais non avec 15 décimales exactes : à plus forte raison, la question de savoir si cette longueur s'exprime, en cm., par un nombre rationnel ou irrationnel n'a absolument aucun sens pour un physicien.

constitue en aucune façon le contraire logique de l'objectivité. Les gens dits "raisonnables" ou "sensés" donnent le plus souvent leur assentiment (subjectif) à un énoncé (objectif) bien corroboré, tel que : "quand une pomme se détache de l'arbre, elle tombe à terre et ne s'envole pas vers les étoiles". En ce sens, évidemment, un énoncé probabiliste, dans la mesure où tel individu lui donne son adhésion, peut toujours être dit subjectif, mais cela n'exclut pas a priori son objectivité (une loi objective, comme la loi de l'attraction universelle, dans la mesure où je la crois "vraie", peut, elle aussi, être dite subjective, puisqu'elle représente, en effet, aussi mon opinion personnelle). Bien que les subjectivistes jouent parfois sur les mots, ce n'est évidemment pas ainsi qu'ils l'entendent. Lorsqu'ils disent que "la probabilité" est subjective, cela veut dire "purement subjective" au sens de dépourvue de toute signification objective : sans d'ailleurs que cela implique une connotation d'arbitraire ou de fantaisie (irrationnelle). Ils se réfèrent plutôt au fait (indéniable) que deux individus différents, placés dans la même situation et possédant la même information, peuvent très bien se comporter de manière différente, sans pour autant que leurs comportements soient irrationnels. Simplement leurs goûts, leurs aspirations etc... diffèrent, et aussi leurs appréciations personnelles de la situation (appréciation qui, selon les bayésiens, peut toujours s'explicitier sous forme d'évaluation de la probabilité (subjective) attribuée aux diverses éventualités).

IL N'Y A PAS DE PROBABILITE EN SOI, SEULEMENT DES MODELES PROBABI-
LISTES.

Abordons maintenant notre problème. Récusons d'abord les affirmations définitives concernant l'objectivité ou la subjectivité en soi de "la" probabilité. Le singulier et l'article défini sont, proprement, aberrants. Sur une σ -algèbre (non triviale ¹),

1 - i.e., contenant au moins un événement distinct de l'évènement impossible (vide) et de l'évènement certain (plein).

on peut toujours construire une infinité de probabilités. Et, dans les applications, il existe bien des manières différentes (et raisonnables) de probabiliser un phénomène donné, selon le point de vue adopté, le but poursuivi, etc.. Ensuite, la notion de probabilité est de nature mathématique, et non empirique, et, comme telle, ne relève pas de la notion d'objectivité telle qu'elle a cours dans les sciences positives ¹. On dira, évidemment, que ce qui est en cause ici ce n'est pas la théorie des probabilités en tant que telle, mais son "application à la réalité". Malheureusement personne n'a jamais appliqué, et n'appliquera jamais à la réalité ni la théorie des probabilités, ni aucune autre théorie mathématique. A la réalité, on ne peut "appliquer" que des opérations réelles (physiques, techniques etc..), et non des opérations mathématiques. Ces dernières ne s'appliquent qu'à des modèles mathématiques de la même nature qu'elles-mêmes. Autrement dit c'est toujours, et seulement, à des modèles probabilistes que l'on applique la théorie des probabilités. Et la question qui se pose est d'examiner si ces modèles probabilistes peuvent, ou non, avoir une signification objective. Je ne conteste pas qu'il soit possible ² et même parfois utile, pour un individu, de mettre de l'ordre dans ses idées ou opinions en les représentant sous forme de modèles probabilistes. Mais cela implique-t-il que ces modèles-ci, ou d'autres que nous pourrions former par d'autres voies - sont nécessairement dépourvus d'objectivité?

Pour nous résumer : il n'y a pas de probabilité en soi. Il n'y a que des modèles probabilistes. La seule question qui se pose réellement, dans chaque cas particulier, est celle de savoir si tel modèle probabiliste, en relation avec tel phénomène réel, présente ou non un sens objectif. Comme nous l'avons vu, cela revient à se demander si ce modèle est falsifiable. La pratique montre

1 - Personne, semble-t-il, ne parle d'espace vectoriel subjectif (ou objectif).

2 - La technique de pari utilisée par les subjectivistes pour leurs évaluations de probabilités ne me semble utilisable en pratique que lorsqu'un petit nombre seulement de variables aléatoires (non indépendantes) interviennent simultanément. Leur procédé serait difficile à mettre en oeuvre dans le cas des fonctions aléatoires, où intervient (théoriquement) une infinité de variables (en pratique, souvent, plusieurs milliers).

que la réponse à cette question peut être positive. Il y a, en effet, des cas où tout le monde, à la lumière des résultats expérimentaux, conviendra d'abandonner le modèle probabiliste qui avait été proposé au départ. Mais qu'est-ce qu'un modèle probabiliste ?

LES MODELES PROBABILISTES.

D'une manière générale, un "modèle probabiliste" est un "espace probabilisé" (Ω, \mathcal{A}, P) plus une convention permettant d'établir une correspondance entre les éléments de cet espace et un certain secteur de la réalité. Les éléments ω de Ω figurent les états (considérés comme) possibles pour le ou les phénomènes que nous voulons décrire. Les éléments A de la σ -algèbre \mathcal{A} (appelés traditionnellement "événements", mais au sens d'événements possibles, non nécessairement réalisés) représentent les énoncés (considérés comme) décidables, c'est-à-dire les constatations (virtuelles) susceptibles d'apparaître comme le résultat éventuel des observations ou expériences que nous pouvons (pourrons, ou pourrions) effectuer à propos de ce phénomène. La probabilité P , enfin, est une fonction définie sur \mathcal{A} (attribuant à chaque événement A de \mathcal{A} un nombre $P(A)$, appelé probabilité (numérique) de A), que nous pouvons, au départ, choisir à notre guise, pourvu seulement qu'elle respecte les axiomes. En pratique, deux cas se présenteront souvent : tantôt, nous choisirons, d'entrée de jeu, une probabilité P , unique, bien définie, et nous dirons que le modèle est (entièrement) spécifié ; tantôt, au contraire, nous nous réserverons une certaine marge de manœuvre, en choisissant seulement une famille $P(\lambda, \mu)$ de probabilités dépendant d'un petit nombre de paramètres $\lambda, \mu \dots$. Dans ce dernier cas, nous dirons que nous avons seulement choisi le type du modèle, et remis à plus tard le problème de sa spécification, c'est-à-dire du choix des valeurs numériques qu'il convient d'attribuer aux paramètres $\lambda, \mu \dots$. Dans l'optique des statisticiens "orthodoxes", le choix du type du modèle constitue une "hypothèse", tandis que le problème de la spécification (des valeurs numériques des paramètres) porte le nom d'"inférence statistique" ou "d'estimation" de ces

paramètres. Cette terminologie véhiculant des présupposés implicites ¹ (concernant l'existence réelle, objective, de ces paramètres) nous utiliserons le terme neutre et purement descriptif de choix (du type de modèle, des paramètres etc... : et de fait, en tout état de cause, c'est toujours nous qui les choisissons).

LE MODELE DE L'ALTERNATIVE REPETEE.

Prenons l'exemple classique du type de modèle appelé "alternative répétée". Ce modèle est destiné à décrire une suite (potentiellement) infinie d'expériences ², dont chacune ne peut amener que l'un ou l'autre de deux résultats (appelés, conventionnellement, "succès" ou "échec"). L'exemple traditionnel est la partie de pile ou face. Mais il peut s'agir de tout autre chose, par exemple de la succession, au cours du temps, des jours de pluie et de beau temps. L'espace Ω est constitué de toutes les suites $\omega = (x_1, x_2, \dots)$ où chaque x_n peut être 1 ou 0, selon que la $n^{\text{ème}}$ épreuve est (a été, ou sera) un succès ou un échec. La valeur numérique de x_n est évidemment déterminée une fois que l'on connaît le résultat complet de la suite d'épreuves c'est-à-dire $\omega = (x_1, x_2, \dots)$. Il existe donc une fonction X_n sur Ω , telle que l'on ait, justement, $x_n = X_n(\omega)$: X_n est la variable aléatoire associée, dans le modèle, au résultat de la $n^{\text{ème}}$ épreuve. L'ensemble \mathcal{A} des constatations possibles comprend, évidemment, les résultats éventuels de chacun des essais, soit les événements $\{X_n = 0\}$ et $\{X_n = 1\}$, pour toutes les valeurs de n , et aussi tous les événements que l'on peut déduire des précédents par un nombre fini ou dénombrable d'opérations logiques. Par exemple, l'évènement $\{X_1 = X_2 = \dots = 1\}$, ou $\{\omega = (1, 1, \dots)\}$, c'est-à-dire une infinité de succès consécutifs, est la conjonction

1 - Sur ce point, j'approuve en grande partie la critique sévère que font les "subjectivistes" de cette terminologie. Voir par exemple B. de Finetti, Theory of Probability, J. Wiley 1974, passim (mais je n'en tire pas les mêmes conclusions).

2 - Je ne dis nullement qu'il s'agit de la "même expérience", ce qui (les subjectivistes ont raison sur ce point) n'aurait, à strictement parler, aucune signification objective.

(ou produit) des événements $\{X_n = 1\}$ pour $n = 1, 2, \dots$

Dans ce type de modèle (l'alternative répétée) les variables X_n sont, par définition ¹, des variables indépendantes, et, par définition également, les valeurs numériques des probabilités de succès des différentes épreuves, sont égales, soit

$$P(X_1 = 1) = \dots = P(X_n = 1) = \dots = p$$

Cette valeur commune p constitue l'unique paramètre dont dépend ce type de modèle. Si nous choisissons à l'avance la valeur numérique de p , disons $p = 1/2$ (ce que nous ferions vraisemblablement dans le cas d'une partie de pile ou face) le modèle (mathématique) est entièrement spécifié. Mais nous ne sommes pas obligés de le faire, et nous pouvons attendre d'avoir recueilli une certaine quantité d'informations expérimentales avant d'achever la spécification de notre modèle. Il se peut même (comme nous le verrons dans un autre chapitre) que nous n'ayons pas réellement besoin de procéder à ce choix ultime. L'essentiel, du point de vue méthodologique, est de distinguer soigneusement les deux rôles très différents que nous attribuons à ce même symbole p . D'un côté, p est un paramètre du modèle. A ce titre, il est susceptible, comme le modèle lui-même, de présenter (ou non) une signification objective et, l'assertion " $p = 1/2$ " peut être (ou non) falsifiable (c'est-à-dire objective ou empirique). D'un autre côté, le même symbole figure dans l'égalité $P(X_{10} = 1) = p$. Dans ce cas, on affirme que la probabilité de succès au 10^{ème} coup (de cette partie-ci) a une valeur numérique déterminée, par exemple $p = 1/2$: énoncé singulier et indécidable, donc certainement dépourvu de signification objective comme tout énoncé relatif à la probabilité d'un événement unique ².

1 - Il s'agit de la définition du modèle mathématique, nullement d'une assertion relative au monde physique. Les subjectivistes (voir par exemple B. de Finetti, op. cit) ont raison de souligner que toute assertion relative à l'indépendance des résultats du premier et du deuxième coup (de cette partie-ci, jouée tel jour, avec telle pièce de monnaie, par M. Durand et Dubois) est dépourvue de signification objective.

2 - Ici encore je suis d'accord avec les "subjectivistes".

A LA RECHERCHE D'UN CRITERE D'OBJECTIVITE.

Il reste maintenant à se mettre d'accord ¹ sur un critère d'objectivité. Ce critère est fortement suggéré à la fois par l'intuition et par la terminologie. Les probabilistes (même dans des études purement mathématiques) appellent, en effet, "presque impossibles" les événements de probabilité nulle, et "presque certains" les événements de probabilité égale à 1. Le mot "presque" indique qu'il ne s'agit pas d'une impossibilité ou d'une nécessité logique. Par exemple, dans l'alternative répétée, l'évènement $\{X_n = 1 \text{ pour tout } n\}$ c'est-à-dire la suite infinie de succès sans aucun échec, est logiquement possible, mais de probabilité nulle (pourvu que p soit strictement plus petit que 1), donc, "presque impossible" mais non impossible. De fait, si nous observions une suite infinie (en pratique, disons, quelques milliers) de succès consécutifs, nous choisirions un modèle déterministe du genre "on gagne toujours à ce jeu-là" : modèle tout-à-fait objectif (puisque falsifiable), que nous abandonnerions éventuellement par la suite, si nous observions quelques échecs, mais certainement préférable, en l'état actuel de notre information, à n'importe quel modèle probabiliste (préférable parce que plus simple). Le critère fondant (au sens strict) l'objectivité des modèles probabilistes serait donc celui-ci : nous conviendrons de déclarer le modèle falsifié si un évènement de probabilité nulle (dans le modèle) se produit en fait (dans la réalité). A ce point crucial de l'exposé, plusieurs remarques s'imposent.

a) La première remarque concerne ce que (depuis Cournot) on a pris l'habitude d'appeler la "loi du hasard", loi selon laquelle les événements de probabilité nulle (ou très faible) ne se produisent jamais. A. Lichnerowicz ² l'énonce ainsi : "Les événements

1 - C'est à dessein que j'utilise cette expression, car tout critère d'objectivité repose, en définitive, sur un consensus.

2 - Remarques sur les Mathématiques et la réalité, dans Logique et Connaissances scientifiques, p. 82.

dont la probabilité est assez faible sont expérimentalement impossibles" et commente en ces termes : "...Cette loi ... demeure encore assez mystérieuse, et elle n'est au fond justifiée que par la coïncidence courante des conséquences théoriques du calcul des probabilités avec les faits expérimentaux interprétés, sans qu'il soit encore possible de percer entièrement le secret de cette coïncidence". C'est sans malice que je relève cette naïveté échappée à un grand mathématicien (qui n'est d'ailleurs nullement un praticien du calcul des probabilités), mais pour montrer avec quelle facilité les meilleurs esprits scientifiques ¹ se méprennent dès qu'il est question de hasard et de probabilités. Cette loi n'a rien de mystérieux, en effet, pour la simple raison que ce n'est pas du tout une loi, mais un critère conventionnel - le seul d'ailleurs qui nous permette de falsifier (de rejeter) un modèle probabiliste.

b) Une difficulté (théorique) provient du fait qu'il existe, en général, (dans le modèle) une infinité non dénombrable d'évènements presque impossibles. Dans le cas de l'alternative, ce n'est pas seulement la suite infinie de succès (1,1,...) mais chacune des suites particulières possibles (x_1, x_2, \dots) qui reçoit une probabilité nulle: quel que soit le résultat d'une suite infinie d'expériences, celui-ci était presque impossible. Pourtant, l'un de ces résultats presque impossibles doit nécessairement se réaliser ².

Cela signifie que l'on doit toujours choisir à l'avance (avant de connaître les données, ou, en tout cas, sans se laisser "influencer" par elles), l'évènement (ou les évènements en nombre fini, ou au plus dénombrable) de probabilité nulle ou très faible destiné à servir de "test". Il y a là, sans doute, un certain arbitraire, gênant du point de vue théorique. Il apparaît à l'usage que le consensus s'établit sans trop de peine dans les applications pratiques.

1 - cf. l'exemple de J. Monod dans le chapitre précédent.

2 - La réunion de ces évènements de probabilité nulle a donc une probabilité 1, ce qui n'est pas interdit par les axiomes, puisqu'il s'agit d'une famille non dénombrable. On voit ainsi que la restriction de l'axiome d'additivité aux familles dénombrables d'évènements disjoints constitue une limitation essentielle.

c) Une autre difficulté est la suivante : le plus souvent, les événements presque impossibles du modèle font figure de cas limite, et ne correspondent pas à des constatations expérimentales réellement possibles (on ne peut pas effectuer réellement une infinité de tirages). En pratique, les statisticiens choisiront donc comme test un événement de probabilité ε petite, mais non nulle. Faut-il prendre $\varepsilon = 10^{-2}$, 10^{-6} , 10^{-9} etc...? Il y a là un danger réel d'arbitraire. Nous y reviendrons.

d) Notre critère étant conventionnel, les subjectivistes restent libres de le refuser, ce qu'ils feront certainement avec la logique de fer qui les caractérise toujours. Ici, toutefois, leur logique de fer risque d'entrer en contradiction avec la pratique scientifique normale, et avec le simple bon sens. J'espère ne pas trop déformer leur pensée en leur prêtant l'argumentation suivante : en premier lieu, ils n'utilisent jamais un modèle entièrement spécifié ($p = 1/2$ par exemple) puisque personne n'est jamais absolument certain que la pièce ne soit pas truquée. Ils utiliseront donc le modèle non spécifié, mais en ajoutant cette restriction importante, qui en modifie la nature probabiliste : p , pour eux, ne sera pas un paramètre de valeur inconnue, quoique déterminée, mais une première variable aléatoire à laquelle ils attribueront une loi de probabilité (subjective), tenant compte de l'information dont ils disposent avant de commencer la partie, mais n'exprimant en définitive rien d'autre que leur propre opinion. Par exemple, ils choisiront pour p une loi uniforme sur le segment $(0,1)$ (ou toute autre loi). C'est, alors, seulement à p fixé que les variables X_n sont indépendantes. Mais p n'est pas fixé, et les X_n ne sont donc pas indépendantes dans leur modèle. Supposons maintenant qu'après une partie très longue, disons $n = 10^6$ coups, on ait obtenu un million de succès consécutifs. Compte tenu de cette information nouvelle, ils remplacent la loi initiale ("a priori") qu'ils attribuaient à p avant l'expérience par la loi conditionnelle que prend cette variable lorsque l'on a observé un million de succès. Dans notre exemple, cette loi admet la densité $(n+1)p^n$ et l'espérance $E(p) = (n+1)/(n+2) = 0,999999$ (mais non pas 1). Cela veut dire que, devant parier sur le $(n+1)^{\text{ème}}$ coup, ils se comporteront comme si celui-ci avait une chance sur un million d'être un échec.

Du point de vue pratique, cela est à peu près équivalent à la certitude du succès, mais il subsiste une différence théoriquement irréductible. On ne peut pas réfuter leur modèle, qui est, par construction, compatible avec tous les résultats expérimentaux possibles : ce qui est logique, puisque, justement, ils refusent à ce modèle toute signification objective. Toutefois, en pareil cas, comme nous l'avons remarqué, l'attitude scientifique la plus recommandable consisterait à admettre (provisoirement) le modèle déterministe : "on gagne toujours à ce jeu", et à chercher pourquoi (est-ce que la pièce ne comporterait pas, par exemple, deux côtés "pile", à supposer que pile représente le succès etc...). Ce qui disqualifie ici l'interprétation subjectiviste, ce n'est pas son manque de logique (elle est parfaitement logique) mais simplement son manque d'intérêt ¹.

RECONSTRUCTION OPERATOIRE DES CONCEPTS PROBABILISTES.

Nous avons vu que c'est le caractère opératoire d'un concept qui fonde, en définitive, son objectivité. Cela signifie qu'un concept (mathématiquement bien défini) intervenant dans un modèle (déterministe ou probabiliste) ne pourra pas être déclaré "objectif" avant d'avoir été entièrement redéfini, ou, mieux, reconstruit en termes strictement opératoires : métamorphose radicale, ou refaçonnement en profondeur de sa personnalité, si l'on peut dire, qui, de son état initial de simple concept mathématique le fait accéder au statut de concept physique. Lors de l'examen critique d'un modèle

1 - Poussée à son terme, cette logique conduit à attribuer une probabilité subjective à chacun des énoncés singuliers déduits des lois physiques. Par exemple, on évaluera à un sur mille (ou un million, ou un milliard, comme on voudra) la probabilité (subjective) pour que ces deux corps ci, demain à telle heure, ne s'attirent pas en raison inverse du carré de leur distance : on peut le faire, mais ce n'est pas très intéressant. Ma thèse est simplement que certains modèles probabilistes (pas forcément tous) et certains des concepts qu'ils mettent en jeu (pas tous) possèdent une objectivité du même genre que celle que le consensus général attribue aux concepts et aux lois de la physique.

probabiliste donné, il sera très important de faire le tri : je veux dire distinguer soigneusement les concepts susceptibles d'être ainsi rendus opératoires, et les autres. Les premiers seuls, ainsi que les énoncés, les paramètres etc.. qui leur sont associés pourront être dits objectifs. Les autres (concepts, énoncés, paramètres) resteront purement conventionnels. Ils auront un sens (mathématique) bien défini dans le modèle, mais sans qu'il leur corresponde de contrepartie univoquement constatable dans le phénomène réel. Cela ne nous interdira nullement de les utiliser, mais uniquement à titre heuristique : pour nous suggérer des méthodes ou des algorithmes, auxquels nous n'aurions pas pensé autrement, non pour justifier nos conclusions. Plus précisément, les conclusions qu'ils nous auront suggérées devront être passées au crible de la critique, reformulées en termes opératoires et soumises à des tests objectifs avant d'être (provisoirement) adoptées... La règle, ici, consistera à s'assurer que toute trace de ces concepts ou paramètres conventionnels a disparu du résultat ultime.

C'est bien ainsi que les physiciens procèdent, dans le cas de l'alternative répétée, pour reconstruire le concept (mathématique) de probabilité p et le transformer en un concept physique très différent, celui de fréquence théorique. Leur démarche peut être ainsi décrite : à toute suite (finie) de n épreuves ayant donné les résultats (x_1, x_2, \dots, x_n) nous associons le nombre moyen de succès qu'elle contient, c'est-à-dire le nombre

$$f_n = \frac{1}{n} \sum_{k=1}^n x_k$$

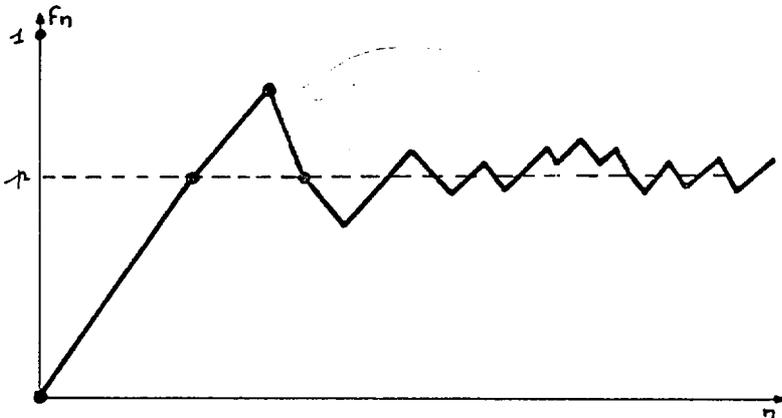
appelé fréquence empirique. Cette fréquence f_n est évidemment une fonction de la suite complète $\omega = (x_1, x_2, \dots)$, donc une variable aléatoire du modèle, admettant (dans le modèle) une loi de probabilité bien définie. Nous pouvons alors considérer la suite infinie (f_1, f_2, \dots) constituée des fréquences empiriques successives. Elle peut être convergente ou non, converger vers p ou vers une autre valeur numérique. L'ensemble des ω pour lesquels la suite f_n converge vers p constitue un événement du modèle (c'est-à-dire appartient

à α), à savoir l'évènement $\{f_n \text{ converge vers } p\}$, ou $\{f_n \rightarrow p\}$.
On démontre (mathématiquement) que cet évènement a (dans le modèle) une probabilité égale à 1 :

$$P(f_n \rightarrow p) = 1$$

Ce théorème, connu sous le nom de "loi forte des grands nombres", nous apprend donc qu'il y a (dans le modèle) une probabilité égale à 1 pour que la suite des fréquences empiriques converge vers le paramètre p .

Cet évènement $\{f_n \rightarrow p\}$ étant presque certain (dans le modèle), nous sommes en droit de le choisir comme critère de falsification. Autrement dit, nous décidons en principe de rejeter le modèle si la suite numérique des fréquences observées dans l'expérience réelle ne converge pas. En pratique, évidemment, nous ne pouvons effectuer qu'un nombre fini N (aussi grand que nous le voulons ou le pouvons, mais il faudra bien s'arrêter) d'expériences. Toutefois, pour un physicien, ce résultat expérimental sera déjà éclairant.



S'il constate que l'amplitude des oscillations, d'abord très fortes, s'atténue de façon plus ou moins régulière, de sorte que, n croissant, la courbe semble bien se stabiliser autour d'une asymptote horizontale d'ordonnée p , il adoptera provisoirement

l'hypothèse " f_n converge vers p " et procèdera à divers contrôles : poursuivre l'expérience au-delà de N ; extraire et étudier séparément des sous-suites (par exemple, la suite impaire (f_3, f_5, \dots) et la suite paire (f_2, f_4, \dots) ; recommencer d'autres¹ séries

1 - Cela suppose, en principe, un élargissement du modèle initial, qui doit maintenant contenir des variables à 2 indices $X_{n,m}$ (représentant le $n^{\text{ème}}$ résultat de la $m^{\text{ème}}$ expérience), mais ce point ne soulève pas de difficultés.

d'expériences etc... Ces divers contrôles et recoupements, s'ils s'accordent pour mettre en évidence l'existence de la même limite asymptotique p , confèrent par là même à celle-ci le statut de concept opératoire : le concept mathématique est maintenant remplacé par le concept physique ainsi construit, auquel on peut donner le nom de fréquence théorique. A ce concept, nous associons la valeur numérique de la limite asymptotique ainsi dégagée par le procédé même qui nous a servi à construire le concept, et nous désignons encore par p le résultat de cette mesure. En fait, p n'est connue qu'avec une certaine approximation. Mais c'est là le sort de tous les concepts et de toutes les mesures physiques. Nous pouvons d'ailleurs, en principe, améliorer cette approximation, si nous le souhaitons, en procédant à de nouvelles expériences. De plus, la rapidité avec laquelle la suite f_n converge vers p fait (dans le modèle) l'objet de théorèmes précis, dont les énoncés peuvent, à leur tour, être reconstruits en termes opératoires et nous aider à apprécier l'ordre de grandeur de l'approximation obtenue.

D'autres concepts encore peuvent bénéficier de cette métamorphose : la stationnarité par exemple. Nous ne pouvons pas "vérifier" que $P(X_n = 1)$ est indépendant de n , puisque c'est là un énoncé qui met en jeu des probabilités d'évènements individuels. Mais, si nous avons effectué $N = 10000$ essais, par exemple, nous pouvons découper cette suite en 100 tranches de 100 essais consécutifs, examiner si chacune des 100 fréquences correspondantes est, ou non, voisine de p , comparer la distribution empirique de ces 100 fréquences avec celle que laisse prévoir le modèle, faire des tests statistiques précis (suggérés par le modèle) etc... A nouveau, tout cet ensemble de contrôle et de recoupements donne naissance à un concept opératoire, celui de stationnarité physique.

De même pour l'indépendance. Nous ne pourrions jamais "vérifier" que X_3 et X_4 sont indépendantes (ce n'est pas là un énoncé objectif). Mais, de même que nous avons construit le concept opératoire de fréquence (d'une suite), nous pouvons construire celui de fréquence de deux succès consécutifs, ou d'un succès suivi d'un échec etc... Nous pouvons ainsi mettre en évidence une loi physique, selon laquelle la fréquence de deux évènements consécutifs est égale au

produit des fréquences de chacun d'eux, loi physique qui, à son tour, sert de définition opératoire au concept physique d'indépendance d'ordre 2. On peut, de même, donner un sens physique à l'indépendance d'ordre 3 ou 4 (mais, comme on l'a vu, on ne pourra pas aller très loin, et l'indépendance d'ordre 200, par exemple, restera inaccessible).

Dans ce cas (idéal), nous avons réussi, à partir du modèle mathématique initial, à reconstruire un modèle physique entièrement objectif. La transposition n'est pas, et ne peut pas être totale. Rien ne nous garantirait jamais (à supposer que cela ait un sens) que les essais 13 et 14 étaient indépendants et avaient la même probabilité que les autres. Dans une partie de pile ou face, il se pourrait que pour ces deux essais-là, justement, à l'exclusion de tous les autres, on ait substitué une pièce truquée à la pièce initiale : l'analyse statistique ne nous permettra jamais de nous en apercevoir. Mais l'essentiel, je veux dire le contenu objectif de notre modèle, a été sauvegardé.

L'HYPOTHESE ANTICIPATRICE ET LE RISQUE D'ERREUR RADICALE.

Nous nous sommes donnés la partie belle, dans cet exemple, en supposant que nous pouvions, en principe, réaliser autant d'expériences que nous le voulions. C'est bien ainsi, du reste, que procèdent, (ou du moins raisonnent) les physiciens, nos maîtres en la matière. Mais en pratique les choses se passent souvent de manière bien différente. On est souvent obligé de prendre immédiatement une décision sur la base d'une information limitée, que l'on n'a pas le temps ou pas les moyens de compléter. Par exemple, supposons que l'on ait, dans l'exemple précédent, procédé à 20 expériences ayant donné les résultats (S) suivants :

(S) : 11001001000011111101

On nous demande, sur la base de ces 20 seules données, de faire une prévision concernant la fréquence moyenne d'apparition

du chiffre 1 dans les 10000 essais qui suivront, et, si possible, de donner l'ordre de grandeur de l'erreur possible.

Le statisticien "orthodoxe" avancera l'"hypothèse"¹ qu'il "s'agit" d'une alternative répétée avec un paramètre p inconnu (dans ce cas il "estimera" p au moyen de la fréquence $f = 11/20 = 0.55$) ou bien (s'il y a par ailleurs des indications en ce sens) avec $p = 1/2$ par exemple. Puis il procédera à des "tests" qui donneront d'ailleurs des résultats positifs, ou plutôt ne donneront pas de résultats négatifs, concernant l'indépendance (d'ordre 2), la valeur $p = 1/2$ etc... S'il a choisi le modèle spécifié $p = 1/2$, il annoncera donc, pour les 10000 essais à venir, un nombre de succès de 5000 ± 100 avec une "probabilité d'erreur" de 5% (correspondant à 2 écarts types). Le subjectiviste, d'ailleurs, aboutirait à une prévision sensiblement équivalente, bien qu'énoncée dans un langage différent.

Naturellement, cette prévision peut se révéler radicalement fausse. Il peut arriver, par exemple, que, le phénomène changeant de nature pour une raison que nous ne pouvions pas prévoir, il n'y ait plus que des 0 à partir du millième essai. Le statisticien a donc pris un risque (inévitabile). Mais, s'il y a risque, c'est-à-dire si sa prévision peut être démentie par la suite, cela signifie que cette prévision était falsifiable, donc avait un sens objectif. Le statisticien avait donc réellement avancé une hypothèse objective (falsifiable) : non pas exactement celle qu'il avait énoncée, qui ne concernait que le modèle mathématique, mais, implicitement, une hypothèse anticipatrice relative à la validité du modèle physique, dont nous avons expliqué ci-dessus la construction opératoire. Il s'agit bien d'une hypothèse objective (puisqu'elle peut se révéler fausse après coup), et d'une anticipation (puisque les tests les plus rigoureux effectués sur les 20 données, à supposer qu'ils corroborent la validité du modèle pour ces 20 données là, ne nous garantissent en aucune façon que le phénomène

1 - Dans notre langage, il ne s'agit pas ici d'une hypothèse, mais du choix d'un type de modèle et, le cas échéant, de sa spécification $p = 1/2$. Mais ce choix implique bien, comme nous le verrons, une hypothèse objective (falsifiable).

ne changera pas de nature par la suite). C'est parce qu'elle est objective (introduit une information supplémentaire, non contenue dans les 20 données) que cette hypothèse nous permet de tirer de ces données plus qu'il n'y est contenu, et d'avancer une prévision. C'est parce qu'elle est anticipatrice (adoptée avant que sa validité n'ait été contrôlée) qu'elle introduit un risque d'erreur radicale¹ : et ce risque est la contrepartie obligée du gain d'information qu'elle introduit. Nous reviendrons longuement sur ce point par la suite.

D'autres éventualités inattendues peuvent se manifester. Supposons que nous ayons l'idée (bizarre) de placer une virgule après les deux premiers chiffres 1, et de lire la suite (S) comme l'écriture binaire d'un nombre $x < 4$: en transcrivant x en écriture décimale, nous trouvons, puisqu'il y a 20 symboles binaires, que x est compris entre 3,141590 et 3,141594. Devons-nous conclure que, (pour une raison inconnue) on a bien $x = \pi$, et avancer comme une prévision "déterministe" que les 10000 essais à venir correspondront aux symboles suivants de l'écriture binaire du nombre π ? C'est là une hypothèse anticipatrice beaucoup plus forte que la précédente (plus facilement falsifiable) mais de même nature et, chose curieuse, nullement incompatible avec elle².

MODELES PANSCOPIQUES ET MODELES MONOSCOPIQUES.

Cette idée d'une hypothèse anticipatrice que nous sommes obligés d'avancer, et du risque d'erreur radicale qu'elle entraîne obligatoirement, nous fait pénétrer au coeur du débat qui oppose "objec-

1 - Dont l'amplitude est d'un autre ordre de grandeur que la "fourche d'erreur à 5%" de ± 100 que prévoit le modèle. C'est pourquoi je parle d'erreur radicale.

2 - En ce sens que les tests usuels, appliqués à la suite des n premiers symboles de l'écriture binaire du nombre π , ne conduiraient vraisemblablement pas au rejet du modèle "alternative répétée". L'évènement lui-même " $(x_1, x_2, \dots, x_n) = n$ premiers symboles binaires de π " a, dans le modèle, la probabilité $1/2^n$, très faible déjà pour $n = 20$ (un millionième), mais ne peut pas, en principe, servir de critère, puisque son choix a été suggéré par les données. De fait, chaque suite de 20 symboles peut être considérée comme le début de l'écriture d'un nombre plus ou moins remarquable (multiples entiers et puissances de π etc... : il suffit d'en trouver un million).

tivistes et subjectivistes". Derrière les querelles de mots, les passions sectaires¹ et les terminologies plus ou moins inadéquates, on devine une opposition plus fondamentale : celle, en gros, qui sépare l'idéal de la "connaissance scientifique", et les nécessités de la "vie réelle" ou de la "pratique". Le scientifique cherche à construire des modèles aussi riches, aussi compréhensifs et aussi bien corroborés qu'il se peut. Il admet que l'on a (au moins en principe) tout à la fois la possibilité et le loisir de réunir toute la documentation nécessaire, de procéder à tous les contrôles souhaitables avant de porter un jugement sur la validité objective de chacune de ses hypothèses. Le praticien, lui, affronte des situations d'urgence. Dans le feu de l'action, en effet, on ne dispose le plus souvent que d'une information fragmentaire, de qualité parfois douteuse, sur la base de laquelle on doit pourtant, impérativement, prendre une décision importante. Il en va ainsi dans le domaine de la vie sociale, économique ou politique, aussi bien que dans la vie professionnelle ou privée de chacun de nous. En outrant le tableau, on opposera le caractère objectif et désintéressé de la connaissance scientifique à la subjectivité et à la rapacité sauvages qui se déchaînent dans le domaine de l'action pratique ; ou, au contraire, on exaltera la vie de l'homme d'action, au détriment du chercheur, accusé de fuir la vie réelle et ses responsabilités. On peut aussi, plus raisonnablement peut-être, penser qu'il s'agit là d'une opposition de degré plus que de nature. Car l'information dont dispose le scientifique n'est jamais parfaite, et lui aussi, par conséquent, doit avancer, sous sa responsabilité, des hypothèses dont la validité, si bien corroborée soit-elle aujourd'hui, peut être à chaque instant remise en question par le résultat inattendu d'une expérience nouvelle. Et, de son côté, le praticien n'est jamais démuné au point que les hypothèses antici-

1 - On rencontre, chez certains subjectivistes, des proclamations fracassantes, concernant par exemple le 21^{ème} siècle, qui sera parait-il un siècle "bayésien". Même un auteur s'exprimant généralement de manière fort sensée et courtoise, comme B. de Finetti, laisse parfois tomber des ukases définitifs. "Speaking of unknown probabilities must be forbidden as meaningless", lisons-nous par exemple dans sa "Theory of Probability" (J. Wiley and Sons, 1974) p. 190. Pourquoi ce bruit de bottes dans la république des savants ?

patrices (explicites ou implicites) sur lesquelles il fonde ses décisions soient totalement dépourvues de valeur objective : sinon, le jeu impitoyable de la sélection naturelle le ferait rapidement disparaître de la scène de l'action pratique.

Il n'en existe pas moins une opposition réelle, entre les objectifs que l'on vise, et les critères que l'on utilise dans ces deux types d'activités. En ce qui concerne les objectifs, d'abord, on peut remarquer que nos sciences et nos techniques semblent hésiter, ou osciller, entre un mode spéculatif ou explicatif, et un mode instrumental ou manipulateur. Le premier, réaliste et "désintéressé"¹, prétend viser la connaissance pour elle-même, comprendre ou expliquer l'objet tel qu'il existerait indépendamment de nous et des applications pratiques que nous pourrions en faire : non qu'il méprise ces applications, mais elles devraient être, en quelque sorte, "données par surcroît", et il ne s'y intéresse vraiment que dans la mesure où elles viennent illustrer et corroborer ses modèles explicatifs. Le second mode, au contraire, veut transformer le monde, plutôt que l'expliquer, manipuler et domestiquer l'objet, sans s'intéresser à ce qu'il est en soi. Il est nominaliste : le modèle n'est pas la réalité, et résolument instrumentaliste : le meilleur modèle est le plus efficace. Cette opposition n'est d'ailleurs pas irréductible : la manipulation n'est efficace que dans la mesure où le modèle utilisé est, d'une manière ou d'une autre, adapté à la réalité, ou à l'aspect de la réalité, auquel nous nous intéressons ; et les théories les plus désintéressées ne sont scientifiques que dans la mesure où elles restent opératoires : c'est, en définitive, à l'efficacité des prévisions qu'elles permettent que l'on juge le degré de leur validité. Mais, si nous prenons en considération le nombre, l'ampleur et la généralité des objectifs poursuivis, nous pouvons dire, en gros, que les modèles ou théories scientifiques se veulent panscopiques² alors que les modèles pratiques, ou techniques, se contentent, plus

1 - "L'action désintéressée est, en réalité, très intéressante, et intéressée, étant admis que..." remarque Nietzsche (Par delà le bien et le mal).

2 - Panscopique : tous les buts ; monoscopique : un seul but.

modestement, d'être polyscopiques, ou même strictement monoscopiques.

En effet, un modèle ou une théorie scientifique, pour être valable, doit, par définition, permettre de prévoir correctement le résultat de toute expérience, de toute observation, que nos moyens techniques et nos connaissances actuelles nous permettent d'imaginer et de réaliser effectivement (du moins dans le domaine où la théorie est opératoire). Et elle doit donc permettre aussi de résoudre efficacement tous les problèmes pratiques, ayant un sens actuel, que nous sommes ou serions capables de nous poser dans ce domaine. Ainsi, le critère de validité est l'efficacité opératoire à l'égard de tous les objectifs envisageables. Nous dirons donc que les théories scientifiques ont un caractère panscopique.

Mais pour atteindre cet idéal scientifique panscopique, il faut du temps, de la réflexion, des moyens, et une masse énorme d'information. Dans les activités pratiques, cet idéal apparaîtra souvent comme un luxe que, dans le feu de l'action, on n'a ni le temps, ni les moyens de s'offrir. D'où le caractère plus ou moins monoscopique des modèles pratiques, qui résulte en quelque sorte d'un principe d'économie. De fait, dans les activités techniques, économiques, etc... l'homme est sans cesse confronté à des problèmes pratiques qui doivent impérativement recevoir une solution (aussi bonne que possible, mais il faut en choisir une), à des décisions qui doivent impérativement être prises sur la base d'une information insuffisante. Dans ces conditions, il n'a que rarement le loisir (ni même le goût) de s'interroger longuement sur les caractéristiques objectives de la situation qui ne concernent pas directement le problème à résoudre, ou la décision à prendre. Ce qu'il attendra donc, avant tout, d'un modèle (probabiliste ou non) c'est une efficacité opératoire réelle dans le seul domaine qui l'intéresse, et peu lui importera que le modèle soit grossièrement faux relativement à d'autres aspects de la situation qui ne l'intéressent pas directement. Son objectif est monoscopique, et cela essentiellement pour des raisons d'urgence et de choix de priorité.

Les critères de validité ne sont évidemment pas les mêmes pour les deux sortes de modèles. Une théorie scientifique (pan-scopique) se trouve automatiquement réfutée si l'une quelconque des conséquences que l'on peut en déduire (dans le domaine où elle est opératoire) est réfutée par l'expérience. Le modèle monoscopique, lui, doit être jugé d'après son adéquation au seul but qu'il poursuit. Mais, corrélativement, on ne doit pas, non plus, attendre de lui des réponses sensées aux questions qu'il n'était pas conçu pour résoudre.

De plus, contrairement aux théories scientifiques, qui présentent toujours un degré plus ou moins élevé d'universalité, le modèle monoscopique concerne le plus souvent un phénomène particulier, une situation unique, que l'on ne retrouvera jamais plus identique à elle-même. La condition de répétabilité, sans laquelle il est difficile de fonder des définitions opératoires et de justifier la validité objective d'un modèle, semble donc ici faire défaut. Si l'on ajoute à cela que le modèle monoscopique est choisi dans le feu de l'action, c'est-à-dire sur la base d'une information insuffisante par définition, et constitue donc toujours une hypothèse anticipatrice, qui pourra éventuellement être contrôlée "après coup", mais dont rien, en toute rigueur, ne permet d'affirmer la validité au moment où on l'adopte, on conçoit que l'on puisse s'interroger sur son statut épistémologique, et même comme les Bayésiens, sur sa signification objective.

Nous reviendrons longuement sur l'importante question de la signification et de la valeur objective des modèles probabilistes monoscopiques. Le point important à retenir, pour l'instant, c'est qu'au moment où on le choisit, ce modèle monoscopique introduit une hypothèse anticipatrice dont la légitimité ne peut en aucune façon être garantie par sa compatibilité avec les données numériques disponibles : car cette hypothèse revient, en somme, justement, à admettre que les caractéristiques structurelles que nous avons induites à partir de ces données peuvent être extrapolées telles quelles aux parties inconnues du phénomène ; ou encore, si l'on veut, que le phénomène se comporte, là où on ne le connaît pas, d'une manière suffisamment analogue à ce que l'on a observé là

où il est connu. Ceci implique deux conséquences : pour choisir une hypothèse de ce genre, il faut soigneusement tenir compte de toutes les sources d'information, numérique ou non, dont on dispose (connaissances générales sur la physique de ce phénomène, expérience acquise sur des cas analogues, etc...) ; d'autre part affaiblir au maximum cette hypothèse, de manière à la réduire au strict minimum indispensable pour permettre d'atteindre l'objectif visé par le modèle monoscopique : valorisation maximale de toutes les sources d'information, et principe d'économie stricte dans le choix des hypothèses anticipatrices.

LES CRITERES EXTERNES OU L'OBJECTIVITE D'UNE METHODOLOGIE.

Lorsqu'il s'agit d'un modèle donné (le plus souvent monoscopique), destiné à représenter un phénomène unique, il est parfois difficile de justifier rigoureusement la signification et la valeur objective de ce modèle particulier. Nous verrons pourtant plus loin qu'il existe des critères internes d'objectivité susceptibles d'apporter une réponse, au moins partielle, même dans le cas d'un phénomène unique. Mais, s'il n'est pas évident que l'on puisse porter un jugement sur chaque cas individuel, il n'en est certainement pas de même de la méthodologie générale que nous utilisons pour choisir des modèles monoscopiques dans chaque cas particulier : elle fera, à la longue, la preuve de sa plus ou moins grande efficacité. En effet, chaque situation, chaque problème est unique et fait l'objet d'un modèle monoscopique ad hoc. Mais il y a des classes de situations et de problèmes, non identiques, mais suffisamment analogues pour que les règles qui dictent le choix du modèle que nous adoptons dans chaque cas puissent être, au moins partiellement, formalisées, et finissent par constituer un système méthodologique : ce système sera soumis à la sanction de la pratique, et devra faire ses preuves ou être abandonné.

En d'autres termes, il est exact que c'est, en définitive, la possibilité de répétition qui fonde l'objectivité (la "relation

répétable" de J. Ullmo¹. Mais cela ne signifie pas qu'il n'y ait pas de science possible de l'unique. D'abord, en effet, c'est toujours en un sens relatif que l'on parle de refaire la "même" expérience. A strictement parler, il n'y a pas deux expériences identiques : elles diffèrent toujours l'une de l'autre par quelques facteurs accessoires (mais c'est nous qui les jugeons tels), et par les conditions de lieu ou de temps. Tout ce que l'on peut dire, c'est que les facteurs qui nous paraissent importants, ont été rendus aussi semblables que nous le permettent nos moyens techniques, et les autres sont ce qu'ils sont. De même, on ne peut pas parler d'un phénomène qui se reproduit, mais seulement d'une classe de phénomènes que nous jugeons suffisamment proches les uns des autres pour les considérer comme équivalents. Cette proximité, ou ressemblance, peut d'ailleurs n'être que qualitative et structurelle, sans aller jusqu'à l'égalité des paramètres numériques descriptifs. En géologie ou en astronomie, par exemple, il n'y a pas deux objets identiques, ce qui n'empêche nullement de fonder l'objectivité sur la répétition du semblable.

De la même manière, les soi-disant "uniques" sont uniques en tant qu'objets ou situations individuelles, mais se laissent classer en espèces ou genres regroupant objets et situations qualitative-ment et structurellement analogues. C'est à l'intérieur de ces classes (à nous de les définir de manière aussi précise que possible, car cette définition jouera un rôle constitutif dans notre discipline!) qu'opère la répétition qui fonde l'objectivité. Dans la polémique qui les oppose sur l'interprétation qu'il convient de donner à la "probabilité", les fréquentistes auraient donc raison contre les subjectivistes, mais en un sens large seulement : c'est la comparaison systématique de la prévision probabiliste et de la réalisation, portant sur une classe d'épreuves éventuellement non identiques, mais appartenant à un même genre, défini de manière plus ou moins large, qui fonde à la longue l'objectivité, non pas peut-être d'une prévision singulière, mais au moins celle de la

1 - Les concepts physiques, in op. cit.

méthodologie qui y a conduit. Si cette prévision ne représentait vraiment rien d'autre que l'état d'âme du praticien ou du preneur de décisions, cela n'intéresserait personne, même pas le preneur de décisions lui-même : car ce qu'il veut surtout, c'est que sa décision soit la plus efficace possible, et cette efficacité dépend évidemment aussi de la situation objective.

Ce critère externe d'objectivité revient donc, en somme, à examiner si l'on a, ou non, "raison en moyenne" d'utiliser telle méthodologie pour tenter de résoudre telle catégorie de problème.

Pour essayer de formuler cela de manière un tout petit peu plus technique, remarquons d'abord que, pour utiliser le modèle de l'alternative répétée, ou les modèles analogues du type¹ "suite de variables aléatoires indépendantes et identiquement distribuées" nous n'avons, à aucun moment, supposé qu'il s'agissait (dans la réalité) de la même expérience indéfiniment répétée. Je n'affirmerais pas de manière aussi catégorique que les subjectivistes qu'une telle expression ne puisse, en aucun cas, présenter la moindre signification objective. Car, après tout, les physiciens me semblent gens à savoir ce qu'ils disent (et ce qu'ils font). Mais, dans le domaine qui est le nôtre, les difficultés seraient très réelles. Il se trouve, heureusement, que nous n'avons en aucune façon besoin d'une hypothèse de ce genre pour fonder notre critère

1 - La classe la plus générale des modèles de ce genre est donnée par un espace probabilisé (Ω, \mathcal{A}, P) défini comme le produit d'une suite infinie d'espaces $(\Omega_n, \mathcal{A}_n, P_n)$ identiques (en langage rigoureux : isomorphes). Chaque espace facteur décrit un phénomène ou une expérience donnée, l'espace produit correspond à l'ensemble de ces phénomènes ou expériences. Du fait que la probabilité P est définie comme le produit des probabilités P_n , les variables X_n, Y_n etc... décrivant la $n^{\text{ème}}$ expérience sont, par construction, indépendantes (dans le modèle) des variables attachées à toutes les autres expériences. Cette classe de modèles constitue la transposition mathématique de la notion physique de répétabilité. On note le caractère formel, ou conventionnel, de l'espace Ω (défini simplement, en somme, comme l'ensemble de tous les résultats possibles de toutes les expériences envisageables) et son caractère modulaire (au sens d'extensible à volonté) : on peut toujours l'enrichir à volonté en lui adjoignant de nouveaux facteurs isomorphes $(\Omega_n, \mathcal{A}_n, P_n)$ chaque fois que le besoin se fait sentir de prendre en compte une nouvelle suite d'expériences possibles. Ceci contraste avec le caractère concret de l'espace Ω des modèles pour lesquels il existe (comme en physique) des critères internes d'objectivité.

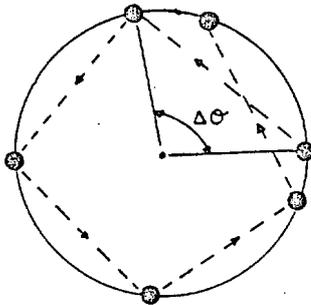
d'objectivité externe (méthodologique). Il peut s'agir, en principe, de n'importe quelles expériences, pourvu seulement que l'on sache, ou puisse, les identifier. Naturellement, en matière scientifique non plus, il n'est pas interdit d'avoir du bon sens, et on n'introduira pas volontiers dans le même modèle les élections françaises de 1978, les gisements de pétrole du Proche Orient, l'âge du capitaine et la chèvre de Monsieur Seguin. Mais il peut s'agir d'expériences réellement très différentes.

Le mieux est de prendre un exemple. Il existe des types étonnamment variés de gisements de Cuivre, petits et gros, très riches ou à teneur moyenne très basse, de structure simple ou compliquée, à teneurs faiblement ou fortement dispersées, etc... D'autre part, l'information dont on dispose pour procéder à leur estimation est, elle aussi, extrêmement variable en nature, quantité et qualité. Passons sur les détails techniques, et supposons univoquement définie la notion de ressource ou quantité Q , exprimée en milliers de tonnes, de cuivre contenu dans un gisement donné. Cette quantité ne sera connue qu'après l'exploitation effective. Supposons (ce qui est le cas) que nous disposions d'une méthodologie nous permettant d'intégrer, dans un modèle probabiliste, toute l'information disponible au moment où l'on doit estimer ce gisement (donc avant sa mise en exploitation). Dans ce modèle, il correspond à Q une variable aléatoire que nous appellerons encore Q , et notre modèle contient l'énoncé : "il y a une chance sur deux pour que Q soit supérieur à q ". Ici, q est une valeur numérique que nous sommes capables de calculer (dans le modèle, q est la médiane de la variable Q prise conditionnellement à l'information disponible). Peu nous importe, pour le moment, que cet énoncé individuel ait ou non un sens objectif. Si nous avons ainsi estimé une centaine de gisements, numérotés de $n = 1$ à $n = 100$, nous pouvons comparer, pour chacun d'eux, ses ressources Q_n , maintenant connues, et la médiane q_n que notre méthodologie avait attribuée à la variable correspondante du modèle. Posant $X_n = 1$ si $Q_n \geq q_n$ et $X_n = 0$ si $Q_n < q_n$, nous sommes en présence d'une suite de succès et d'échecs à laquelle nous associons le modèle (spécifié) de l'alternative répétée avec $p = 1/2$: nous pouvons alors tester ce modèle (plus précisément,

contrôler la validité du modèle physique que ce modèle probabiliste permet de reconstruire). Si l'on juge que 100 épreuves ne suffisent pas (en pratique, je veux dire : aux yeux des praticiens de l'industrie minière, elles suffiront certainement), on peut attendre d'en avoir 200 ou 1000. mais nous sommes (ou serons) en mesure de porter un jugement sur la valeur de notre méthodologie. Celle-ci est donc objective, puisqu'elle peut être, et sera, disqualifiée si elle conduit, en moyenne, à des erreurs notables de prévision¹.

CRITERES D'OBJECTIVITE INTERNE, LIES AU CARACTERE CONCRET DE L'ESPACE Ω .

D'après le texte déjà cité² de J. Ullmo, l'"idée de hasard" c'est-à-dire, en réalité, l'usage des modèles probabilistes, s'introduit en physique "lorsque des conditions initiales inséparables expérimentalement sont suivies ultérieurement d'une séparation manifeste des phénomènes observés". Le mieux que nous puissions faire ici est d'illustrer cette conception profonde par un exemple simple. Considérons le cas (idéal) d'un billard circulaire. La boule,



initialement lancée sans effet de rotation, rebondit sur la bande circulaire en respectant les lois de la réflexion, de sorte que, d'un choc au suivant, l'angle θ qui fixe sa position augmente d'une quantité (en principe) constante $\Delta\theta$. Nous admettons de plus que les pertes d'énergie peuvent être négligées, de sorte que la boule circule indéfiniment sur le billard. Introduisons une notion physique simple, celle de fréquence de visite $f_n(\ell)$ d'un arc

1 - On peut aller plus loin. A chaque Q_n notre méthodologie associe la loi de probabilité F_n de la variable qui lui correspond dans le modèle. On sait que, dans le modèle, $F_n(Q_n)$ est alors une variable uniformément distribuée sur $(0,1)$. On peut donc, sur la suite $X_n = F_n(Q_n)$ des valeurs numériques obtenues, tester la validité du modèle "suite de variable indépendantes et uniformément distribuées sur $(0,1)$ ".

2 - Les concepts physiques, dans Logique et Connaissance scientifique, par J. Piaget et al., Paris, 1967, p. 649.

(quelconque) de longueur ℓ comptée en radians sur la bande, au cours de n rebondissements consécutifs. Autrement dit, $f_n(\ell) = k/n$, où k est le nombre des occasions où le rebondissement s'est produit en un point de l'arc ℓ . On démontre, alors, le théorème suivant : lorsque n tend vers l'infini, la fréquence de visite tend vers le rapport $\ell/2\pi$, pourvu seulement que $\Delta\theta/2\pi$ ne soit pas un nombre rationnel. Si, par contre, $\Delta\theta/2\pi$ est un nombre rationnel, la boule repassera indéfiniment par les mêmes points, en nombre fini, et les autres ne seront jamais visités. Ce théorème suggère irrésistiblement que la loi de probabilité uniforme sur le cercle doit, d'une manière ou d'une autre, intervenir dans la description du comportement de la boule, à l'exception possible (importante) des cas où $\Delta\theta/2\pi$ serait "par hasard" un nombre rationnel. Nous avons le sentiment intuitif que ces cas d'exception doivent être très peu nombreux, ou même, en pratique, ne se produire jamais... Si nous choisissons, "au hasard", c'est-à-dire avec une densité uniforme, un arc $\Delta\theta$ compris entre 0 et 2π , il y a, en effet, une probabilité nulle pour que $\Delta\theta/2\pi$ soit un nombre rationnel. Pour un physicien, s'agissant d'un phénomène macroscopique, cette éventualité est tout simplement exclue.

Partant de la position θ_0 (supposée, pour simplifier, exactement connue), la boule, après n rebondissements, occupe la position $\theta_n = \theta_0 + n \Delta\theta$. Si précise que soit notre mesure initiale de $\Delta\theta$, connue avec une erreur de l'ordre d'un nombre ε très petit, cette erreur va s'amplifier démesurément, puisque nous ne pouvons prévoir θ_n qu'à $\pm n \varepsilon$ près. Dès que n est assez grand, cette erreur $n \varepsilon$ dépasse 2π , où un multiple donné, mettons 10 fois 2π : ce qui revient à dire que le $n^{\text{ème}}$ rebondissement aura lieu n'importe où sur la bande. Les conditions initialement inséparables de J. Ullmo ont fini par être séparées. Ceci suggère un modèle probabiliste (que l'on pourra contrôler en répétant l'expérience) dans lequel θ_n serait représentée par une variable uniformément distribuée sur le cercle unité.

Indépendamment même de tout contrôle expérimental, le choix de cette loi uniforme est univoquement imposé par la nature physique de ce modèle. De fait, supposons que nous choissions pour θ_n

une autre loi que la loi uniforme (pourvu seulement qu'elle admette une densité¹). On démontre alors sans peine que, pour n' assez grand, la loi de $\theta_{n+n'}$ diffère aussi peu que l'on veut de la loi uniforme. Il suffit donc de remplacer n par $n+n'$ pour obtenir, à nouveau, la même loi : ce phénomène de convergence vers une loi limite univoquement imposée est une manifestation de la propriété connue sous le nom d'ergodicité.

La manière la plus simple de définir le modèle probabiliste consiste à faire correspondre à ε une variable aléatoire ayant une loi quelconque, pourvu toujours qu'elle admette une densité. Alors $\Delta\theta$ est de la forme $\Delta\theta = \Delta\theta_0 + \varepsilon$, avec un $\Delta\theta_0$ numériquement connu, et, dans le modèle $\theta_n = \theta_0 + n \Delta\theta_0 + n \varepsilon$ est une variable aléatoire dont la loi, à nouveau, converge vers la loi uniforme. L'intérêt de cette formulation est que, maintenant, les cas d'exception ($\Delta\theta/2\pi$ rationnel) ont, dans le modèle, une probabilité nulle (quelle que soit la loi de ε admettant une densité).

On peut encore améliorer le modèle en remarquant que, le plan du billard et sa bande présentant forcément des petites irrégularités, les accroissements successifs $\Delta\theta_n = \theta_{n+1} - \theta_n$ ne pourront pas être rigoureusement égaux. Même si, par hasard, l'un d'eux était un multiple rationnel de 2π , ils ne le seraient pas tous. De sorte que les cas d'exception sont, effectivement, exclus par cette considération physique. En prenant cette fois $\Delta\theta_n = \Delta\theta + \varepsilon_n$, avec (dans le modèle) des variables aléatoires ε_n quelconques (non nécessairement indépendantes, il suffit qu'elles ne soient pas trop corrélées), on démontre alors que $\theta_n, \theta_{2n}, \theta_{3n}$ etc... pour n assez grand sont, dans le modèle, des variables indépendantes, admettant toujours cette même loi uniforme. Et la valeur de ce modèle peut, maintenant, être contrôlée expérimentalement, sans même qu'il soit nécessaire de répéter l'expérience.

1 - Du point de vue physique, cette restriction n'est pas gênante : car pour attribuer un atome bien localisé à cette loi, il faudrait avoir une information extraordinairement précise (expérimentalement impossible) sur l'erreur ε .

Je n'insisterai pas davantage sur cet exemple¹, dont le but était seulement de montrer comment des considérations physiques concrètes imposent, en certains cas, le choix d'un modèle probabiliste unique et bien déterminé : c'est l'analyse physique du phénomène lui-même qui nous a fourni les critères d'objectivité interne (c'est-à-dire permettant de juger de l'intérieur l'objectivité du modèle probabiliste, sans recourir à la répétition de la "même" expérience).

Toutefois, dans ces indications exemplaires que nous offre la physique, il ne s'agit pas vraiment de phénomènes uniques, puisque la condition de répétabilité est toujours présente, et nous permet de renforcer et de recouper par des contrôles externes, basés sur la répétition, les critères proprement internes de l'objectivité de ces modèles. Donnons donc encore un exemple, d'apparence tout-à-fait banale, mais qui se révélera pourtant éclairant lorsque nous aborderons, dans la 3ème partie, la recherche de modèles strictement objectifs, ou, comme nous les appellerons, de représentations probabilistes permettant de décrire des phénomènes uniques.

Soit S un domaine borné de l'espace usuel (à 2 dimensions par exemple) et $f(x)$ une fonction² définie sur S . Donnons-nous également sur S une loi de probabilité, par exemple la loi de densité uniforme sur S . Dans le modèle où \underline{x} désigne le point aléatoire obtenu en "tirant au sort" un point x dans S selon cette loi de probabilité³, la fonction $f(x)$ est une variable aléatoire. La loi de cette variable est définie par sa fonction de répartition F :

1 - La mécanique statistique fournirait, en plus grandiose, un exemple comparable, et conduirait à des conclusions de même nature : c'est le caractère concret du modèle ou de la théorie physique qui impose ici de manière univoque le choix du modèle probabiliste.

2 - Il s'agit, bien sûr, d'une fonction mesurable.

3 - En langage plus correct, il s'agit du modèle (Ω, \mathcal{A}, P) où $\Omega = S$, \mathcal{A} est l'ensemble des boréliens de S , et P la probabilité uniforme sur S . Concrètement, S pourrait représenter un gisement minier, $f(x)$ la teneur en un point x et \underline{x} un échantillon (un sondage par exemple) implanté "au hasard" dans S .

$F(a)$, probabilité d'avoir $f(\underline{x}) \leq a$, est la mesure (surface, si l'espace est à 2 dimensions) du sous-ensemble de S défini par la condition $f(x) \leq a$. Cette loi a donc un sens absolument concret et, en particulier, objectif : une fois que l'on connaît la fonction f , tout énoncé probabiliste relatif, dans ce modèle, à la variable aléatoire $f(\underline{x})$ est, en effet, décidable au sens strict. Tous les paramètres associés à cette loi ont, eux aussi, une stricte signification objective. Par exemple, l'espérance mathématique m de la variable $f(\underline{x})$ de ce modèle est, par construction, égale à la valeur moyenne dans S de la fonction f , soit

$$m = \frac{1}{S} \int_S f(x) dx$$

grandeur possédant une signification parfaitement objective (décidable).

Supposons maintenant que nous "choisissions au hasard" dans S , indépendamment les uns des autres, des points particuliers x_1, x_2, \dots (ce qui est techniquement réalisable au moyen de tables de nombres au hasard, ou grâce aux procédés "arithmétiques" de génération de tels nombres que nous avons envisagés au chapitre précédent). Les valeurs numériques obtenues $f(x_1), f(x_2) \dots$ peuvent alors être interprétées dans le cadre du modèle "suite de variables indépendantes admettant la même loi F ". En particulier, si la fonction f , quoique parfaitement déterminée, ne nous est pas connue, de sorte que les seules données sont constituées d'une suite finie de valeurs numériques $f(x_1), f(x_2) \dots f(x_n)$, nous reconstituons dans un cadre absolument concret, les conditions du problème théorique de l'"inférence statistique" et de l'"estimation" de m : avec cet avantage énorme que nous sommes assurés ici, d'entrée de jeu, de la signification objective de la loi que nous cherchons à inférer, ou du paramètre m que nous cherchons à estimer. De plus, des considérations très terre à terre, relatives au calcul numérique approché de l'intégrale d'espace définissant m à partir d'un petit nombre de points x_1, x_2, \dots, x_n , viennent jeter une lumière crue, mais appréciable, sur la nature quelque peu mystérieuse de

cette fameuse "inférence statistique". Dans ce cadre, il est permis de parler sans guillemets de l'estimation¹ du paramètre m , puisque ce m là existe réellement ailleurs que dans notre modèle probabiliste.

1 - Je rappelle que je réserve le terme "estimation" à l'évaluation d'une grandeur, dont nous ne connaissons pas la valeur exacte, mais qui n'en n'existe pas moins, dans la réalité, indépendamment de nous. C'est donc seulement lorsque je suis assuré, comme dans le cas présent, de la signification objective d'un paramètre que je parle d'estimer ce paramètre. Dans les autres cas, j'utilise le verbe choisir. Gardons toujours présente à l'esprit la distinction entre énoncés objectifs, choix méthodologiques et critères de contrôle : on estime des grandeurs objectives, on choisit des méthodes et on convient de critères.

ESTIMER ET CHOISIR

DEUXIEME PARTIE

LES CRITERES DE L'OBJECTIVITE INTERNE DANS LE CAS
DES PHENOMENES UNIQUES

"Omnis res positiva extra animam eo ipso est singularis"

G. d'Ockham

Traduction approximative : "Toute chose existant en dehors de la pensée est,
par cela même une chose singulière".

E S T I M E R E T . C H O I S I R

- Essai sur la Pratique des Probabilités -

DEUXIEME PARTIE

Table des Matières

<u>CHAPITRE III - LA FORET POISSONNIENNE</u>	65
Le paramètre existe-t-il ? Est-il utile?	66
Les trois étapes du choix du modèle	68
Quelques fils directeurs	71
Un critère d'objectivité : les grandeurs régionales	73
Le point de vue de la pratique	74
<u>CHAPITRE IV - LE CHOIX ET LA HIERARCHIE DES MODELES</u>	76
Un premier critère : la décidabilité à l'aide des régionales	77
Le modèle constitutif	78
Le modèle générique : type et spécification	80
Premières indications sur les critères de choix	85
Le modèle primaire	87
Seuil de robustesse et seuil de réalisme	91
Un exemple de contrôle de la robustesse de type	93
La robustesse vis-à-vis des données	96
<u>CHAPITRE V - FAIRE LE TRI</u>	99
Les "fluctuations" des régionales	99
L'ergodicité	103
La portée	106
La micro-ergodicité	109
L'estimation, in praxi, des grandeurs régionales	114
Exemple de l'espérance conditionnelle	116
Exemple du krigeage	117

- CHAPITRE III -

LA FORET POISSONNIENNE

Dans les pages qui suivent et, sauf exception, jusqu'à la fin de cet ouvrage, nous concentrerons notre attention sur le cas d'un objet ou phénomène unique - forêt, tel gisement minier, telle chaîne de montagnes etc...- occupant une portion bien délimitée et bornée de l'espace où nous vivons. Ces objets sont intéressants par eux-mêmes. Ils peuvent être, et sont effectivement, étudiés du point de vue panscopique qui est celui de la connaissance scientifique. Mais ils sont aussi utiles. Le praticien chargé de les gérer et de les exploiter -le forestier, le mineur- rencontre des problèmes, de nature essentiellement pratique, qu'il doit résoudre, d'une manière ou d'une autre, pour mener sa tâche à bien : le mineur, par exemple, doit estimer les différentes parties de son gisement, avant de décider lesquelles exploiter et lesquelles laisser en place, comme trop pauvres. Pour faire cette estimation, il ne dispose, par la force des choses, que d'une information limitée et fragmentaire. Son point de vue sera donc, essentiellement, monoscopique : non qu'il se désintéresse des autres aspects de son objet, on observe souvent, au contraire, chez les professionnels, une sorte d'enthousiasme contenu ; mais il doit penser d'abord à remplir sa tâche. Dans ce qui suit, j'adopterai tantôt l'un et tantôt l'autre de ces points de vue, mais la balance penchera plutôt en faveur des modèles monoscopiques: c'est à propos de ces derniers, en effet, que le problème de l'objectivité se pose de la manière la plus aigue ; c'est d'eux aussi que j'ai acquis, par une longue pratique, le plus d'expérience directe, d'eux par suite que je puis espérer parler en sachant à peu près ce que je dis.

L'objectivité externe, c'est-à-dire celle de la méthodologie générale qui sert à construire les modèles individuels, ne soulève pas de problèmes majeurs : c'est, tout simplement, la sanction de la pratique. La Géostatistique, par exemple, ayant servi à estimer plus d'une centaine de gisements miniers, a fait ses preuves, et peut être considérée comme objectivement fondée. Il s'agira donc, dans ce qui suit essentiellement de l'objectivité interne : celle des énoncés concernant ce gisement ci, cette forêt là. Nous commencerons par un exemple introductif, celui de la "forêt poissonnienne"

LE PARAMETRE θ : EXISTE-T'IL, EST-IL UTILE ?

Pour entrer concrètement dans le vif du sujet, considérons le modèle poissonien parfois utilisé par les forestiers pour représenter la distribution des arbres (assimilés à des points) dans l'étendue de la forêt (J.P. MARBEAU,¹). Naturellement, dire que la forêt "est" poissonienne est une manière abrégée d'indiquer que les données disponibles ne sont pas incompatibles avec cette interprétation poissonienne. Dans l'optique statistique "orthodoxe", le problème le plus important qui se poserait alors serait celui de "l'inférence statistique", c'est-à-dire ici, celui de "l'estimation" de la densité inconnue θ de ce processus poissonien. En effet, une fois ce paramètre connu, on peut calculer toutes les autres caractéristiques du processus. Ce point de vue attribue, implicitement, une existence réelle, objective à la densité θ : même si notre information ne nous permet qu'une estimation approximative de la "vraie" valeur de θ , celle-ci n'en existerait pas moins quelque part dans la nature, et pourrait faire l'objet d'une mesure rigoureuse si notre information était parfaite. Or, il n'est pas certain que cette affirmation ait un sens opératoire : pour déterminer θ rigoureusement, il faudrait que la forêt s'étende à l'infini (en restant poissonienne), alors que son étendue réelle est évidemment limitée. Cette fausse évidence (l'existence de θ) repose sur une identification sommaire du modèle (le processus de Poisson) et de la réalité (la forêt) : une telle confusion, fréquente chez les statisticiens, est un véritable court-circuit épistémologique. Car, si bien adapté que soit le modèle à son objet, nous n'avons en aucun cas la garantie que toutes ses caractéristiques soient le reflet fidèle de propriétés objectives qui leur correspondraient biunivoquement dans la réalité.

Si nous maintenons fermement cette distinction nécessaire entre le modèle et la réalité, le problème (en toute rigueur insoluble) de l'inférence statistique perd une grande partie de son importance. Le paramètre θ resterait en partie indéterminé, même si nous connaissions parfaitement la forêt : on peut donc mettre sérieusement en doute son "existence objective", et par conséquent aussi ne pas juger trop grave son indétermination. De fait, le problème pratique intéressant le forestier n'est, en aucun cas, l'estimation de la densité poissonienne θ du modèle théorique. : celle-ci, (si on y a recours) n'interviendra jamais que comme un intermédiaire commode de calcul, et disparaîtra toujours du résultat final. En réalité, le forestier souhaite, par exemple, estimer le nombre $N(S)$ d'arbres présents dans une surface S donnée,

1 - Géostatistique Forestière, Thèse, Fontainebleau 1976.

connaissant le nombre $N(s)$ des arbres comptés dans la surface s échantillonnée. Ou bien, ce qui revient au même, il s'intéresse au nombre moyen $\theta(S) = N(S)/S$ d'arbres à l'hectare dans la surface S . Pour estimer $\theta(S)$ et $N(S)$, il calculera les quantités :

$$(1-1) \quad \theta^*(s) = \frac{1}{s} N(s) \quad ; \quad N^*(S) = \frac{S}{s} N(s)$$

Il est remarquable que l'interprétation poissonnienne ne joue, en réalité, à peu près aucun rôle dans la formation de ces "estimateurs", qui semblent bien les plus "naturels" possibles. De fait, ces estimateurs apparaissent comme non biaisés dans le cadre d'une interprétation beaucoup moins restrictive que le modèle poissonien. Il suffit, par exemple, de postuler une certaine forme de stationnarité (dans la terminologie relative au modèle probabiliste), c'est-à-dire une certaine forme d'homogénéité du phénomène dans l'espace, une certaine constance de la densité moyenne (terminologie plus vague, mais relative à la réalité physique, et non plus au modèle, et présentant un sens pour les praticiens).

Par contre, le caractère poissonien du modèle intervient effectivement au moment du calcul de la variance d'estimation σ^2 c'est-à-dire la variance (dans le modèle) de l'erreur $N(S) - N^*(S)$. En supposant la surface échantillonnée s incluse dans la surface S à estimer ($s \subset S$), on trouve sans difficulté l'expression suivante :

$$\sigma_E^2 = \theta \frac{S(S-s)}{s}$$

Mais, à dire vrai, le paramètre θ qui figure dans cette formule n'est pas connu exactement ; on le remplace donc par son estimation $N(s)/s$, ce qui conduit à prendre :

$$(1-2) \quad \sigma_E^2 = N(s) \frac{S}{s} \left(\frac{S}{s} - 1 \right)$$

relation où ne figure plus aucun paramètre du modèle poissonien hypothétique, et où interviennent seulement des quantités expérimentalement connues.

Ainsi, le paramètre θ du modèle ne figure nulle part, ni dans l'expression (1-1) des estimateurs, ni dans l'expression (1-2) de la variance que nous leur attribuons. Ce paramètre n'étant jamais réellement utilisé, et sa signification objective étant par ailleurs plutôt douteuse, on commence à soupçonner que le fameux problème de l'inférence statistique est sans doute un faux problème. Par contre, pour passer de (1-1) à (1-2), c'est-à-dire pour obtenir la variance d'estimation, on utilise explicitement "l'hypothèse" poissonnienne, c'est-à-dire le fait que le modèle probabiliste que nous avons choisi est du type poissonien, mais sans qu'il soit à aucun

moment nécessaire de spécifier le modèle, c'est-à-dire de fixer la valeur numérique de θ . Mais choisir un modèle générique de type poissonien (considéré indépendamment des valeurs du paramètre θ qui le spécifie) revient simplement à formuler dans un langage probabiliste - avec tous les avantages conceptuels et toutes les possibilités opératoires que celui-ci fournit - une propriété physique objective que l'on ne pourrait décrire autrement que dans un langage plus vague, à savoir :

i) la densité "moyenne" d'arbres à l'hectare ne semble pas présenter de variations systématiques dans l'espace, et

ii) le fait qu'une zone s soit très (ou très peu) fournie en arbres n'implique pas "en moyenne", qu'une zone s' voisine de s soit plus (ou moins) fournie que les autres.

Ces deux propriétés (même s'il est difficile d'en donner un énoncé précis dans un langage non probabiliste) correspondent indéniablement à des caractéristiques objectives du phénomène réel. Elles paraîtront certainement pleines de sens au praticien (le forestier) qui, en général, n'hésitera pas à dire dans chaque cas particulier si elles lui semblent, ou non, plausibles.

On voit qu'il ne s'agit plus exactement ici du problème classique de l'inférence statistique (estimation de la densité θ d'un processus de Poisson) mais bien plutôt du choix d'un modèle adéquat pour représenter une réalité physique donnée. Une fois admise - et c'est le point crucial - une certaine interprétation physique de la réalité (homogénéité dans l'espace, et "indépendance" des zones disjointes) qui conduit à adopter le modèle générique poissonien, le problème de la spécification de ce modèle (l'estimation de θ) ne présente plus qu'une importance limitée, ou, même ne se pose pas réellement : ni les expressions (1-1) des estimateurs, ni la variance (1-2) qu'on leur attribue, n'en dépendent. C'est donc uniquement pour simplifier le langage que l'on dira "cette forêt peut être représentée par un processus de Poisson admettant la densité $\theta =$ tant d'arbres à l'hectare" (et la valeur numérique attribuée à θ coïncidera avec celle de l'estimateur $\theta^x(S)$ de la densité moyenne dans S).

Les 3 étapes du choix d'un modèle

Or, cet énoncé résume trois étapes dont le statut épistémologique est bien différent :

1) Il y a d'abord un choix épistémologique : on décide de recourir à des techniques probabilistes pour représenter le phénomène (la forêt). C'est une décision, non une hypothèse : décision constitutive (elle constitue pour nous la forêt en tant qu'objet d'étude, définit le cadre général dans lequel nous allons travailler et

détermine le choix des instruments que nous allons utiliser), et non pas hypothèse expérimentalement contrôlable (il n'est ni vrai ni faux de dire "cette forêt est une réalisation d'un processus stochastique" en ce sens qu'il n'existe pas d'expérience ou d'observation concevable dont le résultat serait susceptible de réfuter cette proposition). A ce niveau, nous parlerons de modèle constitutif (ici, probabiliste).

2) Nous rencontrons ensuite une hypothèse de nature physique sur le phénomène en question - homogénéité dans l'espace, absence d'influence entre zones voisines - qui conduit au choix d'un modèle générique : ce processus est un processus de Poisson. Contrairement au choix précédent (qui ne se justifie que par son efficacité, par les succès auxquels il conduit, et sur lequel on ne peut porter de jugement qu'à la longue, après avoir traité un grand nombre de cas) ce second choix découle d'une hypothèse physique objectivement contrôlable. Il peut, en conséquence, être confirmé ou infirmé par les données expérimentales - soit au moyen de tests statistiques, faciles à mettre en oeuvre dans ce cas particulier très simple, soit par telle ou telle autre méthode, y compris le jugement du praticien qui connaît bien cette forêt. (Ce recours à l'intuition du praticien ne dissimule aucun mysticisme ; il traduit plutôt la priorité épistémologique attribuée à la réalité sur le modèle mathématique, ou statistique, que l'on a choisi pour la décrire).

Il convient de souligner l'importance capitale de cette étape. C'est en effet, pour l'essentiel, ici que nous incorporons au modèle des hypothèses ayant une signification objective, et véhiculant une information positive non contenue dans les données numériques brutes. Seul, cet apport positif permet de comprendre que nous puissions tirer (en apparence) des données plus qu'il n'y est réellement contenu (une prévision plus une variance d'estimation). La contrepartie de ce petit miracle, c'est que notre modèle est maintenant vulnérable, et que nos prévisions peuvent être démenties par l'expérience si les hypothèses sur lesquelles nous les avons fondées ne sont pas objectivement valables. Précisons bien qu'il ne suffit pas de s'assurer (par exemple par des tests) que ces hypothèses sont compatibles avec les données, c'est-à-dire sont vérifiées sur les parcelles échantillonnées. Il faut aussi admettre qu'elles restent valables dans les zones non échantillonnées, et cela, c'est après coup seulement que nous le saurons. Le choix du modèle constitue une hypothèse anticipatrice, et implique toujours un risque d'erreur radicale, et c'est pourquoi il doit impérativement tenir compte, non seulement des données numériques, mais également de toutes les autres sources d'information dont on peut disposer (connaissances générales sur ce type de phénomène, expérience des praticiens, etc...).

Pour mieux localiser l'apport d'information positive et l'apparition corrélative de la vulnérabilité du modèle, il y aura souvent intérêt à distinguer deux étapes dans le choix du modèle générique : choix d'un modèle générique au sens large, puis choix d'un type particulier dans ce modèle. Ici, le modèle générique (au sens large)

serait par exemple "processus ponctuel stationnaire". Le mot ponctuel signifie que nous avons décidé de considérer, en première approximation, les arbres comme des points dans le plan. C'est une décision (constitutive) n'impliquant ni apport d'information positive ni risque de démenti expérimental. Par contre, le mot stationnaire est associé à une hypothèse d'homogénéité dans l'espace, et implique apport positif et vulnérabilité, mais à un degré relativement faible. Par type de modèle, nous entendrons un modèle auquel ne manque plus pour être complètement spécifié que la donnée des valeurs numériques d'un petit nombre de paramètres. Ici, le type de modèle choisi est "le processus de Poisson", avec un seul paramètre indéterminé (le paramètre θ). Le choix de ce type est lié à une hypothèse très forte (absence d'influence entre zones voisines) ; il introduit une information très riche (qui permet, entre autres, le calcul des variances d'estimation) et donc aussi des risques de démenti expérimental. C'est là une circonstance générale : le plus souvent, c'est le choix d'un type dans un modèle générique qui constitue la décision cruciale, celle qui ouvre le plus de possibilités opératoires, et, corrélativement, introduit le plus de risques d'erreur.

3) La dernière étape, enfin, est le choix du modèle spécifique ou, comme nous dirons aussi, la spécification du modèle (ici, le choix de la valeur numérique θ = tant d'arbres à l'hectare). Alors que la statistique mathématique attribue une importance absolument prédominante à ce troisième aspect du "choix du modèle", auquel elle donne le nom d'inférence statistique (i.e. estimation numérique des paramètres), il ne joue ici qu'un rôle très effacé, ou même nul, puisque les résultats essentiels (l'estimation et la variance) s'expriment, en définitive, à l'aide des seules données expérimentales, sous une forme qui fait intervenir les choix 1) et 2) (et surtout le choix du type poissonien) mais en aucune façon le choix 3) de la valeur numérique de θ . C'est, en somme, seulement par commodité de langage, et pour fixer les idées, que nous procédons à la spécification du modèle, alors que le choix 1) (épistémologique) et le choix 2) (le modèle générique et son type), joints à l'information numérique - ici $N(s)$ - fournissent une base opératoire suffisante pour résoudre les problèmes que nous nous posons.

La distinction précise de ces trois aspects du choix du modèle paraît élémentaire et peu intéressante dans un cas aussi simple que celui de la forêt "poissonienne". Mais les choses changeront vite au fur et à mesure que nous examinerons des modèles plus riches. Une fois admis le choix du modèle constitutif (décision épistémologique), le problème capital reste le choix n° 2, celui du modèle générique et plus particulièrement du type de modèle : car, une fois adopté un type de modèle convenablement adapté aux données expérimentales, il est en général raisonnablement possible de le spécifier (si on le désire vraiment) conformément à ces mêmes données. Le problème important n'est donc plus l'inférence statistique, mais le choix du modèle générique

et de son type. Notons bien qu'il ne s'agit pas simplement ici de "tester une hypothèse" à l'aide des données. Ce point de vue, qui est celui de la statistique orthodoxe, laisse échapper l'essentiel : ce n'est pas seulement de l'accord avec les données disponibles que nous avons besoin, mais de l'hypothèse beaucoup plus forte que le type choisi est compatible aussi avec les données non disponibles, c'est-à-dire avec les parties inconnues ou inaccessibles du phénomène. C'est cette dernière hypothèse qui explique la fécondité de la méthode, et aussi, nous l'avons vu, sa vulnérabilité. La compatibilité avec les données est, évidemment, nécessaire, mais elle ne suffit en aucun cas à nous garantir contre un désaccord toujours possible avec ce qui n'est pas donné (et que justement nous cherchons à estimer).

Nous examinerons au chapitre 4, dans un cadre plus général, la hiérarchie des modèles "topo-probabilistes", et le problème du choix tel qu'il se pose à chacune des trois ou quatre étapes que nous avons distinguées. Auparavant, nous allons procéder à un examen critique de notre exemple élémentaire, ce qui nous permettra de dégager quelques fils directeurs.

Quelques fils directeurs

Voyons comment les choses se présentent dans le cas de la forêt "poissonienne". Pour examiner l'objectivité de ce modèle, nous devons distinguer nettement deux problèmes bien différents. Il faut en premier lieu faire abstraction du fait (somme toute contingent) que nous ne disposons que d'une information partielle, et supposer (ce qui ne nous fait pas sortir du cadre opératoire) que nous disposions d'un relevé topographique complet donnant l'implantation de chacun des arbres de la forêt : seuls les concepts accessibles à partir de cette base expérimentale maximale auront un sens opératoire et une signification objective. En second lieu, il convient d'examiner les complications et incertitudes supplémentaires qu'introduit le caractère lacunaire de notre information réelle, et le problème qui se pose est celui du degré de validité ou même de la possibilité d'une estimation effectuée sur cette base : ce second problème est le plus important en pratique, mais du point de vue de la méthodologie, c'est plutôt le premier point qui l'emporte.

a) La forêt étant supposée parfaitement connue, pouvons-nous, en un sens objectif, la qualifier de "poissonienne", et, en particulier, pouvons-nous donner une définition opératoire de la densité poissonienne θ ? La réponse est : oui, mais dans certaines limites seulement. La forêt ayant, par exemple, une surface $S = 10.000$ hectares, peut être divisée en parcelles jointives de même surface s : s ne doit pas être trop petite (disons au moins 400 m^2), pour que l'assimilation des arbres à des points demeure raisonnable, ni trop grande (disons au plus 100 hectares) pour que l'on dispose d'un effectif "statistique" suffisant. Pour chaque valeur de s , nous disposons d'une popu-

lation de $n = S/s$ valeurs numériques, à savoir, les nombres des arbres comptés sur chacune des parcelles s , et nous pouvons examiner si l'histogramme correspondant est, ou non, compatible avec une interprétation poissonnienne. Il est vrai que les réponses que donnent les tests statistiques sont toujours équivoques. Mais, s'agissant de populations comportant au moins 100 individus, et en général beaucoup plus (jusqu'à 250.000 pour la plus petite valeur de s), ces réponses seront suffisamment nettes pour que nous puissions négliger cette marge d'ambiguïté ou d'indétermination (inhérente, nous l'avons vu, à la définition de tout concept opératoire). Si ces populations sont suffisamment poissonniennes, et si les variances (expérimentales) correspondantes $\sigma^2(s)$ sont approximativement proportionnelles à la surface s (et le degré d'approximation admissible peut également, dans une certaine mesure, être précisé sur la base de considérations statistiques), nous avons mis en évidence l'existence d'une loi (approximative) $\sigma^2(s) = \theta s$, qui constitue justement la définition opératoire de la "densité poissonnienne" θ .

Ceci dit, il faut immédiatement ajouter

i) que cette densité θ n'est, non seulement mesurée, mais même, plus profondément, définie qu'avec une certaine marge d'indétermination correspondant au caractère approximatif de la loi $\sigma^2(s) = \theta s$ qui constitue la définition de θ , et aussi

ii) que le concept de densité poissonnienne perd toute validité objective pour des surfaces s trop petites ou, surtout, trop grandes. Par exemple, pour $s = S/2$ (2 parcelles) ou $s = S$ (une parcelle unique) on ne peut plus parler de population poissonnienne ni de variance proportionnelle à s . En particulier, considérer le nombre total $N(S)$ des arbres de la forêt "comme une variable poissonnienne de moyenne θ " ne présente plus aucune espèce de signification objective : nous sommes complètement sortis du domaine de validité du concept opératoire.

On peut faire encore d'autres remarques. Le modèle poissonnien, dans ses limites de validité opératoire, n'est pas le seul possible, ni même le meilleur. En effet, pour raisonner un instant dans un cadre purement probabiliste, on sait que, quelle que soit la valeur du paramètre θ du processus de Poisson, une fois fixé le nombre total $N(S)$ des points tombés dans la surface S , ces $N(S)$ points sont répartis indépendamment les uns des autres dans S avec une densité de probabilité uniforme : autrement dit, conditionnellement à $N(S)$ fixé, cette loi de répartition est la même quelle que soit la valeur de θ . Toujours à $N(S)$ fixé, les parcelles disjointes s ne sont plus tout à fait indépendantes (puisque la somme des nombres d'arbres qu'elles contiennent doit être égale à $N(S)$), mais, pour $S/s > 100$ leur degré de dépendance peut être considéré comme négligeable. Les lois correspondantes sont binomiales (et non plus poissonniennes, encore qu'elles diffèrent peu des lois poissonniennes), espérances et variances étant données par :

$$(1-3) \quad E [N(s)] = \frac{s}{S} N(S) \quad ; \quad \text{Var} [N(s)] = \frac{s}{S} \left(1 - \frac{s}{S}\right) N(S)$$

Comme s/S est $\leq 1/100$, la variance est pratiquement égale à $(s/S) N(S)$, c'est-à-dire à l'espérance : nous obtenons le même résultat qu'avec une loi poissonnienne d'espérance $s \theta(S)$ (au lieu de $s \theta$), $\theta(S) = N(S)/S$ représentant justement "l'estimation" classique de la densité θ .

Pour un statisticien, cette élimination du paramètre θ provient du fait que $\theta(S) = N(S)/S$ constitue une "statistique exhaustive", c'est-à-dire résume la totalité de l'information dont nous disposons pour "estimer" θ , ce qui se traduit ici par le fait que la loi "conditionnelle" des variables $N(s)$ associées aux différentes parcelles ne dépend plus de θ lorsque $N(S)$ est fixé. Sous cette forme, il s'agit d'une particularité du processus de Poisson, qui ne subsisterait pas dans le cas de processus plus généraux. Mais cette manière de présenter les choses revient à renverser les rôles, je veux dire à porter un jugement sur la réalité en se plaçant du point de vue du modèle. Je donnerai plus loin un énoncé général du critère d'objectivité dont nous avons besoin. Pour l'instant, poursuivons l'étude de notre forêt poissonnienne.

Un critère d'objectivité : les grandeurs régionales

A ce stade de l'analyse, nous commençons à soupçonner que le paramètre θ du modèle théorique est peut être bien dépourvu d'existence objective, puisque tous les caractères objectifs que nous lui avons attribués, y compris sa définition opératoire, sont aussi bien pris en charge par le prétendu estimateur $\theta(S) = N(S)/S$ (et, en fait, mieux, puisque la formule (1-3) de la variance conserve un sens pour $s = S$: la variance de $N(S)$ est effectivement nulle, puisque $N(S)$ est fixé). On voit ainsi se renverser le point de vue de la statistique orthodoxe : ce qui était considéré comme une réalité, inconnue mais existant objectivement et devant faire l'objet d'une "inférence statistique" (le paramètre θ), se révèle une fiction inutile, et c'est la prétendue estimation (la densité moyenne $\theta(S) = N(S)/S$ de la forêt réelle) qui reçoit seule, maintenant une signification objective.

Convenons d'appeler grandeurs régionales, ou, simplement, régionales les quantités qui, comme $N(S)$, ou aussi bien les nombres $N(s)$ associés à toutes les parcelles s possibles, se trouvent définies sans équivoque lorsque le phénomène régionalisé auquel on s'intéresse (ici, la forêt) est suffisamment connu dans la totalité de son extension spatiale (ou champ) S . Le paramètre poissonnien θ n'est pas une grandeur régionale, et, aussi bien, nous l'avons vu, il est dépourvu de signification objective. Nous pouvons énoncer un premier principe, utile pour nous orienter dans le choix de modèles probabilistes destinés à décrire des phénomènes uniques. :

- Principe des grandeurs régionales : Parmi les paramètres d'un modèle topo-probabiliste destiné à décrire un phénomène unique, seuls ont une signification objective ceux que l'on peut exprimer en termes de grandeurs régionales. De même, parmi les concepts introduits par ce modèle, seuls ont un sens objectif ceux qui peuvent recevoir une définition opératoire en termes de grandeurs régionales, c'est-à-dire une définition constituée par un système de relations expérimentalement vérifiables (dans un domaine de validité donné, avec une approximation donnée) entre de telles grandeurs.

Ainsi, le concept de loi de distribution (poissonienne ou binomiale) du nombre $N(s)$ des arbres présents sur les parcelles de taille s a, indéniablement, un sens opératoire pourvu que s ne soit ni trop petit ni trop grand. Par contre, pour $s = S$, il n'y a aucun sens opératoire à considérer $N(S)$ comme une variable poissonienne d'espérance θS .

Le point de vue de la pratique

b) Venons-en maintenant au second problème, celui de la possibilité d'une estimation sur la base d'une information partielle. En particulier, nous devons nous demander si le concept de variance d'estimation introduit plus haut est susceptible de recevoir une définition opératoire en termes de régionales. La réponse est évidemment différente si l'on se place "après coup" (c'est-à-dire si l'on suppose la forêt parfaitement connue, et si l'on compare à la réalité les estimations que l'on aurait pu faire sur la base de telle ou telle information partielle que nous pouvons choisir à volonté) ou "dans le vif de l'action" (c'est-à-dire si l'on suppose que l'on n'a réellement qu'une information partielle donnée à sa disposition).

Après coup, la réponse est certainement positive : on peut définir la variance d'estimation en termes de grandeurs régionales. (Dans le cas général, il y a même plusieurs façons non équivalentes de le faire, et cette difficulté retiendra plus tard notre attention. Mais, dans le cas "poissonien", toutes ces définitions sont équivalentes, et l'ambiguïté disparaît). La formule (1-3), en effet, qui donne la variance de $N(s)$ (à $N(S)$ fixé) exprime une relation, expérimentalement vérifiable, entre grandeurs régionales : dans la mesure où elle a été effectivement vérifiée, nous sommes en droit d'en déduire la variance de l'erreur associée à l'estimation de $N(S)$ sur la base de l'information $N(s)$ fournie par l'échantillonnage d'une surface (de forme quelconque) de taille s , ce qui donne :

$$(1-4) \quad \sigma^2_E = \left(\frac{S}{s} - 1\right) N(S)$$

Si, maintenant, nous nous plaçons "dans le vif de l'action", $N(S)$ est inconnue (c'est justement cette grandeur que nous voulons estimer) et on ne peut donc que lui substituer son estimation, qui est $(S/s)N(s)$. On retrouve ainsi la formule (1-2), qui constitue une approximation raisonnable de (1-4), pourvu que s ne soit pas trop petit.

La valeur objective de la formule (1-2) - la seule que l'on puisse utiliser en pratique - est donc inférieure, mais comparable à celle de la définition opératoire (1-4). Seulement la validité de cette dernière ne peut être définitivement établie qu'après coup (lorsque l'on connaît toute la forêt) et non sur la base de l'information partielle disponible "dans le vif de l'action". Par conséquent aussi, la valeur de (1-2) dépend de la validité d'une hypothèse anticipatrice (qui peut se révéler fautive après coup) selon laquelle le caractère poissonien, que nous avons pu raisonnablement tester sur la portion s échantillonnée, peut être étendu à toute la surface totale S .

C'est là une circonstance générale : nous ne sommes jamais à l'abri de mauvaises surprises. Le fait qu'un modèle soit en bon accord avec un échantillonnage partiel ne nous garantit jamais qu'il restera compatible avec la totalité du phénomène. Par conséquent toute prévision (y compris celle d'une variance d'estimation ou de tout autre concept exprimable en termes de grandeurs régionales) faite sur la base d'une information partielle et d'un modèle peut se révéler grossièrement fautive. Le degré de confiance que nous pouvons lui attribuer est lié à la valeur objective d'une hypothèse anticipatrice dont, justement, nous ne pourrions juger qu'après coup. En matière scientifique, on peut toujours suspendre son jugement en attendant d'avoir une information complémentaire. Mais le praticien, lui, est obligé de prendre une décision dans le vif de l'action, c'est-à-dire sur la base d'une information en général insuffisante, et cela implique toujours un risque d'erreur radicale.

LE CHOIX ET LA HIERARCHIE DES MODELES

Il faut maintenant définir avec un peu plus de précision les niveaux que nous avons rencontrés à propos de l'exemple simple de la forêt "poissonienne" : le modèle constitutif, le modèle générique, son type et sa spécification. La situation de référence est la suivante : nous nous intéressons à un certain phénomène ou objet défini sur un domaine borné S de l'espace (euclidien, à 1, 2 ou 3 dimensions, éventuellement 4 si l'on désire aussi faire intervenir le temps). Nous supposons que ce phénomène se laisse décrire de manière satisfaisante par la donnée d'une (ou éventuellement plusieurs) fonction z définie sur S , que nous appellerons, d'un terme général, variable régionalisée (V.R.). L'expression "après coup" signifiera toujours que nous nous plaçons, par la pensée, dans la situation idéale où l'on connaît la valeur numérique de $z(x)$ en chaque point $x \in S$; l'expression "in praxi"¹, au contraire, désigne la situation, habituelle en pratique, où l'on ne dispose que d'un échantillonnage partiel de la V.R., par exemple des N valeurs numériques $z_\alpha = z(x_\alpha)$ prises par celle-ci en N points expérimentaux donnés, x_α , $\alpha = 1, 2, \dots, N$. (On peut aussi, naturellement, envisager le cas où les z_α sont les valeurs numériques de telles ou telles fonctionnelles de z , par exemple les valeurs moyennes de la V.R. sur des ensembles donnés $s_\alpha \subset S$, celui où ces valeurs sont entachées d'erreurs de mesure etc...). Mais il est entendu que, même "in praxi", nous avons aussi, outre cette information numérique, certaines connaissances de nature qualitative et structurelle sur la physique du phénomène, certaines idées (intuitives ou rigoureuses, éventuellement erronées d'ailleurs) sur le type de comportement qu'il est susceptible de présenter, une expérience acquise pour avoir déjà traité des cas similaires, les avis (éventuellement contradictoires) de divers spécialistes et praticiens etc... Ces diverses sources d'information, malgré leur nature hétérogène et leur qualité inégale, pèsent d'un grand poids au moment où l'on doit choisir un modèle, c'est-à-dire adopter "in praxi" une hypothèse anticipatrice. Nous les désignerons d'un terme général, comme l'information non numérique, ou qualitative.

1 - in praxi : dans l'action. Avec une pointe de lyrisme, on peut entendre :
dans le feu de l'action.

Un premier critère : la décidabilité à l'aide des régionales

Nous appelons grandeur régionale, ou simplement régionale toute fonctionnelle de la variable régionalisée z définie sur S , c'est-à-dire toute grandeur dont la valeur se trouve déterminée par la donnée de toutes les valeurs numériques $z(x)$ lorsque x parcourt S . Dans un premier temps du moins, nous utiliserons les régionales comme critère d'objectivité. Dans un cadre (trop strict, comme nous le verrons) un énoncé a une valeur objective si, et seulement si, il est décidable après coup : c'est-à-dire s'il peut être déclaré, sans ambiguïté, vrai ou faux une fois que l'on connaît toutes les valeurs $z(x)$, $x \in S$. Nous utilisons (provisoirement) la décidabilité, et non, comme l'aurait suggéré le critère de K.R. POPPER, la seule falsifiabilité, car, après coup, par définition, tous les "falsificateurs virtuels" de l'énoncé sont eux-mêmes décidés, c'est-à-dire classés univoquement vrais ou faux. Dans ces conditions, d'ailleurs, la "valeur de vérité" de l'énoncé (1 s'il est vrai, 0 s'il est faux) apparaît elle-même comme une fonctionnelle de z , donc est elle-même une grandeur régionale.

Par la suite, nous assouplirons ce critère trop rigide, et nous reviendrons au critère Popperien fondé sur la seule falsifiabilité. C'est qu'en effet la variable régionalisée z n'est nullement identique à la réalité, mais constitue elle-même un premier modèle. Il n'existe, originellement, dans la nature, rien de tel que des points $x \in S$ ou des valeurs numériques $z(x)$: c'est nous qui les introduisons, à la suite d'un premier élagage ou d'une première simplification opérée dans le foisonnement du réel. Cette différence originelle reste inaperçue la plupart du temps, mais réapparaît en certaines occasions critiques, lorsque le besoin se fait sentir d'une analyse plus fine de la signification physique de certains concepts mathématiques (la continuité ou la dérivabilité par exemple). Nous reviendrons sur cette difficulté dans le paragraphe consacré au modèle primaire, et nous raisonnons, pour l'instant, comme si nous avions le droit d'identifier la V.R. et la réalité.

D'après ce critère, des énoncés comme : "la variable régionalisée z est une réalisation d'une fonction aléatoire" ou même, plus précisément, "d'une fonction aléatoire stationnaire" n'ont pas de signification objective : ils ne pourraient être déclarés ni vrais, ni faux, même si nous connaissions toutes les valeurs numériques de z , c'est-à-dire le maximum d'information empirique qu'il nous soit possible d'acquérir. De même, un énoncé comme : "il y a 99 chances sur 100 pour que la valeur moyenne de z dans S soit supérieure à 2,5" n'a pas de signification objective, du moins d'après le présent critère, qui est un critère d'objectivité interne.

En pratique, on utilise pourtant des énoncés de ce genre. On a raison de le faire, et on y est d'ailleurs à peu près obligé : mais il faut alors invoquer un critère différent, celui de ce que nous avons appelé l'objectivité externe, ou méthodologique.

Parmi les paramètres servant à spécifier notre modèle, seuls auront une signification objective ceux qu'il nous sera possible d'identifier à des grandeurs régionales. Nous les appellerons paramètres objectifs, et les autres seront dits paramètres auxiliaires ou conventionnels. Par exemple, dans un modèle où la V.R. z est considérée comme une réalisation d'une F.A. stationnaire d'espérance $m = E [Z(x)]$, cette espérance n'est pas un paramètre objectif, mais un paramètre conventionnel. De même, un concept, même s'il admet dans le modèle une définition mathématique, ne sera considéré comme opérateur (dans un domaine de validité défini) que s'il est possible de le redéfinir au moyen d'un système de relations expérimentalement contrôlables après coup (avec une approximation donnée, dans un domaine donné) entre grandeurs régionales.

Le modèle constitutif

Dans le cas qui nous occupe, le modèle "constitutif" est celui de fonction aléatoire (F.A.). Choisir ce modèle signifie que nous décidons de considérer la V.R. $z(x)$, $x \in S$ comme (la restriction à S d') une réalisation d'une F.A. $Z(x)$, que nous devons évidemment spécifier par la suite, mais à ce stade il s'agit seulement de la notion (mathématique) de F.A. en général. Il s'agit donc uniquement d'un choix méthodologique, d'une décision que nous prenons, et non pas d'une hypothèse vérifiable ou falsifiable. Nous avons vu, d'ailleurs, que la proposition "la V.R. z est une réalisation d'une F.A." est dépourvue de signification objective. Cela ne signifie pas que cette décision soit arbitraire. Si nous la prenons, c'est que nous avons de bonnes raisons de penser que c'est la meilleure décision possible, la seule peut-être qui nous fournisse un cadre conceptuel dans lequel nous puissions espérer formuler et résoudre au mieux le, ou les problèmes qui nous occupent. Avant de la prendre, nous aurons certainement essayé, dans ce cas-ci ou dans d'autres analogues, d'ajuster un modèle déterministe ou fonctionnel, qui nous aurait certes permis des prévisions plus précises, mais toujours sans succès (ceci s'entend, évidemment, in praxi : après coup, nous disposons toujours d'un modèle fonctionnel très précis, à savoir la fonction z elle-même. Mais, même après coup, ce modèle parfait risque d'être beaucoup trop compliqué pour que l'on puisse l'utiliser commodément).

Ce choix n'est pas constitutif à l'égard du phénomène lui-même (qui existe et qui est ce qu'il est indépendamment des décisions que nous prenons à son sujet), ni même à l'égard de la V.R. $z(x)$ elle-même : celle-ci représente d'ailleurs un premier modèle constitutif, antérieur à la F.A., car elle délimite et constitue un premier objet d'étude, en opérant une première réduction ou simplification dans le foisonnement du réel. Mais ce choix (du modèle "F.A.") est constitutif en tant qu'il fonde ou constitue l'objet que nous allons effectivement étudier (qui est maintenant : la V.R. en tant que réalisation d'une F.A.), et détermine ainsi à l'avance le choix des méthodes que nous utiliserons pour cette étude (c'est-à-dire : les techniques probabilistes).

L'intuition qui guide ce choix fondamental est évidemment celle de la coexistence, au sein du même phénomène, d'un aspect structuré et d'un aspect "aléatoire" ou chaotique. Nous soupçonnons l'existence de certains traits structuraux, masqués par la variabilité chaotique que manifeste le phénomène au niveau local, et qui exclut toute représentation fonctionnelle simple.

Cette intuition probabiliste est relativement fugitive, et difficile à expliciter. Elle peut même disparaître complètement avec un choix différent du modèle constitutif. Par exemple, nous pourrions choisir de représenter la V.R. $z(x)$ par un développement du type

$$z(x) = \sum a_n f_n(x) \quad (x \in S)$$

où les f_n sont des fonctions "orthogonales" sur S (par exemple des fonctions trigonométriques, si nous souhaitons de faire l'analyse harmonique du phénomène, ou bien des polynômes orthogonaux, etc...). Les f_n étant choisies, l'interprétation probabiliste consisterait à considérer en outre les coefficients a_n comme des réalisations de variables aléatoires A_n . Mais cette interprétation sent l'artifice. Si l'on connaît la totalité de ce que l'on peut savoir sur la V.R. $z(x)$, c'est-à-dire les valeurs numériques des coefficients a_n , tout ce que l'on peut dire de chaque V.A. A_n , c'est qu'un tirage au sort (unique) lui a attribué la valeur numérique a_n . Dans l'optique classique, on est dans un cas où l'inférence statistique est complètement indéterminée (on ne peut pas reconstruire la loi de la V.A. A_n à partir de la seule valeur numérique a_n). Dans la présente terminologie, nous dirons que les lois des A_n ne correspondent à aucun concept opératoire. Ce modèle (les A_n aléatoires) en effet, n'est pas spécifiable en termes de grandeurs régionales.

A dire vrai, nous n'aurions nul besoin de probabiliser les coefficients a_n du développement de $z(x)$. Une fois la V.R. parfaitement connue, ces coefficients sont numériquement déterminés, et constituent donc des paramètres objectifs. Si notre but est d'effectuer une analyse spectrale (énergétique), les $|a_n|^2$ nous fournissent ce que nous cherchons. Si nous désirons une représentation simplifiée, lissée, du phénomène, nous pouvons décomposer z en deux termes :

$$D = \sum_{n=0}^N a_n f_n \quad ; \quad R = \sum_{n=N+1}^{\infty} a_n f_n$$

dérive et résidu (le choix de N est évidemment arbitraire, mais en pratique, et compte-tenu de l'objectif que l'on poursuit, il n'y a qu'assez peu d'ambiguïté). A ce stade, dire que D représente la partie déterministe du phénomène et R la partie aléatoire n'est qu'une manière de parler assez creuse, puisque ces deux termes sont de même nature et ont exactement le même statut épistémologique. L'intuition probabiliste a bel et bien évacuée.

In praxi, des difficultés apparaîtraient. Sur la base d'une information limitée (quelques points expérimentaux), nous pourrions peut être évaluer de manière approximative les premiers coefficients a_n , par exemple ceux qui figurent dans l'expression D . Mais, en l'absence de modèle probabiliste, nous pourrions plus difficilement apprécier l'ordre de grandeur de l'erreur commise dans cette estimation.

Le modèle générique - type et spécification

La démarche principale va maintenant consister à choisir, parmi toutes les F.A. possibles, celle que nous allons charger de représenter le phénomène auquel nous nous intéressons. Nous désignerons cette fonction aléatoire par Z , ou $Z(x)$, $x \in S$. Pour faire ce choix, nous disposons d'une part d'une information qualitative sur le phénomène, de l'autre de la V.R. $z(x)$, $x \in S$ (connue partiellement seulement in praxi) considérée comme une réalisation de Z . C'est ici que nous nous heurtons au difficile problème que la terminologie classique désigne sous le nom d'inférence statistique. D'habitude, en effet, on définit une F.A. Z par sa loi spatiale, c'est-à-dire par la donnée de toutes les lois de distribution (en nombre infini) de la forme :

$$P[Z(x_1) \leq z_1 ; Z(x_2) \leq z_2 ; \dots ; Z(x_n) \leq z_n]$$

pour tous les entiers $n \geq 1$, et tous les choix possibles des points d'appui $x_1, \dots, x_n \in S$ et des nombres réels z_1, \dots, z_n . Or, il n'est pas possible de procéder à l'estimation de toutes ces lois, à partir de l'information, disponible in praxi, qui est constituée d'un nombre fini d'échantillons prélevés dans S . Contrairement à ce qui se passe en physique par exemple, où l'on a en principe la possibilité de répéter aussi souvent que l'on veut une même expérience, et où l'on peut ainsi disposer de plusieurs réalisations d'une même F.A., nous avons affaire ici à une réalisation unique échantillonnée en un nombre fini de points seulement. L'inférence statistique de la loi spatiale (dans sa totalité) n'est certainement pas possible "in praxi". Et, ce qui est plus important, il est facile de se rendre compte qu'elle ne l'est pas davantage "après coup", c'est-à-dire même si l'on connaît $z(x)$ en tout $x \in S$. Du point de vue classique, nous nous heurtons à un problème insoluble ou indéterminé. Pour nous, compte-tenu de notre critère d'objectivité en termes de grandeurs régionales, nous concluerons qu'il s'agit plutôt d'un faux problème : la loi spatiale n'est pas univoquement définie par la V.R., et se trouve donc dépourvue de signification objective.

Par conséquent, nous devons d'une manière ou d'une autre introduire des conditions limitatives en vue de lever l'indétermination. Du point de vue classique, ces conditions limitatives sont le plus souvent considérées comme des "hypothèses" (par exemple "l'hypothèse" de la stationnarité) qui, si elles sont vérifiées, permettent de diminuer le nombre des paramètres à estimer et rendent raisonnablement possible l'inférence statistique in praxi. Encore faudrait-il qu'elles soient effectivement controlables après coup. Or, même du point de vue classique, une hypothèse trop générale comme la stationnarité ne peut pas faire l'objet de tests statistiques significatifs. En effet, n'importe quelle fonction z définie sur un domaine borné S peut toujours être considérée comme la restriction à S d'une fonction périodique, dont les périodes (selon les axes de coordonnées) définissent un parallélépipède grand vis-à-vis du champ S . Moyennant une translation par un vecteur aléatoire (uniformément réparti dans le parallélépipède des périodes) cette fonction périodique devient, en toute rigueur, une F.A. stationnaire dont la fonction z donnée sur S peut, très raisonnablement, être considérée comme une réalisation. D'où il suit qu'aucun test portant sur la "stationnarité" en général ne pourra jamais conduire à une réponse négative.

Cette conclusion ne doit pas nous surprendre : elle signifie que la "stationnarité en général" n'est pas décidable¹ en termes de grandeurs régionales, et n'a donc pas de signification objective. Par conséquent, parmi les conditions limitatives que nous introduirons, nous devons très soigneusement distinguer entre celles qui correspondent réellement à des hypothèses objectives (décidables² après coup) et celles qui constituent seulement des choix méthodologiques. Les premières, seules, véhiculent cette information supplémentaire substantielle qui nous permet de tirer des données numériques disponibles in praxi plus qu'il n'y est réellement contenu. Ce sont elles qui constituent, à proprement parler, l'hypothèse anticipatrice, mentionnée à plusieurs reprises, qui peut se trouver démentie après coup. Cette vulnérabilité est la contrepartie obligée de leur fécondité, et c'est pour diminuer, autant qu'il est possible, cette vulnérabilité que nous devons les choisir in praxi selon un principe d'économie stricte : dans l'optique monoscopique qui est ici la nôtre, une hypothèse objective ne doit être adoptée in praxi que si elle est absolument indispensable pour atteindre le but que nous poursuivons (et, naturellement, si elle n'est pas incompatible avec l'information - numérique et qualitative - dont nous disposons). Les secondes, au contraire, n'introduisant ni information positive, ni risque de vulnérabilité, relèvent de simples décisions méthodologiques : la règle sera donc de choisir celles qui conduisent à la solution la plus efficace du problème que l'on cherche à résoudre. (Notons aussi, comme chaque fois qu'interviennent des choix méthodologiques, que la manière dont nous effectuons ces choix dans chaque cas particulier relève d'une méthodologie plus générale qui sera soumise à la longue à la sanction de la pratique, et n'est donc pas dépourvue d'objectivité à long terme, bien que chaque choix particulier reste en un sens arbitraire).

Quoi qu'il en soit, une fois introduites ces conditions limitatives (hypothèses objectives et choix méthodologiques), nous n'avons plus à chercher notre F.A. $Z(x)$ dans l'ensemble infiniment trop vaste de toutes les F.A. possibles, mais seulement dans la famille \mathcal{F} plus restreinte définie par ces conditions, à l'intérieur de laquelle le nombre des paramètres qui définissent une F.A. individuelle se trouve substantiellement réduit (bien qu'il reste, le plus souvent, encore infini). Naturellement, si nous ajoutons des conditions limitatives supplémentaires, nous obtenons une hiérarchie

1 - Elle n'est même pas falsifiable : le concept de stationnarité est donc dépourvu de signification objective, non seulement au sens du critère fort (décidabilité) que nous utilisons pour l'instant, mais aussi au sens du critère Popperien plus faible (falsifiabilité) auquel nous reviendrons par la suite.

2 - ou au moins falsifiables, lorsque nous aurons assoupli notre critère.

de modèles et de sous-modèles $\mathcal{F}_0 \supset \mathcal{F}_1 \supset \dots$ allant du plus général au plus particulier. En pareil cas, nous disons que \mathcal{F}_0 est un modèle générique, et un sous-modèle $\mathcal{F}_1 \subset \mathcal{F}_0$, suffisamment restreint pour que l'identification d'un individu dans \mathcal{F}_1 ne fasse plus intervenir qu'un petit nombre de paramètres (deux ou trois par exemple, rarement beaucoup plus), constitue ce que nous appellerons un type du modèle générique \mathcal{F}_0 .

Par exemple, la famille des F.A. stationnaires et gaussiennes constitue un modèle générique \mathcal{F}_0 . Pour définir une F.A. $Z \in \mathcal{F}_0$, il faut se donner une espérance m et une fonction de covariance $\sigma(h)$. Le sous-modèle \mathcal{F}_1 des F.A. stationnaires gaussiennes à covariance exponentielle $\sigma(h) = \sigma^2 \exp(-a|h|)$ constitue un sous-modèle ou type de F.A. dans \mathcal{F}_0 , dans lequel un individu est identifié par trois paramètres (l'espérance m , la variance σ^2 , et le paramètre d'échelle a de la covariance).

Au lieu de F.A., on considèrera très souvent des classes d'équivalence de F.A., et on appellera encore modèle générique une famille \mathcal{F} de telles classes d'équivalence. Par exemple, dans les problèmes d'estimation linéaire, on n'a besoin de connaître que les moments d'ordre 1 et 2 d'une F.A. On appellera alors F.A. d'ordre 2 la classe de toutes les F.A. ayant une espérance $m(x)$ et une covariance $\sigma(x,y)$ données. Ainsi, la famille des F.A. stationnaires d'ordre 2 constituera un modèle générique \mathcal{F}_0 ou un individu est spécifié par la donnée d'une espérance m et d'une covariance stationnaire $\sigma(h)$. Les F.A. stationnaires d'ordre 2 à covariance exponentielle constituent alors un type \mathcal{F}_1 dans \mathcal{F}_0 , où un individu est défini par 3 paramètres seulement (m , σ^2 et a) : mais cet "individu" est en réalité lui-même une classe de F.A. (que l'on ne distingue pas les unes des autres) : nous dirons qu'il constitue un modèle spécifique ou une spécification du modèle générique \mathcal{F}_0 (et du type \mathcal{F}_1).

Cette hiérarchie à 3 niveaux (modèle générique, type et spécification) peut paraître arbitraire. Elle se relève utile à l'usage, parce que, le plus souvent, le choix du modèle générique relève d'une simple décision méthodologique, tandis que le choix du type a presque toujours une portée objective : c'est en général au moment où l'on choisit (in praxi) le type dans un modèle générique que l'on introduit l'hypothèse anticipatrice la plus vulnérable et la plus féconde. La spécification (in praxi), c'est-à-dire si l'on veut, l'estimation des quelques paramètres dont dépend le type, à partir de l'information disponible, entraîne évidemment elle aussi, des risques sérieux d'erreurs (c'est le problème classique de "l'inférence statistique" mais de nature peut-être moins fondamentale. En effet, une fois que nous avons choisi le type de modèle, la statistique mathématique (appliquée dans le cadre de ce type) peut nous indiquer si l'estimation des paramètres peut être envisagée dans des conditions raisonnables ou non sur la base de l'information actuelle. Si elle ne l'est

pas, cela veut dire que nous ne pouvons pas résoudre le problème qui nous intéresse dans le cadre de ce modèle générique et de ce type, et nous devons chercher une autre formulation. Notons encore que, pour des raisons faciles à comprendre, l'information qualitative (non numérique) intervient de manière essentielle au moment du choix du type, tandis que la spécification se fait, le plus souvent, sur la base de la seule information numérique.

En résumé, un modèle générique se caractérise par son extension et ses critères de spécification. L'extension : c'est simplement l'ensemble des F.A. appartenant à l'une des classes du modèle. Par exemple, dans le modèle générique "F.A. stationnaire d'ordre 2", l'extension est constituée de toutes les F.A. admettant des moments stationnaires d'ordre 1 et 2. De leur côté, les critères de spécification servent à définir (ou spécifier) les classes du modèle générique, appelées aussi modèles spécifiques. Dans notre exemple, un modèle spécifique est constitué de toutes les F.A. admettant une même espérance et une même covariance, de sorte que les critères de spécification sont constitués par la donnée d'une constante m et d'une fonction $\sigma(h)$ de type positif. A mi-chemin, nous rencontrons la notion de type, ou sous-modèle, dans lequel les critères de spécifications sont réduits à un petit nombre de paramètres (par exemple : les F.A. stationnaires d'ordre 2 à covariance exponentielle).

En ce qui concerne les critères de spécification, il est important de savoir s'ils ont, ou non, une signification objective. De ce point de vue, l'exemple précédent n'est pas très heureux, car ni m ni $\sigma(h)$ ne sont strictement identifiables à des grandeurs régionales. De même, les paramètres m , σ^2 et a , qui interviennent dans le type à covariance exponentielle, ne sont pas des paramètres strictement objectifs (sauf peut-être si le champ S est très grand vis-à-vis de $1/a$). Par contre, nous verrons que le produit $a\sigma^2$ (qui représente la pente du variogramme à l'origine) a une signification objective.

Du point de vue méthodologique, notons aussi qu'adopter un modèle générique donné implique que l'on renonce à distinguer entre elles les différentes F.A. appartenant à un même modèle spécifique, c'est-à-dire répondant aux mêmes critères de spécification. Comme conséquence immédiate, on n'a plus le droit d'utiliser par la suite d'autres instruments de travail que les critères de spécification eux-mêmes. Par exemple, dans le cas du modèle générique "F.A. stationnaire d'ordre 2", les seuls instruments de travail autorisés seront l'espérance m et la covariance $\sigma(h)$; cela signifie, par exemple, en ce qui concerne le problème de l'estimation, que l'on

devra se restreindre à la classe des estimateurs linéaires affines (de la forme $Z^* = \lambda_0 + \sum \lambda_i Z(x_i)$), et, parmi les caractéristiques probabilistes des estimateurs de cette forme, n'utiliser que leurs espérances et leurs variances, qui, seules, s'expriment à l'aide de m et de $\sigma(h)$. En toute rigueur, il conviendrait même d'être plus sévère encore, et, parmi les estimateurs de ce type, se limiter à ceux dont les caractéristiques (espérance et variance) sont susceptibles d'être redéfinies après coup en termes de grandeurs régionales. Dans la pratique, il y a souvent bien des accommodements avec cette règle un peu trop stricte, mais, en gros, on s'efforce de la satisfaire de façon plus ou moins approchée.

Premières indications sur les critères du choix.

Il n'est pas possible, ici, d'examiner de manière approfondie les règles et les précautions qu'il convient d'observer au moment de choisir (in praxi !) un modèle, et nous nous contentons de donner quelques indications sommaires. Il est entendu, aussi, que nous nous plaçons résolument du point de vue monoscopique, qui est en général celui qu'adopte le praticien lorsqu'il doit prendre une décision. Dans une recherche plus désintéressée, on viserait plutôt un modèle aussi panscopique que possible, donc présentant un degré de spécification aussi poussé que cela est possible sur la base de l'information disponible : car un modèle est évidemment d'autant plus riche et mieux apte à représenter des aspects plus variés de la réalité qu'il contient davantage de critères de spécification. Le praticien, au contraire, pour se garantir au mieux contre le risque de prendre une décision erronée, doit affaiblir le plus possible les hypothèses anticipatrices sur lesquelles il est bien obligé de se fonder. Il recherche, par conséquent, le degré de spécification le plus faible possible compatible avec la résolution effective du problème posé.

Dans tous les cas, le modèle doit vérifier les quatre conditions suivantes :

a - Il doit être opérationnel : une fois spécifié, le modèle doit permettre d'obtenir une solution effective, numérique, du problème que l'on veut résoudre. Il en résulte (ce qui est bien conforme à notre point de vue monoscopique) que la définition du modèle à utiliser est liée étroitement à la nature de l'objectif que l'on poursuit. C'est là un point méthodologique fondamental, sur lequel nous reviendrons. Le principe d'économie qui préside au choix des modèles monoscopiques peut s'énoncer sous la forme suivante : les critères de spécification doivent contenir tout ce qu'il faut pour résoudre le problème posé, et il est inutile qu'ils contiennent davantage.

b - Il doit être spécifiable. Sous une forme un peu vague, que nous essaierons de préciser un peu par la suite, cela signifie qu'il doit être "raisonnablement" possible, sur la base de l'information (numérique et qualitative) disponible, de choisir le type du modèle et d'estimer les valeurs numériques des paramètres qui le spécifient.

c - Il doit être compatible avec les données (numériques et qualitatives) en un sens que nous devons préciser progressivement. Les problèmes que posent cette condition et la précédente sont analogues (mais non identiques) à ceux que la statistique classique connaît sous le nom de "tests d'hypothèses" et "inférence statistique" des paramètres. Ils ne relèvent qu'en partie de la statistique mathématique. Car, d'une part, si un paramètre est dépourvu de signification objective, il ne doit pas non plus influencer sensiblement la solution d'un problème réel (car un problème réel s'énonce et se résout en termes opératoires, donc de grandeurs régionales). Par suite, ce paramètre peut être choisi de façon relativement arbitraire : le fait que l'inférence statistique ne soit pas possible perd une grande part de son importance. D'autre part, en sens inverse, si un "test d'hypothèse" conclut à l'acceptation du type de modèle choisi, c'est-à-dire à sa non incompatibilité avec les données actuelles, cela ne signifie pas qu'une autre sorte de test n'aurait pas conclu au rejet. Et surtout, rien ne nous garantit que les données futures, celles dont nous ne disposons pas encore, confirmeront notre conclusion. Le risque d'erreur radicale est toujours présent. Enfin, la vérité nous oblige à reconnaître qu'il est rare, dans la pratique, que l'on puisse procéder à des tests vraiment sophistiqués : soit, simplement, parce qu'il n'en existe pas ; ou, s'il en existe, parce qu'ils sont trop compliqués ; ou encore, parce que leur mise en oeuvre nécessiterait une information dont nous ne disposons pas (pour tester des moments d'ordre deux, il faudrait, par exemple, déjà connaître les moments d'ordre quatre etc...).

d - Il doit être efficace, c'est-à-dire conduire à une solution correcte (en un sens au moins statistique) du problème posé. Ce point ne peut, évidemment, être vérifié qu'après coup. Mais l'expérience acquise à l'occasion de cas analogues joue aussi un rôle important : la sanction de la pratique, semblable à la sélection naturelle, élimine à la longue impitoyablement les méthodologies inefficaces et les modèles inadéquats.

En ce qui concerne le point a/ , notons l'importance décisive du choix de la méthode adoptée pour résoudre un problème donné. Le problème est ce qu'il est,

c'est une contrainte. Mais nous avons en général le choix, pour le résoudre, entre plusieurs méthodes, plus ou moins puissantes, plus ou moins approchées et nécessitant aussi, pour leur mise en oeuvre, plus ou moins de prérequisites. La première tâche à entreprendre est de procéder, pour chaque méthode possible, à une analyse stricte des prérequisites minimaux qu'elle exige réellement (cet élagage révèle souvent que certains paramètres, que l'on croyait indispensables, sont en fait tout-à-fait superflu).

Comme on a intérêt à avoir un modèle aussi peu spécifié que possible, on cherchera à faire coïncider, autant que faire se peut, les critères de spécification du modèle avec les prérequisites minimaux exigés par la méthode choisie : ceci est une des règles directrices dans le choix du modèle générique. Les critères de spécification doivent contenir les prérequisites minimaux, et il est inutile qu'ils contiennent rien d'autre.

Plus une méthode est puissante, plus elle exige de prérequisites minimaux, et plus, par conséquent, le modèle utilisé doit être spécifié. Mais il y a toujours une limite, puisque les données disponibles ne permettent jamais de dépasser un certain degré de spécification. D'où la deuxième règle, relative cette fois au choix de la méthode à utiliser pour résoudre un problème donné : choisir la méthode la plus puissante parmi celles dont les prérequisites minimaux peuvent être raisonnablement spécifiés à partir des données disponibles.

La troisième règle est, évidemment, celle de la non-incompatibilité du modèle avec les données expérimentales et notre intuition physique du phénomène. Sa compatibilité avec la réalité elle-même, par contre, n'est jamais absolument assurée, même si l'examen critique de la possibilité de la spécification (i.e. de l'estimation des grandeurs régionales requises) a conduit à une réponse positive. On n'est jamais complètement à l'abri de mauvaises surprises. Mais on n'a pas le choix, il faut bien, à un moment donné, faire une hypothèse permettant de relier d'une manière ou d'une autre la structure de la partie inconnue du phénomène à celle des données disponibles.

Le modèle primaire

Dans les pages qui précèdent, nous avons raisonné comme s'il était possible d'identifier la réalité à la simple collection des valeurs numériques de la variable régionalisée. Mais celle-ci n'épuise jamais la richesse du réel. Il y a toujours

autre chose que les valeurs numériques ponctuelles. Une chaîne de montagnes peut évidemment se définir comme une collection de cotes topographiques, mais ce point de vue, analytique et exhaustif, n'est pas forcément le plus intéressant : il y a aussi la structure géologique d'ensemble de la chaîne, son histoire, sa tectonique, etc... La simple carte topographique ne suffit plus, et nous devons introduire autant de V.R. qu'il y a de grandeurs géologiques à considérer, nous devons aussi conserver constamment en mémoire l'aspect évolutif du phénomène, sa genèse, même si l'observation directe ne nous fournit rien de tel. En un sens, donc, le réel est inépuisable. Toutefois, les problèmes - techniques et modestes - que nous avons en vue sont de nature estimatoire ou interpolatoire, et s'accrochent assez bien d'une réduction de toute la richesse de la réalité à la donnée d'une ou plusieurs fonctions z sur S (les V.R.). Le reste - les idées que nous pouvons nous faire sur la genèse et la structure du phénomène, et, plus généralement l'intuition physique que nous nous en formons - n'en continue pas moins à jouer en coulisse un rôle des plus importants. C'est en général dans ce trésor archétypique que nous aurons des chances de trouver les schèmes ou les principes moteurs de modèles vraiment adaptés. Nous nous garderons donc du purisme qui consiste à identifier le réel à une collection de valeurs numériques. D'abord, parce que les valeurs numériques ne sont pas le réel, mais une première image (analytiquement très riche, structurellement très pauvre) de celui-ci. Ensuite, parce que les schèmes de l'intuition physique nous fournissent exactement le fil directeur dont nous avons besoin pour nous orienter dans le problème du choix d'un modèle. Mais la réduction analytique du réel à une collection de valeurs numériques, seule, fournit les bases d'une méthodologie opératoire. Parmi les différents modèles possibles, et les concepts qu'on peut leur rattacher, nous ne retiendrons comme objectifs (non métaphysiques) que ceux qui sont entièrement explicites en termes de grandeurs régionales (i.e. : déterminés par la donnée d'une ou plusieurs V.R. sur le champ S).

Il en est ici un peu comme en logique : on sait bien que, dans un concept, il y a beaucoup plus que son extension, à savoir l'élément moteur, le mouvement dialectique, etc... Mais le traitement formalisé rigoureux n'est possible que si l'on accepte d'identifier le concept à son extension. C'est à ce prix seulement que nous évitons les illusions métaphysiques et que nous obtenons un formalisme opératoire. Le riche contenu du concept, fascinant et insaisissable, s'est évanoui, en apparence, pour faire place à une méthodologie rigoureuse, opératoire, glacée, et un peu ennuyeuse. Mais en réalité rien n'est perdu. Car l'élément moteur du concept continue secrètement à inspirer nos démarches (et sans lui, sans doute, nous n'entreprendrions rien du tout). Mais ces démarches, et leurs résultats, sont aussitôt traduits dans la langue du formalisme le plus épuré, le plus rigoureux, seul officiellement reconnu.

Quoiqu'il en soit, la variable régionalisée n'est pas identique à la réalité, mais constitue déjà elle-même un modèle primaire. C'est pourquoi : le critère de simple décidabilité en termes de régionales se révèle insuffisant dès que le besoin d'une analyse plus fine se fait sentir : il faut alors revenir au critère Popperien, fondé sur la seule falsifiabilité.

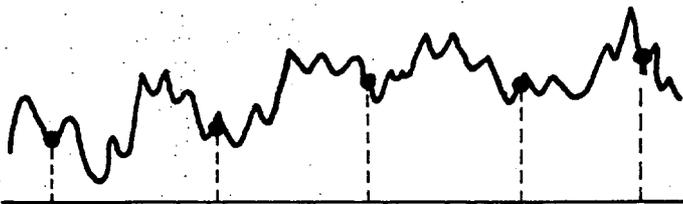
Cette nécessité se manifeste de manière impérieuse, chaque fois qu'il faut tenir compte de la signification physique des grandeurs que nous manipulons, et des limites à l'extérieur desquelles les concepts qui leur correspondent cessent d'être opératoires. De fait, il n'existe dans la nature ni points $x \in S$, ni valeurs numériques $z(x)$ univoquement affectées à chaque point x . Pour un physicien, en effet, il n'y a pas de points, au sens mathématique, mais seulement des éléments de volume, petits vis-à-vis de l'échelle où l'on a choisi de travailler, mais grands vis-à-vis des échelles inférieures pour lesquelles le phénomène que l'on étudie cesse d'être définissable en termes opératoires. Si, par exemple, $z(x)$ est censée représenter la teneur en un métal donné en un "point" x donné d'un gisement minier, le volume élémentaire δv de référence sera le plus souvent choisi comme grand vis-à-vis des dimensions granulométriques. Si l'on veut étudier le phénomène à cette dernière échelle, ce qui est possible, la teneur $z(x)$ prendra une allure en tout ou rien, à peu près nulle dans les grains de stérile, à peu près constante dans les grains minéralisés. Dans tous les cas, elle dépendra de la constitution minéralogique du minerai. Mais, même à cette échelle granulométrique, les volumes élémentaires δv , petits vis-à-vis des grains, restent très grands vis-à-vis des dimensions atomiques ou infra atomiques : à l'échelle de ces dernières, d'ailleurs, on ne pourrait plus observer qu'un chaos de fonctions d'ondes, le modèle usuel de l'espace euclidien à 3 dimensions cesserait d'être opératoire et il ne saurait plus être question ni de points $x \in S$ ni de teneur $z(x)$ univoquement définie.

Ainsi, comme tout concept physique, celui de V.R. est soumis à des limitations assez strictes en dehors desquelles il cesse d'être opératoire. De ce point de vue, nous ne devons pas prendre au pied de la lettre des énoncés purement mathématiques concernant par exemple la continuité ou la dérivabilité de la V.R., ou plus généralement des propriétés topologiques faisant intervenir tous les voisinages, si petits soient-ils, d'un point x donné. Il s'agit là, toujours, de caractéristiques du modèle, et non, en toute rigueur, de propriétés de la réalité physique, puisque, en deçà d'un certain seuil de petitesse, le phénomène s'évanouit en quelque sorte ou change de nature.

Cependant, le contraste des échelles est tel que nous pouvons, en première approximation, négliger ces effets de flou (qui ne se manifestent en aucune façon à notre échelle de travail) et considérer, en première approximation, le modèle "V.R." comme bien représentatif de la réalité. Mais, même ainsi, des difficultés sérieuses subsistent. En effet, l'ensemble des valeurs numériques $\{z(x), x \in S\}$ est alors infini, et par conséquent n'est pas accessible à une connaissance expérimentale exhaustive. Ainsi, un énoncé comme "z est continu, ou dérivable, au point x" faisant intervenir une infinité (et même non dénombrable) de points et de valeurs numériques, échappe, en toute rigueur, à toute possibilité de contrôle expérimental effectif, et par conséquent doit être considéré comme dépourvu de signification objective. De fait, quelle que soit la "vraie" fonction z (si tant est qu'elle soit définissable), nous pouvons toujours trouver une fonction ϕ continue, dérivable etc... qui l'approche suffisamment pour en être expérimentalement indiscernable. Car on ne connaîtra jamais réellement les $z(x_i)$ que sur une maille $\{x_i\}$ aussi serrée que l'on veut, mais toujours finie, et on pourra toujours les interpoler par une fonction ϕ aussi régulière que l'on veut.

Naturellement, du point de vue physique, la notion clé est ici celle d'échelle. La fonction ϕ ainsi choisie sera très régulière (dérivable, etc.. au sens strict des mathématiciens), mais seulement à l'échelle de la maille très serrée définie par les $\{x_i\}$. Cette très haute micro-régularité n'empêche en rien cette fonction de se comporter, pratiquement, à une échelle plus élevée comme une fonction très irrégulière, voire comme un pur bruit blanc.

Si l'on envisage le phénomène à cette échelle, il serait physiquement faux (bien que mathématiquement vrai) de considérer la fonction ϕ comme continue et dérivable.



Ainsi, les notions mathématiques les plus usuelles, comme la continuité, la dérivabilité etc..., doivent faire l'objet d'une reconstruction opératoire. Pour traiter un problème qui se pose (par exemple) à l'échelle du mètre ou de la centaine de mètres, il suffit de connaître $z(x)$ à une maille centimétrique, ou mieux, de connaître, à cette maille, la régularisé de $z(x)$ sur un support centimétrique : les données physiques, en effet, correspondent à de petites intégrales, plutôt qu'à des valeurs strictement ponctuelles. On peut ainsi en principe, obtenir une définition physiquement exhaustive de la V.R., au moyen d'un nombre, peut-être élevé, mais fini de valeurs numériques, expérimentalement mesurables. C'est à cette échelle (centimétrique sur notre exemple) que l'on doit procéder à la redéfinition opératoire des concepts mathématiques.

Seuil de robustesse et seuil de réalisme

La "dialectique" de la V.R. et du modèle présente ensuite un aspect important : c'est que le modèle, une fois choisi, va tendre à vivre de sa vie propre, qui nous échappe en partie, et ne coïncide plus qu'en partie avec notre intuition physique de la réalité et les caractéristiques des V.R. que nous avons choisies. Le modèle, en effet tend à déborder la réalité, en ce sens que, si nous nous abandonnons sans esprit critique au formalisme mathématique, si nous cédon à la tentation du passage à la limite, nous nous laisserons entraîner au delà du domaine où le modèle est opératoire. Autrement dit, certains résultats tirés du modèle par simple déduction mathématique correspondent à des propriétés réelles, d'autres non : et ceci, notons-le bien, même si le modèle a pu être spécifié après coup (sur la base de toutes les valeurs numériques de la V.R.). Ce seuil d'objectivité au-delà duquel on n'a pas le droit de pousser la déduction mathématique, sous peine d'engendrer artefacts et faux problèmes, est parfois difficile à définir avec précision, mais on conçoit bien qu'il existe toujours, et que le sens physique du chercheur joue, en pratique, un rôle important de son appréciation. La situation est aggravée in praxi du fait que nous risquons de nous tromper en adoptant des hypothèses anticipatrices objectives qui peuvent se révéler fausses après coup. Il apparaît ainsi, in praxi, un seuil de robustesse que l'on ne devrait pas franchir, ou peut-être deux seuils, puisque l'on peut distinguer une robustesse de type et une robustesse spécifique, liées respectivement à la possibilité d'erreurs commises dans le choix du type et dans l'estimation des paramètres qui le spécifient. D'une manière générale, une opération peut être dite robuste (vis à vis d'une cause d'erreur) si son résultat n'est que faiblement affecté par cette cause d'erreur. La robustesse spécifique peut être étudiée de manière quantitative, puisque l'on peut toujours faire varier les paramètres du modèle autour des valeurs choisies, et examiner dans quelle mesure ces écarts se répercutent sur les résultats d'une opération donnée. La robustesse de type, liée à l'incertitude d'un choix structural, est plus difficile à étudier, car il existe en général un nombre illimité de types imaginables, à peu près également compatibles avec les données et l'information disponibles. Ici aussi, le bon sens, l'intuition physique et l'expérience acquise ont leur mot à dire.

En ce qui concerne ces seuils, ce qui importe avant tout, c'est de toujours conserver présente à l'esprit la distinction claire entre caractéristiques du modèle et propriétés physiques du phénomène réel. Dans le cadre de l'interprétation probabiliste, ceci revient à peu près à distinguer soigneusement la F.A. elle-même (notion mathématique), définie par sa loi spatiale, ou caractérisée simplement, en

pratique, par une fonction de covariance ou un variogramme, d'une part, et de l'autre la réalisation particulière, unique et limitée à un domaine S, connue dans sa totalité (après coup), ou, plus souvent (in praxi) en partie seulement. Cette réalisation n'est pas autre chose que la V.R. initiale, réinterprétée dans un cadre probabiliste. Mais, même dans ce cadre, il n'y a nullement identité entre la F.A. et sa réalisation : même si le modèle théorique est stationnaire et ergodique, du seul fait que la réalisation est restreinte à un domaine S borné, il n'y a pas identité entre les grandeurs régionales (interprétées maintenant comme des réalisations de variables aléatoires) et leurs valeurs "théoriques" dans le modèle, qui sont les espérances correspondantes. Or, seules les grandeurs régionales, calculées sur la réalisation, représentent la réalité physique et ont une signification objective. Leur contrepartie théorique n'est qu'un paramètre conventionnel (non objectif, à moins que, par chance, ce paramètre ne se trouve coïncider presque sûrement dans le modèle avec la valeur régionale correspondante).

On regrettera peut être que nous ayons mis l'accent sur les aspects pragmatiques, au détriment des aspects théoriques de la démarche topo-probabiliste. Il est indéniable que les problèmes évoqués dans ces pages, les méthodes et les modèles que l'on y décrit, semblent beaucoup plus proches du pôle monoscopique et instrumental que du pôle spéculatif et panscopique. Mais la manipulation n'est efficace que dans la mesure où le modèle est, d'une manière ou d'une autre, grosso modo adapté à la réalité ou à l'aspect de la réalité auquel nous nous intéressons. Les critères de spécification de nos modèles génériques, ou plutôt les grandeurs régionales qui leur correspondent (et que nous pouvons toujours considérer pour elle-mêmes indépendamment de l'interprétation probabiliste que nous choisissons ou non d'en faire) expriment aussi des propriétés structurelles du phénomène réel, et possèdent par elles-mêmes une riche signification physique. Ainsi, le comportement d'un variogramme au voisinage de l'origine exprime la plus ou moins grande régularité (continuité) d'une variation spatiale. Le variogramme du modèle théorique est, certes, dépourvu d'existence empirique, mais il n'en est pas de même de sa version régionale. Certes, le variogramme régional n'est pas donné tel quel. Il résulte d'une première réduction, ou modélisation du phénomène, auquel on a tout d'abord imposé le statut de V.R., puis d'une manipulation effectuée sur les données numériques ainsi fabriquées, (ou extraites,) à partir de la réalité. Il est réel, mais d'une réalité technique plutôt que naturelle. Il résulte de l'application d'un schème (ou procédé de construction). Les éléments moteurs de ce schème sont issus de notre intuition physique du phénomène, qui nous suggère le type de modèle (probabiliste ou non, etc...) auquel nous devons recourir. Ensuite, le modèle théorique s'anime de sa vie propre (qui est celle d'une fiction régulatrice), et nous suggère en retour le choix de la forme des

instruments mathématiques (les critères de spécification) les mieux aptes à donner une représentation numérique précise des caractéristiques physiques qui ont retenu notre attention. Le choix du modèle joue donc un rôle heuristique capital, mais, en définitive, ne sert que d'intermédiaire. Car, finalement, nous effectuerons des manipulations numériques sur des données numériques. Quelles que soient les intuitions physiques et les interprétations théoriques qui ont conduit au choix de ces opérations, ces manipulations peuvent et doivent être analysées pour elles-mêmes : nos problèmes se ramènent toujours à choisir des grandeurs régionales et à (essayer de) les estimer sur la base d'une information limitée. De ce point de vue, pas de miracle : quelle que soit la richesse de notre intuition et la subtilité de nos modèles, ce sont toujours les mêmes données numériques que nous malaxons et pétrissons indéfiniment, sans que ces opérations puissent, par elles-mêmes, engendrer la moindre information supplémentaire. De même qu'il existe in praxi un seuil de robustesse, il doit exister un seuil de réalisme qu'il serait illusoire de prétendre dépasser. Sur la base d'une information trop pauvre et limitée, on pourra former une solution approximative de certains problèmes simples. Mais on devra renoncer honnêtement à en résoudre d'autres plus complexes. Reconnaître l'existence de ces seuils de robustesse et de réalisme, et savoir les apprécier correctement dans chaque cas particulier, c'est là le propre du bon praticien, et ce sens physique est en définitive beaucoup plus important que la théorisation mathématique.

Un exemple de contrôle de la robustesse de type

Voici un exemple de la manière dont on peut contrôler l'influence du choix du type de modèle sur le résultat final que l'on a en vue. Le test porte sur la "variance d'extension" d'un point dans un segment de droite ou un carré, c'est-à-dire la variance de l'erreur commise en estimant la teneur moyenne du segment ou du carré à partir d'un échantillon ponctuel. Cette variance s'exprime à l'aide d'un modèle de covariance ou de variogramme dont on doit choisir le type et spécifier les paramètres. J'ai construit mon exemple de manière à obtenir un variogramme présentant, au voisinage de l'origine, un comportement de type inhabituel, et envoyé à divers géostatisticiens le texte qu'on lira ci-dessous : les démarches successives correspondent à la pratique réelle. L'usage, en particulier, de demander

des croix de sondages supplémentaires à maille plus serrée, pour préciser le comportement du variogramme au voisinage de l'origine est tout-à-fait courant, et les exploitants, le plus souvent accèdent à cette demande de géostatisticien malgré les frais que cela entraîne. C'est que, même sans être géostatisticien, un praticien comprend parfaitement l'importance de la structure aux courtes distances, et la nécessité de disposer d'une information relative à cette échelle. Voici ce texte :

"Test de robustesse"

" Ceci n'est pas un piège, mais un test pour évaluer la robustesse de nos méthodes d'ajustement de variogramme. Je demande à chaque destinataire d'ajuster un modèle de variogramme, et de donner les variances d'extension correspondantes, à partir des valeurs numériques suivantes :

A -

$\gamma_1 = 0,63$	Au-delà de 6, les valeurs diffèrent peu de
$\gamma_2 = 0,80$	1. L'unité de longueur choisie est égale à
$\gamma_3 = 0,89$	la maille des sondages (soit $u = 180$ m pour
$\gamma_4 = 0,94$	fixer les idées). Le nombre N des sondages
$\gamma_5 = 0,97$	est grand (plusieurs centaines), de sorte
$\gamma_6 = 0,99$	que les points expérimentaux peuvent être
	considérés comme très bons.

" Je demande à chacun de faire son ajustement exactement de la manière dont il procède dans la pratique réelle, et de le faire indépendamment des autres géostatisticiens et sans se laisser influencer par eux.

" Avec le modèle ajusté, je demande de calculer la variance d'extension d'un point central dans le segment unité et dans le carré unité.

B -

" A la demande du géostatisticien, une croix de 20 sondages à maille 1/9 (soit 20 m) est exécutée sur une zone bien représentative, et donne :

$\gamma_{1/9} = 0,25$	
$\gamma_{2/9} = 0,34$	
$\gamma_{3/9} = 0,40$	Modifiez-vous votre ajustement initial ?
$\gamma_{4/9} = 0,46$	Quelles sont les nouvelles variances d'ex-
$\gamma_{5/9} = 0,50$	tension ?

C -

"Par la suite, on resserre la maille, qui passe à 90 m, soit 1/2 avec notre unité, et on trouve

$$\gamma_{1/2} = 0,49$$

Variances d'extension pour la maille 1/2 ?

N.B. : Ces valeurs numériques sont déduites d'un schéma de type inhabituel. L'intérêt du test est donc double : il mettra en évidence la dispersion des évaluations individuelles, et leurs écarts avec celles que l'on déduit du schéma en question (ces dernières ne sont pas les "vraies" valeurs, puisque, dans un cas réel, rien ne nous garantirait que ce modèle particulier soit le "vrai" modèle)."

Les 10 réponses que j'ai reçues présentent des ajustements assez notablement différents. En général, le premier modèle était du type "effet de pépite" plus un schéma exponentiel ou sphérique (ou même 2 sphériques superposés). Après obtention de la croix de sondages, le modèle s'enrichit et devient soit : pépite + sphérique + sphérique, ou pépite + exponentiel + sphérique ou pépite + exponentielle + exponentielle (et même pépite + 3 schémas sphériques) avec d'ailleurs des paramètres forts différents d'un modèle à l'autre. D'autres modèles aussi ont été proposés, dont l'un différait peu du mien. Graphiquement, toutefois ces différents ajustements sont peu différents et conduisent à des valeurs fort peu dispersées pour les variances d'extension.

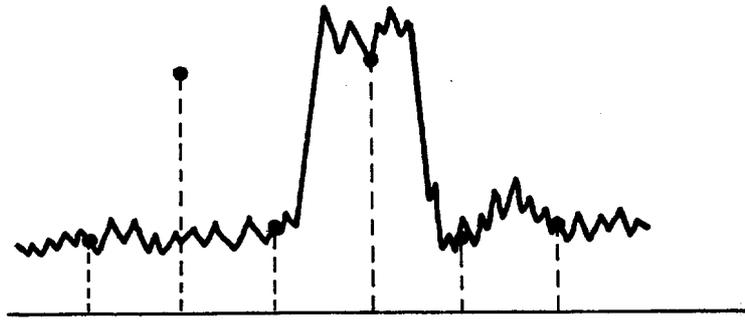
Ces valeurs sont données ci-après. Pour le premier modèle, ajusté sans tenir compte de la croix de sondages, elles sont assez dispersées, et en général un peu trop fortes : en l'absence d'information sur les microstructures, le géostatisticien choisit, pour des raisons de sécurité, l'hypothèse d'un effet de pépite assez fort, donc surestime délibérément les variances d'estimation. Dès le réajustement effectué à l'aide de la croix de sondages, les modèles utilisés, quoique très différents les uns des autres, conduisent pratiquement aux mêmes valeurs numériques.

<u>Modèle A</u>		<u>Modèle B</u> (maille 1)		<u>Modèle C</u> (maille 1/2)	
					
0.377	0.405	0.309	0.364	0.231	0.272
0.377	0.412	0.321	0.381	0.243	0.288
0.330	0.370	0.330	0.370	0.232	0.284
0.372	0.405	0.307	0.370	0.227	0.273
0.409	0.430	0.302	0.376	0.232	0.262
0.370	0.405	0.303	0.367	0.225	0.272
0.396	0.430	0.303	0.380	0.227	0.265
0.353	0.388	0.304	0.364	0.225	0.262
0.347	0.372	0.306	0.368	0.237	0.266
0.369	0.402	0.313	0.346	0.232	0.272

La robustesse vis-à-vis des données

D'une manière générale, une opération est robuste si son résultat est peu influencé par les erreurs que nous avons pu commettre soit dans le relevé des données expérimentales, soit dans le choix et (ou) dans la spécification du modèle générique. Il y a donc trois types de problèmes et trois types de robustesse à envisager : l'influence des données dites aberrantes, celle d'une erreur dans la spécification du type de modèle, celle enfin d'une erreur dans le choix du type lui-même, et corrélativement : robustesse des données, robustesse spécifique, robustesse de type. Nous avons suffisamment parlé des deux dernières. Il nous reste à dire quelques mots au sujet de la première, c'est-à-dire examiner la robustesse de données et le problème des données "aberrantes".

Parmi les données disponibles, certaines peuvent être entachées d'erreurs grossières (erreurs de transcription, virgule mal placée, mesures fausses pour une raison ou pour une autre) et doivent autant que possible être éliminées. Il est souvent commode, lorsque les données sont nombreuses, de disposer de procédés automatiques, permettant d'individualiser les mesures "suspectes" candidates à une élimination éventuelle. Des procédés très simples (histogrammes, nuages de corrélation) donnent souvent de très bons résultats. Mais une donnée "suspecte" n'est pas forcément fausse ou "aberrante" et ne doit jamais être éliminée automatiquement. Elle peut très bien refléter le comportement réel du phénomène (ex. de la figure 3).



Que ce comportement nous paraisse "anormal" ou "aberrant" c'est là de notre part un jugement purement subjectif. La réalité est ce qu'elle est, mais le modèle (plus ou moins stationnaire) qui nous sert, implicitement, de pierre de touche peut être mauvais. Ce n'est pas, ici, la donnée qui est aberrante, mais plutôt la stationnarité qui est mise en défaut. Eliminer cette donnée serait une erreur, qui nous ôterait par la suite toute possibilité de prévoir l'occurrence de semblables accidents en des zones non informées.

La robustesse des données a fait l'objet d'études très approfondies, dont les résultats sont sans doute en partie transposables aux modèles topo-probabilistes. Il ne s'agit pas seulement, d'ailleurs, de se prémunir contre l'effet des données aberrantes, mais aussi contre l'influence très exagérée qu'exercent parfois, sur les estimateurs usuels, les valeurs expérimentales extrêmes (très fortes ou très faibles) qui se manifestent lorsque l'on a des lois de distribution très éloignées du type gaussien. Par exemple, dans le cas de certaines distributions symétriques autour d'une valeur centrale θ , on montre que la médiane est un estimateur de θ plus robuste (moins sensible à l'effet des valeurs extrêmes) que la moyenne arithmétique.

Toutefois, si on a amélioré la robustesse vis-à-vis des données, on a peut-être aussi détérioré la robustesse vis-à-vis du modèle (nous devons maintenant "supposer" la loi symétrique etc...). C'est là une objection sérieuse contre ces estimateurs robustes, sophistiqués, que nous présente la statistique mathématique : ils sont, souvent, moins robustes, vis-à-vis du modèle, que des estimateurs plus simples. De plus, ils sont toujours conçus pour estimer un paramètre (en général non objectif) du modèle (une valeur centrale etc...) et non une grandeur régionale. De sorte qu'il n'est certainement pas possible de les transposer tels quels à nos modèles topo-probabilistes. Il faudrait étudier le problème plus général de la robustesse de l'estimation d'une grandeur régionale à la fois vis-à-vis des données, du type et de la spécification. Ce n'est certainement pas un problème simple.

On se rend compte qu'un estimateur robuste (vis-à-vis des données) comme la médiane, n'est pas forcément adéquat pour l'estimation, disons, par exemple, de la moyenne régionale $\bar{z} = (1/S) \int_S z(\mathbf{x}) \, d\mathbf{x}$, en remarquant ceci : notre estimateur, de la forme :

$$z^{\mathbf{x}} = \phi(z_1, \dots, z_n)$$

s'il est optimal en un sens quelconque, doit certainement vérifier la condition suivante : si la maille constituée dans S par les points expérimentaux se resserre de plus en plus (N tendant vers l'infini, et les points expérimentaux recouvrant à la limite le champ S entier), l'estimateur $z^{\mathbf{x}}$ doit converger vers la valeur réelle \bar{z} , et la variable $z^{\mathbf{x}}$ qui lui est associée dans le modèle doit converger presque sûrement vers \bar{z} : en effet, lorsque tous les points sont connus, le meilleur estimateur de \bar{z} est \bar{z} lui-même, qui est devenu expérimentalement accessible. Or, manifestement, un estimateur comme la médiane, si robuste soit-il, ne vérifie pas cette condition. Donc, à supposer que la médiane soit réellement un meilleur estimateur que la moyenne arithmétique lorsque la maille est grande et que le nombre N des points expérimentaux est petit, il arrive certainement un moment où, la maille étant plus serrée et N plus grand, la conclusion se renverse : la moyenne arithmétique reprend le dessus, du moins en tant qu'estimateur de la régionale \bar{z} .

- CHAPITRE V -

FAIRE LE TRI

Lorsque l'on a choisi un modèle (générique) il reste à faire le tri, je veux dire examiner, parmi les divers critères et paramètres spécifiques du modèle, lesquels correspondent à des réalités et lesquels ne jouent qu'un rôle purement conventionnel. Pour la commodité de l'exposé, je distinguerai deux étapes dans cette recherche de l'objectivité. La première, qui constitue le tri proprement dit, est examinée dans ce chapitre. C'est une critique interne du modèle, effectuée en se plaçant du point de vue du modèle lui-même. Elle consiste à rechercher, parmi les variables aléatoires que le modèle associe aux diverses régionales, celles qui, dans le modèle, coïncident presque sûrement (= avec une probabilité unité) avec leur espérance mathématique : ces espérances constituent alors, en effet, des paramètres objectifs. Ce point de vue nous conduira à examiner deux notions importantes : l'une classique, est l'ergodicité ; l'autre, moins classique, mais non moins importante pour notre propos, recevra le nom de microergodicité. Il nous faudra naturellement aussi examiner comment les choses se présentent in praxi, c'est-à-dire aborder le problème de l'estimation de ces grandeurs objectives (c'est-à-dire régionales) à partir d'une information fragmentaire. Après ce tri, ou critique interne du modèle, viendra la deuxième étape, celle de la reconstruction des concepts et de la reformulation du modèle lui-même en termes opératoires. Ce sera l'objet de la troisième partie.

Les "fluctuations" des régionales

Du point de vue du modèle, la variable régionalisée z , définie sur le champ S , est considérée comme une réalisation d'une fonction aléatoire Z . Mais cela ne signifie nullement que la priorité épistémologique soit passée de la première à la seconde : c'est toujours la V.R., et les régionales qui s'en déduisent, qui fondent l'objectivité. En cas de "désaccord" entre le modèle et la V.R., c'est cette dernière qui "a raison" : le modèle, lui, ne dit rien de plus que ce que suggèrent les décisions (peut-être maladroitement) et les hypothèses (peut-être fausses) qui ont guidé son choix. Par exemple, dans un modèle non stationnaire, on interprète souvent l'espérance

$$m(x) = E[Z(x)]$$

de la F.A. au point x comme représentant la "tendance" ou la "dérive" de la variable régionalisée. Cette interprétation est parfois utile. Mais, du moins tant que l'on n'a pas reconstruit cette notion en termes opératoires, il ne s'agit que d'une caractéristique conventionnelle du modèle. Il ne faudrait pas croire, en particulier, que $m(x)$ représente la "vraie" valeur de la V.R., dont la valeur numérique réelle

$z(x)$ ne s'écarterait que par suite d'une "erreur de la nature". La situation est, toutefois, en partie différente lorsque les valeurs dont on dispose in praxi sont des mesures, entachées d'erreur, d'une V.R. inaccessible.

Ce que nous venons de dire à propos des valeurs ponctuelles s'applique de la même façon à toutes les grandeurs régionales. Considérons, en effet, une régionale donnée $a = \mathcal{O}(z)$, où \mathcal{O} est une fonctionnelle de la variable régionalisée z sur S . Dans ce qui suit, nous considérerons surtout les deux exemples simples suivants : celui de la moyenne (régionale) de la V.R. dans son champ S , soit :

$$(1) \quad \bar{z} = \frac{1}{S} \int_S z(x) \, dx$$

et celui du variogramme régional γ_R : il s'agit de la fonction donnant, pour chaque valeur de l'argument vectoriel h , la demi-valeur moyenne du carré de la différence des valeurs prises par la V.R. en deux points x et $x+h$, soit

$$(2) \quad \gamma_R(h) = \frac{1}{2K(h)} \int_{S(h)} [z(x+h) - z(x)]^2 \, dx$$

Comme les deux points x et $x+h$ doivent appartenir à S , le domaine d'intégration n'est plus S lui-même, mais l'intersection de S avec l'ensemble qui s'en déduit par la translation $-h$, domaine désigné par $S(h)$. Quant à $K(h)$, il représente la mesure (surface, si l'espace est à 2 dimensions, volume s'il en a trois etc...) du domaine $S(h)$ d'intégration. Comme l'usage réserve la notation γ au variogramme "théorique" c'est-à-dire (dans le modèle "F.A. intrinsèque") à l'espérance :

$$\gamma(h) = \frac{1}{2} E [(Z(x+h) - Z(x))^2]$$

j'ai mis la lettre R en indice dans $\gamma_R(h)$ pour rappeler qu'il s'agit de la grandeur régionale associée au variogramme. Il ne faut pas dire que γ_R constitue la version régionale de γ , ce qui serait inverser la priorité épistémologique, mais plutôt au contraire que γ est la version théorisée de γ_R .

Si, maintenant, dans la définition $a = \mathcal{O}(z)$, nous substituons la F.A. Z à la V.R. z , nous obtenons une variable aléatoire $A = \mathcal{O}(Z)$, au lieu d'une grandeur numérique a . Dans le cadre de l'interprétation probabiliste, a est alors considérée comme une réalisation de la V.A. ainsi définie, $A = \mathcal{O}(Z)$. Cela ne signifie en aucune façon que l'espérance $\alpha = E(A)$ de cette V.A. (qui est une caractéristique du modèle) doit être considérée comme "plus vraie" ou plus "réelle" que la valeur effective

$a = \mathcal{Q}(z)$ (seule objective), ce qui reviendrait, en somme, à admettre que la réalité a tort et que le modèle a raison. Dans notre premier exemple, la moyenne régionale \bar{z} est associée à la V.A. \bar{Z} dont l'espérance (si le modèle choisi est stationnaire) est égale à l'espérance constante $m = E(Z(x))$:

$$\bar{z} \rightarrow \bar{Z} = \frac{1}{S} \int_S Z(x) dx \quad ; \quad E(\bar{Z}) = m$$

De même, au variogramme régional γ_R , est associée, dans le modèle, la variable aléatoire Γ_R dont l'espérance, dans le modèle "F.A. intrinsèque", est égale au variogramme "théorique" γ :

$$\gamma_R(h) \rightarrow \Gamma_R(h) = \frac{1}{2K(h)} \int_{S(h)} [Z(x+h) - Z(x)]^2 dx$$

$$E[\Gamma_R(h)] = \gamma(h)$$

Il ne faut jamais oublier que α , m , $\gamma(h)$ etc... sont des paramètres du modèle et dépendent de la manière dont celui-ci a été choisi : ce sont, en général, en effet, des paramètres conventionnels sans signification objective. En toute rigueur, le seul cas où α soit un paramètre objectif est celui où l'on a presque sûrement (dans le modèle) $A = \alpha$, c'est-à-dire justement le cas où la valeur théorique α coïncide avec la valeur objective a (l'estimation de α sur la base d'une information fragmentaire pose une autre sorte de problème).

Conformément à la terminologie habituelle, convenons d'appeler fluctuation la V.A. (du modèle) égale à $A - \alpha$, dont l'écart $a - \alpha$ entre la grandeur régionale a et la caractéristique α du modèle qui lui est associée est interprété comme une réalisation. Si le paramètre α a un sens objectif, alors par définition la fluctuation est p.s. égale à 0 dans le modèle. Inversement, si la fluctuation n'est pas p.s. égale à 0, cela signifie que le paramètre α n'est pas objectif.

A la fluctuation $A - \alpha$, d'espérance nulle, sont attachées (dans le modèle) d'autres caractéristiques, par exemple sa variance $\text{Var}(A - \alpha)$ etc... Convenons de dire que $\text{Var}(A - \alpha)$ est la variance de fluctuation (dans le modèle) de la grandeur régionale. Si cette variance est grande, nous ne devons pas dire que "l'inférence statistique" du paramètre α est difficile ou mauvaise. Nous devons dire, au contraire, que ce paramètre n'a que peu, ou pas du tout, de signification objective. En effet, A , ou plutôt sa réalisation $a - \alpha$ épuise complètement le contenu objectif que l'on peut attribuer à l'opération définie par l'application de la fonctionnelle \mathcal{Q} à la V.R. z . Autrement dit, la variance de fluctuation donne une mesure, non pas de la difficulté de l'inférence statistique du paramètre α du modèle, mais bien plutôt du manque de

signification objective de ce paramètre.

Dans notre premier exemple, si l'on désigne par $\sigma(h)$ la covariance centrée (du modèle stationnaire), on démontre que la variance de fluctuation de la moyenne locale est (dans le modèle) :

$$(3) \quad \text{Var} (\bar{Z}-m) = \frac{1}{S^2} \int \sigma(h) K(h) dh$$

C'est donc seulement dans la mesure où cette variance est très petite, que l'espérance m (paramètre conventionnel) prend une signification objective (approximative). Nous discutons ce point dans le paragraphe consacré à l'ergodicité.

De la même façon, dans le modèle intrinsèque, la fluctuation du variogramme régional admet une variance, dont l'expression explicite fait d'ailleurs nécessairement intervenir les moments d'ordre 4 des accroissements de la F.A. Dans le cas (particulier) où ces accroissements obéissent à des lois de Gauss (modèle "F.A. intrinsèque gaussienne"), on montre que cette variance de fluctuation est donnée par l'expression

$$(4) \quad \text{Var} [\Gamma_R(h) - \gamma(h)] = \frac{1}{2(K(h))^2} \int_{S(h)} \int_{S(h)} [\gamma(x-y+h) + \gamma(x-y-h) - 2\gamma(x-y)^2] dx dy$$

Lorsque le module $|h|$ du vecteur h n'est pas très petit, cette variance de fluctuation peut devenir énorme.¹ Cela ne veut pas dire que "l'inférence statistique" du variogramme soit impossible ; mais seulement que ce variogramme γ n'a en dehors du voisinage de l'origine, qu'une signification conventionnelle. Conséquence importante : parmi les caractéristiques du modèle (les variances d'estimation par exemple) qui s'expriment à l'aide du variogramme γ , nous ne pourrions utiliser pour résoudre un problème réel (par exemple l'estimation de \bar{z}) que celles qui ne dépendent, pour l'essentiel, que du comportement du variogramme γ pour les petites valeurs de $|h|$, et sont relativement peu affectées par l'allure que présente cette fonction en dehors d'un certain voisinage de l'origine. Cette condition de robustesse, seule, nous garantit l'objectivité des grandeurs que nous calculons à l'aide du variogramme. Au contraire, pour $|h|$ petit, la variance définie en (4) est, en général, elle-même très petite, et cette circonstance nous permettra d'illustrer la notion de microergodicité

1 - J'ai développé ce point dans "les variables régionalisées et leur estimation" Paris, Masson, 1965, Chapitre XIII.

L'ergodicité

Dans l'optique classique, c'est toujours, en définitive, sur une propriété ergodique que se fonde la possibilité de "l'inférence statistique". Cela est vrai, évidemment, dans le cas simple d'une épreuve indéfiniment répétée. Si l'on admet l'indépendance des épreuves successives, et, surtout, que la probabilité p de succès reste constante, on a vu, en effet, que la "fréquence empirique" converge vers p avec une probabilité unité. De ces deux "hypothèses" la première peut être affaiblie. Il suffit que les corrélations entre épreuves successives s'amortissent suffisamment vite avec la distance dans le temps. La seconde est essentielle, et implique avant tout la stationnarité : on peut la réaménager, en disant par exemple que la probabilité $p(t)$ de succès au temps t est elle-même une F.A. stationnaire et ergodique; dans tous les cas, la stationnarité devra s'introduire.

Considérons maintenant le cas du modèle "F.A. aléatoire", et, par exemple, de "l'estimation" de la covariance $C(x,y)$. S'il s'agit d'un phénomène répétable, c'est-à-dire si l'on dispose d'autant de réalisation que l'on veut de la même F.A., la situation est substantiellement la même que dans le cas de l'épreuve en tout ou rien. On verra bien (en termes de physicien sinon de mathématicien) si la valeur moyenne du produit $Z(x) Z(y)$ converge ou non vers une limite $C(x,y)$ lorsque le nombre des réalisations augmentent "indéfiniment". Il n'est d'ailleurs pas nécessaire de supposer l'indépendance des épreuves successives, mais seulement leur stationnarité dans le temps (et leur ergodicité). Par exemple, $Z(x,t)$ peut être la température observée au temps t en une station météo implantée au point x du territoire européen. On peut admettre (ce n'est d'ailleurs pas évident, des tendances "séculaires" interviennent sans doute, incompatibles avec un modèle réellement stationnaire) que la moyenne des températures observées en x le 1er Janvier de chaque année à 0h. tend vers une limite lorsque la série temporelle dont on dispose s'allonge indéfiniment. On peut aussi, pour tenir compte, en première approximation, des tendances séculaires, admettre que celles-ci ne font sentir leurs effets qu'à long terme, et qu'à l'échelle de la dizaine d'années (disons entre 1970 et 1980), tout se passe comme si la température était stationnaire, avec une moyenne $m = m(t)$ qui serait, en réalité, une fonction très lentement variable du temps t . C'est, en somme, déjà notre modèle de stationnarité locale¹. On peut aussi utiliser des modèles du type classique "dérive séculaire (fonctionnelle) + F.A. stationnaire". L'ergodicité est ici un concept opérateur, parce que, bien qu'actuellement finie, la suite des observations est virtuellement infinie et augmente avec le temps, de sorte qu'un accroc à la stationnarité finira certainement à la longue par être décelé, corrigé en première approximation par l'ajustement d'une dérive séculaire, puis à leur tour les résidus se révéleront non-stationnaires etc...

Lorsqu'il s'agit d'un phénomène unique, d'extension bornée, (comme un gisement minier) la situation est sensiblement différente, puisqu'il n'y a cette fois plus aucune possibilité de répétition. Pour évaluer $E(Z_x Z_y)$, à supposer que nous arrivions à donner un sens opératoire à cette espérance, nous ne disposons en tout et pour tout que de la seule valeur numérique $z_x z_y$. D'où, bien sûr, appel à une "hypothèse" de stationnarité, permettant de remplacer la non répétabilité dans le temps par la répétition dans l'espace. Nous choisirons donc (pourvu que cela ne soit pas contraire à notre information qualitative sur le phénomène) un modèle stationnaire, et naturellement aussi "ergodique".

Ici, nous ne pouvons éviter une courte parenthèse un peu plus technique. Parmi les différentes acceptions que peut revêtir le terme "ergodicité", nous nous référons uniquement à la propriété suivant laquelle les moyennes spatiales, prises sur des domaines S de plus en plus grands, convergent vers l'espérance mathématique. Mais il y a lieu de distinguer nettement un théorème ergodique, tout à fait général, et une "hypothèse" ergodique, qui constitue une particularité de certains modèles de F.A. stationnaires, mais non pas de tous. Le théorème ergodique nous dit ceci : si $Z(x)$ est une fonction aléatoire stationnaire, et si l'on se donne une suite de domaines S_n tendant vers l'infini (en ce sens que toute boule de rayon donné est contenue dans S_n dès que n est assez grand), alors la suite des moyennes d'espace

$$Z_n = \frac{1}{S_n} \int_{S_n} Z(x) dx$$

converge vers une limite M (en un sens bien défini, par exemple la "convergence presque sûre" ou, dans le cas des F.A. d'ordre 2, la "convergence en moyenne quadratique" etc...). cette limite M ne dépend pas de la suite S_n choisie. En particulier, elle est invariante par translation. Mais le théorème n'affirme nullement que M soit identique à l'espérance m . En général, cette limite M , qui existe pour "presque toutes" les réalisations possibles de la F.A., prend des valeurs différentes pour les différentes réalisations. C'est donc une variable aléatoire, non une grandeur numérique, dont on peut dire seulement qu'elle est "invariante par translation", non qu'elle est constante. Dans les modèles vérifiant ce que l'on appelle la propriété, ou "l'hypothèse" ergodique, par définition, les seules variables aléatoires invariantes par translation sont les constantes, et dans ce cas, évidemment, $M=m$, ou, si l'on préfère, les moyennes spatiales Z_n convergent vers leur espérance m .

Quelle que soit son importance, en mécanique statistique par exemple¹, la distinction entre le théorème et "l'hypothèse" ergodiques ne peut jouer aucun rôle dans le cas d'un phénomène unique : en effet, puisque nous ne disposons et ne disposerons jamais que d'une seule réalisation de la F.A. (à savoir la variable régionalisée z) nous ne saurons jamais si la limite M aurait pris, ou non, une valeur différente sur une autre réalisation : la distinction entre M et m ne présente, dans ce contexte, aucune signification objective.

Par conséquent, si (bizarrement) nous avons choisi pour Z un modèle ne possédant pas la propriété ergodique, nous devrions nous empresser de le remplacer par un autre modèle où cette propriété serait vérifiée (il suffirait, par exemple, de conditionner la F.A. initiale par la variable M , ce qui assure automatiquement l'égalité $M=m$ sans détruire la stationnarité). Ce qui revient à choisir comme définition de l'espérance m la limite de la suite des moyennes spatiales Z_n . Si m et M n'ont qu'un statut purement conventionnel, il n'y a aucun inconvénient à les identifier. Si elles ont une signification objective, cette signification est obligatoirement la même pour m et M : à savoir, pour l'une comme pour l'autre, la limite de la moyenne spatiale lorsque S tend vers l'infini.

En toute rigueur, d'ailleurs, cette définition ne peut être que conventionnelle : du fait que le champ réel est irréductiblement borné, il n'est en aucune façon possible de faire "tendre" réellement S vers l'infini, et la notion d'espérance m ou de limite ergodique M , dont la définition fait intervenir le comportement (fictif) du phénomène à l'infini est, par là même, dépourvue² de signification objective, du moins au sens strict. Cependant, il ne faut pas non plus être trop puriste. Si le champ S est vraiment très grand, la variance donnée par la formule (3), sans être nulle, peut être suffisamment petite pour être considérée comme négligeable, et le paramètre correspondant du modèle, sans être rigoureusement identique à la grandeur régionale, peut cependant être confondu avec elle avec une bonne approximation.

1 - Voir dans cette collection, l'analyse approfondie que donne F.Fer, in "l'irréversibilité, fondement de la stabilité du monde physique", 1977.

2 - Lorsque nous procéderons, dans la 3^{ème} partie, à la reconstruction opératoire de nos modèles, la situation sera bien différente : le paramètre m sera alors défini comme une certaine moyenne spatiale, et son objectivité sera ainsi automatiquement garantie.

Nous disons, en pareil cas, que l'ergodicité est pratiquement atteinte, en ce qui concerne ce paramètre, et celà nous autorise à en faire légitimement usage.

La portée

Du point de vue pratique, la notion capitale est ici celle de la portée de la covariance, c'est-à-dire, en termes approximatifs, de la distance à partir de laquelle les corrélations s'éteignent ou deviennent négligeables. C'est seulement dans la mesure où le champ S est grand vis à vis de la portée que l'on peut espérer atteindre pratiquement, l'ergodicité, et donner ainsi un sens objectif à l'espérance m.

L'examen de la formule (3) permet de comprendre pourquoi. Dans cette formule, $K(h)$ représente la mesure du domaine $S(h)$, intersection du champ S et de son translaté par le vecteur h. En particulier, $K(o)$ est égal à la mesure de S. Il est alors facile de montrer que, pour S assez grand, la variance de $\bar{Z}-m$ est peu différente de $(1/S) \int \sigma(h) dh$. Ceci suggère la définition suivante : appelons portée intégrale la quantité

$$(5) \quad A = \frac{1}{\sigma^2} \int \sigma(h) dh$$

(la notation traditionnelle σ^2 représente la variance, dans le modèle, de la F.A. stationnaire $Z(x)$. On a, évidemment, $\sigma^2 = \sigma(o)$, valeur en $h = 0$ de la covariance $\sigma(h)$). A est une longueur, une surface ou un volume selon que l'espace a une, deux ou trois dimensions. Lorsque S est grand, posons $N=S/A$: N représente donc, si l'on veut, le nombre de pavés disjoints de mesure A contenus dans le champ S. Avec ces notations, la formule (3) donne asymptotiquement, c'est-à-dire lorsque S devient grand :

$$\text{Var } (\bar{Z}-m) = \frac{\sigma^2}{N}$$

En ce qui concerne "l'estimation" de m, tout se passe donc, dans ce modèle, comme si l'estimateur \bar{Z} était obtenu en prenant la moyenne de N variables indépendantes de variance σ^2 , La portée intégrale A représente donc bien l'élément de référence vis à vis duquel il y a un sens à dire que le champ S est grand. Plus le nombre $N=S/A$ est grand, plus, en effet, la variance de $\bar{Z}-m$ est petite, et plus, par conséquent, le paramètre m présente de signification objective.

Mais ici une objection sérieuse se présente. En définissant la portée intégrale A par la relation (5), c'est-à-dire par une expression où intervient de manière cruciale le comportement de la covariance $\sigma(h)$ pour $|h|$ tendant vers l'infini, nous faisons appel à des caractéristiques du modèle auxquelles ne peuvent correspondre aucune propriété objective du phénomène réel. En particulier, on peut donner des exemples de covariances différentes, pratiquement identiques dans le domaine utile ($|h|$ inférieur au plus grand diamètre de S), mais conduisant à des valeurs extrêmement différentes pour les portées A , voir même à une valeur infinie (l'intégrale (5) peut très bien être divergente pour certains choix de la covariance $\sigma(h)$).

Pour lever cette objection, il faut, comme d'habitude, procéder à la reconstruction opératoire du concept incriminé. C'est ici que s'introduit une régionale particulière, que nous appellerons "variance de s dans S " et désignerons par $v(s/S)$. Nous ne donnerons pas la définition rigoureuse, que l'on trouvera dans la littérature spécialisée, mais le lecteur concevra facilement que l'on puisse diviser, au moins par la pensée, un champ donné S en un certain nombre d'éléments s égaux entre eux (ou à peu près) à une translation près. Il est alors facile de définir, numériquement, la variance de la population (finie) constituée par les valeurs moyennes de la V.R. z sur chacun de ces éléments s dont la réunion constitue S : c'est, par définition, la variance $v(s/S)$ de s dans S , expression dont la valeur numérique constitue une grandeur régionale, accessible (après coup) à une détermination expérimentale. A cette régionale $v(s/S)$ est associée, comme d'habitude, dans le modèle, une variable aléatoire $V(s/S)$ dont l'espérance $E[V(s/S)] = \sigma^2(s/S)$ peut être appelée, si l'on veut, variance théorique de s dans S . Cette dernière quantité est un paramètre du modèle, qui s'exprime facilement à l'aide du seul variogramme γ (explicitement, $\sigma^2(s/S) = \bar{\gamma}(S) - \bar{\gamma}(s)$, écriture ou $\bar{\gamma}(s)$, par exemple, désigne la valeur moyenne de $\gamma(x-y)$ lorsque les deux points x et y parcourent, indépendamment l'un de l'autre, le domaine s). Ces expressions n'ont pas forcément de signification objective, puisque le calcul de $\bar{\gamma}(S)$, par exemple, met en jeu les valeurs de $\gamma(h)$ pour des arguments h atteignant le diamètre de S . Mais, après coup au moins, nous pouvons contrôler si la relation $v(s/S) = \bar{\gamma}(S) - \bar{\gamma}(s)$ est, ou non, approximativement vérifiées. Remarquons que c'est exactement de cette manière que nous avons procédé au chapitre 3 pour reconstruire la notion de densité poissonnienne. On peut ainsi mettre en évidence l'existence d'une loi physique valable, avec une certaine approximation, dans un certain domaine de variation de s et de S (il faut, évidemment, que s soit suffisamment petit devant S , pour que nous ayons, à notre disposition, un effectif statistique pas trop faible, disons au moins 10 individus).

Or la théorie nous fait prévoir ceci : si s (et à plus forte raison S) est grand devant la portée A , telle que nous l'avons définie par la relation (5), alors $\bar{\gamma}(s)$ diffère très peu de $\sigma^2 - A/s$. Par conséquent, si notre modèle est bon, nous pouvons nous attendre à observer une relation de la forme :

$$(6) \quad v(s/S) = A \left(\frac{1}{s} - \frac{1}{S} \right)$$

Cette relation constitue une loi physique, puisque $v(s/S)$ est une régionale. Si nous pouvons (après coup) contrôler expérimentalement qu'au delà d'une certaine valeur de s cette loi est effectivement valable, nous avons, par là même, réussi à construire en termes opératoires le concept de portée intégrale : c'est maintenant la loi physique (6) qui constitue elle-même la définition opératoire du concept.

Cette reconstruction n'est, en principe, réalisable qu'après coup. Le plus souvent, il est possible, même in praxi, de se rendre compte si le champ S est suffisamment grand pour que l'on puisse considérer l'ergodicité comme "atteinte" ou "réalisée" : un indice sûr est fourni par l'allure du "variogramme expérimental" (c'est à-dire l'estimation du variogramme régional que l'on peut former, in praxi, à partir d'un nombre limité d'échantillons). Si ce variogramme expérimental atteint un palier, ou asymptote horizontale, qui semble stable (abstraction faite des fluctuations résiduelles que l'on doit toujours s'attendre à observer) on peut, sans grand risque d'erreur, avancer l'hypothèse que la portée existe bien (objectivement) et en faire une mesure approximative. La figure 4 donne un exemple assez parlant. Sur la figure 5 au contraire, on observe une circonstance singulière : les quatre variogrammes expérimentaux, construits dans quatre directions différentes, coïncident à peu près jusque vers $h = 500$ m et l'allure de deux de ces courbes semble annoncer l'existence d'un palier et d'une portée de l'ordre de 700 m. Mais au delà de $h = 500$ m, les quatre courbes divergent violemment, et l'allure de deux d'entre elles est absolument incompatible avec l'existence d'un palier. Pour un praticien, cette divergence est un indice très sûr de l'existence de ce que l'on appelle une "dérive", notion sur laquelle nous reviendrons. Le fait que les quatre courbes expérimentales coïncident assez bien, jusqu'à cette valeur critique $h = 500$ m où les divergences se révèlent brusquement, est bien typique de ce que nous appellerons la stationnarité locale. Nous verrons, au chapitre 7, la définition précise des modèles correspondants ou, modèles locaux. Mais, d'ores et déjà, on conçoit qu'il sera, dans ce cas, légitime de faire un usage local (i.e : ne faisant jamais intervenir simultanément que des points séparés par des distances inférieures à 500 m) du modèle de variogramme dont l'ajustement est présenté sur l'agrandissement que l'on peut voir dans la moitié droite de la figure.

Sur la figure 6, on peut voir un exemple expérimental de variance $v(s/S)$ de s dans S . Ici, s (échantillons de taille fixe) est constant, et, on a porté, en fonction de $\log S/s$ (de $S/s = 2$ à plus de 1000) Les variances des teneurs dans des zones S de plus en plus grandes d'un gisement d'or. Il n'y a aucun signe annonçant l'existence d'une asymptote horizontale. La relation (6) n'est certainement pas vérifiée, et rien ne permet d'attribuer à ce phénomène une variance σ^2 finie. La courbe expérimentale observée s'explique très bien, par contre, dans le cadre du modèle :

$$\gamma(h) = \alpha \log |h|$$

(variogramme logarithmique, ou "de Wijsien"). Il s'agit d'un variogramme indéfiniment croissant, donc sans palier ni portée. On peut évidemment, imaginer que la courbe finirait bien par s'infléchir, au delà du dernier point expérimental observé, et introduire, par force, un palier (ou variance σ^2) et une portée finis : mais il s'agirait alors de paramètres purement conventionnels, auxquels aucune réalité observable ne pourrait être associée. Cela n'aurait d'ailleurs aucune importance : car ce modèle à palier et portée finis, pourvu qu'il coïncide, à peu près, avec le modèle de Wijsien jusqu'à la plus grande valeur possible de $|h|$, conduirait exactement aux mêmes conclusions que ce dernier, en ce qui concerne du moins les problèmes réels que nous pourrions nous poser au sujet de ce phénomène : (s'il s'agit d'un problème réel, il ne peut pas mettre en jeu des distances supérieures à celles qui existent entre les points de l'objet réel.

La microergodicité

Supposons que nous ayons mesuré les valeurs de la variable régionalisée z en n points implantés, par exemple, selon une maille carrée de côté a , dans le champ S (supposé bidimensionnel). Il y a alors deux manières bien différentes de faire tendre n vers l'infini. Ou bien, la maille a restant constante, on augmente le champ d'investigation, de sorte que la densité d'échantillonnage reste constante, tandis que la surface échantillonnée tend vers l'infini - ou bien, au contraire, on laisse fixe la surface échantillonnée et l'on resserre la maille, de sorte que c'est cette fois a qui tend vers 0, ou la densité d'échantillonnage qui tend vers l'infini. Dans le premier cas (et moyennant certaines restrictions peu importantes), on retrouve l'essentiel, la situation "ergodique" classique, telle que nous l'avons analysée ci-dessus.

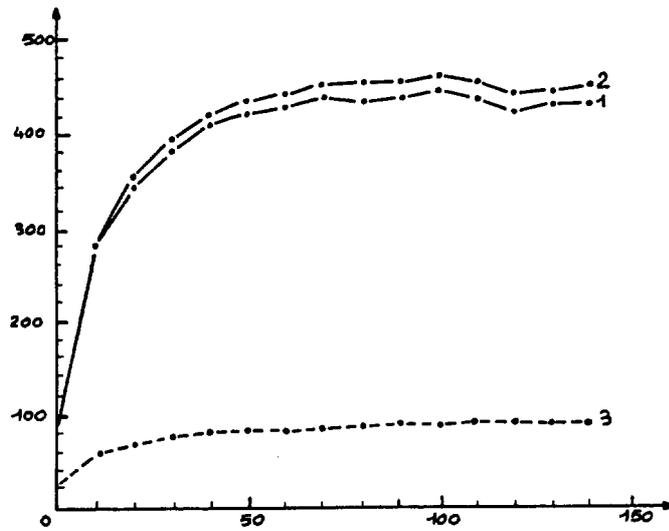


Fig.4 - Variogrammes associés à des faciès observés sur des parements de mine :
1, concrétions calcaires ; 2, minerais interconcrétions ; 3, joints argileux.
D'après J. Serra¹

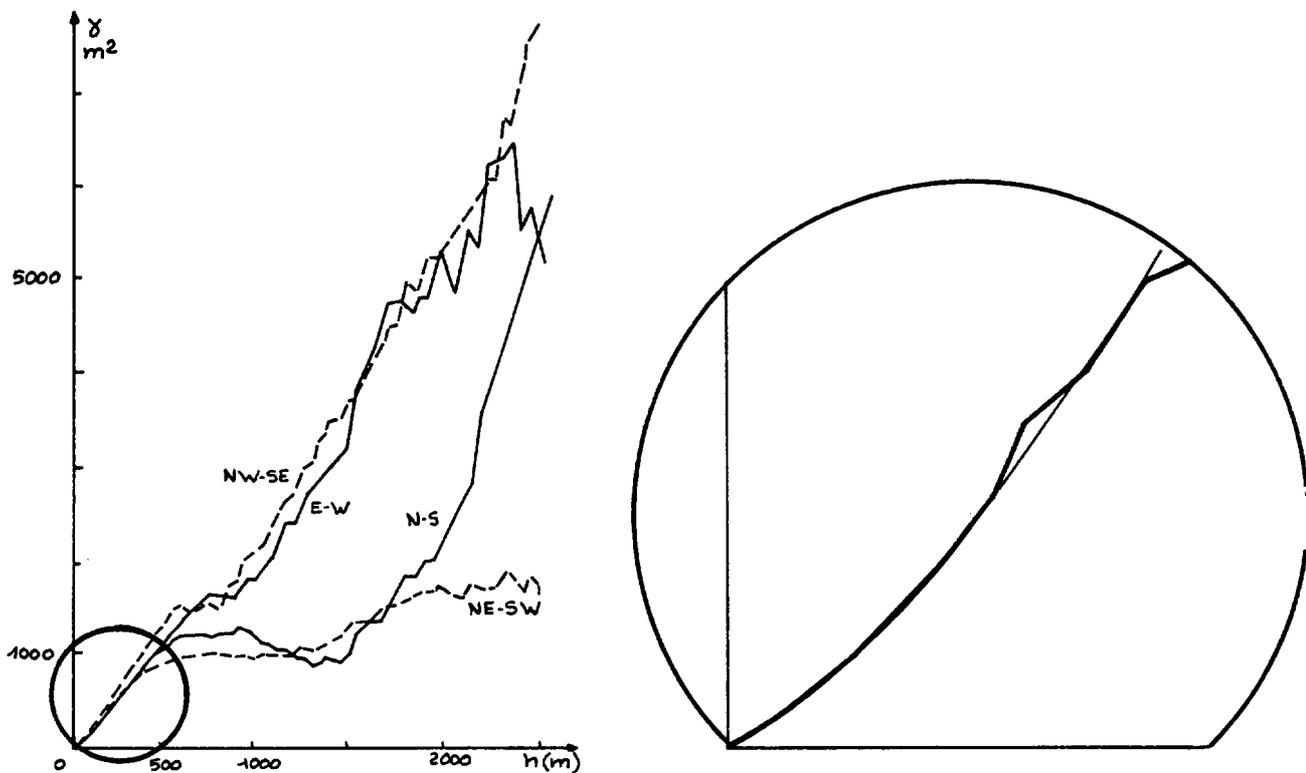


Fig.5 - Variogrammes expérimentaux des cotes topographiques (faille Noiretable, 1/25000). D'après J.P. Chilès²

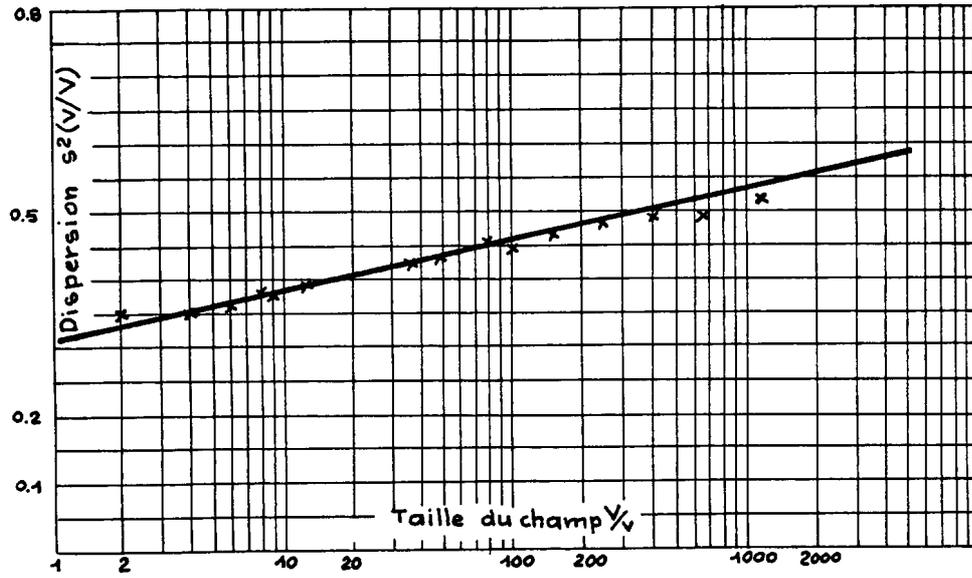


Fig.6 - Variances expérimentales des teneurs en or de blocs de taille fixe dans des zones de plus en plus grandes d'un gisement d'Afrique du Sud.
D.G. Krige, cité in A. Journel.

-
- 1 - Morphologie Mathématique et g n se des concr tions carbonat es des minerais de fer de Lorraine - *Sedimentology*, 10 (1968), 183-208.
 - 2 - G ostatistique des ph nom nes non stationnaires dans le plan, th se, Nancy 1977.

Dans le second cas, au contraire, on retrouve, à la limite, pour $a = 0$, la situation idéale dite "après coup" dont nous avons déjà noté l'intérêt épistémologique, celle où la valeur numérique de la V.R., interprétée maintenant comme réalisation de la F.A. du modèle, est connue en tous les points du champ S. Nous avons longuement insisté sur le fait que les grandeurs régionales, seules, possèdent une signification objective, et que par suite les paramètres du modèle qui ne se laissent pas identifier à de telles grandeurs sont, au sens strict, dépourvus de réalité. Parmi ces paramètres, donc, il y a un tri à faire, et ce sera, du point de vue technique, l'un des aspects les plus importants du choix et de la spécification du modèle : pour un modèle donné de F.A. et un domaine S, quels sont les paramètres du modèle qui se trouvent rigoureusement (ou, avec une approximation donnée) déterminés si l'on connaît une (seule) réalisation de la F.A. sur S - et quels sont ceux qui ne le sont pas ? Seuls les premiers ont un sens objectif et sont susceptibles d'une définition opératoire en termes de "a tendant vers 0". Nous dirons qu'ils sont micro-ergodiques. Quant aux autres, étant dépourvus de réalité opératoire, ils peuvent dans une très large mesure être choisis arbitrairement (puisque nous n'aurons jamais d'indication expérimentale vraiment contraignante à leur égard).

Cette propriété de micro-ergodicité est tout à fait distincte de la propriété ergodique habituelle, et mériterait à mon sens une étude systématique. Même dans le cas d'un modèle stationnaire et ergodique, ni la moyenne ni la variance ne sont micro-ergodiques. Par contre, le comportement du variogramme au voisinage de $h = 0$ est micro-ergodique, pourvu, comme nous allons le voir, qu'il ne soit pas trop régulier. Bien mieux, et moyennant certaines hypothèses assez lâches, cette micro-ergodicité subsiste même dans le cas non stationnaire (il s'agit alors évidemment du variogramme moyen, défini comme la moyenne dans S du variogramme non stationnaire). Ainsi, la stationnarité et l'ergodicité classique n'entraînent nullement la micro-ergodicité, et celle-ci, inversement, peut être vérifiée même en l'absence de stationnarité.

Examinons, en effet, la formule (4) qui donne (dans le modèle "F.A. intrinsèque" la variance de la "fluctuation" du variogramme régional $\Gamma_R(h)$ (cette formule n'est valable que dans le cas gaussien, mais les conclusions que nous allons dégager ont, évidemment, une signification plus générale). Nous avons vu que, à moins que $|h|$ ne soit petit (ou, ajoutons le, puisque nous le pouvons maintenant, à moins que le phénomène n'admette une portée finie, petite vis-à-vis du champ S), cette variance pouvait être très grande, privant ainsi de signification objective le comportement du variogramme théorique γ aux grandes valeurs de $|h|$.

Pour $|h|$ petit, au contraire, deux cas bien différents peuvent se présenter. Supposons, pour abrégier la discussion, que le variogramme γ du modèle soit "isotrope" (c'est-à-dire ne dépende que du module $|h|$ du vecteur h et non de sa direction) et que, pour $|h|$ petit, $\gamma(h)$ soit équivalent¹ à une expression de la forme $a|h|^\lambda$. La théorie prévoit que deux éventualités seulement sont possibles :

~ ou bien $\lambda = 2$, et dans ce cas la F.A. est "dérivable en moyenne quadratique" : ce genre de modèle est bien adapté à la description de phénomènes très réguliers dans leur variation spatiale

~ ou bien $\lambda < 2$: la F.A. est alors continue, mais non dérivable "en moyenne quadratique", et le modèle convient pour la représentation de phénomènes beaucoup plus irréguliers et discontinus. L'éventualité $\lambda > 2$ est exclue par la théorie.

On démontre alors (dans le modèle) que la variance du rapport $\Gamma_R(h)/\gamma(h)$ est, pour $|h|$ petit, équivalente à une expression de la forme

$$C |h|^{4-2\lambda} + B |h|^n$$

(n est le nombre des dimensions de l'espace, soit $n = 1, 2$ ou 3).

Ainsi cette variance relative tend vers 0 si et seulement si λ est inférieur à 2, c'est-à-dire si la F.A. n'est pas continue en moyenne quadratique. Ou encore, dans le modèle, on a la convergence :

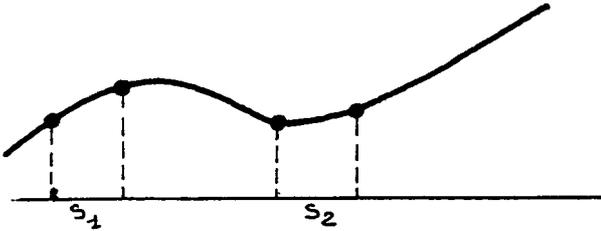
$$(7) \quad \frac{\Gamma_R(h)}{|h|^\lambda} \rightarrow a$$

(au sens "en moyenne quadratique" ou, aussi bien, au sens "presque sûr") pourvu que λ soit inférieur (strictement) à 2, et le paramètre a est alors microergodique. Par contre, si $\lambda = 2$, cette convergence n'a pas lieu, et la microergodicité n'est pas non plus assurée.

Ainsi se dégage une conclusion très instructive : les paramètres qui définissent le comportement du variogramme au voisinage à l'origine sont microergodiques, et donc, en particulier, possèdent une signification objective, pourvu seulement que le phénomène (ou plutôt la F.A. qui lui correspond dans le modèle) ne soit pas trop régulier.

1 - Autrement dit, $\gamma(h)/|h|^\lambda$ tend vers a pour $|h|$ tendant vers 0.

Cette conclusion ne doit pas surprendre. Le champ S peut être très petit (il suffit qu'il contienne un ensemble ouvert, par exemple une boule de rayon aussi petit que l'on veut). Si $\lambda = 2$, la réalisation de Z est une fonction très continue, assimilable, sur n'importe quel petit domaine S , à une parabole, par exemple : mais, suivant l'implantation S_1, S_2 de ce petit champ S , on voit que l'on peut observer à peu près n'importe quel morceau de parabole. Dans le langage classique, l'inférence statistique n'est pas possible. Ce qui veut dire en réalité, que le modèle théorique n'a que peu de signification objective. Pour $\lambda < 2$, au contraire, l'information fournie par les points (en nombre



infini, même s'ils sont très voisins les uns des autres) appartenant à S est beaucoup moins redondante, du fait même que la F.A. est moins régulière, d'où la microergodicité et la possibilité d'accéder expérimentalement aux paramètres a et λ . Le lecteur qui connaît les propriétés du processus de Wiener-Levy comprendra sans peine ce que nous ne pouvons dire ici qu'en termes un peu vagues.

Notons, pourtant, que la convergence (7), pour $\lambda < 2$, est un événement (presque sûr) du modèle, mais que sa contrepartie écrite en termes de régionales :

$$\frac{\gamma_R(h)}{|h|^\lambda} \rightarrow a$$

n'est pas, réellement, contrôlable expérimentalement, puisqu'il s'agit d'une limite, mettant en jeu une infinité de points. Nous sommes donc dans un cas où il est nécessaire de distinguer entre le modèle primaire (la V.R.) et la réalité physique.

Pour préciser le contenu physique de la notion, apparemment simple, du "comportement du variogramme au voisinage de l'origine" il sera nécessaire de procéder à une véritable reconstruction opératoire. Nous reviendrons sur ce point dans la 3^{ème} partie.

L'estimation, in praxi, des grandeurs régionales

Le problème de l'estimation (in praxi) d'une grandeur régionale à partir d'une information fragmentaire est un problème bien réel, qui se pose constamment en pratique, et que l'on ne doit absolument pas confondre avec celui de la fluctuation de

(ou plutôt : que le modèle attribue à) la grandeur régionale autour de sa version théorisée. En particulier, dans le cas limite où l'information tend à devenir parfaite, l'estimation coïncide avec la vraie valeur régionale, tandis que la fluctuation subsiste. Dans le cas le plus simple, le problème est le suivant : connaissant les valeurs numériques $z_\alpha = z(x_\alpha)$ de la V.R. aux points N expérimentaux x_α , on cherche une fonction $a^x = Q^x(z_1, \dots, z_N ; x_1, \dots, x_N)$ de ces variables qui réalise au mieux (en un sens à définir) l'estimation de la régionale $a = Q(z)$. Dans le modèle, après substitution de $Z(x)$ à $z(x)$, a et Q^x deviennent des V.A. (non indépendantes), soient A et A^x respectivement. Dans la version théorisée du problème de l'estimation, on est conduit, donc à rechercher une fonction Q^x de N variables telle que la V.A. $A^x = Q^x(z_1, \dots, z_N)$ soit aussi proche que possible (en un sens à définir) de la V.A. que l'on veut estimer, à savoir $A = Q(Z)$. Par exemple, on peut chercher à minimiser la variance d'estimation $\text{Var}(A - A^x)$ sous la condition de "non biais" $E(A - A^x) = 0$. Cette variance et cette espérance sont des caractéristiques du modèle, d'où une première limitation : on doit se contenter de chercher la fonction Q^x dans la classe Φ des fonctions (mesurables) ϕ telles que les quantités

$$E[A - \phi(Z_1, \dots, Z_N)] \text{ et } \text{Var}[A - \phi(Z_1, \dots, Z_N)]$$

s'expriment en fonction des seuls critères de spécification du modèle choisi. Le problème serait donc :

$$\left\{ \begin{array}{l} \text{Trouver } \phi \in \Phi \text{ minimisant } \text{Var}[A - \phi(Z_1, \dots, Z_N)] \text{ sous la contrainte} \\ E[A - \phi(Z_1, \dots, Z_N)] = 0. \end{array} \right.$$

A cette limitation évidente (ne chercher ϕ que dans la classe des estimateurs dont nous soyons effectivement capables de calculer l'espérance et la variance que notre modèle leur attribue) viennent s'ajouter des restrictions concernant la signification objective de cette espérance et de cette variance. L'examen de cette dernière question nécessitant, en fait, la reconstruction opératoire du modèle, ne sera abordé que dans la 3^{ème} partie, et nous notons seulement ici quelques remarques préliminaires. Du point de vue pratique, l'essentiel se résume en une seule ligne :

ATTENTION AUX SEUILS DE REALISME ET DE ROBUSTESSE

Le mieux à ce stade, est sans doute d'examiner quelques exemples.

Exemple de l'espérance conditionnelle.

Plaçons-nous, un instant, dans le cas idéal d'un modèle complètement spécifié, c'est-à-dire où l'on a choisi explicitement la loi spatiale de la F.A. A chaque régionale a , on peut alors associer la loi de la V.A. correspondante A , prise conditionnellement aux données disponibles Z_α , $\alpha = 1, 2, \dots, N$. Et ceci résout le problème de l'estimation, puisque le meilleur estimateur possible est alors l'espérance conditionnelle $a^x(Z_1, \dots, Z_N) = E(A|Z_1, \dots, Z_N)$.

Il s'agit d'ailleurs d'un cas purement idéal, car en pratique les données disponibles ne permettent pour ainsi dire jamais de spécifier complètement la loi spatiale de la F.A. théorique. En pratique, on remplace le plus souvent l'espérance conditionnelle inaccessible par un estimateur linéaire optimal, ou par un estimateur plus élaboré (du type "disjonctif" par exemple, cf. chapitre 8), mais exigeant moins de prérequisites que l'espérance conditionnelle.

En fait, si l'on n'utilise pas les lois conditionnelles, ce n'est pas seulement (pour reprendre la terminologie traditionnelle) parce que les données disponibles permettent rarement l'inférence statistique de la loi spatiale. Il arrive, en effet, quelquefois que le modèle soit complètement spécifié, par exemple s'il s'agit d'une F.A. stationnaire gaussienne ou lognormale dont l'espérance et la covariance ont pu faire l'objet d'une estimation sérieuse. Une autre considération intervient, liée à ce que nous avons appelé la robustesse de modèle (robustesse de type et robustesse spécifique). En termes théoriques (probabilistes) cela revient à se demander ce qui se passe si la F.A. "réelle" n'est pas tout à fait conforme au modèle choisi (pas tout-à-fait gaussienne, pas tout-à-fait stationnaire, a une covariance réelle légèrement différente de celle que nous avons retenue etc...). Il y a tout lieu de craindre que l'espérance conditionnelle soit gravement affectée par ces variations, et qu'ainsi nos seuils de robustesse ne soient largement dépassés. Nous reviendrons sur ce point dans le chapitre 8.

Plus généralement, en se plaçant d'un point de vue purement théorique (et à supposer toujours que l'on ait réussi à spécifier complètement le modèle de F.A.) on pourrait croire que le meilleur modèle possible à utiliser dans le problème de l'estimation est celui de la F.A. initiale $Z(x)$ prise conditionnellement à l'information disponible, i.e. à $Z(x_\alpha) = z_\alpha$, $\alpha = 1, \dots, N$ fixés. Mais le trait le plus saillant de la loi spatiale ainsi conditionnée est que, par construction, elle ne contient plus aucun élément accessible à un contrôle expérimental direct : en effet, toutes les données expérimentales sont passées du côté des variables conditionnantes, et jouent

le rôle de paramètres dans l'expression de la loi conditionnée. On conçoit qu'il faille vraiment accorder une confiance de fer au modèle choisi pour oser utiliser, les yeux fermés, cette loi conditionnée. Il y a tout lieu de redouter qu'à ce stade, non seulement le seuil de robustesse mais aussi le seuil de réalisme aient été franchis depuis longtemps.

Exemple du "krigeage"¹

Pour prendre un exemple moins ambitieux, considérons maintenant le cas du modèle générique "F.A. stationnaire d'ordre 2", où les critères de spécification sont donc : une espérance m et une covariance centrée $\sigma(h)$ stationnaire. Supposons que nous voulions estimer la valeur $z(x)$ de la V.R. en un point x différent des points expérimentaux x_α ($\alpha = 1, 2, \dots, N$), ou, plus généralement, une "moyenne pondérée" du type

$$z(p) = \int p(dx) z(x)$$

où p est une "mesure" de somme unité et à support dans S , c'est-à-dire une grandeur régionale définie par une fonctionnelle linéaire de la V.R. z . La classe Φ des estimateurs possibles (caractérisables à l'aide des seuls critères de spécification du modèle) est constituée des combinaisons linéaires affines de la forme $a + \sum \lambda^\alpha z_\alpha$. La version théorisée du problème de l'estimation est donc la suivante :

Trouver les constantes a et λ^α , $\alpha = 1, 2, \dots, N$, minimisant $E[Z(p) - a - \sum \lambda^\alpha z_\alpha]^2$.

(Nous n'ajoutons pas la condition de non biais $E(Z^* - Z(p)) = 0$, qui, dans cette formulation, est automatiquement remplie). La solution, comme on sait, est obtenue en prenant l'estimateur

$$z^* = m + \sum \lambda^\alpha (z_\alpha - m)$$

avec les poids λ^α vérifiant le système

$$\sum \lambda^\alpha \sigma(x_\alpha - x_\beta) = \int p(dx) \sigma(x - x_\beta)$$

1 - Ce terme, qui tire son origine de la Géostatistique minière, fait allusion aux travaux de D.G. Krige, et signifie "meilleur estimateur linéaire vérifiant telle ou telle condition de non biais".

L'expression de cet estimateur dépend, évidemment, des critères de spécification choisis, m et $\sigma(h)$. De plus, si les points expérimentaux x_{α} sont en nombre élevé ($N = 100$, ou 1000 etc...), on peut se demander si le système est "bien conditionné". Enfin, s'il s'agit d'estimer $z(x)$ en un point x donné, ou une moyenne $z(p)$ prise au voisinage immédiat de ce point, on se rend compte, en examinant ces formules, que l'on risque d'attribuer un poids exagérément important à des points expérimentaux très éloignés du point à estimer. Il paraît physiquement peu plausible que des points aussi éloignés exercent une influence aussi forte. Il faudrait ici encore, accorder une confiance vraiment totale au modèle "F.A. stationnaire d'ordre 2" pour appliquer ces formules les yeux fermés. Ici, le sens physique doit intervenir, et rectifier ce qu'il y a de trop rigide dans le modèle théorique. Certes, la réalité est à peu près "stationnaire" (en un sens physique, difficile à préciser), mais à peu près seulement, et nous devons nous interdire d'utiliser des estimateurs aussi peu robustes que celui-ci vis-à-vis de la stationnarité du modèle théorique. En pratique donc :

i) on imposera aux coefficients la condition $\sum \lambda^{\alpha} = 1$, et on minimisera la variance d'estimation sous cette contrainte : ceci a l'avantage de filtrer le moment m d'ordre 1, qui disparaît de l'expression de l'estimateur (et donc peu importe que ce m là, quel que soit son statut épistémologique, soit ou non réellement constant dans l'espace) ; on peut encore voir dans cette disparition de m une application de la règle générale selon laquelle les opérations réellement effectuées doivent toujours être exprimées (ou exprimables) en termes de grandeurs régionales - et

ii) plus drastiquement encore, nous excluons de notre estimateur les points expérimentaux les plus éloignés, pour ne conserver que ceux qui appartiennent à un voisinage raisonnable du point à estimer (le choix de ce voisinage peut apparaître arbitraire, mais un praticien expérimenté, qui a une bonne intuition de son phénomène, n'hésitera pas longtemps). Moyennant une perte (théorique) de précision, d'ailleurs en général peu considérable, on augmente beaucoup la robustesse de l'estimateur (vis-à-vis du choix du type du modèle et de sa spécification). Et, accessoirement, on allège considérablement les calculs (ce qui n'est pas du tout négligeable dans la pratique réelle, où les considérations de prix de revient des calculs jouent toujours un rôle important).

De plus, ayant ainsi restreint la classe des estimateurs dans laquelle nous cherchons notre optimum, nous nous apercevons que notre modèle initial était trop spécifié. En effet, le paramètre conventionnel m s'est révélé inutile, et la cova-

riance $\sigma(h)$ peut être (avantageusement) remplacée par le variogramme $\gamma(h)$. Notre estimateur est maintenant donné par l'expression :

$$z^* = \sum \lambda^i z_i$$

où la sommation n'intéresse plus que les points expérimentaux x_i situés dans le voisinage de la zone à estimer ($N = 10$ à 20 par exemple) et les poids λ^i vérifient le système :

$$\sum \lambda^i \gamma(x_i - x_j) = \int p(dx) \gamma(x-x_j) + \mu$$

$$\sum \lambda_i = 1$$

En somme, au modèle initial "F.A. stationnaire d'ordre 2", nous avons substitué le modèle "F.A. intrinsèque" (F.A.I.), et même, en réalité, le modèle (considérablement moins spécifié) "F.A. localement intrinsèque" (voir chapitre 7). Ainsi, nous n'avons pas seulement amélioré la robustesse de notre estimateur, nous avons aussi considérablement affaibli l'hypothèse anticipatrice sur laquelle il est fondé, dans la mesure même où nous avons réduit les critères de spécification (disparition de m , nécessité de connaître $\gamma(h)$ seulement au voisinage de la zone à estimer, c'est-à-dire justement cette partie du variogramme qui a le plus de signification objective).

A cette estimation, le modèle associe une "variance d'estimation" qui permet de se faire une idée de l'ordre de grandeur des erreurs possibles. Nous reviendrons longuement, dans les chapitres qui viennent, sur le sens objectif de cette variance d'estimation et sa reconstruction en termes opératoires. Donnons seulement ici un exemple¹ emprunté à la Géostatistique Minière. Il s'agit d'un grand gisement de cuivre, dans lequel 810 panneaux de même dimension ont été estimés, par krigeage, à l'aide des sondages voisins. La maille étant régulière, la géométrie respective du panneau et des sondages était invariable, de sorte que le modèle attribuait la même variance d'estimation (théorique) à chacun des panneaux. Chose exceptionnelle dans l'industrie minière, il a été ici possible, après exploitation, de reconstituer les teneurs réelles des 810 panneaux, et donc de calculer l'erreur moyenne (elle est pratiquement nulle) et la variance expérimentale de ces 810 panneaux. On a trouvé (en Cu %)

variance théorique : 0.117

variance expérimentale 0.118

1 - D'après A. Maréchal et I. Ugarte, cité in A. Journel, Mining Geostatistics

Des exemples comme celui-ci ne laissent place à aucun doute en ce qui concerne la signification objective (et la validité) de la notion de variance d'estimation. Mais, en pratique, on rencontre une difficulté bien réelle : c'est que la variance d'estimation, telle qu'on peut l'évaluer in praxi, c'est-à-dire sur la base d'une information partielle et d'une estimation approximative des paramètres du modèle choisi, est, en fait, déjà elle-même une estimation : il faudrait donc, à nouveau, évaluer la variance de cette estimation, et ainsi de suite à l'infini.

Si le modèle a été bien choisi, la variance d'estimation que l'on veut évaluer s'exprime, en principe, à l'aide des seuls paramètres objectifs du modèle, c'est-à-dire ceux qui s'expriment en termes de grandeurs régionales. Mais la difficulté subsiste : pour apprécier la possibilité de l'estimation des régionales spécifiant le modèle, il est en général nécessaire de faire intervenir d'autres instruments de travail que ceux auxquels nous avons droit dans ce cadre (les critères de spécification). D'où obligation :

i) de faire appel à un modèle plus spécifié que celui que nous avons en vue,

ii) de procéder au préalable à l'estimation d'autres régionales encore, en plus de celles dont nous désirons critiquer l'estimation. Par exemple, pour porter un jugement sur l'estimation d'un moment d'ordre 2 (covariance ou variogramme) et calculer, disons, la variance d'estimation correspondante, il faut faire appel à des moments d'ordre 4. L'estimation directe de ces moments d'ordre 4 est parfois possible (en règle générale, dans des conditions moins favorables que celle des moments d'ordre 2 eux-mêmes), mais nous oblige à adopter un modèle plus spécifié (spécifié par les moments d'ordre 2 et 4). De plus, se pose le problème de la critique de cette nouvelle estimation (celle des moments d'ordre 4), et l'on voit ainsi s'amorcer un désastreux regressus ad infinitum (désastreux, car expérimentalement on ne peut jamais aller très loin, et d'ailleurs les seuils de robustesse et de réalisme seraient vite dépassés). Souvent, d'ailleurs, l'estimation directe des moments d'ordre 4 n'apparaît même pas comme raisonnablement possible. On peut alors chercher à se tirer d'affaire en introduisant une hypothèse beaucoup plus forte, permettant de déduire, par le calcul, les moments d'ordre 4 à partir des moments d'ordre 2 (par exemple, on admettra que la loi spatiale est gaussienne, ou lognormale, etc...). Mais cette hypothèse supplémentaire (qui nous impose d'entrée de jeu le choix d'un modèle encore plus spécifié que précédemment) doit à son tour être justifiée. Et c'est de nouveau la course au

mauvais infini. En pratique, cependant, on ne se préoccupe pas trop de cette difficulté théorique : car une erreur sur la variance d'estimation est évidemment beaucoup moins grave qu'une erreur sur l'estimation elle-même (et ainsi de suite).

ESTIMER ET CHOISIR

TROISIEME PARTIE

LA RECONSTRUCTION OPERATOIRE

ESTIMER ET CHOISIR

- Essai sur la Pratique des Probabilités -

TROISIEME PARTIE

Table des Matières

<u>CHAPITRE VI - LES MODELES GLOBAUX</u>	122
Les représentations, ou modèles probabilistes strictement objectifs	124
Les représentations transitives	127
L'estimation du covariogramme transitif	130
Les formules d'approximation	133
Cas des mailles irrégulières	135
Passage aux modèles probabulistes usuels	137
Modèles globaux non stationnaires	140
<u>CHAPITRE VII -LES MODELES LOCAUX</u>	147
Les représentations glissantes	147
Priorité de la méthode	151
F.A. localement stationnaire d'ordre 2	156
F.A. localement intrinsèque d'ordre 0	157
F.A. localement intrinsèque d'ordre k	161
<u>CHAPITRE VIII -</u>	163
L'espérance conditionnelle est-elle opératoire?	163
Après coup : l'objectivité des lois conditionnelles	164
Quelques ordres de grandeurs	167
In praxi : l'estimation des lois conditionnelles	169
Les estimateurs disjonctifs	170

- CHAPITRE VI -

Les modèles globaux

Dans les trois derniers chapitres de cet ouvrage, je voudrais surtout m'attacher à mettre en lumière le contenu physique implicite des modèles probabilistes que nous utilisons pour représenter les phénomènes régionalisés. La richesse d'un concept physique dépend du nombre et de la variété des phénomènes entre lesquels il permet d'établir des rapprochements, donc de la richesse du réseau de relations et de lois physiques dans lesquelles il intervient, et qui constituent à proprement parler la définition même de ce concept en tant qu'opérateur. C'est donc à une véritable reconstruction opératoire de nos modèles probabilistes que nous devrions nous livrer : tâche immense, qu'il n'est évidemment pas question d'accomplir dans sa totalité. Je me contenterai donc d'examiner quelques cas particuliers typiques, sans d'ailleurs prétendre en aucune façon que la démarche que j'adopterai (passage par l'intermédiaire des "représentations probabilistes") soit la seule possible, ni même forcément la meilleure.

L'exemple du variogramme nous retiendra longtemps. Précisons bien qu'il ne s'agit pas du variogramme γ du modèle, dont le statut objectif est encore en partie incertain, mais de ce que nous avons appelé le variogramme régional γ_R : non pas l'espérance, mais la valeur moyenne dans l'espace physique du carré de la différence $z(x+h) - z(x)$. Réécrivons sa définition, c'est-à-dire la formule (2) du chapitre précédent :

$$(1) \quad \gamma_R(h) = \frac{1}{2K(h)} \int_{S(h)} [z(x+h) - z(x)]^2 dx$$

Dans cette écriture, rien n'évoque les probabilités. La régionale γ_R est une réalité physique, que nous pouvons considérer pour elle-même, indépendamment des interprétations (probabilistes ou non) que nous lui adjoindrons par la suite. Elle constitue, par elle-même, une sorte de résumé des caractéristiques structurales de la V.R., et véhicule une information déjà très riche, de nature physique et non probabiliste. Nous avons vu combien son comportement aux grandes valeurs de $|h|$ était révélateur pour un praticien : si les courbes observées dans les différentes directions de l'espace se stabilisent autour d'une asymptote horizontale commune (ou palier), nous avons un indice sûr de la "stationnarité" (en un sens physique), c'est-à-dire de l'homogénéité du phénomène dans l'espace. A cette réalité physique est associée un concept opératoire, celui de portée, dont nous avons vu la construction à partir d'une loi physique (variance de s dans S). Au contraire, si les courbes divergent violemment dans les différentes directions, nous disons qu'il y a une "dérive" :

concept plus difficile à cerner, sur lequel nous devons revenir, mais qui correspond pourtant à quelque forme de réalité physique, à une certaine allure de la variation spatiale de la V.R.

Plus riche encore d'enseignements, peut-être, se révèle le comportement du variogramme au voisinage de l'origine. De par sa définition même, d'ailleurs, γ_R caractérise les propriétés de régularité et de continuité du phénomène : ou, plus précisément, il nous donne une image moyenne de ces propriétés. Plus rapide est sa croissance, lorsque $|h|$ augmente, plus vite se détériore en moyenne "l'influence" d'un point x sur son voisin $x+h$, et donc plus discontinu et irrégulier se révèle le phénomène lui-même. On est amené à représenter le comportement de γ_R autour de $h = 0$ par des modèles simples, qui sont en fait des approximations, utilisables seulement à une certaine échelle de travail : par exemple, comportement parabolique (en $|h|^2$), ou linéaire (en $|h|$) ou "effet de pépite" (discontinuité à l'origine). Ces types de comportement répondent à des propriétés réelles du phénomène. Dans le cas parabolique, on a un phénomène très régulier, qu'il est possible de cartographier à cette échelle, avec tous ses détails. Dans le cas linéaire, la cartographie est encore, en gros possible, mais laisse échapper beaucoup de détails, liés à la trop grande irrégularité locale. Enfin, s'il y a un effet de pépite, on sait d'avance que toute carte sera, en grande partie, illusoire. Ces observations, purement empiriques, s'imposent au praticien avec une évidence ou une nécessité inéluctable : des lois physiques sont ici à l'oeuvre, bien que non formalisées encore, et vont permettre la reconstruction opératoire de la notion "d'allure du variogramme autour de l'origine".

Pour pousser plus loin l'analyse, nous allons nous ménager une étape intermédiaire : entre la V.R. initiale et le modèle probabiliste, nous allons insérer ce que nous appellerons des "représentations probabilistes", c'est-à-dire des modèles parfaitement objectifs qui ne font rien d'autre que de présenter sous forme probabiliste la totalité de l'information contenue dans la V.R. : modèles tautologiques, accessibles seulement après coup, et inutilisables en tant que tels. Leur intérêt sera méthodologique. Ils serviront de critères pour juger les modèles plus simples utilisés en pratique (stationnaires, intrinsèques etc...). Ces derniers ne tomberont plus tout faits du ciel des Idées. Leur statut sera celui de simples approximations anticipées de ces modèles strictement objectifs inaccessibles que nous appelons représentations. Le mystérieux problème de "l'inférence statistique" apparaîtra, lui aussi, sous un jour un peu nouveau. Dès lors, en effet, que l'essentiel de la réalité physique dont nous voulons rendre compte se trouve synthétisé dans le variogramme régional, le problème n'est plus "d'inférer" le mystérieux γ du modèle théorique, mais d'estimer l'intégrale d'espace qui figure dans la formule (1), sur la base des données dont nous

disposons in praxi : il ne s'agit donc plus "d'inférence statistique" mais du calcul numérique approché d'une intégrale à partir d'un petit nombre de points expérimentaux. Le problème n'est pas résolu pour autant, mais l'obscurité qui nous voilait sa véritable nature s'est quelque peu dissipée.

Dans le paragraphe suivant, nous définirons deux sortes de représentations probabilistes d'une V.R. : les représentations transitives, utiles pour l'étude des problèmes globaux (par exemple l'estimation de l'intégrale $Q = \int z(x) dx$), et les représentations glissantes, qui nous permettront d'aborder les problèmes locaux (par exemple, l'estimation de la valeur $z(x)$ en un point x donné). Le reste de ce chapitre sera consacré aux problèmes et aux modèles globaux, les modèles locaux faisant l'objet du chapitre suivant.

Les représentations, ou modèles probabilistes strictement objectifs

Avec le principe d'objectivité extrêmement strict que nous avons adopté, on peut à bon droit se demander s'il est seulement possible, pour une V.R. donnée, de trouver (après coup) un modèle probabiliste parfaitement objectif, c'est-à-dire dont tous les paramètres puissent être identifiés à des grandeurs régionales. Ce serait d'ailleurs peut-être une exigence excessive. Nous ne recherchons, en effet, que des modèles monoscopiques, c'est-à-dire élaborés expressément en vue de résoudre un problème donné, au moyen d'une méthode que nous avons choisie : la définition de ce problème et de cette méthode fait, en général, appel à certains paramètres, indépendants de la V.R., mais déterminés sans aucune équivoque. Nous les appellerons paramètres méthodologiques. Par exemple, supposons que le problème posé consiste à estimer l'intégrale $Q = \int z(x) dx$ d'une V.R. z dans l'espace à deux dimensions, à partir des résultats $z(x_\alpha)$ fournis par l'échantillonnage d'un réseau régulier de points expérimentaux x_α implantés selon une maille carrée ($a \times a$) : ce paramètre a , ou maille, n'est pas lié à la V.R., mais il est imposé par le problème à résoudre : c'est un paramètre méthodologique, que l'on peut, sans arbitraire, incorporer au modèle monoscopique à titre de paramètre auxiliaire.

Convenons de dire qu'un modèle est strictement objectif si ses critères de spécification ne font intervenir que des paramètres objectifs (identifiables à des grandeurs régionales) et des paramètres méthodologiques (imposés sans ambiguïté par le problème à résoudre et la méthode choisie) à l'exclusion de toute autre sorte de paramètres conventionnels. Nous allons voir qu'il est effectivement toujours possible, après coup, de définir des modèles strictement objectifs, qui nous auraient donc

permis de résoudre de manière parfaite le problème posé s'il nous avait été possible de les spécifier correctement in praxi : naturellement, il n'y a pas de miracle, et la spécification parfaite d'un modèle objectif reste hors d'atteinte, in praxi, précisément parce que nous ne connaissons pas les valeurs exactes des régionales associées aux paramètres objectifs. Théoriquement, c'est un cercle vicieux. Il est clair d'ailleurs qu'un modèle strictement objectif - et justement parce qu'il est strictement objectif - ne peut être que tautologique : il ne fait que présenter sous une autre forme la même information, celle qui est contenue dans la donnée de la V.R. elle-même, et qui n'est accessible qu'après coup. Il s'agit simplement d'une représentation de la même V.R.

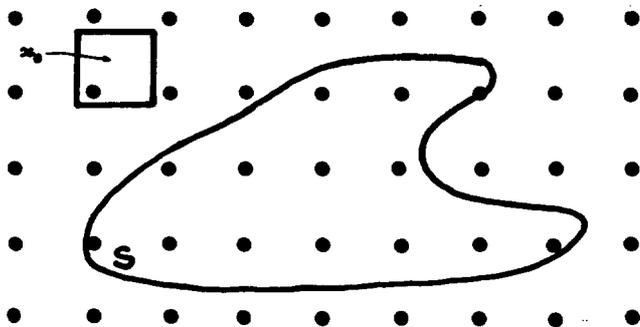
En pratique, pourtant, l'introduction de certaines hypothèses raisonnables d'approximation ou de simplification permet de briser le cercle : c'est ici que nous allons retrouver nos fameuses hypothèses anticipatrices (avec leur fécondité et la vulnérabilité qu'elle implique), et que nous pourrions essayer de préciser un peu mieux leur statut épistémologique, qui est celui d'une approximation anticipée : à la représentation rigoureuse de la V.R. (modèle trop riche pour être spécifiable in praxi) nous allons substituer une représentation approchée beaucoup plus simple, c'est-à-dire un type de modèle dont les paramètres puissent faire l'objet d'une estimation raisonnable in praxi, et conduise cependant à une solution du problème posé pratiquement équivalente à celle que l'on aurait choisie si l'on avait pu disposer de la représentation rigoureuse. Naturellement, c'est, comme toujours, après coup seulement que l'on peut affirmer cette équivalence pratique. Le risque d'erreur in praxi ne sera pas éliminé, il sera simplement mieux circonscrit.

Ceci dit, la représentation, ou modèle strictement objectif, est de nature probabiliste, alors que la situation réelle est unique, sans rien d'aléatoire, du moi au départ. Elle est caractérisée par deux éléments : d'une part, la V.R. z , définie sur un champ S , et connue après coup seulement ; d'autre part le réseau des points expérimentaux x_α , $\alpha = 1, 2, \dots, N$ en lesquels sont données in praxi les valeurs numériques $z_\alpha = z(x_\alpha)$. Pour probabiliser, nous devons entourer, en quelque sorte, cette situation actuelle d'un halo d'états virtuels considérés comme "possibles", et admettre, d'une façon ou d'une autre, que l'état actuel n'est que l'un, quelconque, de ces états possibles, et ne se singularise par rien de particulier. On achèvera la construction du modèle en se donnant une loi de probabilité sur cet ensemble d'états considérés comme possibles. Ici, le principal danger à éviter est évidemment l'arbitraire. Dans la définition et la probabilisation de ces états virtuels, nous ne pouvons

utiliser (si nous voulons rester strictement objectifs) rien d'autre que la V.R. elle-même et la structure du réseau d'information. Mais nous devons en même temps redonner une certaine mobilité relative à ces deux éléments du problème, afin précisément d'engendrer une famille d'états virtuels réalisables. L'idée directrice, très simple, consiste à imaginer que le réseau d'information se déplace (ou se déforme) dans le champ S de la V.R.

Pour fixer les idées, considérons le cas le plus simple possible, celui où les points expérimentaux x_α forment un réseau régulier, à maille carrée ou rectangulaire dans l'espace à deux dimensions, et où le problème posé est celui d'une estimation linéaire.

Les choses se présentent d'ailleurs de façon assez différente selon qu'il s'agit d'une estimation globale ou locale.



a - Estimation globale : Il s'agit, par exemple, d'estimer l'intégrale

$$Q = \int (x) dx$$

Fig. 7. Estimation globale

Fig. 7

La forme exacte du champ S de la V.R. est inconnue (in praxi), mais le problème de l'estimation globale ne se pose, raisonnablement, en pratique, que si le réseau régulier des points réels x_α , en nombre nécessairement fini, déborde en fait suffisamment le champ S pour que l'on puisse sans inconvénient le prolonger à l'infini (en attribuant la valeur zéro à ces points fictifs supplémentaires). Ce point de vue correspond aux représentations transitives¹ : l'un, quelconque, des points expérimentaux, soit x_0 , choisi comme origine du réseau, est supposé se déplacer dans le rectangle défini par la maille, en entraînant avec lui l'ensemble du réseau. Ce déplacement définit l'ensemble des états virtuels, et il suffit ensuite de considérer ce point x_0 comme un point aléatoire uniformément distribué dans le rectangle de maille pour achever la définition du modèle.

1 - Le mot "transitif" est employé, ici, comme un équivalent spatial de "transitoire". Il indique que nous désirons tenir compte, dans notre modèle, des phénomènes de discontinuité qui s'observent lorsque l'on franchit les frontières du champ S.

b - Dans un problème d'estimation locale, il s'agit au contraire d'estimer

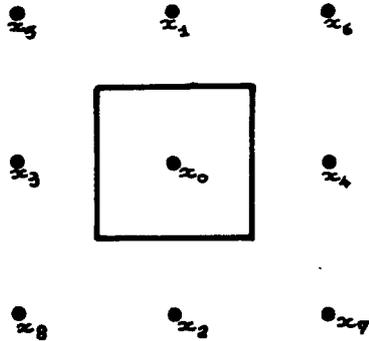


Fig. 8 - Estimation locale

Fig. 8

une figure liée au réseau de prélèvements, par exemple la valeur moyenne de $z(x)$ dans la "zone d'influence" $\Pi(x_0)$ d'un point x_0 du réseau (c'est-à-dire dans le rectangle de maille centré en ce point x_0). Compte tenu des considérations sur la robustesse exposées précédemment on a décidé de n'utiliser, pour cette estimation, que les points expérimentaux les plus proches, soit par exemple x_0 lui-même et ses huit voisins immédiats x_1, \dots, x_8 . Dans ces

conditions, c'est la configuration complète formée par $\Pi(x_0)$ et les neuf points x_0, x_1, \dots, x_8 qui sera censée se déplacer dans l'espace : le point x_0 , entraînant avec lui l'ensemble de la configuration, décrira un domaine S_0 (choisi par nous : S_0 , en général, sera un sous-ensemble du champ total S , sélectionné, in praxi, au vu des résultats de l'échantillonnage lui-même). Il suffira, ensuite, de considérer x_0 comme un point aléatoire uniformément distribué dans S_0 pour achever la probabilisation. Ce second point de vue correspond à ce que nous appellerons les représentations glissantes. Dans le premier point de vue, la zone à estimer (identique au champ total S) restait fixe dans l'espace, tandis que le réseau de prélèvements se déplaçait par translation (dans la terminologie minière, ce dispositif est dit du type "champ fixe et implantation flottante"). Maintenant, au contraire, c'est la configuration complète, constituée par les points expérimentaux et la zone à estimer, qui se déplace dans son ensemble (dans la terminologie minière, on parle, d'un dispositif du type "champ flottant et implantation préférentielle" : étant entendu que le champ flottant correspond à la zone à estimer $\Pi(x_0)$ et non plus au champ total S qui reste évidemment fixe dans l'espace).

Les représentations transitives

Outre la V.R. z elle-même, définie sur un champ S (inconnu) et prolongée à l'infini en posant $z(x) = 0$ lorsque x n'appartient pas à S , nous considérerons un réseau de "points de prélèvement" $x_i = x_0 + h_i$ implantés selon une maille régulière. Pour fixer les idées, prenons le cas d'une maille rectangulaire dans l'espace à deux dimensions (on peut, évidemment, envisager d'autres types de maille, et traiter aussi le cas de l'espace à 3 dimensions). La maille est définie par un rectangle de base π , de côtés a_1 et a_2 , centré à l'origine des coordonnées. Les vecteurs h_i de la maille

admettent comme composantes, selon les deux axes de coordonnées, les multiples entiers de a_1 et a_2 respectivement. L'origine x_0 du réseau est "implantée au hasard dans π " (ce qui veut dire que nous prenons comme espace Ω le rectangle π , et, comme probabilité P , la loi uniforme dans π). Les données expérimentales $z(x_i)$, implantées aux points $x_i = x_0 + h_i$, deviennent donc elles-mêmes, dans le modèle, des variables aléatoires (ce sont, en effet, des fonctions de $\omega = x_0$).

Ce schéma correspond bien à la situation, très fréquente en pratique, où on ne connaît pas (in praxi) les limites exactes du champ S de la V.R., et où le réseau, dans son ensemble, a été implanté "n'importe où" par rapport à ce champ. L'espace des états virtuels est constitué des diverses implantations possibles du même réseau, considérées comme "également probables", d'où le choix d'une densité uniforme dans Π . Cette probabilisation correspond, en somme, à une translation aléatoire affectant l'ensemble du réseau. On pourrait penser "enrichir" le modèle en ajoutant aussi une rotation aléatoire. Mais ce ne serait pas réaliste. Car les directions préférentielles, s'il en existe, que présente la variation spatiale du phénomène sont, en général, fort bien décelées expérimentalement dès que l'on dispose des résultats d'une maille régulière. Il ne serait donc certainement pas correct de considérer, parmi les états virtuels, toutes les orientations possibles du réseau par rapport à la V.R. comme également probables (puisque en fait l'une d'elles a été réalisée et qu'on la connaît). Pour ce qui est des translations affectant l'origine du réseau, par contre, la situation est bien différente : on peut certes penser que certaines implantations du point $x_0 \in \Pi$ conduiront à des résultats exceptionnellement favorables ou défavorables, mais on ne sait (in praxi) absolument pas lesquelles. Il est donc bien "naturel" de considérer toutes ces implantations possibles comme également probables. Il s'agit, bien sûr, d'un choix méthodologique, mais dicté sans équivoque par la nature du problème posé (estimation globale) et la structure de l'information (réseau régulier implanté en l'absence d'information précise sur les limites du champ).

La quantité à estimer ($Q = \int z(x) dx$) est une grandeur régionale à caractère global. L'estimateur linéaire le plus simple possible que l'on puisse envisager est :

$$(2) \quad Q^*(x_0) = v \sum z(x_0 + h_i)$$

où v est la mesure du motif de base Π (longueur, surface ou volume, selon que l'espace a 1, 2 ou 3 dimensions). Comme x_0 est (dans le modèle) aléatoire dans Π , cet estimateur $Q^*(x_0)$ est une variable aléatoire définie sur l'espace $\Omega = \Pi$ muni de la probabilité uniforme dans Π . D'après la manière même dont cette V.A. a été définie, toutes ses propriétés s'exprimeront à l'aide des seules caractéristiques de Π (paramètres

méthodologiques) et de la V.R. z (paramètres objectifs). Il s'agit donc bien d'un modèle strictement objectif (ou représentation).

L'espérance de cette variable est, par définition

$$E[Q^*] = \frac{1}{v} \int_{\pi} Q^*(x_0) dx_0$$

En remplaçant $Q^*(x_0)$ par son expression (2), on voit sans peine que la sommation sur tous les vecteurs h_i de la maille, jointe à l'intégration dans le rectangle de base, reconstitue l'intégrale étendue à l'espace tout entier, de sorte qu'il reste

$$E(Q^*) = \int z(x) dx = Q$$

Il s'agit donc (dans le modèle) d'un estimateur sans biais du paramètre (objectif) Q que nous cherchons à estimer. Un calcul tout à fait analogue montre que le moment d'ordre 2, c'est-à-dire l'espérance de $(Q^*)^2$ est :

$$E[(Q^*)^2] = v \sum_i \int z(x) z(x+h_i) dx$$

Nous sommes ainsi conduits à associer à notre V.R. z la fonction g , appelée covariogramme transitif, définie par la relation :

$$(3) \quad g(h) = \int z(x) z(x+h) dx$$

Le moment d'ordre 2 s'écrit alors simplement $E[(Q^*)^2] = v \sum g(h_i)$. En intégrant en h la fonction définie en (3), on fait apparaître le carré de la quantité $Q = \int z(x) dx$ qu'il s'agit d'estimer, soit :

$$\int g(h) dh = Q^2$$

Or, dans le modèle, la variable aléatoire Q^* admet une variance $\text{Var } Q^* = E[(Q^*)^2] - Q^2$, que nous appellerons variance d'estimation. D'après ce qui précède, elle est donnée par la relation :

$$(4) \quad \text{Var } Q^* = v \sum_i g(h_i) - \int g(h) dh$$

Cette formule mérite un commentaire. Elle relie une notion objective, définie sans ambiguïté, la variance d'estimation, à une fonction très simple, le covariogramme transitif g : celui-ci, comme on le voit sur la formule (3), constitue en quelque sorte l'équivalent "déterministe" de la covariance $C(h) = E [Z(x) Z(x+h)]$ d'une F.A. stationnaire Z . Il rendra, en tout cas, les mêmes services dans la pratique. Mais il n'a été nécessaire d'introduire aucune "hypothèse" de stationnarité concernant la V.R. ; celle-ci n'a d'ailleurs à aucun moment été interprétée comme la réalisation d'une fonction aléatoire. La stationnarité qui se manifeste ici n'est pas une propriété du phénomène, mais seulement une caractéristique de notre réseau de prélèvements. Notre covariogramme transitif contient exactement la même information structurale qu'une covariance stationnaire, mais présente sur cette dernière un avantage décisif : il est défini sans ambiguïté par la relation (3), et sa signification est purement objective. Si nous nous contentons, comme on le fait le plus souvent en pratique, de caractériser la précision de l'estimation globale par la seule variance d'estimation, la formule (4) montre qu'il s'agit d'un problème bien posé, admettant une solution définie sans équivoque, du moins après coup.

La signification physique de cette formule (4) est d'ailleurs simple et facile à saisir. La variance d'estimation apparaît, en effet, comme la différence entre la valeur approchée et la valeur exacte d'une même intégrale $\int g(h) dh$, donc, si l'on veut, comme l'erreur que l'on commettrait en évaluant cette intégrale à l'aide d'une sommation discrète sur les points h_i de la maille. D'après ce que l'on sait du calcul numérique approché des intégrales, on doit donc s'attendre à ce que cette variance soit d'autant plus petite que la maille est plus serrée, ce qui est bien naturel ; et d'autant plus petite, aussi, que la fonction g est plus régulière. Comme la régularité de g reflète les propriétés moyennes de continuité de la variable régionalisée z elle-même, il apparaît donc que, pour une maille donnée, l'estimation sera d'autant plus précise que la V.R. sera elle-même plus régulière et continue dans sa variation spatiale : ceci est bien conforme à ce que l'intuition nous suggère, et donne au contenu de cette intuition une expression mathématique rigoureuse.

L'estimation du covariogramme transitif

Il reste à examiner comment les choses se présentent in praxi. Du fait que les paramètres méthodologiques (v et les vecteurs h_i) sont donnés au départ, le seul problème qui se pose (mais il est sérieux) est celui de l'estimation du covariogramme transitif g .

De prime abord, en effet, cette estimation semble poser un problème plus difficile que le problème initial (l'estimation de Q), et comporte deux parties : d'une part, estimer les valeurs $g(h_i)$ du covariogramme pour les arguments h_i correspondant aux vecteurs du réseau (les seuls pour lesquels on ait une information expérimentale directement utilisable). D'autre part, interpoler entre ces valeurs estimées, de manière à obtenir une évaluation de $g(h)$ pour les autres vecteurs h .

Pour un vecteur donné h_i , l'estimation de $g(h_i) = \int z(x) z(x+h_i) dx$ est analogue à celle de Q (la V.R. $z(x)$ étant simplement remplacée par $z(x) z(x+h_i)$), et ceci conduit à prendre l'estimateur :

$$g^{\mathbf{x}}(h_i) = v \sum_j z(x_o+h_j) z(x_o+h_i+h_j)$$

A cause de l'invariance du réseau par translation, les points $(x_o+h_i+h_j)$ sont, en effet, disponibles expérimentalement en même temps que les (x_o+h_j) . Pour h_i fixé, le produit $z(x) z(x+h_i)$ est lui-même une fonction de x , c'est-à-dire une V.R. Toutefois, le champ de cette nouvelle V.R. est l'intersection $S \cap S-h_i$, donc plus petit que le champ initial, et les données utiles (différentes de 0) sont donc moins nombreuses que celles dont on dispose pour estimer Q . On pourrait songer, pour apprécier la précision de cette estimation, à appliquer la formule (4) au covariogramme associé à la V.R. produit $z(x) z(x+h_i)$, et ainsi de suite : ceci nous conduirait à une désastreuse régression à l'infini, introduisant des covariogrammes d'ordre de plus en plus élevé, qu'il faudrait estimer à partir de données de moins en moins nombreuses (du fait de la diminution du champ utile), et d'ailleurs dans des conditions de plus en plus douteuses (on sait bien, en statistique, que l'estimation des moments d'ordre un peu élevé perd vite toute robustesse). Il ne sera donc pas possible d'aller bien loin dans cette direction, et, en pratique, on renoncera souvent à associer une variar à l'estimateur $g^{\mathbf{x}}(h_i)$.

Le second problème est celui de l'interpolation de g entre les valeurs (estimées $g^{\mathbf{x}}(h_i)$). A dire vrai, d'après la formule (4), nous n'avons besoin - en apparence - que de l'intégrale $\int g(h) dh = Q^2$, et on pourrait donc penser lui substituer son estimation $(Q^{\mathbf{x}})^2$.

Malheureusement, cette substitution conduirait à attribuer à $\text{Var } Q^{\mathbf{x}}$ la valeur 0, comme cela résulte de l'identité facile à vérifier :

$$v \sum g^{\mathbf{x}}(h_i) - (Q^{\mathbf{x}})^2 = 0$$

Ce résultat se comprend facilement : en substituant à Q et aux $g(h_i)$ leurs estimations Q^* et $g^*(h_i)$ obtenues à partir du seul réseau de points $x_0 + h_i$, nous avons en somme remplacé le problème de l'estimation de l'intégrale $Q = \int z(x) dx$ par celui de la somme discrète $Q^* = v \sum z(x_0 + h_i)$ - et l'identité ci-dessus signifie que $Q^*(x_0)$ est un excellent estimateur de $Q^*(x_0)$ lui-même. En termes moins triviaux, cela veut dire qu'on ne peut pas déduire des mêmes données à la fois une estimation et la précision de cette estimation.

La conclusion est donc (comme toujours) que l'on doit nécessairement introduire une hypothèse anticipatrice d'approximation, c'est-à-dire remplacer le vrai $g(h)$ (inconnu) par une fonction $\bar{g}(h; \lambda, \mu)$ d'un type convenablement choisi et dépendant d'un nombre peu élevé de paramètres λ, μ, \dots , que l'on spécifiera au mieux à l'aide des données expérimentales (les $g^*(h_i)$). Ce choix présente une importance cruciale, puisque en définitive le résultat que nous obtiendrons en appliquant la formule (4) au modèle \bar{g} (une fois spécifié) n'aura de valeur que dans la mesure où ce modèle simplifié représentera une approximation acceptable du vrai $g(h)$ inconnu. La pire des choses à faire ici serait de s'en remettre aveuglément à quelque procédé automatique pour interpoler entre les valeurs expérimentales $g^*(h_i)$.

Pour faire ce choix en connaissance de cause, il convient d'examiner de plus près la structure de la formule (4), de manière à identifier, autant qu'il est possible, les facteurs qui exercent le plus d'influence sur la variance d'estimation. D'après sa définition (3), le covariogramme transitif g est une fonction de "type positif". C'est d'ailleurs cette condition qui nous garantit que l'expression (4) est nécessairement positive, comme il convient à une variance. Le modèle \bar{g} que nous choisirons devra donc, lui-même, être une fonction de ce type. Or, les irrégularités de ces fonctions se localisent, pour l'essentiel, au voisinage de l'origine, et voici donc, à nouveau, notre attention attirée sur le comportement du covariogramme autour de $h=0$. De fait, une analyse fine montre que la valeur numérique fournie par la formule (4) dépend principalement du comportement du $g(h)$ sur un voisinage de l'origine dont les dimensions sont comparables à celle du motif de base Π du réseau de prélèvement. Le point crucial concerne donc l'interpolation de $\bar{g}(h)$ entre $g^*(0)$ et les premiers points expérimentaux disponibles au voisinage de 0, et le type de comportement analytique que présente le modèle choisi \bar{g} dans cette zone (par exemple : linéaire avec effet de pépite, ou comportement en $|h|^\lambda$, $0 < \lambda \leq 2$: c'est en fait le choix de ce type de comportement au voisinage de $h=0$ qui constitue l'essentiel de l'hypothèse anticipatrice.

Les formules d'approximation

Il y a un très grand intérêt, du point de vue physique, à étudier le comportement de la variance d'estimation lorsque la maille devient très petite, et à le relier à celui que présente le covariogramme g autour de $h=0$. Plaçons nous d'abord dans le cas de l'espace à une seule dimension (la droite) et désignons par L (au lieu de S) la longueur du champ de la V.R. : L est aussi la portée de $g(h)$, c'est-à-dire la distance au delà de laquelle cette fonction devient identiquement nulle. En $h = L$, $g(h)$ présente certaines irrégularités analytiques, liées à la manière, plus ou moins brutale, dont se fait en ce point le raccordement avec l'axe des h . Désignons par a la maille, ou équidistance des points de prélèvements, et par $\sigma^2(a)$ la variance d'estimation, telle qu'elle est donnée par la formule (4). On trouve alors que cette variance est la somme de deux termes.

$$\sigma^2(a) = T(a) + T'(a)$$

Le premier terme, $T(a)$ est lié au comportement de g en $h = 0$; le second, $T'(a)$ au comportement de cette fonction autour de la portée. $T'(a)$ dépend, pour l'essentiel, de la quantité $\varepsilon = L/a$ modulo 1 (c'est-à-dire $\varepsilon = (L-na)/a$, où n est l'entier tel que $na \leq L < (n+1)a$). C'est une fonction périodique de ε , de période 1 et sans terme constant, donc de moyenne nulle. On l'appelle, pour cette raison, terme fluctuant (ou Zitterbewegung). Son amplitude peut être considérable, comme le montre la Fig. 9.

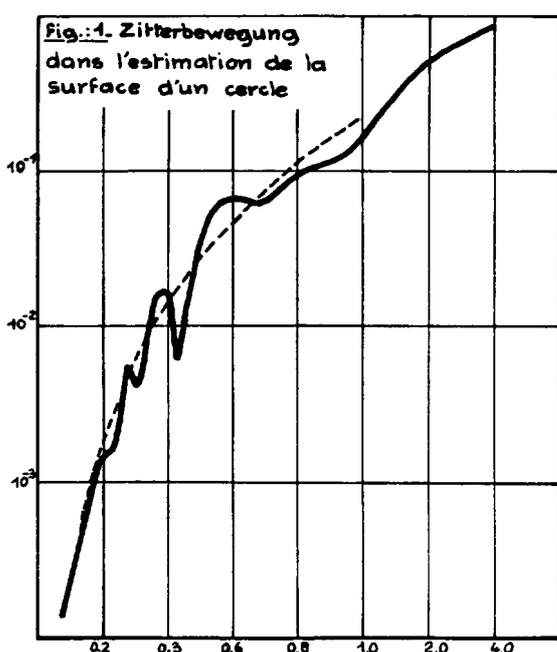


Fig. 9

Zitterbewegung dans l'estimation de la surface du cercle de diamètre unité à l'aide d'un réseau à maille carrée. En abscisse la maille a . En ordonnée la variance d'estimation correspondante $\sigma^2(a)$. La courbe en trait plein représente la valeur exacte, calculée à partir de la formule exacte (IV,1,12). La courbe en pointillé représente la formule $\sigma^2(a) = 0.2276 a^3 + 0.00477 a^5$ obtenue en négligeant le Zitterbewegung et en retenant les deux premiers termes du développement limité donné par le principe de correspondance.

Bien que l'on sache former son expression théorique, on ne peut jamais, dans les applications, prévoir sa valeur exacte : car, justement, la valeur exacte de la portée L n'est connue qu'à $\pm a$ près, et ε , qui est L/a modulo 1 est donc totalement indéterminé. On s'autorise alors du fait que la valeur moyenne, en ε , de ce terme est égale à 0 pour le négliger purement et simplement. On notera le caractère probabiliste camouflé de cette approximation. Tout se passe, en effet, à peu près comme si l'on considérait $\sigma^2(a)$ comme une variable aléatoire, dont l'espérance serait le terme régulier $T(a)$, le terme fluctuant $T'(a)$ représentant la partie purement aléatoire et imprévisible. De fait, la Fig. 9 évoque la situation pré-aléatoire analysée par J. Ullmo, où des conditions initiales inséparables entraînent ultérieurement une séparation des phénomènes observés.

Quant au terme régulier $T(a)$, il est lié au comportement de $g(h)$ autour de $h=0$. Par exemple, dans un modèle où $g(h)$ est de la forme $g(h) = g(0) - A|h|^\lambda + \dots$ ($0 < \lambda \leq 2$), on trouve, lorsque la maille a est petite, une expression asymptotique du type :

$$T(a) = A T_\lambda a^{1+\lambda}$$

(T_λ est un coefficient numérique connu, par exemple $T_\lambda = 1/6$ si $\lambda = 1$ etc...). Désignons par n le nombre des points de prélèvements x_i où $z(x_i)$ n'est pas nul : on a, à peu près, $n = L/a$. En remplaçant a par L/n , on voit que la formule ci-dessus nous amène à prévoir (pour n grand) une variance d'estimation de la forme :

$$(5) \quad \sigma^2(a) = \frac{C}{n^{1+\lambda}}$$

Dans le cas (fréquent) d'un covariogramme "linéaire" au voisinage de $h = 0$, soit $\lambda = 1$, la variance varie donc en raison inverse du carré du nombre des échantillons utiles, du moins lorsque ce nombre n est grand.

Dans le cas de l'espace à deux dimensions et d'une maille rectangulaire $a_1 \times a_2$, on obtient des résultats analogues. Il y a encore un terme fluctuant, nullement négligeable comme le montre la figure 9, et un terme régulier, qui est cette fois de degré $2+\lambda$ (plus précisément, si $a_1 \leq a_2$, il comporte un terme en $(a_2)^{2+\lambda}$ et un autre en $a_2(a_1)^{1+\lambda}$). Prenons, par exemple, le cas d'une maille carrée $a_1=a_2=a$: la variance est proportionnelle à $a^{2+\lambda}$. Comme le nombre n des données utiles est à peu près $n = S/a^2$, on trouve cette fois une loi de la forme :

$$\sigma^2(a) = \frac{C}{n^{1+\frac{\lambda}{2}}}$$

Par exemple, dans le cas fréquent où λ est égal à 1, la variance d'estimation est, cette fois, en raison inverse de la puissance 3/2 du nombre des données utiles.

Ces relations (5) ou (6) constituent, en fait, des lois physiques. Après coup du moins, il est possible, en effet, de les contrôler expérimentalement. Elles nous montrent de quelle manière l'ordre de grandeur de la variance d'estimation se relie au nombre des échantillons. Et, puisqu'il s'agit de lois physiques, le paramètre λ qu'elles mettent en jeu reçoit, du même coup, un statut opératoire : elles nous garantissent que le modèle de comportement que nous avons attribué au covariogramme g représente autre chose qu'une simple approximation graphique : à savoir un concept physique (celui du comportement de g au voisinage de l'origine, représenté par un "degré de régularité" λ) ainsi que les lois physiques qui constituent ce concept et permettent de le mesurer.

Cas des mailles irrégulières

Le covariogramme transitif $g(h)$, qui s'introduit de manière toute naturelle dans l'expression (4) de la variance d'estimation, présente de grandes analogies formelles avec une covariance stationnaire, et le problème que pose son estimation in praxi n'est pas non plus sans rappeler les difficultés que soulève "l'inférence statistique d'une fonction de covariance à partir d'une réalisation d'une F.A. stationnaire. Pourtant, nous n'avons introduit aucune hypothèse de stationnarité relative à la V.R. qui n'a d'ailleurs même pas été considérée comme une réalisation d'une F.A., et a conservé son statut "déterministe". Par contre, le réseau régulier de prélèvements, compte-tenu de la loi de distribution uniforme dans le motif de base Π attribuée à son origine x_0 , a été en fait traité comme un processus ponctuel stationnaire. C'est la stationnarité du réseau qui nous a permis, en l'absence de toute hypothèse portant sur la V.R., de bénéficier des circonstances très avantageuses que l'on croit généralement liées à la stationnarité du phénomène lui-même, ou de la V.R. qui le représentent. Nous reviendrons sur cette remarque importante : la stationnarité peut souvent être introduite, non pas à titre d'hypothèse relative à la réalité physique, mais simplement comme une caractéristique de la méthode d'estimation choisie par nous.

Dans le cas, maintenant, où les points expérimentaux sont implantés de manière quelconque dans le champ S et à son voisinage, et ne constituent plus une maille régulière, nous ne pouvons plus, sans artifice, imaginer que ce réseau se déplace dans l'espace sans se déformer. L'idée qui vient donc à l'esprit consiste à considérer ce réseau comme une réalisation d'un processus ponctuel. Si de plus ce processus peut être considéré comme stationnaire, on trouvera de nouveau, pour la variance d'estimation, des formules où la V.R. n'interviendra que par l'intermédiaire de son covarogramme transitif $g(h)$: ici encore le principe de la démarche consiste à attribuer la stationnarité au dispositif informant (le réseau) et non à la réalité physique (la V.R.).

Naturellement, il convient de s'interroger sur la signification d'une hypothèse, où plutôt d'une décision méthodologique, de ce genre. Il s'agit, en effet, d'une décision constitutive, qui définit, en particulier, la notion de variance d'estimation comme une fonction des caractéristiques que nous choisissons d'attribuer à ce processus ponctuel : pour que la notion ainsi construite puisse être considérée comme réaliste et utilisable en pratique, il faut que le réseau de reconnaissance présente un degré suffisant d'homogénéité dans l'espace pour qu'il soit au moins plausible de l'assimiler à une réalisation d'un processus ponctuel stationnaire admettant telles et telles caractéristiques.

Notons bien qu'en général (à la différence de ce qui se passe lorsqu'il s'agit d'avancer une "hypothèse" de stationnarité concernant la réalité physique) nous avons en main, in praxi, tous les éléments d'appréciation si nous avons nous-même implanté le réseau, ou si nous disposons à tout le moins de l'historique de la reconnaissance. Il y a des cas sans équivoque (par exemple, celui où les implantations des points ont été réellement tirées au sort selon un schéma aléatoire pur ou aléatoire stratifié). Souvent aussi le réseau se présente comme très hétérogène, mais peut sans trop d'ambiguïté se laisser diviser en deux ou plusieurs sous-zones homogènes, qu'il y a alors tout intérêt à traiter séparément. Cette circonstance se rencontre fréquemment en géostatistique minière, lorsque telle ou telle partie du gisement a fait l'objet d'une reconnaissance plus poussée que le reste (en raison de sa richesse particulière, ou simplement de la planification de l'exploitation). Mais il y a aussi des cas beaucoup plus difficiles, par exemple lorsque l'on sait que l'implantation du réseau a été influencée par des hypothèses (vraies ou fausses d'ailleurs) concernant la V.R. Ainsi, en matière minière, on sait qu'au stade de la première reconnaissance les sondages sont souvent implantés en fonction d'un objectif structurel, par exemple, pour vérifier une hypothèse de nature géologique etc... Cet effet préférentiel est générateur de biais importants mais difficiles à apprécier, et interdit pratiquement toute estimation quantitative précise.

Du point de vue mathématique, ce modèle (réseau de prélèvements considéré comme un processus ponctuel stationnaire) conduit à des formules analogues à (4). Ici encore on peut mettre en évidence le rôle crucial que joue le comportement du covariogramme au voisinage de l'origine. Par exemple, dans le cas d'une maille "aléatoire stratifiée" on trouve, pour la variance d'estimation, l'expression :

$$\text{Var } Q^x = v [g(0) - g(\pi)]$$

où $g(\pi)$ représente la valeur moyenne de $g(x-y)$ lorsque les deux points x et y décrivent le rectangle Π définissant la maille : seules interviennent dans cette formule les valeurs que prend g sur un voisinage de l'origine dont les dimensions sont du même ordre de grandeur que la maille elle-même.

Passage aux modèles probabilistes usuels

Dans les applications pratiques, on utilise rarement les représentations transitives elles-mêmes, et on préfère leur substituer des modèles probabilistes de type usuel plus faciles à mettre en oeuvre. Les représentations transitives, en effet, travaillent avec des intégrales, plutôt qu'avec des valeurs moyennes, d'où une certaine lourdeur dans leur maniement. Considérons, par exemple, la définition (3) du covariogramme transitif : l'intégrale est, en réalité, étendue au domaine $S(h)$ où $z(x)$ et $z(x+h)$ sont toutes deux différentes de 0 : $S(h)$ est l'intersection du champ S et du domaine qui s'en déduit par la translation $-h$. Nous désignerons par $K(h)$ la mesure de $S(h)$. Les valeurs numériques du covariogramme sont très influencées par cette fonction $K(h)$ qui, à son tour, reflète les propriétés géométriques du champ S plutôt que la variabilité de la V.R. $z(x)$ elle-même. Il est donc intéressant de tenter de séparer ces deux sortes d'effet, en remplaçant l'intégrale par la valeur moyenne correspondante, c'est-à-dire en posant :

$$(7) \quad C_R(h) = \frac{1}{K(h)} \int_{S(h)} z(x) z(x+h) dx = \frac{g(h)}{K(h)}$$

Dans certains cas, cette séparation est illusoire. Par exemple, si la V.R. présente une allure zonale typique, caractérisée par une décroissance plus ou moins continue à partir d'un coeur riche, la géométrie du champ est liée de manière trop intime à l'allure de la variation spatiale de la V.R. pour qu'il soit possible de distinguer ces deux facteurs. Dans d'autres cas, par contre, on a l'impression que le

champ S découpe, comme à l'emporte pièce, une régionalisation qui aurait pu, aussi bien, se poursuivre bien au delà : dans ce dernier cas, qui est celui des phénomènes présentant une certaine forme de stationnarité (au sens physique), $c_R(h)$ représente, mieux que $g(h)$, les propriétés de la V.R. z "elle-même", considérée indépendamment de la géométrie de son champ.

Cette fonction $C_R(h)$ a, évidemment, le sens d'une covariance. De fait, si nous choisissons le modèle "F.A. stationnaire", l'expression $C_R(h)$ écrite en (7) représente exactement "l'estimateur" que l'on utiliserait après coup pour procéder à ce que le point de vue classique appelle "l'inférence statistique" de la "vraie" covariance $C(h)$ (non centrée) du modèle. Nous savons, en fait, que la moyenne d'espace $C_R(h)$ épuise le contenu objectif de la notion de covariance, puisqu'aucune information expérimentale supplémentaire ne nous permettra jamais de remonter plus loin en direction de la covariance idéale. Nous pourrions donc identifier ces deux fonctions, et choisir justement $C_R(h)$ comme covariance du modèle¹. Mais cette manière de voir laisse échapper un point important. De fait, le covariogramme transitif, à supposer que nous le connaissions après coup, présenterait de petites ondulations, des points singuliers, toute une structure de détail qui fournirait autant d'information ou presque, que la donnée de la V.R. elle-même. Mais cette riche structure de détail est absolument inaccessible in praxi. Même après coup, d'ailleurs, nous serions amenés à simplifier, à remplacer le vrai $g(h)$, trop complexe, par un modèle plus accessible. En particulier, comme nous l'avons vu, nous introduisons une hypothèse concernant le comportement (en $|h|^\lambda$ par exemple) de cette fonction au voisinage de $h=0$: hypothèse ayant une signification parfaitement objective², dans la mesure où elle entraîne des relations, comme (5) et (6), qui ont le sens de lois physiques, et se prêtent à un contrôle expérimental. Mais on peut, naturellement, aussi interpréter ce remplacement de la vraie fonction g , ou C_R , par un modèle plus simple comme un passage à l'espérance mathématique : c'est alors ce modèle simplifié, ayant un sens objectif, qui peut servir de définition à la covariance $C(h)$.

1 - Sous réserve de vérifier que cette fonction est bien de type positif ; g et K sont de type positifs, par construction, mais non pas forcément leur rapport.

2 - Au sens du critère popperien (falsifiabilité), et non au sens du critère trop strict de décidabilité en termes de régionales (cf. ch. 4, le paragraphe consacré au modèle primaire).

D'autre part, et du moins tant que nous ne nous intéressons qu'à l'estimation globale, nous n'avons aucune raison de limiter notre choix aux seules modèles stationnaires : les résultats essentiels que nous a fournis l'étude des représentations transitives ne dépendent, en effet, en aucune façon d'une hypothèse quelconque concernant la stationnarité physique du phénomène. Ils sont liés seulement à la stationnarité du réseau de prélèvements. Ceci suggère la démarche suivante: sans risque aucun de démenti expérimental, nous pouvons considérer la V.R. $z(x)$ comme une réalisation d'une F.A. non stationnaire d'ordre 2, $Z(x)$, caractérisée par une covariance $C(x; y) = E [Z(x) Z(y)]$ dépendant séparément des points d'appui x et y , et non plus seulement de leur différence $x-y$. Dans l'optique classique, cette covariance stationnaire ne peut pas faire l'objet d'une "inférence statistique", puisque nous ne disposons que d'une seule réalisation. Nous préférons dire qu'elle ne présente pas de signification objective, et joue dans notre modèle un rôle purement conventionnel. Cela n'a aucune importance, car nous n'avons pas réellement besoin de la connaître. En effet, dans ce modèle, le covariogramme transitif devient une fonction aléatoire de h , à savoir

$$G(h) = \int_{S(h)} Z(x) Z(x+h) dx$$

L'espérance de cette V.A. $G(h)$ peut être identifiée, non pas à $g(h)$ lui-même, mais au modèle simplifiée $g(h)$ que nous lui avons substitué, celui qui comportait (par exemple) un comportement en $|h|^\lambda$ au voisinage de l'origine, et dont nous avons souligné la signification physique. Il est facile de voir que cette espérance $E[G(h)] = \bar{g}(h)$ s'exprime de manière simple à l'aide de la covariance non stationnaire $C(x,y)$. Désignons, en effet, par $\bar{C}(h)$ la valeur moyenne de la covariance $C(x, x+h)$ relative à deux points x et $y = x+h$ distants de h , lorsque le point x parcourt le domaine $S(h)$. On trouve la relation simple :

$$\bar{g}(h) = K(h) \bar{C}(h)$$

Cette relation montre que l'on a besoin de connaître, non pas la totalité de la fonction $C(x,y)$, mais seulement la "covariance moyenne" $\bar{C}(h)$. Plus précisément, même, nous n'avons besoin de connaître que le comportement de $\bar{C}(h)$ autour de $h=0$, et son allure analytique. Nous pouvons aussi bien utiliser le variogramme moyen de ce modèle, soit :

$$\bar{\gamma}(h) = \bar{C}(0) - \bar{C}(h)$$

Entre ce variogramme $\bar{\gamma}$ et la régionale définie en (1), il y a le même rapport qu'entre le modèle $\bar{g} = E(G)$ et le covariogramme transitif $g : \bar{\gamma}$ est une version simplifiée de γ_R , présentant par exemple en $h=0$ un comportement analytique en $|h|^\lambda$, tout en conservant le sens objectif et le riche contenu physique que nous avons analysé.

D'un point de vue purement monoscopique, le modèle que nous proposons pour résoudre le problème de l'estimation globale peut être défini comme suit : le modèle générique est constitué des classes de F.A. non stationnaires admettant un même variogramme moyen $\bar{\gamma}$ dans S , et les critères de spécification du modèle sont fournis par $\bar{\gamma}$ lui-même, ou même simplement par les paramètres qui définissent son comportement au voisinage de l'origine. En définitive donc, nous avons identifié et rassemblé dans une même rubrique les paramètres qui servent à spécifier le modèle, ceux qui permettent de résoudre le problème posé et ceux qui possèdent le plus riche contenu physique.

Une remarque pour terminer. A la variance d'estimation(4) des représentations transitives se trouve maintenant associée, dans le modèle, la variable aléatoire obtenue en substituant $G(h)$ à $g(h)$. L'espérance $\sigma^2 = E(\text{Var } Q^*)$ de cette V.A. constitue ce que nous pourrions appeler la "variance d'estimation théorique". Comme $E(G)$ est égal, par définition, au modèle \bar{g} de covariogramme, cette variance théorique se calcule à partir de \bar{g} exactement de la même manière que $\text{Var } Q^*$ à partir de g . Autrement dit, cette variance théorique s'identifie au terme régulier, lié au comportement du covariogramme en $h=0$, tandis que la fluctuation $\text{Var } Q^* - \sigma^2$, variable aléatoire d'espérance nulle dans le présent modèle, coïncide avec le terme fluctuant lié au comportement de $g(h)$ au voisinage de la portée : l'aspect pré-aléatoire, et imprévisible in praxi, de ce terme fluctuant se trouve ainsi, lui aussi, pris en charge dans la nouvelle formulation.

Modèles globaux non stationnaires

Comme nous l'avons vu, l'absence de stationnarité (physique) du phénomène ne suscite pas de difficultés majeures tant que l'on se limite à un problème d'estimation globale. Mais on poursuit souvent d'autres objectifs. Il peut arriver que notre but soit, justement, d'étudier la non-stationnarité elle-même et d'en donner une représentation globale précise, visualisée sous la forme de cartes. Les géophysiciens, par exemple, chercheront à séparer "l'anomalie régionale" de "l'anomalie locale", la

la première reflétant les structures profondes du sous-sol, la seconde liée à des accidents superficiels, d'intérêt purement local : cette distinction est peut être difficile à préciser, mais présente certainement une signification objective. L'intuition physique qui sert ici de guide est celle d'un phénomène "auto-régulé" : en l'absence de perturbations, l'évolution du phénomène dans l'espace se laisserait décrire par une fonction $m(x)$ suffisamment régulière, à notre échelle de travail, pour que nous puissions la considérer comme "déterministe", c'est-à-dire en donner une représentation fonctionnelle suffisamment simple et précise. Du fait des irrégularités et accidents superficiels, la V.R. $z(x)$ réelle s'écarte pourtant çà et là de cette position d'équilibre $m(x)$, mais ces écarts ne sont jamais ni très amples, ni très durables : tout se passe comme si une force de rappel contraignait bientôt $z(x)$ à revenir au voisinage de $m(x)$, et la figure qui en résulte est celle d'une sorte d'oscillation autour de la position d'équilibre représentée par $m(x)$. Ceci suggère le modèle simple :

$$(8) \quad Z(x) = m(x) + Y(x)$$

où $Y(x)$ est une fonction aléatoire stationnaire, d'espérance nulle et de portée finie (on peut, évidemment, envisager aussi des modèles plus complexes). Nous dirons que $m(x)$ est la dérive et $Y(x)$ le résidu, ou la fluctuation. Dans ce modèle, la dérive $m(x)$ est l'espérance (non constante) de la F.A. $Z(x)$ associée à la V.R. $z(x)$. A la fluctuation, ou F.A. $Y(x)$ du modèle, correspond de même une nouvelle V.R., soit $y(x) = z(x) - m(x)$. Il est entendu que $y(x)$ représente, elle aussi, une réalité physique, et nullement quelque "erreur de la nature" (dans notre exemple : l'influence des structures superficielles), et la F.A. $Y(x)$ du modèle n'est pas un simple "bruit" elle possède, elle aussi, des caractéristiques structurelles (une covariance, une portée, etc...), mais à une échelle plus modeste que la dérive $m(x)$.

Ce premier exemple n'est pas trop ambigu, parce qu'un modèle physique préalable nous garantit que la dichotomie dérive + résidu correspond à une réalité (structures profondes et accidents de surface). Mais il n'en est pas toujours ainsi. En cartographie sous-marine, on s'attend bien à ce que la profondeur augmente lorsqu'on s'éloigne des côtes. Il s'agit, manifestement, d'un phénomène non stationnaire. Mais nous sommes déjà bien en peine de définir avec quelque précision la notion de dérive qu'il convient d'utiliser ici. Nous pouvons, conventionnellement, choisir d'appeler dérive le résultat fourni par tel ou tel procédé de lissage (moyenne mobile, filtrage de hautes fréquences etc...). Mais il s'agit là d'une définition purement instrumentale, et non pas d'un concept. La difficulté provient sans doute de ce que la représentation

intuitive que nous avons en tête, lorsque nous parlons de dérive, est essentiellement liée à une échelle de travail. Pour prendre un autre exemple, considérons une montagne : Si nous travaillons à l'échelle de la dizaine ou de la centaine de mètres, la montagne se présente comme une dérive fonctionnelle, et les accidents locaux du relief (petits ravins, rochers isolés etc..) font figure de résidus. Mais si, travaillant à l'échelle de la dizaine de kilomètres, nous étudions maintenant l'ensemble de la chaîne à laquelle appartient notre montagne, celle-ci ne nous apparaît plus que comme une fluctuation locale que rien de bien particulier ne permet de distinguer des autres montagnes qui l'entourent. Le modèle a changé. Ce qui tout à l'heure nous semblait une dérive est pris en charge, à l'échelle de la chaîne, par une fonction aléatoire éventuellement stationnaire.

De fait, entre les deux termes de la dichotomie (8), un certain balancement est toujours possible : nous pouvons, dans une certaine mesure, choisir d'incorporer telle ou telle caractéristique structurale du phénomène, à volonté, soit dans la dérive, soit dans le résidu : d'où plusieurs modèles possibles à peu près équivalents, les uns présentant une dérive complexe et un résidu pauvre, de portée très courte, d'autres au contraire une dérive très simple (linéaire ou quadratique, par exemple) et, en contrepartie un résidu de structure plus riche, une fonction de covariance plus compliquée, une portée plus grande etc... Il n'y a pas ici, de critère univoque d'objectivité, et le point de vue monoscopique s'impose. C'est à l'usage que nous apprenons à choisir, dans chaque cas particulier, le type de modèle qui nous permet de résoudre au mieux le problème que nous nous posons.

Ainsi, la notion de dérive, ou tendance, correspond à une fausse évidence, et se révèle à l'analyse singulièrement équivoque. On oscille entre deux pôles. Le premier correspond à une simple attraction visuelle : On se représente, instinctivement, la dérive comme la courbe ou la surface "régulière" qui passe aussi près que possible des points expérimentaux. Par surface "régulière", on entend le plus souvent, soit un polynôme de degré donné (c'est à ce point de vue que répond la technique d'ajustement par moindres carrés), soit une surface soumise à des contraintes de courbure (par exemple) qui lui imposent un certain type de régularité (c'est le point de vue de la technique des fonctions spline). Ces techniques sont parfaitement défendables, et rendent de grands services dans certains types de problèmes. Il n'y a rien à leur objecter pourvu :

i) que l'on reconnaisse clairement leur caractère conventionnel et purement instrumental, et que l'on n'aille prendre les résultats numériques auxquels elles conduisent ni pour la "réalité physique", ni pour la réalisation d'un concept, et

ii) que l'on ne confonde pas les résultats obtenus in praxi en appliquant ces techniques aux seules données expérimentales avec ce que l'on obtiendrait (après coup) en les appliquant à la totalité des valeurs numériques $z(x)$ disponibles dans S .

Ce dernier point est important. Ce que l'on obtient en appliquant, par exemple, la technique des moindres carrés aux données expérimentales $z(x_\alpha)$ ne constitue en aucune façon le meilleur estimateur du résultat que donnerait la même technique appliquée à la V.R. z connue en tous les points x de S . Pour obtenir cette meilleure estimation, il convient d'estimer ces moindres carrés "futurs" sur la base de l'information disponible. Par exemple, on effectuera un krigeage global, si l'on dispose d'un modèle global, ou plus simplement un recollement de krigeages locaux si l'on ne dispose que d'un modèle local. Il est facile de voir que cela revient exactement à appliquer la technique des moindres carrés à la surface obtenue en krigeant les valeurs inconnues de $z(x)$. Autrement dit, on doit ajuster par moindres carrés, non pas les valeurs expérimentales $z(x_\alpha)$, mais les valeurs krigées $z^*(x)$, $x \in S$.

L'autre point de vue (qui risque fort de se révéler purement métaphysique, à moins que l'on ne dispose d'un modèle physique préalable) correspond à la croyance spontanée en l'existence d'une sorte de tendance régulière, ou fond continu ("trend") par rapport auquel les déviations que présente le phénomène réel constitueraient des "erreurs de la nature" ou, au mieux, des "anomalies" à signification purement locale. En fait, on ne voit pas bien sur quelle base on pourrait accuser la nature de commettre des erreurs (je ne parle pas ici du problème bien différent, et tout-à-fait réaliste, qui se pose lorsque les données expérimentales sont entachées d'erreurs de mesure). Le concept d'anomalie est certainement plus riche et plus intéressant, mais très difficile à élucider.

De toute façon, avant de choisir un estimateur, il faut définir de manière précise et opérateur ce que l'on veut estimer. Concrètement, cette définition sera constituée par un algorithme applicable, après coup, à la V.R. z . C'est toujours une grandeur régionale que l'on cherche à estimer, c'est-à-dire à approcher au mieux à l'aide d'une fonction des données disponibles. Si donc des géophysiciens, par exemple

nous demandent, au vu des résultats discrets d'une campagne de mesures, de séparer "l'anomalie locale" et "l'anomalie régionale", nous leur poserons la question cruciale habituelle : si vous connaissiez votre V.R. $z(x)$ en tous les points x du champ qui vous intéresse -donc tout ce qu'il est en principe expérimentalement possible de connaître au sujet de ce phénomène - comment procéderiez-vous pour séparer ces deux composantes locales et régionales ? Cette distinction résulte-t-elle d'un modèle précis, qui vous est suggéré par les connaissances que vous pouvez avoir sur la physique de ce phénomène, et vous conduirait-elle, en cas de connaissance expérimentale exhaustive, à une détermination rigoureuse de chacune de ces deux composantes, ou au moins à une procédure d'estimation bien définie ? ou bien vous contenteriez-vous d'appliquer mécaniquement telle ou telle procédure de filtrage ou d'ajustement par moindres carrés ? Dans tous les cas, nous exigeons de vous que vous nous précisiez, sous votre responsabilité, la procédure numérique - l'algorithme - que vous utiliseriez si vous connaissiez parfaitement votre phénomène.

Le résultat (actuellement inconnu puisqu'en fait z n'est donné qu'en quelques points) auquel conduirait l'algorithme choisi par vous constitue la grandeur régionale qu'à notre tour, et pour votre service, nous pouvons vous proposer d'estimer de la meilleure manière possible, compte-tenu de l'information dont nous disposons et du modèle que nous avons réussi à spécifier sur cette base. Mais, pour notre part, nous ne disposons pas de règle universelle permettant, d'entrée de jeu, de séparer "dérive en soi" et "résidu en soi", "lisse" et "rugueux" et ainsi de suite.

Dans certains cas, la notion de dérive se prête à une reconstruction opératoire, analogue à celle qui nous a permis de définir la portée, et présente alors une signification objective sans équivoque. Mais, plus souvent peut-être la dichotomie (8) apparaît irréductiblement arbitraire. Or, nous savons qu'une solution sensée d'un problème réel ne peut pas dépendre d'un choix purement conventionnel. Autrement dit, si la notion de dérive n'est pas susceptible d'une définition objective, nous n'en aurons non plus jamais besoin pour résoudre un problème réel. Cette remarque conduit à rechercher des modèles plus synthétiques que (8), dans lesquels on renonce à séparer "dérive" et "résidu", mais qui conservent cependant une forme atténuée de stationnarité : stationnarité assez faible pour être compatible avec la réalité, mais suffisante pour permettre "l'inférence statistique" classique. Un bon exemple de modèle de ce genre est fourni par les "fonctions aléatoires intrinsèques d'ordre k "¹. Il s'agit d'une F.A. $Z(x)$ non stationnaire, mais dont les "accroissements généralisés", eux, sont stationnaires. Parmi les combinaisons linéaires de la forme $\sum \lambda_i Z(x_i)$, seules sont station-

1 - Voir mon article "intrinsic random functions", Adv. in App. Prob., Déc. 1973, et l'exposé de P. Delfiner "linear estimations of non stationary spatial phenomena", in Advanced Geostatistics in the Mining Industry, ed. by M. Guarascio et al, 1976, D. Reidel,

naires les "combinaisons linéaires autorisées", c'est-à-dire filtrant (annulant) les polynômes de degré inférieur ou égal à k . La fonction de covariance de Z est de la forme :

$$(9) \quad C(x, y) = K(x-y) + \sum a_1(x) f^1(y) + \sum a_1(y) f^1(x)$$

Dans cette relation, les f^1 représentent les monômes de degré inférieur ou égal à k , et les $a_1(x)$ des fonctions quelconques (inconnues), inaccessibles in praxi, et même peut être après coup. La partie non stationnaire de la covariance $C(x,y)$, c'est-à-dire les deux sommes du type $\sum a_1 f^1$, a donc une signification purement conventionnelle. Au contraire, la partie stationnaire $K(x-y)$, ou "covariance généralisée" est accessible (après coup) et peut faire (in praxi) l'objet d'une estimation : car les variances des combinaisons linéaires autorisées ne dépendent que de cette fonction K , et non de la partie non stationnaire de la covariance. C'est donc la covariance généralisée K qui constitue le critère de spécification du modèle générique "F.A. intrinsèque d'ordre k ". Corollaire obligé, si nous choisissons ce modèle nous devons nous astreindre à ne manipuler que des combinaisons linéaires autorisées : limitation réelle, mais qui ne nous empêche nullement d'apporter des solutions sensées aux problèmes bien posés.

La notion de dérive ne figure pas parmi les critères de spécification et les outils de travail autorisés par ce modèle. Elle est cependant prise en charge, mais de manière implicite seulement. En effet, à une même covariance généralisée K (modèle spécifique) correspond toute une classe de F.A., différant les unes des autres par des polynômes à coefficients aléatoires. Si $Y(x)$ est l'une de ces F.A., les autres sont de la forme :

$$Z(x) = Y(x) + \sum A_1 f^1(x)$$

où les f^1 sont les monômes, et les A_1 des variables aléatoires, non indépendantes de $Y(x)$ en général. Toute combinaison linéaire autorisée, filtrant, par définition, les polynômes de degré k , définit donc une seule et même variable $\sum \lambda_i Z(x_i) = \sum \lambda_i Y(x_i)$ indépendante du choix de Y ou de Z . On peut, si l'on veut, appeler dérive le polynôme

aléatoire $\sum A_1 f^1(x)$ qui est ainsi, automatiquement, filtré par les combinaisons linéaires autorisées. Mais cette "dérive" n'est pas définie, ni même définissable, car il n'existe pas, en général, dans la classe des F.A. associées à une même covariance généralisée K , de fonction aléatoire privilégiée (stationnaire par exemple) pouvant servir de référence fixe.

Dans le cas particulier $k=0$, on obtient des classes de F.A. à accroissements stationnaires, définies à une constante (aléatoire) près, et, au signe près, la covariance généralisée s'identifie au variogramme, soit $K-\gamma$. Nous avons vu plus haut que le comportement du variogramme au voisinage de l'origine présentait une signification objective indéniable, riche de contenu physique, et se prêtant à une reconstruction opératoire. Nous pouvons espérer qu'il sera de même pour la covariance généralisée $K(h)$ lorsque k est plus grand que 0. Pour des raisons de commodités, c'est dans le chapitre suivant, consacré aux modèles locaux, que nous examinerons ce point.

Les modèles locaux

Dans ce chapitre consacré aux modèles locaux - c'est-à-dire aux modèles dans lesquels on a choisi de ne jamais manipuler simultanément que des données provenant de points relativement peu distants les uns des autres - notre tâche sera à peu près la même que dans le chapitre précédent : fonder l'objectivité, esquisser la reconstruction opératoire. Cette tâche sera, en un sens, moins ardue, puisque la répétition dans l'espace qui, faute de répétition dans le temps, nous fournit notre critère d'objectivité interne, s'introduit de manière particulièrement naturelle lorsque nous adoptons le point de vue des modèles locaux. La démarche, ici encore, consistera à partir d'un modèle strictement objectif, ou représentation, qui servira ensuite de critère de comparaison pour juger l'objectivité et la valeur des modèles usuels. Et, à nouveau, le choix de ces derniers, tel qu'on peut l'effectuer in praxi, nous apparaîtra comme tributaire d'une hypothèse d'approximation anticipée.

Les représentations glissantes

Pour introduire de manière naturelle notre second groupe de modèles strictement objectifs, que nous appellerons représentations glissantes, laissons-nous guider par un problème simple, qui se présente d'ailleurs fréquemment en pratique, celui de l'estimation linéaire locale. La situation typique est la suivante : nous avons affaire à une V.R. z définie dans un champ S . Comme d'habitude, nous prolongeons z à l'espace entier en posant $z(x) = 0$ pour tout x n'appartenant pas à S . Nous voulons, sur la base des points expérimentaux x_α , $\alpha = 1, 2, \dots, N$, estimer la valeur $z(x)$ en chaque point x d'un domaine S_0 choisi par nous. Plus généralement, on pourrait aussi envisager l'estimation d'une moyenne mobile de type $z_p(x) = \int z(x+y) p(dy)$ où p serait une mesure dont le support serait contenu dans le voisinage mobile V défini ci-après : mais, dans cet exemple introductif, nous pouvons nous limiter au cas le plus simple, celui de l'estimation des valeurs ponctuelles $z(x)$.

Pour procéder à cette estimation, nous avons (à tort où à raison, nous reviendrons sur ce point) pris un certain nombre de décisions méthodologiques :

i) nous avons choisi de n'utiliser que des estimateurs linéaires du type $z^*(x) = \sum \lambda^\alpha z_\alpha$, et, plus précisément, comme dans le cas du krigeage, nous avons décidé, pour estimer un point x donné, de ne retenir que les points expérimentaux les plus proches du point à estimer. Cette décision a essentiellement pour but de renforcer la robustesse de notre procédure. D'une manière précise, nous avons fait choix d'un voisinage V de l'origine, appelé voisinage glissant, et, pour estimer

$z(x)$, nous ne retenons que les points x_{α} qui tombent dans le voisinage V_x (translaté de V par la translation x) du point x . Autrement dit, en désignant par x_i les points expérimentaux retenus, et en posant $h_i = x_i - x \in V$, nous nous limitons à la classe des estimateurs linéaires du type :

$$(1) \quad z^{\mathbf{x}}(x) = \sum_i \lambda^i z(x+h_i)$$

ii) De plus, (et du point de vue méthodologique, c'est là la décision la plus importante) nous décidons de n'utiliser, pour former nos estimateurs $z^{\mathbf{x}}(x)$, que des algorithmes invariants par translation. Ce qui veut dire ici que les poids λ^i qui figurent dans l'expression (1) pourront dépendre des vecteurs h_1, h_2, \dots du voisinage mobile V , mais non du point x à estimer. Autrement dit, ces poids seront invariants par translation : si nous appelons configuration la figure constituée par le point x à estimer et les points $x_i = x+h_i \in V_x$ sélectionnés pour former l'estimateur (1), les poids λ^i ne seront pas modifiés par une translation affectant l'ensemble de la configuration.

Il s'agit, répétons-le, d'une décision de notre part (judicieuse ou non) et nullement d'une hypothèse relative à la réalité physique. C'est là un deuxième exemple de la démarche qui consiste à faire de la stationnarité une caractéristique (choisie par nous) de la classe des estimateurs que nous utilisons, plutôt qu'une hypothèse (vulnérable) relative à la réalité.

Pour apprécier (après coup) la valeur de l'estimateur (1) où les poids λ^i et les vecteurs h_i sont supposés fixés, nous choisissons conventionnellement (mais c'est une convention assez "naturelle") le critère : valeur quadratique moyenne de l'erreur $z^{\mathbf{x}}(x) - z(x)$ lorsque le point x parcourt le domaine S_0 auquel nous avons choisi de limiter notre intérêt. Posons donc, par définition :

$$||z^{\mathbf{x}}-z||^2 = \frac{1}{S_0} \int_{S_0} [z^{\mathbf{x}}(x) - z(x)]^2 dx$$

En remplaçant $z^{\mathbf{x}}(x)$ par son expression (1), on voit apparaître la fonction $C(h, h')$ définie par la formule

$$(2) \quad C(h, h') = \frac{1}{S_0} \int_{S_0} z(x+h) z(x+h') dx$$

et notre erreur quadratique moyenne peut s'écrire :

$$||z^* - z||^2 = c(0,0) - 2 \sum_i \lambda^i c(0, h_i) + \sum_{i,j} \lambda^i \lambda^j c(h_i, h_j)$$

Si la fonction $C(h, h')$ était connue (in praxi, évidemment, il faudra l'estimer) la suite des opérations (choix des poids λ^i optimaux etc...) se déroulerait exactement comme dans le cas d'un krigeage. Cette fonction $C(h, h')$, d'après sa définition (2), a du reste bien la signification d'une covariance. Ceci suggère la définition tout-à-fait générale suivante :

Etant donné un voisinage glissant V , un domaine S_0 et une V.R. z , nous dirons que la F.A. définie par :

$$(3) \quad Z(h) = z(\underline{x} + h) \quad (h \in V)$$

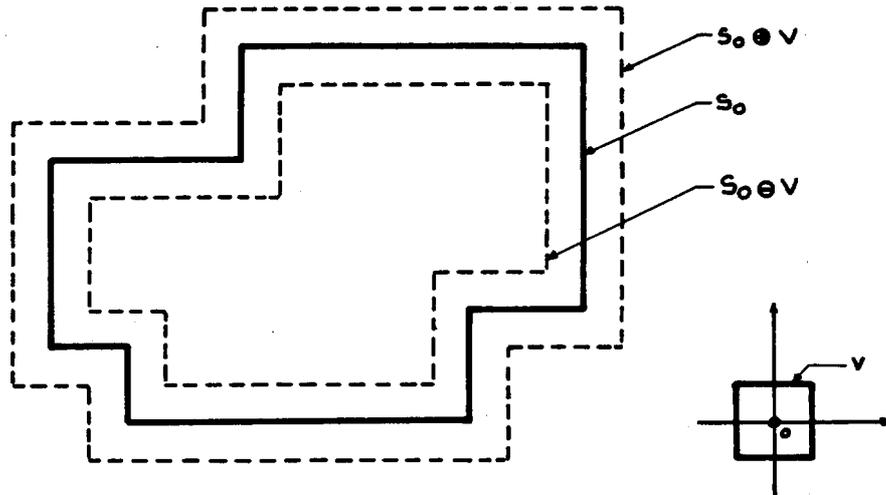
où \underline{x} est le point aléatoire obtenu en munissant S_0 de la loi de probabilité uniforme constitue la représentation glissante de la V.R. z dans S_0 (pour le voisinage glissant V).

Avec cette définition, la relation (2) peut se réécrire sous la forme $C(h, h') = E[Z(h) Z(h')]$, de sorte que $C(h, h')$ est bien la fonction covariance associée à la F.A. $Z(h)$ que nous venons de définir.

Le voisinage glissant V et le domaine S_0 , choisis par nous, sont des paramètres méthodologiques. D'après la définition ci-dessus, toutes les caractéristiques de la F.A. $Z(h)$, $h \in V$, s'exprimeront en fonction de S_0 , V et de grandeurs régionales, à l'exclusion de tout autre paramètre. Il s'agit donc bien d'un modèle strictement objectif, ou représentation, d'ailleurs parfaitement tautologique, puisque nous ne faisons rien d'autre pour l'instant que de présenter la V.R. z sous une forme à peine différente. Notons pourtant les points suivants, concernant les relations mutuelles de V , S_0 et du champ réel S à l'extérieur duquel z est identiquement nul.

Les points y de l'espace seuls impliqués dans la définition (3) sont les points de la forme $y = x+h$, x décrivant le domaine S_0 , et h le voisinage V choisis par nous. Ils forment un ensemble, que nous pouvons désigner par $S_0 \oplus V$ et appeler "dilaté¹ de S_0 par V ". Deux cas présentent un intérêt particulier :

1 - Je suppose ici que V est un voisinage symétrique de l'origine O des coordonnées, de manière à ne pas avoir à distinguer entre V et son symétrique par rapport à l'origine.



- Fig. 10 -

Si l'ensemble dilaté $S_0 \oplus V$ est entièrement contenu dans le champ réel S de la variable régionalisée (champ S en dehors duquel z est identiquement nulle), la définition (3) de la F.A. $Z(h)$ n'utilise pas réellement le prolongement (fictif) de la V.R. z à l'extérieur de son champ naturel. Nous dirons qu'il s'agit d'une représentation glissante interne, modèle particulièrement bien adopté à l'étude de la variabilité "interne" de z , c'est-à-dire à l'intérieur de son champ S et abstraction faite de la géométrie de celui-ci.

La F.A. $Z(h)$ s'annule, identiquement, sur V chaque fois que le point aléatoire x tombe à l'extérieur du dilaté $S \oplus V$ du champ S par le voisinage V . Donc, si le domaine S_0 choisi par nous contient $S \oplus V$, ou, ce qui revient au même, si le champ S est contenu dans l'érodé $S_0 \ominus V$ (voir figure 10), toutes les configurations non triviales que l'on peut observer sur les voisinages glissants V_x interviennent effectivement dans la définition de la F.A. $Z(h)$. Nous disons, dans ce cas, qu'il s'agit d'une représentation externe ou exhaustive de la variable régionalisée.

On peut démontrer le théorème suivant: toute représentation glissante externe est une fonction aléatoire stationnaire dans V . En particulier, si l'on désigne par $g(h)$ le covariogramme transitif de la V.R. z , la fonction de covariance de la F.A. stationnaire sur V qui lui est associée dans une représentation externe, est :

$$C(h, h') = \frac{1}{S_0} g(h - h')$$

Cependant, et malgré ce joli théorème, les représentations externes ne sont qu'assez rarement intéressantes dans les applications. Elles correspondent, en effet, à une situation où l'on aurait décidé a priori d'utiliser le même algorithme invariant par translation aussi bien pour les configurations internes (V_x inclus dans le champ de la V.R.) que pour celles qui chevauchent la frontière : attitude en général peu réaliste car, dans la plupart des cas, il y a un intérêt évident à traiter à part, de manière différente, les configurations de bordure (qui contiennent des données nulles, extérieures au champ réel). Souvent, d'ailleurs, le domaine S qui nous intéresse est une partie, sélectionnée par nous, d'un champ total plus vaste, et dans ce cas la représentation glissante est nécessairement interne. Ces représentations glissantes externes ne sont en réalité pas autre chose qu'une version localisée des représentations transitives. Elles ne peuvent donc, comme ces dernières, présenter d'intérêt que lorsque la géométrie du champ de la V.R. présente une grande importance pour le problème à résoudre, par exemple si ce champ S est constitué d'un assez grand nombre de composantes connexes disjointes ou présente de nombreuses enclaves stériles.

Priorité de la méthode

Le point de vue des représentations glissantes revient, en somme, à essayer de faire de la stationnarité une caractéristique (limitative) choisie par nous de la classe d'estimateurs à utiliser plutôt que de la considérer comme une propriété de la réalité physique (la V.R.) ou même comme une caractéristique d'un modèle de F.A. (non strictement objectif) choisi a priori. L'opération réussit parfaitement dans le cas des représentations externes (voir le théorème ci-dessus), partiellement seulement dans les autres cas. Nous y reviendrons. Notre décision méthodologique : n'utiliser que des estimateurs à caractère local, associés à des algorithmes invariants par translation, ne présuppose en aucune façon la stationnarité ou l'homogénéité du phénomène dans l'espace. Elle paraîtra sans doute plus judicieuse si le phénomène est effectivement plus ou moins homogène dans l'espace, mais ne sera pas pour autant absurde en cas d'hétérogénéité.

Certes, en cas d'hétérogénéité, s'il est raisonnablement possible (à partir des données disponibles in praxi) de diviser le champ total en sous-zones où le phénomène reste à peu près homogène (et de spécifier un modèle différent dans chacune de ces sous-zones !) on aura intérêt, en général, à utiliser des algorithmes différents, ajustés spécialement pour chacune des sous-zones : l'erreur sera moindre, en moyenne, et de plus les variances d'estimation seront localisées (c'est-à-dire différeront d'une sous-zone à l'autre, conformément à ce que suggère l'intuition physique).

Il n'est, malheureusement, pas toujours possible de définir, dans des conditions acceptables, cette partition du champ, ni de spécifier des modèles différents adaptés à chacune des zones : ceci, tout simplement, parce que les données disponibles se raréfient, au fur et à mesure que ces sous-zones se multiplient, de sorte que les erreurs commises dans l'estimation des paramètres objectifs deviennent de plus en plus grandes, et que l'on risque fort de dépasser le seuil de robustesse - et même le seuil de simple réalisme.

A la méthode puissante, utilisant des algorithmes différents dans chacune des sous-zones, mais nécessitant des prérequisites trop nombreux, pratiquement inaccessibles in praxi (autant de modèles spécifiques que de sous-zones) nous devons souvent conformément à nos règles méthodologiques générales, substituer une méthode moins puissante (un seul algorithme invariant par translation, du moins tant que x appartient à S), mais nécessitant aussi moins de prérequisites. Nous aurons, certes, ainsi une variance d'estimation plus forte et non localisée : la valeur numérique que nous lui attribuerons représentera seulement la moyenne des valeurs différentes qu'elle prend dans chacune des sous-zones que nous avons renoncé à distinguer. En contrepartie, nous retombons en dessous du seuil de robustesse, et la spécification du modèle redevient accessible in praxi.

Notons pourtant un point important : aucune hypothèse concernant l'homogénéité physique du phénomène n'est requise ici. Par contre, nos représentations glissantes sont réalistes que si la reconnaissance (le réseau des points expérimentaux), est suffisamment homogène (dans le domaine $S_0 \oplus V$). En effet, pour une configuration donnée centrée en x , nous calculons la variance d'estimation en prenant la moyenne quadratique de l'erreur lorsque x décrit S_0 : ce qui n'est sensé que si l'on peut s'attendre à rencontrer des configurations identiques, ou du moins suffisamment analogues à celle-ci, pour des implantations de x réparties de manière à peu près uniforme dans S_0 . Dans le cas contraire, en effet, nous risquons d'utiliser des caractéristiques globales de la V.R., bien différentes peut-être des caractéristiques locales qui auraient convenu au traitement de cette configuration trop particulière. En cas d'hétérogénéité, la solution consistera, en principe, à découper le champ en sous-zones reconnues de façon homogène (ce qui est plus facile, ou moins risqué, que le découpage en sous-zones où le phénomène lui-même est homogène) et à traiter séparément chacune de ces sous-zones.

Au stade où nous sommes parvenus, il convient d'introduire une notion générale, celle de régionale glissante associée à une représentation glissante, notion dont l'estimateur (1) constitue un cas particulier. La V.R. z ayant été prolongée (par des valeurs identiquement nulles à l'extérieur de S), considérons une fonctionnelle quelconque $\mathcal{A}(z ; V)$ définie par la donnée de la restriction de z au voisinage V de l'origine O . Pour chaque x , posons $\mathcal{A}(z ; V_x) = \mathcal{A}(z_{-x} ; V)$, où z_{-x} est la V.R. déduite de z par la translation $-x$, soit $z_{-x}(y) = z(x+y)$: nous dirons que la fonction $x \rightarrow \mathcal{A}(z ; V_x)$ ainsi définie constitue la régionale glissante associée à l'algorithme $\mathcal{A}(z ; V)$ invariant par translation. La restriction de cette fonction à S_0 (muni de la loi de probabilité uniforme) est donc une variable aléatoire associée à la F.A. $Z(h)$ définie par la représentation glissante : à savoir la V.A. $\mathcal{A}(Z ; V)$ obtenue en substituant la F.A. Z à la V.R. z dans l'algorithme de définition $\mathcal{A}(z ; V)$. Le formalisme des représentations glissantes permet donc, en principe, d'étudier n'importe quelle régionale glissante, linéaire ou non, et donc, en particulier, d'aborder les problèmes d'estimation (locale) non linéaires, au moyen d'espérances conditionnelles, ou d'estimateurs de type disjonctif etc... : sous la réserve (essentielle) que la spécification in praxi soit raisonnablement accessible.

En toute généralité, le problème de l'estimation locale est celui de l'estimation de la valeur en tout $x \in S_0$ d'une régionale glissante $\mathcal{A}(z ; V_x)$. Si les points expérimentaux $x_i = x+h_i$ disponibles dans le voisinage glissant V_x du point x ont fourni les valeurs $z(x_i) = z_i$, nous cherchons à former une fonction dépendant des vecteurs h_i et des valeurs z_i , c'est-à-dire un algorithme $\mathcal{A}^*(z_i ; h_i)$ lui-même invariant par translation et pouvant jouer le rôle d'estimateur. Pour choisir un tel algorithme, nous pouvons adopter différents critères. Le plus simple d'entre eux consiste à minimiser (dans le modèle) la variance d'estimation, ou même simplement l'erreur quadratique moyenne :

$$E [(\mathcal{A}(z, V) - \mathcal{A}^*(z_i, h_i))^2]$$

lorsque la fonction \mathcal{A}^* décrit une famille Φ choisi par nous. In praxi, évidemment, nous choisirons une famille d'autant plus large que nous pourrions mieux spécifier notre modèle. Par exemple, Φ pourra être la classe des estimateurs linéaires, vérifiant telle ou telle condition "d'universalité", ou bien celle des estimateurs "disjonctifs" (voir ch. 8) etc...

La régionale glissante que nous devons estimer peut être à peu près quelconque (sous réserve que sa définition ne fasse intervenir que des valeurs observables sur

sur un même voisinage V). Par exemple, il peut s'agir d'une moyenne mobile

$$z_p(x) = \int z(x+y) p(dy)$$

de la V.R. pondérée par une mesure p dont le support est contenu dans V : le cas le plus fréquent en pratique est celui de la teneur moyenne d'un "panneau glissant" $v_x \subset V_x$. Mais Q peut aussi représenter une fonctionnelle non linéaire en z . Par exemple, si l'on ne s'intéresse pas seulement à la teneur moyenne des panneaux glissants, mais également à leurs lois de distributions, on cherchera pour chaque $x \in S$ et chaque nombre réel u à estimer la quantité

$$\theta_p(x; u) = \begin{cases} 1 & \text{si } z_p(x) \geq u \\ 0 & \text{si } z_p(x) < u \end{cases}$$

Le problème à résoudre (la régionale $Q(z, V)$ à estimer) est une donnée imposée, sur laquelle nous ne pouvons pas agir. En ce qui concerne le choix de la méthode (c'est-à-dire le choix de la classe Φ des estimateurs à utiliser) nous disposons d'une assez grande latitude. Plus grande sera cette classe Φ , plus puissant sera notre estimateur, mais plus nombreux aussi seront les prérequisites relatifs à la loi spatiale de la représentation glissante $Z(h)$ dont la connaissance sera impérativement requise pour déterminer effectivement l'estimateur optimal. Si la représentation glissante était complètement spécifiée (ce qui serait le cas après coup), on devrait prendre pour Φ la classe la plus large possible, c'est-à-dire celle de toutes les fonctions mesurables. L'estimateur optimal serait en conséquence, l'espérance conditionnelle. Mais il n'est pour ainsi dire jamais possible de spécifier suffisamment in praxi, la représentation pour que l'espérance conditionnelle soit accessible dans des conditions de robustesse raisonnables. La tactique à observer va donc consister à restreindre la classe Φ des estimateurs à utiliser, suffisamment pour que la spécification des prérequisites nécessaires soit raisonnablement possible, mais pas trop cependant, pour ne pas diminuer plus qu'il n'est nécessaire la puissance de notre estimateur.

Nous retrouvons donc ici, très exactement, le problème du choix et de la spécification d'un modèle de F.A., tel qu'il a été exposé dans le chapitre 4, mais avec une différence capitale : s'agissant maintenant de représentations, c'est-à-dire de modèles strictement objectifs, les paramètres de ce modèle (sauf S et V qui sont choisis une fois pour toutes) ont tous une signification purement objective.

Bien mieux, l'appartenance de $Z(h)$ à tel ou tel modèle générique est elle-même décidable après coup, de sorte qu'un énoncé comme "la représentation $Z(h)$ est une F.A. stationnaire d'ordre 2" constitue maintenant une hypothèse objective, vraie ou fausse, mais décidable après coup. Après coup, en effet, nous pouvons calculer les expressions :

$$\left\{ \begin{array}{l} m(h) = E [Z(h)] = \frac{1}{S_0} \int_{S_0} z(x+h) dx \\ c(h,h') = E [Z(h) Z(h')] = \frac{1}{S_0} \int_{S_0} z(x+h) z(x+h') dx \end{array} \right.$$

et constater si elles sont, ou non invariantes par translation. En pratique, elles ne seront jamais (rigoureusement) stationnaires. Mais nous n'en demandons pas tant, et nous sommes tout disposés à nous contenter d'une stationnarité approchée. Si nous trouvons, avec une approximation acceptable (pour h, h' appartenant à V) :

$$(4) \quad \left\{ \begin{array}{l} m(h) \neq m \\ c(h,h') \neq \bar{c}(h-h') \end{array} \right.$$

nous remplacerons, sans faire de difficultés, la vraie représentation $Z(h)$ par le modèle stationnaire d'ordre 2 défini par m et $\bar{c}(h-h')$. Quant à ce qu'il faut entendre par une approximation "acceptable", cela dépend évidemment du problème à résoudre et des données disponibles : à partir de ces données, le modèle exact et le modèle approché doivent conduire à des estimateurs peu différents, et leur attribuer des variances d'estimation peu différentes.

Naturellement, c'est après coup seulement que nous pourrons constater si cette condition est remplie et si le modèle générique "F.A. stationnaire d'ordre 2" constitue une approximation acceptable de la représentation $Z(h)$. Adopter ce modèle in praxi constitue toujours une hypothèse anticipatrice (à la fois féconde et vulnérable), comme dans le chapitre 4, mais dont la signification est maintenant purement objective (décidable après coup), et dont le statut épistémologique est parfaitement clair : le choix in praxi d'un modèle générique pour la représentation glissante constitue simplement une hypothèse d'approximation anticipée.

Nous allons maintenant passer en revue quelques-uns des modèles génériques que l'on utilise le plus souvent en pratique à titre d'approximation des représentations glissantes. Nous dirons que ce sont des modèles locaux. L'adjectif "local" ne doit pas faire illusion. Le modèle dit "local", en effet, ne prétend en aucune façon distinguer les unes des autres les caractéristiques que possède la V.R. dans les différentes parties de son champ, mais présente seulement une image moyenne de ces diverses propriétés localisées. Il est dit "local" parce qu'il permet d'effectuer des estimations locales, et n'est d'ailleurs défini que sur le voisinage V de l'origine. Mais les paramètres qui le définissent ont tous une signification globale (ils sont défini par des intégrales dans S).

D'où cette conséquence (qui n'est paradoxale qu'en apparence) importante pour la méthodologie : la spécification in praxi d'un modèle local pose toujours un problème d'estimation globale. Pour aborder ce problème, il y a en général intérêt à distinguer deux étapes :

i) en premier lieu, se placer (par la pensée) dans la situation "après coup" et se demander quel algorithme constituerait alors la meilleure définition d'un paramètre donné.

ii) se demander ensuite comment l'algorithme en question (inaccessible in praxi) peut être estimé sur la base de l'information actuellement disponible.

F.A. localement stationnaire d'ordre 2

C'est, par définition, le modèle constitué par les hypothèses d'approximation (4) : $m(h)$ à peu près constante et $C(h, h')$ à peu près invariante par translation (dans V). Il est adapté à la recherche du meilleur estimateur affine d'une régionale glissante linéaire en z (par exemple une moyenne mobile). Pour définir (après coup) m , on peut, par exemple, poser

$$m = \frac{1}{V} \int_V m(h) \, dh$$

Sous forme explicite, m apparaît comme une régionale, et constitue une moyenne des valeurs de $z(x)$ dans $S_0 \oplus V$, pondérées par une fonction tenant compte de la distance du point x à la frontière du domaine S_0 . Lorsque le voisinage mobile V est relativement petit vis à vis de S_0 (ce qui est le cas le plus usuel) m diffère peu de la moyenne (ordinaire) de $z(x)$ dans S_0 .

De même, pour la covariance $\bar{C}(u)$, on peut la définir comme la valeur moyenne de $C(h, h+u)$ lorsque h parcourt le domaine $V(u)$, égal à l'intersection de V et de son translaté par $-u$. Sous forme explicite, la régionale $\bar{C}(u)$ apparaîtra comme une valeur moyenne dans $S_0 \oplus V$ de l'expression $z(y)z(y+u)$, pondérée par une fonction tenant compte de la proximité des frontières. Ici encore, si V est petit devant S_0 , des simplifications apparaîtront, et $\bar{C}(u)$ ne différera guère de la moyenne non pondérée de la même expression prise dans le domaine $S_0(u)$: c'est, en somme, la "covariance moyenne" $\bar{C}(h) = g(h)/K(h)$ du dernier paragraphe du chapitre précédent.

Il résulte de ceci que le problème de l'estimation, in praxi, de m et de $\bar{C}(u)$ se présentera, formellement, dans les mêmes conditions que celui de "l'inférence statistique" classique. Avec pourtant cette différence, importante du point de vue méthodologique, que nous savons exactement ce que nous cherchons à estimer : des intégrales d'espace, et que nous connaissons la nature des erreurs que nous commettons : celles qui accompagnent nécessairement le calcul numérique approché d'une intégrale à partir d'un nombre fini de points.

F.A. localement intrinsèque (d'ordre 0)

Pour des raisons de robustesse, comme on l'a déjà indiqué, on remplace le plus souvent la classe des estimateurs affines par celle des estimateurs linéaires autorisés à l'ordre 0. Le moment d'ordre 1, $m(h)$, ne figure plus parmi les prérequisites, et la covariance $C(h, h')$ est remplacée par le variogramme (non centré, non intrinsèque en général).

$$\gamma(h, h') = \frac{1}{2S_0} \int_{S_0} [z(x+h) - z(x+h')]^2 dx$$

Le modèle "F.A. localement intrinsèque (d'ordre 0)" est alors défini, très simplement, par l'hypothèse d'approximation :

$$\gamma(h, h') \neq \bar{\gamma}(h-h')$$

où $\bar{\gamma}$ désigne le variogramme moyen, dont l'expression explicite fait apparaître une moyenne pondérée des $[z(x+u) - z(x)]^2$, d'ailleurs peu différente, si V est relativement petit devant S_0 , du variogramme régional γ_R défini dans le chapitre 7 (avec cette différence que le champ réel S est remplacé par le domaine S_0 choisi par nous).

L'estimation in praxi se fait dans les mêmes conditions que celle du variogramme d'une F.A. intrinsèque à partir d'une réalisation unique : ayant choisi des "classes d'angle et de distance" Δu_i de centres u_i , on forme une expression de la forme

$$(5) \quad \gamma^*(u_i) = \frac{1}{N_i} \sum \frac{[z(x_\alpha) - z(x_\beta)]^2}{2}$$

où la somme est étendue aux N_i couples de points expérimentaux tels que $x_\alpha - x_\beta \in \Delta u_i$. On peut, éventuellement, aussi introduire des poids pour tenir compte des "zones d'influence" des différents échantillons. Il reste ensuite à ajuster à ces valeurs expérimentales discrètes un modèle $\gamma(u)$ ($u \in V$) judicieusement choisi : l'élément décisif est ici, comme d'habitude, le choix du type de comportement à l'origine.

Ainsi, une fois dépouillé de toute interprétation métaphysique, le fameux problème de "l'inférence statistique" comporte deux aspects :

i) examiner dans quelle mesure la somme discrète (5) constitue une approximation acceptable de l'intégrale d'espace correspondante ; c'est là un problème d'estimation global de type banal. On n'est, certes, jamais à l'abri de mauvaises surprises, mais l'expérience indique que, dans la plupart des cas, pourvu que les points expérimentaux ne soient pas trop peu nombreux ni trop mal répartis, cette estimation est possible dans des conditions raisonnables. On pourrait d'ailleurs associer à cette estimation une variance $\text{Var}(\gamma^* - \gamma)$, définie dans le cadre d'une nouvelle représentation, transitive ou glissante, qu'il est inutile d'explicitier ici. Mais le calcul de cette variance d'estimation nécessiterait, à son tour, la spécification d'un modèle approché pour cette nouvelle représentation, et amorcerait ainsi la mauvaise régression à l'infini.

ii) choisir un type de modèle (c'est-à-dire essentiellement un type de comportement de γ au voisinage de l'origine) et l'ajuster aux valeurs expérimentales $\gamma^*(u_i)$. L'ajustement lui-même ne pose guère de problèmes, mais il n'en est pas de même du choix du type : ce choix constitue réellement une hypothèse anticipatrice, vulnérable et féconde à la fois. Ici encore cette hypothèse a le statut d'une approximation anticipatrice. La situation est au fond la même que dans le cas de l'estimation du covariogramme transitif (Ch.6). Le vrai $\bar{\gamma}(u)$, si nous le connaissions, présenterait certainement un comportement très compliqué au voisinage de l'origine, et ne se laisserait certainement pas ramener (en toute rigueur) à l'un ou l'autre des types très simples dont nous faisons habituellement usage. Mais nous pourrions, en première approximation

lui substituer un variogramme γ de type plus simple, et juger si cette approximation est acceptable en examinant ses conséquences sur le problème qui nous occupe : si cette substitution ne modifie que très peu les poids et la variance du krigeage des configurations qui nous intéressent, l'approximation ainsi faite pourra être considérée comme objectivement valable. Dans la situation réelle, in praxi, c'est cette même hypothèse d'approximation que nous avançons à titre d'anticipation, à nos risques et périls d'ailleurs, puisque cette fois nous pouvons nous tromper, et que c'est seulement après coup que nous pourrions vérifier sa validité.

D'un point de vue épistémologique, on peut noter qu'une hypothèse d'approximation de ce genre (remplacement du vrai $\bar{\gamma}(u)$ par un $\gamma(u)$ de type plus simple) est non seulement légitime mais même, dans un sens, obligatoire. Car, si l'on revient à la signification physique des grandeurs mises en jeu, on s'aperçoit qu'au delà d'une certaine échelle de petitesse, la notion même de V.R. devient floue et mal définie, de sorte qu'à cette échelle le concept de comportement analytique du vrai $\bar{\gamma}(u)$ cesse d'être opératoire. Mais, si la simplification ou typification du variogramme est obligatoire pour des raisons physiques, il est certain que le type que nous choisissons in praxi est beaucoup plus simple que celui que nous pourrions adopter après coup, de sorte que nous introduisons réellement une hypothèse (anticipée) d'approximation.

A côté du problème de "l'inférence statistique", dont nous venons de préciser le statut épistémologique, la Statistique Mathématique considère également celui des "tests d'hypothèses". Dans le point de vue très "positiviste" qui est ici le nôtre, nous nous refusons à poser ces problèmes au niveau des généralités : nous pensons qu'il n'y a pas grand sens à estimer un paramètre qui n'a pas fait au préalable l'objet d'une définition opératoire précise, et, de même, nous nous refusons à tester des hypothèses dépourvues de signification objective. Dans le cas qui nous occupe, le variogramme à estimer est défini de manière précise par une intégrale d'espace. De même, "l'hypothèse intrinsèque" qu'il s'agirait de tester est l'hypothèse d'approximation $\gamma(h, h') \neq \bar{\gamma}(h-h')$. Elle a un sens objectif, puisque nous pouvons toujours après coup juger s'il s'agit d'une approximation acceptable (à l'égard de la solution d'un problème donné). Mais il n'est pas facile de la soumettre in praxi à un contrôle précis, utilisant des critères bien définis : parce que, s'il est possible en général d'estimer raisonnablement in praxi le $\bar{\gamma}$ (qui ne dépend que d'un argument u), il n'en est pas du tout de même du $\gamma(h ; h')$, qui dépend de deux arguments : en général, les données disponibles ne permettent pas d'estimer de manière significativement différente $\gamma(h ; h')$ et son translaté $\gamma(h+u ; h'+u)$.

De fait, les accrocs éventuels à la stationnarité, ou au caractère intrinsèque, de la représentation glissante sont imputables uniquement à l'allure que présente la V.R. dans la zone de bordure, comprise entre le dilaté $S_0 \oplus V$ et l'érodé $S_0 \ominus V$: or cette zone contient, en général, trop peu de points expérimentaux pour qu'il soit possible de tester avec précision son homogénéité ou son hétérogénéité. En pratique, cependant, un géostatisticien averti n'hésitera pas beaucoup. D'un côté, en effet, l'information non numérique (qualitative) qu'il possède sur le phénomène physique lui indique, en général, s'il est ou non raisonnable d'admettre que la proximité de la frontière n'exerce pas d'influence sensible. Ensuite, l'aspect des "variogrammes bruts" (les $\gamma^*(u_i)$) dans les différentes directions est en général assez révélateur : des types d'anisotropie incohérents et instables, des allures violemment convexes contrastant avec la croissance plus modérée que l'on peut observer dans d'autres directions constituent des indices très sûrs de non stationnarité. Au contraire, des variogrammes relativement stables, d'allure concave ou au plus linéaire, des anisotropies de type simple etc... inciteront à adopter le modèle intrinsèque local.

Dans ce qui précède, nous avons adopté, pour simplifier, le critère fort d'objectivité (décidabilité, au lieu de la seule falsifiabilité poppérienne) selon lequel un énoncé est objectif s'il est susceptible d'être déclaré univoquement vrai ou faux après coup, c'est-à-dire une fois $z(x)$ connu en tout $x \in S$. Mais la question du comportement au voisinage de l'origine du "vrai" variogramme $\bar{\gamma}$ nous oblige à rectifier cette position, et à revenir au critère poppérien de la falsifiabilité. De fait, pour un $\gamma(h, h')$ donné (non intrinsèque) associé à la représentation glissante d'une V.R. z , il existe plusieurs manières possibles de définir un $\bar{\gamma}$ moyen, acceptable vis-à-vis d'un problème donné avec une approximation donnée. Celle que nous avons choisie ne constitue que l'une des définitions possibles de ce $\bar{\gamma}$, même si c'est la plus simple possible. En d'autres termes, le modèle "F.A. localement intrinsèque" n'est pas univoquement défini, et, parmi les différents $\bar{\gamma}$ également compatibles avec la réalité (après coup), rien n'indique que l'un soit plus particulièrement "vrai" que les autres (tout dépendrait du critère choisi). Un modèle donné de variogramme $\gamma(h)$ peut certainement être déclaré faux (après coup) s'il conduit à des estimations grossièrement différentes de celles que l'on formerait sur la base du vrai $\gamma(h, h')$. Mais il ne peut être déclaré vrai que dans un sens relatif, et non exclusif, s'il conduit à des estimations peu différentes. On a vu au chapitre 4 un exemple frappant où un même ensemble de données numériques, soumis à différents auteurs, a été interprété par des modèles extrêmement différents, mais conduisant à des solutions pratiquement équivalentes du problème posé. De ces différents modèles, on ne peut pas dire qu'ils ont été vérifiés, puisqu'un seul d'entre eux pourrait être déclaré vrai, mais

seulement qu'ils ont été corroborés (non réfutés). Par conséquent, c'est bien le critère de falsifiabilité qu'il convient d'appliquer ici.

D'autre part, à ce $\bar{\gamma}$, quelle que soit sa définition, nous avons, en réalité substitué un modèle γ , caractérisé par un type de comportement au voisinage de l'origine (effet de pépite, terme en $|h|^\lambda$ etc...) : cette notion de type (analytique) de comportement est bien défini mathématiquement, mais doit être, du point de vue physique, entièrement reconstruite en termes purement opératoires. La définition mathématique fait, en effet, intervenir des passages à la limite auxquels rien ne répond dans la réalité physique. Cette reconstruction opératoire se fait exactement de la même manière que dans le cas d'un covariogramme transitif, et il n'y a pas lieu d'y revenir.

F.A. localement intrinsèque d'ordre k

Il s'agit cette fois d'une adaptation locale du modèle non stationnaire rencontré à la fin du chapitre 7. On y a recours lorsque le modèle "F.A. localement intrinsèque d'ordre 0" ne convient décidément pas. Il reste alors la ressource de restreindre encore la classe Φ de nos estimateurs linéaires, en imposant à leurs coefficients des "conditions d'universalité" à un ordre k donné. Ces conditions expriment simplement que l'erreur d'estimation est une combinaison linéaire autorisée à cet ordre (c'est-à-dire filtrant, ou annulant, les polynômes de degré inférieur ou égal à k). Ces estimateurs constituent, pour ces polynômes, des interpolateurs exacts (en ce sens qu'ils donnent les valeurs exactes de $z(x)$ lorsque la V.R. est elle-même un polynôme de degré $\leq k$). On peut dire encore que l'erreur d'estimation reste invariante lorsque l'on ajoute à z un polynôme quelconque de degré $\leq k$.

On a vu, à la fin du chapitre précédent, qu'une fonction aléatoire intrinsèque d'ordre k doit plutôt être considérée comme une classe de F.A. admettant des covariances de la forme :

$$(6) \quad C(h, h') = K(h-h') + \sum a_1(h) f^1(h') + \sum a_1(h') f^1(h)$$

où K est une "covariance généralisée", caractéristique du modèle, tandis que les fonctions a_1 dépendent de la version choisie et restent inaccessibles in praxi. Mais la variance d'une combinaison linéaire autorisée, et, en particulier notre variance d'estimation, ne dépend que de K , et non de la partie non stationnaire de la covariance

Ceci suggère la définition du modèle "F.A. localement intrinsèque d'ordre k " : ce modèle consiste à admettre que la covariance $C(h, h')$ de la représentation glissante vérifie (approximativement) la relation (6) pour une covariance généralisée K donnée. Il s'agit donc, comme dans tous les modèles locaux, d'une hypothèse d'approximation, que l'on adopte in praxi de manière anticipée et que l'on peut en principe contrôler après coup en vérifiant qu'elle ne modifie que peu la solution du problème qui nous intéresse.

Notons, en passant, qu'avec cette définition les conditions d'universalité, que nous imposons à nos estimateurs, ne peuvent plus être interprétées comme des conditions de "non biais". Le non biais est, d'ailleurs, une propriété définie dans le modèle seulement, et ne pourrait recevoir une signification objective qu'après avoir subi la reconstruction habituelle. Nos conditions sont, en termes physiques, des conditions de filtrage. Elles ont pour effet d'éliminer l'influence d'une "dérivée éventuelle, c'est-à-dire d'une composante suffisamment régulière pour que l'on puisse localement (c'est-à-dire sur un voisinage V_x de chaque point x) l'assimiler à un polynôme de degré $\leq k$.

Ce modèle est spécifié par la seule covariance généralisée $K(h)$, qu'il suffit d'ailleurs de définir sur la voisinage $2V$ (double de V) de l'origine. Nous n'avons aucunement besoin de connaître les fonctions $a_1(h)$ de la formule (6). Il est possible après coup, de les définir de façon objective, mais certainement pas de les estimer, in praxi, dans des conditions raisonnables. En ce qui concerne la covariance généralisée, sa signification objective est la même que celle d'un variogramme local, et sa reconstruction opératoire peut être effectuée de manière analogue. Son estimation in praxi (à partir des combinaisons linéaires autorisées que l'on peut former à l'aide des données expérimentales) est, très souvent, possible dans des conditions acceptables.

Ces questions, déjà plus complexes ne peuvent pas être abordées ici, et le lecteur devra se reporter à la littérature spécialisée.

L'espérance conditionnelle est-elle opératoire ?

Les modèles locaux que nous venons de passer en revue sont tous des modèles d'ordre deux (c'est-à-dire spécifiés par la donnée d'une covariance, d'un variogramme ou d'une covariance généralisée) et ne permettent d'aborder que des problèmes d'estimation linéaires : la classe Φ à laquelle nous décidons de limiter la recherche d'un estimateur optimal, du fait même que nous adoptons l'un de ces modèles, est la classe très pauvre des estimateurs linéaires autorisés à tel ou tel ordre. Mais la pratique pose souvent des problèmes dont la solution nécessite le recours à des estimateurs plus puissants, non linéaires. Citons simplement deux exemples :

~ Lorsque l'on krige, par exemple, un point x à partir des points expérimentaux disponibles dans le voisinage mobile V_x , on assortit cette estimation d'une variance de krigeage que l'on sait calculer. Mais, comme nous l'avons remarqué, cette variance est un paramètre global, et non local. Elle n'est pas liée, spécialement, aux particularités structurelles que présente la V.R. sur ce voisinage V_x , mais représente simplement la valeur quadratique moyenne des erreurs que l'on commettrait en estimant chacun des points de S à l'aide de la même configuration de points expérimentaux. S'il se trouve qu'au voisinage du point x considéré la V.R. présente un comportement plus erratique ou, au contraire, plus régulier qu'en moyenne, on s'attend bien, physiquement, à ce que l'erreur soit plus forte, ou au contraire, plus faible en ce point. Mais notre variance d'estimation, n'ayant qu'un sens global, ne peut pas rendre compte de cet effet. Dans le cadre d'une représentation glissante, il n'est, par construction, pas possible de donner une signification locale à la variance d'estimation. Mais, faute de pouvoir localiser l'erreur (c'est-à-dire exprimer la variance en fonction du point x à estimer) nous pourrions espérer la "conditionnaliser" (c'est-à-dire l'exprimer en fonction des valeurs z_i observées aux points expérimentaux $x_i \in V_x$). Cette variance conditionnelle constituerait un bon substitut de la variance localisée que nous ne pouvons atteindre dans le cadre de ce modèle. En effet, si la V.R. présente au voisinage de x une dispersion plus forte ou plus faible qu'ailleurs, les valeurs expérimentales z_i observées dans V_x sont elles-mêmes plus, ou moins, dispersées qu'en moyenne, et par conséquent, si le modèle est bon, la variance conditionnelle peut rendre compte de cet effet local.

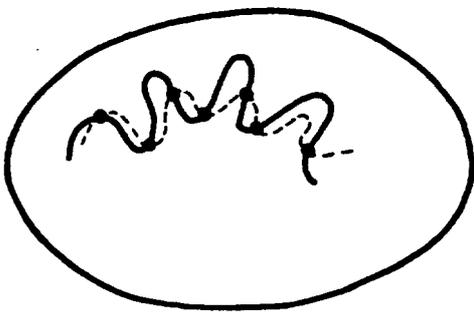
~ le deuxième exemple est emprunté à la pollution (on aurait une formulation analogue, mais plus complexe, en mines avec le problème de l'exploitation sélective). Ayant mesuré les teneurs z_α en polluant un certain nombre de points x_α , on s'inquiète de la probabilité pour que la teneur moyenne $z(v_{x_0}) = (1/v) \int_v z(x_0+x) dx$ en polluant dans un voisinage v_{x_0} d'un point x_0 donné dépasse un certain seuil d'alerte z_0 : en représentation glissante (en supposant $v \subset V$) il s'agit donc d'évaluer la probabilité $P(Z(v) \geq z_0)$ conditionnée par les observations $Z_i = z_i$ disponibles dans le voisinage mobile v .

Ces problèmes, et d'autres analogues, nécessitent le recours aux lois conditionnelles. Comme d'habitude, avant de faire usage d'un concept mathématique, nous devons nous poser deux questions : est-il possible de redéfinir ce concept d'une manière purement opératoire (point de vue après coup), et pouvons nous, sur la base de l'information actuelle, en obtenir une estimation raisonnable (point de vue in praxi) ? La réponse à la première question ne sera que partiellement positive ; à la deuxième question, elle sera catégoriquement négative dès qu'il y a plus d'un ou deux points conditionnants, et notre conclusion sera la suivante : même si nous disons et nous croyons utiliser l'espérance conditionnelle, c'est-à-dire chercher notre estimateur optimal dans la classe immensément vaste de toutes les fonctions mesurables, ce que nous faisons en réalité est tout différent, et l'espace Φ des fonctions que nous manipulons réellement est toujours beaucoup plus restreint. L'espérance conditionnelle, telle qu'on peut la définir après coup en termes opératoires, est toujours inaccessible in praxi (même si nous avons l'illusion de l'atteindre) et nous lui substituons toujours (consciemment ou non) des expressions beaucoup plus simples, plus ou moins grossièrement approchées.

Après coup : l'objectivité des lois conditionnelles

Considérons notre représentation glissante, c'est-à-dire la F.A. définie par $Z(h) = z(\underline{x}+h)$ avec $h \in V$, et \underline{x} uniformément distribué dans S , et demandons-nous quelle figure revêtent, dans ce cadre strictement objectif, les diverses lois conditionnelles auxquelles nous pouvons nous intéresser. Sans entrer dans le détail du formalisme mathématique, il est clair que, pour un $h \in V$ et un nombre réel ζ donnés,

fixer la valeur $Z(h) = \zeta$ de la V.A. $Z(h)$ revient à obliger le point générique x à décrire l'ensemble $\{x : z(x+h) = \zeta\}$ au lieu de S entier. Cet ensemble est le translaté $L_{-h}(\zeta)$ par $-h$ de la ligne (dans R^2) ou surface (dans R^3) de niveau $L(\zeta) = \{x : z(x) = \zeta\}$. Après le conditionnement, donc, le point générique x n'est plus distribué uniformément dans S , mais se trouve affecté d'une loi conditionnelle concentrée sur cette ligne $L_{-h}(\zeta)$. Si maintenant f est une V.A. du modèle, c'est-à-dire une fonction $f(x)$ mesurable sur S , son espérance conditionnelle n'est plus $(1/S) \int_S f(x) dx$: elle est donnée par une intégration effectuée sur la ligne $L_{-h}(\zeta)$.



-Fig.: 11-

De la même manière, si nous conditionnons sur deux points h_1 et h_2 , fixer $Z(h_1) = \zeta_1$ et $Z(h_2) = \zeta_2$ revient à assujettir le point générique à décrire l'intersection :

$$L_{-h_1}(\zeta_1) \cap L_{-h_2}(\zeta_2)$$

des translatées par $-h_1$ et $-h_2$ des deux lignes de niveau correspondantes. Dans R^2 , cette intersection est en général constituée de points isolés, en nombre infini, si nous attribuons à la V.R. z un microcomportement très pathologique (analogue, par exemple, à celui de la trajectoire d'un mouvement brownien), mais, le plus souvent, fini (si nous faisons une hypothèse physiquement plausible sur ce microcomportement). Dans R^3 , la même circonstance se produit à partir de trois points conditionnants, et, de manière générale, dans R^n à partir de n points. Si nous prenons un nombre p , pas nécessairement très élevé, mais supérieur à n (par exemple : $p = 9$ dans R^2 ou R^3) de points conditionnants, nous devons nous attendre à ce que l'intersection

$$L_{-h_1}(\zeta_1) \cap \dots \cap L_{-h_p}(\zeta_p)$$

soit vide ou réduite à un point unique (sauf peut-être pour certains choix exceptionnels des valeurs ζ_1, \dots, ζ_p). Cela veut dire qu'il y a, en général, au plus une seule implantation du point x telle que l'on ait $z(x+h_1) = \zeta_1, \dots, z(x+h_p) = \zeta_p$. Autrement dit, les lois conditionnelles sont dégénérées. Pour toute fonction f , il existe une fonction $F(\zeta_1, \dots, \zeta_p)$ telle que l'on ait (sauf peut être en quelques points x exceptionnels)

$$f(x) = F(z(x+h_1), \dots, z(x+h_p))$$

puisque la donnée des valeurs numériques $\zeta_i = z(x+h_i)$, $i = 1, \dots, p$ suffit à déterminer le point x (donc aussi la valeur $f(x)$ de la fonction f en ce point). Naturellement, cette fonction F aura une structure très compliquée, et nous n'aurons aucun espoir de la déterminer in praxi.

En fait, cette loi conditionnelle (dégénérée) sera très instable, et on peut s'interroger sur sa réalité physique. Le point x , en effet, apparaît comme une fonction $x(\zeta_1, \dots, \zeta_p)$ des valeurs conditionnantes. Mais cette fonction n'est pas définie partout, puisque toutes les configurations $z(x+h_i) = \zeta_i$ n'existent pas nécessairement. De plus, on peut s'attendre à ce que de faibles variations de l'une ou de l'autre des valeurs conditionnantes se répercutent par des variations considérables du point x , qui pourrait par exemple passer d'une extrémité à l'autre du domaine S .

Les caractéristiques des lois conditionnelles que nous mettons ainsi en évidence sont liées de la manière la plus étroite au microcomportement de la V.R. $z(x)$, et il y a tout lieu de penser qu'à l'échelle extraordinairement fine à laquelle nous sommes conduits à travailler, nous avons en fait dépassé depuis longtemps le seuil au-dessous duquel la notion même de V.R. cesse d'être opératoire. En effet, les propriétés fines des lignes de niveau (et des intersections de leurs translatées) ne sont pas accessibles ni même réellement définissables sur une base expérimentale. Exactement comme dans le cas de l'examen critique de la continuité ou de la dérivabilité nous devons revenir au critère poppérien - et l'appliquer à la V.R. elle-même.

Dans ce cadre poppérien, les choses se simplifient un peu, mais les conclusions essentielles - l'extrême instabilité des lois et espérances conditionnelles - demeurent. Tout d'abord, pour tenir compte du fait que les valeurs numériques de $z(x)$ ne sont connaissables (et même définissables) expérimentalement qu'avec un petit nombre de chiffres significatifs, nous devons discrétiser les paramètres ζ , c'est-à-dire ne retenir qu'un nombre fini de classes de valeurs C_i . Les lignes de niveau $L(\zeta)$ sont ainsi remplacées par des bandes $L(C_i) = \{x : z(x) \in C_i\}$ présentant une certaine épaisseur. De même, faute de pouvoir atteindre expérimentalement une infinité de points, nous devons également discrétiser le domaine S lui-même. Par exemple, S sera considéré comme réunion d'un grand nombre de petits carrés ou petits cubes jointifs, et les valeurs expérimentalement accessibles représenteront, non plus des valeurs ponctuelles, mais plutôt des moyennes prises sur ces petits carrés ou petits cubes.

Dans cette vision discrétisée des choses, nous ne pouvons plus affirmer que le point x soit déterminé par la donnée des valeurs discrétisées des ζ_i . Cela dépendra du nombre des classes retenues. Une intersection du type

$$L_{-h_1}(C_{i_1}) \cap \dots \cap L_{-h_p}(C_{i_p})$$

pourra, en effet, contenir éventuellement plusieurs petits carrés. Mais, dans ce cas, le plus souvent ces petits carrés seront très proches les uns des autres. Parfois, pourtant, pour certains choix des valeurs conditionnantes discrétisées, on verra apparaître plusieurs paquets de petits carrés, implantés dans des endroits très différents. De plus, il suffira souvent que l'un des ζ_i passe d'une classe de valeurs à la classe voisine pour que ce ou ces paquets de petits carrés subissent dans S des déplacements considérables, ou disparaissent, ou qu'apparaissent de nouveaux paquets en des endroits inattendus etc... Cette extrême instabilité peut être visualisée à l'analyseur de texture. (Voir fig. 12).

Du fait de son instabilité et de son extrême irrégularité, cette loi ou cette espérance conditionnelle (bien que définissable en principe) serait - si on la connaissait - beaucoup trop compliquée pour qu'il soit possible de l'utiliser en pratique. De plus, il est clair qu'il ne sera jamais possible de l'estimer in praxi. Ce caractère illusoire est facilement mis en évidence par un examen des ordres de grandeurs.

Quelques ordres de grandeurs

Pour dégager des ordres de grandeurs, considérons l'exemple suivant, qui schématise un cas banal de reconnaissance minière (à 2 dimensions). Les données disponibles (sondages) représentent des teneurs moyennes sur un support $s = 20 \text{ cm}^2$ et sont implantées à maille $100 \times 100 \text{ m}$. La surface S qui nous intéresse représente 200 hectares, soit $S = 10^9 s$. Le voisinage mobile V est un carré de $300 \times 300 \text{ m}$. et contient donc 9 sondages. Sa surface est $V = 4,5 \cdot 10^7 s$. L'espace du gisement est discrétisé en petits carrés de 20 cm^2 (et on admet que les sondages s sont identifiables à de tels petits carrés). Les valeurs de la teneur sont discrétisées en C classes (par exemple $C = 5, 10$ ou 20).

Notre voisinage mobile contenant 9 sondages, nous cherchons à calculer des espérances conditionnelles en Z_1, \dots, Z_9 . Or, 9 variables, prenant chacune C valeurs, donnent naissance à C^9 configurations possibles :

$$(C)^9, \text{ soit : } \begin{array}{l} 5.10^{11} = 500 \text{ milliards si } C = 20 \\ 10^9 = \text{un milliard si } C = 10 \\ 2.10^6 = \text{deux millions si } C = 5 \end{array}$$

Le nombre des configurations réellement présentes dans le gisement est $S/s =$ un milliard. C'est donc seulement pour $C = 5$ (c'est-à-dire pour une discrétisation extrêmement grossière), que nous pouvons espérer qu'un nombre important des $C^9 =$ deux millions de configurations possibles soient représentées dans le gisement de façon suffisamment substantielle pour que le raisonnement statistique prenne un sens physique (après coup : in praxi, nous ne disposons que de 200 configurations, et aucune estimation raisonnable n'est possible).

Ce nombre $(C)^9$ est aussi la dimension de l'espace des fonctions mesurables de 9 variables prenant C valeurs chacune. Ainsi, avec $C = 10$ classes (ce qui dans bien des cas est déjà très grossier) l'espace de travail requis pour la détermination d'une espérance conditionnelle est de dimension $N =$ un milliard. Pouvons-nous sérieusement prétendre que nous travaillons réellement dans un espace aussi riche, et, en particulier, nous imaginer qu'il soit possible in praxi d'estimer un milliard de paramètres à partir de 200 données expérimentales ?

Dans le krigeage usuel (à l'ordre 0), l'espace de travail est de dimensions 8 (6 pour le krigeage d'ordre 1). De 8 à un milliard il manque manifestement un intermédiaire. Dans le krigeage disjonctif (voir ci-dessous) la classe Φ des estimateurs est constituée des fonctions de la forme

$$f_1(z_1) + f_2(z_2) + \dots + f_N(z_N)$$

somme de N fonctions d'une seule variable (N est le nombre des données disponibles dans V, ici $N = 9$). Avec C classes de valeurs, l'espace correspondant a la dimension C^N (au lieu de C^9 pour l'espérance conditionnelle), soit ici, avec 9 données et 10 classes, une dimension 90, déjà élevée, mais nettement plus raisonnable qu'un milliard. Le krigeage disjonctif représente bien l'intermédiaire cherché.

In praxi : l'estimation des lois conditionnelles

Les ordres de grandeurs précédents sont décisifs. Il est hors de question d'estimer réellement in praxi une loi ou une espérance conditionnelle si le nombre des points conditionnants dépasse 1 ou 2. On peut les calculer, mais cela suppose l'emploi d'un modèle infiniment plus spécifié que les données disponibles ne le permettent réellement. D'ailleurs les modèles (illégitimement) ultra spécifiés que nous pouvons utiliser en pratique conduisent tous à des expressions relativement simples de l'espérance conditionnelle, beaucoup plus simples, en tous cas, que celles que l'on obtiendrait après coup, en **appliquant** la définition opératoire, et sans grand rapport peut être avec elle. Il semble donc bien qu'en effectuant ces calculs nous ayons largement transgressé tous les seuils de réalisme et de robustesse.

Mais il y a, heureusement, une autre manière de présenter les choses. Renonçant complètement à l'idée de déterminer la véritable espérance conditionnelle, nous pouvons nous contenter de faire du modèle ultra spécifié, ou plutôt de son type, un usage purement heuristique. Ce modèle, en effet, nous suggère une classe Φ d'algorithmes, dépendant d'un petit nombre de paramètres (ceux qui spécifient le type du modèle) dont chacun représente l'expression (dans le modèle correspondant) de notre espérance conditionnelle. Oubliant cette interprétation, nous nous fixons simplement pour tâche de choisir dans cette classe Φ (telle qu'elle est) l'algorithme qui convient le mieux à notre problème.

Cette manière, très monoscopique, de procéder soulève évidemment un certain nombre de questions. Et tout d'abord, puisque le type du modèle ultra spécifié ne joue qu'un rôle heuristique et que nous ne prétendons plus lui attribuer de valeur objective, qui nous garantit qu'un autre choix ne nous aurait pas conduit à un meilleur estimateur ? Plus généralement, il n'y a pas de raison décisive qui nous contraigne à nous limiter aux classes Φ d'algorithmes associés aux espérances conditionnelles de tel ou tel type de modèle. Nous pourrions, aussi bien, choisir a priori une classe Φ arbitraire, pourvu qu'elle ne dépende que d'un nombre limité de paramètres et ne conduise pas à des calculs trop compliqués. C'est, effectivement, ainsi que l'on pourrait présenter le krigeage disjonctif (voir ci-dessous). Mais bien d'autres choix seraient possibles, et cela ouvre à la recherche des horizons assez vastes.

Ceci dit, il faut aussi tenir compte de ce qui est raisonnablement possible in praxi. Et, en premier lieu, nous devons nous demander jusqu'à quel point (et moyennant quelles hypothèses d'approximation anticipatrices) nous pouvons espérer procéder à une reconstitution approximative de la loi spatiale de la représentation glissante $Z(h)$. La réponse, je crois, est assez nette : moyennant une hypothèse d'approximation de type stationnarité locale, nous pouvons espérer reconstituer assez bien la loi à une seule variable $Z(h)$, obtenir quelques indications sur les lois à deux variables $Z(h)$, $Z(h')$ et rien de plus : nous n'aurons, en général, aucun espoir de tirer de nos données des renseignements crédibles sur les lois à plus de deux variables. Cette conclusion me conduit à dire quelques mots des estimateurs de type disjonctif, déjà évoqués ci-dessus.

Les estimateurs disjonctifs

Dans bien des cas, on souhaite disposer d'un estimateur plus puissant que les estimateurs linéaires examinés dans les chapitres précédents, mais ne nécessitant pas la connaissance de paramètres aussi nombreux, aussi peu accessibles in praxi (et peut être aussi dépourvus de significations objectives) que l'espérance conditionnelle. Donnons un exemple, emprunté à la pollution. Supposons que nous ayons mesuré les teneurs $z(x_\alpha) = z_\alpha$ en un certain polluant, d'échantillons prélevés aux points x_α , $\alpha = 1, 2, \dots, N$. Au dessus d'un certain seuil, disons z_0 , on parle de contamination. Nous souhaitons estimer, à l'intérieur d'une surface S_0 donnée, ce que l'on pourrait appeler le pourcentage de contamination. La définition précise est la suivante. Pour chaque point x , posons $\theta(x) = 1$ si $z(x)$ dépasse le seuil d'alerte z_0 , et $\theta(x) = 0$ dans le cas contraire. La superficie contaminée dans S_0 a pour valeur :

$$S(z_0) = \int_{S_0} \theta(x) \, dx$$

et le pourcentage de contamination est défini par le rapport $S(z_0)/S_0$, exprimée en %. Comme $\theta(x)$ est une variable régionalisée, associée à $z(x)$, la superficie contaminée $S(z_0)$ est elle-même une grandeur régionale, et c'est cette régionale que nous souhaitons estimer à partir de nos quelques données numériques. Faute de disposer d'un modèle stationnaire global raisonnablement adopté à notre problème, nous avons toujours la ressource d'estimer $\theta(x)$ en chaque point x à l'aide d'un modèle stationnaire local et de procéder ensuite à un recollement de ces estimations locales, de manière à reconstituer un estimateur de $S(z_0)$: si le voisinage de travail V (qui doit pour cela contenir un nombre suffisant de points expérimentaux) est relativement petit vis à vis

de S_0 , l'hypothèse d'approximation anticipée que nous avançons en choisissant ce modèle à des chances raisonnables de se révéler acceptable après coup, et nous pouvons même déjà la contrôler en partie in praxi. Nous sommes ainsi ramenés au problème de l'estimation locale d'une V.R. "en tout ou rien" (c'est-à-dire ne prenant que les deux valeurs 0 et 1). Or, les estimateurs linéaires sont particulièrement mal adaptés au cas des V.R. en tout ou rien, comme le savent bien tous ceux qui ont eu l'occasion de traiter, en pratique, ce genre de problèmes. D'un autre côté, l'espérance conditionnelle de $\theta(x)$ est inaccessible, et peut-être illusoire. D'où l'idée de recourir à "l'estimateur optimal de type disjonctif" appelé, plus brièvement "krigeage disjonctif". Cet estimateur est choisi dans la classe Φ des fonctions de la forme suivante (où z_1, \dots, z_n représentent les n données disponibles dans le voisinage V_x du point x) :

$$f(z_1, \dots, z_n) = f_1(z_1) + f_2(z_2) + \dots + f_N(z_N)$$

(Il s'agit donc des fonctions qui se présentent comme la somme de N fonctions mesurables d'une seule variable. Nous avons déjà remarqué plus haut que cet espace fonctionnel Φ représentait, du point de vue de sa dimension, un compromis raisonnable entre l'espace vraiment trop pauvre des combinaisons linéaires, et celui, infiniment trop riche, des fonctions mesurables de N variables. Parmi les fonctions de la classe Φ , nous choisirons, pour jouer le rôle d'estimateur, celle qui minimise (dans notre modèle localement stationnaire) la variance d'estimation correspondante. Il apparaît alors que cette variance s'exprime à l'aide des seules lois marginales d'ordre deux, c'est-à-dire des seules lois à deux variables $(Z(x), Z(x_1))$ ou $(Z(x_1), Z(x_j))$ du modèle. Et nous venons de voir qu'il n'était pas déraisonnable de chercher à estimer in praxi ces lois à deux variables seulement : nous sommes en présence d'un problème bien posé. Le tableau suivant montre la situation intermédiaire qui est celle du krigeage disjonctif :

	Krigeage	Krigeage disjonctif	Espérance conditionnelle
Estimateur optimal parmi les fonctions de la forme :	$\sum \lambda_i z_i$	$f_1(z_1) + \dots + f_n(z_n)$	$f(z_1, \dots, z_n)$
Prerequisites	La matrice des covariances	Les lois marginales à deux variables	La loi à n+1 variables

La théorie montre alors que, dans la classe Φ défini ci-dessus, notre estimateur optimal est déterminé, de manière univoque, par un système d'équations intégrales qu'il n'est pas nécessaire de reproduire ici. Bien que ces équations intégrales soient d'un type relativement simple, on ne sait pas, en général, trouver leur solution sous forme explicite. On peut, évidemment, envisager un calcul numérique approché, mais il faudrait résoudre un système linéaire d'assez grande dimension : cela coûterait cher, d'une part, et d'autre part ce système risque d'être assez instable, en raison même de sa taille. Nous sommes, en effet, sans doute déjà tout proche du seuil de robustesse. Peut être même l'avons nous déjà dépassé, et notre modèle a-t'il déjà commencé, à notre insu, à fonctionner de façon partiellement heuristique.

Puisque le pas est vraisemblablement déjà franchi, nous sommes tentés d'aller un tout petit peu plus loin, et d'avancer encore une hypothèse qui, sans grand risque supplémentaire, nous permettra de simplifier considérablement notre système d'équations intégrales, et nous conduira à un algorithme particulièrement économique et facile à mettre en oeuvre. Cette hypothèse consiste à choisir un "modèle isofactoriel". Voici, en quelques mots, de quoi il s'agit.

Considérons d'abord le cas de (n+1) variables aléatoires, non indépendantes, Y_0, Y_1, \dots, Y_n , gaussiennes, d'espérance nulle et de variance unité, telles aussi que toutes les lois marginales à 2 variables soient elles-mêmes des lois de Gauss à deux variables, caractérisées par leur coefficient de corrélation. (On peut aussi considérer le cas un tout petit peu plus général où ces lois à deux variables sont du type "hermitien"). Il existe alors une suite de polynômes H_p , appelés "polynômes d'Hermite", qui possède deux propriétés remarquables : si $p \neq m$, et si Y_i et Y_j sont deux quelconques de nos variables (j peut être égal à i, et dans ce cas il s'agit de la même variable), les variables aléatoires $H_p(Y_i)$ et $H_m(Y_j)$ sont orthogonales

(c'est-à-dire : admettent une covariance, ou un coefficient de corrélation, égal à 0). De plus, si Y est l'une de nos variables et f une fonction mesurable quelconque (telle seulement que le moment d'ordre 2, $E[f(x)^2]$ soit fini), $f(x)$ admet un développement convergent (en moyenne quadratique) de la forme :

$$(1) \quad f(Y) = \sum f_p H_p(Y)$$

avec des coefficients f_p que la propriété d'orthogonalité permet de calculer très facilement. Nous dirons que les polynomes H_p sont les "facteurs" de notre modèle.

On démontre alors ceci : pour chaque indice p , si l'on veut estimer $H_p(Y_0)$ à l'aide des n variables Y_1, \dots, Y_n , l'estimateur disjonctif optimal H_p^x est une combinaison linéaire des seuls polynomes de même degré p , $H_p(Y_1), \dots, H_p(Y_n)$, et les coefficients de cette combinaison linéaire sont donnés par un système linéaire de rang r (identique au système habituel du krigeage). De plus, l'estimateur optimal de la variable $f(Y_0)$ s'obtient en remplaçant, dans le développement (1), chaque facteur H_p par son krigeage disjonctif H_p^x , soit

$$f^x = \sum f_p H_p^x$$

En bref, dans ce modèle, donc, on krige séparément chacun des facteurs. Au lieu d'avoir à résoudre un seul système linéaire de rang extraordinairement élevé (théoriquement infini, puisqu'il s'agit d'équations intégrales), il suffit de résoudre séparément autant de systèmes de rang n (beaucoup plus modeste) que l'on a retenu de termes dans le développement (1). En pratique, on a rarement besoin de plus d'une dizaine de termes, et le krigeage disjonctif se ramène à une dizaine de krigeages ordinaires.

Revenons maintenant à nos variables régionalisées. Les variables $Z(x), Z(x_1) = Z_1, \dots, Z(x_n) = Z_n$ associées, dans notre modèle localement stationnaires, au point x à estimer et aux données expérimentales x_1, \dots, x_n , ne sont évidemment pas gaussiennes, sauf cas particulier. Mais on peut trouver des gaussiennes Y_0, Y_1, \dots, Y_n et une transformation croissante, ou anamorphose, telle que $Z(x) = \phi(Y_0), Z_1 = \phi(Y_1), \dots, Z_n = \phi(Y_n)$. Ces variables Y , inversement, se déduisent des variables Z par la transformation réciproque. Chacune d'elles, prise séparément, obéit à une loi de Gauss à une seule variable. Cela n'implique pas, en général, qu'elles obéissent, prises deux à

deux, à des lois de Gauss à deux variables. Cependant, nous admettrons qu'il en est ainsi. Ou, plus précisément dit, nous choisissons un modèle dans lequel cette propriété (à savoir : que l'anamorphose ϕ restitue des lois de Gauss à 2 variables) est vérifiée, et c'est ce modèle générique que nous appelons modèle isofactoriel. La suite des opérations se déroule ensuite comme dans le cas gaussien, puisque toute fonction d'une variable Z devient, par la transformation ϕ , une fonction de la variable Y correspondante : à nouveau, le krigeage disjonctif se réduit à une dizaine de krigeages simples.

Cette hypothèse supplémentaire (celle d'une "anamorphose gaussienne d'ordre 2") que nous avons introduite à seule fin de nous ramener au cas simple d'un modèle isofactoriel, n'est pas dépourvue de signification objective. On peut, effectivement contrôler après coup, et même déjà in praxi, si les lois marginales des variables transformées sont, ou non, à peu près des lois de Gauss à deux variables. Mais, à supposer que notre hypothèse soit, effectivement, à peu près vérifiée, nous transgressons pourtant certainement le seuil de robustesse, et peut être même le seuil d'objectivité, en déduisant de cette hypothèse que tous les facteurs sont, réellement deux à deux orthogonaux. En toute rigueur, le modèle isofactoriel fonctionne déjà de façon simplement heuristique.

D'un point de vue monoscopique, un modèle heuristique peut être considéré comme un réservoir d'algorithmes, réservoir dans lequel nous allons chercher tel ou tel élément qui nous semble convenir à la solution du problème que nous cherchons à résoudre. Il est d'ailleurs parfaitement clair qu'en l'absence des suggestions de la théorie, nous n'aurions jamais pensé à former justement cet algorithme là : c'est en ce sens qu'il s'agit d'un usage heuristique du modèle. Usage nullement illégitime d'ailleurs, à condition que nous n'accordions pas une confiance aveugle au modèle, et que nous procédions, après coup, et déjà, si possible, in praxi, à un contrôle expérimental aussi sévère que possible. Après coup, nous avons tout loisir de comparer les valeurs réelles aux estimations que l'on aurait pu en former, par krigeage disjonctif, sur la base de tels et tels points expérimentaux ; de confronter les variances d'estimation prévues par le modèle à leurs valeurs réelles, expérimentales ; de mettre en concurrence, enfin, les estimateurs suggérés par différents modèles heuristiques : krigeage disjonctif, espérance conditionnelle etc... In praxi, les possibilités de contrôle sont plus réduites. Pourtant, les lois à deux variables peuvent déjà être examinées, et l'hypothèse d'une anamorphose gaussienne d'ordre 2 peut faire l'objet de tests. De plus, nous pouvons pour chaque point expérimental, faire comme s'il était inconnu, et l'estimer, par krigeage disjonctif, à partir de ses

voisins : d'où, à nouveau, possibilité de comparer les valeurs vraies et estimées, les variances expérimentales et théoriques etc... Moyennant ces contrôles, nous conférons à nos estimateurs une signification objective, qui n'est d'ailleurs plus exactement celle que leur prêtaient les modèles heuristiques qui nous les avaient suggérés, et nous pouvons ensuite procéder à la reconstruction opératoire des concepts physiques qu'ils mettent, implicitement, en jeu.

