

# LA CARTOGRAPHIE ET L'EXTRAPOLATION DE L'INFORMATION

Jacques BERTIN

*Directeur d'Etudes à l'Ecole Pratique des Hautes Etudes*

Je tiens tout d'abord à remercier Gilles SAUTER, Jean Paul TROUCHAUD et l'ORSTOM de m'avoir permis d'être aujourd'hui parmi vous. C'est pour moi l'occasion de connaître enfin cette Afrique Noire sur laquelle nous avons beaucoup dessiné ; d'avoir le plaisir de retrouver les chercheurs que j'avais rencontrés à Paris, de les voir dans leur travail, aux prises avec des difficultés qu'on imagine mal au Laboratoire. C'est de plus l'occasion de définir une très utile distinction entre deux objectifs fort différents de l'intervention cartographique dans le processus de la connaissance et aussi de redécouvrir une réalité que j'avais certainement délaissée, sous la pression des problèmes de la « civilisation industrielle ».

Après les exposés que nous venons d'entendre, il m'apparaît clairement que le but de la transcription cartographique n'est pas le même suivant que l'on se trouve devant une information insuffisante, ou devant une information pléthorique comme c'est le cas dans les pays dits « développés ».

En effet, lorsqu'on est, comme ici, devant une information réduite à quelques sondages, la cartographie est le moyen d'extrapoler, d'étendre la connaissance au-delà du connu, avec le meilleur coefficient de probabilité. Par contre, lorsqu'on est, comme en France, devant une information surabondante, la cartographie n'est que l'un des moyens de « réduire » cette information, de découvrir les groupements utiles dans une masse d'indicateurs qui dépasse la limite de la mémorisation humaine.

Cette distinction devrait avoir comme premier résultat de préciser dans chaque cas les processus opérationnels appropriés, de donner au chercheur une conscience tranquille de ses problèmes et d'éviter des discussions ayant pour origine une confusion entre deux situations informationnelles différentes.

Dans la plupart des travaux qui ont été présentés, et particulièrement ceux de MM. AVENARD et BONVALLOT en géographie physique, de M. de la SOUCHÈRE en pédologie, de M. HAERINGER en économie agricole et sociale, la totalité de l'information disponible peut être analysée de la manière suivante :

1) Une information très réduite en nombre de phénomènes identifiés, mais étendue à toute la surface géographique. C'est, entre autre, l'information qui est décelable sans ambiguïté sur la couverture photographique aérienne. Appelons-la information *S* (spatiale).

2) Une information étendue en nombre de phénomènes, mais réduite dans l'espace géographique à un petit nombre de points (enquêtes, sondages, transects). C'est l'information *P*.

Quels que soient le domaine et la discipline du chercheur, l'objectif est partout le même. Devant l'impossibilité matérielle de multiplier les enquêtes *P* (impossibilité qui est universelle, car l'étendue du pensable est toujours sans commune mesure avec l'étendue du matériellement possible), la nécessité de connaître, de comprendre et de décider conduit à extrapoler les informations *P* à toute la surface géogra-

phique, en se servant des informations  $S$ . Cette opération est rigoureusement logique puisque toute information  $P$  contient l'information  $S$  pour le point étudié et qu'en tout point non  $P$ , on ne connaît, par définition, rien d'autre que  $S$ . Le problème est de s'assurer du meilleur coefficient de probabilité au cours de cette extrapolation. Cette analyse de l'information et de l'objectif facilite grandement l'élaboration d'un processus de travail.

*Première question : comment accroître  $S$  en l'absence de tout  $P$  ?*

C'est un problème de documentation. La pénurie d'information doit conduire ici à faire feu de tout bois, à utiliser pour un domaine donné les informations recueillies par toutes les autres disciplines, à prendre comme objectif la recherche de coïncidences entre des « ruptures de pentes » climatiques, végétales, géologiques, pédologiques, humaines, ethniques, topographiques, etc. L'interdisciplinaire est de rigueur et la réunion d'Abidjan le prouve amplement. Toute information est bonne à prendre, même si elle paraît incertaine, contrairement à ce que l'on disait autrefois, car les méthodes de corrélation mettront en évidence aussi bien l'erreur que la part de vérité qu'elle contient. Un exemple en est fourni par l'étude de Ph. HAERINGER, qui enquête auprès des praticiens de l'agriculture, parce que leur expérience constitue, dans son domaine, la seule information  $S$  connue. On souligne qu'il a, en cette occasion, suivi la meilleure méthode : donner des cartes à corriger et non pas demander de préciser l'extension de tel phénomène. L'utilisation des anciennes couvertures photo, leur comparaison avec les plus récentes, les nouvelles méthodes électroniques ou optico-chimiques de détection à distance font évidemment partie de cette question. Mais on voit déjà que ces méthodes ne peuvent être complètement exploitées sans recours à l'information  $P$ .

*Deuxième question : où faut-il implanter  $P$  ?*

Evidemment en fonction de ce que l'on sait déjà c'est-à-dire en fonction de  $S$ . Et l'on voit ici tout l'intérêt des travaux de M. de la SOUCHÈRE, qui commence par définir, à partir de  $S$ , des zones relativement homogènes, différentes des zones voisines et à l'intérieur desquelles la probabilité d'extension d'un phénomène donné est logiquement supérieure, en l'absence de toute autre information. On voit, par là même, l'erreur proprement mathématique du sondage aléatoire, qu'il soit spatial ou, plus encore, qu'il ne le soit pas, puisqu'il considère comme égal à zéro un indice de

probabilité dont cent années de sciences géographiques démontrent sans discussion possible qu'il est loin d'être nul ! Un sondage aléatoire n'est défendable qu'à partir du moment (ou de l'échelle) où il n'existe plus aucun indice de probabilité connu pour un phénomène donné.

*Troisième question : comment accroître  $S$  à partir de  $P$  ?*

C'est en réalité l'objectif de toute l'opération.

1) Tout d'abord en allant contrôler en  $P$  des indices visibles en  $S$ , mais dont on ignore la signification. C'est une raison supplémentaire pour choisir  $P$  en fonction de  $S$ . C'est ainsi par exemple que l'on peut découvrir la signification de telle fausse couleur, de telle forme caractéristique, etc. Il faut rappeler ici les premiers essais d'utilisation extensive de la couverture aérienne, expérimentés au Laboratoire par J. LETARTRE sous la direction de G. SAUTTER et développés depuis par J.P. GILG. Celui-ci, par un jeu de vérifications successives entre  $P$  et  $S$ , parvient à déceler en  $S$  le témoignage très probable de plusieurs dizaines de phénomènes de toute nature, y compris des facteurs humains diversifiés.

2) En établissant des corrélations entre toutes les informations  $S$ , avec contrôle en  $P$ . C'est ainsi que la corrélation automatique, point par point, de diverses sortes de détections à distance détermine la présence de telle végétation, de telle culture précise, de telle structure géologique, de tel type de sol, invisibles sur chacun des systèmes de détection pris en particulier, et dont la carte se construit automatiquement.

3) En établissant, à travers tous les points  $P$ , l'indice de corrélation entre toutes les observations  $P$  et toutes les observations  $S$ . On dégage ainsi l'analogie entre certains  $P$  et certains  $S$ , ce qui autorise, en l'absence de toute autre information, l'extrapolation de ces  $P$ . C'est ici qu'interviennent les méthodes mathématiques de corrélation, les méthodes mécano-graphiques (fiches perforées) mais aussi les méthodes graphiques, généralement plus abordables et dont les résultats sont comparables à ceux obtenus par les algorithmes les plus élaborés, telle que l'analyse factorielle des correspondances, de J.P. BENZECRI (A.F.A.CO.). Les méthodes de traitement graphique de l'information ont de plus l'avantage de conserver à tout instant l'information élémentaire dans sa rigueur et dans son intégralité. Sa présence constante relie le traitement logique à la réalité. Elle permet de tenir compte des multiples observations hétérogènes non introduites dans le traitement et d'en enrichir les corrélations découvertes.

On peut résumer la méthodologie graphique à partir d'un exemple théorique très simplifié.

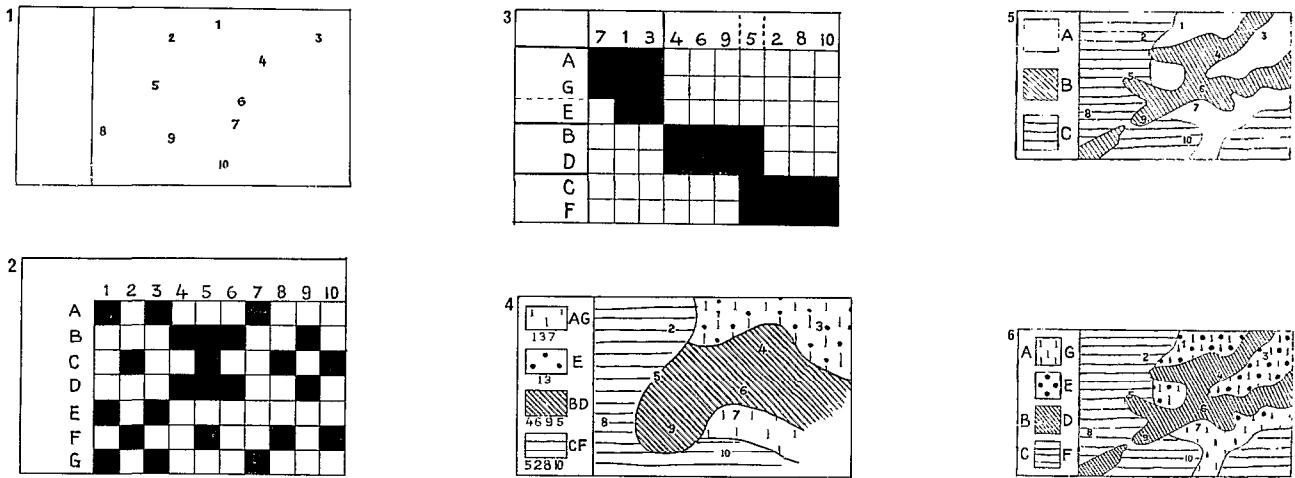


FIG. 1, 2, 3, 4, 5, 6.

#### A - UTILISATION DE LA SEULE INFORMATION P

Soit dix stations  $P$  (1, 2, 3...) implantées géographiquement suivant la carte (1). Pour ces stations on connaît la présence ou l'absence de 7 indicateurs  $A B C D E F G$ . Cette information construit la matrice (2).

La permutation des lignes et des colonnes permet de « resserrer » (d'optimiser) la matrice c'est-à-dire d'obtenir (3) et de découvrir les corrélations contenues dans l'information, à s'avoir

- corrélation entre  $B$  et  $D$ , caractérisant les stations 4, 6, 9 et 5
- corrélation entre  $C$  et  $F$ , caractérisant les stations 5, 2, 8 et 10
- corrélation entre  $A$ ,  $E$  et  $G$ , caractérisant les stations 1, 3 et 7

Si l'on ne se fonde que sur la géographie des points  $P$  (1), la cartographie ne peut fournir que la distribution (4).

#### B - UTILISATION DE L'INFORMATION P+S

Supposons que les indicateurs  $A$ ,  $B$  et  $C$  soient des

informations  $S$  décelées par exemple sur la couverture photographique et qu'elles permettent de construire la carte (5).

L'observation des corrélations visibles sur la matrice (3) autorise la construction de la carte (6) et permet d'affirmer que la distribution de  $G$ ,  $E$ ,  $D$  et  $F$  suivant (6) est infiniment plus probable que la distribution (4).

On notera que la station 5 est à cheval sur deux zones, ce qui apparaît sur la matrice 3 et que l'absence de  $E$  en la station 7 conduit à l'expression d'une probabilité réduite de  $E$  dans le sud de la carte.

Bien sûr, si la réalité était aussi simple, il n'y aurait pas lieu de construire une matrice. Le simple report de l'information sur la carte suffirait à résoudre ce problème. Les matrices se justifient au contraire quand l'information est plus abondante et surtout quand les corrélations ne sont pas aussi nettes. Les matrices offrent alors le moyen d'établir des typologies dont la discussion, appuyée à la fois sur la matrice et la carte, peut faire intervenir des observations complémentaires et constitue la partie proprement scientifique du travail du chercheur.

Manuscrit reçu au SCD le 24 février 1972

## DISCUSSION

M. SAUTTER

Nous savons tous quelles difficultés administratives il faut affronter pour avoir accès au gros matériel de traitement de l'information. D'où l'intérêt de ces techniques qui ne demandent que peu de moyens. Toutefois, les problèmes concernant la valeur de l'information de base restent entiers.

M. BERTIN

Les traitements que je viens de décrire permettent souvent de déceler les défauts de l'information.

M. SAUTTER

Certes, mais les chiffres sont très persuasifs.

M. GOUROU

Pour moi l'intérêt des techniques de M. BERTIN est grand, en raison des possibilités qu'elles offrent pour la recherche des corrélations, dont le géographe est assoiffé.

M. DE LA SOUCHÈRE

Ces techniques sont-elles applicables aux sciences naturelles ?

M. BERTIN

Bien sûr, vous voyez bien que, logiquement, elles s'appliquent à toutes les disciplines à condition que vous puissiez disposer votre information en un *tableau à double entrée*.

M. DUCHEMIN

Je me demande si, lorsque la matrice est très grande, elle reste visuellement analysable.

M. BERTIN

Les limites ne sont pas visuelles, elles sont mécaniques. Il est difficile actuellement de construire des matrices supérieures à  $100 \times 200$  environ. Au delà, il faut procéder à des réductions de type mathématique (classements automatiques, typologies, analyses factorielles) qui permettent de ramener une information trop longue à des dimensions acceptables par le matériel. Il est vrai que le temps passé pour réduire une matrice est fonction des dimensions de cette matrice. Mais c'est parce que celle-ci donne au chercheur l'occasion de réfléchir, de définir ses options, d'utiliser ses connaissances et de faire œuvre d'imagination au *cours de la réduction* et non *après la réduction*, comme c'est le cas avec l'utilisation des routines mathématiques.

M. DUCHEMIN

Ce qui me gêne justement, c'est que le choix des classes, des familles, est à faire par le chercheur. Ne pourrait-il pas être automatique ?

M. BERTIN

Méfiez-vous de l'automatisme. Le classement peut être automatique. Pas l'appréhension de ce classement. On ne pourra jamais automatiser l'opération de M. HAERINGER.

Ph. H.