CAD

Stabilité structurale, théorie des catastrophes et analyse qualitative

Les cahiers de l'analyse des données, tome 3, nº 2 (1978), p. 235-238

<http://www.numdam.org/item?id=CAD_1978__3_2_235_0>

Les Cahiers de l'Analyse des Données
Vol. III - 1978 - n° 2 - p. 235-238

MÉMOIRE REÇU

Stabilité structurale, théorie des catastrophes et analyse qualitative

[CATASTROPHES]

d'après un exposé de R. Thom
Analyse par J.-P. Benzécri (1)

S'adressant en Juin 1976 à une conférence organisée par la S.I.A.M. *(Society for Ind. & Appl. Math.)*, René Thom a choisi pour thème la stabilité structurale et la théorie des catastrophes. Son exposé (publié in *S.I.A.M. Rev.* Vol. 19 pp. 189-201), comporte une section *"The qualitative analysis"* qui intéresse l'analyse des données, et sur laquelle deux statisticiens travaillant à l'université de Cincinnati - Madame Hoang Thu et Monsieur Pham Son - ont eu l'obligeance d'attirer notre attention.

Bien que diffusés dans une collection destinée au grand public *(Modèles mathématiques de la morphogénèse*, a paru dans la collection 10/18 de l'Union générale d'éditions ; Paris 1974) les travaux de R. Thom sur la morphogénèse, se laissent difficilement pénétrer sans une connaissance approfondie des applications différentiables entre espaces multidimensionnels ; connaissance qui manque ordinairement au statisticien . Fort heureusement, le § de R. Thom consacré à l'analyse qualitative est d'une lecture assez facile ; et nous en rendrons compte ici et tenterons ensuite de le situer dans cette "théorie des catastrophes" que l'honnête homme de la fin du XX° siècle est avide de comprendre.

A la question *"What finally is statistics about?"* R. Thom répond:"Si nous soumettons un système à une série d'expériences, nous aboutissons à un nuage de points dans l'espace euclidien V des observables. La statistique n'est rien d'autre que l'interprétation d'un nuage de points". Le problème étant ainsi posé, R. Thom affirme ensuite que "d'ordinaire, la statistique oscille entre deux points de vue : ou bien le nuage donné devrait se concentrer le long d'une sous-variété de V, décrite par des lois quantitatives régissant les observables ; ou bien le nuage devrait être concentré en son centre, et seul le bruit explique les écarts par rapport au centre". Bien que R. T. n'entre pas dans les détails, on reconnaît dans le premier point de vue l'analyse des données, et dans le second la statistique paramétrique. Cependant pour R. T. l'approche la plus naturelle est autre, et il s'étonne que "la statistique semble ne pas concevoir ainsi sa tâche". Pour "interpréter le nuage donné dans V, il faut construire un espace M de paramètres cachés puis dans le produit $M \times V$, former un système dynamique déterministe (ergodique) Γ tel que la projection sur V d'une distribution invariante équibrobable sur $\Gamma \subset M \times V$ engendre la distribution des points du [nuage] sur V".

Nous avons dit ailleurs ([Principes] TII A n° 1) que l'ergodicité est le fondement de l'application des probabilités en statistique : en bref c'est parce qu'un système réalise dans des conditions stables toutes les combinaions possibles, que l'on peut attribuer à celles-ci des probabilités, comme aux positions d'un point qui au cours du temps enroule uniformément sa trajectoire sur une variété Γ. En tant que modèle différentiable, la suggestion de R. T., nous convient donc. Cependant on doit se demander si ce modèle offre plus qu'une justification de fond à la statistique ; s'il peut aussi dans un cas déterminé conduire à un déchiffrage de la structure ? Nous y reviendrons, mais considérons d'abord

(1) *Professeur de statistique, université Pierre et Marie Curie ; Paris.*

sur un exemple comment R. T. suggère de découvrir des paramètres cachés (coordonnées non-mesurées de M) qui avec les variables observables (coordonnées de V) engendrent un espace où le nuage des données soit intelligible, avec sa perspective propre, et non écrasé, privé de son relief naturel.

Voici tel qu'il apparaît figuré sur son mémoire américain, l'exemple de R. Thom. Nous avons seulement ajouté à la main des indications d'axes.

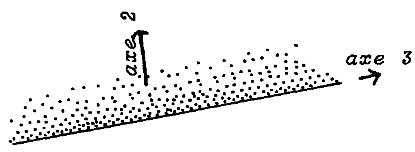
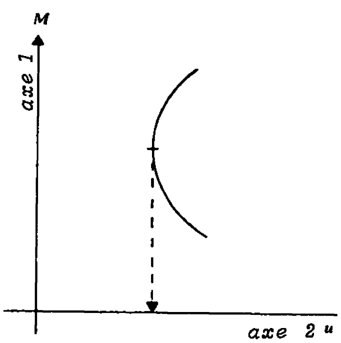


FIG. 2A.



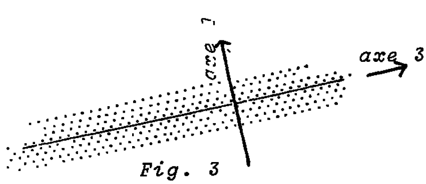


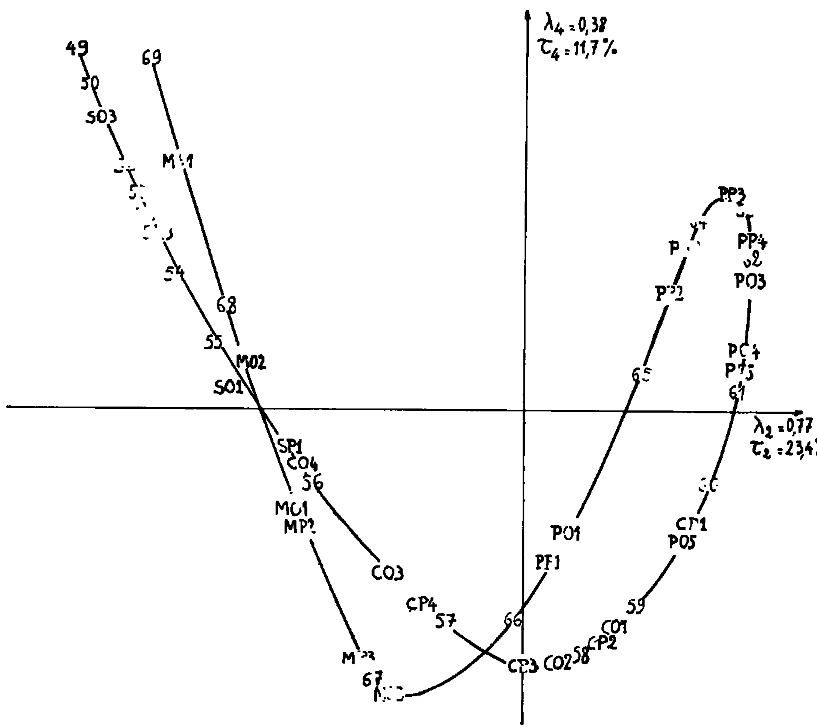
Fig. 3

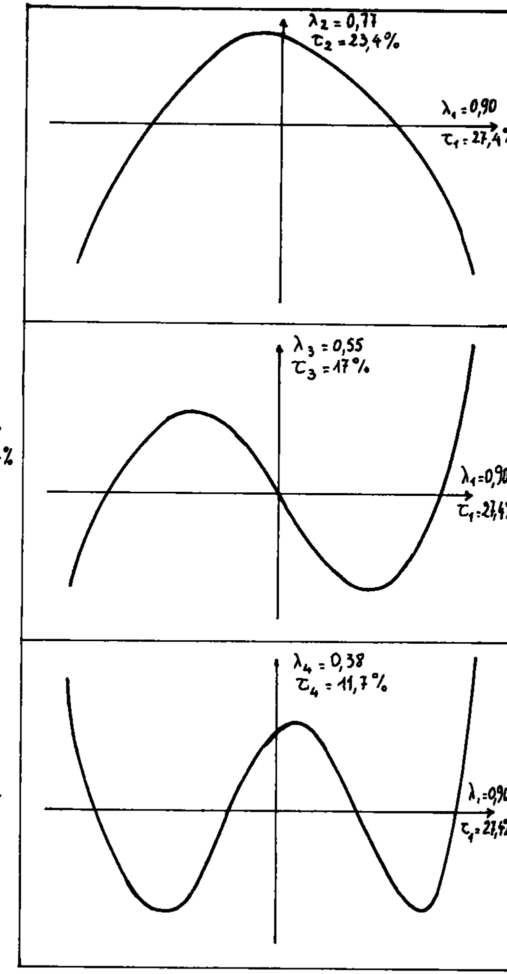
R. T. écrit "Qu'arrive-t'il si le nuage présente une morphologie tranchée avec par exemple des frontières, des coins ou des points triples? La statistique est impuissante devant de telles situations, et le spécialiste doit inventer des explications *ad hoc*... . La théorie des catastrophes qui connaît l'aspect générique en projection des valeurs critiques , peut interpréter ces accidents comme des singularités de la projection de M x V [l'espace enrichi par des paramètres cachés] sur V [l'espace des seules variables observables]". "Par exemple, si dans un plan $\mathbb{R}^2$ un nuage de points montre une frontière tranchée ; avec accroissement de la densité vers la frontière (fig 2A) au lieu d'un étalement bilatéral(fig3) il sera plus naturel de définir le nuage comme la projection d'un pli. .." (fig 2B et fig 3).

Or la situation qu'imagine R.Thom est assez classique en analyse des données : elle correspond à un effet Guttman net entre les facteurs 1 et 2 (F_2 fonction quadratique de F_1) avec un troisième facteur indépendant des deux premiers. C'est l'interprétation que suggèrent les indications d'axes ajoutées par nous. Lorsque l'effet Guttman est très fort mettant en jeu e.g. quatre facteurs, comme cela arrive parfois, on peut voir apparaître des figures encore plus énigmatiques, telles que ce plan 2 x 4 extrait de la thèse de A. Akiki.

Toutefois l'analyse factorielle fournit elle-même la clef, sans qu'il soit besoin de deviner des paramètres cachés. En statistique multidimensionnelle en effet, pourvu que les données soient assez riches toutes les propriétés fondamentales de l'objet se révèlent comme des dimensions du nuage ; la tâche de l'interprétation étant de formuler ces propriétés fondamentales dans les termes des variables observées, principalement en s'aidant de la représentation simultanée des deux ensembles en correspondance.

Mais revenons aux conceptions de Thom sur la morphogénèse. En tête de son § "*The quantitative analysis*", R. T. rappelle que la théorie des catastrophes (entendez : changements brusques, sauts qualitatifs) prétend énoncer des propositions telles que "si dans l'intervalle de temps (t_0 , t_1), le système a fait voir une morphologie (M_0^1), alors on doit s'attendre à ce qu'ultérieurement, dans l'intervalle (t_1 , t_2) il présente une certaine morphologie (M_1^2)". Pour parvenir à une telle conclusion, R. T. (in *Modèles...*, 1974, p. 24) "attribue l'apparition de la structure à une hypothèse de *généricité* : en toute circonstance la nature réalise la morphologie locale la moins complexe compatible avec les données initiales locales". Ce que nous commenterons





Analyse par A. Akiki du recensement scolaire au Liban en 1973 : I, ensemble d'années de naissance ; J, ensemble des types de classes ; et, e.g. k(67, MP2) : nombre des enfants nés en 1967 et fréquentant, en 1973 une classe maternelle de 2e année d'une école privée. Ci-dessus, dans le plan 2x4 une courbe singulière ; à droite, en réduction, les plans 1x2, 1x3 et 1x4 montrant la relation fonctionnelle étroite qui lie au premier facteur les facteurs de rang supérieur.

ainsi : est *générique* un caractère morphologique qu'une petite perturbation est impuissante à faire disparaître, sinon en le compliquant. Un exemple suggérera que derrière ce langage qualitatif on peut placer des énoncés mathématiques . Soit la fonction $y = x^2$, considérée au voisinage de l'origine $x = 0$; cette fonction présente en 0 un minimum, et toute fonction $f(x)$ voisine de x^2, présentera également un minimum en un point voisin de 0 : c'est là une propriété structurelle stable. Mais la fonction $y = x^3$ présente en 0 un point où $y' = y'' = 0$, singularité que des fonctions arbitrairement voisines ne présentent pas. Cependant la famille de fonction $y = x^3 + u\,x$, considérée dans le voisinage de $x = 0$, et pour des valeurs petites du paramètre u, comprend la fonction $y = x^3$; et de quelque manière que l'on perturbe cette famille, la famille $y = f(x,u)$ obtenue comprendra une fonction $y = f(x,u_o)$ (u_o convenable) qui en un certain point x_o soit telle que $f'_x(x_o, u_o) = f''_{x^2}(x_o, u_o) = 0$. On voit qu'une même propriété (avoir en un point $y' = y'' = 0$) sera stable ou non selon le contexte dans le - quel on la situe : les contraintes dimensionnelles : nombre de variables, nombre de paramètres jouent un rôle essentiel.

Pour qui a examiné, même sans compétence biologique, des schémas d'embryologie, la pertinence des conceptions de Thom apparaît à l'évidence : parvenu à un point critique de son développement, tel tube-à moins de subir une compression invraisemblable en restant sur place - doit se nouer en basculant (solution générique) ; et l'on conçoit qu'une suite de catastrophes analogues engendrent une forme déterminée avec la quasi-certitude d'un programme... On doutera toutefois que ce modèle soit universel. R. Thom écrit (*Modèles...*, 1974, p. 8) : "L'espace sur lequel apparaît une morphologie sera dit l'espace substrat... .Cet espace sera pratiquement toujours un ouvert d'un espace euclidien, comme un domaine de l'espace-temps usuel". Bien que R. T. reconnaisse l'importance d'espaces de dimension infinie, il semble que la contrainte dimensionnelle (dimension finie) soit requise pour la validité des théories du moins telles qu'il les a présentées jusqu'ici. Or singulièrement dans les sciences humaines, il n'est pas généralement possible d'assigner au processus qu'on étudie un espace substrat déterminé : les dimensions mentales sont plutôt potentiellement infinies, et certes non fixées *a priori*.

Reconnaissant dans le modèle de la mécanique des gaz "l'un des paradigmes les plus parfaits de l'explication scientifique" R. Thom (in *Modèles...*, 1974, pp. 22-23) y distingue "deux algorithmes insuffisamment dissociés dans l'esprit du savant : l'emploi du calcul différentiel comme paradigme de processus déterministique, et l'usage d'atomes en interaction". Il affirme que l'emploi du calcul différentiel est quoiqu'on dise, d'une validité beaucoup plus étendue que le recours aux atomes. Bien que le choix des unités (en biologie : cellule ou molécule ? demande R. T.) ne soit pas toujours aisé ; nous estimons au contraire que le modèle certes nébuleux des atomes en conflit, peut valoir dans des disciplines où le calcul différentiel n'a pas de prise. Là même où le modèle différentiel s'applique parfaitement, il est frappant que les véritables catastrophes (destructrice, et non génératrice de structure) soient livrées aux fluctuations locales : un ballon qu'on gonfle croît régulièrement ; la ligne suivant laquelle il se déchire est aléatoire (même si cette ligne a des caractéristiques statistiques stables...). A plus forte raison en doit-il être ainsi des crises de l'homme.

Sera-t-il jamais possible de découvrir après ces analogies une science qui calcule et prévoit. Il serait oiseux d'en débattre ici. Pour l'heure dans de nombreux domaines, on n'a pas même reconnu où sont les principales distances ou les similitudes. On ne soupçonne pas quelles sont les dimensions de l'*espace substrat* où pourrait s'inscrire une théorie quelle qu'elle soit. A découvrir ces dimensions s'emploie l'analyse des données.