

CAD

**Mémoire reçu : COMPCLUS, a FORTRAN  
programm for rapid initial clustering  
of large data sets**

*Les cahiers de l'analyse des données*, tome 4, n° 3 (1979),  
p. 379

[http://www.numdam.org/item?id=CAD\\_1979\\_\\_4\\_3\\_379\\_0](http://www.numdam.org/item?id=CAD_1979__4_3_379_0)

© Les cahiers de l'analyse des données, Dunod, 1979, tous droits réservés.  
L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme  
Numérisation de documents anciens mathématiques  
<http://www.numdam.org/>

## MÉMOIRE REÇU

*Mémoire reçu : COMPCLUS, a FORTRAN program for rapid initial clustering of large data sets ; Hugh G. Gaudr Jr. ; Ecology and Systematics, Cornell University, Ithaca, N.Y. 14850 ; Fev. 1979 (60 p.).*

Comme en témoigne le titre de son mémoire, l'auteur pense comme nous, que les programmes d'agrégation ont pour rôle majeur de préparer des analyses ultérieures (phase initiale d'une étude). Les données visées sont des tableaux de relevés écologiques croisant un ensemble I de relevés avec un ensemble J d'espèces ; le but étant (comme dans le travail de Briane, Lazare et Salanon, auquel H.G.G. fait référence : cf *Cahiers* Vol III n° 2 ; 1978) d'agréger en un nombre restreint de classes l'ensemble I des relevés.

Quant à la méthode, il est particulièrement opportun que nous puissions en rendre compte dans ce même cahier où sont publiés les travaux de G. Flamenbaum. En effet, l'un et l'autre auteur imposent aux classes une contrainte de rayon. Mais tandis que G.F. a recours au principe de E. Diday - agrégation autour de centres variables - pour aboutir itérativement à des classes satisfaisantes, H.G.G. procède en deux étapes successives, ainsi que nous l'expliquons brièvement ci-dessous d'après le mémoire (p. 11).

*Etape 1* : Dans la première phase on tire au sort un premier individu  $i$  auxquels sont agrégés tous ceux  $i'$ , situés à une distance de  $i$  inférieure à une distance d'un seuil  $R$  fixé ; et chaque fois qu'il est nécessaire, on introduit de nouveaux centres. Comme il peut apparaître ainsi des classes de très faible effectif, on s'applique à éliminer celles-ci dans une deuxième étape.

*Etape 2* : Un individu appartenant à une classe de faible effectif (ici encore il faut fixer un seuil) est rattaché à la base d'effectif élevé la plus proche, pourvu que sa distance au centre de celle-ci soit inférieur à  $kR$  ; le rapport  $k$  devant être supérieur à 1, mais en être proche, ainsi qu'il ressort de l'expérience acquise par l'auteur.

L'un des mérites du programme est qu'en publiant des histogrammes d'effectif des classes etc., il permet à l'utilisateur d'aboutir en quelques essais à des valeurs satisfaisantes de divers seuils.