

## **État des recherches sur la classification ascendante hiérarchique exposées dans ce cahier**

*Les cahiers de l'analyse des données*, tome 7, n° 2 (1982),  
p. 162

[http://www.numdam.org/item?id=CAD\\_1982\\_\\_7\\_2\\_162\\_0](http://www.numdam.org/item?id=CAD_1982__7_2_162_0)

© Les cahiers de l'analyse des données, Dunod, 1982, 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/>

Etat des recherches sur la classification  
ascendante hiérarchique exposées dans ce cahier

[ETAT C.A.H.]

L'article [BORNES NIV. C.A.H.] ainsi que les sept articles qui le suivent sont consacrés à la classification ascendante hiérarchique. Il nous paraît bon de profiter d'une page blanche pour préciser les rapports entre ces huit articles.

[BORNES NIV. C.A.H.] concerne le choix d'un seuil dans l'algorithme des voisinages réductibles (cf. M. Bruynooghe) ; algorithme éventuellement susceptible de perfectionnements, mais qui, dans son état actuel est nettement plus lent que les algorithmes programmés par J. Juan dans [PROG. C.A.H. RECIP.] et [PROG. C.A.H. CHAINE RECIP.].

[C.A.H. DIST. TRI] suggère une idée nouvelle qui paraît pouvoir accélérer grandement la C.A.H. ; et donne sans programme de calcul les premiers résultats expérimentaux.

[C.A.H. FLUX] offre un nouveau critère d'agrégation : ce critère peut servir dans l'algorithme de base de la C.A.H., mais aussi dans les algorithmes accélérés d'agrégation des voisins réciproques, éventuellement en chaîne : il est en particulier immédiatement utilisable avec le programme ALGOL donné dans [C.A.H. CHAINE RECIP.].

[PROG. C.A.H. RECIP.] en utilisant le principe des voisins réciproques (dont C. de Rham nous a montré l'intérêt) ce programme est actuellement, à notre connaissance, celui qui réalise les meilleures performances moyennes en agrégation suivant la variance.

[PROP. RECIP.] on voit que sur un cas modèle (qui semble typique) [PROG. C.A.H. RECIP.] a un temps d'exécution de l'ordre de  $n^2$  (n étant le nombre d'individus).

[CHAINE COMP. C.A.H.] mais on peut construire des exemples où [PROG. C.A.H. RECIP.] a, comme l'algorithme de BASE, un temps de l'ordre de  $n^3$  ; d'où l'intérêt des algorithmes en "CHAINE".

[C.A.H. CHAINE RECIP.] en recherchant individuellement les paires de voisins réciproques, on a un temps de calcul toujours en  $n^2$ , quelles que soient les données. L'algorithme donné ici, opérant sur un tableau de distances, est apte à traiter tous les critères usuels, notamment [C.A.H. FLUX].

[PROG. C.A.H. CHAINE RECIP.] applique le principe de la recherche en CHAINE avec le critère d'agrégation suivant la variance ; et opère sur un tableau de coordonnées factorielles (ou de mesures), ou un tableau de correspondance. Il est en moyenne un peu moins rapide que [PROG. C.A.H. RECIP.], mais est toujours en  $n^2$  (cf. [CHAINE COMP. C.A.H.]).