

# Carctérisation des accessions de safoutiers.

Par MBOGNING TCHINDA Cyprien

Master de Statistique Appliquée  
ENSP - Université de Yaoundé 1

22 Octobre 2007

- 1 Introduction.
- 2 Données.
- 3 Analyse multidimensionnelle des données.
- 4 Présentation des méthodes utilisées.
- 5 Application des méthodes et résultats.
- 6 Conclusion.

## Contexte

Le 2ème séminaire international sur la valorisation du safoutier et autres oléagineux non conventionnels effectué à l'ENSAI de Ngaoundéré rapporte que le développement et la promotion de la culture du safoutier se heurte à de nombreuses contraintes, à savoir

## Contexte

- l'insuffisance des connaissances scientifiques et techniques en ce qui concerne l'amélioration génétique, l'agronomie, la lutte contre les maladies, les technologies post-récoltes ;
- le caractère très périssable du fruit, l'inexistence des technologies de conservation et de transformation, constituent des entraves majeures à la commercialisation ;
- l'absence d'informations sur le marché (quantités, prix), lacune aggravée par la non disponibilité de données statistiques fiables.

Les perspectives de développement de la filière safou s'articulent en priorité autour de la recherche de solutions aux contraintes sus-mentionnées.

## Problématique

L'une des étapes dans la recherche sur l'amélioration d'une espèce végétale est la caractérisation qui a pour but la valorisation des différents écotypes de cette espèce. Le safoutier, arbre originaire des zones tropicales humides et sub-humides d'Afrique, dont le fruit est le safou, d'une grande importance aussi bien économique qu'alimentaire, n'échappe pas à cette règle. C'est conscient de cette lacune que l'IRAD (Institut de Recherche Agricole pour le Développement) au Cameroun a procédé à une collection non ciblée d'accessions de safoutiers (ceci pour une conservation de la diversité des génotypes dans un endroit sécurisé) pour une caractérisation botanique de celles-ci ; ce qui permettra à long terme d'avoir de meilleures connaissances sur les différents écotypes.

# Problématique

## Objectif

Dégager les éventuels groupes relativement similaires et les caractériser à l'aide de variables pertinentes.

# Enjeu

Le travail que nous réalisons permettra aux chercheurs dans la filière safou d'avoir de meilleures connaissances sur les différents écotypes et surtout d'effectuer des sélections variétales ; ce qui n'était pas fréquent jusqu'à nos jours. Ce travail servira à améliorer l'espèce *dacryodes edulis*.

## Origine des données

Les données sont issues de la banque des gènes de Barombi-Kang dans la province du Sud-Ouest au Cameroun.

# Résultats de l'ACP réalisée sur les variables quantitatives

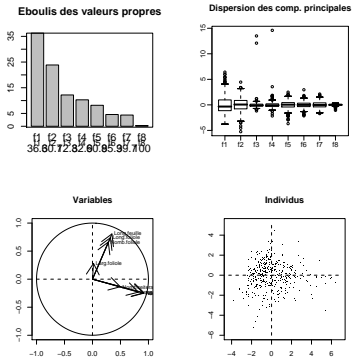


Figure: Récapitulatif de l'ACP

## Résultats de l'AFCM réalisée sur les variables qualitatives

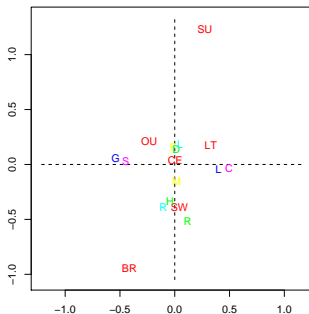


Figure: AFM du tableau de Burt

## AFCM

Méthode d'analyse multidimensionnelle nécessitant des variables catégorielles et permettant d'identifier les éventuels groupes de variables

## Méthode des k-centroïdes

C'est une méthode de classification automatique permettant de regrouper les individus en classes tout en minimisant la variance intra-classe. L'algorithme mettant en oeuvre cette méthode non-implémentée sous le logiciel R est le suivant

- Tirer aléatoirement  $k$  individus de l'échantillon; ils constituent les centres des  $k$  premières classes.
- Pour chaque individu de l'échantillon, calculer les distances entre l'individu considéré et tous les centres et allouer l'individu  $N^{\circ}i$  à la classe dont il est le plus proche du centre.
- Chaque fois qu'un individu est alloué dans une classe le nouveau centre doit être calculé (cest l'individu le plus proche des autres dans la classe considérée).
- Répéter ce processus jusqu'à la convergence (en effet, la variance intra-classe décroît avec les itérations et fini par se stationnariser.)
- Algorithme sous optimal car est lié à l'initialisation des centres donc nécessité de l'exécuter plusieurs fois avec des centres distincts et retenir celui qui fournit une variance intra-classe minimale.

Plan de l'exposé

Introduction

Données

Analyse multidimensionnelle

Méthodes utilisées pour la caractérisation

Application des méthodes au jeu de données sur le safou

Conclusion

Application de la méthode des  $k$ -centroïdes

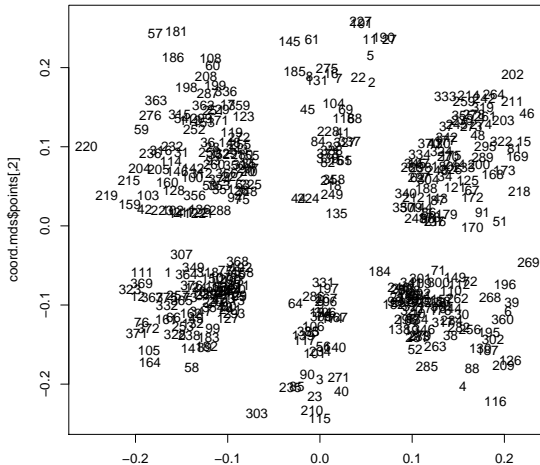
## Application de l'AFCM.

Les groupes obtenus sont contenus dans le tableau (TAB.1)

Groupes	Composition des groupes
Groupe1	<b>LI1</b> : Long.inflore $\leq 17$
	<b>NFL1</b> : Nomb.fleurs $\leq 63$
	<b>nfl1</b> : Nfleurs.inflo $\leq 62$
	<b>F</b> : Sexe féminin
	<b>NR1</b> : Nomb.reiterations $\leq 1$
Groupe2	<b>LF3</b> : Long.feuille $> 50.7$
	<b>LAFO2</b> : Larg.foliole $> 6.02$
	<b>NFO3</b> : Nomb.foliole $> 16$
	<b>LFO3</b> : Long.foliole $> 18.8$
	<b>S</b> : Couleur vert-sombre
	<b>G</b> : Aspect.limbe gaufré

Groupe3	<b>C</b> : Couleur vert-clair
	<b>L</b> : Aspect.limbe lisse
	<b>LAFO1</b> : Larg.foliole $\leq 6.02$
	<b>NFO1</b> : Nomb.foliole $\leq 14$
	<b>LFO1</b> : Long.foliole $\leq 16.8$
Groupe4	<b>LF1</b> : Long.feuille $\leq 43.8$
	<b>NR2</b> : Nomb.reiterations $> 1$
	<b>NFL3</b> : Nomb.fleurs $> 127$
	<b>nfl3</b> : Nfleurs.inflo $> 118$
	<b>LI3</b> : Long.inflo $> 25$

L'application de la méthode nécessitant un nombre de classes fixé à l'avance, nous l'avons estimé à l'aide d'un positionnement multidimensionnel. Ce dernier est illustré sur la figure (FIG.4).



## Classes obtenues

On voit d'après la figure (FIG.4) que les individus se répartissent en 6 groupes. Les classes obtenues sont contenues dans le tableau (TAB.2)

Classes	Individus
G1	2 17 18 27 37 43 48 65 68 81 82 104 108 118 167 171 181 191 194 195 198 199 202 203 213 215 216 218 223 228 230 236 240 266 269 278 297 317 319 327 334 336 337 338 339 342 345 350 351 352 353 355 356 358 363 364 365 366 367 369 373 374 375 376
G2	5 7 8 9 11 20 22 24 29 40 51 53 54 62 63 70 73 75 78 84 87 93 112 117 121 126 130 131 132 133 134 138 139 148 159 160 161 165 166 168 170 173 178 179 180 185 200 205 209 211 214 217 219 220 227 231 244 245 246 248 258 259 260 267 272 276 286 287 288 290 303 305 306 307 312 314 315 318 322 323 326 328 348 357 370

G3	1 3 12 13 14 15 16 21 49 55 56 57 58 60 67 89 92 94 95 96 99 100 103 107 123 124 127 128 129 140 142 150 154 158 162 163 164 174 182 183 184 189 192 193 197 201 221 222 229 232 234 235 237 238 239 241 250 253 255 275 309 316 332 333 346
G4	23 25 30 31 39 52 59 61 64 66 72 74 76 88 105 110 111 115 116 187 226 254 256 257 261 262 263 264 265 268 271 274 277 279 281 282 284 285 292 296 301 302 308 310 313 320 321 330 335 347 360 371 372 377
G5	26 28 32 33 35 36 41 42 44 45 47 50 69 77 80 102 113 114 119 122 136 141 143 145 146 151 153 155 172 186 190 204 206 207 208 212 224 225 243 247 249 251 252 280 289 293 294 299

G6	4 6 10 19 34 38 46 71 79 83 85 86 90 91 97 98
	101 106 109 120 125 135 137 144 147 149 152
	156 157 169 175 176 177 188 196 210 233 242
	270 273 283 291 295 298 300 324 325 331 340
	341 343 344 349 354 361 368

**Table:** Répartition des individus dans les classes

# Discrimination des classes

Les variables discriminantes des classes sont contenues dans le tableau (TAB.3).

# Discrimination des classes

Classes	Variables caractéristiques
G1	$9 \leq \text{Long.inflo} \leq 39$ $0 \leq \text{Nomb.reiterations} \leq 9$ $21 \leq \text{Nomb.fleurs} \leq 240$ $19 \leq \text{Nfleurs.inflo} \leq 177$ Sexe (M) Mode.de.ramification (D)
G2	$0 \leq \text{Nomb.reiterations} \leq 9$ $15 \leq \text{Nomb.fleurs} \leq 192$ $10 \leq \text{Nfleurs.inflo} \leq 177$ Aspect.ecorce (L) Sexe (M)

G3	$24 \leq \text{Nomb.fleurs} \leq 219$ $20 \leq \text{Nfleurs.inflo} \leq 192$ Aspect.ecorce (L) Aspect.limbe (L) Couleur (S)
G4	$9 \leq \text{Long.inflore} \leq 32$ $0 \leq \text{Nomb.reiterations} \leq 45$ $21 \leq \text{Nomb.fleurs} \leq 108$ $4.1 \leq \text{Larg.foliole} \leq 7.7$ Sexe (F)

G5	$9 \leq Long.inflor \leq 3$ $0 \leq Nomb.reiterations \leq 8$ $48 \leq Nomb.fleurs \leq 146$ Mode.de.ramification (D), Aspect.ecorce (L) Aspect.limbe (L)
G6	$10 \leq Long.inflor \leq 29$ $4.1 \leq Larg.foliole \leq 7.8$ Mode.de.ramification (D) Aspect.limbe (G) Couleur (S)

**Table:** Tableau représentant les classes et les variables permettant de les expliquer.

Le but de notre étude était de caractériser les safoutiers en utilisant les données collectées sur les safoutiers par l'IRAD dans la banque des gènes de Barombi-Kang.

Pour ce faire, nous avons utilisé 2 méthodes : Une méthode d'analyse multidimensionnelle (AFCM) et une méthode de classification automatique ( $k$ -centroïdes). La deuxième a l'avantage qu'elle sort du cadre d'une analyse descriptive contrairement à la première.

Un résultat intermédiaire obtenu des 2 méthodes est que la zone écologique de provenance des safoutiers n'a pas d'influence sur leur comportement. Ce phénomène est peut-être dû aux facteurs externes. Une caractérisation dans les zones écologiques de provenance donnerait peut-être des résultats différents.