

Contribution à l'étude de l'échantillonnage des agrumes

II - Problème de la stratification d'un clémentinier

S. STRIZYK, P. LOSSOIS et J.-P. GAILLARD*

CONTRIBUTION A L'ETUDE DE L'ECHANTILLONNAGE
DES AGRUMES

II.- PROBLEME DE LA STRATIFICATION D'UN CLEMENTINIER

S. STRIZYK, P. LOSSOIS et J.P. GAILLARD

Fruits, mars 1976, vol. 31, n°3, p. 177-215.

RESUME - Prévoir une récolte de clémentines, en suivre l'évolution qualitative dans le temps, en effectuer une description pomologique, nécessite le prélèvement d'échantillons pour effectuer les observations correspondantes bien avant la cueillette totale. Certaines méthodes, comme le Frame count et le Limb count, en usage dans certains pays agrumicoles, ne sont pas sans présenter d'inconvénients. La technique la meilleure pour connaître la production d'un clémentinier serait de cueillir tous les fruits d'un arbre, d'en mesurer les paramètres de poids, d'extrait sec, d'acidité... Cette technique étant destructive et longue, force est de recourir à l'échantillonnage dont on essayera de définir les caractéristiques. Après avoir décrit, par la technique d'analyse en

composantes principales, la population pour toutes les variables quantitatives sur l'arbre et sur les clémentines, nous nous sommes attachés à définir les tendances générales de la nature à répartir certaines variables de production (poids, extrait sec, acidité) selon les différents critères de positions géographiques (hauteur, distance, orientation), par la technique de l'analyse des correspondances.

Il apparaît que la hauteur, la distance, l'orientation, jouent un rôle dans la détermination du poids, de l'extrait sec et de l'acidité. Leurs différentes combinaisons permettent de cerner des régions homogènes de l'arbre aux fruits à caractéristiques différentes de poids, d'extrait sec et d'acidité. Ces dernières variables, dites de production, étant liées entre elles, l'extrait sec et l'acidité s'avèrent les meilleures pour déterminer le groupe de poids auquel appartient une clémentine (démonstration par la technique d'analyse discriminante). L'opération qui consiste à répartir la population des fruits d'un arbre en groupes de mêmes caractéristiques s'appelle la stratification ; c'est le prélude à la détermination d'un échantillon.

Cette publication fait suite à un premier article : « Monographie de la récolte d'un clémentinier » paru dans *FRUITS*, vol. 31, n°1, 1976.

Le lecteur est informé que cette étude générale est une approche de l'étude de l'échantillonnage spécifique au clémentinier en Corse. L'importance et la diversité des problèmes posés ont nécessité de découper l'exploitation des résultats d'observation en trois parties, dont nous développerons ici celle ayant trait à la détermination des caractères de stratification par les techniques d'analyse multidimensionnelle.

Afin de faire la liaison entre la description des variables prises en compte pour positionner chaque fruit sur l'arbre et l'analyse des données proprement dite, on rappellera succinctement les conclusions essentielles du premier article pouvant servir de support à la suite de nos travaux et, par là même, fournir des éléments de base à la création d'un échantillon.

Sur un clémentinier pris au hasard dans une parcelle expérimentale de la Station de Recherche agronomique de Corse, on a décrit la récolte totale de l'arbre en retenant pour chaque clémentine un certain nombre de paramètres qualitatifs et quantitatifs susceptibles d'expliquer cette récolte.

A ce stade de l'étude descriptive, on a pu dégager à partir de tableaux à double entrée, les conclusions partielles suivantes :

- L'orientation des fruits semble agir sur leur poids, leur couleur, leur acidité, leur extrait sec.

- La hauteur des fruits au-dessus du sol et leur distance à l'axe vertical de l'arbre ont une incidence sur ces mêmes caractères.

Des recoupements ont permis d'imaginer une certaine partition de l'arbre en couples hauteur x distance. Partant de cette représentation schématique on a pu situer une zone où vraisemblablement (compte tenu de la distribution de chaque variable) la prise d'échantillon sera, non seulement facile à réaliser, mais représentative de la récolte. Toutefois, il faudra préciser en fin de cette étude la taille de l'échantillon et les modalités exactes de sa constitution.

Cette première étude paraît cependant insuffisante pour

* - S. STRIZYK, Association de Coordination technique agricole,
149, rue de Bercy, 75579 PARIS Cedex 12
P. LOSSOIS et J.P. GAILLARD, IRFA, 6, rue du Général Clergerie,
75116 PARIS.

prouver que telle ou telle zone de l'arbre convient mieux qu'une autre ; d'une part, parce que nous avons toujours travaillé dans un espace à deux dimensions, ignorant de ce fait les interactions multiples entre les différentes variables et, d'autre part, parce que nous avons travaillé sur un seul arbre sans avoir pu vérifier nos premières conclusions. Les deux articles faisant suite à cette première communication, et dont nous développerons ci-après la seconde, s'efforceront de répondre aux problèmes soulevés.

La recherche entreprise étant de nature méthodologique et exploratoire pourra servir à une certaine conception de l'échantillonnage.

Nous devons avoir en permanence à l'esprit le souci d'efficacité pratique et, par conséquent, veiller à la souplesse de la méthode et porter attention à l'ensemble du contexte technique, économique et humain.

Les domaines, caractère de l'étude et choix de la population sur laquelle nous effectuons notre recherche étant précisés, il nous faut répondre à certaines questions afin de définir le problème.

DEFINITION ET CONCEPTION DE L'ÉCHANTILLONNAGE

Pourquoi effectuer un échantillonnage ?

La technique la plus absolue pour connaître la production d'un clémentinier et par extension, d'une parcelle ou d'un verger, est de peser toute la récolte.

Dans le même esprit, si l'on veut exprimer les critères définissant la qualité, il faut analyser tous les fruits.

Outre qu'un tel travail est souvent destructif, il n'est pas réalisable chaque fois qu'on désire des informations sur un verger, sur une parcelle expérimentale ou sur un arbre, vu le nombre très élevé d'individus constituant la population parente.

Pour ces raisons, il est nécessaire d'effectuer un échantillonnage, c'est-à-dire d'élaborer une méthode permettant de constituer un ou des échantillons qui, sous certaines contraintes, devront être représentatifs de la population dont ils sont issus.

Si l'on parvenait à démontrer que quelques fruits suffisent à décrire l'aspect qualitatif et quantitatif de toute une récolte, on aurait alors apporté une contribution essentielle à la connaissance de celle-ci.

L'intérêt immédiat en serait double :

a) techniquement :

- la facilité de travail sur quelques fruits permet la répétition d'un échantillonnage dans le temps pour suivre l'évolution d'une récolte

- l'échantillonnage permet également des études comparatives en expérimentation agronomique.

b) économiquement :

- l'échantillonnage doit permettre de faire des prévisions de récolte et de ce fait, donner la possibilité d'organiser la commercialisation.

Qu'est-ce qu'un échantillon ?

Nous pouvons définir un échantillon comme étant un ensemble d'individus qui doit être représentatif de l'ensemble dont il est prélevé pour les critères que l'on cherche à exprimer.

Le mot échantillon provient du vieux français «eschantillon» qui signifie «échelle pour mesurer». Cette définition semble satisfaisante car elle inclut l'idée d'étalon de mesure qui doit posséder certaines propriétés.

Quelles sont les qualités d'un bon échantillon ?

a) *un échantillon doit être correct*, c'est-à-dire que tous les éléments constitutifs des lots à échantillonner (clémentiniers) doivent avoir une égale probabilité d'être prélevés. C'est une propriété de l'opération d'échantillonnage seul.

b) *un échantillon doit être juste*, c'est-à-dire que la moyenne de la distribution du critère étudié de l'échantillon doit être égale à la moyenne du critère étudié de la population parente. On dit alors que l'échantillon n'est pas **biaisé**. L'écart enregistré entre les mesures du caractère pris sur la population et sur l'échantillon doit être nul. Cet écart, s'il existe, s'appelle biais ou erreur systématique ou erreur d'échantillonnage. Dans la réalité on cherchera à minimiser cette dernière erreur. La justesse est une propriété du résultat de l'échantillonnage : elle dépend, d'une part, de la correction de l'échantillonnage, (condition nécessaire mais pas suffisante) et d'autre part, de la population des éléments constitutifs sur laquelle on n'a aucune prise. Un échantillon peut donc être correct sans être juste.

c) *un échantillon doit être fidèle*, c'est-à-dire que l'erreur d'échantillonnage doit être peu dispersée autour de sa moyenne, qu'elle soit nulle ou non.

d) *un échantillon est précis ou représentatif* lorsqu'il est à la fois juste et fidèle, c'est-à-dire lorsque l'erreur d'échantillonnage est présumée faible quelles que puissent être ses origines et sa nature.

Quand faut-il échantillonner ?

L'époque à laquelle est effectué un échantillonnage dépend de ce qu'on cherche à exprimer, ou bien une évolution d'un ou plusieurs caractères dans le temps (ex : échantillonnage à fréquence mensuelle), ou bien l'expression d'un ou plusieurs caractères à un stade végétatif donné (ex : échantillonnage à maturité des fruits).

Dans le cas qui nous préoccupe, on se propose d'échantillonner la production du clémentinier en début de récolte, pour obtenir le maximum d'information sur la qualité et la quantité des clémentines produites. En Corse, cette période se situe généralement début décembre.

Qui va pratiquer l'échantillonnage ?

Le premier article paru sur cette étude a montré que deux observateurs pouvaient apprécier différemment des caractères qualitatifs.

Pour minimiser ce risque d'erreur on s'efforcera :

- de définir dans la méthodologie des caractères précis afin d'éviter toute ambiguïté dans leur interprétation.

- de choisir une même personne ou une même équipe pour échantillonner chaque verger ou chaque série de vergers dans des conditions identiques.

- la personne ou l'équipe qui aura la tâche d'effectuer l'échantillonnage devra s'astreindre à la rigueur des principes de la méthode et ne pas laisser libre cours à son intuition. De la qualité de son travail dépendra la qualité des résultats finaux.

Comment réaliser l'échantillonnage ?

C'est le but de l'étude. Avant de proposer une ou des méthodes d'échantillonnage, il faut étudier dans le détail la récolte du clémentinier observé.

L'idée qui vient tout de suite à l'esprit quand on songe à remplacer un ensemble (univers) par un échantillon c'est, selon P. THIONET, de constituer des groupes homogènes à l'intérieur de cet univers avant d'y prélever l'échantillon, afin d'amortir les fluctuations que le hasard peut apporter à sa constitution.

Ces «groupes homogènes» constituent les strates et l'opération s'appelle stratification. Plus précisément la stratification consiste à découper l'univers en groupes homogènes appelés strates, chaque unité appartenant à une strate et à une seule, puis à procéder à un tirage aléatoire dans chaque strate. On démontre par ailleurs que le fait de stratifier une population, s'il n'est pas toujours rentable, n'est du moins pas néfaste ; à la limite, le gain en précision de la stratification peut être nul mais l'erreur ne peut être supérieure à celle que l'on aurait obtenue par simple tirage. Le gain est d'autant plus important que la variable stratifiante, dite encore variable d'échantillonnage, est en bonne liaison avec la variable que l'on veut étudier, dite de production.

De fait, les variables (se reporter à l'article précédent pour la définition de celles-ci) peuvent se répartir en deux groupes :

- Variables de production ou variables à expliquer, déterminant une hiérarchie sur les clémentines :

- poids	symbole	POID
- extrait sec	symbole	ES
- acidité	symbole	AC
- nombre de pépins	symbole	PEPI
- épaisseur de la peau	symbole	EPAI

- Variables d'échantillonnage susceptibles d'expliquer les précédentes :

- coloration externe du fruit	symbole	COL
- orientation du fruit sur l'arbre		OR
- hauteur du fruit au-dessus du sol		HAUT
- distance du fruit à l'axe vertical de l'arbre		DIST

Les résultats de cette étude nous amèneront à effectuer une classification des variables dites explicatives ou d'échantillonnage. Cette classification dépendra de leur capacité à expliquer tel ou tel niveau de poids, tel ou tel niveau d'extrait sec, tel ou tel niveau d'acidité.

MÉTHODES STATISTIQUES EMPLOYÉES ET PLAN DE L'ÉTUDE

Le cheminement statistique choisi pour «analyser» les clémentines fait appel à différentes méthodes que nous décrirons à l'usage du lecteur non initié.

Les résultats seront présentés avec pour ossature :

Description statistique des variables quantitatives (poids, extrait sec, acidité, ES/AC, hauteur, distance, nombre de pépins, épaisseur de la peau).

● description classique des variables quantitatives et corrélations entre elles :

- moyenne,
- variance,
- coefficient de variation,
- tableau de corrélations linéaires.

● analyses en composantes principales sur la matrice des corrélations précédentes : présentation de la méthode, explication des résultats.

Essai de stratification du clémentinier par la méthode d'analyse des correspondances :

- présentation de l'analyse des correspondances,
- application au problème posé,
- construction des tableaux de contingence :
 - (poids x acidité) x [(coloration) x (orientation) x (hauteur x distance)],
 - (poids x extrait sec) x [(coloration) x (orientation) x (hauteur x distance)].

Détermination du pouvoir discriminant des variables : hauteur, distance, extrait sec, acidité, sur le poids, par la méthode d'analyse factorielle discriminante.

Conclusion : Propositions pour la recherche d'un échantillon représentatif des fluctuations de poids, d'acidité et d'extrait sec.

DESCRIPTION STATISTIQUE DES VARIABLES, QUANTITATIVES

● Description classique des variables quantitatives (tableaux 1 et 2).

Nous ne nous attarderons pas sur le tableau 1 des moyennes, variances, laissant le lecteur y jeter un regard. Ces premiers résultats dont on tire peu de conclusions donnent une image globale et imparfaite du clémentinier.

Le tableau 2 des corrélations montre que, excepté la corrélation négative forte entre l'acidité et le rapport extrait sec/acidité (-0,85), les autres corrélations offrent intrinsèquement un intérêt réduit pour l'interprétation agronomique.

Par nature, le coefficient de corrélation traduit une mesure de l'intensité des liaisons linéaires entre les variables, mais il ne peut en aucun cas suffire à les interpréter. Nous avons l'avantage de posséder une population importante (1.828 clémentines) et, de ce fait, un certain nombre de coefficients de corrélation, quoique faibles, sont significatifs. Cependant, pour porter un jugement sur une population, nous effectuerons de préférence une régression linéaire, avec l'analyse de variance correspondante, plus riche d'informations que le coefficient de corrélation qui en est la syn-

TABLEAU 1 - Description des variables.

variables	effectifs	moyenne	variance	écart type	coefficient de variation
hauteur des fruits	1828	159 cm	4.973,09	70,3	44
distance du fruit	-	123 cm	1.410,63	37,6	30
poids	-	51 cm	288,75	17,0	33
nombre de pépins	-	1,14	196,06	2,1	4,5
épaisseur de la peau	-	2,4mm	0,81	0,9	38
extrait sec	-	10,5	0,52	0,7	6,9
acidité	-	1,0	0,19	0,1	13
ES/AC	-	10,4	2,23	1,5	14

TABLEAU 2 - Coefficients de corrélation entre variables.

	distance	poids	nombre de pépins	épaisseur peau	extrait sec ES	acidité AC	ES/AC
hauteur	-0,064	0,342	0,080	0,110	0,367	-0,020	0,369
distance	1	0,229	0,028	0,121	0,270	-0,229	0,317
poids		1	0,375	0,544	0,329	-0,411	0,524
nbre pépins			1	0,170	0,046	0,137	0,101
épaisseur peau				1	0,016	-0,295	0,247
extrait sec					1	0,001	0,452
acidité						1	-0,852
ES/AC							1

thèse.

Soit un coefficient de corrélation de 0,50; il signifie que 25 p. cent de la variabilité totale est expliquée par la liaison linéaire entre les caractères analysés, c'est-à-dire, que la plus grande partie des fluctuations n'est pas expliquée par cette liaison. Aussi, sans rejeter un coefficient de corrélation de 0,50, nous procéderons à un examen plus approfondi de la population pour les caractères concernés.

Le cas d'un coefficient très proche de 1 correspond à une très forte liaison linéaire. Par contre, si un coefficient de zéro indique qu'il n'existe aucune liaison linéaire, il n'exclut pas qu'il puisse en exister une autre.

De fait, nous porterons une attention particulière sur les corrélations suivantes : extrait sec et acidité (0,0012), hauteur et distance (-0,06). Ces corrélations signifient une pratique indépendance linéaire entre les éléments de chacun des couples. Ainsi à un niveau d'extrait sec donné correspondent des niveaux d'acidité variés et réciproquement ; à un niveau donné de distance correspondent des niveaux variés de hauteur et réciproquement. Nous constaterons et décrirons mieux ces différents résultats par l'analyse en composantes principales qui va suivre.

● **Description par l'analyse en composantes principales effectuées sur les variables quantitatives citées dans le tableau 2.**

Cette méthode d'analyse multi dimensionnelle permettra de contrôler le clémentinier par la connaissance des différentes mesures quantitatives effectuées sur ses fruits. Elle n'est certainement pas la plus appropriée à l'étude, car les caractères qui sont offerts à l'analyse sont d'ordres quantitatif et qualitatif. Cependant, elle nous permettra de mieux

connaître les variables quantitatives retenues et d'avoir une première description de l'arbre qui ne sera ni sommaire, ni décisive, ni définitive.

principe de la méthode et application au problème posé

L'analyse en composantes principales a pour but de décrire un ensemble d'individus repérés au moyen de caractères ou mesures quantitatives.

L'étude directe du tableau 3 rectangulaire de données est laborieuse. Basée sur la méthode du tri à plat, elle a fait l'objet de la précédente publication.

Pour effectuer cette étude par la méthode de l'analyse en composantes principales, on a le choix entre deux espaces de départ :

- *espace des individus* : les huit «point-caractères» mesurés seront représentés dans un espace à 1828 dimensions, ce qui revient à dire que chaque caractère est repéré par un vecteur dont les 1828 coordonnées sont les mesures effectuées sur les 1828 clémentines correspondantes.

- *espace des caractères* : les 1828 «point-individus» ont chacun huit coordonnées qui sont les valeurs des huit caractères leur correspondant. L'individu peut être considéré comme la réalisation d'un vecteur aléatoire qui aura huit composantes et sera donc représenté dans un espace à huit dimensions.

On désire représenter, dans la mesure du possible, ces données statistiques dans un espace à nombre réduit de dimensions avec le minimum de perte d'information. On résume par exemple les huit variables mesurées par deux facteurs, chaque facteur étant formé par une ou plusieurs variables mesurées. On peut passer ainsi d'un espace à huit

TABLEAU 3 - Exemple type des données quantitatives.

individus	caractères observés							
	1	2	3	4	5	6	7	8
	hauteur	distance	poids	nbre pépins	épaisseur peau	extrait sec	acidité	ES/AC
1	180	120	52	1	3,5	12	1,3	9,2
2	150	80	45	0	2,0	15	2,0	7,5
3								
·								
·								
1828								

dimensions à un espace à deux dimensions. Sans citer de formule, nous expliquerons succinctement le principe du passage de l'espace à huit dimensions à l'espace à nombre réduit de dimensions. L'illustration qui va suivre vaut pour toute analyse multidimensionnelle.

principe de détermination des facteurs ou axes dans le cas d'un espace à huit dimensions (huit caractères mesurés) (figure 1).

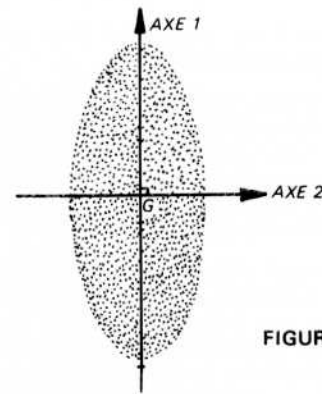
- Les nouveaux axes passeront par le centre de gravité G du nuage de points. Ce point G a pour coordonnées les différentes moyennes des caractères 1, 2, 3 ... 8.

- Le premier axe, ou composante principale, est choisi de telle façon que l'inertie du nuage de points (par analogie avec les systèmes dynamiques) par rapport à cet axe soit minimum. On dira aussi qu'il minimise la somme quadratique des distances de chaque point à lui-même : cet axe ajuste au mieux le nuage de points.

- Le deuxième axe (figures 1 et 2) est choisi de telle façon qu'il soit orthogonal au premier tout en minimisant l'inertie du nuage de points par rapport à lui-même.

- Plan des deux premiers axes.

Ce plan représenté par la figure 2 est celui par rapport auquel l'inertie est minimum : il est appelé plan principal.



G = centre de gravité

FIGURE 2

- Autres axes : ils sont obtenus à partir du même principe.

Nous avons ainsi trouvé des axes factoriels qui ont la propriété d'extraire progressivement le plus d'informations possible concernant les proximités entre les points «individus» de l'espace à huit dimensions. La qualité de cette information peut être mesurée par l'importance de la variance expliquée par les différents facteurs (tableau 4).

L'analyse en composantes principales ne fait appel à aucune hypothèse particulière préalable. Cette méthode

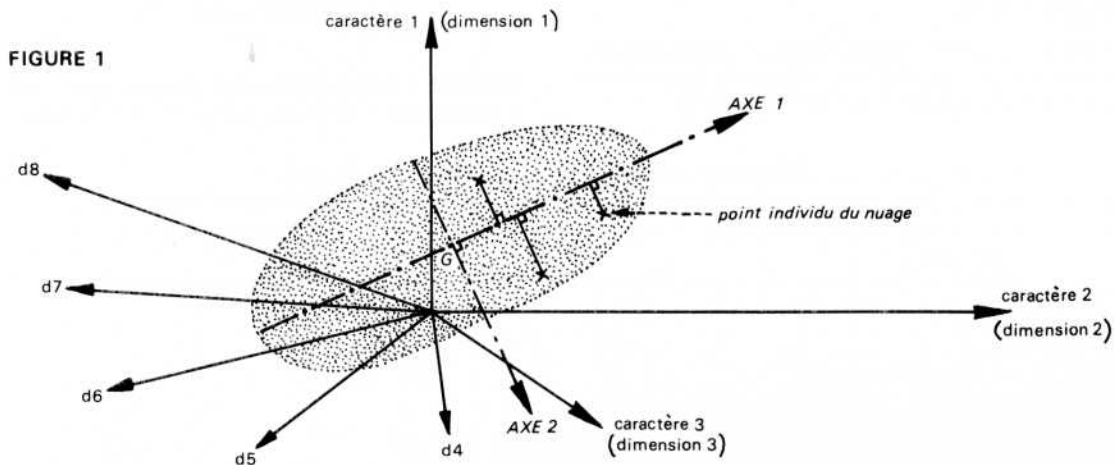


FIGURE 1

TABLEAU 4 - Interprétation des résultats

composante ou axe factoriel	variance de la composante	part de variance expliquée par la composante	part cumulée
1	2.95135	.36892	.36892
2	1.31653	.16457	.53349
3	1.10694	.13837	.67185
4	.99237	.12405	.79590
5	.80691	.10086	.89676
6	.49118	.06140	.95816
7	.29230	.03654	.99470
8	.04241	.00530	1.00000

permet uniquement une description de la population par l'intermédiaire des caractères mesurés.

Les deux premiers facteurs expliquent respectivement 37 p. cent et 16 p. cent de la variance totale, soit 53 p. cent à eux deux. Ces deux premières composantes représentent une dispersion organisée indépendante des fluctuations liées au choix de l'arbre, sauf bien sûr si ce dernier avait été choisi nettement aberrant.

Le plan principal affichera des tendances générales de comportement expliquant la plus grande part de variabilité des clémentines. Les individus ayant des caractéristiques différentes de la majorité des autres n'auront pas leur tendance représentée en ce plan. C'est pourquoi il serait inexact de penser que ces facteurs n'extraient que 53 p. cent de l'information utilisable ; comme nous le verrons par la suite, la quantité d'informations apportées par ce plan est bien plus importante que ne pourrait le faire croire ce pourcentage.

Par ailleurs, il serait incorrect d'arrêter l'analyse à ces deux premiers axes. En effet, les autres axes apportent 47 p. cent de la variance totale et permettent d'analyser la population dans son ensemble. L'étude montre qu'il n'est pas nécessaire d'aller au-delà des quatre premiers axes, les variables les caractérisant suffisant à analyser le phénomène de dispersion.

Corrélations des composantes principales avec les variables (tableau 5).

Le tableau 5 indique les corrélations des variables étudiées avec les axes 1 à 6 et permet une analyse de ce que nous avons précédemment appelé « dispersion organisée ».

L'axe 1 (figure 3) est défini principalement par l'opposition entre : d'une part, le poids et le rapport ES/AC et l'acidité, d'autre part. Les corrélations respectives de ces caractères avec cet axe sont de 0,81-0,88 et -0,71. Le poids et l'ES/AC sont pratiquement totalement expliqués par cet axe (leurs corrélations avec les autres axes étant faibles). Nous verrons que l'acidité est aussi expliquée en partie par

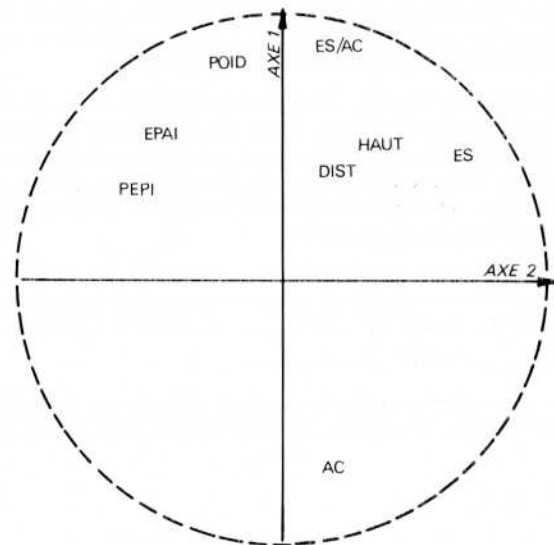


FIGURE 3 - Corrélation des variables avec les nouvelles composantes : 2 en abscisse et 1 en ordonnée.

TABLEAU 5 * Corrélations des composantes principales avec les variables.

axe principal ou composante	hauteur	distance	poids	nbre pépins	ép. peau	extrait sec ES	acidité AC	ES/AC
	1	2	3	4	5	6	7	8
1	.502	.407	<u>.813</u>	.314	.527	.476	<u>-0.708</u>	<u>.876</u>
2	.335	.197	-0.199	<u>-0.529</u>	<u>-0.485</u>	<u>.722</u>	.239	.183
3	<u>-0.569</u>	<u>.586</u>	-0.221	-0.370	-0.073	-0.203	-0.394	.228
4	-0.309	<u>.549</u>	.215	.314	.186	.309	<u>.499</u>	-0.266
5	.023	<u>-0.910</u>	.115	<u>-0.603</u>	<u>.621</u>	.006	.143	-0.124
6	-0.463	-0.378	.101	.010	-0.012	.305	-0.012	.174

* - les valeurs soulignées sont les plus fortes sur l'axe correspondant.

l'axe 4. Cette première information peut se résumer de la façon suivante : la population de clémentines varie principalement en fonction du poids, de l'acidité et du rapport ES/AC.

Sur l'arbre étudié et pour une grande partie de la population, nous constatons que lorsque le poids est fort, l'acidité est faible et réciproquement. Une autre partie moins importante ne répond pas à ce phénomène et sera expliquée par le quatrième axe. Le fait que le rapport ES/AC varie en sens contraire de l'acidité est la conséquence du champ de variation de la mesure de l'acidité qui est nettement plus grand que celui de l'extrait sec. Pour cette raison nous ne prendrons pas en considération cette variable dans la suite de notre commentaire.

Les autres variables expliquent l'axe 1 dans une moindre mesure ; leurs corrélations sont pour la plupart inférieures ou égales à 0,50. Toutefois pour une certaine partie de la population, l'extrait sec (corrélé à 0,48) augmente avec le poids s'opposant ainsi à l'acidité.

La hauteur et la distance ont aussi une corrélation non négligeable avec cet axe (0,50 et 0,41). Nous analyserons ce phénomène le moment venu (troisième axe). Notons déjà que l'ES augmente avec la hauteur et la distance, l'acidité décroissant alors.

L'axe 2 est défini par l'extrait sec (ES) principalement. La corrélation de cette variable avec cet axe est de 0,72.

L'épaisseur (EPAI) et le nombre de pépins (PEPI) expliquent cet axe dans une moindre mesure.

Nous retiendrons l'explication de cet axe par l'extrait sec qui est la variable de qualité représentant la seconde source de variation principale du lot de clémentines.

Le plan des axes 1 et 2 (figure 3) fait apparaître l'indépendance entre les critères suivants :

- extrait sec et épaisseur (corrélation = 0,016)
- extrait sec et nombre de pépins (corrélation = 0,046)
- extrait sec et acidité (corrélation : 0,0012). Ce résultat jouera dans la suite de notre étude car ces deux variables sont des critères de qualité ; il indique que pour un niveau d'extrait sec donné, nous avons des niveaux d'acidité variés et réciproquement.

L'axe 3 (figures 4 et 5) est défini par l'opposition entre la hauteur et la distance (corrélations respectives avec cet axe : - 0,57 + 0,58), ce qui signifie que la position des clémentines sur l'arbre est très variée. Cependant les corrélations à 0,50 et 0,41 de la hauteur et de la distance avec l'axe 1 traduisent le fait que pour une certaine partie de la population la distance à l'axe central de l'arbre augmente avec la hauteur (graphique 1).

Le plan des axes 1 et 3 (figure 4) indique entre hauteur et distance une indépendance (corrélation = 0,064) qui confirme la répartition variée des clémentines sur l'arbre ; cette particularité sera un élément important pour l'étude de l'échantillonnage.

L'axe 4 (figure 6) est principalement caractérisé par la distance et l'acidité et dans une faible mesure par la hauteur. Cet axe indique que pour une faible partie de la population, lorsque la distance est forte et la hauteur faible, l'acidité est

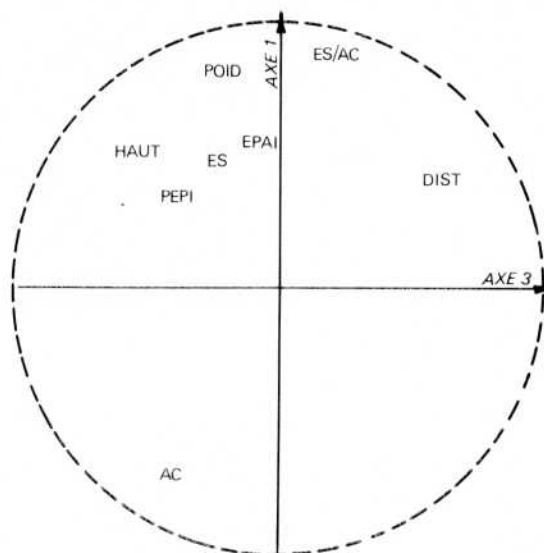


FIGURE 4 - Corrélation des variables avec les nouvelles composantes : 3 en abscisse et 1 en ordonnée

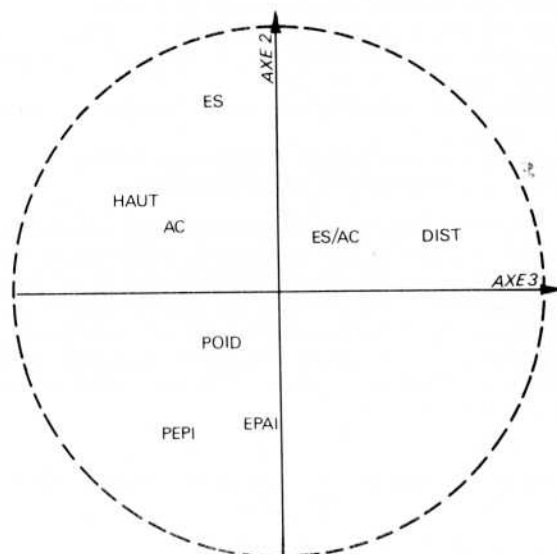


FIGURE 5 - Corrélation des variables avec les nouvelles composantes : 3 en abscisse et 2 en ordonnée.

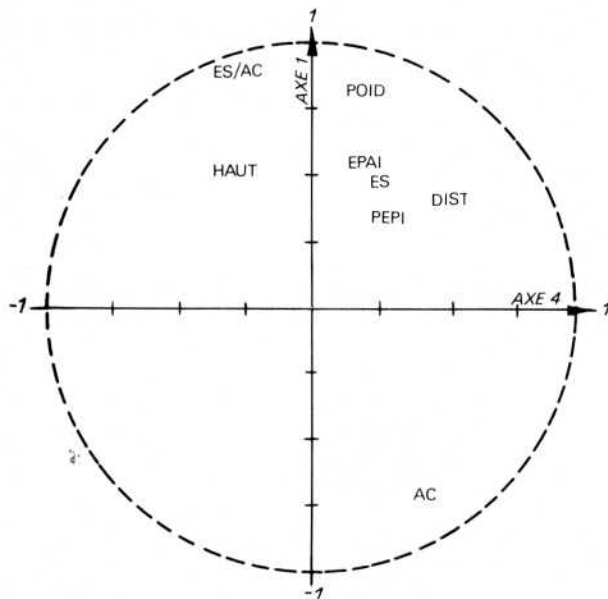


FIGURE 6 - Corrélation des variables avec les nouvelles composantes : 4 en abscisse et 1 en ordonnée.

élevée.

En résumé, de l'analyse en composantes principales se dégagent les éléments suivants :

- poids, acidité, extrait sec, sont les variables expliquant le plus les fluctuations des clémentines observées,
- indépendance entre extrait sec et acidité ($r = 0,0012$)
- indépendance entre extrait sec et épaisseur de la peau ($r = 0,016$)
- indépendance entre extrait sec et nombre de pépins ($r = 0,046$)
- indépendance entre hauteur et distance ($r = 0,064$)
- répartition variée des clémentines sur l'arbre.

Le nombre d'individus étant trop important, nous n'avons pas sorti la projection des individus sur les différents plans (interprétation visuelle quasi-impossible). De ce fait, nous proposerons de constituer des groupes d'individus afin d'essayer de les différencier d'après les critères de position géographique sur l'arbre et les critères de qualité ou de production, ce qui nous amène à effectuer une analyse des correspondances.

ESSAI DE STRATIFICATION DU CLÉMENTINIER OBSERVÉ

● Présentation de la méthode d'analyse des correspondances.

L'analyse des correspondances s'apparente à l'analyse en composantes principales par le fond de la méthode et l'obtention des résultats. Elle conduit à trouver des axes factoriels qui ont la propriété d'extraire le maximum d'information (inertie) concernant la proximité des variables entre elles.

Elle a toutefois le privilège, non seulement de décrire un

tableau rectangulaire, mais aussi d'expliquer des variables par d'autres variables et de prendre en considération indifféremment des variables qualitatives et quantitatives. Ainsi, les tableaux qui seront offerts à l'analyse des correspondances seront des tableaux multiples (ou à multiples entrées). Dans notre cas, nous analyserons par exemple le tableau suivant :

(poids \times acidité) \times [(coloration) \times (orientation) \times (hauteur \times distance)].

Sa forme est décidée a priori en fonction des résultats que l'on cherche à démontrer. Désignons-le symboliquement par :

(P \times AC) \times (C \times O \times H \times D) ou (ensemble J) \times (ensemble I)

Chaque ensemble ci-dessus est décomposé en classes. On aura un tableau $K_{I \times J}$ de quantités de clémentines dans lequel $k(ij)$ représente le nombre de clémentines de poids p et d'acidité a , ces dernières auront la coloration c , l'orientation géographique o , la hauteur par rapport au sol h et la distance à l'axe central d .

Pour faciliter les explications qui vont suivre nous nous appuyerons sur l'analyse du tableau simple K_{HP} , où P représente l'ensemble des vecteurs classes de poids, H l'ensemble des vecteurs classes de hauteurs, P et H étant respectivement les variables expliquées et explicatives.

Nous voulons représenter les ensembles P et H sur un même plan de façon à ce que leurs proximités naturelles soient mises en évidence. L'analyse des correspondances se propose d'arriver à ce but en se basant sur les notions suivantes :

1. Deux classes de poids p et p' seront proches l'une de l'autre, si les distributions dans les colonnes p et p' du tableau K_{HP} sont presque proportionnelles.
2. Deux classes de hauteurs h et h' seront proches si les lignes h et h' du tableau sont semblables, quels que puissent être leurs totaux marginaux respectifs. Dans le premier cas comme dans celui-ci, nous dirons que les profils sont équivalents.
3. Une classe de poids p est d'autant plus proche d'une classe de hauteur h que cette dernière intervient davantage dans cette classe de poids p .

Afin de bien définir le tableau numérique de départ et le but principal recherché nous nous permettrons quelques formules mathématiques, soit :

- le tableau K_{HP}

- les effectifs de cases :

$k(h,p)$ nombre de clémentines de poids p et de hauteur h

- les valeurs marginales

$k(p) = \sum_h k(h,p)$ nombre de clémentines de poids p

$k(h) = \sum_p k(h,p)$ nombre de clémentines de hauteur h

$K = \sum_{h,p} k(h,p)$ nombre total des clémentines

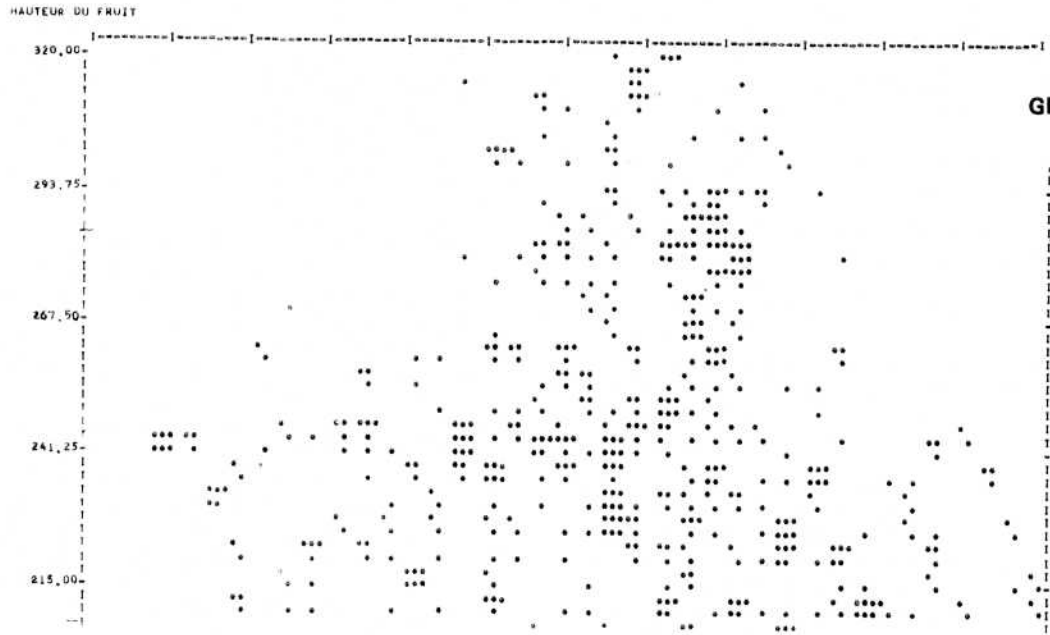
- les fréquences de cases :

$f(h,p) = k(h,p)/K$ p. cent de clémentines de poids p et de hauteur h dans l'ensemble de clémentines,

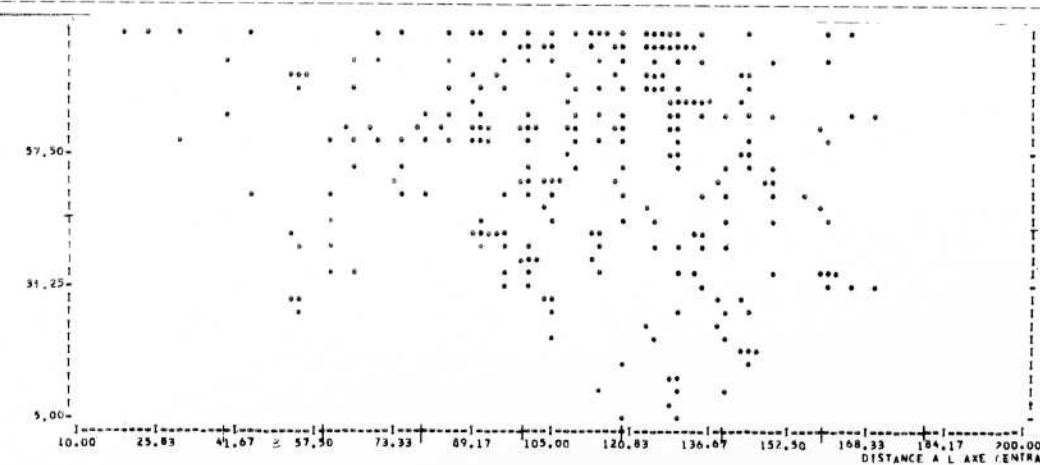
- les fréquences conditionnelles de p si h et de h si p respectivement :

$f(p \text{ si } h) = p.$ cent de clémentines de poids p et de hauteur h dans l'ensemble des clémentines de hauteur h

$f(h \text{ si } p) = p.$ cent de clémentines de hauteur h et de poids



GRAPHIQUE 1



- p dans l'ensemble des clémentines de poids p
- les fréquences marginales :
 - . de colonne :
 - f (p) = k (p)/K : p. cent de clémentines de poids p dans l'ensemble de clémentines
 - . de ligne :
 - f (h) = k (h)/K : p. cent de clémentines de hauteur h dans l'ensemble de clémentines

Si, quels que soient h et p, on avait l'identité rigoureuse :
 $f(h,p) = f(h) \times f(p)$

nous aurions l'indépendance statistique entre la hauteur et le poids, autrement dit le poids ne dépendrait pas de la hauteur, alors on ne pourrait rien extraire du tableau soumis à l'analyse.

En terme statistique, nous dirons que nous avons choisi l'indépendance comme hypothèse nulle H_0 . En analyse des correspondances, il ne s'agit pas de savoir s'il faut ou non rejeter cette hypothèse nulle, mais de voir apparaître aussi clairement que possible comment les données diffèrent de H_0 , c'est-à-dire dans quelle direction elles s'écartent de l'indépendance. En supposant les données conformes à un modèle provenant d'une ou de plusieurs lois diverses, les facteurs extraits au moyen de l'analyse permettent de découvrir le modèle et d'en calculer ses paramètres.

● Position du problème et choix des variables.

Le choix des variables dérive principalement du problème posé. Or, il paraît logique de cataloguer les clémentines en fonction des variables de production. Nous tenterons donc d'expliquer le poids, l'extrait sec et l'acidité par d'autres mesures rendant compte de leurs variations respectives constatées et permettant à la fois de partitionner l'arbre et de déterminer des zones d'iso-poids, d'iso-extrait sec et d'iso-acidité.

Variables introduites.

- Variables explicatives ou d'échantillonnage ou de contrôle : coloration, orientation, hauteur, distance
- Variables de production ou à expliquer : poids, extrait sec, acidité

• *Choix des tableaux* sur lesquels seront effectuées les analyses.

La classification des variables dans les deux groupes ci-dessus conduit à analyser les tableaux suivants :

- (POID x AC) = [(COL) x (OR) x (HAUT x DIST)] (tableaux 10 et 11)
- (POID x ES) = [(COL) x (OR) x (HAUT x DIST)] (tableaux 13 et 14).

En effet, l'acidité et l'extrait sec évoluant d'une façon indépendante, il nous a semblé préférable de les substituer l'un à l'autre dans les tableaux précités, afin de pouvoir comparer les analyses et cataloguer facilement les clémentines se trouvant dans tel ou tel secteur géographique. De plus, le croisement des différentes variables nous aidera à dégager des conclusions. Cette façon de procéder n'est certainement pas exclusive.

Limites de classes des différentes variables.

Le choix des limites de classes a été déterminé de telle façon que l'on obtienne des tableaux le maximum d'informations sur les différentes tendances générales à l'indépendance.

C'est en visualisant le tableau de contingence à analyser que nous avons cherché des regroupements de classes, effectués de telle façon que des différences trop importantes entre lignes ou entre colonnes ne masquent pas la description sur le graphe de phénomènes beaucoup plus généraux. On aurait pu regrouper la colonne P3A3, dont l'effectif est petit avec P3A2. Toutefois nous avons voulu conserver les différentes classes d'acidité pour le poids P3, afin de mieux juger de l'interaction poids x acidité. Nous aurions pu effectuer plus de classes de poids, d'extrait sec, d'acidité, de hauteur, de distance, mais partant du principe que ces différentes variables étaient continues, la décomposition en trois classes était suffisante pour faire apparaître aussi clairement que possible la direction vers laquelle nos données s'écarteraient de l'indépendance.

Par ailleurs, le regroupement de l'orientation peut paraître surprenant sur le plan géographique. Il faut préciser que lors de l'étude précédente, nous avons constaté une tendance des fruits situés au nord à avoir un poids relativement faible. Pour essayer de donner à l'orientation la dimension à laquelle elle peut postuler et donc de ne pas effectuer de regroupement contre nature, nous avons dû effectuer une analyse des correspondances préliminaire sur le tableau de contingence poids x orientation (9 classes de poids et 8 classes d'orientation) exprimé en p. cent du total (tableau 6).

L'inertie expliquée par chacun des axes est indiquée dans le tableau 7.

Le produit (inertie totale x nombre d'individus) est un Chi^2 à (9 classes de poids - 1) x (8 classes d'orientation - 1) degrés de liberté, soit ici :

$$0,0634 \times 1754 = 111$$

est un Chi^2 à 56 degrés de liberté

D'après la table du Chi^2 , cette valeur correspond à un seuil de probabilité inférieur à 0,005.

On peut en conclure raisonnablement qu'à certaines orientations correspondent certaines catégories de poids plutôt que d'autres. Bien que l'orientation ait une influence sur le poids inférieure à celle de la hauteur ou de la distance, comme nous le verrons, nous avons jugé utile de conserver les tendances constatées sur le plan des axes 1 et 2 (figure 7).

Ainsi nous remarquons les oppositions suivantes :

- axe 1 : opposition entre nord-ouest et (nord-est, est, sud-est)
- axe 2 : opposition entre nord et (sud-ouest, sud, et dans une moindre mesure ouest)

Dans l'ensemble, le nord s'oppose à toutes les autres directions.

A partir de ces remarques, nous avons adopté le découpage géographique indiqué figure 8.

Nous avons dû effectuer également l'analyse des corres-

TABLEAU 6 - Poids x Orientation en pourcentage

Classes de poids en grammes		P1	P2	P3	P4	P5	P6	P7	P8	P9	total
		10-22	22-34	34-46	46-58	58-70	70-72	72-94	94-106	106-118	
1	nord	3,03	1,57	1,91	1,51	1,12	1,23	0,28	0,67	0,29	11,54
2	nord-est	0,67	0,67	0,67	0,62	0,56	0,67	0,56	0,78	0,06	5,26
3	est	2,02	1,57	1,63	1,85	1,57	1,63	1,40	1,46	0,28	13,41
4	sud-est	2,19	1,63	2,52	1,62	1,40	1,35	1,34	1,18	0,56	13,79
5	sud	1,34	1,63	1,34	1,68	1,96	1,74	1,46	1,12	0,67	12,94
6	sud-ouest	0,84	0,84	1,23	1,01	1,24	1,18	1,12	1,01	1,01	9,48
7	ouest	2,13	2,97	2,30	2,19	1,85	2,13	1,46	1,74	0,78	17,55
8	nord-ouest	2,13	2,18	2,13	3,25	2,13	1,57	1,01	1,01	0,62	16,03
total		14,35	13,06	13,73	13,73	11,83	11,50	8,63	8,97	4,20	1,00

TABLEAU 7

axe ou facteur	inertie	p. cent inertie expliquée par chacun des axes	p. cent cumulé
1	0,0346	55	55
2	0,0108	17	72
3	0,0098	15	87
4	0,0039	6	93
5	0,0030	5	98
6	0,0012	2	100
7	0,0001	0	100
8	0,0000	0	100
total	0,0634	100	100

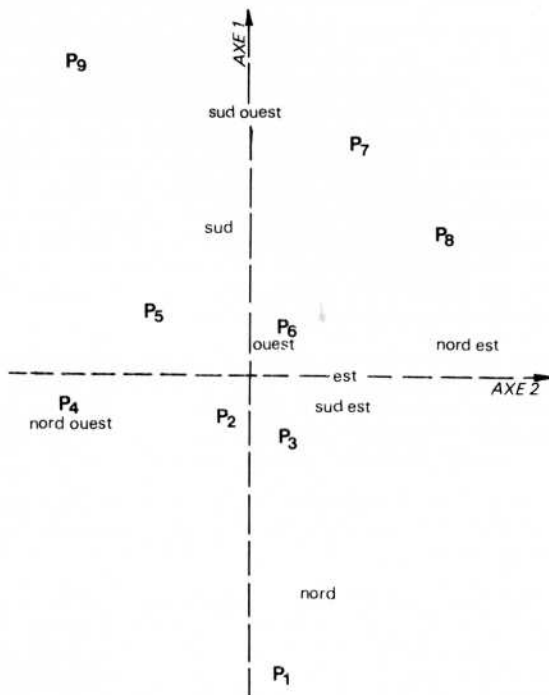


FIGURE 7 - Tendances constatées poids x orientation.

pondances sur les tableaux suivants :

POID x HAUT avec 9 classes de poids et 8 classes de hauteur
 POID x DIST avec 9 classes de poids et 8 classes de distance

On a remarqué que les sommes des inerties des deux premiers axes relatives à chaque tableau sont :

	axe 1	axe 2	
hauteur	0,1298	+	0,0225 = 0,1523
distance	0,1048	+	0,0140 = 0,1188

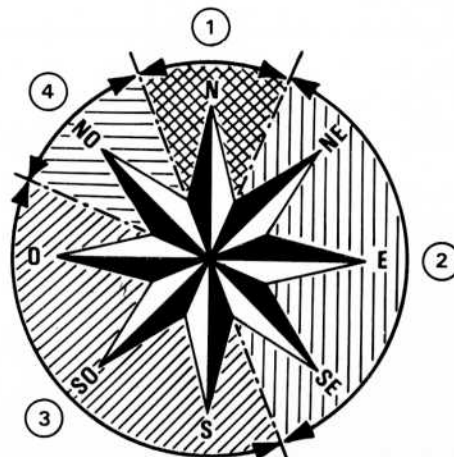


FIGURE 8 - Découpage géographique en fonction du poids x orientation.

TABLEAU 8 - Limites de classes réalisées.

poids	P ₁ [10-46[P ₂ [46-64[P ₃ [64 et plus	
extrait sec	ES ₁ [8,0-10,0[ES ₂ [10,0-10,7[ES ₃ [10,7 et plus	
acidité	A ₁ [0,7-1,0[A ₂ [1,0-1,1[A ₃ [1,1 et plus	
coloration	1 fond vert-tournant	2 orange	3 rouge	
orientation	1 (N)	2 (NE+ E +SE)	3 (S+SO+O)	4 (NO)
hauteur	H ₁ [0-120[H ₂ [120-180[H ₃ [180 et plus	
distance	D ₁ [0-120[D ₂ [120-160[D ₃ [160 et plus	

Ces inerties sont nettement supérieures à celles trouvées pour l'orientation, et nous pouvons déjà classer les variables d'échantillonnage en fonction de leur pouvoir discriminant du poids.

- 1 - hauteur 0,23
- 2 - Distance 0,12
- 3 - Orientation 0,04

Les limites de classes effectivement réalisées sont données dans le tableau 8 pour les différentes variables.

Dans les analyses qui vont suivre :

Nous avons supprimé les groupes d'individus dont les effectifs sont inférieurs ou égaux à trois. Un groupe est caractérisé par le vecteur aléatoire à 9 composantes $x_1 \dots x_9$. Supposons que les différentes coordonnées soient les suivantes : 2,0,0,0,0,0,0,0, la somme est alors inférieure à trois. Ce groupe apparaîtrait en extrémité du graphe en direction de x_1 . Ce graphe étant limité en surface, on risquerait de voir d'autres fluctuations apparaître moins clairement. Ce sont donc des raisons pratiques d'interprétation qui nous poussent à supprimer ce groupe.

Nous avons exprimé la situation géographique des groupes d'individus en fonction de neuf combinaisons de classes de hauteur x distance, définies comme suit (tableau 9).

TABLEAU 9 - Combinaisons hauteur x distance en cm

	classes de distance			
		[0 - 120[[120 - 160[[160 et plus
classes de hauteur				
[180 et plus		7	8	9
[120 - 180[4	5	6
[0 - 120[1	2	3

Analyse du tableau de contingence (POID x AC) x [(COL) x (OR) x (HAUT x DIST)]

Le lecteur aura intérêt à se reporter au tableau 10 des effectifs et au tableau 11 des fréquences.

Pour faciliter la compréhension du phénomène, nous avons effectué trois extraits du graphe général des correspondances (plan formé par les axes 1 et 2).

Ces différents extraits proviennent d'une seule et même analyse du tableau 10 de contingence. Ils diffèrent entre eux par les variables explicatives choisies pour repérer les groupes d'individus. Ces trois extraits sont intitulés :

- analyse (poids x acidité) x (hauteur x distance) (graphe 4)
- analyse (poids x acidité) x (orientation) (graphe 5)
- analyse (poids x acidité) x (coloration) (graphe 6).

Analyse générale (tableau 12).

TABLEAU 12 - Inerties expliquées par chacun des axes.

axe ou facteur	parts d'inertie	p. cent inertie expliqué par chaque axe	p. cent cumulé
1	0,291	0,35	0,35
2	0,179	0,21	0,56
3	0,101	0,12	0,68
4	0,073	0,09	0,77
5	0,069	0,08	0,85
6	0,053	0,06	0,91
7	0,042	0,05	0,96
8	0,034	0,04	1,00
total	0,842		

Les parts d'inerties expliquées par chacun des trois premiers axes sont importantes, ce qui signifie qu'entre groupes ou combinaisons des variables explicatives (coloration, orientation, hauteur, distance), doivent exister des différences de profils. La chute des parts d'inertie est nette du premier axe au second, puis du second au troisième ; entre axes successifs suivants, elle l'est beaucoup moins. Ces derniers axes expliquent une faible part d'inertie que l'on attribuera aux fluctuations propres au matériel végétal. Ainsi nous limiterons l'analyse aux trois premiers axes qui extraient 68 p. cent de l'inertie totale, part attribuable à la « dispersion organisée » ou tendance générale.

Nous examinerons successivement la représentation des neuf classes de (poids x acidité) et relativement à elles, la constellation des groupes de clémentines sur l'arbre, représentée dans les graphes successifs pour les classes de hauteur x distance, d'orientation et de coloration.

Il est possible d'interpréter successivement les trois axes (ou en d'autres termes, les projections des points « variables à expliquer » sur les trois axes de nos figures).

Le premier axe (graphe 1) est principalement un axe de poids. Il oppose les poids faibles (P₁) situés à gauche du centre du graphe aux poids forts (P₃) situés à droite du centre du graphe. Les poids moyens (P₂) occupent une position intermédiaire entre les poids faibles et les poids forts, c'est-à-dire au centre du graphe. Les poids croissent donc sur cet axe de gauche à droite.

Le second axe (graphe 1) est un axe d'acidité. Les faibles acidités A₁ se situent en haut de cet axe, au-dessus du centre du graphe tandis que les fortes acidités A₃ se situent très au-dessous du point 0 de cet axe. Ainsi l'acidité varie en sens inverse de la direction de l'axe 2.

La séparation en zones «d'iso-poids», par des droites d'allure générale perpendiculaire au premier axe, donc parallèles au second, est nette, par contre, celle des zones «d'iso-acidité» l'est moins quoique relativement claire.

Pour la première droite choisie d'iso-poids (P₁A₁ - P₁A₂ - P₁A₃), on constate une opposition, entre P₁A₁ et P₁A₃, le point P₁A₂ étant pratiquement un point-milieu sis sur l'axe 1. Pour les deux autres droites d'iso-poids, on constate un rapprochement entre P₂A₂ et P₂A₃, et entre P₃A₂ et P₃A₃. L'opposition entre P₂A₁ et (P₂A₂, P₂A₃) et entre P₃A₁ et (P₃A₂, P₃A₃) est plus importante que dans le cas précédent, ce qui nous amène à examiner l'axe 3.

Le troisième axe (graphe 2) est principalement un axe de croissance d'acidité à poids constant P₃. Toutefois, il met en évidence l'interaction Poids x Acidité. Sur cet axe on constate un profil identique entre P₁A₃ et P₃A₁, P₃A₂ et P₁A₂ et dans une moindre mesure P₃A₃ et P₁A₁. Autrement dit, la nature nous donne une probabilité sensiblement identique d'obtenir des clémentines de poids P₁ et d'acidité A₃ ou de poids P₃ et d'acidité A₁.

Pour les couples P₃A₃ et P₁A₁, le phénomène probabiliste précédent se vérifie dans une moindre mesure. L'acidité A₂ s'associe indifféremment aux trois classes de poids. Cet axe n'explique pas de relation privilégiée entre P₂ et l'une ou l'autre des classes d'acidité et ne nous permet pas de comparer le profil P₂A aux autres profils P₁A et P₃A.

L'axe 4 (graphe 3) est expliqué par les combinaisons de P₃ et, dans une moindre mesure de P₂ avec les différents niveaux d'acidité.

Nous retiendrons principalement le parallélisme des lignes brisées d'iso-poids P₂ et P₃ dû au comportement identique de certains secteurs se situant principalement sur le cône extérieur de l'arbre. La ligne d'iso-poids P₁ en fonction de l'acidité diffère des précédentes et est le fait d'individus situés dans le cylindre central.

Analyse du premier extrait (graphe 4 et tableau 21)

Explication du poids et de l'acidité des clémentines par la hauteur au-dessus du sol et par la distance à l'axe vertical de l'arbre.

Pour une meilleure clarté des graphes, nous les compléterons d'une représentation de l'arbre en ses différentes combinaisons de hauteur x distance déjà décrites (figure 9).

Cette représentation permet de positionner les clémentines dans un secteur précis de l'arbre ; une répartition claire se dégage :

- les individus distribués autour de la droite P₁ du graphe sont caractérisés presque exclusivement par les couples hauteur x distance 1, 2 et 4 définissant la zone A du graphe 4.

Les clémentines se situant dans cette partie de l'arbre ont des degrés d'acidité variés. Toutefois elles tendent à associer

les niveaux d'acidité A₃ (voire A₂) aux faibles poids P₁.

- les individus distribués autour de la droite P₃ du graphe sont presque exclusivement caractérisés par les couples de hauteur x distance 6, 8 et 9.

Ils définissent la zone C du graphe.

Cette partie de l'arbre porte les fruits les plus lourds (P₂ et P₃) ; il semble que l'association P₃A₁, voire P₂A₁, soit privilégiée dans cet ensemble.

- la zone B, située entre les deux droites brisées d'iso-poids P₁ et P₃, est une zone intermédiaire qui assure en quelque sorte le phénomène de continuité des poids. Elle concerne les régions 3, 5 et 7 de l'arbre et comprend des fruits de tous poids à dominante P₂ et P₁ et d'acidité dominante A₁.

Finalement, l'arbre est partitionné en différentes régions permettant de relier les variations de poids des fruits à leur position sur l'arbre. Ces positions ne sont encore définies que dans le système de repère hauteur x distance, c'est-à-dire qu'elles se situent à l'intérieur de cylindres ou manchons découpés dans le volume de l'arbre. L'orientation permettant de reconnaître quatre secteurs par cylindre ou manchon, nous essaierons de déterminer son rôle dans l'explication du poids et de l'acidité.

Analyse de l'extrait n 2 (graphe 5)

Explication du poids et de l'acidité par l'orientation.

A l'origine, l'orientation avait été mesurée selon les huit directions normales. Rappelons les regroupements afin de gagner en homogénéité d'analyse :

- 1 - nord
- 2 - nord-est + est + sud-est
- 3 - sud + sud-ouest + ouest
- 4 - nord-ouest

Deux zones se dégagent du graphe :

- une zone A formée essentiellement des groupes situés au

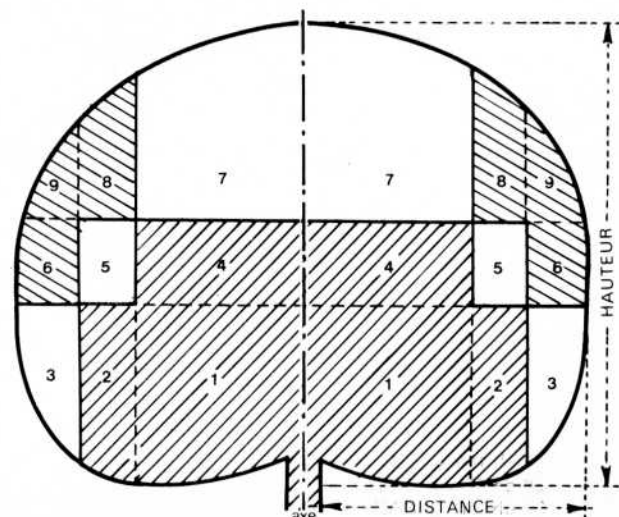
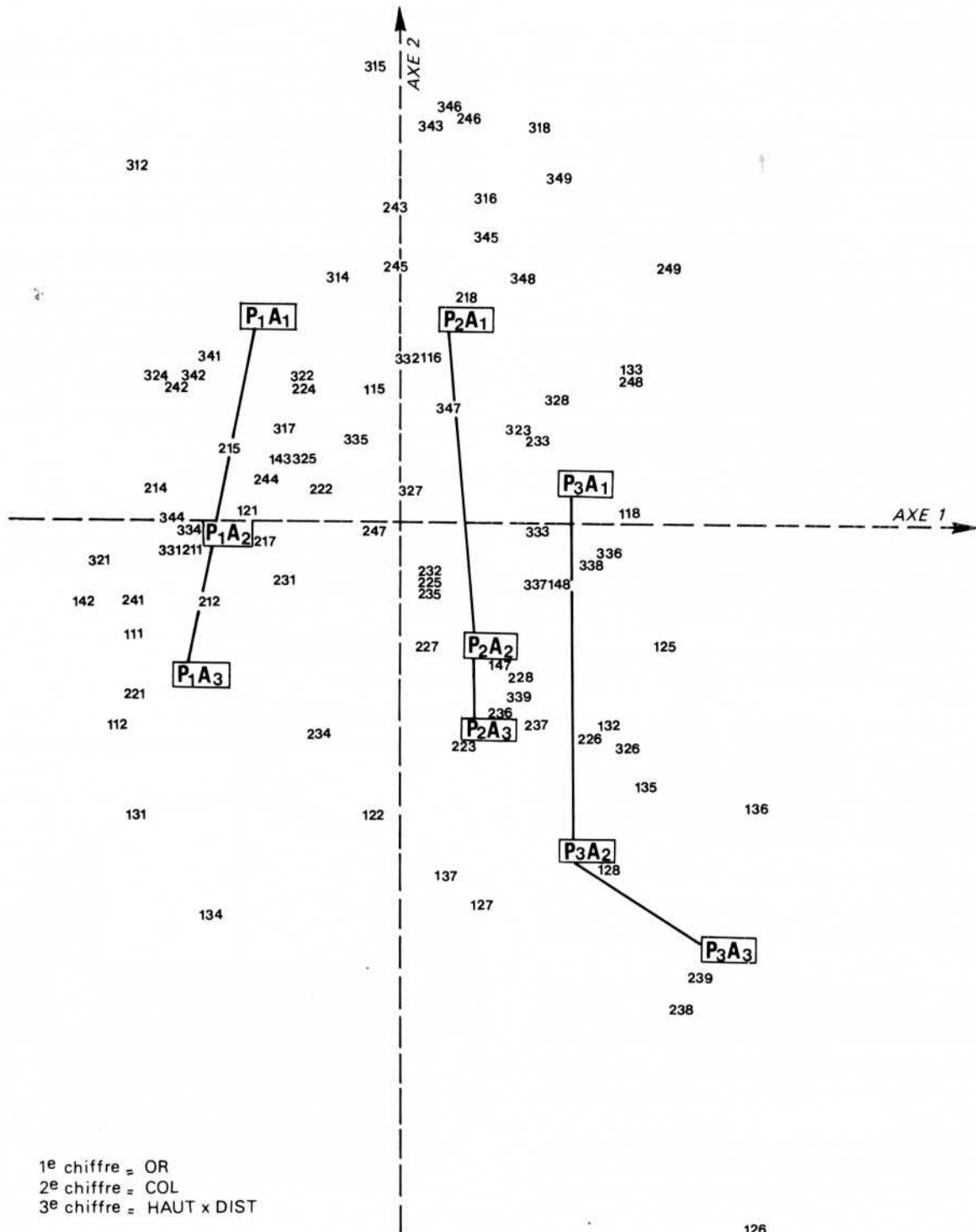
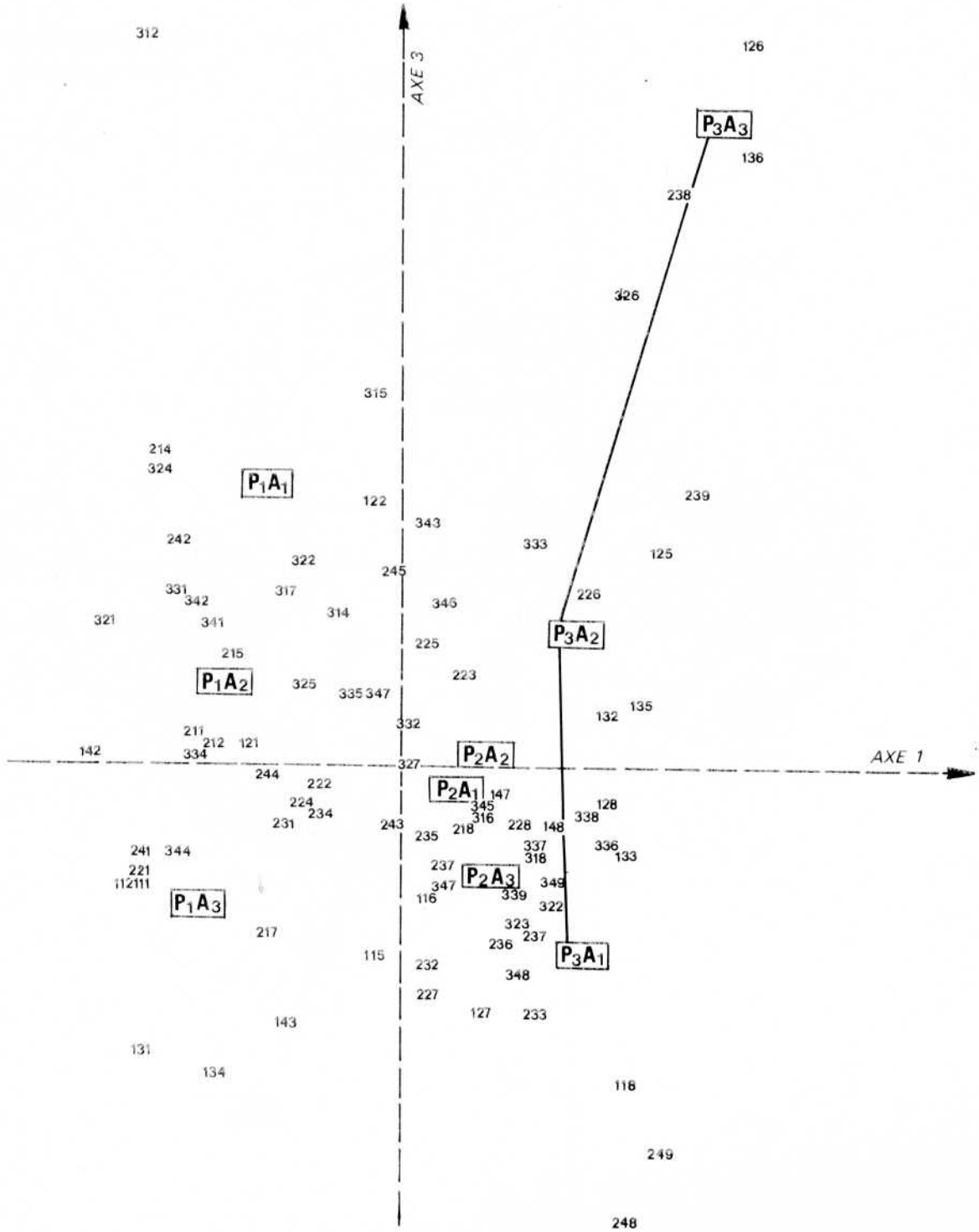


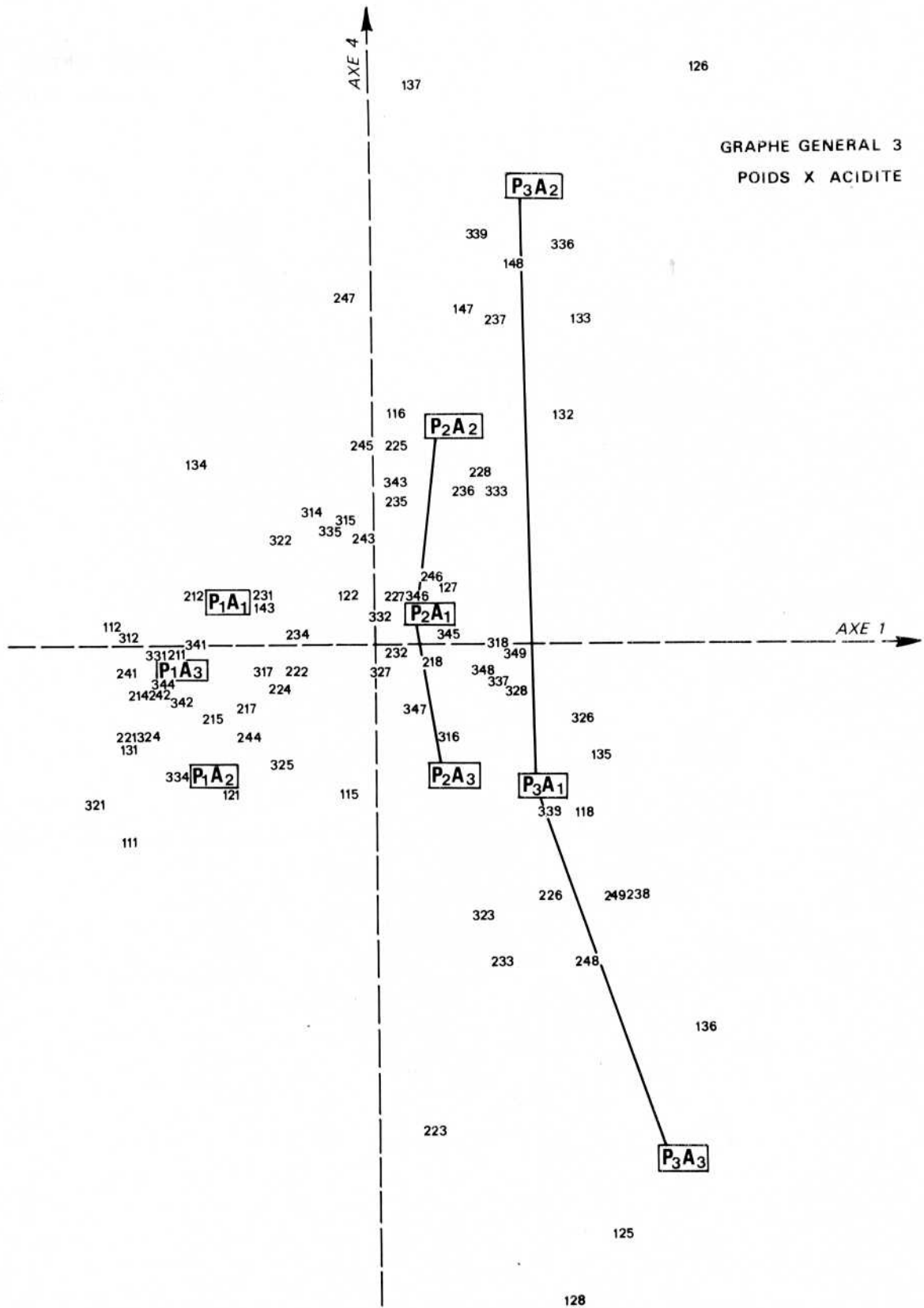
FIGURE 9 - Différentes combinaisons de hauteur x distance.

GRAPHE GENERAL 1
POIDS X ACIDITE

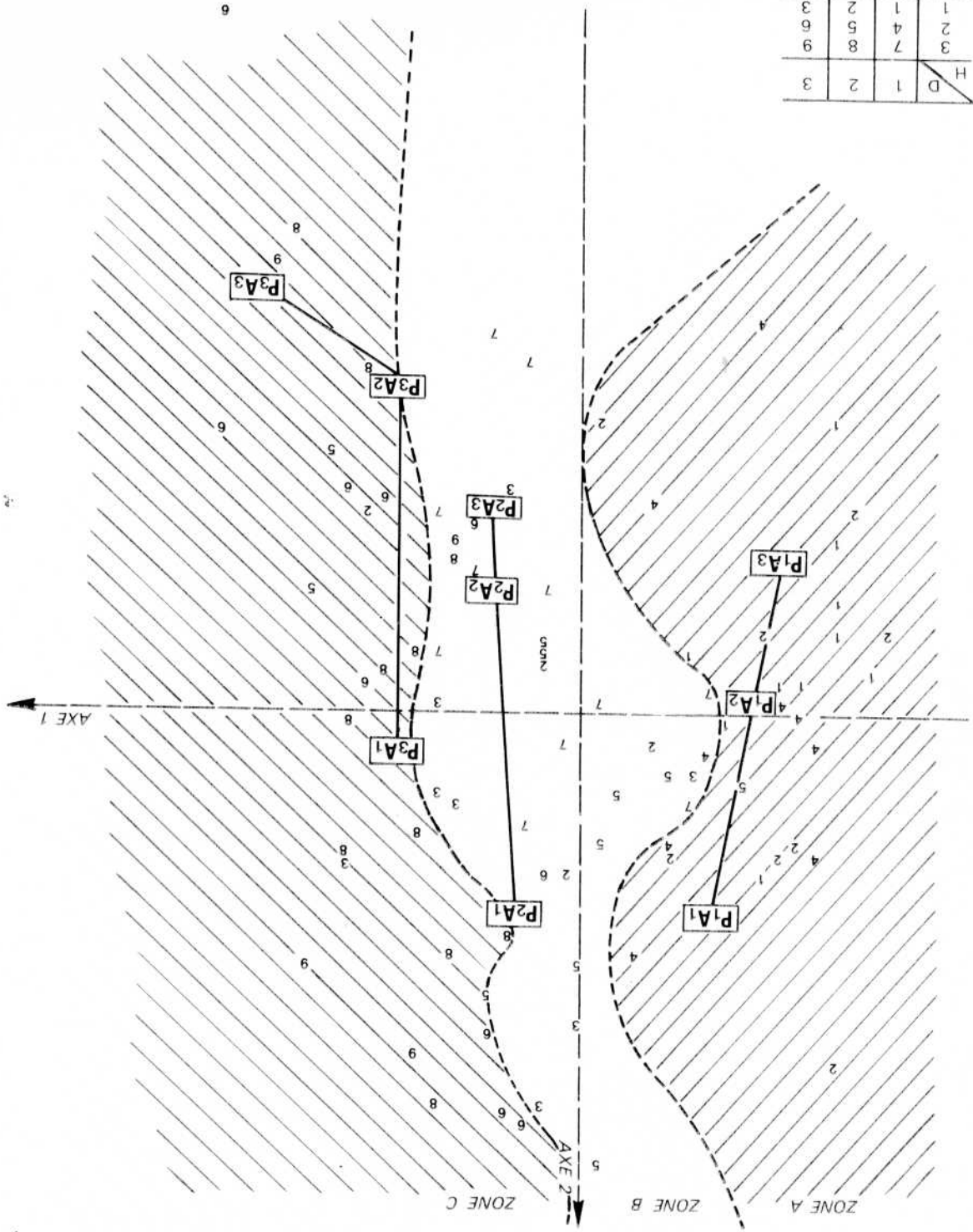


GRAPHE GENERAL 2
POIDS X ACIDITE





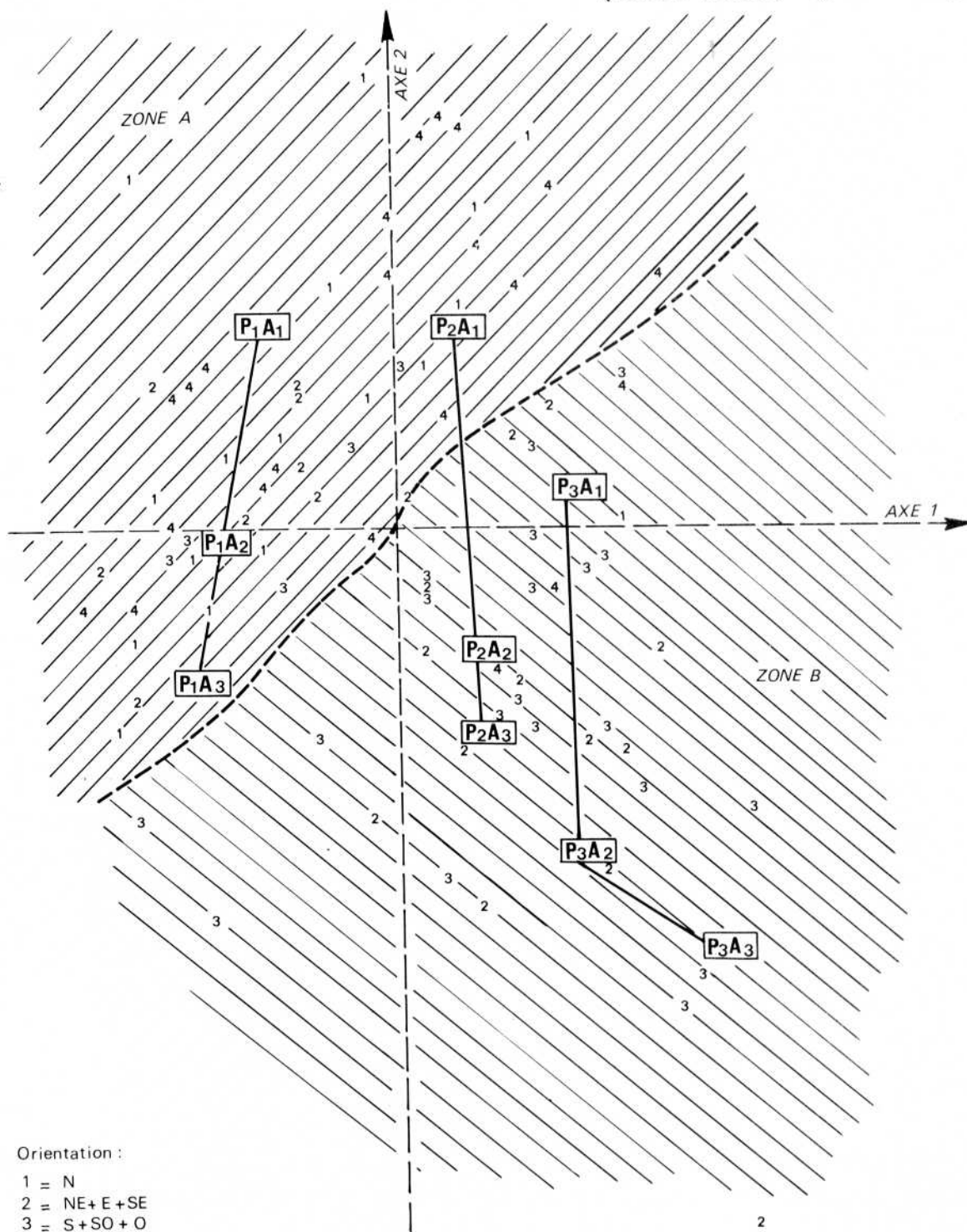
	D	1	2	3
H	3	1	2	3
	3	7	8	9
	2	4	5	6
	1	1	2	3



(POIDS x ACIDITE) x (HAUTEUR x DISTANCE)

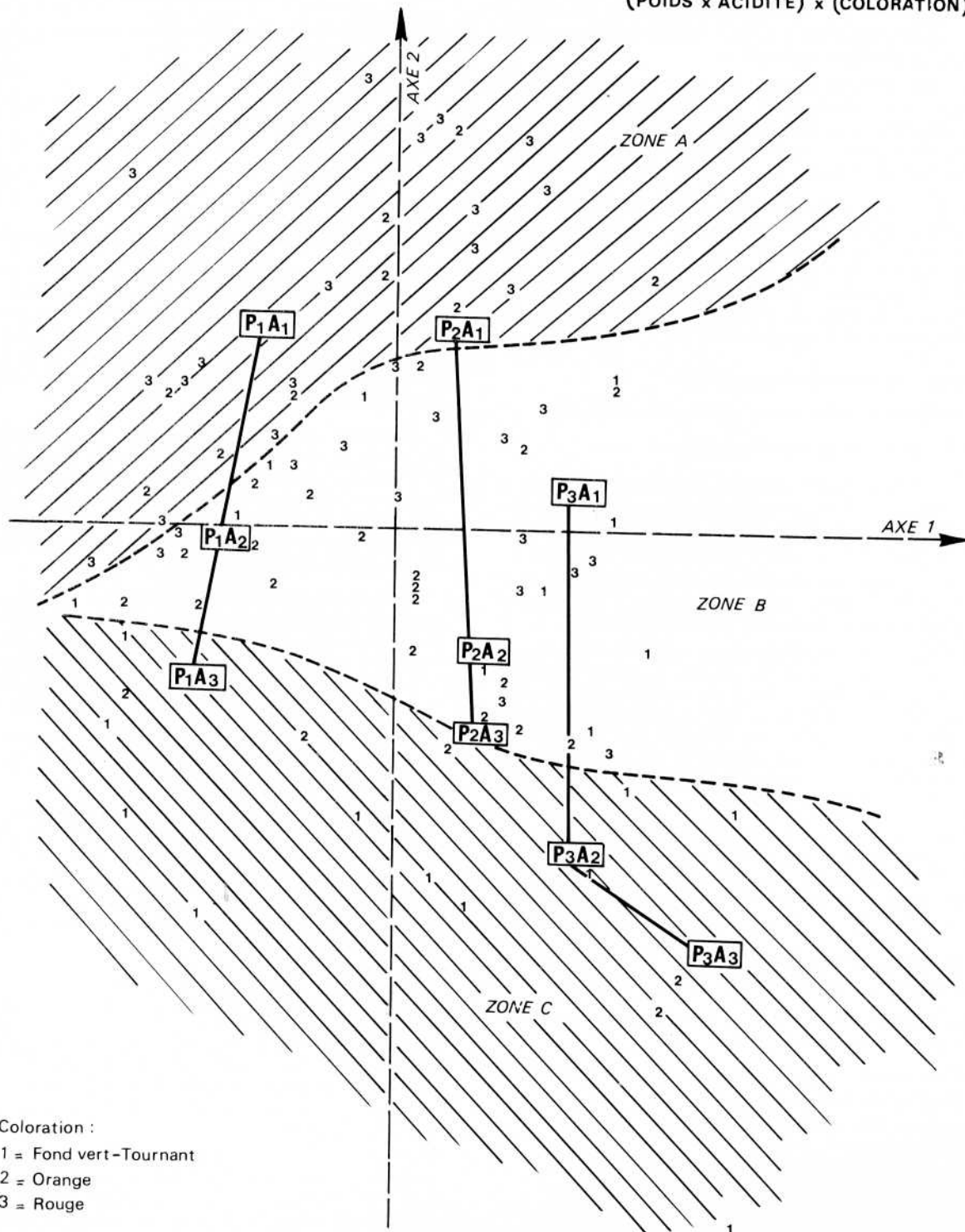
EXTRAIT DE GRAPHE 4

EXTRAIT DE GRAPHE 5
(POIDS x ACIDITE) x (ORIENTATION)



EXTRAIT DE GRAPHE 6

(POIDS x ACIDITE) x (COLORATION)



nord et au nord-ouest ; leurs fruits ont un poids faible à moyen, une acidité à tendance faible. Le couple hauteur x distance les situe principalement dans les régions 3, 6, 8 et 9, ou cône extérieur de l'arbre pour les moins acides et de poids moyens et 1, 2, 4, ou région intérieure de l'arbre pour ceux d'acidité moyenne à forte et de poids faible. La région 5 a un comportement intermédiaire.

- une zone B formée surtout des groupes situés au nord-est, au sud et à l'ouest, dont les fruits ont tendance à avoir des poids moyens et surtout forts et une acidité forte ; leurs classes de hauteur x distance ne font apparaître aucune localisation particulière.

Exception faite des fruits proches de l'axe vertical de l'arbre, l'orientation semble jouer un rôle comme source de variation de l'acidité et du poids. Ce rôle est beaucoup moins net que celui joué par la hauteur x distance pour le poids. Il se traduit par un axe privilégié nord-sud qui explique le gradient de l'acidité des clémentines situées dans le cône périphérique de l'arbre.

Analyse de l'extrait n 3 (graphe 6)

Explication du poids et de l'acidité par la coloration.

Trois zones peuvent être distinguées :

- La zone A comprend les régions géographiques à fruits les plus colorés, oranges et rouges qui ont un degré d'acidité inférieur à tous les autres. Ces fruits sont de poids faible et moyen et se situent au nord et au nord-ouest de l'arbre.

- la zone C comprend les clémentines (de coloration allant de fond vert à orange) qui sont les plus acides ; leur situation géographique est à l'opposé de la précédente.

- dans la zone intermédiaire, le phénomène intéressant réside dans le fait que les clémentines ayant une coloration forte n'ont pas obligatoirement une acidité faible ; en effet, dans cette zone, à toute coloration peut correspondre toute classe d'acidité. D'après leur position sur le graphe, il semblerait que ces fruits n'aient ni une classe de poids préférentielle, ni une orientation dominante, ni une classe de hauteur x distance particulière.

Avant de pouvoir tirer les informations essentielles de cette analyse, nous essayerons de dégager les éléments intéressants du tableau de contingence (POID x ES) x (OR) x (COL) x (HAUT x DIST) où l'extrait sec est substitué à l'acidité.

● Analyse du tableau de contingence (POID x ES) x [(COL) x (OR) x (HAUT x DIST)].

Nous conseillons au lecteur de consulter le tableau 13 des effectifs et le tableau 14 des fréquences.

Comme précédemment, nous avons effectué trois extraits du graphe général qui sont intitulés :

- analyse (poids x extrait sec) x (hauteur x distance)
(graphe 10)

Analyse générale (graphe 7)

TABLEAU 15 - Inertie expliquée par chacun des axes

axe ou facteur	inertie	p. cent d'inertie expliquée par chaque axe	p. cent cumulé
1	0,511	0,44	0,44
2	0,192	0,17	0,61
3	0,120	0,10	0,71
4	0,097	0,08	0,79
5	0,082	0,07	0,86
6	0,071	0,06	0,92
7	0,056	0,05	0,97
8	0,032	0,03	1,00
total	1,161		

Ces parts d'inertie expliquées par chacun des axes sont bien plus importantes que dans l'analyse précédente. Nous constaterons de fortes différences de profils. La chute des parts d'inertie est nette entre le premier facteur et le second, puis entre le second et le troisième. Nous analyserons les trois premiers axes qui expliquent 71 p. cent de la part d'inertie totale.

Le premier axe est principalement caractérisé par l'extrait sec, dans une moindre mesure par le poids ; la part d'inertie qu'il explique est de 0,51 contre 0,29 dans le cas de l'acidité. Il apparaît que l'interaction poids x extrait sec est beaucoup plus nette que celle poids x acidité. L'axe 1 oppose principalement P₁ES₁ aux groupes P₁ES₃, P₂ES₃, P₃ES₃ qui ont des profils différents.

Le deuxième axe est principalement un axe de poids et, dans une moindre mesure d'extrait sec ; toutefois ce second axe oppose nettement la région d'iso-poids P₁ aux régions P₂ et P₃, opposition due principalement à l'influence de l'extrait sec qui, combiné au poids, se comportera différemment selon qu'il est associé au poids P₁ ou P₂ et P₃.

Le troisième axe (graphe 8) confirme ce raisonnement en traduisant les différents comportements de l'interaction poids x extrait sec.

En effet, P₁ES₁, P₂ES₃ et P₃ES₃ se comportent de la même façon et s'opposent principalement à P₁ES₂ et P₃ES₁ et dans une moindre mesure à P₂ES₂ et P₃ES₂. Ces derniers, bien qu'occupant une place intermédiaire sur le graphe, ont un profil très proche de P₁ES₂ et P₃ES₁. L'examen des effectifs du tableau 16 de contingence permet de l'expliquer.

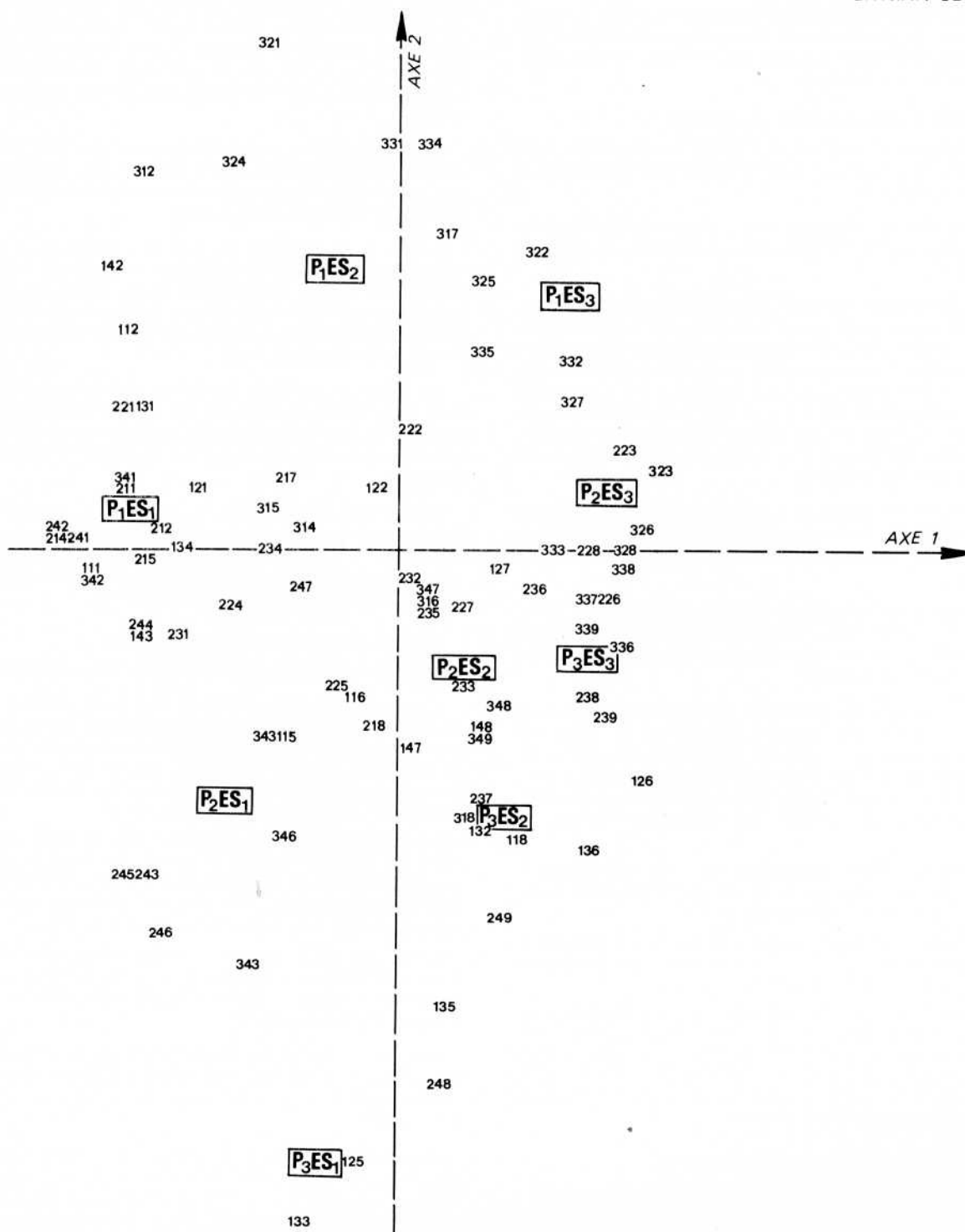
TABLEAU 16.

classes de poids classes extrait sec effectifs	P ₁			P ₂			P ₃		
	ES ₁	ES ₂	ES ₃	ES ₁	ES ₂	ES ₃	ES ₁	ES ₂	ES ₃
	321	257	156	163	246	251	34	125	230

On a vu que l'axe 1 oppose P₁ES₁ aux groupes P₁ES₃, P₂ES₃ et P₃ES₃.

Cette opposition est probablement due au critère position

GRAPHE GENERAL 7
POIDS X EXTRAIT SEC



acidités fortes sont plutôt combinées aux poids faibles et réciproquement. Dans ce plan 1 et 3, on constate un effet Guttman type, c'est-à-dire que le facteur 3 est une régression quadratique du facteur 1.

L'axe 4 (graphe 9) oppose la combinaison P₃ES₁ à toutes les autres, c'est-à-dire que la nature donne une probabilité très faible aux clémentines de posséder à la fois un poids fort et un extrait sec faible.

Analyse du premier extrait (graphe 10).

Explication du poids et de l'extrait sec des clémentines par la hauteur au-dessus du sol et la distance à l'axe vertical de l'arbre (figure 9).

Le graphe de l'analyse des correspondances peut se décomposer en trois régions correspondant à une partition de l'arbre :

Le phénomène constaté lors de l'analyse précédente montrait que le poids variait en fonction de la combinaison hauteur x distance. Cette tendance se dégage encore ici et met en relief l'interaction poids x extrait sec. Les fruits des régions 1, 4 et 2 de l'arbre, ont tendance à être caractérisés par un poids plutôt faible P₁, et dans une moindre mesure P₂. Le titre en extrait sec, associé à ce groupe de clémentines, oscille entre ES₁ et ES₂. Cette tendance est matérialisée sur le graphe par la zone A.

Les fruits des régions 6, 8 et 9 de l'arbre, tendent à être de poids P₂, P₃ et à avoir un extrait sec élevé ES₃, et dans une moindre mesure ES₂. Ils correspondent à la zone C du graphe.

La zone B est une zone de tout poids à dominante P₂ P₁ et de tout extrait sec ; elle se situe dans les régions 3, 5 et 7 de l'arbre ; elle est intermédiaire et la même que celle du graphe 4 relatif à l'acidité.

Sur l'ensemble du graphe, le phénomène d'interaction poids x extrait sec apparaît nettement ; les poids P₁ tendent à s'associer à ES₁, les poids P₃ à ES₃, ce qui se matérialise par la forte densité des groupes ayant ces combinaisons. Sans grands risques, nous pouvons donc proposer les regroupements suivants des différentes régions de l'arbre :

- régions 1, 2, 4 : dans lesquelles les fruits ont tendance à avoir un poids P₁, un extrait sec ES₁ (voire ES₂) et des niveaux d'acidité à dominante A₃.
- régions 6, 8, 9 : dans lesquelles les fruits ont tendance à avoir un poids P₂ et P₃ et un extrait sec à dominante ES₃, et une acidité associée A₁.
- régions 3, 5, 7 : dans lesquelles les fruits sont de tous poids à dominante P₂ P₁. Une acidité faible A₁ y est associée aux P₂ (voire P₃), un extrait sec ES₂ ou ES₃ l'étant à P₃ (voire P₂) ; on ne remarque pas de différenciation en ce qui concerne les niveaux d'extrait sec et d'acidité associés au poids P₁.

Analyse du deuxième extrait (graphe 11)

Explication du poids et de l'extrait sec par l'orientation.

On constate, sur le graphe, que les extraits secs les plus forts ES₃ se trouvent principalement à l'est, au sud et au sud-ouest, orientations matérialisées sur le graphe par les

chiffres 2 et 3. Les poids associés sont surtout P₃ et P₂. L'ensemble de ces groupes définit une zone A sur le graphe.

Les clémentines d'orientation nord et nord-ouest sont presque uniquement représentées dans la zone B supplémentaire de A. Leur dominante de poids est très nettement P₁, en association avec un extrait sec de tendance ES₁, voire ES₂. Dans cette zone, les deux orientations se distinguent légèrement par le fait que le nord porte des fruits ES₁ et ES₂, alors que le nord-ouest plus chargé a une prédominance ES₁.

Sur ce graphe, la région d'iso-extrait sec ES₁ et ES₂ comprend des fruits de toutes orientations. Cette remarque nous amène à rechercher une liaison entre l'orientation et la classe des hauteurs x distances.

Si l'on se rappelle que sur le graphe des hauteurs x distances les fruits P₁ES₁ et P₁ES₂ correspondent à la partie centrale de l'arbre, on en déduit, après l'étude du graphe orientation que l'on vient de faire, que les fruits P₁ES₂ et P₁ES₁ orientés au nord se trouvent surtout à l'intérieur de l'arbre. Leur acidité est forte.

De même, on a vu que les régions extérieures de l'arbre, 6, 8 et 9, étaient données par les combinaisons P₃ES₃ et P₂ES₃ ; on constate maintenant que les orientations autres que nord et nord-ouest (2 et 3 sur le graphe) y participent largement.

Enfin, dans les régions 3, 5, 7 de l'arbre, on distingue deux groupes d'orientation : l'un allant du nord-est à l'ouest par le sud avec dominante ES₃ associée à toute classe de poids ; l'autre qui lui est supplémentaire, associe un extrait sec faible ES₁ aux poids P₁ et P₂.

Analyse du troisième extrait (graphe 12).

- Explication du poids et de l'extrait sec par la coloration.

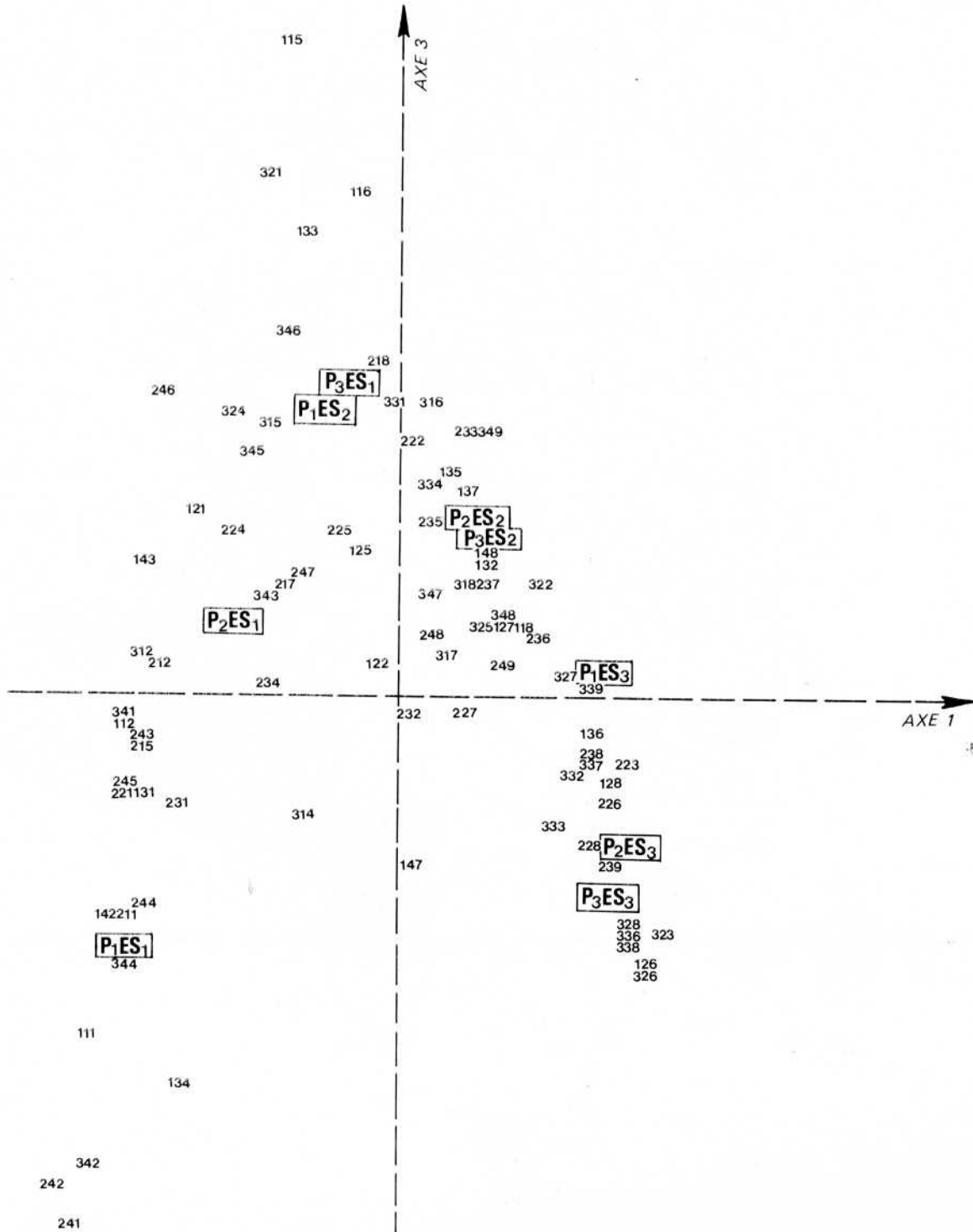
Ce graphe permet de distinguer deux zones A et B d'où il résulte que les niveaux d'extrait sec semblent avoir beaucoup plus de capacité à décrire les différentes colorations que les colorations à décrire l'extrait sec, résultat peu satisfaisant étant donné que la coloration aurait pu être une variable de contrôle de l'extrait sec. Toutefois, on note :

- qu'une probabilité non négligeable associe (zone A) aux fruits de coloration rouge un extrait sec élevé ES₃, quel qu'en soit leur poids. Ces fruits sont d'orientation autre que nord et nord-ouest ; leurs classes dominantes de hauteur x distance sont 6, 8, 9, situées en périphérie de l'arbre.
- que les colorations vertes et oranges sont surtout associées à des extraits secs de niveaux ES₁ et ES₂ (zone B) ; leurs classes de hauteur x distance dominantes sont 1, 2, 4 situées à l'intérieur de l'arbre.

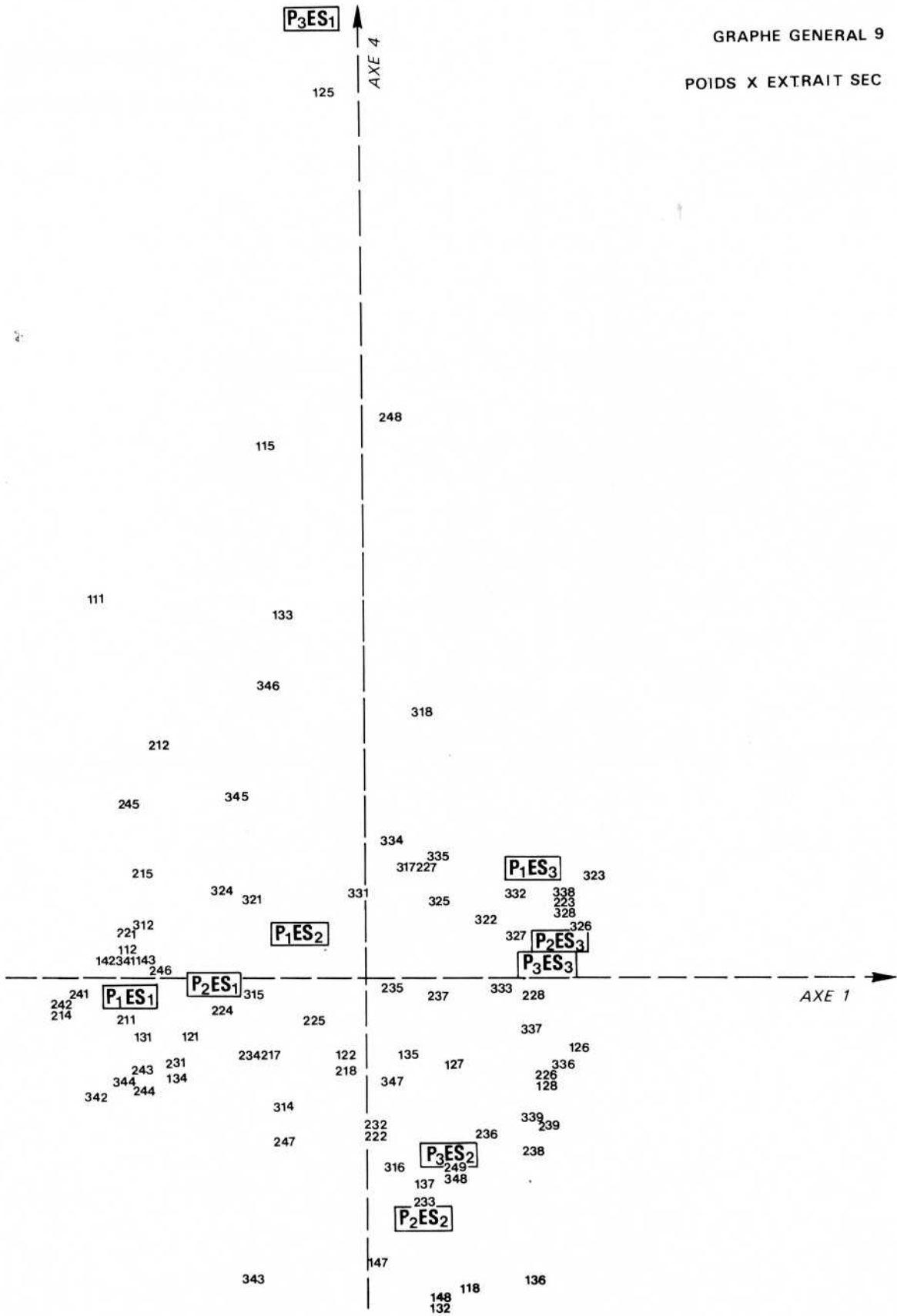
On peut penser qu'une normalisation plus stricte du code des couleurs et de leur mode d'observation améliorerait sensiblement la capacité de la coloration à décrire l'acidité et l'extrait sec.

Au cours de ces deux chapitres, nous avons tenté de décrire, comme nous nous étions proposés de le faire, les variables de production (poids, extrait sec, acidité) par les variables d'échantillonnage ou de stratification : coloration, orientation, hauteur, distance. Nous avons analysé la dispersion organisée ou tendance générale. Il nous reste à envisager

GRAPHE GENERAL 8
POIDS X EXTRAIT SEC

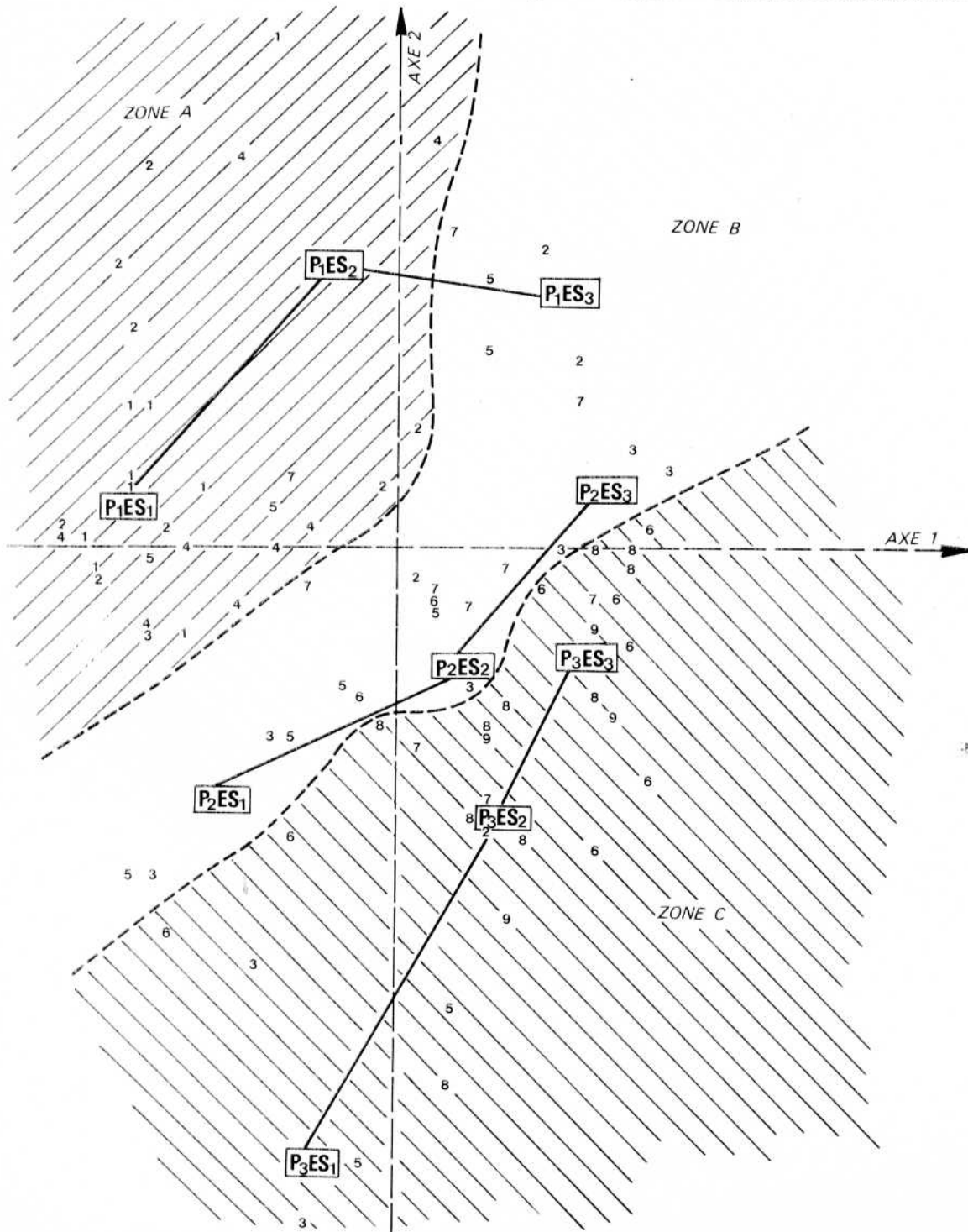


GRAPHE GENERAL 9
POIDS X EXTRAIT SEC



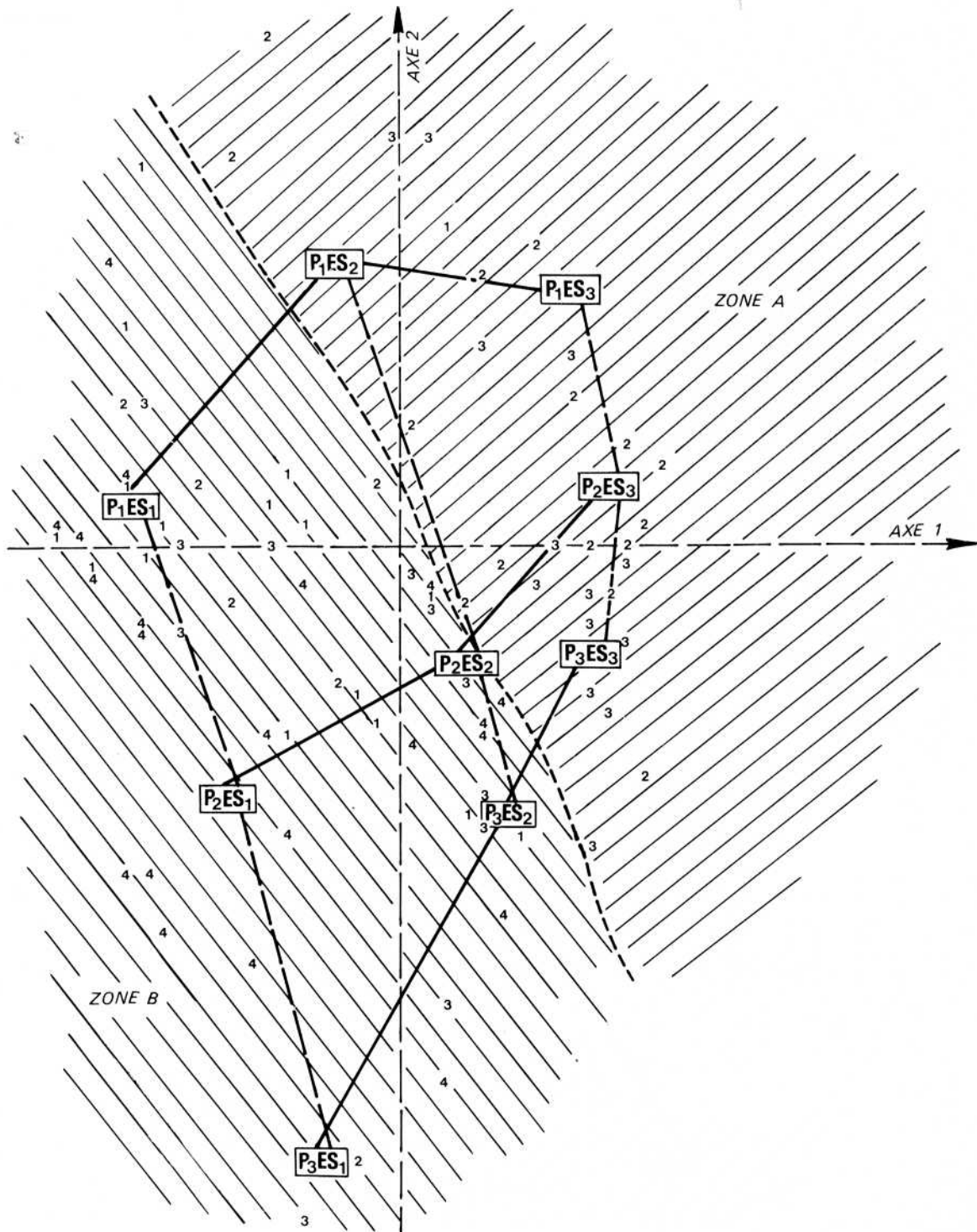
EXTRAIT DE GRAPHE 10

(POIDS x EXTRAIT SEC) x (HAUTEUR x DISTANCE)



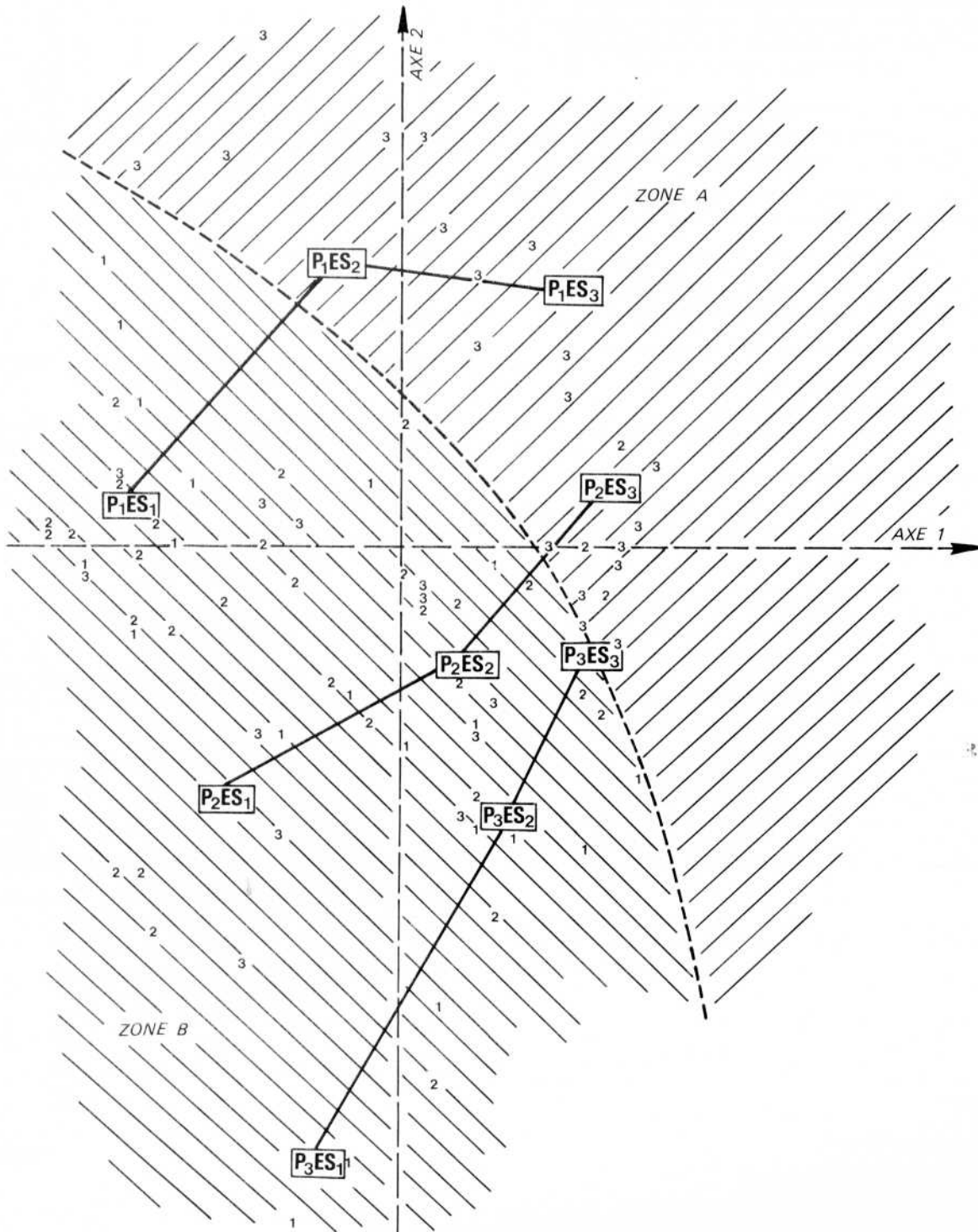
EXTRAIT DE GRAPHE 11

(POIDS x EXTRAIT SEC) x (ORIENTATION)



EXTRAIT DE GRAPHE 12

(POIDS x EXTRAIT SEC) x (COLORATION)



ger un dernier aspect, celui du pouvoir discriminant des variables d'échantillonnage.

DÉTERMINATION DU POUVOIR DISCRIMINANT SUR LE POIDS DES VARIABLES : HAUTEUR, DISTANCE, EXTRAIT SEC, ACIDITÉ, PAR LA MÉTHODE D'ANALYSE DISCRIMINANTE

● Présentation de la méthode d'analyse discriminante.

Soit un ensemble d'individus répartis en plusieurs classes (groupes de poids pour notre recherche) définies a priori.

Soit un ensemble de variables mesurées sur chacun des individus (hauteur, distance, extrait sec, acidité).

On se pose alors la question suivante :

Ces variables permettent-elles de différencier les classes définies a priori, et de les discriminer entre elles ?

Tel est le problème général de la discrimination.

En fonction de ces buts, nous différencierons :

- la discrimination à but descriptif,
- la discrimination à but décisionnel (identification)

La discrimination à but descriptif.

Étant donné différentes classes d'individus, on désire mettre en évidence le pouvoir à les discriminer de variables mesurées sur ces individus.

On est donc intéressé à tester les valeurs discriminantes de ces variables et à les comparer. En outre, on s'efforce de définir dans l'ensemble des variables retenues, le sous-ensemble permettant la meilleure discrimination entre classes.

La discrimination à but décisionnel ou d'identification.

C'est le type d'analyse auquel on se réfère le plus communément lorsqu'on parle de discrimination. Connaissant les valeurs d'un ensemble de variables mesurées sur un individu, dont on sait ou suppose qu'il appartient à une classe parmi plusieurs définies a priori, on cherche à affecter cet individu à l'une de ces classes. En d'autres termes, il s'agit d'identifier la classe de l'individu à partir de variables mesurées sur lui.

Pour caractériser ces classes, ou du moins, les différencier les unes des autres, il faut disposer pour chacune d'échantillons d'individus sur lesquels on mesure un même ensemble de variables.

La discrimination à but décisionnel est l'étape succédant nécessairement à celle de la discrimination descriptive.

Nous traiterons dans l'immédiat le problème sous l'aspect descriptif.

● Analyse du problème.

TABLEAU 17.

groupe	borne inférieure (g)	borne supérieure (g)
1	10	34
2	34	40
3	40	46
4	46	52
5	52	58
6	58	64
7	64	et plus

Variables introduites pour différencier éventuellement les groupes :

	symbole
hauteur	H
distance	D
extrait sec	ES
acidité	A

Descriptions statistiques des groupes par les variables étudiées (tableau 18 et graphique 2).

L'examen de ces différents tableaux et du graphique 2 confirme certains résultats de l'analyse des correspondances :

- le poids est une fonction croissante de la hauteur et de la distance,
- le poids est une fonction croissante de l'extrait sec,
- le poids est une fonction décroissante de l'acidité.

Résultats de l'analyse et interprétation (tableaux 19 et 20 et figure 10).

TABLEAU 19 - Inertie expliquée par chacun des axes

axe ou facteur	inertie	p. cent inertie expliquée par chacun des axes	p. cent cumulé
1	0,4092	0,893	0,893
2	0,0431	0,094	0,987
3	0,0015	0,010	0,997
4	0,0013	0,002	1,000

TABLEAU 20 - Corrélation des facteurs avec les variables

axe ou facteur	hauteur	distance	extrait sec	acidité
1	-0,549	-0,510	-0,524	0,794
2	0,357	-0,210	0,805	0,490
3	0,167	0,696	0,048	0,360
4	0,737	-0,459	-0,272	-0,007

Les inerties respectives expliquées par les premier et deuxième axes sont de 0,41 et 0,04, c'est dire que le premier axe discrimine pratiquement à lui seul les groupes

TABLEAU 18.

Poids en g	variables	effectifs	moyennes	écarts-types	coefficient de variation	variances
[10-34[GROUPE I	poids	256	27,64	4,50	16	20,2604
	hauteur	"	121	6,42	53	41,2051
	distance	"	105	3,49	33	12,1863
	extrait sec	"	10,3	7,08	7,0	50,1059
	acidité	"	1,17	13,73	12	188,5458
[34-40[GROUPE II	poids	233	37,16	1,80	4,8	32,3500
	hauteur	"	130	6,09	47	37,0721
	distance	"	110	3,39	31	11,5124
	extrait sec	"	10,2	6,25	6,1	39,0829
	acidité	"	1,10	12,56	11	157,6475
[40-46[GROUPE III	poids	245	43,26	1,70	3,9	28,8900
	hauteur	"	143	6,66	47	44,3020
	distance	"	122	3,53	29	12,4679
	extrait sec	"	10,4	6,51	6,3	42,3239
	acidité	"	1,04	13,70	13	187,5603
[46-52[GROUPE IV	poids	244	49,05	1,71	3,5	29,8210
	hauteur	"	156	6,44	41	41,5269
	distance	"	134	3,25	24	10,5627
	extrait sec	"	10,4	6,69	6,4	44,7253
	acidité	"	9,97	11,29	11	127,4752
[52-58[GROUPE V	poids	211	54,91	2,29	5,3	32,5220
	hauteur	"	170	7,20	42	51,8091
	distance	"	132	3,50	27	12,2634
	extrait sec	"	10,6	7,23	7,0	52,3068
	acidité	"	9,71	10,33	11	106,7036
[58-64[GROUPE VI	poids	255	61,77	2,29	3,7	52,6090
	hauteur	"	181	6,87	38	47,1476
	distance	"	135	3,41	25	11,5976
	extrait sec	"	10,8	7,01	6,5	49,1986
	acidité	"	9,76	9,97	10	99,4284
[≥64[GROUPE VII	poids	339	76,98	10,41	14	108,4108
	hauteur	"	192	6,38	33	40,7243
	distance	"	137	3,08	23	9,5108
	extrait sec	"	10,9	6,51	5,9	42,3247
	acidité	"	9,68	11,39	12	129,6793
Population totale	hauteur	1783	158	7,04	45	49,5874
	distance	1783	126	3,57	28	12,7449
	extrait sec	1783	10,5	7,28	6,9	52,9559
	acidité	1783	1,03	13,96	14	194,7663

Parmi ces variables la hauteur et la distance discriminent les poids de la même façon que l'extrait sec. Il apparaît, au vu de cette analyse et en observant de plus près les corrélations des variables précitées avec l'axe 2, que l'acidité et l'extrait sec suffisent à reconnaître avec une certaine probabilité l'appartenance des individus à leur classe de poids : mais ces variables ne pouvant être mesurées qu'après récolte, elles ne peuvent en aucun cas être assimilées à des variables d'échantillonnage.

Premier axe : opposition entre les groupes à faibles poids

caractérisés par une acidité forte et un extrait sec faible et les groupes à forts poids, caractérisés par une acidité faible et un extrait sec fort.

Deuxième axe : opposition entre les groupes à poids moyens (g2 à g5) caractérisés par un extrait sec et une acidité sous-moyens et les groupes de poids forts (g5 et g7) caractérisés par un extrait sec et une acidité sur-moyens, auxquels se rattache pour cet axe 2 le groupe de poids faible g1. Vu la faible part d'inertie extraite par l'axe 2, ce résultat ne contredit pas le résultat des analyses précédentes.

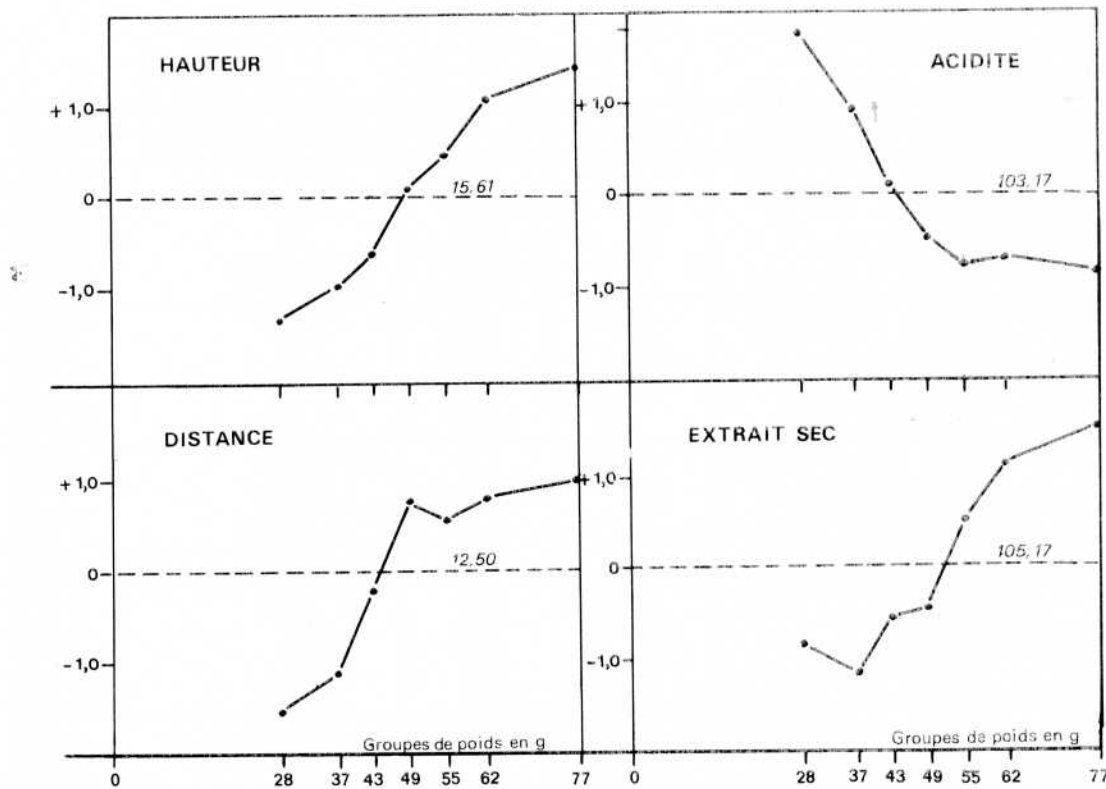
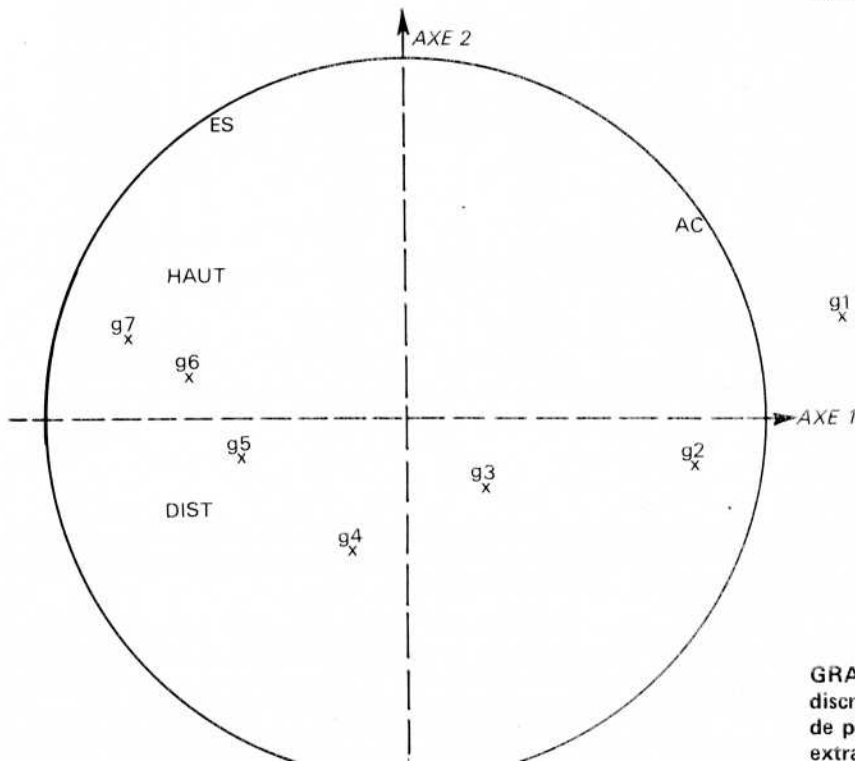


FIGURE 10 - Corrélation des variables avec les nouvelles composantes : 1 en abscisse et 2 en ordonnée.



GRAPHIQUE 2 - Variables discriminantes des groupes de poids (hauteur, distance, extrait sec et acidité, sont

CONCLUSIONS

APPROCHE D'UNE METHODE D'ECHANTILLONNAGE

Nous avons montré que les combinaisons de classes de hauteur et de distance judicieusement choisies permettaient de partitionner l'arbre en différentes régions possédant chacune certaines caractéristiques de poids, d'acidité et d'extrait sec. On a montré également que l'orientation pouvait apporter un complément d'informations sur les fluctuations de ces mêmes caractéristiques. Enfin, la coloration, bien que décrivant imparfaitement les niveaux d'extrait sec et d'acidité, peut apporter une contribution non négligeable dans le choix des clémentines à échantillonner.

Si d'une année à l'autre, les fluctuations du poids, de l'extrait sec et de l'acidité diffèrent, il n'en demeure pas moins vrai que les tendances générales ou dispersions organisées évoluent vraisemblablement de la même façon.

Aussi, serait-il intéressant de disposer d'une table donnant pour les régions de l'arbre précédemment définies la probabilité d'obtenir des clémentines de caractéristiques données. Dans cette optique, et conscients que cette partition n'est peut-être pas optimale, nous avons dressé des tableaux de répartition de fruits pour les régions géographiques suivantes :

- 1) région A formée par le couple (hauteur, distance) : 1,2,4
- 2) région B formée par le couple (hauteur, distance) : 3,5,7
- 3) région C formée par le couple (hauteur, distance) : 6,8,9

Les fruits de ces régions diffèrent nous l'avons vu, par leurs caractéristiques de poids et d'acidité.

En désignant par $P(X)$ les probabilités d'obtenir des fruits possédant la caractéristique X, nous retiendrons :

- **Toutes régions réunies :** Région A région B région C (tableaux 21 et 22). 0,15 0,13

$$P(P_1A_3) > P(P_1A_2) \geq P(P_1A_1)$$

0,19	0,11	0,11
------	------	------

$$(P_1ES_3) < P(P_1ES_2) < P(P_1ES_1)$$

0,09	0,14	0,18
------	------	------

$$P(P_2A_3) < P(P_2A_2) < P(P_2A_1)$$

0,05	0,10	0,22
------	------	------

$$(P_2ES_3) \geq P(P_2ES_2) > P(P_2ES_1)$$

0,14	0,14	0,09
------	------	------

$$P(P_3A_3) < P(P_3A_2) < P(P_3A_1)$$

0,03	0,04	0,15
------	------	------

$$(P_3ES_3) > P(P_3ES_2) > P(P_3ES_1)$$

0,13	0,07	0,02
------	------	------

Combinaisons ayant une faible probabilité d'apparaître :

- $P_2A_3 = 0,05$
- $P_3A_3, P_3A_2 = 0,03, 0,04$
- $P_3ES_1, P_3ES_2 = 0,02, 0,07$

- **Région par région (figure 11)**

Rappelons que pour tout couple d'événements incompatibles (c'est-à-dire d'intersection nulle) $(*) P(\text{événement 1} \cup \text{événement 2}) = P(\text{événement 1}) + P(\text{événement 2})$.

(*) (\cup = union \cap = intersection)

Dans les paragraphes qui vont suivre, on donnera :

- les probabilités des couples dominants

$$\text{ex : } P(P_2A_1) = 0,33$$

$$P(P_3A_1) = 0,25$$

$$P(P_2A_2) = 0,11$$

- la somme des probabilités des deux couples ayant en commun une variable de même niveau à condition que cette somme atteigne au moins 0,50

$$\text{ex : } P(P_2A_1 \cup P_3A_1) = 0,58$$

- la somme des probabilités des trois couples ayant en commun une variable de même niveau à condition que cette somme atteigne au moins 0,60

$$\text{ex : } P(P_2A_1 \cup P_2A_2 \cup P_3A_1) = 0,63$$

Dans le cas où cette somme de trois couples se situe entre 0,50 et 0,60 on donnera aussi la plus forte d'entre celles des deux autres niveaux

$$P(P_2A_1 \cup P_2A_2 \cup P_2A_3) = 0,51 \quad P(P_3) = 0,35$$

- **Région A - 1,2,4 (tableaux 21 et 22)**

poids x acidité

$$P(P_1A_3) = 0,32$$

$$P(P_1A_2) = 0,18$$

$$P(P_1A_1) = 0,14$$

$$P(P_2A_1) = 0,15$$

$$P(P_1A_3 \cup P_1A_2) = 0,50 \quad P(P_1) = 0,64$$

poids x extrait sec

$$P(P_1ES_1) = 0,36$$

$$P(P_1ES_2) = 0,22$$

$$P(P_2ES_1) = 0,10$$

$$P(P_2ES_2) = 0,10$$

$$P(P_1ES_1 \cup P_1ES_2) = 0,58$$

- **Région C - 6,8,9 (tableaux 21 et 22)**

poids x acidité

$$P(P_2A_1) = 0,33$$

$$P(P_3A_1) = 0,25$$

$$P(P_2A_2) = 0,11$$

$$P(P_2A_3) = 0,07$$

$$P(P_2) = 0,51 \quad P(P_3) = 0,35$$

$$P(P_1A_1) = 0,05$$

$$P(P_2A_1 \cup P_3A_1) = 0,58 \quad P(A_1) = 0,63$$

poids x extrait sec

$$P(P_3ES_3) = 0,26$$

$$P(P_2ES_3) = 0,26$$

$$P(P_2ES_2) = 0,19$$

$$P(P_1ES_3) = 0,07$$

$$P(P_3ES_3 \cup P_2ES_3) = 0,52 \quad P(ES_3) = 0,59$$

$$P(ES_2) = 0,32$$

- **Région B - 3,5,7 (tableaux 21 et 22)**

poids x acidité

Les probabilités d'apparition des différentes combinaisons poids x acidité sont voisines de celles de l'ensemble de l'arbre (se reporter aux tableaux).

poids x extrait sec

Les probabilités d'apparition de différentes combinaisons poids x extrait sec sont voisines de celles de l'ensemble de l'arbre.

FIGURE 11 - Dominantes poids, acidité et extrait sec, par région.

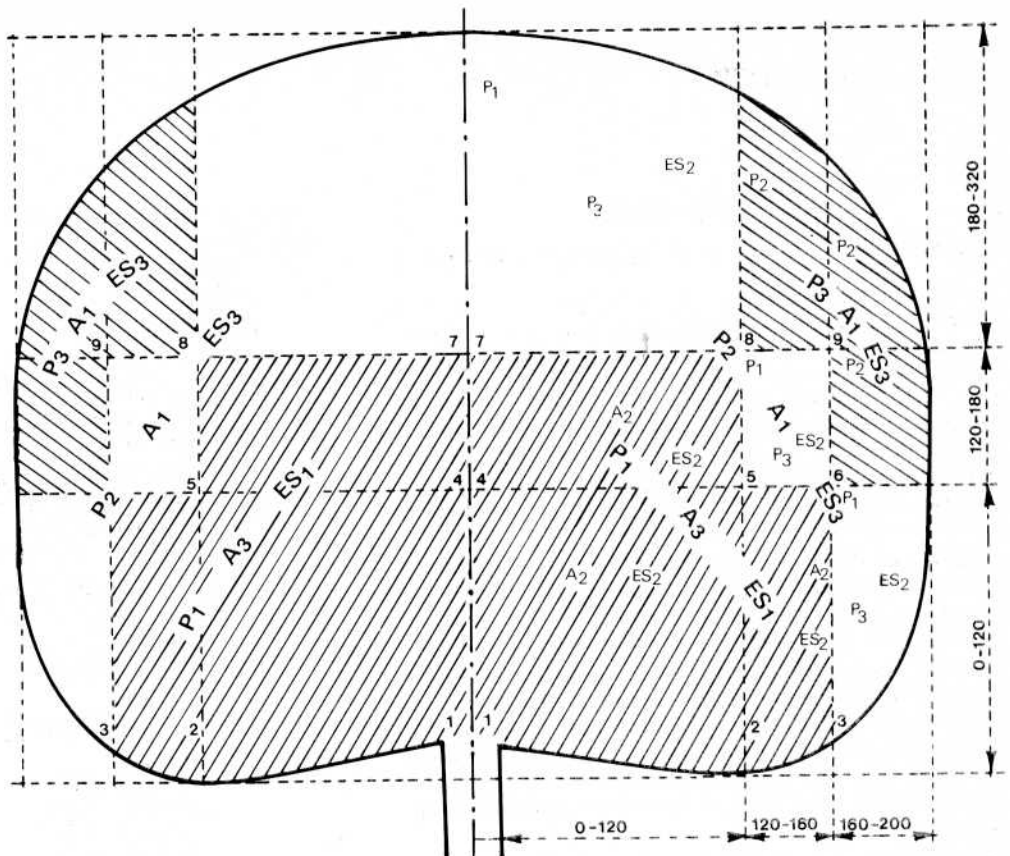


TABLEAU 21 - Répartition des clémentines par région et par classes de poids x acidité.

REGIONS		A (1 - 2 - 4)			B (3 - 5 - 7)			C (6 - 8 - 9)			Total		
		effectifs	fréquence ligne	fréquence colonne	effectifs	fréquence ligne	fréquence colonne	effectifs	fréquence ligne	fréquence colonne	ligne	col.	
P1	A1	94	0,48	0,14	83	0,42	0,13	20	0,10	0,05	197	1	0,11
	A2	119	0,64	0,18	54	0,29	0,08	14	0,08	0,03	187	1	0,11
	A3	214	0,64	0,32	101	0,30	0,15	22	0,06	0,05	337	1	0,19
P2	A1	103	0,26	0,05	151	0,38	0,23	140	0,36	0,33	394	1	0,22
	A2	50	0,30	0,08	71	0,42	0,11	48	0,28	0,11	169	1	0,10
	A3	30	0,35	0,04	26	0,31	0,04	29	0,34	0,07	85	1	0,05
P3	A1	36	0,14	0,05	120	0,46	0,18	105	0,40	0,25	261	1	0,15
	A2	17	0,22	0,03	39	0,50	0,06	22	0,28	0,05	78	1	0,04
	A3	4	0,09	0,01	18	0,41	0,03	22	0,50	0,05	44	1	0,03
Total		667	0,38	1	663	0,38	1	422	0,24	1	1752	1	1

TABLEAU 22 - Répartition des clémentines par région et par classes de poids x extrait sec

P1	ES1	238	0,75	0,36	74	0,23	0,11	5	0,02	0,01	317	1	0,18
	ES2	146	0,58	0,22	85	0,34	0,13	20	0,08	0,05	251	1	0,14
	ES3	43	0,28	0,06	79	0,52	0,12	31	0,21	0,07	153	1	0,09
P2	ES1	69	0,42	0,10	69	0,42	0,10	24	0,15	0,06	162	1	0,09
	ES2	68	0,28	0,10	89	0,37	0,13	82	0,34	0,19	239	1	0,14
	ES3	46	0,19	0,07	91	0,37	0,14	111	0,45	0,26	248	1	0,14
P3	ES1	5	0,15	0,01	22	0,65	0,03	7	0,21	0,02	34	1	0,02
	ES2	25	0,20	0,04	65	0,53	0,10	33	0,27	0,08	123	1	0,07
	ES3	27	0,12	0,04	90	0,40	0,14	110	0,48	0,26	227	1	0,13

Une conclusion importante se dégage :

La distribution, dans la zone 3,5,7, des probabilités d'obtenir les différentes combinaisons de poids x extrait sec et poids x acidité est plus proche que les deux autres de celle de l'ensemble de l'arbre.

• Région par région x orientation

- Région A - 1,2,4 (tableaux 23 et 24)

Orientation 1

poids x acidité

$$P(P_1A_3)=0,34$$

$$P(P_1A_1)=0,22$$

$$P(P_1A_2)=0,20$$

$$P(P_1A_3 \cup P_1A_1)=0,56 \quad P(P_1)=0,76$$

poids x extrait sec

$$P(P_1ES_1)=0,56$$

$$P(P_1ES_2)=0,20$$

$$P(P_2ES_1)=0,11$$

$$P(P_1)=0,76$$

$$P(ES_1)=0,70$$

Orientation 4

poids x acidité

$$P(P_1A_3)=0,39$$

$$P(P_1A_2)=0,19$$

$$P(P_1A_1)=0,21$$

$$P(P_2A_1)=0,18$$

$$P(P_1A_3 \cup P_1A_2)=0,58 \quad P(P_1)=0,79$$

poids x extrait sec

$$P(P_1ES_1)=0,68$$

$$P(P_1ES_2)=0,09$$

$$P(P_1ES_3)=0,01$$

$$P(P_2ES_1)=0,17$$

$$P(P_1ES_1 \cup P_1ES_2)=0,77 \quad P(P_1)=0,78$$

$$P(P_1ES_1 \cup P_2ES_1)=0,85 \quad P(ES_1)=0,85$$

Orientation 2

poids x acidité

$$P(P_1A_3)=0,32$$

$$P(P_1A_2)=0,23$$

$$P(P_1A_1)=0,15$$

$$P(P_2A_1)=0,15$$

$$P(P_1A_3 \cup P_1A_2)=0,55 \quad P(P_1)=0,70$$

poids x extrait sec

$$P(P_1ES_1)=0,30$$

$$P(P_1ES_2)=0,33$$

$$P(P_2ES_2)=0,11$$

$$P(P_1ES_1 \cup P_1ES_2)=0,63$$

$$P(ES_2)=0,46$$

$$P(ES_1)=0,40$$

Orientation 3

poids x acidité

$$P(P_1A_3)=0,30$$

$$P(P_2A_1)=0,15$$

$$P(P_1A_2)=0,14$$

$$P(P_1A_1)=0,11$$

$$P(P_1A_3 \cup P_1A_2)=0,44 \quad P(P_1)=0,55$$

$$P(A_1)=0,35 \quad P(A_2)=0,28 \quad P(A_3)=0,37$$

poids x extrait sec

$$P(P_1ES_1)=0,27$$

$$P(P_1ES_2)=0,19$$

$$P(P_2ES_2)=0,12$$

$$P(P_2ES_3)=0,10$$

$$P(P_1ES_1 \cup P_1ES_2)=0,46$$

$$P(ES_1)=0,36 \quad P(ES_2)=0,37 \quad P(ES_3)=0,26$$

- Région C - 6,8,9 (tableaux 25 et 26).

Orientation 1

poids x acidité

$$P(P_2A_1)=0,48$$

$$P(P_3A_1)=0,22$$

$$P(P_2A_1 \cup P_3A_1)=0,70 \quad P(A_1)=0,79$$

$$P(P_2)=0,62$$

poids x extrait sec

$$P(P_2ES_2)=0,37$$

$$P(P_3ES_3)=0,13$$

$$P(P_1ES_2)=0,13$$

$$P(P_2ES_2 \cup P_1ES_2)=0,50 \quad P(ES_2)=0,54$$

$$P(ES_3)=0,24$$

Orientation 4

poids x acidité

$$P(P_2A_1)=0,52$$

$$P(P_3A_1)=0,28$$

$$P(P_2A_1 \cup P_3A_1)=0,80 \quad P(A_1)=0,85$$

poids x extrait sec

$$P(P_2ES_2)=0,30$$

$$P(P_2ES_1)=0,15$$

$$P(P_3ES_2)=0,15$$

$$P(P_3ES_3)=0,11$$

$$P(P_2)=0,60$$

$$P(ES_2)=0,50 \quad P(ES_3)=0,29$$

Orientation 2

poids x acidité

$$P(P_2A_1)=0,26$$

$$P(P_3A_1)=0,26$$

$$P(P_2A_2)=0,12$$

$$P(P_2A_1 \cup P_3A_1)=0,52 \quad P(A_1)=0,58$$

$$P(A_2)=0,22$$

poids x extrait sec

$$P(P_2ES_3)=0,34$$

$$P(P_3ES_3)=0,34$$

$$P(P_2ES_3 \cup P_3ES_3)=0,68 \quad P(ES_3)=0,80$$

$$P(P_2)=0,46$$

$$P(P_3)=0,39$$

Orientation 3.

poids x acidité

$$P(P_3A_1)=0,23$$

$$P(P_2A_1)=0,22$$

$$P(P_2A_2)=0,16$$

$$P(P_3A_1 \cup P_2A_1)=0,45 \quad P(A_1)=0,47$$

$$P(A_2)=0,28$$

poids x extrait sec

$$P(P_3ES_3)=0,33$$

$$P(P_2ES_3)=0,32$$

$$P(P_3ES_3 \cup P_2ES_3)=0,65 \quad P(ES_3)=0,72$$

$$P(P_2)=0,49 \quad P(P_3)=0,41$$

Les orientations 1 et 4 portent des clémentines ayant des caractéristiques voisines.

Cependant les probabilités de ces fruits d'être situés au nord-ouest sont supérieures à celles d'être au nord.

TABLEAU 23 - Régions 1 - 2 - 4.

ORIENTATIONS	1		2		3		4		total	
	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne
P1A1	14	0,22	28	0,15	35	0,11	17	0,21	94	0,14
P1A2	13	0,20	44	0,23	47	0,14	15	0,19	119	0,18
P1A3	22	0,34	61	0,32	99	0,30	32	0,39	214	0,32
P2A1	8	0,12	29	0,15	51	0,15	15	0,18	103	0,15
P2A2	4	0,06	11	0,06	33	0,10	2	0,02	50	0,07
P2A3	1	0,02	9	0,05	20	0,06	0	0,00	30	0,04
P3A1	1	0,02	4	0,02	30	0,09	1	0,01	36	0,05
P3A2	1	0,02	3	0,02	13	0,04	0	0,00	17	0,03
P3A3	0	0,00	2	0,01	2	0,01	0	0,00	4	0,01
Total	64	1	191	1	330	1	81	1	667	1

TABLEAU 24 - Régions 1 - 2 - 4.

poids x extrait sec	1		2		3		4		total	
	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne
P1ES1	36	0,56	58	0,30	88	0,27	56	0,68	238	0,36
P1ES2	13	0,20	63	0,33	63	0,19	7	0,09	146	0,22
P1ES3	0	0,00	12	0,06	30	0,09	1	0,01	43	0,06
P2ES1	7	0,11	18	0,09	30	0,09	14	0,17	69	0,10
P2ES2	4	0,06	21	0,11	40	0,12	3	0,04	68	0,10
P2ES3	2	0,03	10	0,05	34	0,10	0	0,00	46	0,07
P3ES1	2	0,03	2	0,01	1	0,00	0	0,00	5	0,01
P3ES2	0	0,00	5	0,03	19	0,06	1	0,01	25	0,04
P3ES3	0	0,00	2	0,01	25	0,07	0	0,00	27	0,04
Total	64	1	191	1	330	1	82	1	667	1

TABLEAU 25 - Régions 6 - 8 - 9.

poids x acidité	1		2		3		4		total	
	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne
P1A1	4	0,09	9	0,06	2	0,02	5	0,05	20	0,05
P1A2	2	0,04	5	0,03	4	0,03	3	0,03	14	0,03
P1A3	2	0,04	9	0,06	7	0,06	4	0,04	22	0,05
P2A1	22	0,48	41	0,26	28	0,22	49	0,52	140	0,33
P2A2	3	0,07	18	0,12	21	0,16	6	0,06	48	0,11
P2A3	3	0,07	12	0,08	13	0,10	1	0,01	29	0,07
P3A1	10	0,22	40	0,26	29	0,23	26	0,28	105	0,25
P3A2	0	0,00	11	0,07	11	0,09	0	0,00	22	0,05
P3A3	0	0,00	10	0,06	12	0,09	0	0,00	22	0,05
Total	46	1	155	1	127	1	94	1	422	1

TABLEAU 26 - Régions 6 - 8 - 9.

poids x extrait sec	1		2		3		4		total	
	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne
P1ES1	1	0,02	0	0,00	0	0	4	0,04	5	0,01
P1ES2	6	0,13	5	0,03	4	0,03	5	0,05	20	0,05
P1ES3	1	0,02	18	0,12	9	0,07	3	0,03	31	0,07
P2ES1	7	0,15	2	0,01	1	0,01	14	0,15	24	0,06
P2ES2	17	0,37	17	0,11	20	0,16	28	0,30	82	0,19
P2ES3	4	0,09	52	0,34	41	0,32	14	0,15	111	0,26
P3ES1	2	0,04	1	0,00	1	0,01	3	0,03	7	0,02
P3ES2	2	0,04	8	0,05	9	0,07	14	0,15	33	0,07
P3ES3	6	0,13	52	0,34	42	0,33	10	0,11	110	0,26
Total	46	1	155	1	127	1	95	1	423	1

TABLEAU 27 - Régions 3 - 5 - 7.

ORIENTATION	1		2		3		4		Total	
	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne	effectifs	fréquence colonne
poids x acidité										
P ₁ A ₁	20	0,24	23	0,10	21	0,08	19	0,18	83	0,13
P ₁ A ₂	15	0,18	17	0,08	18	0,07	4	0,04	54	0,08
P ₁ A ₃	15	0,18	45	0,20	27	0,11	14	0,13	101	0,15
P ₂ A ₁	18	0,22	47	0,21	48	0,19	38	0,36	151	0,23
P ₂ A ₂	5	0,06	18	0,08	37	0,15	11	0,10	71	0,11
P ₂ A ₃	1	0,01	11	0,05	14	0,06	0	0,00	25	0,04
P ₃ A ₁	8	0,10	42	0,17	54	0,22	15	0,15	120	0,18
P ₃ A ₂	0	0,00	11	0,05	25	0,10	3	0,03	39	0,06
P ₃ A ₃	0	0,00	11	0,05	7	0,03	0	0,00	18	0,03
Total	82	1	225	1	251	1	105	1	663	1

TABLEAU 28 - Régions 3 - 5 - 7.

poids x extrait sec										
P ₁ ES ₁	21	0,25	17	0,07	12	0,05	24	0,23	74	0,11
P ₁ ES ₂	24	0,29	26	0,11	25	0,10	10	0,09	85	0,13
P ₁ ES ₃	5	0,06	42	0,19	29	0,12	3	0,03	79	0,12
P ₂ ES ₁	14	0,17	15	0,07	13	0,04	27	0,26	69	0,10
P ₂ ES ₂	5	0,06	25	0,11	40	0,16	19	0,18	89	0,13
P ₂ ES ₃	6	0,07	36	0,16	46	0,19	3	0,03	91	0,14
P ₃ ES ₁	2	0,02	10	0,04	7	0,03	3	0,03	22	0,03
P ₃ ES ₂	3	0,03	18	0,08	35	0,14	9	0,09	65	0,10
P ₃ ES ₃	3	0,04	36	0,16	44	0,17	7	0,06	90	0,14
Total	83	1	225	1	251	1	105	1	664	1

Les orientations 2 et 3 offrent des distributions différentes des précédentes, mais sont voisines entre elles.

Les probabilités d'apparition de chacune des caractéristiques définies pour chaque orientation, peuvent être considérées comme étant différentes de celles de la totalité de la région.

- Région B - 3,5,7, (tableaux 27 et 28).

Orientation

poids x acidité

$$P(P_1A_1) = 0,24$$

$$P(P_2A_1) = 0,22$$

$$P(P_1A_2) = 0,18$$

$$P(P_1A_3) = 0,18$$

$$P(P_1) = 0,60$$

$$P(A_1) = 0,56 \quad P(A_2) = 0,24$$

poids x extrait sec

$$P(P_1ES_2) = 0,29$$

$$P(P_1ES_1) = 0,25$$

$$P(P_2ES_1) = 0,17$$

$$P(P_1ES_2 \cup P_1ES_1) = 0,54$$

$$P(ES_1) = 0,44 \quad P(ES_2) = 0,39$$

Orientation 4

poids x acidité

$$P(P_2A_1) = 0,36$$

$$P(P_1A_1) = 0,18$$

$$P(P_3A_1) = 0,15$$

$$(P_2A_1 \cup P_1A_1) = 0,54 \quad P(A_1) = 0,69$$

poids x extrait sec

$$P(P_2ES_1) = 0,26$$

$$P(P_1ES_1) = 0,23$$

$$P(P_2ES_2) = 0,18$$

$$P(P_2ES_1 \cup P_1ES_1) = 0,49 \quad P(ES_1) = 0,52$$

$$P(P_2) = 0,47 \quad P(P_1) = 0,35$$

Orientation 2

poids x acidité

$$P(P_2A_1) = 0,21$$

$$P(P_1A_3) = 0,20$$

$$P(P_3A_1) = 0,17$$

$$P(A_1) = 0,48 \quad P(A_3) = 0,30$$

poids x extrait sec

$$P(P_1ES_3) = 0,19$$

$$P(P_2ES_3) = 0,16$$

$$P(P_3ES_3) = 0,16$$

$$P(ES_3) = 0,51 \quad P(ES_2) = 0,30$$

Orientation 3

poids x acidité

$P(P_3A_1)=0,22$

$P(P_2A_1)=0,19$

$P(P_2A_2)=0,15$

$P(P_2)=0,40$ $P(P_3)=0,35$

$P(A_1)=0,49$ $P(A_2)=0,32$

poids x extrait sec

$P(P_2ES_3)=0,19$

$P(P_2ES_2)=0,16$

$P(P_3ES_3)=0,17$

$P(P_3ES_2)=0,14$

$P(P_1ES_3)=0,12$

$P(P_1ES_2)=0,10$

$P(ES_3)=0,48$ $P(ES_2)=0,40$

La distribution des probabilités dans toutes les orientations sont différentes entre elles. Cependant nous pouvons les regrouper selon deux directions principales (figure 12).

- 1) le nord et le nord-ouest,
- 2) leur complémentaire dans la rosace des orientations

Ces regroupements correspondent à une même tendance observée des couples poids x acidité et poids x extrait sec.

L'étude détaillée des effets de l'orientation nous a montré que celle-ci jouait un rôle principalement dans le cône périphérique de l'arbre formé par les régions 6,8,9 et que ce rôle s'atténuait pour les régions plus internes (3,5,7) et (1,2,4).

De ce fait, il devra en être tenu compte dans la définition de l'échantillon.

- Probabilité de définir un niveau d'acidité et un niveau d'extrait sec par la coloration (tableau 29).

Les probabilités calculées confirment que les fruits rouges ont dans leur majorité un extrait sec fort et une acidité faible ; que les fruits verts et tournants ont dans leur majorité un extrait sec faible à moyen et une acidité fortes ; que les fruits oranges ont des distributions en classes d'acidité et d'extrait sec plus proches des fruits verts et tournants que des fruits rouges.

Au stade de la récolte, la probabilité de tirer un fruit rouge et d'extrait sec fort était de 0,26 ; celle de tirer un fruit rouge et d'acidité faible était de 0,29 ; la probabilité de tirer toute autre combinaison de couleur x extrait sec ou de couleur x acidité était au maximum de 0,17.

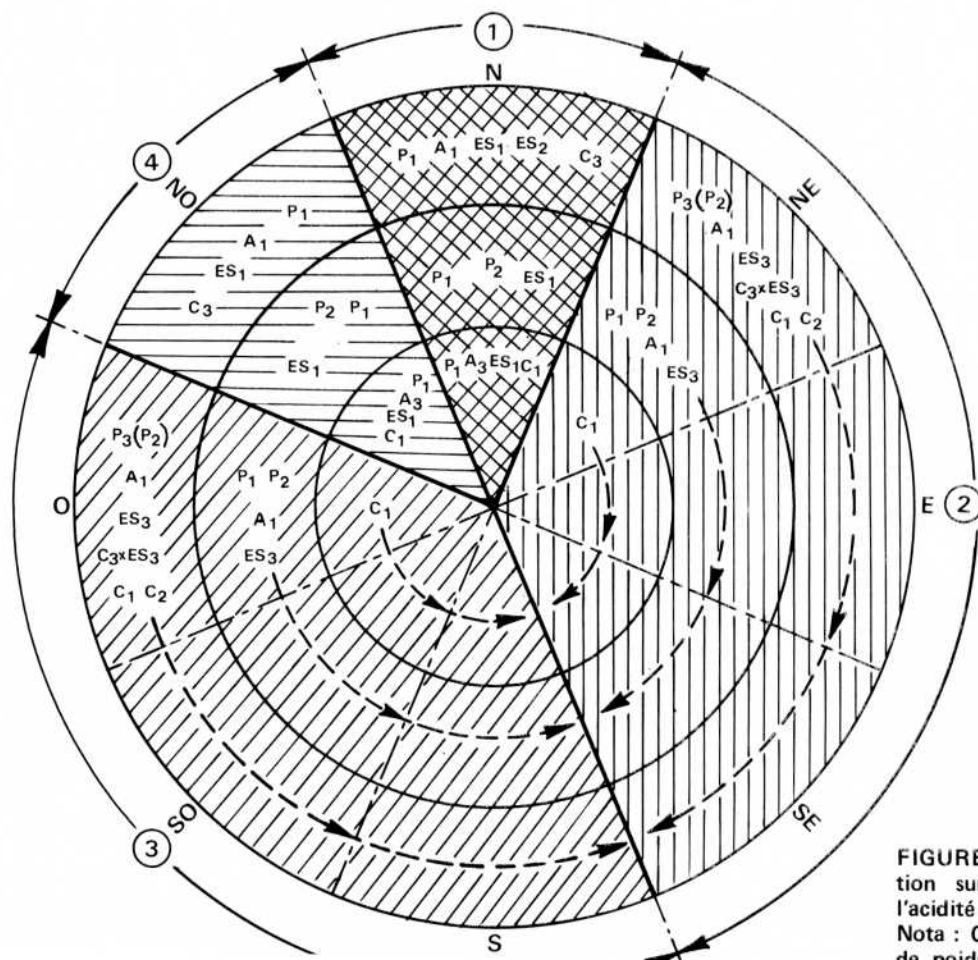


FIGURE 12 - Influence de l'orientation sur le poids, l'extrait sec, l'acidité et la coloration, par région. Nota : C3 est associé à toute classe de poids, d'orientation et de couleur.

TABLEAU 29 Probabilité de définir un niveau de AC et de ES par la coloration.

	acidité				extrait sec			
	A1	A2	A3		ES1	ES2	ES3	
vert + tournant	0,32	0,26	0,42	1,00	0,40	0,40	0,20	1,00
	0,04	0,03	0,06	0,13	0,05	0,05	0,03	0,13
orange	0,39	0,29	0,32	1,00	0,46	0,33	0,21	1,00
	0,15	0,11	0,12	0,38	0,17	0,13	0,08	0,38
rouge	0,59	0,22	0,19	1,00	0,12	0,34	0,54	1,00
	0,29	0,11	0,09	0,49	0,06	0,17	0,26	0,49

Rappelons que le tableau 29 est une image de l'arbre à un instant donné. La coloration des fruits évoluant naturellement dans le temps.

Il faudra donc être prudent quant à l'utilisation de la couleur comme variable d'échantillonnage.

Au cours de notre recherche, nous avons déterminé les tendances générales permettant de rattacher le poids, l'extrait sec, l'acidité, aux différentes positions géographiques

définies par les classes de hauteur x distance et par l'orientation.

Si la stratification nous a permis d'obtenir cette conclusion, il n'en demeure pas moins vrai que le souci d'efficacité pratique devra être préservé.

A la lumière de ces résultats, nous nous efforcerons, au cours d'une publication à paraître, de proposer une méthode d'échantillonnage du clémentinier qui précisera la position géographique des prélèvements et leur importance.

BIBLIOGRAPHIE

- LEBART (L.) et FÉNELON (J.P.)
Statistique et informatique appliquées
- SNEDECOR (G.W.) et COCHRAN (W. Gertrude)
Méthodes statistiques.
- BENZECRI (J.) et Coll.
L'analyse des données.

