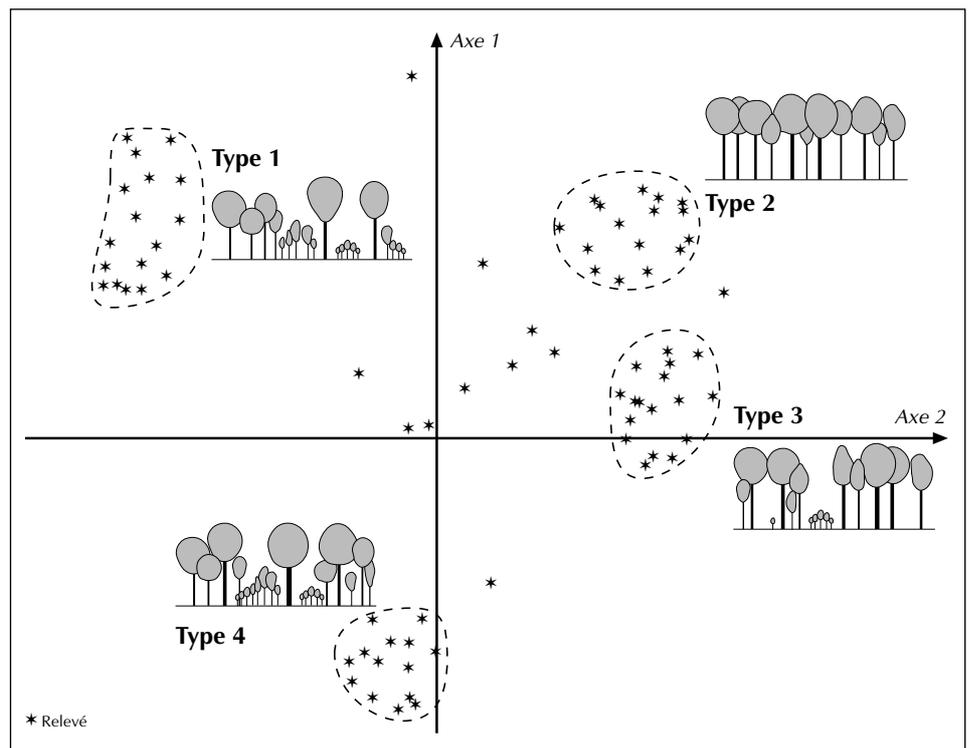


L'approche typologique et son utilité en foresterie



I. Principe et intérêt de l'approche typologique

I.1. Quelques exemples de typologie	1
<i>I.1.1. Les stades de futaie régulière</i>	1
<i>I.1.2. La typologie des peuplements de futaie jardinée de Herbert et Rébeiro (1981)</i>	2
<i>I.1.3. Le catalogue des types de stations de la Plaine de la Saône (Bailly, 1995)</i>	3
<i>I.1.4. Synthèse concernant ces exemples</i>	3
I.2. Les principes typologiques	4
<i>I.2.1. La simplification d'une réalité complexe</i>	4
<i>I.2.2. Principes d'élaboration des typologies</i>	4
<i>I.2.3. Les éléments d'une typologie</i>	6
<i>I.2.4. Les limites de l'outil typologique</i>	6

II. Les méthodes de création des typologies

II.1. Les méthodes anciennes	7
<i>II.1.1. Méthodes manuelles</i>	7
<i>II.1.2. Méthodes utilisant des indices</i>	9
II.1.2.1. Le coefficient de communauté floristique (distance de Jaccard)	9
II.1.2.2. L'analyse différentielle de Czekanowski	9
II.2. Les méthodes informatiques modernes	10
<i>II.2.1. Les analyses factorielles</i>	10
II.2.1.1. Fondements théoriques des techniques utilisées	10
II.2.1.1.1. Approche des techniques d'analyses factorielles à l'aide d'un problème en deux dimensions	10
II.2.1.1.2. Généralisation à plusieurs dimensions	12
II.2.1.1.3. D'une lecture simple à une lecture élaborée...	13
II.2.1.2. Quelques exemples d'analyses factorielles	14
II.2.1.2.1. L'analyse en composantes principales (ACP)	14
II.2.1.2.2. L'analyse factorielle des correspondances (AFC)	14
<i>II.2.2. Les classifications automatiques</i>	14
II.2.2.1. Principe et intérêt	14
II.2.2.2. Les classifications non-hiérarchiques	15
II.2.2.3. Les classifications hiérarchiques	15
<i>II.2.3. Méthodologie générale en analyse des données</i>	16
II.2.3.1. Les dix points chauds de l'analyse des données (d'après Fénelon, 1981)	16
II.2.3.2. Les limites de l'analyse des données	19

Conclusion	20
-------------------	-----------

Bibliographie	20
----------------------	-----------

L'approche typologique et son utilité en foresterie

Objectif :

Etre capable d'appréhender une démarche de typologie.

De nos jours, les forestiers utilisent couramment des outils typologiques que ce soit pour les études de stations ou pour la gestion des peuplements. Même si l'on ne cherche pas à ce que chaque titulaire d'un BTSA Gestion Forestière sache réaliser une typologie en sortant de formation, il est bon de connaître quelques rudiments concernant l'élaboration des typologies pour utiliser au mieux ces outils et bien connaître leurs intérêts et leurs limites.

I. Principe et intérêt de l'approche typologique

On se propose ici de chercher à connaître les principaux traits communs des typologies à l'aide de l'étude de quelques exemples intéressants les forestiers.

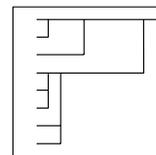
I.1. Quelques exemples de typologie

1.1.1. Les stades de futaie régulière

Un exemple de typologie très simple peut être donné par les stades de futaie régulière. En effet, un peuplement de futaie évolue de façon continue dans le temps et on a cherché à diviser cette évolution en stades afin d'avoir des repères et un langage de description commun.

On trouve généralement pour définir les stades de la futaie régulière des descriptions voisines de celles qui suivent :

- **Semis** : les arbres ont de 0 à 1 m ;
- **Fourré** : les arbres ont de 1 à 3 m ;
- **Gaulis** : les arbres ont de 3 à 8 m et un diamètre dominant inférieur à 10 cm, les tiges y sont encore flexibles ;
- **Bas-perchis** : les arbres ont de 8 à 15 m de hauteur et un diamètre dominant compris entre 10 et 20 centimètres ;



- **Haut-perchis** : les arbres ont de 15 à 25 m et un diamètre dominant compris entre 20 et 30 cm ;
- **Jeune futaie** : la croissance des arbres en hauteur est beaucoup plus lente qu'au stade perchis, ils ont un diamètre dominant supérieur à 30 cm ;
- **Futaie adulte** : le diamètre dominant est supérieur à 45 cm, ce stade correspond pour certaines essences à l'aptitude à fructifier ;
- **Vieille futaie** : ce stade est caractérisé par l'état végétatif des arbres qui commencent à dépérir.

1.1.2. La typologie des peuplements de futaie jardinée de Herbert et Rébeiro (1981)

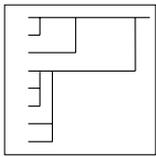
Des typologies de peuplements ont été présentées dans le cadre des cours de sylviculture et d'aménagement. La plus ancienne de ces typologies a été construite par Herbert et Rébeiro (1981) et concerne les peuplements résineux jardinés de la Haute-Chaîne du Jura.

Une présentation simplifiée de cette typologie (Allégrini, 1981) a été réalisée par le CRPF de Franche-Comté et a été distribuée lors d'une sortie en futaie jardinée.

On constate que :

- Les peuplements d'allure jardinée ont été classés en **divers types** :
 - A, B et C pour les peuplements jardinés,
 - D, E et F pour les peuplements régularisés,
 - G pour les peuplements à capital résineux très faible et envahis de feuillus.
- Pour chaque type de peuplement, il existe une **fiche descriptive** résumant les principales caractéristiques sylvicoles et dendrométriques.
- Une **clef de détermination** permet à la suite d'une visite de terrain ou d'un inventaire de déterminer les types de peuplements présents.
- Un **diagramme d'évolution** naturelle entre types présente les évolutions possibles entre types de peuplements en cas de non-gestion.
- Des conseils concernant la **sylviculture et la gestion** sont également fournis par type de peuplement (règles concernant les travaux, prélèvements...).

La typologie présentée est un outil de gestion intégré. Elle permet en effet de nommer efficacement les peuplements grâce à un langage commun de description, de les décrire et de donner leurs principales caractéristiques. Elle permet aussi de planifier la gestion en donnant les grandes orientations sylvicoles possibles et en anticipant les évolutions potentielles.



1.1.3. Le catalogue des types de stations de la Plaine de la Saône (Bailly, 1995)

Outre des données générales (cadre géographique, climat, géologie, types de sols...), ce document présente les types de stations forestières de la région naturelle de la Plaine de la Saône.

Les types de stations sont construits en fonction :

- de données **géologiques** et **géomorphologiques**,
- de données **pédologiques**,
- de données **floristiques**.

Ainsi, une station est une étendue relativement homogène en fonction des données citées ci-dessus. Le catalogue des types de stations présente toutes les stations potentielles de la région naturelle choisie. Pour chaque type, on résume les éléments caractéristiques du type (groupes floristiques rencontrés, types de sols, forme d'humus, position topographique...) et on donne les facteurs favorables et les contraintes pour la végétation forestière. Cela permet de donner la liste des essences à favoriser dans le cadre de la gestion, ou à introduire lors d'un reboisement.

Une clef de détermination fournie dans ce document permet, sur le terrain, de déterminer le type de station sur lequel on se trouve.

1.1.4. Synthèse concernant ces exemples

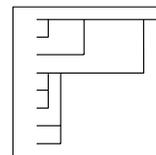
Ces quelques exemples permettent de dégager quelques idées générales concernant les typologies :

- Créer une typologie revient à faire des *groupes homogènes* pour clarifier ou simplifier une réalité complexe.

- La typologie permet une approche plus facile de cette *réalité complexe* en donnant notamment un *langage de description commun*. Ainsi, on ne dit plus « *je suis en présence d'un peuplement plus ou moins jardiné dans lequel on a majoritairement des bois moyens* », mais « *je suis en présence d'un type E* ». On en sait plus en parlant d'un type E qu'en faisant de longues phrases car celui-ci est clairement défini. De plus, si la typologie est bien construite, toutes les personnes ayant affaire à ces peuplements les appelleront des peuplements de type E. Sans typologie, il est difficile de comparer plusieurs peuplements entre eux.

- Dans certains cas, des *clefs de détermination* sont fournies avec les typologies. Elles permettent de classer de nouveaux individus.

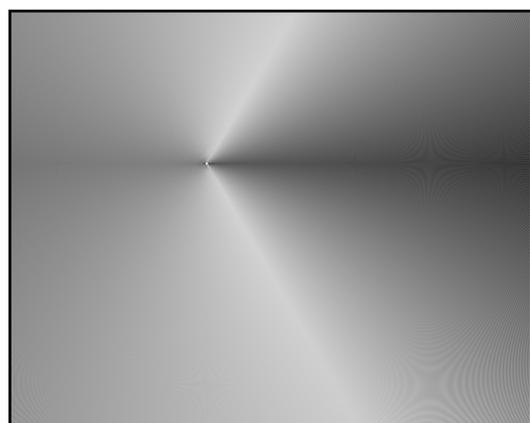
Les typologies sont utilisées dans de très nombreux domaines. On en trouve par exemple en sciences humaines, dans des traitements d'enquêtes, en économie... Elles sont également utilisées dans le cadre de la gestion forestière, notamment en **phytosociologie**, en **aménagement forestier** (typologie de peuplements) ou bien pour construire des **catalogues de stations**.



I.2. Les principes typologiques

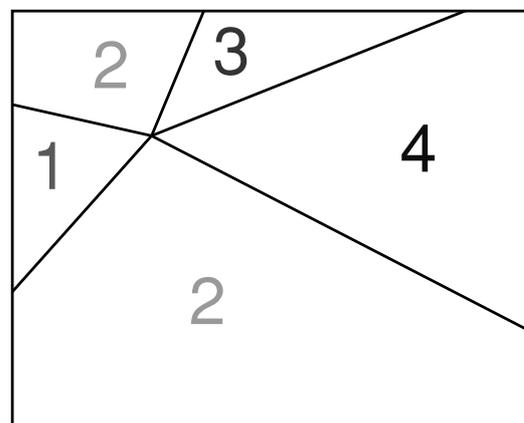
I.2.1. La simplification d'une réalité complexe

Dans la majorité des cas, les typologies correspondent à la simplification et à l'interprétation d'une réalité complexe. Le schéma ci-dessous illustre ceci.



Ensemble possédant une réalité complexe

Création de
la typologie



Simplification de cette réalité par la typologie

La typologie permet ainsi de se repérer plus facilement. Elle fait souvent la synthèse de nombreuses variables (par exemple volume, surface terrière, densité, répartition en PB, BM et GB, composition en essences... dans le cas d'une typologie de peuplements). Elle transforme une réalité dans laquelle les variations sont continues en variables synthétiques discrètes que sont les types.

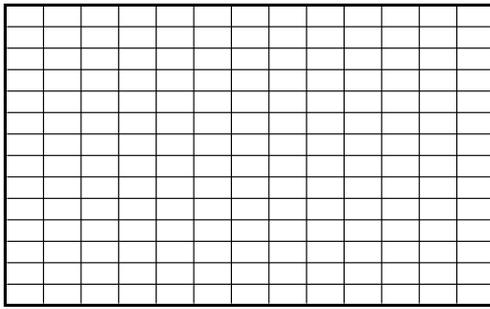
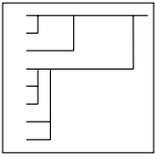
René Doussot (1990) explique cela en prenant l'exemple des typologies de peuplements : « *Créer une typologie de peuplements, c'est réunir dans un effort de synthèse, sous une même appellation, des peuplements ayant en commun certaines caractéristiques jugées déterminantes en ce qui concerne à la fois les objectifs à leur assigner à long terme et les règles sylvicoles à leur appliquer dans le présent* ».

I.2.2. Principes d'élaboration des typologies

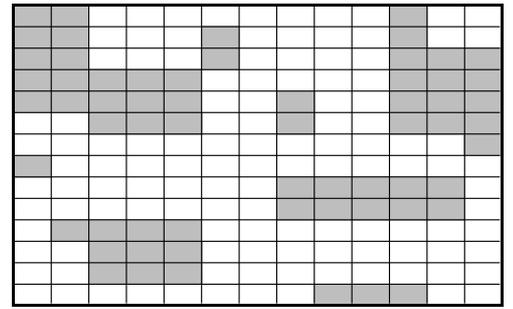
Les méthodes d'élaboration des typologies sont assez nombreuses. On distingue des méthodes « manuelles » et des méthodes statistiques utilisant l'ordinateur. Les progrès en efficacité et en convivialité de l'informatique font que les méthodes utilisant l'ordinateur sont de plus en plus accessibles.

On distingue également deux grands types de typologies :

- Les **typologies *a priori***, construites en essayant à l'avance de prévoir tous les types possibles et en vérifiant l'existence de ces types potentiels sur le terrain.



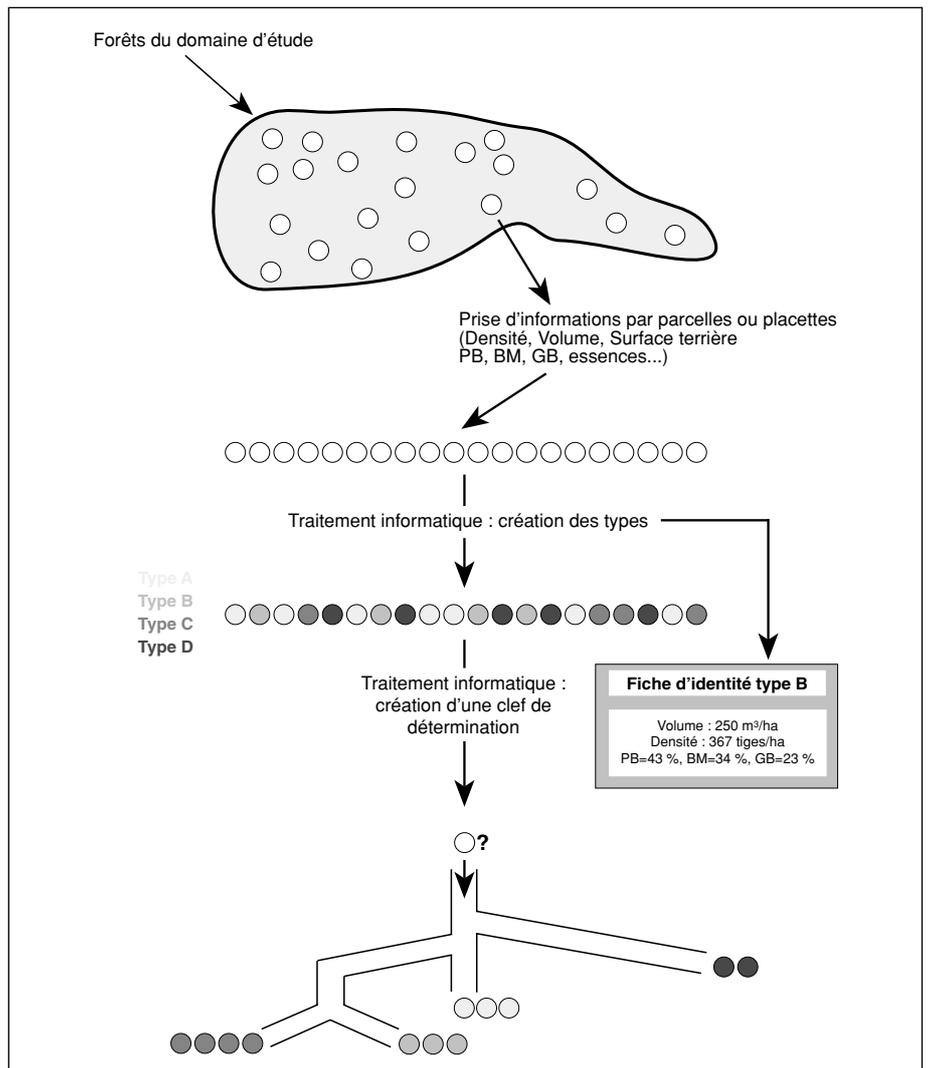
Création de tous les types possibles

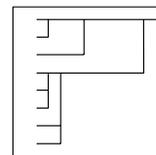


Vérification sur le terrain de l'existence des types (les types existant réellement sont grisés)

Ces typologies sont assez rares et utilisées surtout pour la typologie des peuplements. On procède le plus souvent ainsi, quand il n'est pas possible matériellement ou financièrement de procéder à un échantillonnage de la zone à étudier.

- Les **typologies construites à partir d'un échantillonnage**. Dans ce cas, on sonde un grand nombre de points d'échantillonnage et on construit la typologie à partir de groupes de placettes ayant suffisamment de points communs. Le schéma suivant résume cette démarche, dans le cas de la construction d'une typologie de peuplements.





1.2.3. Les éléments d'une typologie

Une typologie présente au minimum une *fiche de description pour chaque type*. Elle permet de connaître les principales caractéristiques du type et de comparer les types entre eux.

Très souvent, on trouve également une *clef de détermination* des types. Cette dernière permet d'associer un type à un relevé effectué en forêt. C'est elle qui permet de rendre opérationnelle la typologie.

Parfois, des éléments concernant la *dynamique des types* sont également fournis. Ils permettent de savoir quelles sont les évolutions possibles entre types. On trouve ce genre de données dans certaines typologies de peuplements.

1.2.4. Les limites de l'outil typologique

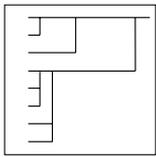
- Une première limite concerne la **zone de validité géographique** de la typologie employée. Les régions naturelles forestières sont très nombreuses et la majorité des typologies utilisées en forêt (typologies de peuplements ou de stations) sont construites à l'échelle d'une ou de quelques régions naturelles. On peut parfois utiliser avec prudence une typologie construite pour une région donnée dans une région semblable et voisine, mais le plus souvent la précision des résultats est alors moins bonne. Ainsi, une typologie est construite pour une zone géographique donnée et cela constitue une première limite pour ce genre d'outil.

- Une seconde limite est liée à la définition même des typologies. En tant qu'outils qui simplifient une réalité complexe, elles entraînent une **perte d'information et de précision**. Cela dit, il vaut mieux pouvoir appréhender une réalité complexe, même en la simplifiant un peu que de chercher à tout percevoir dans les moindres détails et n'aboutir à rien.

- Une troisième limite de l'outil typologique est liée à la **qualité de sa construction**. En effet, il n'existe jamais pour un problème donné une seule typologie, mais plusieurs possibles. Il faut savoir choisir correctement les types pour que la typologie soit correcte et efficace. Il faut faire un découpage le plus juste possible avec les critères les plus pertinents.

De plus, même avec un bon choix des types, il peut toujours exister une confusion entre les types proches d'où des incertitudes liées à la détermination des types sur le terrain. Cette incertitude est augmentée quand les clefs de détermination ne sont pas bien construites ou n'utilisent pas les bonnes variables.

- La quatrième et dernière limite concerne l'**utilisateur de la typologie**. Plus ce dernier est compétent, plus la typologie sera efficace. Il apparaît évident qu'une personne n'ayant pas reçu un enseignement de base en pédologie et en botanique aura du mal à utiliser à bon escient un catalogue de stations.



Ainsi plus l'utilisateur est entraîné et compétent, plus les résultats qu'on peut attendre de la typologie seront bons. Un utilisateur chevronné devra même être capable de définir les limites de l'outil qu'il utilise afin de mettre en place une stratégie optimale d'utilisation.

II. Les méthodes de création des typologies

II.1. Les méthodes anciennes

Depuis longtemps, les écologues ont cherché à construire des typologies et cela bien avant la possibilité d'utiliser l'outil informatique. Par exemple, toute personne qui s'intéresse à la botanique finit par constater que certaines plantes se trouvent souvent ensemble et sur certains milieux. On a ainsi cherché à décrire et classifier les associations végétales ; c'est le domaine de la *phytosociologie*.

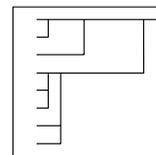
On verra donc dans ce chapitre quelles sont les méthodes qui ont été utilisées avant l'outil informatique, notamment par les phytosociologues.

II.1.1. Méthodes manuelles

Imaginons que l'on ait effectué plusieurs relevés de végétation dans des endroits différents. Si l'on nomme R1, R2, R3... les différents relevés, on peut obtenir un tableau synthétisant les différentes informations. Sur ce tableau, un point dans une case indique que la plante est présente dans le relevé concerné.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
Anemone nemorosa			•		•		•					•
Clematis vitalba								•	•	•		
Cytisus scoparius	•			•		•					•	
Dicranum scoparium				•		•						
Fissidens taxifolius			•		•							
Glechoma hederacea							•					•
Lathyrus vernus									•	•		
Lonicera xylosteum		•						•		•		
Luzula sylvatica	•					•					•	
Sambucus nigra			•				•					•
Stachys sylvatica					•		•					
Teucrium scorodonia	•			•							•	

Même si ce tableau n'est qu'une simplification de ce que peut être un tableau phytosociologique, on constate qu'il est difficile de dire quels sont les relevés qui se ressemblent et quelles sont les plantes que l'on trouve souvent ensemble. Pour mieux visualiser cela, on peut permuter lignes et colonnes pour rapprocher les plantes qui vont ensemble (en travaillant sur les lignes) et les relevés qui se ressemblent (en travaillant sur les colonnes). Si l'on prend soin de bien manipuler une ligne entière ou une



colonne entière, on ne change rien aux données initiales. On ne fait que les présenter d'une autre façon.

Une telle manipulation sur les colonnes donne les résultats suivants :

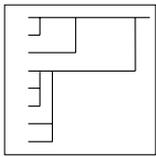
	R1	R6	R11	R4	R5	R12	R7	R3	R9	R10	R2	R8
Anemone nemorosa								
Clematis vitalba									.	.		.
Cytisus scoparius								
Dicranum scoparium		.		.								
Fissidens taxifolius					.			.				
Glechoma hederacea						.	.					
Lathyrus vernus									.	.		
Lonicera xylosteum										.	.	.
Luzula sylvatica	.	.	.									
Sambucus nigra						.	.	.				
Stachys sylvatica					.		.					
Teucrium scorodonia	.		.	.								

Si l'on déplace ensuite les lignes, on obtient :

	R1	R6	R11	R4	R5	R12	R7	R3	R9	R10	R2	R8
Cytisus scoparius								
Dicranum scoparium								
Luzula sylvatica								
Teucrium scorodonia								
Anemone nemorosa								
Fissidens taxifolius					.			.				
Glechoma hederacea						.	.					
Sambucus nigra						.	.	.				
Stachys sylvatica					.		.					
Clematis vitalba									.	.		.
Lathyrus vernus									.	.		
Lonicera xylosteum										.	.	.

Sur cet exemple simplifié, on constate que certaines cases ont été grisées. Cela signifie qu'il s'agit de *zones homogènes*. On peut ainsi dire que les relevés R1, R4, R6 et R11 se ressemblent car ils possèdent des plantes communes qu'on ne trouve pas dans les autres relevés. De même, on peut dire que *Cytisus scoparius*, *Dicranum scoparium*, *Luzula sylvatica* et *Teucrium scorodonia* sont des plantes que l'on trouve souvent ensemble.

L'opération qui a été faite s'appelle une *diagonalisation*. Sur des cas concrets, elle est plus difficile à effectuer et l'on n'obtient pas de blocs parfaits. Autrefois, on diagonalisait à la main (en découpant le tableau en bandelettes de papier). Maintenant, on peut utiliser un tableur pour faciliter cette opération.



II.1.2. Méthodes utilisant des indices

II.1.2.1. Le coefficient de communauté floristique (distance de Jaccard)

On peut utiliser des indices pour voir si deux relevés se ressemblent. Par exemple, on définit le coefficient de communauté floristique (ou distance de Jaccard) de la façon suivante :

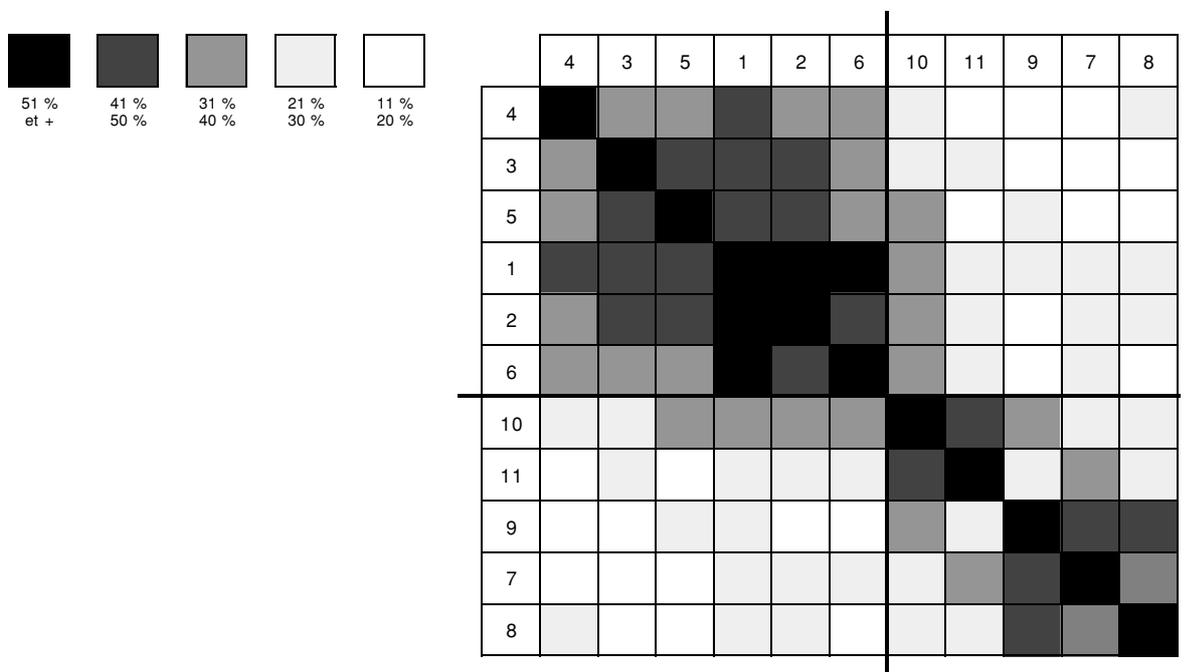
$$I_{AB} = \frac{\text{Card} (A \cap B)}{\text{Card} (A \cup B)}$$

où card (A ∩ B) est le nombre de plantes communes aux relevés A et B et card (A ∪ B) est le nombre d'espèces présentes dans le relevé A ou dans le relevé B.

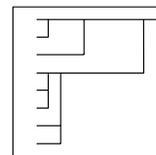
Ce coefficient varie entre 0 et 1 (ou entre 0 et 100 %). Plus il est élevé, plus les deux relevés ont de ressemblances.

II.1.2.2. L'analyse différentielle de Czekanowski

L'utilisation de la distance de Jaccard peut permettre de faire une analyse différentielle de Czekanowski. On répartit d'abord les coefficients en classes. Ensuite, on trace un tableau dans lequel on représente les relevés en ligne et en colonne. A chaque intersection, on fait figurer la convention graphique correspondant à la classe. Enfin, on fait permuter lignes et colonnes pour diagonaliser le tableau et faire apparaître des groupes de relevés homogènes.



Dans l'exemple ci-dessus (d'après Guinochet et Casal, *in* Lacoste et Salanon, 1969), on distingue deux groupes de relevés. Cette méthode, même si elle présente l'avantage d'être accessible à tous, est longue et fastidieuse. Elle n'est quasiment plus utilisée de nos jours.



II.2. Les méthodes informatiques modernes

II.2.1. Les analyses factorielles

II.2.1.1. Fondements théoriques des techniques utilisées

II.2.1.1.1. Approche des techniques d'analyses factorielles à l'aide d'un problème en deux dimensions

Si les plans factoriels peuvent sembler assez simples au premier abord, les méthodes qui permettent de les élaborer ne le sont pas toujours. Pour essayer d'aborder ces méthodes de manière progressive, un exemple simple va d'abord être traité.

• Définition de l'exemple traité

On s'intéresse au poids et à la taille relevés sur un certain nombre d'individus. Les résultats peuvent être présentés sous la forme d'un tableau de données (communiqué ci-contre). Une colonne correspond au poids, l'autre à la taille. Chaque ligne correspond à un individu.

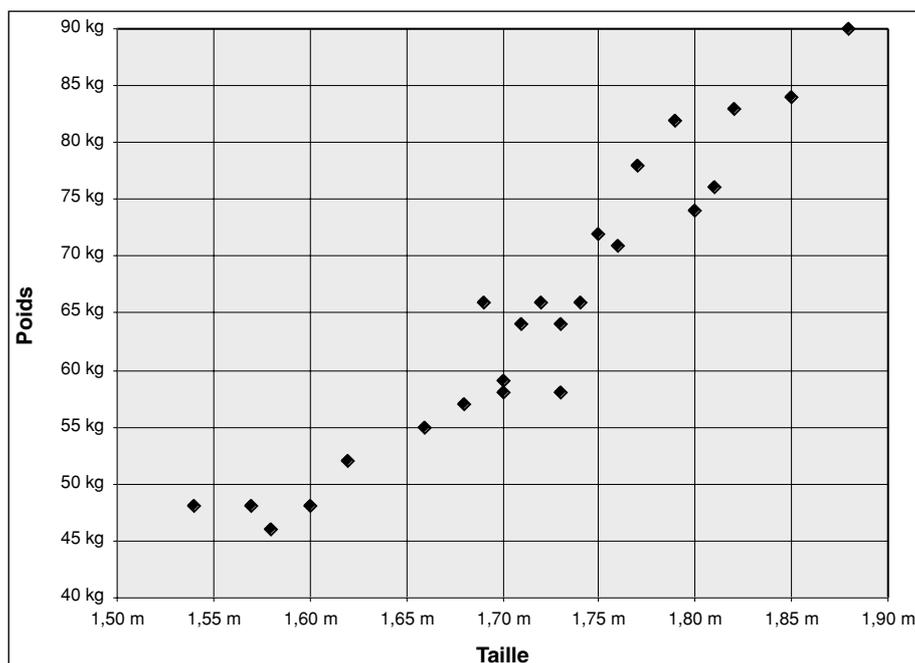
• Présentation matricielle des données

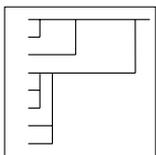
Supposons que l'on veuille représenter graphiquement ces résultats. On va définir un axe qui correspondra à la taille (par exemple l'axe des abscisses) et un autre au poids (l'axe des ordonnées). Un individu donné sera représenté par un couple (taille, poids) : ce sont ses coordonnées. Ainsi, le tableau des données peut être perçu comme l'ensemble des coordonnées des individus dans un plan (espace à deux dimensions). On a ici *deux* coordonnées, donc il faut *deux* dimensions pour représenter le nuage de points.

• Représentation graphique des données

On peut construire le graphique représentant les couples (taille, poids). On constate qu'il existe une certaine relation linéaire entre la taille

Taille	Poids
1,75 m	72 kg
1,73 m	58 kg
1,81 m	76 kg
1,54 m	48 kg
1,82 m	83 kg
1,72 m	66 kg
1,73 m	64 kg
1,60 m	48 kg
1,58 m	46 kg
1,69 m	66 kg
1,68 m	57 kg
1,70 m	58 kg
1,77 m	78 kg
1,71 m	64 kg
1,79 m	82 kg
1,70 m	59 kg
1,57 m	48 kg
1,85 m	84 kg
1,80 m	74 kg
1,74 m	66 kg
1,88 m	90 kg
1,76 m	71 kg
1,62 m	52 kg
1,66 m	55 kg

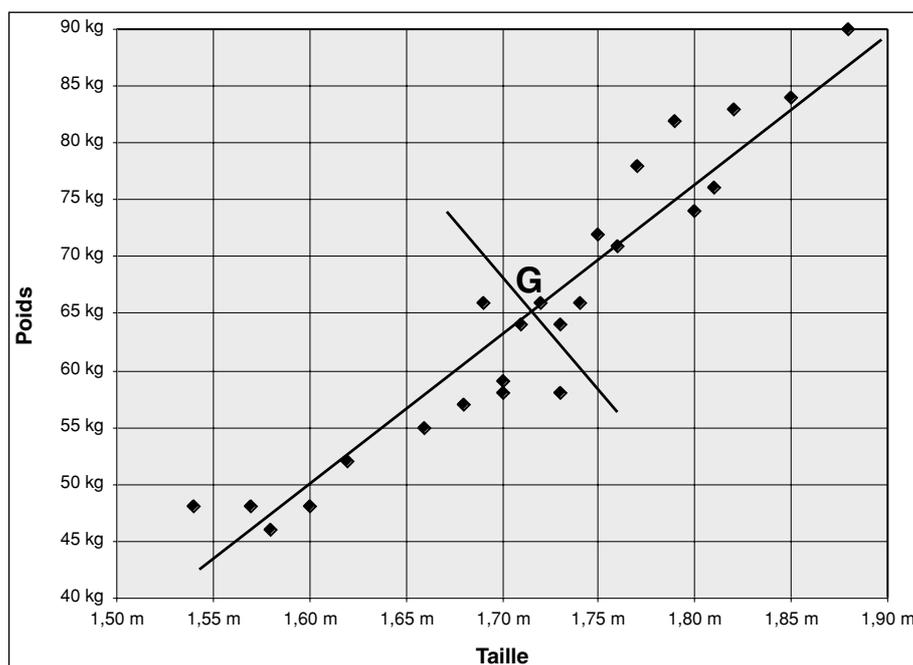




et le poids. On pourrait construire une droite de régression qui aurait une bonne fiabilité.

- **Création et interprétation des axes factoriels**

On peut construire un axe passant par le centre d'inertie G du nuage de points qui suit du mieux possible l'alignement des points. C'est cet axe qui décrirait le mieux le nuage, si l'on ne pouvait disposer que d'un seul axe pour le décrire. On dit que c'est lui qui traduit le mieux *l'inertie* du nuage (par exemple ici, 83 % de l'inertie). Ainsi, si l'on ne considère que les coordonnées des points sur cet axe (une seule dimension), on simplifie la réalité et l'on garde 83 % de "l'information" contenue dans les données de départ.

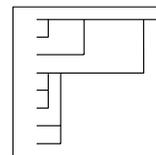


On peut ensuite tracer un second axe orthogonal au premier et passant toujours par le barycentre du nuage de points. Cet axe, comme il est perpendiculaire au premier, traduit le reste de l'inertie du nuage (17 %).

On vient donc en créant ces deux axes de construire un nouveau repère (on dit également faire un changement de base). L'avantage de ce nouveau repère est qu'il *traduit mieux la forme du nuage*. Son premier axe est orienté dans le sens de l'élongation maximale du nuage, le second traduit l'inertie restante.

- **Pour résumer...**

On vient d'associer à un tableau de données contenant *deux colonnes*, un plan (*espace à deux dimensions*) où sont représentés graphiquement les individus. Dans ce plan, on a créé un nouveau repère dont le but est de mieux traduire la forme du nuage de points.



II.2.1.1.2. Généralisation à plusieurs dimensions

• Quand les choses se compliquent un peu...

Supposons que le tableau présenté page 10 ne soit pas complet et qu'il manque des colonnes. Le nouveau tableau plus complet est donné ci-contre. Deux nouvelles variables sont présentes, la pointure des chaussures et le tour de tête donnant la pointure pour les chapeaux.

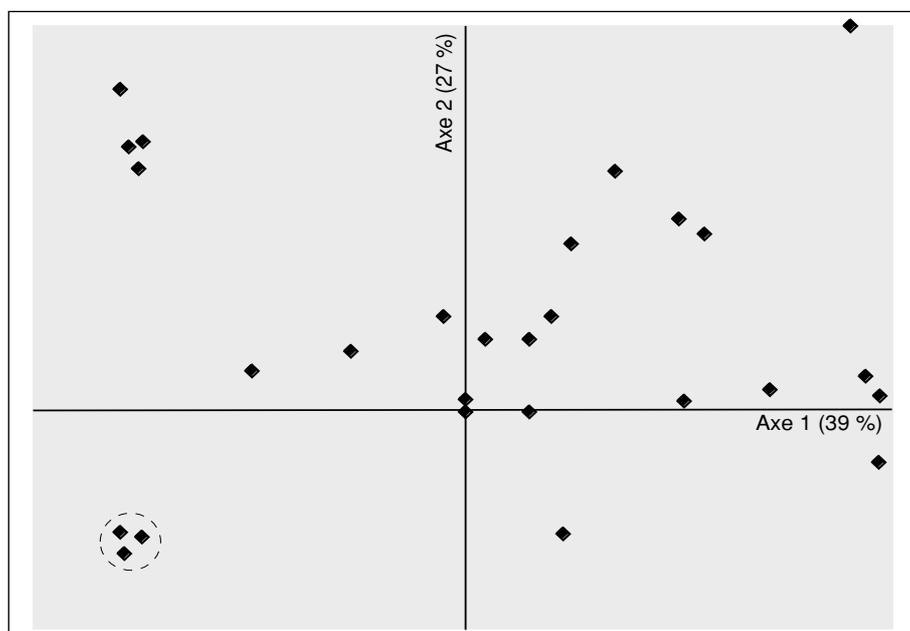
On ne peut plus ici représenter graphiquement le nuage de points car il faudrait 4 dimensions ! Il n'est donc pas facile de visualiser les différents individus en fonction des quatre critères choisis et on peut dire qu'en général, plus le nombre de critères choisis est grand, plus la visualisation est difficile... Toutefois, on peut quand même calculer les coordonnées de l'individu moyen dans un espace à 4 dimensions. Cela donne (1,72 m ; 65 kg ; pointure 41 ; 55 cm).

Des outils mathématiques permettent de faire le même changement de repère que lorsqu'il y avait deux dimensions. Ainsi, on peut créer quatre axes passant par le centre d'inertie du nuage, le premier traduisant la plus grande élongation du nuage. Le second est orthogonal au premier et traduit le maximum d'inertie restant. Le troisième est orthogonal aux deux premiers et traduit le maximum d'inertie laissé par les deux premiers...

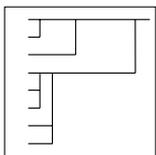
• Les plans factoriels

Construire des axes dans des espaces à n dimensions ne sert pas à grand chose si l'on ne peut toujours pas visualiser de résultats concrets. C'est pour cela qu'on projette orthogonalement les points sur les plans formés par les axes principaux.

Ainsi, si l'on projette nos données sur le plan factoriel formé par l'axe 1 et l'axe 2, on peut obtenir une figure voisine de la suivante :



Taille	Poids	Pointure chaussure	Pointure chapeau
1,75 m	72 kg	38	55
1,73 m	58 kg	42	57
1,81 m	76 kg	45	57
1,54 m	48 kg	42	54
1,82 m	83 kg	46	61
1,72 m	66 kg	36	55
1,73 m	64 kg	39	56
1,60 m	48 kg	39	52
1,58 m	46 kg	36	52
1,69 m	66 kg	39	54
1,68 m	57 kg	40	53
1,70 m	58 kg	38	52
1,77 m	78 kg	44	57
1,71 m	64 kg	42	60
1,79 m	82 kg	44	55
1,70 m	59 kg	41	54
1,57 m	48 kg	38	54
1,85 m	84 kg	44	54
1,80 m	74 kg	44	60
1,74 m	66 kg	43	57
1,88 m	90 kg	44	57
1,76 m	71 kg	46	55
1,62 m	52 kg	40	56
1,66 m	55 kg	39	51



Le plan factoriel 1-2 est le plan qui donne la vision la moins déformée de la réalité. En effet, l'axe 1 et l'axe 2 sont les axes qui traduisent le mieux l'inertie du nuage. La projection sur ces deux axes est donc la meilleure. On peut extraire des projections sur d'autres axes (plan 1-3, plan 2-3...), mais elles donnent une vision plus déformée du nuage de points. On a au plus autant d'axes factoriels qu'il y a de variables au départ.

Etant donné que les axes factoriels sont orthogonaux entre eux, l'inertie du plan factoriel correspond à la *somme des inerties* des axes. Par exemple, ici l'inertie du plan 1-2 vaut 66 % (39+27). Plus l'inertie du plan factoriel est forte, moins ce plan déforme le nuage de points dans la représentation qu'il en donne.

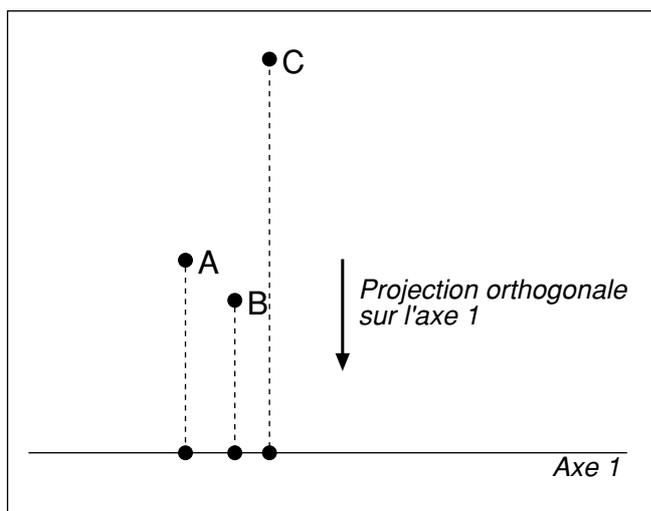
II.2.1.1.3. D'une lecture simple à une lecture élaborée...

C'est une chose de construire un plan factoriel, c'en est une autre de l'interpréter. Si l'on s'intéresse au plan factoriel donné ci-contre, une vision intuitive peut nous donner un certain nombre d'informations.

- Par exemple, si l'on cherche à établir une typologie, les trois points entourés en pointillé au sud-ouest du plan factoriel sont très proches, on peut alors en déduire que les individus que ces points représentent ont des caractéristiques physiques (taille, poids, pointure...) voisines. Autrement dit, la distance qui les sépare est faible.

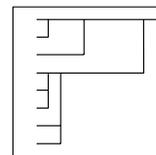
Cette approche intuitive doit être parfois fortement nuancée. Il ne faut pas en effet oublier qu'on ne visualise pas la réalité sur un plan factoriel, mais une projection de la réalité. Il peut exister des *erreurs de perspective*. En effet, comme l'illustre la figure ci-dessous, les projetés de deux points peuvent être proches sans que les deux points le soient.

Sur l'axe 1, on peut penser que les trois points A, B et C sont proches car leurs projetés le sont. Or, on constate que seuls A et B sont proches alors que C est distant de ces deux points. Ainsi, il faut se méfier d'interprétations dangereuses quand on lit un plan factoriel.



Fort heureusement, les logiciels permettent de juger de la qualité de la projection sur les plans factoriels ce qui évite les erreurs.

- Un autre problème réside dans la *notion de distance*. Jusqu'à présent, on a sous-entendu qu'il n'existait qu'une distance, la distance euclidienne



usuelle. Or les mathématiciens peuvent définir d'autres distances ! Ainsi, suivant les méthodes d'analyse, on aura affaire à une distance ou à une autre.

II.2.1.2. Quelques exemples d'analyses factorielles

L'objectif n'est pas de passer ici en revue toutes les techniques d'analyses multivariées, mais de seulement présenter les plus communes utilisées en sciences et techniques forestières.

II.2.1.2.1. L'analyse en composantes principales (ACP)

L'analyse en composantes principales traite les tableaux de réels comme celui présenté page 12. On peut d'une part représenter le nuage des individus sur divers plans factoriels et d'autre part le cercle des corrélation concernant les variables. Normalement, on n'a pas de représentation simultanée des individus et des variables sur le même graphique. Certains logiciels le font, mais il faut alors connaître les conventions utilisées.

L'ACP détecte des relations linéaires entre les individus ou les variables. Souvent, on procède à une ACP normée pour que les variables exprimées par de grands nombres (par exemple 35 mètres pour une hauteur) « n'étouffent » pas celles exprimées par de petits nombres (par exemple 0,35 m de diamètre ou 1,204 m³...). Pour cela, on remplace chaque donnée par sa valeur centrée réduite :

$$\frac{x - \bar{x}}{\sigma}$$

II.2.1.2.2. L'analyse factorielle des correspondances (AFC)

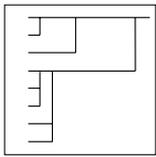
L'analyse factorielle des correspondances traite les tableaux de contingence ou les tableaux disjonctifs complets (comme celui présenté page 7). Les variables et les individus sont représentés sur le même plan factoriel. On peut donc avoir différentes lectures en ne regardant que les individus ou que les variables ou bien les deux en même temps. Lorsqu'on fait des interprétations entre individus et variables, il faut bien avoir à l'esprit qu'un individu est positionné en fonction de toutes les variables et qu'une variable est positionnée en fonction de tous les individus.

II.2.2. Les classifications automatiques

II.2.2.1. Principe et intérêt

Si les analyses factorielles peuvent contribuer à l'élaboration de typologies en montrant quels sont les facteurs premiers de variation et comment se répartissent les individus dans l'espace, elles ne peuvent que rarement à elles seules permettre de construire une typologie.

Des outils de classification automatique peuvent servir à aider à créer la typologie. Généralement, ils sont utilisés conjointement avec les analyses factorielles. Il est difficile, pour construire une typologie de ne se



servir que des analyses factorielles ou que des méthodes de classification automatique.

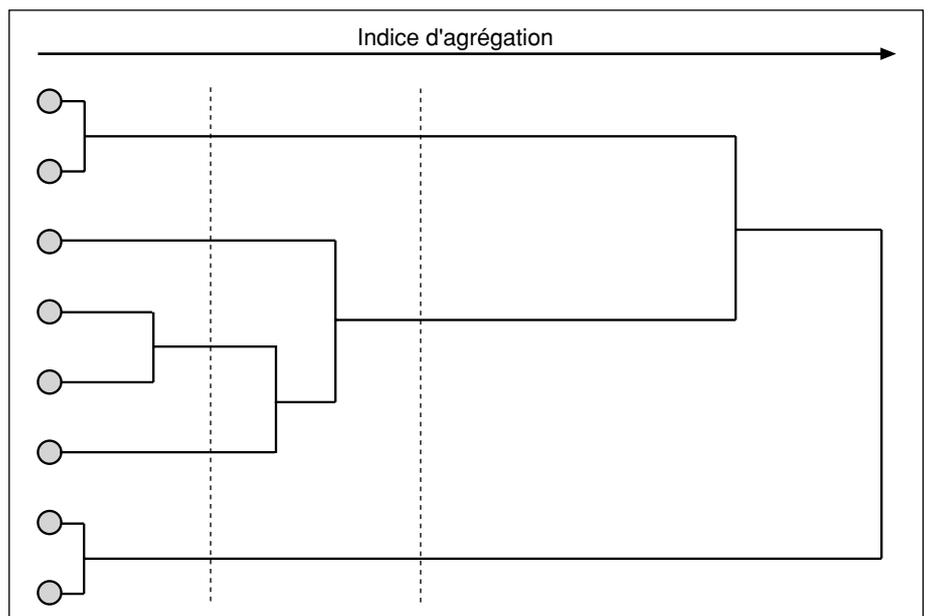
Les outils de classification automatique proposent des groupes les plus homogènes possibles. Ils visent à maximiser la variance inter-classe et à minimiser la variance intra-classe. Diverses méthodes existent pour cela.

II.2.2.2. Les classifications non-hiérarchiques

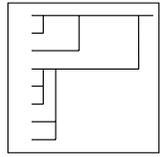
Dans le cadre de ces méthodes (méthode des centres mobiles ou des nuées dynamiques, par exemple), on fixe un nombre de classes (plus ou moins arbitrairement) et le logiciel regroupe les individus en fonction de leur proximité. La notion de choix d'une distance apparaît donc aussi ici, comme pour les analyses factorielles. Le problème de ces classifications tient au fait que le nombre de classes est imposé au départ et cela parfois arbitrairement.

II.2.2.3. Les classifications hiérarchiques

Les classifications hiérarchiques ne donnent pas a priori un nombre déterminé de classes, mais une série de partitions qui permet de partir des individus isolés jusqu'à un groupe regroupant tous les individus, en proposant de nombreuses partitions intermédiaires.



Ainsi, on peut représenter une classification hiérarchique sous la forme d'un arbre. L'indice d'agrégation sert à mesurer la pertinence des groupes formés. En effet, il correspond à l'inertie intra-classe. Ainsi, plus il est élevé, plus les groupes formés sont hétérogènes. L'utilisateur est libre de fixer avec ce genre d'analyse le niveau de rupture qui donnera les groupes (deux exemples de rupture ont été figurés en pointillés sur le dendrogramme fourni ici). Il doit trouver un compromis entre le nombre de groupes formés et la validité statistique des groupes (variance intra-groupe la plus faible possible).

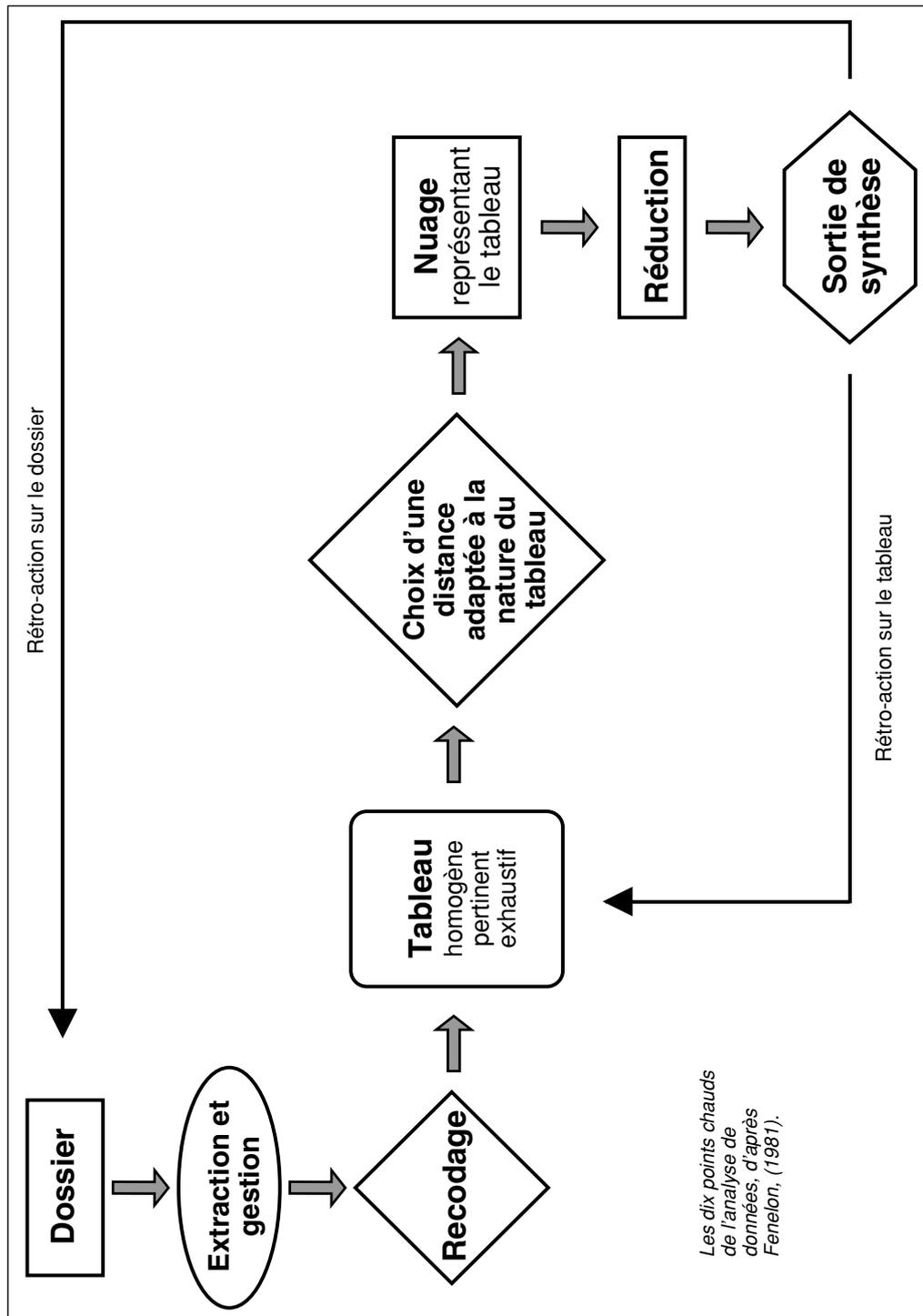


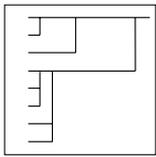
Les classifications hiérarchiques présentent l'avantage d'être moins arbitraires que les classifications non-hiérarchiques. On est toutefois dépendant du choix d'une distance et rien n'interdit de confronter les résultats de ces deux familles de méthodes pour se donner tous les moyens d'aboutir à une typologie fiable.

II.2.3. Méthodologie générale en analyse des données

II.2.3.1. Les dix points chauds de l'analyse des données (d'après Fénelon, 1981)

Le schéma suivant présente les 10 points chauds (ou 10 étapes) d'une démarche d'analyse de données.





- **Le dossier d'étude**

Le dossier d'étude correspond à l'ensemble des données recueillies pour être analysées. Il peut s'agir des résultats d'une enquête, de l'ensemble de relevés effectués en forêt... Ce dossier d'étude correspond à des objectifs bien précis comme par exemple établir une typologie des peuplements ou montrer les relations entre structure des peuplements et flore du sous-bois.

- **Extraction des données**

Dans cette étape, on prépare les données au traitement. On peut mettre un identificateur pour chaque individu (chaque placette porte un numéro qui la différencie des autres) et on choisit, parmi celles présentes dans le dossier, les données que l'on soumettra au traitement.

- **Codage des données**

La plupart du temps, les variables récoltées nécessitent un recodage. On distingue en effet un grand nombre de familles de variables différentes :

- les **variables amorphes**. Ce sont par exemple les variables qui servent à numéroter les individus. L'individu 8 n'est pas deux fois plus petit que l'individu 16 ni quatre fois plus que l'individu 32 !

- les **variables ordinales**. Elles informent sur une relation d'ordre. On peut par exemple coder 1 les individus ayant eu une mention bien au BTS, 2 ceux ayant eu une mention assez bien, 3 ceux ayant eu une mention passable...

- les variables de **présence-absence**. On peut mettre 1 quand une plante est présente dans un relevé et 0 quand elle est absente.

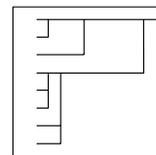
- les **variables continues**. Elles résultent de mesures effectuées et prennent des valeurs réelles, par exemple 35 % de PB, $V = 205 \text{ m}^3/\text{ha}$...

La façon dont sont codées ou recodées les variables est particulièrement importante pour les analyses multivariées. En effet, du bon codage dépend la qualité des résultats obtenus et certains codages ne permettent de faire que certaines analyses. Le codage des données peut prendre beaucoup de temps. Une très grande partie de l'arbitraire des résultats tient du codage.

- **Le tableau de données**

Le tableau de données est un extrait du dossier initial où certaines variables ont été recodées. C'est ce tableau qui va être soumis aux analyses.

Il existe de nombreuses variétés de tableaux (tableaux de contingences, tableaux de mesures, tableaux de notes, tableaux de présence-absence...).



Ces tableaux peuvent être soumis à des analyses diverses, ils possèdent leurs particularités et leurs traitements potentiels.

Un tableau doit être :

- homogène,
- pertinent,
- exhaustif.

L'homogénéité semble aller de soi, on ne doit pas mélanger des données qui n'ont aucun rapport entre elles. Il faut aussi vérifier ce que signifie l'addition des lignes et des colonnes ou bien imaginer ce que représente l'ensemble des termes d'une ligne. On doit porter une attention particulière au fait que les relations entre les nombres présents dans le tableau traduisent bien les relations qui existent dans la réalité.

La pertinence signifie que l'objet du relevé est bien défini et que le problème a un sens. L'exhaustivité signifie que les différentes zones du domaine d'investigation sont bien recouvertes par les relevés.

- **Choix d'une distance**

Le choix d'une distance (au sens mathématique, tel qu'il a été abordé page 13) est conditionné par la nature du tableau et le type d'analyse qu'on veut effectuer. Souvent les logiciels passent cet aspect sous silence et la distance est choisie en fonction du type d'analyse effectué.

- **Nuage de points**

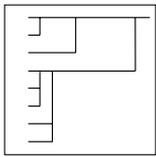
Une distance ayant été choisie, on peut construire un nuage de points où les individus sont placés en fonction de leurs coordonnées. Sauf dans quelques cas rares et simplistes ce nuage de points ne peut être visualisé car il se situe dans un espace possédant bien plus de dimensions que le nôtre.

- **La réduction des données**

Le nuage de points n'étant pas visualisable, il n'est pas intéressant sous sa forme originelle. Les techniques d'analyses multivariées procèdent donc à une réduction du nuage. Cela peut correspondre à la création des axes factoriels et à un changement de base dans le cadre d'une analyse factorielle, à l'agglomération en différentes classes dans celui d'une classification automatique.

- **Sorties de synthèse**

La réduction des données permet de simplifier un tableau souvent assez complexe. Des sorties de synthèse peuvent donc être réalisées après la réduction des données. Cela peut consister en des tableaux qu'on pourra analyser, ou bien en des sorties graphiques (plans factoriels, arbres de classification...). C'est à partir de ces résultats qu'on peut construire une



typologie, mais les étapes précédentes ne doivent en aucun cas être négligées.

- **Rétroaction sur le tableau**

Il se peut qu'arrivé au stade des sorties de synthèse, on constate que certains individus ont un rôle aberrant ou trivial. On peut alors traiter à nouveau le tableau initial en supprimant ces individus, ou en leur donnant un poids nul. Ils sont alors présents sur les plans factoriels, mais n'ont pas contribué à leur élaboration. On parle ainsi de variables ou d'individus supplémentaires. Malheureusement, tous les logiciels ne gèrent pas les variables ou les individus supplémentaires.

- **Rétroaction sur le dossier**

Tout comme on peut rétro-agir sur le tableau, on peut rétro-agir sur le dossier. On peut par exemple faire à nouveau une analyse factorielle avec des individus ou des variables supplémentaires qui n'étaient pas dans le tableau originel. On peut également utiliser les outils classiques de statistique (comparaison de moyenne, intervalle de confiance, régression) sur les résultats obtenus, pour par exemple valider les types obtenus lors de l'élaboration d'une typologie.

Cette étape de l'analyse n'est en rien superflue. Elle permet souvent de mieux interpréter les résultats et les enrichit considérablement.

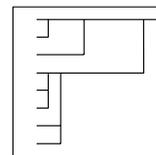
II.2.3.2. Les limites de l'analyse des données

Comme on peut le constater à la lecture de ce qui précède, les analyses multivariées font partie d'un domaine des statistiques dans lequel les erreurs peuvent facilement et rapidement se commettre. En effet, l'obtention de résultats cohérents et reflétant la vérité passe par de nombreuses étapes dans lesquelles la rigueur est de mise. Le codage, le choix des distances, la forme des tableaux traités sont des étapes cruciales si l'on veut obtenir de bons résultats.

Il faut de plus savoir relativiser les résultats, ne pas hésiter à faire plusieurs analyses contradictoires avant de décréter le travail achevé. L'intervention humaine en analyse de données importe beaucoup plus que ce qu'on pourrait penser au départ. Il faut savoir faire des choix et être prudent dans ses interprétations, savoir relativiser. Ce n'est pas l'ordinateur qui décide de tout.

Les 10 points chauds présentés au paragraphe précédent permettent de ne rien laisser au hasard et de raisonner correctement sa pratique. Si l'on n'y prend garde, on peut obtenir des résultats sans réels fondements.

On n'a traité ici qu'une simple approche des analyses multivariées. Sachant que les méthodes employées sont assez complexes, on aura intérêt à consulter des ouvrages traitant de ce thème si l'on veut se spécialiser.



Il ne faut pas qu’une utilisation informatique rendue accessible conduise à des résultats aberrants.

Conclusion

Diverses méthodes sont utiles pour construire des typologies fiables et efficaces. Quelques exemples ont été rapidement passés en revue ici, mais d’autres méthodes, notamment en analyses multivariées, existent. Il faut concevoir qu’un travail de création de typologie n’est pas simple dès qu’on se trouve en conditions réelles. Des soins qu’on apportera à sa réalisation dépendra sa qualité. Il ne faut pas en effet oublier qu’on se trouve, lors d’un travail typologique, dans des domaines où rien n’est tranché et il est impossible d’appliquer une simple méthode rigide sans discernement.

Bibliographie

- ALLEGRINI C. 1981 - Les futaies jardinées privées du Haut-Jura, Besançon, CRPF, 9 p. (résumé des travaux de Herbert et Rébeiro)
- AUBRY S., DRUELLE P. 1988 – Vers une meilleure connaissance des peuplements feuillus. Typologie en région centre, Mémoire de troisième année ENITEF, Les Barres, 87 p. et annexes.
- BAILLY G. 1995 – Catalogue des types de stations forestières de la Plaine de la Saône, CETEF de Côte d’Or, Dijon, 311 p.
- BECKER M. 1979 – Influence du traitement sylvicole sur la flore forestière : cas de la futaie et du taillis-sous-futaie, *Vegetatio*, vol. 40, n°3, pp. 155-161.
- BOUROCHE J.M., SAPORTA G. 1980 – L’analyse des données, quatrième édition (juin 1989), Que sais-je ? (n°1854) PUF, 124 p.
- BRUCIAMACCHIE M. 1989 – Typologie des Peuplements, R.F.F., n°6, pp. 507-512.
- BRUCIAMACCHIE M. 1991 – Cours sur l’analyse des données, document interne ENITEF, Les Barres, 12 p. et annexes.
- DOUSSOT R. 1990 – Cours d’aménagement, document interne ENITEF, Les Barres, 50 p. et annexes.
- FENELON J.P. 1981 – Qu’est-ce que l’analyse des données ? Paris, éditions Lefonen, 311 p.
- GUINOCHET M. 1973 – Phytosociologie, Masson, Paris, 227 p. + carte
- HERBERT I., REBEIROT F. 1981 – Les futaies jardinées privées des hautes chaînes du Jura, Mémoire de troisième année ENITEF, Les Barres, 125 p. et annexes.
- IDF 1985 – Typologie des stations forestières - Vocabulaire IDF, Paris, 243 p.
- LACOSTE A., SALANON R. 1969 – Eléments de biogéographie et d’écologie, Fernand Nathan, Paris, 189 p.
- OZENDA P. 1982 – Les végétaux dans la biosphère, Douin, Paris, 431 p.