



**HAL**  
open science

# Comparaison de diverses méthodes d'interprétation statistique de liaison entre le milieu et la production du Pin sylvestre en Sologne

N. Decourt, M. Godron, F. Romane, R. Tomassone

► **To cite this version:**

N. Decourt, M. Godron, F. Romane, R. Tomassone. Comparaison de diverses méthodes d'interprétation statistique de liaison entre le milieu et la production du Pin sylvestre en Sologne. *Annales des sciences forestières*, 1969, 26 (4), pp.413-443. 10.1051/forest/19690401 . hal-00882008

**HAL Id: hal-00882008**

**<https://hal.science/hal-00882008>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## COMPARAISON DE DIVERSES MÉTHODES D'INTERPRÉTATION STATISTIQUE DE LIAISON ENTRE LE MILIEU ET LA PRODUCTION DU PIN SYLVESTRE EN SOLOGNE

N. DECOURT \*, M. GODRON \*\*, F. ROMANE \*\* et R. TOMASSONE \*\*\*

\* *Station de Sylviculture et Production,  
Centre national de Recherches forestières, 54 - Nancy  
Institut national de la Recherche agronomique*

\*\* *Centre d'Etudes Phytosociologiques et Ecologiques, 34 - Montpellier  
Centre national de la Recherche scientifique*

\*\*\* *Station de Biométrie,  
Centre national de Recherches zootechniques, 78 - Jouy-en-Josas  
Institut national de la Recherche agronomique*

---

### SOMMAIRE

Ce travail est essentiellement consacré à la présentation de méthodes d'interprétation. Pour étudier la corrélation entre plusieurs variables du milieu et la production du Pin sylvestre en Sologne trois techniques statistiques de l'analyse multivariée sont utilisées :

- l'analyse des composantes principales,
- la régression, et surtout la régression orthogonale,
- l'analyse discriminante.

Des extensions sont prévues pour des études futures ; les liens entre l'aspect méthodologique et le traitement sur l'ordinateur sont évoqués.

Tous les programmes FORTRAN utilisés sont disponibles.

---

### INTRODUCTION

Des tables de production ont été construites pour le Pin sylvestre en Sologne par DECOURT (1965) à partir d'une centaine de « placettes temporaires ». Ces tables distinguent des classes de fertilité qu'il était souhaitable de relier à des caractères écologiques déjà étudiés par ailleurs (GODRON *et al.*, 1964). C'est pourquoi des observations écologiques ont été effectuées dans chaque placette (§ 1) et interprétées aussi soigneusement que possible (§ 2).

## I. — CHOIX DES VARIABLES FORESTIERES ET ECOLOGIQUES

## 1.1. — Variables forestières

Puisque l'étude entreprise a pour objectif de préciser les rapports entre les variables écologiques et la production du Pin sylvestre, il est indispensable de retenir certaines variables « forestières ».

— La plus importante est bien sûr *l'indice de fertilité* (Site Index des auteurs anglo-saxons), qui caractérise le potentiel productif du peuplement. Il s'agit ici de la hauteur moyenne théorique qu'aurait le peuplement à cinquante ans. Cet indice a été établi selon la méthode décrite par BRUCE et SCHUMACHER (1950). Le choix de la hauteur moyenne (hauteur de l'arbre de surface terrière moyenne) plutôt que celui de la hauteur dominante, lié à une certaine insuffisance des données de base, ne modifie pas la hiérarchie des placettes, ces deux hauteurs étant liées par une relation linéaire extrêmement étroite.

— La seconde est *la surface terrière*, excellent indice de densité du peuplement, et donc à la fois du volume sur pied et du degré de couvert.

— La troisième est *l'âge*, qui joue un rôle capital dans la construction des tables de production et dans l'évolution du type d'humus. De plus, il était intéressant d'introduire cette variable qui, théoriquement, devrait être indépendante de l'indice de fertilité.

Un certain nombre de remarques doivent être faites sur ces variables et sur l'échantillon de peuplements retenu.

a) *L'indice de fertilité* joue dans cette étude un rôle fondamental puisqu'il s'agit essentiellement d'en expliquer les variations à l'aide des autres. Cette importance justifie qu'on en fasse une rapide critique.

L'estimation de l'indice suppose qu'à partir de mesures de hauteur faites à des âges variés dans divers peuplements, il est possible d'estimer les hauteurs de ces peuplements à un âge de référence standard, au moyen par exemple d'une famille de courbes de croissance. L'erreur qu'on introduit ainsi est inconnue et dépend de la précision avec laquelle les courbes utilisées peuvent représenter la croissance des peuplements individuels.

Il est fort possible que la forme même de ces courbes soit différente selon les milieux écologiques [voir par exemple VINCENT (1961)]. L'usage, même pour ce travail, d'un indice de fertilité est donc relativement arbitraire et susceptible de limiter le succès de cet essai.

De plus, il est clair que « le milieu » ne suffit pas à lui seul à expliquer les variations de l'indice. La variabilité génétique du Pin sylvestre et les provenances géographiques fort diverses — et inconnues — des peuplements étudiés limitent au départ, sans doute de façon non négligeable, les résultats que l'on peut espérer obtenir dans notre étude.

b) *Une autre critique concerne l'échantillonnage des peuplements.* Nos placettes devraient être réparties également dans tout l'éventail des âges et des stations. Ayant choisi des peuplements existants, là où on les trouvait et avant de définir les stations, nous ne pouvons être assurés de la qualité de notre échantillon. Il y a là une difficulté à laquelle on peut difficilement échapper, et le résultat même de l'étude entreprise

pourra être de souligner certaines insuffisances et d'orienter la recherche ultérieure, au besoin vers la récolte d'un supplément d'information sur le terrain.

Ces remarques montrent bien le caractère limité de cette étude et son intérêt essentiellement méthodologique.

### 1.2. — Variables climatiques

Aucune variable climatique n'a été retenue. En effet, à l'échelle de la Grande Sologne, le climat général varie très progressivement, selon un gradient est-ouest : la partie orientale, adossée aux contreforts du Sancerrois, est légèrement plus humide et plus froide que la partie occidentale. Mais cette variation ne semble pas avoir d'effet statistiquement décelable sur la croissance du Pin sylvestre ; ceci est en accord avec les renseignements collectés par J. POISSONET (inédit) qui conduisent à penser que, en Sologne, la croissance du Pin sylvestre n'est pas entravée par des facteurs climatiques limitants.

### 1.3. — Variables biotiques

Un troisième groupe de variables est celui des variables « biotiques », qui sont relatives aux interactions entre êtres vivants. Ces interactions peuvent elles-mêmes être réparties en plusieurs sous-groupes ; trois d'entre eux méritent plus spécialement d'être discutés :

— La concurrence entre végétaux appartenant à des espèces différentes : de ce point de vue, les Pins ont dû lutter, aussitôt après leur plantation, contre la végétation herbacée ou buissonnante, mais l'effet de cette concurrence est difficilement décelable dans un peuplement adulte. Par acquit de conscience, les « recouvrements » des Phanérogames herbacées et des ligneux bas ont été cependant estimés dans chacune des placettes, mais ils n'ont pas permis d'établir de relations intéressantes. Par contre, le recouvrement des Mousses a été maintenu dans les synthèses statistiques, car il s'est révélé utile pour caractériser le milieu.

— La concurrence entre les Pins eux-mêmes : c'est un facteur capital, qui dépend évidemment du régime d'éclaircies pratiqué par le sylviculteur : son effet a été étudié lors de l'élaboration des tables de production (N. DECOURT, 1965). Il est inutile d'y revenir ici.

— L'action des animaux, et en particulier du lapin, est importante au cours des premières années de la vie du peuplement, mais on ne peut pratiquement pas la déceler sur un peuplement adulte.

### 1.4. — Variables édaphiques

#### 1.41. Remarques préliminaires.

L'étude des facteurs liés au sol pose deux problèmes délicats :

— Il faut tenir compte de la *profondeur* à laquelle une observation est faite, pour l'interpréter correctement.

Ainsi, l'étude statistique a montré que, dans les placettes étudiées, l'humus brut est plus épais quand la texture est plus argileuse en surface : au contraire, l'humus brut est plus mince quand la texture est plus argileuse en profondeur.

— Des phénomènes très divers, tels que le lessivage, l'engorgement par l'eau, la transformation et la migration de l'humus, interviennent simultanément dans le sol. En conséquence, une observation simple, comme la teneur en matière organique d'un horizon, n'a pratiquement aucun sens en elle-même ; on ne peut l'interpréter qu'en fonction de l'épaisseur de cet horizon, de sa position relative et des autres caractères du sol. L'interprétation statistique doit donc prendre en considération *l'ensemble* des variables édaphiques observées.

Pour résoudre ces difficultés, on effectue habituellement une étude complète du sol (cf., par exemple, le « Code pour le relevé méthodique de la végétation et du milieu », C.E.P.E., 1968). Dans le cas présent, cette solution débouchait sur une impasse. En effet, l'étude pédologique globale d'une parcelle à reboiser ne suffit pas pour en connaître la productivité et cela pour deux raisons :

- en Sologne, les sols sont extrêmement variables d'un point à un autre, sur une courte distance ;
- les sols observés sont généralement polyphasés et même souvent polygéniques. Aussi le type de sol actuel ne suffit pas pour caractériser sa fertilité.

Si l'on désire pouvoir observer efficacement les variables édaphiques dans les parcelles à reboiser, ces variables doivent être définies par un protocole simple et analytique (la synthèse de ces variables analytiques a été effectuée progressivement au cours de l'interprétation, cf. § 2).

#### 1.42. Procédure adoptée.

Finalement, la solution proposée par MM. DUCHAUFOUR et BONNEAU s'est révélée la plus judicieuse. Elle consiste à faire des observations et à prélever des échantillons pour quatre horizons choisis en raison des possibilités de diagnostic qu'ils offrent :

— Pour l'humus brut superficiel ( $A_0$ ), qui comporte souvent encore des éléments de matière organique non décomposés, l'épaisseur moyenne est observée sur le terrain et notée en centimètres. Un échantillon est récolté à la main, en vue de l'analyse chimique.

— Pour l'humus associé aux particules minérales du sol, l'épaisseur de l'horizon  $A_1$  a été estimée ; mais, du fait que la limite inférieure de cet horizon est rarement nette, le prélèvement a été effectué systématiquement dans les cinq premiers centimètres du profil.

— La majorité des racines et radicelles nourricières du Pin sylvestre est située dans la partie supérieure du sol, et les racines profondes jouent plutôt un rôle dans l'alimentation en eau. En outre, la plupart des parcelles ont autrefois été labourées, et les horizons superficiels ont alors été mélangés. C'est pourquoi un échantillon de terre comprenant les vingt premiers centimètres du profil a été prélevé, en vue d'effectuer les analyses physiques et chimiques classiques.

— Pour les horizons profonds, le problème est assez complexe, et il a été nécessaire de schématiser l'opération, pour que les prélèvements soient aussi comparables que possible d'une station à l'autre et pour que la procédure puisse être appliquée par un sylviculteur non spécialisé en pédologie : l'échantillon est prélevé à l'aide d'une tarière hélicoïdale, dans l'horizon homogène situé à une profondeur voisine d'un mètre.

Une telle procédure a une signification particulière en Grande Sologne, car les études antérieures (DENOZOT, 1927 ; VATAN, 1937 ; HOREMANS, 1961 ; LE HOUEYOU, 1961 ; WACQUANT, en préparation) ont montré que, dans la plupart des cas, on trouve à cette profondeur, soit le Burdigalien peu altéré, soit des dépôts postérieurs, tels que ceux des terrasses quaternaires ; ainsi, la texture de cet horizon profond permet généralement de caractériser la « série » géomorphologique (WACQUANT, en préparation).

#### 1.43. Profondeurs.

Pour préciser la signification des observations précédentes, il est utile d'ajouter les profondeurs où se trouvent trois horizons caractéristiques ; ce sont :

- la profondeur de l'horizon engorgé en permanence (cette profondeur n'a pu être mesurée que sur douze placettes qui ont été séparées de l'ensemble du lot, cf. paragraphe 2.03) ;
- la profondeur de l'horizon engorgé en hiver et sec en été, dont l'importance vient encore d'être soulignée par J. POISSONET (1968) ;
- la profondeur à laquelle apparaît le premier horizon nettement argileux (cf. J. POISSONET, 1968).

1.44. Enfin, la situation topographique et le type de substrat géomorphologique ont été notés dans chacune des placettes.

#### 1.5. — Variables phyto-écologiques

Il était intéressant de comparer les observations précédentes à celles qui peuvent être déduites de la présence des *groupes d'espèces indicatrices*, définis lors des études écologiques faites antérieurement en Sologne (GODRON *et al.*, 1964, p. 101 à 125). Ces groupes, symbolisés par des lettres, ont permis de classer chaque placette en fonction de deux critères :

— L'humidité de la station, qui est très liée à la profondeur de l'horizon engorgé en hiver, mais tient compte, en outre, du microclimat. De ce double point de vue, les groupes écologiques sont ordonnés de la manière suivante :

1	2	3	4	5	6	7
L	T et F	P	N et H	J	CC	S

— L'activité biologique stationnelle (cf. DUCHAUFOUR, 1960), qui est, en général, d'autant plus grande que l'humus est plus actif et que le sol est plus frais. Elle est approximativement proportionnelle au nombre d'Angiospermes non éricacées observées sous la ligne de segments qui a servi à mesurer les fréquences phytosociologiques. Les groupes écologiques peuvent être ordonnés par ordre d'activité biologique croissante :

1	2	3	4	5	6
S	L	V et F	T	N	P, J et H



## 1.6. — Code des observations

Les observations ont été codées de la manière suivante :

- X<sub>1</sub> Age du peuplement principal (en années).
- X<sub>2</sub> Surface terrière moyenne (en mètres carrés).
- X<sub>3</sub> Position topographique.
- 0 terrain plat.
  - 2 sommet arrondi et croupe.
  - 3 haut de versant.
  - 4 mi-versant.
  - 5 bas de versant.
  - 6 fond de vallée.
  - 7 cuvette fermée.
- X<sub>4</sub> Recouvrement des mousses (en %).
- Ce chiffre est déduit de « points quadrats ».
- X<sub>5</sub> Epaisseur de l'horizon A<sub>0</sub> (en centimètres).
- X<sub>6</sub> Valeur du rapport C/N dans l'horizon A<sub>0</sub>.
- X<sub>7</sub> Valeur du rapport C/N dans l'horizon A<sub>1</sub>.
- X<sub>8</sub> Pourcentage de matière organique dans les 20 premiers centimètres du sol.
- X<sub>9</sub> Pourcentage d'argile (0 à 2 μ), dans les 20 premiers centimètres du sol.
- X<sub>10</sub> Pourcentage de limon (2 à 50 μ), dans les 20 premiers centimètres du sol.
- X<sub>11</sub> Pourcentage de refus (2 mm et au-dessus), dans les 20 premiers centimètres du sol.
- X<sub>12</sub> Pourcentage d'argile dans l'échantillon profond.
- X<sub>13</sub> Pourcentage de limon dans l'échantillon profond.
- X<sub>14</sub> Pourcentage de refus dans l'échantillon profond.
- X<sub>15</sub> Profondeur de l'horizon engorgé en hiver (en centimètres).
- Si cet horizon n'a pas été observé, il a été noté conventionnellement à 150 cm.
- X<sub>16</sub> Profondeur de la couche d'argile (en centimètres).
- Si cette couche n'a pas été observée, elle a été notée à 250 cm.
- X<sub>17</sub> Humidité équivalente (en %), dans les 20 premiers centimètres du sol.
- X<sub>18</sub> pH dans les 20 premiers centimètres du sol.
- X<sub>19</sub> Capacité d'échange (en milliéquivalents pour 100 g de terre fine séchée à l'air).
- X<sub>20</sub> Calcium échangeable (en milliéquivalents pour 100 g de terre fine séchée à l'air).
- X<sub>21</sub> Magnésium échangeable (en milliéquivalents pour 100 g de terre fine séchée à l'air).
- X<sub>22</sub> Potassium échangeable (en milliéquivalents pour 100 g de terre fine séchée à l'air).
- X<sub>23</sub> Sodium échangeable (en milliéquivalents pour 100 g de terre fine séchée à l'air).
- X<sub>24</sub> Acide phosphorique (P<sub>2</sub>O<sub>5</sub>) assimilable (en %).
- X<sub>25</sub> Série géomorphologique (cf. WACQUANT, en préparation).
- 1 m<sub>2</sub> Burdigalien.
  - 2 m<sub>20</sub> Cailloutis culminants.
  - 3 m<sub>20</sub> Cailloutis préligériens.
  - 4 Terrasses de la Loire.
  - 5 Terrasses Bouvron-Cosson.
  - 6 Dunaire sur Burdigalien.
  - 7 Argiles à silex.
  - 8 Terrasses des Saultres.
  - 9 Dunaire sur terrasses des Saultres.
  - 10 Dunaire sur argiles à silex.
- X<sub>26</sub> Humidité de la station (cf. GODRON *et al.* 1964, p. 140 et 141).
- 1 groupe écologique L.
  - 2 groupe écologique T (et F).
  - 3 groupe écologique P.
  - 4 groupe écologique N et H.
  - 5 groupe écologique J.

6 groupe écologique U.

7 groupe écologique S.

X<sub>27</sub> Activité biologique (cf. GODRON *et al.* 1964, p. 140 et 141).

1 groupe écologique S.

2 groupe écologique L.

3 groupe écologique U (et F).

4 groupe écologique T.

5 groupe écologique N.

6 groupe écologique P, J et H.

X<sub>28</sub> Indice de production.

(hauteur à 50 ans en décimètres).

## II. — LES DIFFÉRENTS TYPES DE MÉTHODES STATISTIQUES UTILISÉES

### 2.1. — GÉNÉRALITÉS SUR LES MÉTHODES ET LES DONNÉES DE BASE

#### 2.11. *Introduction*

L'étude détaillée de l'effet individuel de chacune des variables écologiques sur la production du Pin sylvestre serait très insuffisante, puisque notre étude se veut synthétique.

Le grand nombre de variables, lorsque nous les étudions séparément, ne permet pas d'avoir une vue d'ensemble des résultats ; il est aussi évident que la masse des analyses que nous pourrions faire sur chacune d'elles, sans nous préoccuper de l'action concomitante des autres, nous conduirait de façon certaine à des redites. Enfin, il serait extrêmement délicat, voire impossible, de dégager les grandes lignes qui caractérisent le milieu, par conséquent encore plus difficile de voir comment le milieu peut influencer la production, et naturellement impossible de chiffrer, même de façon approximative, cette action.

La nature et le nombre des données sur lesquelles nous devons travailler justifiaient *a priori* l'utilisation de l'analyse statistique à plusieurs variables. Nous avons pensé que peu de cas concrets d'interprétation avaient été publiés et qu'il était bon de faire part de notre expérience de l'application de méthodes statistiques, devenues assez banales par ailleurs, à un domaine où elles pourraient être d'un grand secours. Ce faisant, nous n'avons pas voulu étudier systématiquement toutes les méthodes d'analyse à plusieurs variables ; nous avons choisi celles que nous pensions être les plus aptes à nous fournir les résultats les plus facilement interprétables ; nous avons aussi choisi celles que nous avons déjà programmées <sup>(1)</sup>.

#### 2.12. *Les différents types d'interprétation*

Pour l'interprétation proprement dite, nous avons eu trois idées directrices :

1° expliquer le milieu au moyen d'un nombre plus restreint de variables synthétiques ;

(1) Les calculs ont été effectués de façon très fragmentée : commencés sur un ordinateur avec des programmes écrits en ALGOL (CAE 510) ils ont été terminés sur un autre avec des programmes écrits en FORTRAN (IBM 1130).

2° traduire les variations de la production de la façon la plus détaillée, soit en fonction de ces facteurs nouveaux, soit en analysant la part propre de chaque variable mesurée ;

3° classer les placettes étudiées en un certain nombre de groupes caractérisés par un ensemble de valeurs des variables ; ce classement permettant ultérieurement de replacer n'importe quelle placette à l'intérieur d'un groupe avec une certaine probabilité.

En outre, dans cette étude, nous considérons que les outils statistiques nous fournissent plus des suggestions que des preuves absolues. Par conséquent, les résultats que nous allons présenter peuvent offrir un double intérêt : *intérêt méthodologique*, en passant au crible certaines méthodes ; nous pensons que c'est là l'intérêt essentiel ; *intérêt pratique*, permettant de fournir, pour le domaine de la Sologne que nous avons étudié, un certain nombre d'éléments chiffrés. Ces éléments pourraient être employés pour une étude plus détaillée et serviraient de données initiales utilisables pour implanter un dispositif expérimental. Ce seraient donc les éléments d'une « enquête-pilote » au sens statistique du terme.

### 2.13. Les données. Analyse préliminaire

Afin de pouvoir appliquer dans les meilleures conditions les différentes méthodes statistiques, il était indispensable, dans un premier temps, d'étudier la distribution de chaque variable séparément ; étant donné le nombre peu élevé de placettes (89) <sup>(1)</sup>, nous avons utilisé les deux coefficients classiques : le coefficient  $\sqrt{b_1}$  de PEARSON, et le coefficient  $a$  de GEARY, PEARSON E. S. et HARTLEY H. O. (1958). Lorsque la distribution était beaucoup trop éloignée de la normale, nous avons effectué sur les variables des transformations simples (racine carrée ou logarithme). Toutefois, pour des variables de même type, par exemple pour les différents éléments chimiques échangeables, nous avons préféré, pour des raisons d'ordre pratique, conserver le même type de variable transformée. Nous ne pensons pas que cette légère entorse puisse se traduire par des résultats faux. Et nous voyons une fois de plus la difficulté qu'il y a à établir un compromis harmonieux entre une méthodologie stricte et des éléments utilisables de façon simple.

Nous pouvons résumer dans le tableau suivant les résultats des différents tests ; quand nous parlerons dans la suite du texte d'une quelconque variable  $x_i$ , c'est à ce tableau que nous référerons.

(1) Nous avons au préalable supprimé 12 placettes qui correspondaient à des terrains engorgés toute l'année. Une analyse de variance sur chaque caractère séparément nous avait permis de déceler des différences significatives entre les deux groupes et nous n'avons pas voulu introduire une cause d'hétérogénéité supplémentaire.

TABLEAU I

*Variables utilisées dans l'analyse*

N°	Caractère écologique	Transformée	Tests		Type de variables
			<i>a</i>	$\sqrt{b_1}$	
1 ...	Age du peuplement principal	Origine	0,8765	0,2235	Forestière
2 ...	Surface terrière moyenne	Origine	0,7734	0,9120	Forestière
3 ...	Position topographique	Origine	0,9220	0,3964	Analytique 1
4 ...	Recouvrement des mousses	Origine	0,8204	0,3498	Analytique 1
5 ...	Épaisseur de $A_0$	Log (1 + x)	0,8236	0,0680	Analytique 1
6 ...	C/N de $A_0$	Log (1 + x)	0,27291	1,5315	Analytique 1
7 ...	C/N de $A_1$	Log (1 + x)	0,7904	0,8618	Analytique 1
8 ...	Matière organique dans 0,20 cm	Log (1 + x)	0,8089	0,7870	Analytique 1
9 ...	Argile dans 0,20 cm	Log (1 + x)	0,7515	0,8807	Analytique 1
10 ...	Limon dans 0,20 cm	$\sqrt{x}$	0,8020	0,4967	Analytique 1
11 ...	Refus dans 0,20 cm	$\sqrt{x}$	0,8130	0,8460	Analytique 1
12 ...	Argile dans l'échantillon profond	Log (1 + x)	0,8748	0,2242	Analytique 1
13 ...	Limon dans l'échantillon profond	$\sqrt{x}$	0,8168	0,1067	Analytique 1
14 ...	Refus dans l'échantillon profond	$\sqrt{x}$	0,8251	0,8262	Analytique 1
15 ...	Profondeur de l'horizon engorgé en hiver	Origine	0,8024	0,9170	Analytique 1
16 ...	Profondeur de l'argile	Log (1 + x)	0,8223	0,1815	Analytique 1
17 ...	Humidité équivalente dans 0,20 cm	Origine	0,8093	1,0877	Analytique 1
18 ...	pH dans 0,20 cm	Log (1 + x)	0,7721	0,7147	Analytique 1
19 ...	Capacité d'échange dans 0,20 cm	Log (1 + x)	0,8043	0,5387	Analytique 2
20 ...	Ca échangeable dans 0,20 cm	$\sqrt{x}$	0,7121	1,5487	Analytique 2
21 ...	Mg échangeable dans 0,20 cm	$\sqrt{x}$	0,8376	0,9437	Analytique 2
22 ...	K échangeable dans 0,20 cm	$\sqrt{x}$	0,7666	0,6544	Analytique 2
23 ...	Na échangeable dans 0,20 cm	$\sqrt{x}$	0,7691	1,3876	Analytique 2
24 ...	P <sub>2</sub> O <sub>5</sub> assimilable dans 0,20 cm	Origine	0,6234	3,1515	Analytique 2
25 ...	Série géomorphologique	Origine	0,8938	0,7877	Synthétique
26 ...	Humidité	$\sqrt{x}$	0,8393	0,0791	Synthétique
27 ...	Activité biologique	Log (1 + x)	0,8412	1,0182	Synthétique
28 ...	Indice	Origine	0,8205	0,1324	Forestière (à expliquer)

Le fichier des données de base est donc constitué par 28 variables observées sur 89 placettes ; avec le format de présentation des données adopté, il est perforé sur 445 cartes à 80 colonnes.

## 2.2. — LA SYNTHÈSE DES VARIABLES

### 2.21. *Les statistiques utilisées*

Que l'on veuille synthétiser ou expliquer, comme nous le verrons plus loin, il faut passer de l'ensemble des données de base (89 fois 28 nombres, soit 2 492 nombres) à un ensemble de statistiques qui leur sont équivalentes, à savoir :

- les 28 moyennes ;
- la matrice des coefficients de corrélation totale, soit  $27 \times 28/2 = 378$  nombres (tableau 2) ;
- les 28 variances qui permettent de reconstituer la matrice des variances et covariances.

Nous obtenons donc un ensemble de 434 nombres, plus la taille de l'échantillon (89) et le nombre de caractères (28) contre les 2 492 nombres du départ.

Dans tout notre exposé, nous nous référerons toujours à ces statistiques supposées suffisantes ; dans la réalité, comme nous avons avancé pas à pas dans les différentes analyses, ces matrices et ces vecteurs ont toujours été « reconstitués » par les divers programmes, ce qui ne représente pas une grande perte de temps, étant donné la façon dont sont confectionnés les divers programmes qui utilisent toujours des données brutes et non pas des données élaborées comme les statistiques précédentes.

### 2.22. *Corrélation partielle et analyse des composantes principales*

Pour essayer de comprendre la nature de leurs liens, nous avons fait, à partir de la matrice des coefficients de corrélation totale des variables du milieu, deux types de calcul :

1° *Calcul de la matrice des coefficients de corrélation partielle* : c'est-à-dire que nous avons calculé les coefficients de corrélation entre toutes les variables prises deux à deux en supposant les 25 autres variables maintenues constantes. En même temps, il était facile d'obtenir le coefficient de corrélation multiple de chaque variable en fonction des 26 autres (cf. RAO C. R., 1966).

2° *Analyse des composantes principales* : nous ne revenons pas sur les détails de la méthode [pour l'aspect théorique, cf. LAWLEY D. N. et MAXWELL A. E. (1963), HARMAN H. H. (1967), et pour l'étude d'un exemple forestier, DEBAZAC E. F. et TOMASSONE R. (1965), ESCOUFIER (1966)]. Nous rappelons que sous sa forme la plus simple cette méthode consiste à effectuer un changement de coordonnées en utilisant comme nouvelles variables les vecteurs propres de la matrice des coefficients de corrélation totale.

Ces nouvelles variables ont deux avantages :

- chacune est non corrélée avec toutes les autres ; on peut donc l'analyser séparément ;

Coefficients de corrélation to

	Variables analytiques I												
	1	2	3	4	5	6	7	8	9	10	11	12	
X <sub>1</sub> ..													
X <sub>2</sub> ..	58												
X <sub>3</sub> ..	7	1											
X <sub>4</sub> ..	-15	-1	2										
X <sub>5</sub> ..	37	20	13	39									
X <sub>6</sub> ..	12	13	14	29	31								
X <sub>7</sub> ..	3	-7	13	29	56	30							
X <sub>8</sub> ..	-1	-10	0	-28	12	-35	32						
X <sub>9</sub> ..	17	5	-12	-40	-18	-27	-40	28					
X <sub>10</sub> ..	5	8	-22	-35	-29	-35	-44	31	69				
X <sub>11</sub> ..	-12	-23	-1	6	-9	-5	-21	-23	1	2			
X <sub>12</sub> ..	-1	19	12	-11	-18	-24	-17	6	21	38	7		
X <sub>13</sub> ..	11	16	1	-18	-16	-36	-22	13	27	38	1	62	
X <sub>14</sub> ..	-15	-27	-3	21	6	4	-12	-8	4	2	43	-38	
X <sub>15</sub> ..	0	-25	-3	2	12	33	10	-24	-28	-45	14	-61	
X <sub>16</sub> ..	7	-16	-6	12	21	32	21	-16	-36	-54	-2	-82	
X <sub>17</sub> ..	7	4	-12	-37	-11	-42	-24	63	70	79	-14	38	
X <sub>18</sub> ..	-28	-14	-11	-3	-54	-24	-59	-32	25	35	14	17	
X <sub>19</sub> ..	-5	-7	-2	-35	-1	-43	7	82	37	39	-23	19	
X <sub>20</sub> ..	-12	2	-12	-36	-52	-48	-48	23	49	61	0	39	
X <sub>21</sub> ..	-8	1	-6	-39	-33	-44	-29	37	46	60	-11	42	
X <sub>22</sub> ..	-5	-7	2	-33	-30	-55	-39	40	50	54	20	43	
X <sub>23</sub> ..	6	-6	4	-28	-13	-30	-11	50	49	50	-5	28	
X <sub>24</sub> ..	-27	-25	-1	-5	-22	9	-15	9	14	9	6	-11	
X <sub>25</sub> ..	14	-14	-4	-9	15	25	22	-5	-19	-24	-13	-53	
X <sub>26</sub> ..	-6	14	6	-1	-16	-29	5	29	16	31	-16	49	
X <sub>27</sub> ..	-8	25	-17	-5	-40	-31	-33	-2	23	32	0	45	
X <sub>28</sub> ..	-24	32	-3	-1	-24	-14	-18	-1	12	4	-23	19	

La dernière ligne représente les coefficients de corrélation

2  
 e les 28 variables analysées (multipliés par 100)

				Variables analytiques II						Variables synthétiques		
15	16	17	18	19	20	21	22	23	24	25	26	27
73												
— 46	— 52											
— 25	— 31	16										
— 36	— 24	76	— 14									
— 47	— 54	61	67	43								
— 47	— 53	70	45	55	81							
— 42	— 47	64	29	52	63	64						
— 33	— 37	64	6	58	46	53	59					
4	11	— 4	14	5	16	3	— 3	10	11			
57	53	— 19	— 23	— 12	— 34	— 25	— 39	— 17	11			
— 69	— 55	36	18	34	37	34	26	22	— 11	— 40		
— 56	— 46	31	35	21	49	35	26	16	— 3	— 35	66	
— 28	— 18	9	4	14	18	6	— 1	1	3	— 23	30	57

avec la variable expliquée (Indice  $X_{28}$ ).

- chaque valeur propre associée à un vecteur propre représente une part positive de la variation dont la somme est strictement égale à  $p$ ; on peut donc ne s'attacher qu'à l'étude des valeurs propres suffisamment grandes. Dans la pratique courante de la méthode, on peut admettre que l'on conserve les valeurs propres supérieures à 1, et pour éviter de négliger des composantes dont les valeurs sont voisines de cette limite, nous prendrons les valeurs supérieures à 0,9<sup>(1)</sup>.

### 2.23. Matrice des coefficients de corrélation partielle

Si on se reporte au groupe des coefficients de corrélation partielle, on voit apparaître deux groupes de variables dont les coefficients de corrélation partielle à l'intérieur des groupes sont significatifs :

- d'un côté, la plupart des variables chimiques et quelques variables physiques ;
- d'un autre, les variables synthétiques et les variables forestières.

A ces groupes se rattachent, grâce à un seul coefficient significatif, des caractères comme le refus, alors que des caractères ne sont absolument pas corrélés, comme la position topographique, le sodium et la série géomorphologique (cf. figure 2).

### 2.24. Résultats de l'analyse des composantes principales

Nous avons essayé de tirer le meilleur parti de l'analyse en la décomposant en deux parties :

1° Les 27 variables sur lesquelles porte cette analyse sont de nature extrêmement différente ; dans l'interprétation des composantes, il pourrait être délicat d'interpréter une composante où les deux facteurs importants seraient, par exemple, la capacité d'échange et le recouvrement des mousses. Aussi, avons-nous effectué trois premières analyses sur les trois groupes « naturels » de variables :

- les variables analytiques 1,
- les variables analytiques 2,
- les variables synthétiques.

Techniquement, nous avons diagonalisé des sous-matrices de la matrice des coefficients de corrélation totale que nous avons donnée au tableau ci-dessus.

Les trois résultats sont ceux des tableaux ci-dessous.

La première composante est nettement liée à la texture du sol. Elle oppose les sols argilo-limoneux humides aux sols filtrants, secs. La deuxième composante caractérise assez bien le type d'humus et la troisième correspond à un lessivage superficiel. Sur la quatrième composante se détachent les sols caillouteux favorables aux mousses (cf. carte phyto-écologique).

Pour la cinquième et la sixième composantes, l'interprétation est beaucoup moins évidente.

(1) On effectue quelquefois sur les premières composantes, choisies selon des critères plus ou moins arbitraires, une *rotation* pour rendre les « facteurs » plus interprétables. Cette technique couramment utilisée chez les psychologues a donné d'excellents résultats dans des études phyto-sociologiques et écologiques récentes.

TABLEAU 3

*Analyse des composantes principales des variables analytiques 1*

Variables	Vecteurs propres	1	2	3	4	5	6
	Valeurs propres	5.01	2.37	1.87	1.34	1.04	0.92
	%	31.3	14.8	11.7	8.4	6.5	5.8
	% cumulé	31.3	46.1	57.8	66.2	72.7	78.4
X <sub>2</sub> (Position topo.)		.05	.15	.23	.07	.63	.50
X <sub>4</sub> (Rec. mousses)		.20	.02	.29	.43	-.45	-.05
X <sub>5</sub> (Ep. A <sub>0</sub> )		.20	.36	-.04	.38	-.11	.12
X <sub>6</sub> (C/N A <sub>0</sub> )		.26	.03	.12	.03	-.15	.60
X <sub>7</sub> (C/N A <sub>1</sub> )		.22	.46	.01	.10	-.04	-.14
X <sub>8</sub> (% Matière organique 0-20 cm)		-.15	.41	-.39	.06	.10	-.18
X <sub>9</sub> (% Argile)	0-20 cm	-.30	-.04	-.30	.09	-.03	.38
X <sub>10</sub> (% Limon)		-.37	-.04	-.20	.13	-.16	.18
X <sub>11</sub> (% Refus)		.01	-.32	.05	.49	.36	-.23
X <sub>12</sub> (% Argile)		-.30	.14	.42	.09	.11	-.06
X <sub>13</sub> (% Limon)	Echantillon profond	-.27	.07	.19	.08	.22	-.23
X <sub>14</sub> (% Refus)		.09	-.28	-.27	.56	.07	.01
X <sub>15</sub> (Prof. engorgée)		.30	-.18	-.25	-.11	.24	-.02
X <sub>16</sub> (Prof. argile)		.35	-.12	-.31	-.13	.05	-.04
X <sub>17</sub> (Humidité équivalente)		-.36	.16	-.29	.14	-.07	.12
X <sub>18</sub> (pH 0-20 cm)		.20	-.42	.16	-.05	-.23	.11

TABLEAU 4

*Analyse des composantes principales des variables analytiques 2*

Variables	Vecteurs propres	1	2
	Valeurs propres	3.32	1.05
	%	55.3	17.5
	% cumulé	55.3	72.8
X <sub>19</sub>	Capacité d'échange.	.41	.20
X <sub>20</sub>	Ca échangeable (0,20 cm)	.45	-.20
X <sub>21</sub>	Mg échangeable (0,20 cm).	.48	-.22
X <sub>22</sub>	K échangeable (0,20 cm)	.46	.00
X <sub>23</sub>	Na échangeable (0,20 cm)	.42	.12
X <sub>24</sub>	P <sub>2</sub> O <sub>5</sub> échangeable (0,20 cm)	.07	.92

La première composante est pratiquement la capacité d'échange du complexe absorbant et son degré de saturation en bases.

La deuxième composante est pratiquement confondue avec la richesse en acide phosphorique.

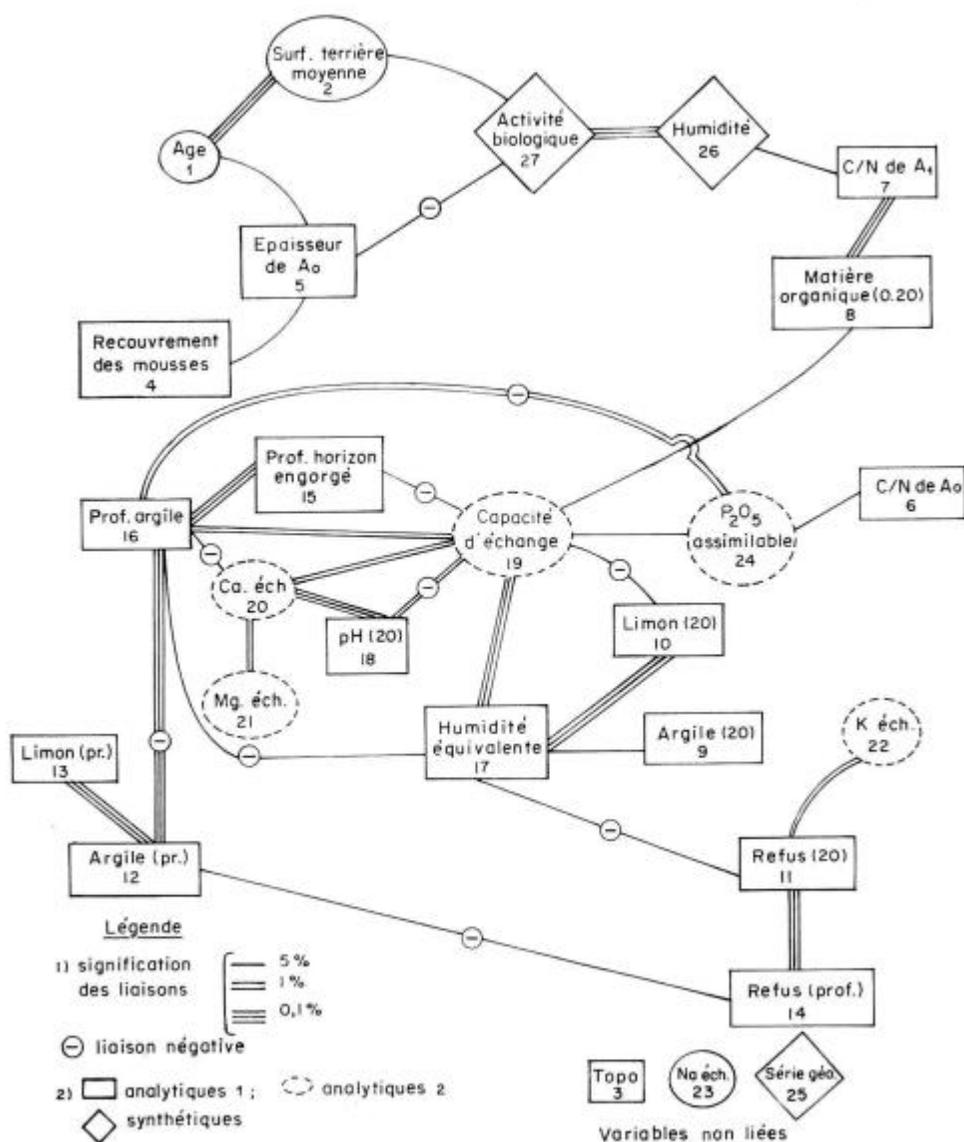


FIG. 2. — Graphe des coefficients de corrélation partielle

TABLEAU 5

*Analyse des composantes principales des variables synthétiques*

Variables	Vecteurs propres	1
	Valeurs propres	1.96
	%	65.3
	% cumulé	65.3
X <sub>25</sub> .....	Série géomorphologique	.49
X <sub>26</sub> .....	Humidité	-.62
X <sub>27</sub> .....	Activité biologique	-.61

Sur la première composante des variables synthétiques, qui absorbe presque les deux tiers de la variabilité totale, la série géomorphologique s'oppose nettement à l'ensemble constitué par l'humidité et l'activité biologique. Les deux autres composantes ont une valeur propre inférieure à l'unité et n'ont pas été conservées.

2° Une fois ces trois analyses partielles effectuées, nous avons fait une analyse des composantes principales sur l'ensemble des 27 variables (les 25 appartenant aux trois groupes précédents, plus l'âge du peuplement principal et la surface terrière moyenne).

Dans l'analyse globale, les valeurs élevées sur la première composante correspondent aux sols les plus argileux, les plus riches, qui sont également des sols humides et biologiquement actifs.

Cette composante est, *grosso modo*, la combinaison des premières composantes des trois analyses partielles précédemment examinées.

La deuxième composante est proche de la deuxième composante des variables analytiques 1, liée au type d'humus et à l'activité biologique.

La signification de la troisième composante n'est pas aussi évidente ; elle semble opposer des stations sur sols à texture grossière en profondeur, secs, à humus relativement doux à des stations plus humides, où le socle burdigalien est plus proche, et contraste avec des horizons superficiels appauvris en phosphore.

Le quatrième groupe essentiellement les deux variables « sylvicoles » (dans les pineraies âgées, la surface terrière est généralement plus grande) et l'interprétation des deux dernières est assez douteuse.

La conclusion la plus remarquable est que les deux premières composantes se placent très naturellement sur le diagramme de classification schématique des milieux forestiers de la Sologne (cf. figure 2) ; la troisième correspond à une « direction » plus rarement observée.

Pour les composantes suivantes, l'interprétation devient plus difficile, sans doute pour trois causes différentes :

— les placettes ne constituent pas un échantillon idéal, puisque leur choix était souvent lié à l'autorisation du propriétaire ;

— du point de vue écologique, elles étaient un peu trop grandes pour être homogènes ;

— la variabilité génétique des Pins sylvestres est très importante et elle introduit du « bruit » dans les relations entre la production et le milieu.

TABLEAU 6

*Analyse des composantes principales  
de toutes les variables réunies*

Variables	Vecteurs p Propres		1	2	3	4	5	6
	Valeurs Propres		8.36	2.94	2.85	1.68	1.54	1.07
	%		31.19	10.95	10.65	6.28	5.73	4.00
	% cumulé		31.19	42.14	52.79	59.07	64.80	68.80
For.	X <sub>1</sub>	Age du peuplement .....	-.02	.18	-.20	.56	-.09	.02
	X <sub>2</sub>	Surface terrière .....	.03	.01	-.35	.41	-.16	.18
Analytiques 1	X <sub>3</sub>	Position topo. ....	-.02	.01	-.16	-.07	.20	.23
	X <sub>4</sub>	Rec. mousses .....	-.14	-.12	-.18	-.15	.38	.10
	X <sub>5</sub>	Epaisseur de A <sub>0</sub> .....	-.14	.30	-.23	.15	.30	-.04
	X <sub>6</sub>	C/N de A <sub>0</sub> .....	-.20	.00	-.09	.12	.09	.46
	X <sub>7</sub>	C/N de A <sub>1</sub> .....	-.14	.31	-.23	-.29	.12	-.05
	X <sub>8</sub>	% Matière organique (0-20 cm)	.15	.44	.03	-.22	.06	-.07
	X <sub>9</sub>	% argile (0-20 cm) .....	.22	.09	.16	.30	.04	.18
	X <sub>10</sub>	% Limon (0-20 cm) .....	.27	.02	.11	.21	.03	.13
	X <sub>11</sub>	% Refus (0-20 cm) .....	-.01	-.21	.19	.14	.45	-.36
	X <sub>12</sub>	% argile dans	.22	-.13	-.26	-.01	.15	-.03
	X <sub>13</sub>	% Limon } l'échantillon	.20	-.05	-.13	.13	.04	-.25
	X <sub>14</sub>	% Refus } profond	-.06	-.06	.30	.06	.45	.13
	X <sub>15</sub>	Profondeur engorgée .....	-.24	-.06	.28	.11	-.09	-.10
	X <sub>16</sub>	Profondeur argile .....	-.26	.11	.21	.03	-.14	-.04
	X <sub>17</sub>	Humidité équivalente .....	.29	.22	.09	.08	.06	-.13
X <sub>18</sub>	pH (0-20 cm) .....	.14	-.39	.16	-.02	-.11	.12	
Analytiques 2	X <sub>19</sub>	Capacité d'échange .....	.21	.36	.07	-.19	.02	.02
	X <sub>20</sub>	Ca échangeable (0-20 cm) .....	.29	-.12	.11	-.01	-.15	.03
	X <sub>21</sub>	Mg échangeable (0-20 cm) .....	.28	.03	.06	-.04	-.15	-.01
	X <sub>22</sub>	K échangeable (0-20 cm) .....	.27	.03	.13	.04	.16	-.23
	X <sub>23</sub>	Na échangeable (0-20 cm) .....	.22	.22	.08	.02	.10	.02
X <sub>24</sub>	P <sub>2</sub> O <sub>5</sub> échangeable (0-20 cm) .....	.02	.02	.28	-.13	.18	.51	
Synthétiques	X <sub>25</sub>	Série géomorphologique .....	-.16	.18	.20	.07	-.25	.15
	X <sub>26</sub>	Humidité .....	.20	-.04	-.28	-.25	.04	.14
	X <sub>27</sub>	Activité biologique .....	.20	-.22	-.15	-.07	-.07	.16

## 2.3. — EXPLICATION DE LA PRODUCTION

2.31. *Les méthodes*

Lorsque l'on cherche à expliquer une variable dite « variable expliquée » en fonction d'un certain nombre d'autres variables, dites « variables explicatives » ou régresseurs, la méthode classique que l'on se propose d'appliquer immédiatement est celle de la régression, c'est-à-dire que l'on calcule les coefficients  $b_0, b_1, \dots, b_{27}$  de l'équation :

$$y = b_0 + b_1x_1 + \dots + b_{27}x_{27}$$

Ce faisant, nous aurons un « résultat » difficilement utilisable pour l'objet de notre enquête, le seul intérêt précis étant l'obtention des coefficients  $b_i$ , comme statistiques permettant de tester la corrélation partielle de l'indice et des 27 régresseurs. Toutefois, ce résultat, pour intéressant qu'il soit, est d'une utilisation limitée ; en effet, deux cas correspondant à des utilisations différentes de l'équation peuvent se présenter :

1° Ou bien nous voulons utiliser l'équation pour estimer l'indice avec une certaine marge d'erreur ; dans ce cas, il est bien évident qu'avec les liens existant entre les variables il y a un trop grand nombre de variables, et que nous pourrions nous contenter d'un nombre beaucoup plus réduit pour une précision équivalente. Il existe plusieurs techniques pour chercher la meilleure équation de régression en fonction d'un nombre fixé de variables, cf. DRAPER N. et SMITH H. (1966) ; nous avons utilisé la technique de la régression progressive qui consiste à introduire les variables pas à pas : la première variable introduite étant celle qui fournit la meilleure explication, c'est-à-dire celle dont le coefficient de corrélation totale avec l'indice est le plus grand en valeur absolue ; puis la seconde variable introduite est celle qui, après la première et avec elle, fournit le meilleur ajustement. Le processus de calcul se poursuit jusqu'à ce que toutes les variables soient introduites. A chaque palier, on peut tester aussi bien la signification de la nouvelle variable introduite que celle de toutes les variables apparues avant elle dans l'équation. Si à un palier une variable préalablement significative ne l'est plus, on peut l'enlever de la régression et calculer une nouvelle régression sans elle ; ceci permet d'améliorer la première méthode et de ne conserver dans la régression définitive que des variables dont la corrélation partielle est supérieure à une valeur fixée à l'avance.

2° Ou bien nous voulons utiliser l'équation comme un moyen d'investigation, c'est-à-dire comme une technique nous permettant d'expliquer l'indice en fonction de facteurs interprétables ; dans ce cas, on peut utiliser une technique plus originale qui consiste à :

a) Transformer les variables initiales en un nouvel ensemble de variables non corrélées ; nous faisons une analyse des composantes principales sur les régresseurs ; puis, pour interpréter plus facilement ces composantes, nous les multiplions par la racine carrée de leur valeur propre ; les coefficients des nouvelles composantes sont alors les coefficients de corrélation de la composante avec les variables initiales.

b) Effectuer la régression de la variable expliquée en fonction de ces nouvelles composantes. On obtient une équation de la forme :

$$y = c_0 + c_1f_1 + \dots + c_p f_p$$

qui représente les avantages suivants :

- Alors que dans la régression multiple classique, il n'est pas possible de dissocier l'effet d'une variable de celui des autres, sauf cas d'espèce, ici, les variables  $f_i$  ( $i = 1, p$ ) étant non corrélées, elles fournissent chacune, à l'explication de  $y$ , une part qui leur est spécifique ; il est donc possible d'analyser leur effet indépendamment.

- On peut ne conserver dans la régression que les variables  $f_i$  dont les coefficients de régression  $c_i$  sont significatifs (la multiplication par la racine carrée de la valeur propre permet d'obtenir des coefficients  $c_i$  qui représentent strictement les coefficients de corrélation — totale ou partielle — des variables  $f_i$  avec  $y$ ).

### 2.32. *Les résultats*

Nous avons procédé en deux temps ; d'abord, nous avons effectué une analyse progressive sur l'ensemble des 27 variables ; puis ces résultats nous ayant fourni une première information, nous avons fait deux nouvelles séries de régressions sur des ensembles réduits des variables initiales : une première série dans laquelle nous avons exclu les « variables peuplement » (âge et surface terrière) et le recouvrement des mousses, et une seconde dans laquelle nous n'avons conservé que les variables « faciles » à mesurer (ont été exclues les variables  $X_6, X_7, X_8, X_{20}, X_{21}, X_{22}, X_{23}, X_{24}$ , et la géomorphologie  $X_{25}$ ).

#### 2.321. *Analyse sur toutes les variables.*

Les variables introduites sont dans l'ordre  $X_{27}$ , puis  $X_{11}, X_1, \dots$  jusqu'à  $X_4$ , et l'on passe d'un coefficient de corrélation multiple 0.32 à un coefficient de 0,73 ; avec neuf variables, le carré du coefficient est presque égal à 0.70 <sup>(1)</sup>. Nous voyons donc qu'à partir d'un certain palier le gain devient négligeable pour le coefficient de corrélation et même l'écart-type résiduel augmente à cause de la diminution du nombre de degrés de liberté résiduels. Il est donc vraisemblable qu'une dizaine de variables au grand maximum sont utiles pour obtenir une précision égale à celle de l'équation avec les 27 variables.

#### 2.322. *Variables peuplement exclues.*

La première analyse sur l'ensemble de toutes les variables nous a donné quelques idées sur l'influence propre à chaque variable ; nous pouvons maintenant nous placer dans des conditions plus proches de la réalité. Nous avons tout d'abord éliminé la variable  $X_4$  (recouvrement des mousses), dont tous les coefficients de corrélation partielle avec les autres variables sont non significatifs et qui n'a aucune influence sur l'indice dans la régression globale. Puis nous avons essayé les deux types de régression (régression progressive et régression orthogonale) sur les deux mêmes sous-ensembles de variables explicatives.

##### 2.3221. *Suppression des variables « peuplement ».*

C'est-à-dire des variables  $X_1$  et  $X_2$  ; il reste dans ce cas 24 régresseurs ; et on aboutit à un carré du coefficient de corrélation multiple égal à 0.52. Les deux variables

(1) Rappelons que le carré du coefficient de corrélation multiple fournit le pourcentage de la variation totale de la variable  $y$  que l'équation de régression permet d'expliquer ; à ce titre le carré est plus intéressant que le coefficient lui-même.

TABLEAU 7

Régression orthogonale sur 24 variables ( $R^2 = 0,52$ ) (les coefficients significatifs des variables dans les composantes sont seuls conservés)

Variables	Vecteurs propres	1	2	3	4	5	6	7	8	9	10
	Valeurs propres	8.19	2.92	2.52	1.42	1.17	1.10	1.04	.87	.75	.63
	% cumulé	34.11	46.28	56.76	62.67	67.53	72.10	76.42	80.04	83.16	85.78
	C Coefficients de régression sur la composante	-0.24 **	-0.10	0.24 **	-0.24 **	-0.21 *	0.16	-0.01	0.17	0.05	-0.32 **
X <sub>3</sub> Position topo .....					.34	.30	.31	-.52	-.29	.38	
X <sub>5</sub> Epaisseur A <sub>0</sub> .....	.14	.32						.33	-.28		
X <sub>6</sub> C/N' de A <sub>0</sub> .....	.20						.47				.31
X <sub>7</sub> C/N de A <sub>1</sub> .....	.15	.39	.22								.43
X <sub>8</sub> % matière organique (0-20 cm) ..	-.15	.46									
X <sub>9</sub> % argile (0-20 cm) .....	-.22			-.24				.37			-.37
X <sub>10</sub> % limon (0-20 cm) .....	-.27							.37			
X <sub>11</sub> % refus (0-20 cm) .....					.58						.32
X <sub>12</sub> % argile dans	-.23		.34								
X <sub>13</sub> % limon } l'échantillon	-.20										
X <sub>14</sub> % refus } profond					.36	-.43			.48	.45	
X <sub>15</sub> Profondeur engorgée .....	.25			-.30						.36	
X <sub>16</sub> Profondeur argile .....	.27			-.25							
X <sub>17</sub> Humidité équivalente .....	-.29										
X <sub>18</sub> pH (0-20 cm) .....	-.15	-.40							-.30		
X <sub>19</sub> Capacité d'échange .....	-.21	.38									
X <sub>20</sub> Ca échangeable (0-20 cm) .....	-.29										
X <sub>21</sub> Mg échangeable (0-20 cm) .....	-.28										.36
X <sub>22</sub> K échangeable (0-20 cm) .....	-.27										
X <sub>23</sub> Na échangeable (0-20 cm) .....	-.22										
X <sub>24</sub> P <sub>2</sub> O <sub>5</sub> échangeable (0-20 cm) .....			-.28				.69		.39		
X <sub>25</sub> Série géomorphologique .....	.17		-.24							.52	
X <sub>26</sub> Humidité .....	-.20		.32			-.40					
X <sub>27</sub> Activité biologique .....	-.20					-.41					

peuplement permettaient d'arriver jusqu'à un carré du coefficient égal à 0.73. Il y a donc 21 % de la variation de l'indice qui est expliquée uniquement par l'action propre de l'âge et de la surface terrière (action propre signifiant action indépendante des autres variables).

Nous allons essentiellement essayer d'interpréter les résultats fournis par les deux méthodes, pour les 24 variables. Evidemment, les coefficients de corrélation multiples globaux sont strictement les mêmes dans les deux méthodes.

D'un côté, la méthode de la régression progressive introduit la variable  $X_{27}$  en tête ; cette variable explique à elle seule 32 % de la variation ; et au dernier palier c'est la seule variable dont l'effet partiel soit encore significatif.

D'un autre côté, cinq composantes sont significatives ; le carré du coefficient de régression sur les composantes fournit le pourcentage d'explication apporté par cette composante (cf. TOMASSONE R., *op. cit.*, 1967). Et la somme de ces cinq explications est égale à 32 %, soit à celle de la seule variable  $X_{27}$ . Mais si nous étudions de plus près ces cinq composantes nouvelles, nous voyons comment détailler l'action des facteurs écologiques principaux. Ces composantes peuvent avoir la signification suivante :

$f_1$  : nous retrouvons ici le caractère majeur détecté par l'analyse des composantes principales. Ce caractère oppose les sols humides, argileux, à humus actif, aux sols secs, sableux, à humus brut. Il regroupe la plupart des variables écologiques conservées.

$f_2$  : l'effet de cette composante sur la production n'est pas significatif.

$f_3$  : elle est très proche de la troisième composante du tableau 6 ; elle souligne l'importance bénéfique de l'humidité due aux horizons argileux profonds du socle burdigalien, même si les horizons superficiels sont lessivés et pauvres en  $P_2O_5$ .

$f_4$  et  $f_5$  : elles montrent que les stations situées sur les plateaux, même si elles sont caillouteuses en profondeur, sont plus favorables au Pin sylvestre que les stations situées dans les vallées au sol caillouteux dès la surface.

$f_{10}$  : cette petite composante n'est pas la moins importante pour la croissance du Pin sylvestre. Elle reprend certains des caractères de la première composante, et souligne l'importance de la texture superficielle et de l'activité biologique de l'humus liée au rapport carbone/azote (DUCHAUFOUR, 1960).

### 2.3222. Suppression des variables difficiles à mesurer.

Les 15 variables qui restent (tableau 8) permettent de retrouver très nettement l'effet de la première composante  $f_1$ , liée à l'humidité du sol, à la texture fine et à l'activité biologique.

Les autres composantes sont modifiées et d'interprétation plus délicate.

$f_2$  dépasse à peine le seuil de signification et son interprétation n'apparaît pas clairement.

$f_4$  oppose les stations basses où l' $A_0$  est épais, les horizons supérieurs caillouteux, les horizons profonds peu sableux et le Pin sylvestre peu productif aux stations opposées où il croît mieux.

$f_7$  montre que les sols mal drainés, pauvres en limon profond, où l' $A_0$  est épais

TABLEAU 8

Régression orthogonale (15 variables) ( $R^2 = .46$ )

Variables	Composante	1	2	3	4	5	6	7	8	9	10	11
	Valeurs propres	5.29	1.90	1.81	1.24	1.02	.87	.69	.55	.44	.32	.27
	% cumulé	35.27	47.94	60.01	68.26	75.04	80.84	85.46	89.11	92.06	94.20	96.00
	$C_i$ (coef. régression)	** .30	.18*	-.11	-.29**	.15	.00	.19*	.01	.31**	.02	.26**
$X_3$	Position topo. ....		.31		.45		-.78					
$X_5$	Rec. mousses ....		.25	.38	.32			-.42	.42		.37	
$X_9$	% argile dans 0-20 cm ....	.25	-.34	.25						.64		
$X_{10}$	% limon dans 0-20 cm ....	.33	-.32									-.40
$X_{11}$	% refus dans 0-20 cm ....		-.32	-.31	.56				-.47			
$X_{12}$	% argile	.32	.25		.27							
$X_{13}$	% limon	.26			.26			.58	.50			
$X_{14}$	% refus		-.46		.32	.50			.33		-.39	
$X_{15}$	Prof. engorgé: ....	-.34						.37				
$X_{16}$	Prof. argile ....	-.36										
$X_{17}$	Humidité équivalente ....	.33		.38								
$X_{18}$	pH (0-20 cm) ....		-.30	-.41					.30	-.38	.51	
$X_{19}$	Capacité d'échange (0-20 cm) ...	.22		.49				.36	-.34	-.32	.31	
$X_{26}$	Humidité ....	.30				.47						-.66
$X_{27}$	Activité biologique ....	.29		-.25		.40				.43		.36

et la capacité d'échange faible, sont moins productifs que les sols qui présentent les caractères opposés.

$f_0$  montre que le Pin sylvestre peut prospérer sur les sols argileux dès la surface, quand les groupes d'espèces indiquant une activité biologique suffisante sont présents, même si l'humus superficiel y est acide et désaturé.

$f_{11}$  permet de compléter les indications données par  $f_1$  : certains sols secs et peu limoneux, qui sont peu fertiles en règle générale, peuvent cependant être favorables au Pin sylvestre, si les espèces présentes en sous-bois témoignent que l'activité biologique reste satisfaisante.

## 2.4. — DISCRIMINATION ET CLASSEMENT

### 2.41. Méthode

La discrimination entre un certain nombre  $k$  de populations est effectuée au moyen de la méthode de l'analyse discriminante ; c'est le seul moment où les statistiques de base ne sont pas celles définies au § 11 ; nous décomposons ici la matrice des variances et covariances de l'ensemble des 89 placettes en deux parties : une liée à la variation entre les  $k$  populations, et une autre à la variation moyenne à l'intérieur des populations. L'ensemble de ces deux matrices et les vecteurs des moyennes de chaque population permettent :

1° de calculer un test global de différences entre les  $k$  populations ;

2° de calculer un ensemble de  $k$  fonctions discriminantes qui permettent de classer un individu à l'intérieur de l'une des  $k$  populations avec une probabilité maximale.

Tous ces problèmes ont déjà été étudiés et font l'objet d'une abondante bibliographie (cf. par exemple ANDERSON T. W., 1958 ; RAO C. R., 1966 ; TOMASSONE R., 1963).

### 2.42. Définition des populations

Nous avons effectué les coupures entre les populations à l'aide de l'indice de production  $X_{28}$  suivant le schéma suivant :

Minimum 10.51	1 <sup>re</sup> population	2 <sup>e</sup> population	...	k <sup>e</sup> population	Maximum 28.20
	L1	L2		Lk-1	

Le choix des limites  $L_1, L_2, \dots, L_{k-1}$  a été délicat. D'abord, nous nous sommes donnés un nombre *a priori* raisonnable de populations, qui sont ici les types de peuplement, trois ou quatre.

Les limites pouvaient être choisies soit pour avoir des intervalles de classe de l'indice égaux, soit des effectifs pour les populations échantillonnées pratiquement égaux. La première procédure consiste à faire sur la variable  $X_{28}$  des coupures à des nombres fixes d'écart-type : les effectifs extrêmes sont plus faibles ; la deuxième conduit à des classes centrales d'amplitude plus faible mais où la valeur des bornes est fonction de l'échantillon.

## 2.43. Résultats

## 2.431. Variables utilisées.

Nous avons étudié deux groupes de neuf variables :

— soit les variables introduites les premières dans la régression de l'indice (sauf  $X_{22}$  qui précédait  $X_6$ ) :

$X_1$	Age	}	Groupe I
$X_2$	Surface terrière		
$X_6$	C/N de $A_0$		
$X_7$	C/N de $A_1$		
$X_9$	% Argile (0 - 20 cm)		
$X_{10}$	% Limon (0 - 20 cm)		
$X_{11}$	% Refus (0 - 20 cm)		
$X_{18}$	pH		
$X_{27}$	Activité biologique		

— soit les variables dont la corrélation partielle était la plus forte une fois toutes les variables introduites (sauf  $X_{22}$  qui précédait  $X_{13}$  et  $X_{10}$ ) :

$X_1$	Age	}	Groupe II
$X_2$	Surface terrière		
$X_3$	Position topographique		
$X_9$	Argile		
$X_{10}$	Limon		
$X_{13}$	Limon profond		
$X_{14}$	Refus profond		
$X_{18}$	pH		
$X_{27}$	Activité biologique		

2.432. Les résultats sont condensés dans le tableau 9 où nous trouvons :

- un test global permettant d'affirmer que les  $k$  populations sont significativement différentes pour l'ensemble des variables ;
- le pourcentage de placettes bien classées à l'aide des fonctions discriminantes.

Et nous pouvons voir :

- que les résultats sont meilleurs pour trois populations que pour quatre ;
- que le découpage en classe d'effectifs égaux est légèrement meilleur ;
- que le choix des variables de l'un ou l'autre des groupes n'est pas très important (mais dans notre exemple, les deux groupes ont six variables communes).

2.44. En regardant le tableau 10, on peut voir la parfaite concordance entre les deux analyses, régression et analyse discriminante : quand on passe des populations 1 à 3, les coefficients sont d'autant plus faibles qu'ils ont une action négative sur l'indice et d'autant plus forts qu'ils ont une action positive.

TABLEAU 9

*Comparaison des différents résultats de l'analyse discriminante*

Groupe de variables			I	II
3 populations	Effectifs égaux .....	Test ( $\chi^2$ à 18 dl)	142.53	143.31
		% total bien classés	67 %	68 %
	Intervalle de classe égal .....	Test ( $\chi^2$ à 18 dl)	96.63	98.21
		% total bien classés	63 %	62 %
4 populations	Effectifs égaux .....	( $\chi^2$ à 27 dl)	159.47	143.71
		% total bien classés	60 %	54 %
	Intervalle de classe égal .....	Test ( $\chi^2$ à 27 dl)	160.49	148.44
		% total bien classés	58 %	55 %

TABLEAU 10

*Fonctions discriminantes pour les 3 populations*

Variabes	Population I faible indice ( $< 16.8$ )	II indice moyen	III fort indice ( $> 20.00$ )	Action dans la régression de l'indice
$X_1$ .....	1.23	1.18	1.05	---
$X_2$ .....	-0.67	-0.49	-0.27	++
$X_{11}$ .....	82.58	80.16	79.45	-
$X_7$ .....	181.67	181.49	179.43	-
$X_9$ .....	10.10	11.00	12.46	++
$X_{10}$ .....	5.06	4.23	3.96	---
$X_{11}$ .....	9.32	9.05	8.47	-
$X_{18}$ .....	808.95	804.34	786.19	---
$X_{27}$ .....	-10.80	-5.66	2.91	++
Constante .....	-1161.96	-1153.62	-1120.20	

### III. — CONCLUSIONS

Au terme de cette étude, nous pouvons, semble-t-il, dégager deux séries de remarques, concernant d'une part les résultats pratiques obtenus et d'autre part la méthodologie. Il nous faut également donner quelques aperçus sur l'extension possible des méthodes que nous avons utilisées, ainsi que sur l'aspect technique de la réalisation des calculs qu'impliquent ces méthodes.

#### 3.1. — Résultats obtenus

Sur le plan pratique, cette méthode souligne les difficultés très grandes qui apparaissent lorsqu'on veut étudier l'action du milieu sur la production, sans se référer à des variables forestières, c'est-à-dire lorsqu'on se trouve dans la position du reboiseur devant un terrain à planter. On constate que, dans ces conditions, 50 % environ de la productivité potentielle reste inexplicée par les 24 variables non forestières. Ce résultat peut sembler décevant. Certes, et nous l'avons signalé, le matériel utilisé pour cette étude méthodologique est loin d'être parfait. Le choix même du Pin sylvestre, avec sa grande variabilité génotypique, nous plaçait d'ailleurs dans les plus mauvaises conditions possibles.

On peut presque considérer qu'à *contrario* ce résultat souligne l'importance du choix des provenances dans le cas de l'espèce étudiée. L'intérêt économique très limité du Pin sylvestre, qu'il convient, de l'avis général, d'utiliser de moins en moins en Sologne, souligne assez l'intérêt essentiellement méthodologique de notre travail. On peut sans doute espérer de meilleurs résultats, et d'un intérêt pratique plus immédiat, en travaillant avec d'autres espèces et dans d'autres régions.

Par contre, la synthèse objective effectuée par l'analyse multivariable permet incontestablement de préciser les types de pineraies qu'on peut trouver en Sologne et les facteurs qui peuvent le mieux les caractériser. L'analyse des composantes principales est à cet égard très significative. Il est intéressant — et rassurant — de voir réapparaître dans les deux premières composantes les critères utilisés, par ailleurs, par les écologistes pour classer les milieux forestiers en Sologne. Il est également assez satisfaisant de voir jouer un rôle particulier et distinct de la « richesse » globale du milieu à l'acide phosphorique. Son rôle privilégié est bien connu, dans la fertilisation du Pin maritime par exemple. De tels résultats peuvent donc éventuellement ouvrir la voie à des recherches plus fines, en attirant l'attention sur certains facteurs particuliers.

#### 3.2. — Efficacité des méthodes employées

Les méthodes employées sont surtout efficaces dans la synthèse d'un grand nombre de variables, et par conséquent dans le classement qu'elles permettent de faire entre peuplements. Il s'agit plus de décrire que d'expliquer, même dans le cas des régressions progressives qui conduisent finalement à trier les variables influençant le plus l'indice de productivité choisi. La comparaison des méthodes est très satisfaisante, puisqu'on n'aboutit à aucune contradiction majeure et qu'au contraire on observe une bonne concordance entre les résultats obtenus par les différentes méthodes. Pouvoir classer les peuplements, sans erreur dans 68 % des cas, en trois groupes objectivement distincts, à partir de neuf caractères seulement, nous semble un résultat intéressant et

sans doute très en deçà des possibilités offertes par la méthode, pour les raisons déjà signalées ci-dessus.

### 3.3. — *Utilisation d'autres méthodes*

3.31. Le but de l'étude était de synthétiser et d'expliquer des caractères, puis de discriminer et de classer des placettes ; le choix des méthodes a été imposé à la fois par leur simplicité et par l'existence de programmes de calcul pour ordinateur. Depuis qu'elle est terminée, nous avons appliqué d'autres méthodes et nous avons mis au point d'autres programmes. C'est à la lumière de cette évolution récente que nous voulons montrer de nouveaux développements possibles. Nous ne reviendrons pas sur la *rotation des facteurs* dont nous avons déjà parlé au cours de l'analyse des composantes principales, cf. paragraphe 2.12 ci-dessus ; cette technique a déjà donné d'excellents résultats et des mises au point concernant son utilisation sont en cours <sup>(1)</sup>. Il reste alors trois grands axes d'approche : l'analyse des coefficients de corrélation canonique, l'application des méthodes de taxinomie numérique et l'analyse des variables canoniques.

#### 3.32. *Coefficients de corrélation canonique.*

Nous avons des groupes de variables déjà bien définis physiquement :

- d'un côté, des variables de milieu avec les sous-groupes : forestier, analytique 1 et 2, synthétique ;
- de l'autre, la variable de production ; il serait d'ailleurs souhaitable d'introduire plusieurs variables de production, dans la mesure où celle-ci peut difficilement être confondue avec un seul paramètre.

Les analyses concernant ces groupes ou ces sous-groupes répondent à deux objectifs :

1° Les groupes sont-ils « indépendants » ? Il existe des tests statistiques permettant de le vérifier pour deux groupes, cf. ANDERSSON T. W. (1958), HOTELLING H. (1936), et même pour plusieurs, cf. KULLBACK S. (1959).

2° Dans le cas de deux groupes non indépendants, comment définir leurs liens de structure ? On recherche les combinaisons linéaires des variables d'un groupe dont les coefficients de corrélation, dits coefficients de corrélation canonique, avec des combinaisons linéaires des variables de l'autre groupe, sont les plus grands possibles.

#### 3.33. *Taxinomie numérique.*

Si nous pouvons définir des mesures de dissemblance <sup>(2)</sup> entre les placettes, il est possible, à l'aide d'algorithmes de calcul, d'effectuer des regroupements progressifs entre placettes voisines et de construire des classifications hiérarchiques relativement objectives. Ces algorithmes existent en très grand nombre dans les disciplines les plus variées

(1) De façon certaine l'analyse factorielle des correspondances donnerait aussi des résultats intéressants ; cf. BENZECRI J.-P. (1965).

(2) On emploie très souvent le terme de distance, bien que les expressions quantitatives utilisées ne soient pas toujours des distances au sens axiomatique du terme ; en particulier, l'inégalité triangulaire classique n'est pas toujours vérifiée. En pratique, on utilise des distances géométriques (la distance au sens cartésien du terme), des distances statistiques (distance de Mahalanobis) ou des coefficients de ressemblance plus généraux, cf. GOWER J.-C. (1968).

(botanique systématique, économie, linguistique, documentation automatique, archéologie, etc...); les principes essentiels sont exposés dans SOKAL R. R. et SNEATH P. H. A. (1963), et notre objectif actuel consiste plus à classer les différentes méthodes qu'à en découvrir de nouvelles, cf. MILLIER C. et TOMASSONE R. (1969).

### 3.34. *Analyse des variables canoniques.*

Si nous pouvons définir des populations contenant un certain nombre de placettes, ou si nous avons plusieurs échantillons par placette <sup>(1)</sup>, l'analyse des variables canoniques permet de déterminer les nouvelles variables et de tester leur signification pour discriminer au mieux les populations. Cette analyse apparaît comme une synthèse de l'analyse de variance à plusieurs variables et de l'analyse des composantes principales; elle a donné, dans des domaines aussi variés que l'agronomie et la linguistique, des résultats excellents, pour l'aspect théorique, cf. RAO C. R. (1966). On démontre que les nouvelles variables, dites variables canoniques, rendent maximum l'ensemble des distances de Mahalanobis entre toutes les populations, cf. GOWER J. C. (1966 a) et GOWER J. C. (1966 b).

### 3.4. — *Traitement automatique des données*

Nous avons vu que l'ensemble des techniques d'analyse étaient bien connues, et leur application théorique ne pose pas de problèmes très délicats, au moins si nous nous limitons à des techniques d'investigation et non à des tests d'hypothèses. Par contre, si nous voulons les appliquer de façon intensive sur de grands nombres de relevés avec beaucoup de variables, nous nous rendons compte que les programmes généralement utilisés dans les bibliothèques d'ordinateurs sont mal conçus et manquent de souplesse; il faudrait, à partir d'un fichier de données de base, pouvoir effectuer tous les calculs de façon entièrement automatique. Il faudrait aussi pouvoir effectuer des analyses complémentaires sur des sous-ensembles de données ou sur de nouvelles données que des résultats partiels auraient suggéré d'analyser. En un mot, il faudrait pouvoir accéder très rapidement à toutes les données, ou à une partie seulement, et les traiter par les méthodes que nous avons exposées. Alors, en quelques heures, c'est-à-dire pour quelques centaines de francs, un petit ordinateur comme le 1130 IBM est capable de fournir tous les résultats; un ordinateur plus gros les fournirait encore plus vite et à un moindre coût.

*Reçu pour publication en juin 1969.*

### SUMMARY

COMPARISON OF VARIOUS METHODS OF DATA'S ANALYSIS AND INTERPRETATION IN THE STUDY OF THE CORRELATION BETWEEN SEVERAL ECOLOGICAL VARIABLES AND THE SCOT'S PINE PRODUCTION IN SOLOGNE.

The paper is for essential devoted to the presentation of methods of data's analysis and interpretation. The study of the correlation between several ecological variables and

(1) Il y aurait alors autant de populations que de placettes, l'essentiel dans cette méthode étant d'avoir deux niveaux de variation: l'un entre populations, l'autre à l'intérieur des populations. Les populations regroupant un certain nombre de placettes pourraient être obtenues par les techniques de taxinomie numérique.

the Scot's pine production in Sologne, is carried out by the use of several multivariate technics.

- principal component analysis,
- regression, with special emphasis on the orthogonalised one,
- discriminant analysis.

Some possible extensions are noted and the special links between method and analysis on computer are outlined.

All the FORTRAN programs are available.

## ZUSAMMENFASSUNG

VERGLEICH ZWISCHEN VERSCHIEDENEN STATISTISCHEN INTERPRETATIONSMETHODEN VON DEN KORRELATIONEN ZWISCHEN DER ERTRAGSLEISTUNG DER WEISSKIEFER IN DER SOLOGNE UND VERSCHIEDENEN STANDORTSVARIABLEN.

Die vorliegende Arbeit behandelt vorwiegend die Darstellung verschiedener Interpretationsmethoden. Um die Korrelationen zwischen der Ertragsleistung der Weisskiefer in der Sologne und verschiedenen Standortvariablen zu untersuchen, wurden verschiedene Methoden der multivariaten Analyse verwendet und zwar :

- die Prinzipalkomponentenanalyse,
- die Regressionsanalyse, vor allem die orthogonale Regressionsanalyse,
- die Diskriminanzanalyse.

Erweiterungen auf zukünftige Untersuchungen sind vorgesehen. Die Zusammenhänge zwischen den rein methodologischen Aspekten und der praktischen Computer-Rechenarbeit werden besprochen. Alle verwendeten Rechenprogramme (FORTRAN IV) können zu Verfügung gestellt werden.

## REFERENCES BIBLIOGRAPHIQUES

- ANDERSON T. W., 1958. *An introduction to multivariate statistical analysis*. John Wiley, New-York, 374 p.
- BENZECRI J. P., 1965. Sur l'analyse factorielle des proximités. *Pub. I.S.U.P.*
- BRUCE O., SCHUMACHER F., 1950. *Forest mensuration*, 3<sup>e</sup> éd. Mc Graw-Hill Book Company, New York, 484 p.
- C.E.P.E., 1968. *Code pour le relevé méthodique de la végétation et du milieu*. C.N.R.S., Paris, 292 p.
- DEBAZAC E. F., TOMASSONE R., 1965. Contribution à une étude comparée des pins méditerranéens de la section halepensis. *Ann. Sci. forest.*, **22** (2), 215-256.
- DECOURT N., 1965. Le Pin sylvestre et le Pin Laricio de Corse en Sologne. Tables de productions provisoires et méthodes utilisées pour les construire. *Ann. Sci. forest.*, **22**, 2, 259-318.
- DENIZOT G., 1927. Les formations continentales de la région orléanaise. *Thèse*, Launoy, Vendôme.
- DIXON W. J., 1965. Regression on principal components, dans « *Biomedical Computer Programs* », 159-168. Univ. Calif., Los Angeles.
- DRAPER N., SMITH H., 1966. *Applied regression analysis*, John Wiley, New-York, 408 p.
- DUCHAUFOUR Ph., 1960. Stations, types d'humus et groupements écologiques, *Rev. forest. fr.*, 484-494.
- ESCOUFIER Y., 1966. Analyse des composantes principales. Utilisation de groupe de variables dans la recherche de la solution. *Thèse de spécialité, Fac. Sc., Montpellier*.
- GODRON M., et al., 1964. Notice détaillée. *Carte phytocéologique et carte de l'Occupation des Terres en Sologne*. C.N.R.S., 192 p., 2 cartes couleur 1/10.000, 75 × 104, 2 cartes en noir 1/10.000, 75 × 104.
- GOWER J. C., 1966 a. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, **53**, 325-338.
- GOWER J. C., 1966 b. A Q-technique for the calculation of Canonical variates. *Biometrika*, **53**, 588-589.
- GOWER J. C., 1968. A general coefficient of similarity and some of its properties. *Rothamsted Experimental Station*.

- HARMAN H. H., 1967. *Modern factor analysis*. Univ. Chicago Press, 474 p.
- HOREMANS P., 1961. Contribution à l'étude pédologique des terrasses alluviales de la Loire moyenne. *Thèse, Paris*, 2 cartes, 164 p.
- HOTELLING H., 1936. Relations between two sets of variates, *Biometrika*, **28**, 321-377.
- KULLBACK S., 1959. *Information theory and statistics*. John Wiley, New-York, 394 p.
- LAWLEY D. N. et MAXWELL A. E., 1963. *Factor analysis as a statistical method*. Butterworths, Londres, 118 p.
- LE HOUEROU H.-N., 1961. Contribution à l'étude des sables et argiles de Sologne. *D.E.S. Montpellier*, 73 p.
- MILLIER C. et TOMASSONE R., 1969. Les méthodes de classification et d'Ordination : leurs possibilités et leur limite. Coll. Intern. C.N.R.S., *Problèmes Mathématiques en Archéologie*, Marseille.
- PEARSON E. S., HARTLEY H. O., 1958. *Biometrika tables for statisticians*. Vol. 1, Cambridge University Press, 240 p.
- POISSONET J., 1968. Essai d'approche quantitative de l'interprétation des thèmes cartographiques de la végétation et de son écologie en Sologne. *Doc. n° 44 - C.E.P.E. - C.N.R.S.*, Montpellier (France).
- RAO C. R., 1966. *Linear statistical inference and its applications*. John Wiley, New-York, 522 p.
- SOKAL R. R. and SNEATH P. H. A., 1963. *Principles of numerical taxonomy*. Freeman, San Francisco, 359 p.
- TOMASSONE R., 1963. Application des fonctions discriminantes à des problèmes biométriques. *Ann. E.N.E.F.*, **20**, 4, 585-617.
- TOMASSONE R., 1967. Une méthode d'investigation : la régression orthogonale. *Ann. Sci. forest.*, **24** 3, 233-258.
- VATAN A., 1937. Sur la présence de sables éoliens sur la bordure sud-est de la Sologne. *C.R. Soc. géol. Fr.*, 179 p.
- VINCENT 1961. Is height/age a reliable index of site? *For. Chron.*, **37**, 144-149.
- WACQUANT J.-P., 1968. *La genèse des sols de Sologne* (en préparation).
-