



MICROFICHE N°

03998

République Tunisienne

MINISTÈRE DE L'AGRICULTURE

CENTRE NATIONAL DE
DOCUMENTATION AGRICOLE
TUNIS

الجمهورية التونسية
وزارة الزراعة

المركز القومي
للتوثيق الفلاحي
تونس

F 1

- INFORMATIQUE ET STATISTIQUE APPLIQUEES -

" Cours International de Statistique et
Informatique Appliquées - Gembloux -
20 Août - 8 Septembre 1978 "

R A P P O R T D E M I S S I O N

Par : EL-AMANI Med. ZAKI , Ingénieur
Principal Pédologue

Septembre 1 9 7 8

=====

- BUT DE LA MISSION -

L'abondance des documents des études pédologiques et encore beaucoup plus les données qu'ils renferment, les possibilités de les archiver facilement et de les étudier méthodiquement, a incité les responsables à créer une banque de données de sols de la Tunisie. Dans le cadre d'initier et familiariser les pédologues à cette nouvelle option de travail, et afin que la banque puisse fonctionner normalement et dans les meilleures conditions, j'étais envoyé à la Faculté Agronomique de Gembloux (Belgique) pour suivre les " Cours de Vacances d'Informatique et Statistique Appliquées et assister à la " Semaine Internationale de Statistique et Informatique en Agronomie " .

- Ces cours suivis des applications m'ont permis de :

- 1)- s'initier à l'informatique et à la statistique appliquées.
- 2)- de se rendre compte des possibilités d'utilisation de ces deux outils pour résoudre les problèmes des traitements de données, pratiquement impossibles de les faire manuellement et en un temps très court.
- 3)- d'avoir une idée sur les avantages, les limites et les contraintes d'utilisation des ordinateurs.
- 4)- de voir de près la banque de données de sols de la Belgique (Service de Sciences de sols à Gembloux).

Nous exposerons succinctement dans ce rapport les thèmes et les principaux sujets évoqués au cours de mon séjour à Gembloux.

- S O M M A I R E -

	Pages
<u>INTRODUCTION A L'INFORMATIQUE</u>	2
Codification et enregistrement de données	2'
1 - Système de numération	2'
2 - Enregistrement de données	7
<u>MATERIEL (Hardware)</u>	10
Généralités	10
Organes de mémoires	10
Organes d'entrée et de sortie	13
<u>LOGICIEL (Software) : Notions générales</u>	17
Programmation	17
Système d'exploitation	24
<u>LE STATICIEN et L'ORDINATEUR</u>	27
Les moyens modernes de calcul et leur emploi	27
Quelques conséquences de l'emploi de l'ordinateur	29
<u>BANQUES ET TRAITEMENT DE DONNEES</u>	32
Banques de données de sols	32
Enregistrement de données de description	36
Edition en clair	36
Traitement de données	36
Exemples de programmes de calculs pédologiques relatif à l'analyse granulométrique	37
<u>BANQUES DE DONNEES FORESTIERES</u>	43
<u>LES PROGRAMMES STATISTIQUES</u>	44
Contenu et organisation des fichiers	44
Caractéristiques et organisation des programmes	45
 <u>CONCLUSION :</u>	
 <u>BIBLIOGRAPHIE :</u>	

- INTRODUCTION -

En Tunisie, les études pédologiques ont été sérieusement entreprises depuis une trentaine d'années. Une bibliothèque des archives se trouve au sein de la Division des Sols, comportant environ 600 études. Ces documents renferment une masse énorme de données, qui pourront être archivées et utilisées ultérieurement.

Vu l'importance des données, et la forme dans laquelle elles existent actuellement, il est difficile de les exploiter. D'où alors le recours à l'ordinateur s'impose :

Ce dernier présente les avantages suivants :

- une capacité de mémoire importante,
- vitesse d'exécution rapide et accès quasi-instantanée,
- possibilité de mise à jour rapide et permanente de l'information mémorisée,
- facilité de communication et d'échange d'information.

C'est dans l'optique de rationalisation des méthodes d'études qu'une banque de données de sols a été lancée au sein de la Division. Elle permettra de résoudre, entre autre grâce à l'emploi de l'ordinateur, le problème d'archivage des études pédologiques. Au moyen des programmes proposés conçus par les utilisateurs, l'exploitation rationnelle de ces données est facilitée.

(*)
- INTRODUCTION A L'INFORMATIQUE -

0.1.- Définitions

Académie française (1967)

L'informatique est la science du traitement rationnel, notamment par machines automatiques, de l'information, considérée comme le support des connaissances et des communications dans les domaines techniques, économique et social.

Dictionnaire de la langue française ("Petit Robert ", 1973)

Informatique : traitement de l'information.

Ordinateur : calculateur électronique doté de mémoires à grande capacité et de moyens de calcul ultra-rapide, pouvant adapter son programme aux circonstances et prendre des décisions complexes.

Programme : ensemble ordonné (et formalité) des opérations nécessaires et suffisantes pour obtenir un résultat; dispositif permettant à un mécanisme d'effectuer ces opérations.

Programmation : élaboration et codification de la suite d'opérations formant un programme sur machine.

Journal officiel (français)

Petit Larousse illustré (1974)

Informatique : technique du traitement automatique de l'information.

Ordinateur : calculateur universel composé d'un nombre variable d'unités spécialisées, commandées par un même programme enregistré, et qui permet, sans intervention humaine en cours de travail, d'effectuer des ensembles complexes d'opérations arithmétiques et logiques.

Hardware : ensemble des éléments qui constituent un calculateur électrique, le calculateur étant considéré du point de vue de sa réalisation.

Software : ensemble des activités qui ont pour objet la conception et l'emploi des calculateurs électriques (codification, organisation, analyse, programmation, etc...).

(*) Tiré du Cours mis à la disposition des étudiants de la Faculté de Sciences (cours informatique et programmation).

Programme : ensemble d'instruction, de données ou d'expressions enregistrées sur un support et nécessaires à l'exécution d'une suite d'opérations déterminées, demandées à un calculateur, à un appareillage automatique ou à une machine-outil.

Programmation : établissement d'un programme.

1.- Codification et enregistrement des données

1.- Systèmes de numération

1.1.- Système binaire pur

Principe

Chiffres : 0 à 1

Nombres : $0_{(10)} = 0_{(2)}$

$1_{(10)} = 1_{(2)}$

$2_{(10)} = 10_{(2)}$

$3_{(10)} = 11_{(2)}$

$4_{(10)} = 100_{(2)}$

⋮
⋮
⋮

Unité binaire : bit (Binary digit).

Puissances de 2

<u>k</u>	<u>2^k</u>
0	1
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1.024
11	2.048
12	4.096
⋮	⋮
⋮	⋮
⋮	⋮

Correspondance approximative : 10 bits = 3 chiffres décimaux ou 1 chiffre décimal = 3,3 bits.

Conversion décimal-binaire

3753	- 2 ¹¹
- 2048	
1705	- 2 ¹⁰
- 1024	
681	- 2 ⁹
- 512	
169	- 2 ⁷
- 128	
41	- 2 ⁵
- 32	
9	- 2 ³
- 8	
1	- 2 ⁰
- 1	
0	

$$3753_{(10)} = (1 \times 2^{11} + 1 \times 2^{10} + 1 \times 2^9 + 0 \times 2^8 + 1 \times 2^7 + 0 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0)_{(2)}$$

$$= 111010101001_{(2)}$$

Conversion binaire-décimal

$$101101010_{(2)} = (1 \times 2^8 + 0 \times 2^7 + 1 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 1 \times 2^1 + 0 \times 2^0)_{(10)}$$

$$= (256 + 64 + 32 + 8 + 2)_{(10)} = 362_{(10)}$$

Addition

0 + 0 = 0	100110	38
0 + 1 = 1		
1 + 0 = 1	+ 10101	+ 21
1 + 1 = 0 + report	111011	59

Soustraction

0 - 0 = 0	1101001	105
0 - 1 = 1 + emprunt		
1 - 0 = 1	- 110101	- 53
1 - 1 = 0	110100	52

Multiplication

1001	9	1101011	107
× 101	× 5	× 11010	× 26
1001	45	0	642
		1101011	214
		1101011	2782
		10101101110	

Division

$$\begin{array}{r}
 101101010 \\
 - 10001 \\
 \hline
 10110 \\
 - 10001 \\
 \hline
 10110 \\
 - 10001 \\
 \hline
 101
 \end{array}$$

$$\begin{array}{r}
 10001 \\
 10101
 \end{array}$$

$$\begin{array}{r}
 362 \\
 - 34 \\
 \hline
 22 \\
 - 17 \\
 \hline
 5
 \end{array}
 \quad
 \begin{array}{r}
 17 \\
 21
 \end{array}$$

Caractères " alphanumériques "

1 caractère (lettre, chiffre décimal, signe de ponctuation, etc...) = 6 à 8 bits.

1.2.- Système " décimal codé binaire "

Principe

1 chiffre décimal = 4 bits (utilisés incomplètement).

Exemple : $1573_{(10)} = 001/0101.0111.0011_{(dcb)}$

Avantages et inconvénients : conversions plus simples (plus rapides),

calculs plus compliqués (plus lents), occupation de mémoire plus importante (perte de place :

$$\frac{4 - 3,3}{3,3} = 20 \%).$$

Caractères " alphanumériques "

1 caractère = 2 chiffres décimaux = 8 bits = 1 octet ou " byte ".

Exemple (code IBM 1620) :

$$A = 41_{(10)} = 0100/0001_{(dcb)}$$

$$3 = 42_{(10)} = 0100/0010_{(dcb)}$$

$$\begin{array}{l}
 \vdots \\
 I = 49_{(10)} = 0100/1001_{(dcb)} \\
 J = 51_{(10)} = 0101/0001_{(dcb)}
 \end{array}$$

$$\begin{array}{l}
 \vdots \\
 R = 59_{(10)} = 0101/1001_{(dcb)} \\
 S = 62_{(10)} = 0110/0010_{(dcb)}
 \end{array}$$

$$\begin{array}{l}
 \vdots \\
 Z = 69_{(10)} = 0110/1001_{(dcb)} \\
 \text{etc...}
 \end{array}$$

1.3.- Système octal

Principe

Chiffres : 0, 1, 2, 3, 4, 5, 6, 7.

1 chiffre octal = 3 bits

Principale utilisation : lecture et écriture simplifiées de nombres binaires purs.

Puissances de 8

<u>K</u>	<u>8^k</u>
0	1
1	8
2	64
3	512
4	4.096
5	32.768
:	:
:	:
:	:

Conversion décimal-octal et octal-décimal

Même principe que pour le système binaire pur :

$$\begin{aligned} 1265_{(8)} &= (1 \times 8^3 + 2 \times 8^2 + 6 \times 8^1 + 5 \times 8^0)_{(10)} \\ &= (512 + 128 + 48 + 5)_{(10)} = 693_{(10)} \end{aligned}$$

Conversion binaire-octal et octal-binaire

Même principe que pour le système "décimal codé binaire":

$$1265_{(8)} = 001/010/110/101_{(2)}$$

1.4.- Système hexadécimal

Principe

Chiffres : 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F,

$$\text{avec : } A_{(16)} = 10_{(10)} = 1010_{(2)},$$

$$B_{(16)} = 11_{(10)} = 1011_{(2)},$$

$$C_{(16)} = 12_{(10)} = 1100_{(2)},$$

$$D_{(16)} = 13_{(10)} = 1101_{(2)},$$

$$E_{(16)} = 14_{(10)} = 1110_{(2)},$$

$$F_{(16)} = 15_{(10)} = 1111_{(2)}.$$

1 chiffre hexadécimal = 4 bits (utilisés complètement).

Puissances de 16

<u>K</u>	<u>16^k</u>
0	1
1	16
2	256
3	4.096
4	65.536
5	1.048.576
.	.
.	.

Conversion décimal-hexadécimal et hexadécimal-décimal

Même principe que pour le système binaire pur :

$$\begin{aligned} 2 \text{ ABC}_{(16)} &= (2 \times 16^3 + 10 \times 16^2 + 8 \times 16^1 + 12 \times 16^0)_{(10)} \\ &= (8.192 + 2.560 + 128 + 12)_{(10)} = 10.892_{(10)} \end{aligned}$$

Conversions binaire-hexadécimal et hexadécimal-binaire

Même principe que pour le système "décimal codé binaire" :

$$2 \text{ ABC}_{(16)} = 0010/1010/1011/1100_{(2)}$$

1.5.- Indication de la position de la virgule (point décimal)

Principe

Virgule fixe et virgule flottante.

Mantisse et exposant (ou caractéristique).

Exemples

$$3,1416 = +,31416\text{E}+01 \text{ ou } +3,1416\text{E}+00$$

$$- 579 = -,579+03 \text{ ou } -5,79\text{E}+02$$

$$0,05 = +,5\text{E}-01 \text{ et } +5,5\text{E}-02$$

avec virgule ou point décimal, avec ou sans zéro devant la virgule ou le point décimal, avec ou sans suppression des signes +.

Autres exemples :

a) 227 805 000 000 : ce nombre peut s'écrire :

,227 805 E12 (exp 12), on remplace la virgule par point ou plus (. +), soit + 0,227 805 E+12.

b) 0,000 000 00 788 peut s'écrire 0,788 E-08

ou 0,788 E-08

2. Enregistrement des données

L'enregistrement des données est réalisé à l'aide de :

2.1.- Cartes perforées (80 colonnes)

Principe

Dimensions : 187 x 83 mm, avec coin(té) coupé(e) ou arrondi(e).

Disposition : 80 colonnes x 12 lignes;

2 lignes " hors-texte " (12 et 11) et 10 lignes numériques (0 à 9).

Codification : voir ci-dessous le code IBM, pour les chiffres décimaux (suivis de dix zéros) et les 26 lettres de l'alphabet; d'autres codes, de même que des signes spéciaux (signes de ponctuation, +, -, x, /, etc.,) peuvent également être utilisés.

Perforation et lecture (mécanique, électrique ou électronique).
Cartes spéciales : impressions diverses, graphitage, préperforation, etc.

Matériel comprend :

- a)- Perforatrice (poinçonneuse), avec ou sans impression, avec ou sans possibilité d'interprétation, avec ou sans programme (saut, reproduction, mise en position numérique ou alphabétique du clavier, etc.).
- b)- Vérificatrice, muni ou non d'un témoin lumineux, et d'en-coche.
- c)- Trieuse, généralement à 13 cases de réception, avec ou sans compteurs.

2.2.- Rubans perforés (bandes perforées)

Principe

Dimensions : 17 à 25 mm de largeur.

Disposition : 5 à 8 canaux, avec généralement 10 perforations par pouce (environ 2,5 mm par perforation).

Codification : voir ci-dessous le code PPTC8 (Paper Tape Transmission Code/8 channels : 1-2-4-8-Check-B-A), pour les chiffres décimaux (suivis de dix zéros) et les 26 lettres de l'alphabet (suivis du signe de fin d'enregistrement).

Perforation et lecture (mécanique, électrique ou électronique).

Matériel :

- a)- Perforatrice, avec ou sans impression.
- b)- Vérificatrice, avec ou sans possibilité de reproduction.
- c)- Appareils téléphoniques (télex) : 5 canaux.
- d)- Appareils enregistreurs divers : balances, saccharimètres, limnimètres, compas forestiers, tables de " digitalisation ", etc...

2.3.- Bandes magnétiques

Principe

Dimensions : 13 mm x 365 (cm) ou 730 m et " minicassettes".

Disposition : 7 à 9 canaux magnétisables, avec 200 à 1.600 bits par pouce (bpi).

Matériel :

- Encodeuse, à un ou plusieurs claviers.
- Appareils enregistreurs divers.

9768 221
THANK

07000 324
TRACE

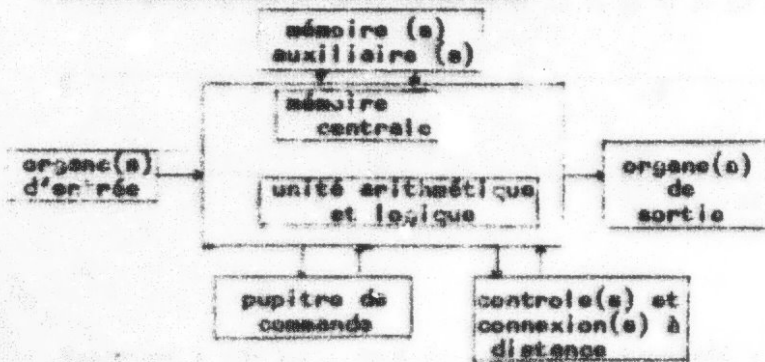
II.- Matériel (hardware)

3. Généralités

3.1.- Les caractéristiques générales des ordinateurs :

- Organes d'entrée, de sortie, de mémoire, etc...
- Fonctionnement électronique.
- Opérations arithmétiques et logiques.
- Programme enregistré en mémoire, avec possibilité pour l'ordinateur de modifier son propre programme.
- Variation discontinue (machines " digitales"), par opposition aux calculateurs analogiques (à variation continue)...

Structure générale des ordinateurs



4.- Organes de mémoire

4.1.- Caractéristiques générales

Organisation

- Chiffre décimal, caractère ou octet (6 ou 8 bits = 1,0 à 2,4, chiffres décimaux ou 1 caractère " alphabétique") ou " mot" de longueur fixe (16 bits, 24 bits, 36 bits, etc.) plus 1 ou 2 bits de contrôle par chiffre décimal, caractère, octet ou mot.
- Avantages et inconvénients : encombrement, vitesses et précision.
- Précision étendue et double précision.
- Adresses.
- Accès direct ou sélectif (" sélectif") et accès séquentiel.

Rapidité :

Temps d'accès et vitesse de transfert (ou débit),

Unités de temps : milliseconde (1 msec = 10^{-3} seconde),

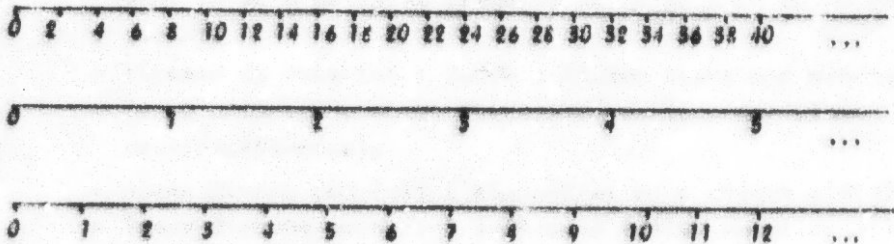
microseconde (1 usec = 10^{-6} seconde), nanoseconde (1 nsec = 10^{-9} seconde),

Capacité :

Unités : bit, chiffre décimal, caractère, octet, mot,

K (1.000 ou 1.024), M (1.000.000 ou 1.048.576).

Correspondance (sans perte de place):



Remarque : une page dactylographiée ou imprimée : environ 3000 caractères.

4.2.- Différents types (1)

Tores magnétiques (ferrites)

- Écriture et lecture par induction (4 ou 5 conducteurs électriques passant dans chaque tore).
- Accès direct.
- Temps d'accès (ou cycle de base = lecture + restauration) : 1 à 10 usec par chiffre décimal, caractère, octet ou mot.
- Capacité : 1.000 à 1.000.000 chiffres décimaux, caractères, octets ou mots.
- Mémoire centrale.

(1) Toutes les caractéristiques techniques sont des ordres de grandeur donnés à titre purement indicatif.

Semi-conducteurs :

- Circuits intégrés MOS (Metal Oxide Semi conducteur).
- Accès direct.
- Temps d'accès : = 1 usec par chiffre décimal, caractère, octet ou mot.
- Capacité : 10,000 à 1,000,000 chiffres décimaux, caractères, octets ou mots.
- Mémoire centrale.

Tambour magnétique

- Cylindre à surface extérieure magnétisable (généralement de 10 à 40 cm de diamètre et de 30 à 60 cm de longueur).
- Vitesse de rotation : 2,000 à 15,000 tours par minute.
- Têtes d'écriture et de lecture fixes (une par piste ou circonférence).
- Accès direct (en fait : séquentiel pour chaque piste).
- Temps d'accès moyen : 2 à 20 msec (délai rotationnel).
- Capacité : 10,000 à 5,000,000 chiffres décimaux, caractères, octets ou mots.
- Mémoire centrale et mémoire auxiliaire.

Disques magnétiques :

- Surfaces circulaires magnétisables à canaux ou pistes concentriques (généralement de 30 à 50 cm de diamètre).
- Vitesse de rotation : 1,000 à 5,000 tours par minute.
- Têtes d'écriture et de lecture mobiles ou fixes.
- Accès direct, avec adressage par disques, faces, secteurs et piste).
- Temps d'accès : 20 à 500 msec (déplacement éventuel de la tête de lecture et délai rotationnel).
- Vitesse de transfert : 50,000 à 500,000 chiffres décimaux, caractères, octets ou mots par seconde.
- Capacité : 500,000 à 50,000,000 chiffres décimaux, caractères, octets ou mots par disque ou groupe de disques ("disk pack").
- Mémoire auxiliaire amovible.

Bandes magnétiques :

- Vitesse d'avancement : 1 à 5 mètres par seconde.
- Têtes d'écriture et de lecture fixes (une par canal).
- Accès séquentiel : blocs numérotés en séquence et séparés par des " blancs ".
- Temps d'accès : Jusqu'à une ou deux minutes (démarrage et avancement de la bande).
- Vitesse de transfert : 10.000 à 500.000 chiffres décimaux, caractères, octets ou mots par seconde.
- Capacité : 3.000.000 à 60.000.000 chiffres décimaux, caractère octets ou mots par bande de 365 ou 730m.
- Mémoire auxiliaire amovible.

Divers :

- Feuilles magnétiques, films minces, bulles, etc...

5.- Organes d'entrée et de sortie

5.1.- Les fonctions sont les suivantes :

- Lecture et écriture.
- Mise en page.
- Codage et décodage (conversions).
- Mémoires-tampons et simultanéités (canaux).

5.2.- Entrée :

Lecteur de cartes perforées

- Lecture électrique ou électronique.
- Vitesse : 100 à 2.000 cartes par minute (= 150 à 2.500 caractères maximum par seconde).

Lecteur de rubans perforés :

- Lecture électrique ou électronique.
- Vitesse : 60 à 2.000 caractères par seconde.

Lecteur de documents :

- Lecture optique (de marques, de chiffres et/ou de lettres) et lecture magnétique.
- Caractères spéciaux : MICR par exemple (caractère Magnétique Codé à 7 bâtonnets, plus ou moins espacés, avec encre magnétisable).
- Vitesse : 500 à 2.000 caractères par seconde.

Unité de disques magnétiques :

- Voir ci-dessus (organes de mémoire).

Dérouleur de bandes magnétiques :

- Voir ci-dessus (enregistrement des données et organes de mémoire).

Pupitre de commande, console, terminal :

- Clavier, interrupteurs (clés ou " switches"), écran cathodique, etc...

Divers :

- Connexion directe avec des appareils enregistreurs, etc...

5.3.- Sortie

Perforateur de cartes :

- Vitesse : 60 à 500 cartes par minute (= 80 à 700 caractères maximum par seconde).

Perforateur de rubans :

- Vitesse : 15 à 300 caractères par seconde.

Ces deux perforateurs ne sont plus utilisés.

Machine à écrire :

- Caractère par caractère.
- Vitesse : 10 à 50 caractères par seconde.

Imprimante :

- Ligne par ligne (80 à 160 caractères par ligne).
- Vitesse : 100 à 2.000 lignes par minute (= 150 à 5.000 caractères maximum par seconde).

Traceur de courbes (" plotter ").

- Etablissement automatique de graphiques.

Unité de disques magnétiques :

- Voir ci-dessus (organes de mémoire).

Dérouleur de bandes magnétiques :

- Voir ci-dessus (organes de mémoire).

Pupitre de commande, console, terminal :

- Voyants lumineux, machine à écrire, écran cathodique (affichage ou visualisation " alphanumérique " et/ou graphique), etc...

Divers :

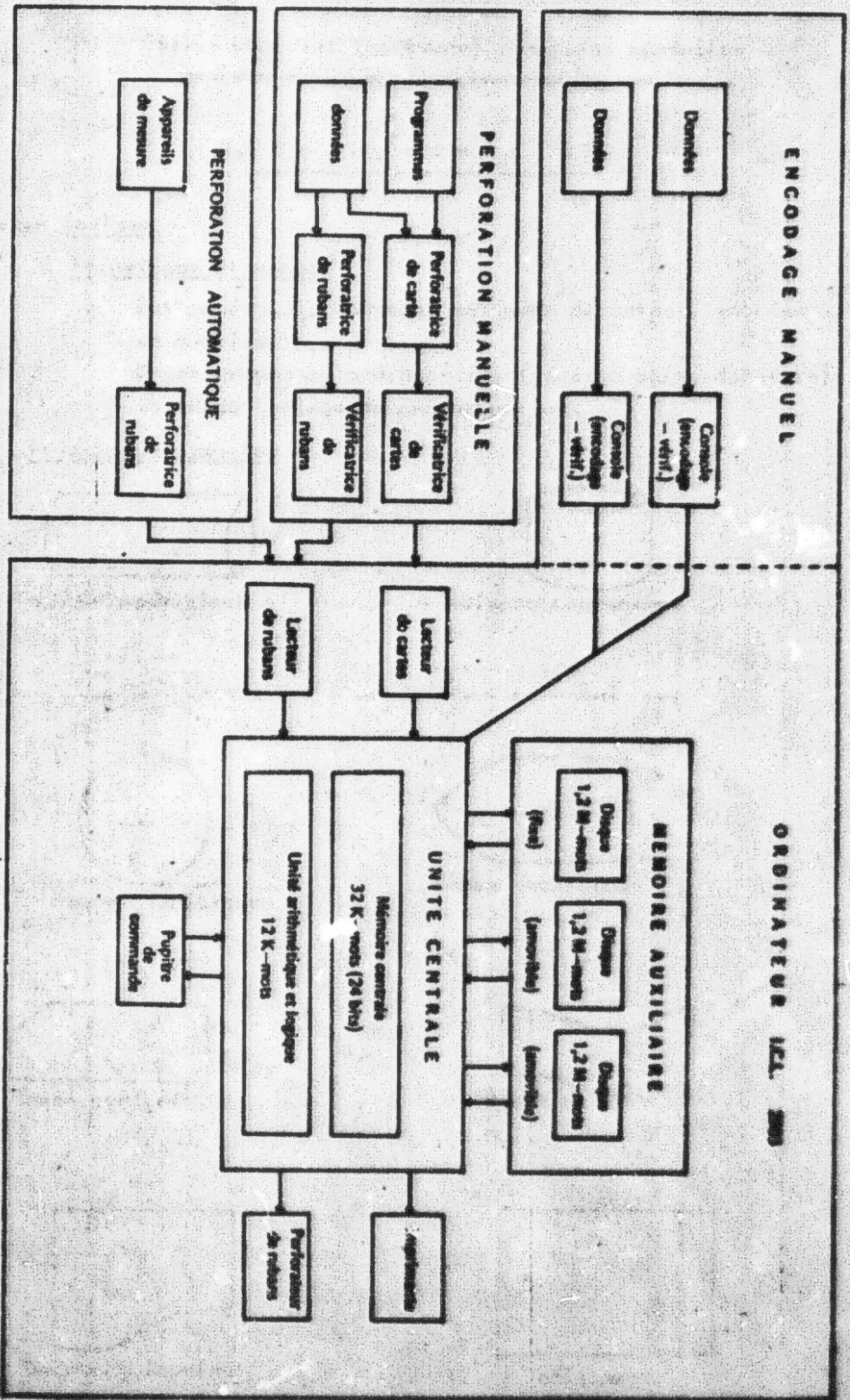
- Microfilm, etc.

5.4.- Quelques exemples ("petite " ordinateurs).

Ordinateur	IBM 650	IBM 1620	IBM 1130	ICL 2903
Cartes perforées :				
Lecture (cartes par minute).....	200	250 ou 500	300 à 600	300 ou 600
perforation (id).....	100	125 " 250	80 ou 160 carts/ann.	100
Rubans perforés :				
Lecture (carac. par sec. perforation (id).....	-	150 15	60 15	250 110
Impression :				
Machine à écr.(carac.p. seconde.....	-	10	15	
Imprimante (lignes.p.min..	150	430 ou 600	80 à 600	150 à 600
Traceur de courbes	-	x	x	x
Lecteur optique	-	-	x	x
Ecran cathodique	-	-	x	x
Connexion à distance	-	-	x	x

5.5.- Résumé des principales vitesses de transfert (ordres de grandeur):

- Mémoire magnétique : 1 à 10 usec par caractère.
- Cartes et rubans perforés (lecture) : 1,000 à 2,000 usec par caractère.
- Imprimante : 1,000 à 2,000 usec par caractère.
- Cartes et rubans perforés (perforation): 5,000 à 10,000 usec par caractère.



III.- Logiciel (software) : notions générales

6.- Programmation

6.1.- Analyse

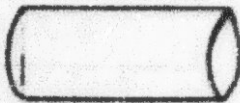
Principaux éléments

But poursuivi, données, méthodes de calcul, entrées et sorties,

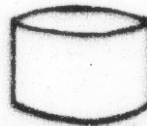
Organigramme ou ordinoگرامme (général et/ou détaillé),

Choix du langage de programmation,

Principaux symboles



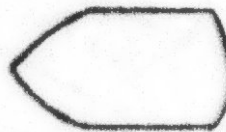
Tambour magnétique



Disque magnétique



Bande magnétique



Ecran cathodique



Carte perforée



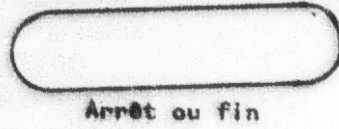
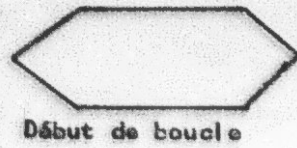
Ruban perforé



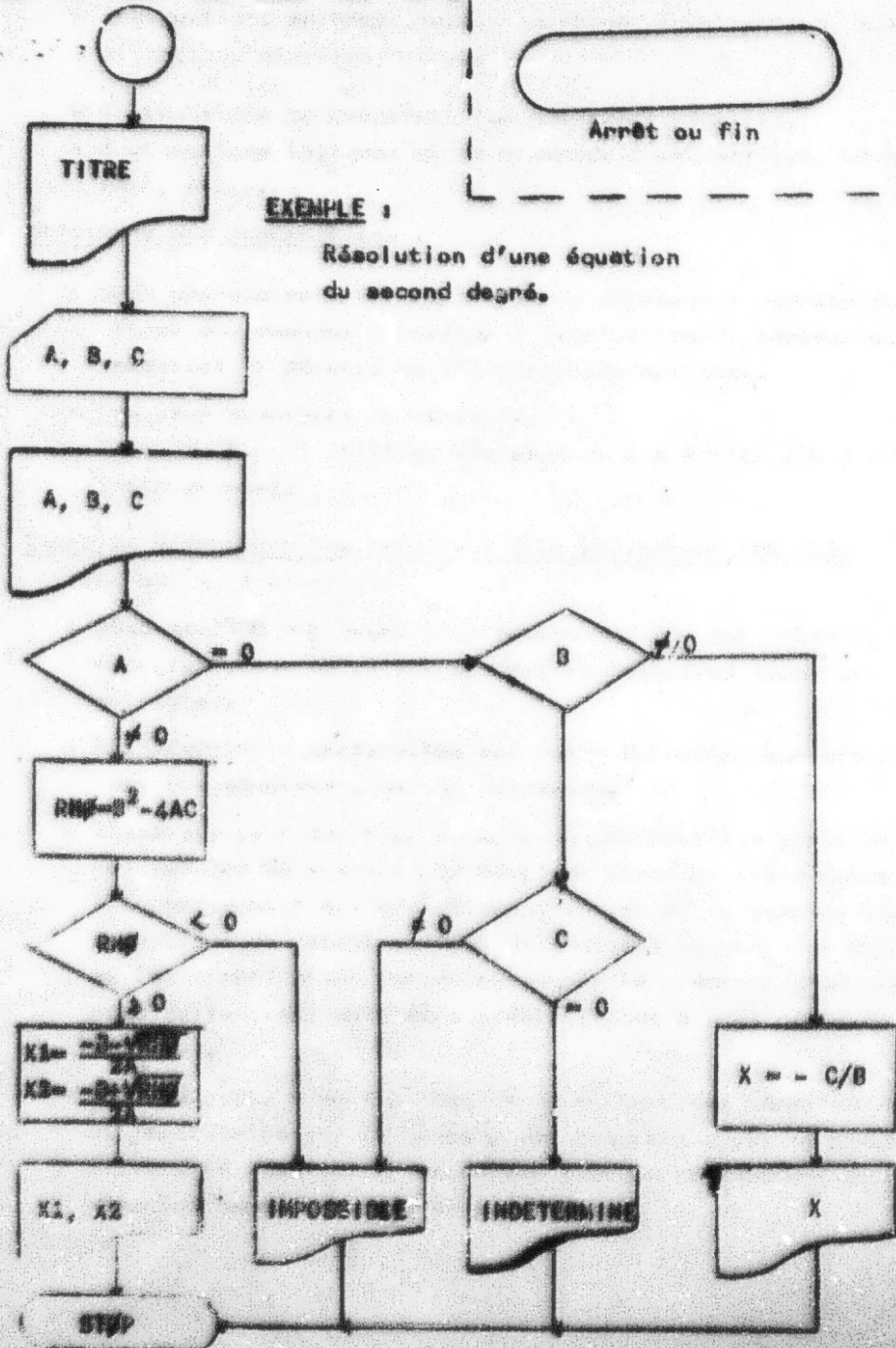
Document imprimé



Traitement

**EXEMPLE :**

Résolution d'une équation
du second degré.



6.2.- Programmation en " Langage machine "

Types d'instructions

- Instructions d'entrée et de sortie : lecture, impression, perforation, etc...
- Instructions arithmétiques : addition, soustraction, multiplication, division, etc...
- Instructions de transfert (de mémoire à mémoire).
- Instructions logiques et de commande : comparaison, branchement, etc...

Structure des instructions :

- Code opératoire + une ou plusieurs adresses : adresse de la ou des données à traiter + éventuellement adresse du résultat et adresse de l'instruction suivante.
- Longueur constante ou variable.
- Codes de 2 ou 3 chiffres décimaux ou 6 à 8 bits (50 à 200 instructions).

Exemples d'instructions relatives à un ordinateur IBM 1620 (décimal codé binaire).

- 36xxxxx00300 = lecture d'un nombre perforé sur ruban et enregistrement de ce nombre dans les mémoires xxxxx et suivantes.
- 30xxxxx00300 = perforation sur ruban du nombre enregistré dans les mémoires xxxxx et suivantes.
- 21xxxxxyyyyy = addition du contenu des mémoires yyyyy et précédentes au contenu des mémoires xxxxx et précédentes (contrairement aux opérations d'entrée et de sortie, les opérations de calcul se font de droite à gauche, les unes et les autres étant interrompues par la présence d'un repère particulier, qui doit être préalablement mis en place en mémoire).
- 23 xxxxxxxyyyyy = multiplication du contenu des mémoires xxxxx et précédentes par le contenu des mémoires yyyyy et précédentes, le résultat de la multiplication se trouve toujours dans les mémoires 00099 et précédentes.

- 26 xxxxyyyyyy = transfert du contenu des mémoires yyyyy et précédentes dans les mémoires xxxxx et précédentes (comme les opérations de calcul, les opérations de transfert se font de droite à gauche).
- 24 xxxxyyyyyy = comparaison du contenu des mémoires xxxxx et précédentes avec le contenu des mémoires yyyyy et précédentes (les opérations de comparaison se font également de droite à gauche).
- 47xxxxx01300 = passage à l'instruction qui est enregistrée dans les mémoires xxxxx et précédentes si le résultat de la dernière comparaison réalisée est négatif.
- 480000000000 = arrêt.

6.3.- Programmation symbolique :

Autocodes :

- Langages symboliques liés aux différents ordinateurs : autocode, langage d'assemblage, assembleur ("assembler").
- Codes mnémoniques et adresses symboliques.
- Exemples : SPS = Symbolic Programming System (EM 1620); assembleur Hewlett-Packard 2114; assembleur IBM 1130; assembleur ICL 2903.

Langages " universels "

- Langages symboliques liés aux différents types de problèmes, avec de légères variantes d'un "niveau" à l'autre et d'un ordinateur à l'autre (entrées et sorties notamment).
- Exemples : FORTRAN (FORMula TRANslation); ALGOL (ALGORITHMIC Language); APL (A Programming Language); BASIC (Beginner's ALL-purpose Symbolic Instruction Code); COBOL COmmon Business Oriented Language); PL/1 (Programming Language 1); GPSS (General Purpose Systems Simulator).
- Avantage et inconvénients : temps de programmation et temps d'exécution.

Traduction, assemblage ou compilation :

- Programme traducteur, programme d'assemblage, compilateur.
- Programme source et programme traduit ou assemblé ou compilé ou objet.
- Eventuellement précompilateur.
- Détection automatique des erreurs d'écriture.
- Introduction de sous-programmes ou sous-routines.

Exemples d'instructions assembler relative à un ordinateur HP2114

- CLA = remise à zéro (Clear) du contenu du registre A.
- CLB = remise à zéro (Clear) du contenu du registre B.
- ADA X = addition (ADD) du contenu de la mémoire appelée X au contenu du registre A (les mémoires correspondant aux différents noms, composés ici de 1 à 5 lettres ou chiffres, doivent également être préalablement définies par des instructions spéciales).
- ADB X = addition (ADD) du contenu de la mémoire appelée X au contenu du registre B.
- LDA X = transfert (Load) du contenu de la mémoire appelée X dans le registre A.
- LDB X = transfert (Load) du contenu de la mémoire appelée X dans le registre B.
- STA X = transfert (Store) du contenu du registre A dans la mémoire appelée X.
- STB X = transfert (Store) du contenu du registre B dans la mémoire appelée X.

Exemple de programme Fortran relatif à un ordinateur IBM 1620
résolution d'une équation du second degré (A non nul et RHP non
négatif).

```
- PRINT 1
1  FORMAT (49H/EQUATION DU SECOND DEGRE (A NON NIL, RHP NON
  NEG.)) ACCEPT 2, A, B, C.
2  FORMAT (F5.0/F5.0/F5.0)
  RHP=30RT(8xx2-4xxxC)
  X1=(-B-RHP)/(2xxA)
  X2=(-B+RHP)/(2xxA)
  PRINT 2,X1,X2
3  FORMAT (E11.4/E11.4)
  PAUSE
  END
```

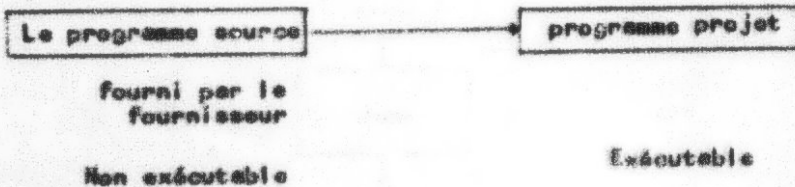

Exemple de programme Fortran relatif à un ordinateur HP2116 :
résolution d'une équation du second degré (A non nul et RND
non négatif).

```

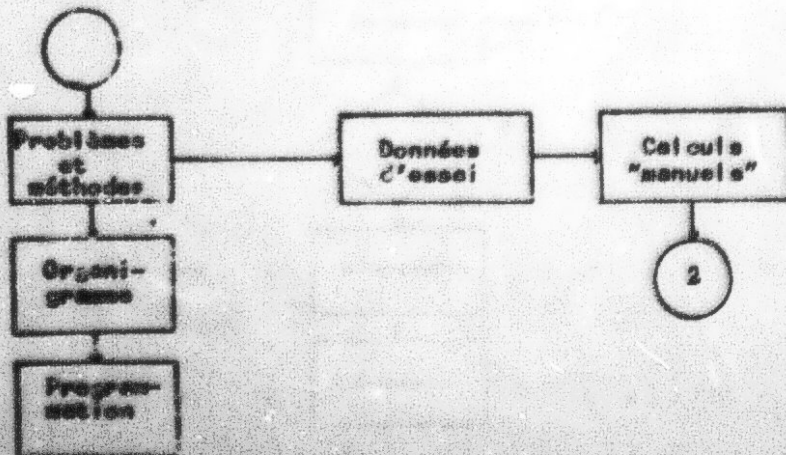
WRITE (2,1)
1  FORMAT ("EQUATION DU SECOND DEGRE (A NON NUL ET RND NON
      NEGATIF)")
  READ (2,2)A,B,C
2  FORMAT (F5.0/F5.0/F5.0)
  RND=0.07T (Bxx2-4.,xAxC)
  X1=(-B-RND)/(2.,xA)
  X2=(-B+RND)/(2.,xA)
  WRITE(2,3)X1,X2
3  FORMAT (E11.4/E11.4)
  PAUSE
  END
  
```

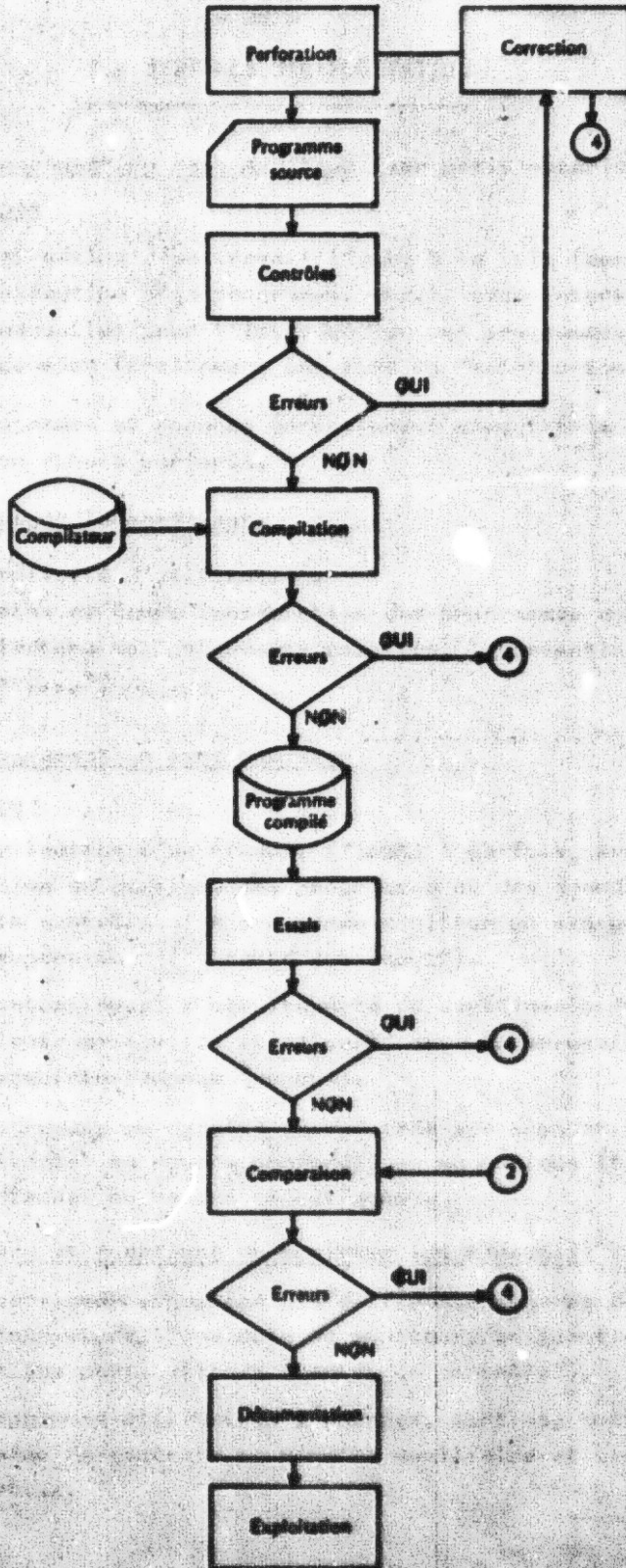
Conclusion :

Un programme écrit dans un Autocode ou dans l' "Universel " n'est pas admis par un ordinateur. Il faut transformer le programme vers le code de la machine : d'où alors recours à des programmes de traduction, assemblage et compilateurs est nécessaire.



6.4.- Schéma récapitulatif de la programmation symbolique (avec
programme source sur cartes perforées, compilateur et pro-
gramme compilé sur disque).





7.- SYSTEMES D'EXPLOITATION

7.1.- Monoprocuration sans moniteur (sur petit matériel)

Principe

Réalisation d'un travail ("job") à la fois (compilation, exécution d'un programme, etc.), avec interventions manuelles pour l'introduction des programmes et des données (traitement par lots ou "batch processing").

Programmes et données généralement enregistrés sur cartes ou rubans perforés.

Avantages et inconvénients

Simplicité d'utilisation.

Pertes de temps (perforation des programmes compilés, lecture des programmes compilés, interventions manuelles, etc...).

7.2.- Monoprocuration avec moniteur

Principe:

Réalisation d'un travail ("job") à la fois, avec intervention automatique des programmes et des données, sous le contrôle d'un programme moniteur ou exécutif ou superviseur ("stacked job input").

Eventuellement : simultanéité de certaines opérations (entrées-sorties et calcul), avec interruptions et mémoires-tampons (casseaux).

Programmes et données enregistrés sur support papier (cartes ou rubans perforés) ou magnétique (tambours, disques ou bandes magnétiques).

Structure et fonctions du moniteur (ou exécutif)

Superviseur : gestion des différents organes de l'ordinateur et de l'ensemble du moniteur, en fonction d'instructions particulières (cartes de contrôle).

Programmes utilitaires : entrées, sorties; transferts de mémoire centrale en mémoire auxiliaire et vice-versa, etc...

Compilateurs : Fortran, Cobol, etc.

Partiellement en mémoire centrale ("résident" ou "squelette")
et partiellement en mémoire auxiliaire.

Avantages et inconvénients

Complexité du logiciel et, dans une certaine mesure, de
l'utilisation.

Encombrement en mémoire centrale.

Absence de pertes de temps générales, mais pertes de temps
pour certains organes.

7.3.- Multiprogrammation et multitraitement

Principes

Réalisation "simultanée" de plusieurs travaux dans plusieurs
"partitions" de la mémoire centrale.

Différentes modalités :

- un travail principal quelconque et un ou plusieurs travaux
secondaires de types ou de dimensions données;
- plusieurs travaux quelconques exécutés parallèlement (avec
interruptions liées notamment à l'occupation des organes
périphériques);
- traitement en temps réel (interruptions d'origine extérieu-
re);
- exécution vraiment simultanée de plusieurs travaux (multi-
traitement ou "multiprocessing").

Moniteur ou exécutif chargé de régler à la fois les problèmes
de gestion des différents organes de l'ordinateur et les
problèmes de priorité des différents travaux.

Avantages et inconvénients

Complexité du logiciel.

Meilleure utilisation de la mémoire centrale et des
organes périphériques.

7.4.- Traitement à distance et traitement en temps partagé

Principes

Traitement à distance (télétraitement ou "teleprocessing") : commande de l'ordinateur ou introduction des données à distance, par l'intermédiaire de lignes téléphoniques ou télégraphiques ordinaires ou spéciales (50 à 10,000 bauds ou bits par seconde).

Traitement à distance différé ("remote job entry" ou "remote batch processing").

Traitement en temps partagé ("time-sharing") : réalisation "simultanée" de plusieurs travaux, commandés à partir de plusieurs pupitres ("terminal"), par tranches successives de très courtes durées (quelques secondes maximum).

Terminal avec différents organes d'entrée et de sortie, "modem" (modulateur-démodulateur) et unité de contrôle des transmissions.

Avantages et inconvénients

Complexité du logiciel.

Simplicité et rapidité d'utilisation.

Coût relativement élevé (en fonction du type de problème).

- LE STATISTICIEN ET L'ORDINATEUR

INTRODUCTION :

La première utilisation de l'ordinateur remonte vers le début des années cinquante vraisemblablement dans un département de statistique d'une institution de recherche en Grande Bretagne (citée par Dagnillie⁽¹⁾), l'emploi ne cesse d'augmenter et ceci en relation avec l'extension des possibilités offertes au statisticien par le développement de l'informatique.

Les moyens modernes de calcul et leur emploi

Les différents matériels :

Des simples machines traditionnelles à calculer mécaniques et électromécaniques (à " quatre opérations ") ont été progressivement remplacées par des machines électroniques, qui forment un ensemble à la fois très déterrifié et très mouvant. Ces dernières comprennent une gamme allant des machines portatives et de bureau, ayant un petit nombre de registres de mémoires et, parfois, un dispositif extérieur d'enregistrement ou de lectures de programmes, jusqu'aux ordinateurs proprement dit, caractérisés par une plus grande capacité de mémoires (plusieurs milliers à plusieurs millions, voire plusieurs dizaines ou centaines de millions de chiffres décimaux), par l'enregistrement des programmes en mémoire sous la même forme que les données et les résultats, et par une plus grande diversité et une plus grande rapidité des fonctionnements des organes périphériques (lecture de données et impression des résultats à des vitesses de l'ordre de 1000 caractères par seconde par exemple).

L'accès à l'ordinateur :

Il se fait soit directement, par traitement à distance et en temps partagé ("tele processing" et " time sharing "), soit indirectement, par un traitement plus classique par lots ("batch processing") soit encore par traitement à distance non immédiat (remote job entry).

Dans le premier cas l'utilisation de l'outil statistique peut obtenir l'exécution quasi-immédiate de certains calculs, sans quitter son lieu de travail habituel s'il y possède un terminal. Dans le deuxième cas, par contre toutes les informations relatives aux calculs à exécuter doivent être transmises au centre de calcul et introduites en machine en suivant la filière habituelle (porteur ou

(1) tiré des documents de Mr. Dagnillie, Centre de calcul, Genève.

courrier, opérateur, etc.,...). Les délais d'exécution peuvent être assez longs.

L'emploi de l'ordinateur :

Dans cette rubrique, nous citons le modèle de comparaison établi par Mr. Bagnelie et auquel il a mis une série d'hypothèses. Il consiste à déterminer le seuil à partir duquel l'emploi de l'ordinateur est opportun ou indispensable.

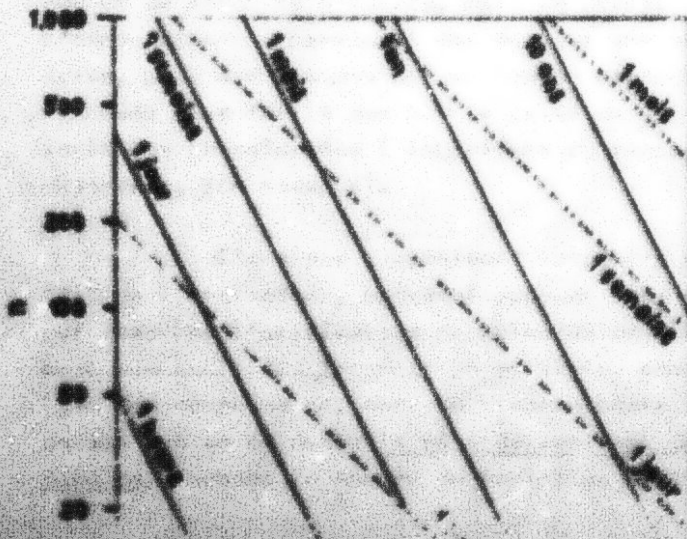
Si le volume de données à analyser et les calculs à effectuer est peu important, le statisticien a intérêt à utiliser la machine portative ou de bureau, programmable ou non programmable, soit l'ordinateur utilisé en temps partagé.

Si le volume de données ou des calculs est important, la décision dépendra des facteurs tels que la durée du travail de perforation ou d'encodage, la durée présumée des calculs, l'urgence de travail. La figure suivante permet de préciser quelque peu les choses, en comparant d'une façon générale des temps de perforation et des temps de calcul dans le cas d'une exécution réalisée à l'aide des machines non programmables.

Les hypothèses émise, étaient les suivantes :

- Les données à analyser sont relatives à variables et individuelles, p allant de 1 à 100 et n de 10 à 1000, ce qui semble recouvrir la majorité des problèmes qui se posent dans le domaine de la statistique " de recherche ".

- Le temps de calcul est proportionnel à np^2 et le temps de perforation est proportionnel à np , ce qui correspond notamment aux calculs classiques de matrices de variances et covariances ou de corrélation.



- Il a été admis ici qu'un jour de travail comporte 8 heures, 1 semaine comporte 5 jours, un mois comporte 20 jours et un an comporte 220 jours de travail. Il est évident que ces hypothèses sont discutables et chacun pourra établir un graphique plus conforme à ses propres conditions de travail. L'essentiel, pour effectuer un choix valable, est de disposer, en fonction du volume de données (p et n), d'estimation sommaires des temps de calcul "manuel et de perforation ou d'encodage", qui puissent être comparés entre eux et avec le degré d'urgence du problème, sans négliger pour autant les temps morts éventuels qui sont liés aux différentes opérations (perforation ou encodage, calculs, etc...).

QUELQUES CONSEQUENCES DE L'EMPLOI DE L'ORDINATEUR :

L'emploi de l'ordinateur est imposé chaquefois qu'on désire effectuer plus rapidement ou plus facilement les calculs qui étaient réalisés antérieurement à l'aide des machines de bureau ou utiliser des méthodes parfois anciennes, mais peu employées en raison de l'abondance des calculs qu'elles nécessitent; signalons que l'ordinateur a provoqué la naissance de méthode et de procédures nouvelles.

L'analyse de la variance, l'analyse de la covariance et les méthodes relatives à la régression et à la corrélation sont incontestablement les outils statistiques les plus classiques, auxquels l'emploi de l'ordinateur a donné, dès le début, une nouvelle dimension.

Le traitement automatique de l'information a tout d'abord permis de supprimer les fastidieux calculs manuels de sommes de carrés et de produits d'écartés. Mais en outre, du fait notamment de l'enregistrement permanent des données sur cartes ou sur rubans perforés, puis sur disques ou sur bandes magnétiques, il s'est avéré beaucoup plus facile que par le passé de calculer et d'analyser des variables transformées (logarithmes ou racines carrées des volumes observés, par exemple).

L'analyse à plusieurs variables a également connu, du fait de l'ordinateur, un développement sans précédents, tant en ce qui concerne l'utilisation de méthodes déjà anciennes, mais relativement peu employées (régression multiple, analyse factorielle et analyse des composantes, analyse discriminante, etc...). Le tableau ci-dessous met en évidence le rôle déterminant de l'ordinateur dans ce domaine, lorsque le nombre de variables atteint ou dépasse la dizaine.

La naissance de l'ordinateur à tout d'abord permis la publication de nombreuses tables numériques nouvelles; plus développées que par le passé, mais la tendance actuelle est la recherche d'algorithmes ou de méthodes de calcul qui permettent de déterminer facilement les valeurs théoriques nouvelles. Signalons que classiquement, pour une série de nombreuses observations :

$$x_1, x_2, x_3, \dots, x_1, \dots, x_n$$

Les formules considérées comme les meilleures sont :

$$\bar{x} = \left(\sum_{i=1}^n x_i \right) / n \quad \text{SCE} = \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 / n \quad \text{et}$$

$$s^2 = \text{SCE} / n$$

Par ordinateur, la formule de définition de la moyenne peut-être utilisée sans inconvénient, mais l'emploi de la formule de détermination de la somme des carrés des écarts peut s'avérer extrêmement dangereux, surtout si le coefficient de variation est élevé et si le nombre des chiffres significatifs utilisés est relativement réduit.

Une première solution peut-être de revenir aux formules de définition de la variance :

$$\text{SCE} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{et} \quad s^2 = \text{SCE} / n$$

Cette solution exige cependant de prendre en considération à deux reprises toutes les valeurs observées, tout d'abord pour calculer la moyenne et ensuite seulement pour calculer la somme des carrés et la variance.

D'autres solutions ont été proposées et l'une des meilleures serait basée sur les formules de récurrence suivantes :

$$S_i = S_{i-1} + x_i \quad \text{et}$$

$$\text{SCE}_i = \text{SCE}_{i-1} + (x_i - S_{i-1})^2 / [i(i-1)]$$

avec comme valeurs initiales :

$$S_1 = x_1 \quad \text{et} \quad \text{SCE}_1 = 0$$

la moyenne et la variance sont alors :

$$\bar{x} = S_n / n \quad \text{et} \quad s^2 = \text{SCE}_n / n$$

Ces formules peuvent être étendues au calcul des sommes des produits d'écart, et, aussi, des sommes des carrés et des produits des écarts factorielles et résiduelles qui interviennent en analyse de la variance et de la covariance [Youngs et Cramer, 1971, cités par Dagnille].

En ce qui concerne la documentation en matière de programmes, une première information dont dispose l'utilisateur d'un ordinateur est également constituée par la documentation du centre de calcul auquel il s'adresse.

Cette documentation peut contenir des ensembles de programmes ou de sous-programmes fournis par les constructeurs ou les services de calcul, tels que les ensembles S, S.P. et STAT - Pack des firmes I B M et UNIVAC etc.,.

D'autres possibilités d'avoir des programmes :

- des revues possédant des rubriques spéciales : Ex, Applied Statistics et The Computer Journal,

- récemment certains livres présentent simultanément des notions de Statistiques et d'Inférentiques,

- Il existe des systèmes intégrés de programmation, qui utilisent des langages spécialement adaptés à l'emploi des méthodes statistiques ("Statistical Systems") : Ex, les systèmes SSCOP Genstat, OMNITAB, P-STAT, S, P, SS, etc.,.

- BANQUE ET TRAITEMENT DE DONNEES -

Comme nous l'avons signalé au début que le problème de traitement de données (nombre considérable) dans certains domaines, peut être résolu par l'utilisation de l'informatique. En effet l'ordinateur présente des caractéristiques très intéressantes (capacité, vitesse, accès, mise à jour) et rend un grand service dans ce contexte.

À la Faculté de Sciences Agronomiques de Gembloux deux Banques de données ont été lancées : Banque de données de sols, " " forestières, une troisième est en voie de préparation c'est la banque de données de gènes.

La réalisation d'une banque de donnée suppose l'existence d'un matériel de traitement disponible (ordinateur) et relativement puissant.

Banque de données de sols :

Une banque doit avoir une structure, à la fois, complète et suffisamment souple vis-à-vis de l'Utilisation.

Saisie des données :

Cette question se présente sous deux aspects :

- Le vocabulaire,
- Le format.

- **Vocabulaire** : Le vocabulaire utilisé est à quelque chose près celui du R.I.T.D.S. ⁽¹⁾.

L'information rassemblée porte sur les chapitres suivants :

- Identification - localisation du profil
- Description de l'environnement du profil
- Description des horizons du profil
- Description synthétique du profil
- Analyse de laboratoire.

- **Format** : En ce qui concerne la description des horizons (à Gembloux), il a été adopté un format fixe, c'est à dire qu'un code est défini à la fois par sa valeur et par la place qu'il occupe sur la carte perforée.

(1) Réseau International de traitement de données de sols

Avant son enregistrement, l'information est transcrite sur un document de base, reproduisant le dessin de la carte perforée. L'avantage réside dans le contrôle de l'information avant son enregistrement, de plus le travail de perforation (ou d'encodage) s'en trouve facilité. A Gembloux, le saisi des données forestières se fait directement à partir de ce document.

- Enregistrement de données

Dans le cas de la banque de données pédologiques de Gembloux, le fichier de base est établi, au moins provisoirement sur cartes perforées ; celles-ci se répartissent comme suit :

- deux cartes pour Identification - localisation,
- deux cartes pour Environnement (plus 2 cartes éventuellement pour des remarques ou notes spéciales).
- trois cartes sur horizon (plus deux cartes éventuelles pour profil, pour des remarques et notes spéciales).
- une carte pour description synthétique,
- trois cartes pour analyse par horizon (plus deux cartes par profil pour des remarques et notes éventuelles ; techniques spéciales par exemple).

Ainsi, l'information complète, pour un profil comportant, par exemple sept horizons, peut-être rassemblée sur une cinquantaine de cartes perforées.

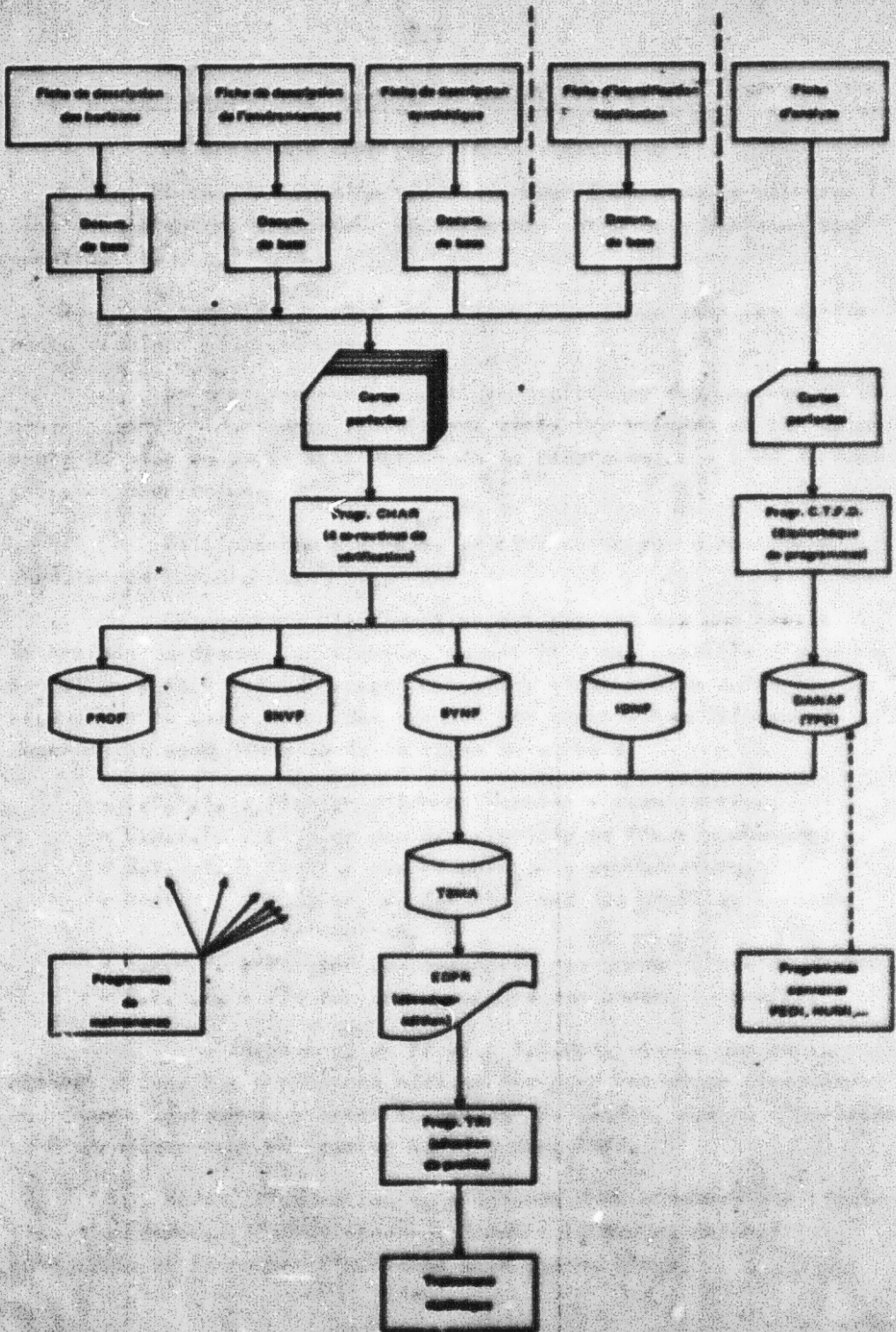


Fig. 1. Organigramme global de la
service de dessin technique.

1.- Enregistrement des données de description :

Le programme CHAR remplit deux missions :

a)- recherche de certains types d'erreur (par exemple valeurs Coûts inexistantes, profondeur décroissante entre deux horizons successifs...),

b)- chargement de données sur disque magnétique dans les différents fichiers pédologiques.

En cas d'erreur, dans l'une quelconque des parties de la description, l'imprimante fournit une liste des erreurs et les documents de base en cause sont retournés au Pédologue, aux fins de contrôle et correction.

En l'absence d'erreur, l'information est directement enregistrée sur disque.

Chaque type d'information est disposé sur une partie déterminée du disque, un fichier, auquel il n'est possible d'accéder que si on y fait spécifiquement allusion, c'est-à-dire qu'on ne risque pas la destruction des données par erreur. Les fichiers de description sont dénommés de la façon suivante :

- I.D.N.F. : Fichier d'identification - localisation,
- E.N.V.F. : Fichier des descriptions de l'environnement,
- S.Y.W.F. : Fichier des descriptions synthétiques,
- Prof. : Fichier des descriptions des profils, horizon par horizon,
- E.R.M.F. : Fichier des remarques concernant l'environnement,
- P.R.M.F. : Fichier des remarques concernant le profil.

Corollairement, un fichier T.B.M.A., table des matières, rassemble tous les paramètres nécessaires pour retrouver directement une donnée quelconque : entre autres, n° du profil, nombre d'horizons, n° d'enregistrement du premier horizon dans PROF.

Enfin, l'exécution du programme CHAR s'achève par l'édition d'un message d'enregistrement donnant le nombre de profils enregistrés et le nombre d'horizons pour chacun d'eux.

2.- Enregistrement des données analytiques :

L'enregistrement, par le programme C.T.P.D., des données de l'analyse en laboratoire se fait sous forme d'un fichier T.P.D. (titres, paramètres, données). Il s'agit d'une structure mise au point au service de Statistique - Informatique de la Faculté de Gembloux et qui permet l'application directe sur les données, d'un grand nombre de programmes statistiques et mathématiques.

3.- Edition en Clair :

Le programme E.D.P.R. réalise le décodage et l'édition en langage naturel des données de description (identification - localisation, environnement, horizons de description synthétique) de tout profil choisi dans la Banque. La traduction passe par l'intermédiaire d'une série de "Fichiers dictionnaires" comportant à la suite l'une de l'autre, les diverses significations des codes. On détruit ainsi un listing prêt à être reproduit après un éventuel passage par un système de réduction graphique.

4.- Traitement de données :

4.1.- Programmes de calculs pédologiques :

A partir des données brutes fournies par le laboratoire une série de programmes ou sous-programmes exécute divers calculs dont les résultats sont directement enregistrés dans le fichier D.A.N.A.F. (de type T.P.D.).

- Exemple :-programme relatif à l'analyse granulométrique
- programme relatif au calcul de C/N, degré de saturation, indice d'le .

4.2. / - Programme de Sélection :

Un programme est en cours d'élaboration permet de rechercher dans les fichiers les horizons ou les profils répondant à une série de critères présentés sous forme d'un " arbre logique ".

Il est possible de retrouver en un temps très bref, par exemple :

- tous les profils situés dans un périmètre défini .
- pente de moins de 5 %, exposé au nord.
- à taux de calcaire total $< 20 \%$

Dans des cas pareils, la programmation est très simple, grâce à l'utilisation des opérateurs logiques.

- EXEMPLE DE PROGRAMMES DE CALCULS PEDOLOGIQUES
RELATIF A L'ANALYSE GRANULOMETRIQUE -

Nous exposons dans ce qui suit l'établissement d'un ordi-nogramme (établi à partir du triangle textural C C S M B), la préparation des données pour le programme, données de laboratoire sur document spécial de travail, listing général du programme et enfin résultat de calcul de données de laboratoire.

- Préparation des données pour le programme de calcul :

Un programme de calcul des résultats de l'analyse granulométrique, PED1, a été préparé par L. BARIDEAU et modifié par H. KINDERMANNS.

Ce programme fournit :

- les teneurs individuelles en chacune des 9 fractions habituelles de l'analyse granulométrique,
- les teneurs cumulatives,
- les trois fractions principales, argiles (0-2), limon (2-50) et sable (> 50),
- la dénomination texturale, au départ des diagrammes triangulaires du Centre de Cartographie des Sols de la Belgique Méridionale (CCSM) et des experts pédologues de la F.A.O. (FAO),
- l'indice d'homogénéité (sur 3 fractions), selon HONNAY,
- la position des points de la courbe granulométrique cumulative, dans le système d'axes : - log diamètre,

L'exécution du programme PED1 exige une présentation normalisée des données de laboratoire. C'est pourquoi, il est vivement recommandé, afin de faciliter le travail d'encodage des données, d'utiliser le document spécial de travail, dont un exemplaire figure en annexe, avec des exemples de notation.

Ce document comporte, successivement :

- Quatre lignes de " titres", chacune de 48 caractères au maximum (1 caractère ou espacement par case). On y mentionne toute information utile à l'identification de la série analytique, certaines précisions techniques, etc...
- Une ligne de " paramètres", provisoirement non utilisée,
- Vingt lignes de " données", sous une série d'en-têtes explicatives,

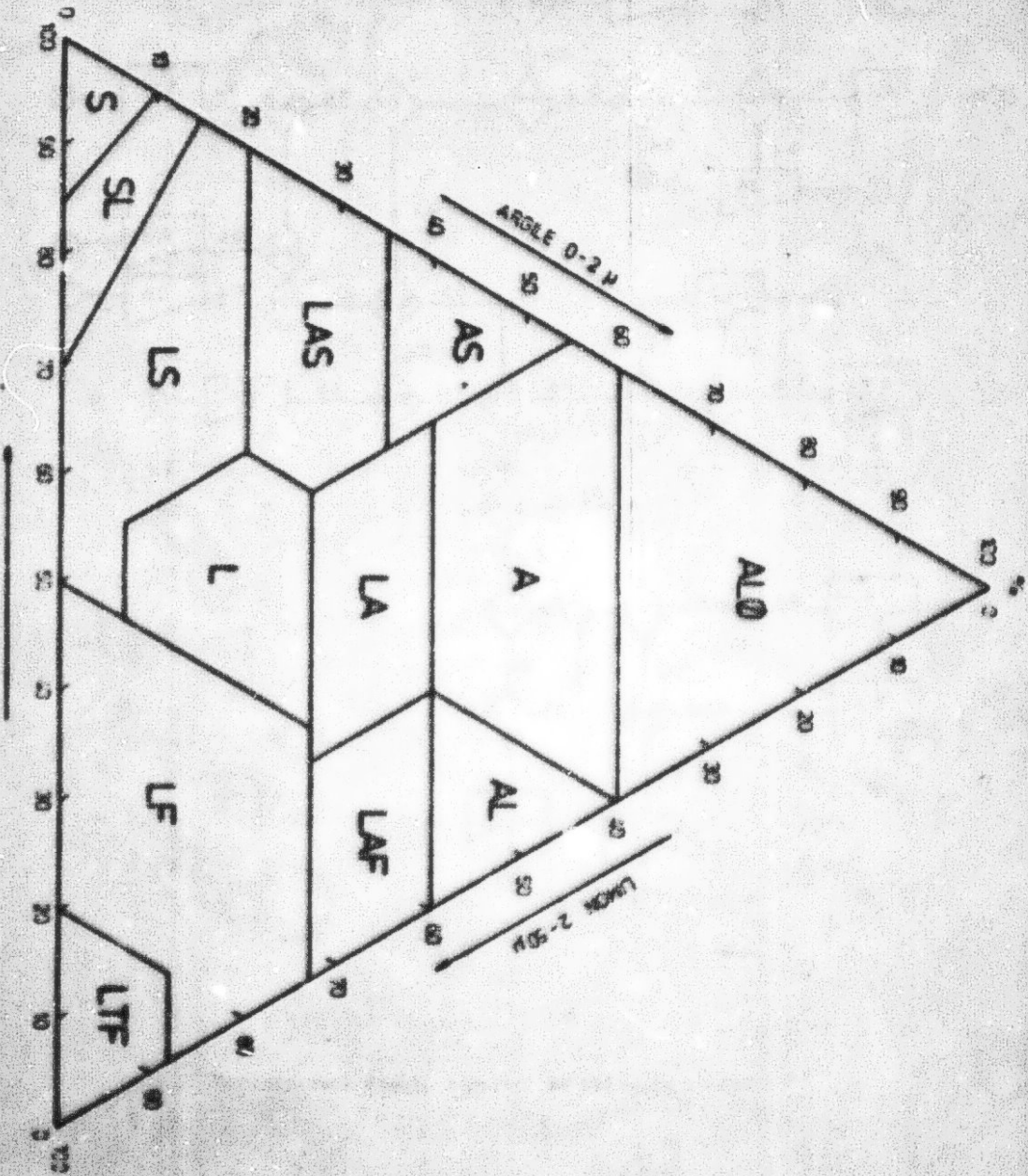
Dans la colonne "échantillon", le pédologue apporte, en format libre, toute indication utile à l'identification de l'échantillon. La colonne "n°" portera un n° d'ordre, soit un n° général d'analyse, un n° dans la série en cause, etc.,. Suivent les données proprement dites. Sont présentées avec une seule décimale : les teneurs en eau, carbonates et mat. organique, ainsi que les graduations lues sur la réglette de l'hydromètre.

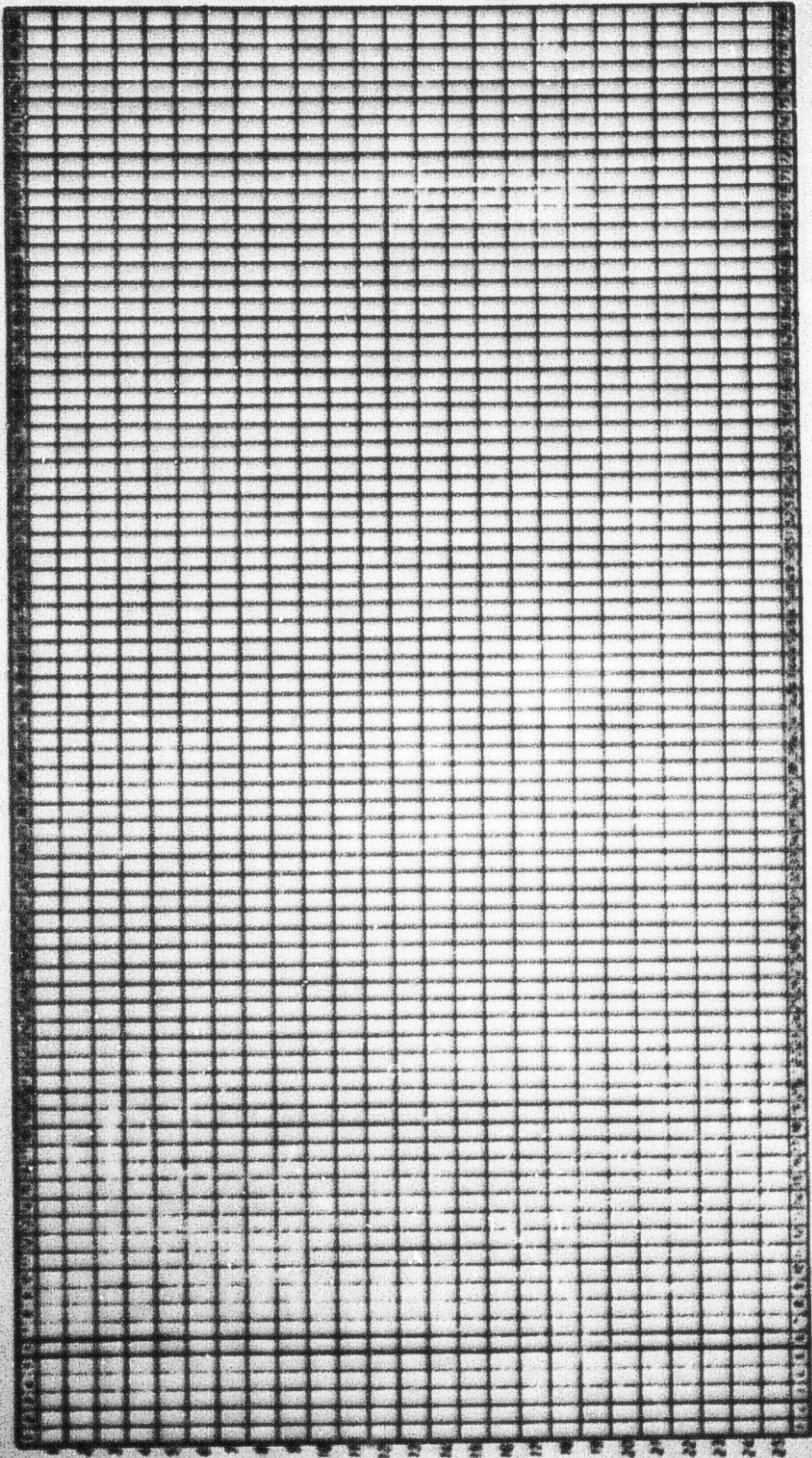
Sont présentées avec deux décimales : les poids des fractions sableuses.

Si la série d'analyse comporte plus de 20 échantillons, il n'est pas nécessaire de répéter les titres sur la (les) feuille(s) complémentaire(s). Toutefois, celles-ci seront numérotées, par exemple, dans le coin supérieur droit.

La dernière donnée sera le nombre 30,000, dans la colonne "N°" (cf exemple de notation).

On ne transcrit sur cartes perforées que les indications figurant à l'intérieur des cadres gras. Pour la perforation, les données numériques sont considérées comme des nombres entiers, de 5 chiffres au maximum.






```

58 C
59 C IMPRESSION DES FRACTIONS
60 C
61 WHITE(MLP,35)NO,ARGILE,X790,X7020,X2050,P
62 35 FORMAT(1H,10,98,980,1)
63 C
64 C DENOMINATION TEXTURELE = DIAGRAMME FAG
65 C
66 IF(ARGILE,GE,60)GO TO 53
67 IF(ARGILE,GE,40,AND,ARGILE,LT,60,AND,SABLE,LT,45,AND,XLIMON,LT,40)
68 56 TO 55
69 IF(ARGILE,GE,40,AND,ARGILE,LT,60,AND,XLIMON,GE,40,AND,XLIMON,LT,
70 56)GO TO 57
71 IF(ARGILE,GE,35,AND,ARGILE,LT,55,AND,SABLE,GE,45,AND,SABLE,LT,45)
72 50 TO 59
73 IF(ARGILE,GE,27,AND,ARGILE,LT,40,AND,SABLE,LT,20)GO TO 73
74 IF(ARGILE,GE,27,AND,ARGILE,LT,40,AND,SABLE,GE,20,AND,SABLE,LT,45)
75 50 TO 43
76 IF(ARGILE,GE,20,AND,ARGILE,LT,35,AND,XLIMON,LT,28,AND,SABLE,GE,45)
77 50 TO 43
78 IF(ARGILE,GE,7,AND,ARGILE,LT,27,AND,XLIMON,GE,28,AND,XLIMON,LT,50,
79 50,SABLE,LT,50)GO TO 67
80 IF(ARGILE,LT,27,AND,XLIMON,GE,50)GO TO 60
81 60 TO 49
82 40 IF(ARGILE,LT,12,AND,XLIMON,GE,50)GO TO 69
83 WRITE(MLP,41)
84 41 FORMAT(1H,758,'LIMON FIN')
85 60 TO 25
86 45 IF(ARGILE+70-SABLE)46,51,51
87 46 IF(ARGILE+370-2*SABLE)47,49,49
88 47 WRITE(MLP,42)
89 45 FORMAT(1H,758,'SABLE')
90 60 TO 25
91 49 WRITE(MLP,51)
92 50 FORMAT(1H,758,'SABLE LIMONFAY')
93 60 TO 25
94 51 WRITE(MLP,52)
95 52 FORMAT(1H,758,'LIMON SABLEFAY')
96 60 TO 25
97 53 WRITE(MLP,54)
98 54 FORMAT(1H,758,'ARGILE LOURDE')
99 60 TO 25
100 55 WRITE(MLP,55)
101 56 FORMAT(1H,758,'ARGILE')
102 60 TO 25
103 57 WRITE(MLP,56)
104 58 FORMAT(1H,758,'ARGILE LIMONFAY')
105 60 TO 25
106 59 WRITE(MLP,60)
107 60 FORMAT(1H,758,'ARGILE SABLEFAY')
108 60 TO 25
109 61 WRITE(MLP,62)
110 62 FORMAT(1H,758,'LIMON ARGILEFAY FIN')
111 60 TO 25
112 63 WRITE(MLP,64)
113 64 FORMAT(1H,758,'LIMON ARGILEFAY')
114 60 TO 25
115 65 WRITE(MLP,65)
116 66 FORMAT(1H,758,'LIMON ARGILE-SABLEFAY')

```

LIST

SEGMENTS(GRAN)
COMPRESS INTEGER AND LOGICAL
COMPACT
INPUT 2 = CR0
OUTPUT 5 = LRG/132
END

MASTER GRAN
DIMENSION ITIT(12),PARAM(10),P(5)
DATA IAST/4H** /
CALL SETDT8
CALL SWON(9)
MCD=2
MLH=5

C LECTURE DES TITRES - DES PARAMETRES - TEST DE FIN D'EXECUTION

5 WRITE(MWR,6)
6 FORMAT(1H1)
DO 10 I=1,4
23 READ(MCD,7)ITIT
7 FORMAT(12A4)
IF(ITIT(1)-IAST)H,500,H
8 WRITE(MWR,9)ITIT
9 FORMAT(1H,12A4)
10 CONTINUE
24 READ(MCD,15)PARAM
15 FORMAT(10F5,0)

C LECTURE D UN ECHANTILLON ET TEST DE FIN DU FORMULAIRE

WRITE(MWR,20)
20 FORMAT(6XNUMBER LIM INF 0 2 10 20 50 100 2
SOL 500 100)
WRITE(MWR,21)
21 FORMAT(1H,12X,'SUP',5X,'2',4X,'10',4X,'20',4X,'50',3X,'100',5X,'2
500',5X,'500',2X,'1000',2X,'2000')
25 READ(MCD,26)NO,H2O,CAC03,XMO,P02,C02,P010,C010,P020,C020,P
26 FORMAT(15,3F5,1,3(F5,0,F5,1),5F5,2)
IF(NO = 5000)27,5,27

C FACTEURS DE TRANSF - CALCUL DES FRACTIONS ARGILE LIMON SABLE

27 F=100./(100.-P20-CAC03-X*0)
ARGILE=F*((P02+C02*9.259)-138.889)/19.852)
X210=(F*((P010+C010*9.259)-138.889)/19.852))-ARGILE
X1020=(F*((P020+C020*9.259)-138.889)/19.852))-ARGILE-X210
F=F*5.
SABLE=F*0.
DO 30 I=1,5
P(I)=P(I)*F
SABLE=SABLE+P(I)
30 CONTINUE
XLIMON=100-ARGILE-SABLE
X2050=XLIMON-X210-X1020

42c

```

117      GO TO 25
118      67 WRITE(PLR,66)
119      68 FORMAT(1H+,75X,'LIMON')
120      GO TO 25
121      69 WRITE(PLR,70)
122      70 FORMAT(1H+,75X,'LIMON TRES FIN')
123      GO TO 25
124      C
125      C      FIN DE L EXECUTION
126      C
127      500 WRITE(PLR,500)
128      501 FORMAT(1H0,'FIN')
129      STOP
130      END
131      FINISH

```

PLANCHETTE OLLOY-SUR-VIROIN 192W

DOUBRES - SOLS SUR CALCAIRE

SANS DECARBONATATION

P, AVRIL - FEVRIER 1978

NUMERO LIM INF	0	2	10	20	50	100	200	500	100
S/P	2	10	20	50	100	200	500	1000	2000
082	4,7	10,6	13,3	47,6	3,5	6,9	11,8	1,2	0,4 LIMON FIN
5382	9,1	11,0	13,1	47,1	8,3	6,3	4,4	0,6	0,0 LIMON FIN
1	3,1	15,8	19,4	59,2	1,3	0,6	0,4	0,3	0,0 LIMON TRES FIN

- BANQUE DE DONNEES FORESTIERES -

Nous ne voulons pas développer ce chapitre, mais nous signalons tout de même qu'un progrès considérable en cette matière a été réalisé et que nous possédons une documentation intéressante et qu'il est possible de la consulter.

Ex. 1:-Constitution d'une banque de données forestières en vue de son exploitation par ordinateur.

-Cartographie par ordinateur, inventaire statistique et gestion forestière.

-Quelques exemples d'utilisations de l'ordinateur dans le domaine de la gestion et de la production forestières.

LES PROGRAMMES STATISTIQUES :

En plus d'initiation en informatique, un cours de statistique appliquée a été organisé. Ce cours est intense et complet. Aux principaux chapitres classiques de statistique sont joints des programmes statistiques très divers. Ces derniers sont tirés des cas pratiques et réels. Depuis 1965, la chaire de statistique et le centre de calcul et d'informatique de la Faculté de Sciences Agronomiques de Gembloux assurent la mise au point et l'exploitation d'un ensemble de programmes statistiques pour ordinateur qui sont très largement utilisés (Dagnillie 1975). Nous citons les programmes statistiques T,P,D, largement utilisés en ce Service.

L'appellation " Programmes statistiques T,P,D. " tire son nom de l'enregistrement de façon systématique sur disques, dans trois fichiers distincts : Fichier T pour les titres, Fichier P pour les paramètres et Fichier D pour les données.

Nous donnons dans ce qui suit des indications générales relatives aux fichiers, aux programmes.

Contenu et organisation des fichiers :

1/- Contenu des fichiers :

Les informations relatives aux différentes variables traitées comportent toujours, conventionnellement, quatre lignes de titres, dix paramètres et les données proprement dites.

Les différentes lignes de titre sont composées chacune d'un maximum de 48 caractères alphabétiques ou numériques. La première ligne contient normalement un numéro de référence et le ou les nom (s) du ou des expérimentateur (s) concerné (s), la deuxième ligne l'indication de l'expérience considérée (dénomination, matériel expérimental, lieu, année, etc...) et la quatrième ligne peut être utilisée pour indiquer le type de transformation de variable qui aurait éventuellement été réalisé.

Les paramètres donnent des indications relatives à la présentation des données sur leur support initial (cartes ou rubans perforés par exemple) et à l'exécution des calculs (effectifs des échantillons par exemple). Ces paramètres sont enregistrés en mémoire comme des nombres réels.

- Les données proprement dites, par contre, sont considérées comme des nombres entiers*. Elles sont suivies pour chaque variable de la valeur de la limite conventionnelle, qui en indique la fin.

2. Organisation des fichiers :

Dans les fichiers relatifs aux titres et aux paramètres (T et P), la place occupée par chaque variable est constante. Par contre, dans le fichier relatif aux données (D), cette place peut être différente d'une variable à l'autre. Le passage de deux premiers au troisième se fait par l'intermédiaire d'un pointeur, qui indique pour chaque variable la position du début des données et qui vient s'ajouter dans chaque cas, dans le fichier P, aux dix paramètres.

La fin de l'ensemble des informations enregistrées est marquée par deux astérisques (**) figurant en tête de la première ligne de titre qui n'est pas utilisée.

La dimension des fichiers doit être adaptée aux possibilités de chaque ordinateur et, éventuellement, aux besoins spécifiques des utilisateurs.

- Caractéristiques et organisation des programmes

1/- Caractéristiques des programmes :

Les programmes** possèdent autant que possible les qualités suivantes : une faible dimension en mémoire centrale, une bonne coordination de l'ensemble, une relative indépendance du système d'exploitation, une grande facilité d'utilisation et une grande facilité de lecture de résultats.

2/- Organisation des programmes :

L'ensemble des programmes a été divisé en plusieurs catégories, dont les principales sont, au cours d'un premier stade, les programmes généraux, les programmes d'analyse de la variance, les programmes de corrélation et de régression et les programmes d'ajustement.

* la limite conventionnelle de ce nombre est adoptée en fonction des contraintes des ordinateurs à mots de n bits (valeurs extrêmes de $\pm 2^n$).

** - rédigés en Fortran

Programmes généraux

Ils permettent essentiellement les " manipulations " de données entre périphériques lents de base (lecture de cartes, imprimante, perforateur de rubans, etc...) et périphériques rapides (disques magnétiques). Ces " manipulations " peuvent se faire dans les deux sens, elles peuvent se limiter à un simple chargement, - déchargement (entrée-sortie) ou être plus élaborées, elles sont toujours accompagnées de quelques éléments de statistique descriptive (calculs de moyennes, valeurs extrêmes, etc...).

Les programmes généraux sont divisés en quatre catégories : programmes du chargement, programme de sortie, programmes de transformation et programmes de description.

Structure générale :

Introduction

Les programmes généraux sont constitués de programmes principaux et de sous-routines.

Programmes :

Programmes de chargement :

Intitulé C T P D (chargement dans T.P.D.) permet de changer sur disque magnétique de données perforées sur cartes ou sur rubans ou encodées aux stations d'entrée directe (D D E : direct data entry) quelle que soit la présentation initiale de ces données (données perforées individu par individu ou variable par variable notamment) une carte-pilote est nécessaire pour indiquer le mode et l'endroit du chargement, de même que la longueur de l'enregistrement. Le premier paramètre, qui suit les quatre lignes de titres, indique toujours de façon conventionnelle, d'une part si les données sont présentées individu par individu ou variable par variable et d'autre part, si le FORMAT d'entrée est lu avant les données proprement dites. Ceci permet d'adapter le chargement au dessin des cartes de données.

Le programme F₁ F₂ (Fichier 1 à Fichier 2) permet de transférer sélectivement des variables contenues dans des fichiers T₁, P₁, D₁, vers des fichiers analogues T₂, P₂, D₂ et ce à n'importe quel endroit.

Programme de sortie :

Intitulé S T P D (sortie de T,P,D), ce programme revêt un triple rôle : Edition, impression et de perforation.

Les cinq opérations permises par ce programme S T P D étant conditionnées par la présence d'un nombre- code à additionner aux éventuels numéros de la variable. Il s'agit de :

- 1000 pour l'édition
- 2000 pour l'impression variable par variable
- 3000 pour l'impression simultanée de plusieurs variables
- 4000 pour la perforation variable par variable
- 5000 pour l'impression et la perforation variable par variable.

Programmes de transformation :

1/- Très souvent, l'analyse statistique des résultats d'expérience nécessite la transformation des variables observées, notamment en vue de stabiliser les variances ou de rendre linéaires les relations existant entre les variables. Parfois on souhaite également analyser des variables dérivées qui doivent être calculées en fonction de deux ou plusieurs variables initiales. Ces transformations et ces calculs de variables dérivées peuvent être réalisés à l'aide du programme F V A R (fonction d'une ou plusieurs variables).

2/- Il y'a lieu parfois de modifier une variable sans en altérer pour autant les valeurs initiales.

Ex : programme Clas (classement) permet de classer les observations relatives à un certain nombre de variables en fonction des valeurs croissantes d'autres variables enregistrées sur disque.

Le programme COMB (Combinaison) permet de reconstituer sur disque une nouvelle variable en sélectionnant différentes valeurs d'une ou plusieurs variables initiales elles-mêmes enregistrées dans les fichiers T, P, D.

3/- Programme CORR (correction): il permet la correction des titres, des paramètres et des données de variables appartenant aux fichiers T, P, D.

Programme de description :

Le programme Hist (Histogramme) est destiné à établir des distributions de fréquences à une dimension (fréquences simples et fréquences cumulées), à les représenter graphiquement sous forme d'histogrammes et à en calculer les principaux paramètres (valeurs extrêmes, moyennes, écarts types, etc.,).

Le programme R E G L (Regression linéaire) permet d'établir des distributions de fréquences à deux dimensions et d'en calculer les principales caractéristiques (voir " programmes de corrélation et de régression).

Exemple :

En matière forestière, le problème se pose assez fréquemment de calculer des tarifs ou des tables de cubage exprimant le volume V des arbres en fonction de leur circonférence C . Ce calcul nécessite la détermination d'une équation de régression simple, qui peut être du type suivant :

$$V = b_0 C^{b_1}$$

ou $\text{Log } V = b'_0 + b_1 \text{ log } C$

En supposant que les circonférences et les volumes d'un certain nombre d'arbres observés (20 arbres dans l'exemple considéré) aient été préalablement perforés ou encodés, et vérifiés, les programmes présentés permettent notamment d'effectuer de façon très simple les opérations suivantes :

- Le chargement sur disque des circonférences et des volumes.
- Le calcul des logarithmes des circonférences et des volumes.
- L'impression des quatre variables C , V , $\text{Log } C$ et $\text{log } V$.
- L'établissement de distributions de fréquence à deux dimensions et la détermination des équations de régression, tant pour les variables initiales que pour les variables transformées.

- CONCLUSION -

L'utilisation de l'ordinateur pour la gestion de données ouvre de nouveaux horizons dans la méthodologie d'exploitation des fichiers de sols. L'établissement d'une banque de données de sols devra permettre aux utilisateurs bien compétents une exploitation rationnelle de ces données. Partout dans le monde, des banques de données de sol sont en voie de création ;

- Citons CANSIS (CANADA), Montpellier, ORSTOM, SOGREHA (France), et celle de Tunisie, etc... Rares sont en fonctionnement. Le point commun et qui est fondamental dans ces banques, est le langage de RITDS⁽¹⁾ (A Gembloux, 95 % des termes utilisés sont ceux de RITDS). D'ailleurs le RITDS prévoit la création d'une "Banque Internationale Centrale", qui pourra avoir son propre langage, mais permettant aussi un échange de données pédologiques entre les différentes banques.

Actuellement les programmes de traitement de données de sols sont pratiquement inexistantes. Ils seront peut-être mieux conçus une fois que ces banques seront fonctionnelles.

Il est évident que la réussite d'une banque de données de sols dépend beaucoup de la compétence des utilisateurs. Il est donc indispensable de divulguer cette nouvelle méthodologie de travail. Il est donc possible de divulguer d'une part le potentiel de données que renferment les fichiers de sols et d'autre part exposer les possibilités offertes par l'ordinateur dans la gestion de données Pédologiques. Cette démarche incite les utilisateurs à formuler leur demande et c'est en fonction de ces demandes que les programmes de traitement de données seront élaborés et par conséquent la banque fonctionnera normalement. D'où la nécessité d'organiser dans les meilleurs délais, des journées d'information sur ce thème.

(1) Réseau international de traitement de données de sols.

- BIBLIOGRAPHIE -

Ils possèdent une documentation fort intéressante sur " Informatique et Statistique Appliquées " et dont nous ne pouvons pas l'exploiter dans ce rapport.

Il est toujours possible de la consulter sur place. La liste des documents est la suivante :

- N° 1 : -Informatique - Par : P. DAGNELIE
- N° 2 : -Dix années d'informatique en Agronomie - Par : P. DAGNELIE
- N° 3 : -Le Statisticien et l'Ordinateur, Par : P. DAGNELIE,
- N° 4 : -Programmes Statistiques T P D (Version 3) : Présentation Générale. Par : P. DAGNELIE,
- N° 5 : -Programmes Statistiques T P D (Version 3) : Programmes Généraux - Par : J.P. GRAYET et G. ROUSSEAU.
- N° 6 : -L'organisation intégrée des essais en champ : Principes Généraux, Par : P. DAGNELIE,
- N° 7 : -L'organisation intégrée des essais en champ : Exemple de réalisation, Par : A. JANVIER, C. MAUDOUX et G. ROUSSEAU.
- N° 8 : -Constitution d'une banque de données forestières en vue de son exploitation par ordinateur, Par J. RONDEUX,
- N° 9 : -Quelques exemples d'utilisations de l'ordinateur dans le domaine de la gestion et de la production forestières,
- N° 10: -Propos sur un système de comptabilité agricole par ordinateur,
- N° 11: -L'ordinateur et l'étude des grands troupeaux d'élevage,
- N° 12: -Banque et traitement de données pédologiques,
- N° 13: -Programmes statistiques T P D (version 3) : Programmes d'analyse de la variance, Par : J.J. CLAUSTRIAUX et R. OGER,
- N° 14: -A propos des transformations de variables, Par : P. DAGNELIE,
- N° 15: -L'analyse de la variance et les plans d'expérience : Principes généraux et exemples, Par : P. DAGNELIE,
- N° 16: -Génération et contrôle de validité de nombres pseudo-aléatoires sur un ordinateur à mots de 16 BITS, J.J. CLAUSTRIAUX,
- N° 17: -Etude de l'échantillonnage (aléat) de trois espèces de pucerons des céréales, Par: L. BARIDEAU, P. DAGNELIE et G. LATTEUR,
- N° 18: -L'estimation de l'intensité des attaques de jaunisse de la Letterave, Par : R. JADOT et R. OGER,
- N° 19: -Contribution à l'étude de l'échantillonnage en futaie feuillue, Par : J.P. GRAYET,
- N° 20: -Cartographie par ordinateur, inventaire statistique et gestion forestière, Par : J.P. GRAYET.

- Bibliographie...2

- N° 21 : - L'échantillonnage appliqué à la détermination des propriétés physiques et mécaniques du bois, Par : R. OGER et LECLERQ,
- N° 22 : - Principes d'expérimentation, par : P. DAGNELIE,
- N° 23 : - Analyse statistique de résultats d'essais à blanc,
- N° 24 : - La fumure du cotonnier : Analyse statistique d'une série d'essais locaux, par : G. T. Van RENTERGHEM et P. DAGNELIE,
- N° 25 : - Programmes statistiques T P D (Version 3) : Programmes de corrélation et de régression, par : C. DEBOUCHE et K. IN,
- N° 26 : Eléments de calcul matriciel, . . .
- N° 27 : - Introduction à l'analyse statistique à plusieurs variables,
- N° 28 : - Un organigramme des méthodes d'analyse statistique à plusieurs variables, par P. DAGNELIE,
- N° 29 : -
- N° 30 : - L'emploi d'équations de régression simultanées dans le calcul de tables de production forestières, par : P. DAGNELIE,
- N° 31 : - Analyse factorielle d'exploitations agricoles,
- N° 32 : - La détermination et la comparaison des fumures optimales par la méthode des variantes systématiques et l'analyse statistique à plusieurs variables. Par : P. DAGNELIE,

FIN

58

VUM