



*Institut de Recherches pour les Huiles et Oléagineux*

Département du Centre de Coopération Internationale  
en Recherche Agronomique pour le Développement (CIRAD)

**Rapport de mission Informatique et Biométrie  
à la station de recherches sur le Palmier à huile  
de LA DIBAMBA (CAMEROUN)  
14 au 19 avril 1989**

**F. BONNOT**

**IRHO/CIRAD  
Division BIOMETRIE  
Doc. N° 2194**

## INTRODUCTION

Cette première mission de M. BONNOT au Cameroun, effectuée du 14 au 19 avril 1989, avait pour but l'étude des problèmes de Biométrie et d'Informatique rencontrés sur la station de recherches sur le palmier à huile de La Dibamba, et l'installation de logiciels spécifiques au traitement des données de palmier à huile. Cette station s'est en effet équipée en 1988 d'un micro-ordinateur compatible IBM-PC en remplacement de son ancien APPLE II, puis plus récemment d'un deuxième appareil du type GOUPIL G5-286 financé par le CIRAD.

Pendant cette mission, une visite a été effectuée à YAOUNDE au cours de laquelle nous avons pu rencontrer M. AYUK TAKEM, directeur de l'IRA, M. MESSENGER, et M. KAISER.

Une visite à la SOCAPALM nous a permis de rencontrer M. MBELLA MBAPPE, directeur de la SOCAPALM, et de travailler avec M. HORNUS sur les problèmes d'informatisation de ses données.

Le travail réalisé à La Dibamba au cours de cette mission a abouti à l'installation de plusieurs logiciels permettant de réaliser sur place une partie de la gestion et du traitement des données des expériences de la station :

- regroupement par arbre et par campagne des données de production mensuelle ;
- gestion du fichier Protocole, du fichier Parcelles, et du fichier Arbres des expériences ;
- analyse statistique des données des expériences.

Nous remercions M. KAMGA FONDJO, directeur de la station de La Dibamba, ainsi que tous nos collègues, pour la qualité de leur accueil.

## 1 - GESTION ET EXPLOITATION DES DONNEES EXPERIMENTALES

### 1.1 - Saisie et transmission des données : programme PREDEC

La saisie des données mensuelles de production par arbre s'effectue depuis plusieurs mois sur micro-ordinateur fonctionnant sous MS/DOS. Pour la campagne 88/89, la saisie s'est effectuée en deux étapes :

- Saisie des données de Juillet à Octobre 88 au moyen de la version 3 du programme PREDEC et envoi des disquettes à Montpellier ;

- Saisie des données à partir de Novembre 88 au moyen du nouveau programme PREDEC (version 4.0) envoyé par la division Biométrie.

La conversion au format du nouveau programme des données de Juillet à Octobre 88 a été réalisée à Montpellier. Les fichiers correspondants ont été installés lors de cette mission sur les micro-ordinateurs de la station qui dispose maintenant de ses données depuis le début de la campagne 88/89 sous une forme homogène et exploitable. Les données de la campagne 87/88 stockées à Montpellier seront également envoyées sur disquettes à La Dibamba après conversion au nouveau format.

L'utilisation à La Dibamba du nouveau programme PREDEC est satisfaisante. Il convient néanmoins de rappeler quelques précautions indispensables à la fiabilité des données et à la bonne gestion des fichiers.

#### a) Gestion des fichiers

Les données d'un mois particulier sont contenues dans un ensemble de 3 fichiers dont les noms sont déterminés par le programme en fonction de la station et de la date de production. Les données de La Dibamba en Février 89 par exemple sont contenues dans les 3 fichiers :

LDPR8902.HDR  
LDPR8902.DAT  
LDPR8902.NDX

S'il existe plusieurs versions différentes de ces 3 fichiers, le programme ne pourra utiliser qu'une seule d'entre elles à la fois et ne pourra donc pas réaliser la fusion des données provenant des différentes versions.

Ce principe a plusieurs conséquences sur la gestion des données :

- Les 3 fichiers d'un même mois peuvent être dupliqués autant de fois que nécessaire, mais si les copies sont modifiées indépendamment les unes des autres il n'y a aucun moyen de

regrouper les différentes modifications au sein d'une seule copie. En particulier si le programme PREDEC est installé sur 2 micro-ordinateurs comme c'est le cas à La Dibamba, la saisie des données et les modifications doivent toujours être effectuées sur le même appareil, le second recevant des copies des fichiers pour exploitation des données.

- Si des erreurs ou omissions se sont produites dans des fichiers envoyés à Montpellier, les corrections doivent être effectuées dans les fichiers d'origine dont une nouvelle copie sera expédiée à Montpellier. Des fichiers ne contenant que des corrections ne pourraient en effet être fusionnés avec des fichiers contenant d'autres données du même mois.

#### b) Saisie des données

Le contrôle des totaux des nombres de régimes et poids de régimes effectués pour chaque ligne de plantation n'empêche pas d'autres erreurs de se produire au moment de la saisie. Il convient en particulier d'être vigilant sur le risque de décalage de données entre arbres d'une même ligne, surtout pour les lignes peu productives.

### 1.2 - Regroupement des données : programme REGROUP

Le programme REGROUP réalisé par la division Biométrie a été installé sur les micro-ordinateurs de La Dibamba.

#### 1.2.1 - Utilité du programme REGROUP

Le programme REGROUP permet de réaliser des regroupements par arbre ou par ligne sur n'importe quelle période d'une durée maximum de 12 mois des données de production mensuelles saisies avec le programme PREDEC. Il est utile pour les opérations suivantes :

- Edition des données mensuelles par arbre et par mois ; la sortie est alors un document présenté selon un format identique à celui des anciens cahiers IKG 7 qui permet une présentation claire et exhaustive des données.

- Edition des totaux par ligne et par mois ; ce document résumé est utile pour la vérification des données puisque les lignes oubliées au cours d'un mois apparaissent avec des productions nulles. Il est conseillé d'éditer cette sortie 6 à 8 mois après le début de la campagne pour une première vérification, puis à la fin de la campagne pour vérification générale lorsque toutes les données sont disponibles. Ce document ne sera plus envoyé par Montpellier puisque sa constitution est désormais possible sur la station.

- Extraction des données par arbre dans un fichier ASCII récupérable par d'autres logiciels permettant différentes formes d'exploitation des données.

### 1.2.2 - Mode d'emploi succinct du programme REGROUP

Une démonstration du programme a eu lieu sur la station. Les principales étapes de son fonctionnement, d'ailleurs très simple, sont décrites ci-dessous. Le mode d'emploi détaillé sera envoyé séparément.

L'utilisation du programme se déroule en 2 phases :

- la phase de création d'un fichier binaire de données regroupées ;

- la phase d'exploitation du fichier binaire.

a) Création d'un fichier binaire de données regroupées

Cette opération s'effectue au moyen de la fonction 2 du menu principal. Elle a pour effet de lire les données enregistrées avec le programme PREDEC et de les regrouper en un fichier binaire unique comprenant les données de production qui doivent être définies dans la grille suivante qui apparaît à l'écran lors de l'appel de la fonction :

Programme REGROUP  
Station : LA DIBAMBA

- 2 -	Création d'un fichier binaire de données regroupées
-------	---

Début de regroupement:  Année : Mois :  Première ligne à extraire:  Parcelle: Ligne :	Fin de regroupement:  Année : Mois :  Dernière ligne à extraire:  Parcelle: Ligne :
---	---

Il est nécessaire de préciser l'année et le mois de début et de fin de regroupement. Selon que la première et la dernière ligne à extraire sont ou non précisées, le programme lira une partie ou la totalité des données et constituera le fichier binaire dont le nom sera toujours LDGROUP.BIN (pour La Dibamba) et qui viendra écraser le précédent fichier du même nom, s'il existe. Si l'on désire conserver plusieurs fichiers binaires sous le même répertoire, il est nécessaire de les renommer au moyen de la commande REN de MS/DOS. La durée de la phase de lecture des données et de création d'un fichier binaire dépend de la quantité des données à extraire et peut atteindre plusieurs dizaines de minutes.

## b) Exploitation du fichier binaire

L'exploitation s'effectue toujours sur le fichier LDGROUP.BIN, qui sera en général le dernier créé. Si l'on a conservé plusieurs fichiers binaires sous des noms différents, il est nécessaire de renommer le fichier à utiliser en LDGROUP.BIN au moyen de la commande REN de MS/DOS. L'exploitation s'effectue au moyen des fonctions 3, 4, ou 5 du menu principal et porte sur les données contenues dans le fichier binaire qui ont été définies lors de sa création.

- La fonction 3 assure l'édition du fichier binaire par arbre ;

- La fonction 4 permet l'édition des totaux par ligne du fichier binaire ;

- La fonction 5 permet la conversion du fichier binaire en fichier texte.

### 1.3 - Exploitation des données : programme STATP

#### 1.3.1 - Présentation du programme STATP

Le programme STATP a été installé sur les micro-ordinateurs de La Dibamba. Il s'agit d'un logiciel destiné à assurer la gestion complète des données des expériences sur palmier à huile, ainsi que leur analyse statistique selon des méthodes classiques :

- Gestion du protocole des essais ;
- Gestion d'un fichier Arbres par essai ;
- Gestion d'un fichier Parcelles expérimentales par essai ;
- Gestion d'un fichier Analyses de régimes (non opérationnel actuellement) ;
- Calcul des données parcellaires à partir des données par arbre, regroupement sur plusieurs campagnes, et réalisation des analyses statistiques des moyennes parcellaires.

Ce logiciel est actuellement en cours de développement et ce n'est qu'à titre expérimental qu'une première version a été installée à La Dibamba. Dans son état actuel il permet néanmoins de réaliser une partie du traitement statistique des données.

Une démonstration du fonctionnement du logiciel a été effectuée sur un essai de sélection (LD-GP 3, Lattice équilibré, 150 parcelles) et un essai d'agronomie (LD-CP 8, factoriel 2 X 2 X 2 X 2, 32 parcelles). Le fichier Protocole, le fichier Parcelles, et le fichier Arbres de ces 2 essais ont été saisis sur les micro-ordinateurs de la station. Les données mensuelles de production de

juillet 88 à février 89 enregistrées avec le programme PREDEC ont pu être transférées dans le logiciel STATP qui a permis de calculer les cumuls par arbre sur le début de la campagne 88/89, puis les moyennes parcellaires à partir des cumuls par arbre. L'analyse statistique du LD-GP 3 a pu être réalisée directement avec le programme STATP. Celle du LD-CP 8 a été effectuée au moyen d'un programme séparé dans lequel les données parcellaires calculées par STATP ont été introduites.

### 1.3.2 - Conseils d'utilisation du programme STATP

Le programme STATP comprend de nombreuses fonctions, mais on peut distinguer 2 phases principales dans son utilisation :

- la phase de création des fichiers d'une expérience ;
- la phase d'enregistrement et d'exploitation des données.

#### a) Création des fichiers d'une expérience

Dans l'état actuel du logiciel, chaque expérience comprend un fichier Protocole, un fichier Parcelles, et un fichier Arbres. Ces fichiers sont créés automatiquement lors du catalogage de l'expérience et sont destinés à contenir les caractéristiques de l'expérience (position des arbres, lignées, etc ...) et les données observées. Il est recommandé de saisir les caractéristiques dans l'ordre Protocole - Parcelles - Arbres, certaines caractéristiques se reportant automatiquement du protocole sur les parcelles et des parcelles sur les arbres. On utilisera donc successivement les fonctions suivantes.

#### Fonction 3 : Saisie / Modification du protocole expérimental.

Les noms des facteurs et des niveaux de chaque facteur doivent être saisis dans la colonne correspondante. En Sélection, on n'étudie en général que le facteur Lignée qui commence au niveau 1 (le niveau 0 est alors inutilisé) : il convient alors de spécifier chaque lignée dans la première colonne, puis de saisir dans la colonne correspondant au facteur 1 les informations qui devront figurer en face de chaque niveau dans les analyses statistiques (lignée, croisement, ...).

#### Fonction 4.2 : Saisie / Modification des données parcellaires.

Cette fonction permet de saisir les caractéristiques de chaque parcelle : répétition, bloc, niveau de chaque facteur, lignée. Si les numéros de lignée correspondant aux niveaux du facteur 1 figurent au fichier Protocole, la lignée de chaque parcelle est déterminée par le niveau du facteur 1 et reportée automatiquement dans le champ correspondant. La fonction 4.2 permet également de visualiser et modifier les données observées sur chaque parcelle, mais en principe ces données sont calculées à partir du fichier Arbres.

### Fonction 5.2 : Création / Remplacement d'arbres.

Cette fonction permet de créer des arbres au fichier et d'en saisir les caractéristiques : année et mois de plantation, numéro de parcelle expérimentale, lignée, numéro de géniteur, variété, état (producteur ou non). Si le numéro de lignée correspondant à la parcelle expérimentale figure au fichier Parcelles, la lignée de chaque arbre est déterminée par le numéro de parcelle et reportée automatiquement dans le champ correspondant. La création ou le remplacement des arbres ont lieu par zones rectangulaires définies par l'intersection de deux bandes délimitées la première par deux numéros de lignes, la seconde par deux numéros d'arbres.

### Fonction 5.3 : Saisie / Modification des données par arbre.

Cette fonction permet de visualiser et modifier arbre par arbre les données créées par la précédente. Elle permet également de visualiser et modifier les données observées sur chaque arbre, mais en principe ces données sont calculées à partir des données mensuelles saisies ligne par ligne.

#### b) Enregistrement et exploitation des données.

Bien que le programme STATP dispose avec la fonction 5.4 de son propre module de saisie des données mensuelles, cette fonction analogue à la fonction de saisie du programme PREDEC ne doit pas être utilisée à La Dibamba pour la saisie qui doit impérativement être effectuée avec le programme PREDEC. La fonction 5.4 peut néanmoins servir à vérifier que les données mensuelles ont été correctement chargées.

La procédure correcte d'enregistrement des données mensuelles consiste à charger ces données à partir d'un fichier externe créé par le programme REGROUP. Cette procédure est décrite ci dessous :

- Avec le programme REGROUP, créer un fichier binaire contenant les données de l'expérience regroupées sur une période correspondant à une campagne entière (juillet à juin). Le fichier binaire peut contenir des arbres excédentaires, mais doit contenir au moins les arbres de l'expérience.

- Toujours avec le programme REGROUP, convertir ce fichier binaire en fichier texte au moyen de la fonction 5. Tous les champs proposés doivent être présents (valider l'ensemble de l'écran avec la touche F2).

- Exécuter ensuite le programme STATP et appeler la fonction 5.8 de chargement externe des données mensuelles. Le nom du fichier de données à charger est celui que l'on aura choisi d'attribuer au moment de l'exécution du programme REGROUP.



L'analyse statistique des données s'effectue au moyen de la fonction 8.2 :

- Saisir le nombre de regroupements souhaités pour chacune des variables.

- Pour chacun des regroupements, préciser toutes les campagnes sur lesquelles il convient de calculer les moyennes. Le programme procède alors au calcul des moyennes sur les périodes et les variables et crée un fichier intermédiaire contenant les données à analyser.

- Choisir le programme d'analyse (actuellement seuls le lattice équilibré et le bloc randomisé sont disponibles).

## 2 - ANALYSES DE REGIMES

### 2.1 - Saisie des données

Jusqu'à maintenant les bordereaux servant à la transcription des données brutes des analyses de régimes étaient envoyés tels quels à Montpellier qui assurait la saisie des données et le calcul des résultats et renvoyait périodiquement les données élaborées sur la station sous forme de listing. Le matériel informatique dont la station est maintenant équipée est capable d'assurer localement ces opérations ainsi que le stockage des données sur support magnétique. Dans ce but la division Biométrie a envoyé en mai 89 un programme permettant la saisie des données brutes, le calcul des résultats des analyses de régimes, leur stockage sur disque dur ou disquettes, et l'interrogation des fichiers. Des copies sur disquettes des données saisies seront envoyées à Montpellier pour archivage et analyse statistique. Actuellement il n'existe pas de programme sur micro-ordinateur permettant de regrouper les données d'analyses de régimes et de les combiner avec les données de production pour calcul et analyse statistique du rendement en tonnes par hectare. Les versions ultérieures du programme STATP offriront cette possibilité.

### 2.2 - Calcul des abaques

La méthode de calcul des abaques a été discutée avec M. WALTER. Le calcul de la teneur en huile  $H$  à partir de la densité corrigée  $D$  est obtenu en assimilant la fonction  $H = f(D)$  à deux droites :

$$H = a_1 D + b_1 \quad \text{si } D \leq D_0$$

$$H = a_2 D + b_2 \quad \text{si } D > D_0$$

La détermination des paramètres  $a_1$ ,  $b_1$ ,  $a_2$ ,  $b_2$ , et  $D_0$  s'effectue de manière relativement empirique en traçant sur un graphe les points  $(D, H)$  issus des mesures systématiques effectuées après l'ouverture d'un nouveau fût d'orthodichlorobenzène, puis en

visualisant le point d'intersection des deux droites dont les équations sont ensuite calculées plus précisément au moyen de deux régressions simples indépendantes. L'inconvénient de cette méthode est qu'elle ne donne pas à coup sûr la meilleure solution, l'équation des droites dépendant du point d'intersection lui même estimé graphiquement. En réalité il s'agit d'un problème de régression non-linéaire consistant à trouver les 5 paramètres qui minimisent la quantité  $S = \sum (H \text{ observé} - H \text{ calculé})^2$ . Le petit nombre de points (quelques dizaines) ne rend pas nécessaire l'utilisation de techniques sophistiquées de régression non-linéaire. Soit en effet N le nombre de points : il existe N-3 manières différentes de séparer les N points en 2 groupes situés de part et d'autre d'une densité donnée et comportant au moins 2 points. Il suffit alors, pour chacune des N-3 manières, d'effectuer une régression simple sur chacun des 2 groupes de points correspondants et de calculer la quantité S. La plus petite des N-3 quantités S correspond aux deux droites procurant la meilleure précision, dont l'intersection est alors calculée à partir des équations.

Un programme permettant la réalisation de ce calcul sera écrit par la division Biométrie et envoyé sur la station.

### 3 - LOGICIEL DF

Le logiciel DF a été développé par l'IRHO pour la gestion sur micro-ordinateur des données de diagnostic foliaire et de fumure sur les plantations industrielles. Outre la saisie, le stockage, la consultation, et le regroupement des données selon différents critères, il permet d'effectuer le calcul automatique des recommandations de fumure à appliquer aux différentes unités en fonction des teneurs foliaires en éléments minéraux.

Le logiciel DF, déjà installé sur les micro-ordinateurs de La Dibamba, a été utilisé avec M. CHAILLARD lors de la présente mission pour la saisie de la structure DF de la station (nom et caractéristiques des divisions, des unités DF, et des parcelles), ainsi que des données de DF et de fumure de 1988. Plusieurs exemples d'édition des données et de calcul automatique des recommandations de fumure ont été effectués.

Il a été convenu que les améliorations qui seront apportées au logiciel porteront sur les points suivants :

- calcul des quantités totales d'engrais apportées sur les divisions et sur la plantation ;

- possibilité d'utiliser des barèmes à 2 entrées, permettant le calcul des recommandations de doses pour un engrais à partir de la donnée des teneurs foliaires de deux éléments minéraux.

## 4 - PROGRAMMES DIVERS

### 4.1 - Paramétrage d'imprimante

Un petit programme appelé IMP.EXE permettant de paramétrer les imprimantes de la station afin d'utiliser différents types de caractères (caractères normaux, condensés, larges, gras) a été écrit et installé sur les micro-ordinateurs.

### 4.2 - Compactage de fichiers

De nouvelles versions des programmes de compactage et décompactage de fichiers PKARC et PKXARC ont été installées. Ces versions sont plus rapides et plus efficaces que les précédentes. Leur utilisation est simple et ne pose pas de problème particulier d'autant que leur mode d'emploi peut être affiché par simple appel des programmes. Les divers fichiers de données qui doivent être envoyés à Montpellier peuvent être enregistrés, si cela s'avère nécessaire, sous forme compactée au moyen de PKARC.

### 4.3 - Gestionnaires de fichiers

Il existe sur le marché de nombreux gestionnaires de fichiers. L'IRHO utilise principalement FOXBASE, qui est équivalent à dBASE III. Le logiciel utilisé le plus fréquemment à La Dibamba et même par l'ensemble des chercheurs du CIRAD au Cameroun semble être OPEN ACCESS. Renseignements pris, il s'agit d'un logiciel puissant capable de gérer de véritables bases de données et qui semble au moins aussi performant que FOXBASE. De plus son utilisation très répandue dans les différents centres de l'IRA au Cameroun fait que ses utilisateurs peuvent bénéficier d'une assistance technique de qualité de la part de la cellule Biométrie-Informatique de l'IRA. Toutes ces raisons poussent à recommander son utilisation, avec la seule contrainte que la division Biométrie ne pourra pas dans l'immédiat apporter son soutien direct aux utilisateurs. Cette situation pourrait être amenée à changer si OPEN ACCESS se développait dans d'autres départements du CIRAD à Montpellier, ce qui semble être la tendance actuelle.

## CONCLUSION

La station de La Dibamba dispose maintenant d'un équipement informatique performant constitué de deux micro-ordinateurs fonctionnant sous le système d'exploitation MS/DOS assurant une compatibilité totale avec le matériel utilisé dans les différents centres de l'IRA et du CIRAD.

Au cours de cette mission plusieurs logiciels plus particulièrement orientés vers la gestion et le traitement des données de palmier à huile ont pu être installés.

Le programme REGROUP permettant le regroupement par arbre et par campagne des données de production mensuelle est venu compléter le programme PREDEC qui fonctionnait déjà efficacement depuis quelques mois pour la saisie, le stockage, et la transmission de ces données.

Le programme STATP permet, grâce à l'utilisation d'un fichier Protocole, d'un fichier Parcelles, et d'un fichier Arbres pour chaque expérience, d'assurer la gestion d'une partie des données expérimentales de palmier à huile et d'en effectuer l'analyse statistique. Il est compatible avec le programme REGROUP dont il peut recevoir des données.

Il a été convenu qu'un programme de calcul des analyses de régimes serait prochainement mis en place.