



Institut de Recherches pour les Huiles et Oléagineux

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

DATA PROCESSING AND BIOMETRY

**Report on the Mission to MARIHAT
from 10th to 24th September 1989**

F. BONNOT

IRHO Biometry Division

Document No.2229^{bis}

January 1990

CONTENTS

SUMMARY AND CONCLUSION	1
1. COMPUTERIZATION OF EXPERIMENTAL DATA	2
1.1. STATPALM software	2
1.1.1. General	2
1.1.2. Aspects specific to Breeding	2
1.1.3. Aspects specific to Agronomy	3
1.1.4. Examples	3
1.1.5. Expected evolution	4
1.2. Changing over from WANG system to the new system	5
2. STATISTICAL PROBLEMS RELATIVE TO EXPERIMENTS	5
2.1. Power of an experiment and the number of replications	5
2.2. Methods of reducing standard deviation σ	7
2.3. Study in the case of a split-split plot trial	7
2.4. Response curves and surfaces	10
3. COMPUTERIZATION OF LEAF ANALYSIS AND FERTILIZER DATA	14
4. MISCELLANEOUS	15
4.1. Organization of data processing at Marihat	15
4.2. Software technical manuals	15
4.3. SAS software	16
4.4. Setting printer parameters	16

ANNEXES

SUMMARY AND CONCLUSION

The main aims of this mission were to improve the data processing system for oil palm experimental data management and exploitation, installed in the Marihat microcomputers since April 1988, and study of various statistical problems concerning experimental design and statistical analysis of data.

The new STATPALM software developed with a view to total management of oil palm experimental data, along with their statistical analysis based on conventional methods was installed in the microcomputers at the station. This software, which was tested on several Marihat experiments, gave satisfactory results. It will make it possible, in the near future, to completely replace the programs used on the old WANG 2200 computer.

The problems relative to experiments and biometry were discussed at a meeting attended by representatives from all the departments. The subjects broached involved the parameters determining the power of experiments, the methods used to increase this power and the experimental designs making it possible to calculate response curves and surfaces.

Finally, there were several demonstrations of the DF software on Marihat agronomical data. This software, developed by IRHO, enables microcomputer management of the data used in mineral nutrition monitoring on commercial oil palm plantations and the automatic computation of recommendations for the fertilizers to be applied.

1. COMPUTERIZATION OF EXPERIMENTAL DATA

1.1. STATPALM Software

1.1.1. General

Up to April 1988, experimental data management and exploitation at Marihat was undertaken on the WANG 2200 computer using programs written during previous IRHO biometry missions. In April 1988, a new program enabling the inputting and storage of production data on the station's IBM-PC compatible microcomputers was installed. Further development of this program led on to the STATPALM software which was installed at Marihat during this mission. Detailed instructions for using this software are given in annex I.

The STATPALM software was developed by IRHO with a view to total management of oil palm experimental data by microcomputer, along with their statistical analysis using conventional methods. It comprises the modules required for the following operations:

- ♦ Management of trial protocol,
- ♦ Management of data per tree in the trials, including monthly production data,
- ♦ Management of data per experimental plot in the trials,
- ♦ Calculation of plot means based on data per tree and grouping together of several campaigns,
- ♦ Statistical analysis of plot means.

This software is both more powerful and simpler to use than the old programs developed on the WANG 2200 computer. The software was translated into English during this mission and various modifications necessary for taking into account aspects specific to Marihat were made. In view of the differences that exist between Agronomy and Breeding as far as data management is concerned, two different versions had to be developed. The first was installed in the Agronomy Department computer and the second in the Breeding Department, and both versions were installed in several Statistics/Data Processing Service computers.

1.1.2. Aspects specific to Breeding

In breeding experiments, bunch harvesting and the recording of production data per tree (bunch number and weight) are carried out row by row, each tree marked with an order number along the row. This is why, in the version of the STATPALM software dedicated to Breeding, the trees are identified by the planting block, the row number and the number of the tree in the row. Each tree also has the number of the experimental plot to which it belongs. New tree creation is in rectangular zones which do not necessarily correspond to experimental plots; it is thus possible to take into account trees that do not belong to any experimental

plot. Monthly production data are entered on the same screen for all the trees in the same row and the data for experimental plots are grouped together automatically in accordance with the allocation of trees to the plots. The corresponding version of the program is contained in the SPSEL.EXE file.

1.1.3. Aspects specific to Agronomy

In agronomy experiments, bunch harvesting and production data per tree (bunch number and weight) are recorded per experimental plot, with each tree having an order number in the plot. This is why, in the version of the STATPALM software dedicated to Agronomy, the trees are identified by planting block, the experimental plot number and the number of the tree in the plot. The number of the tree in the planting row is not recorded. New trees are created per experimental plot and it is not possible to take into account trees that do not belong to any experimental plot, unless they are given an imaginary plot number. Monthly production data are entered on the same screen for all the trees in the same experimental plot. The corresponding version of the program is in the SPAGRO.EXE file.

1.1.4. Examples

Each of the two STATPALM software versions was tested at Marihat, the first on an agronomy experiment (MA-PM 1) and the second on a breeding experiment (BJ-SS 8). The 2 experimental designs are complete randomized blocks. The characteristics of these trials are summarized in the following table:

	MA-PM 1	BJ-SS 8
Number of treatments	6	18
Number of experimental plots	18	108
Number of trees	288	1728
Planting year	76	82
Production years entered	84-85	86

An experimental protocol was entered for each of these trials, along with experimental plot and tree characteristics and monthly production data for one or two years. The data for a period of one or two years per tree and per experimental plot were then calculated by the software and the corresponding statistical analyses were printed out. These analyses are given in annex II.

1.1.5. Expected evolution

The STATPALM software is continually evolving and is destined to be completed or improved, especially in the field of bunch analyses, management of progenies, parents and crosses and of statistical analysis programs.

1.1.5.1. Bunch analyses

Later versions of the software will make it possible to record and store bunch analyses. Contrary to the older programs developed on the WANG, which only made it possible to take into account 40 trees per cross subjected to a maximum of 4 analyses, the new program will not fix any limits to the number of bunch analyses.

At the moment, the results of bunch analyses used by Breeding come from the analysis laboratory, where they are calculated from raw data. In order to ensure better data control by breeders, it would be better if this calculation were carried out on the Data Processing Service microcomputer, which would then receive the raw data from the laboratory. It was therefore agreed that the STATPALM software would allow the acquisition and storage of computed data for the time being, whilst offering the subsequent possibility of calculating analysis results from raw data which would then be entered using the software.

1.1.5.2. Progenies, parents and crosses

The later versions of the software will ensure the management of several files containing information relative to progenies, parents and crosses. The corresponding data will be calculated from the data per tree entered into the software. These files will make it possible to obtain data on the genetic ancestry of the material in the experiment and on its agricultural potential.

1.1.5.3. Statistical analysis programs

It was agreed that the statistical analysis programs which would be given priority to complete those already existing, should make it possible to analyze factorial designs of the following type:

- 3 x 3 x 3
- 3 x 3 x 3 subdivided
- 4 x 2 x 2 x 2

Other designs will be included in the program when necessary.

1.2. Changing over from the WANG system to the new system

Transferring the data entered in the WANG to the new MS-DOS microcomputers would probably be possible using a device consisting of electronic boards and software enabling the two types of computer to be linked, provided special programs are written for transferring the data to the STATPALM software. However, so far the WANG representative in Medan has not confirmed that it is possible to obtain such equipment. In the absence of further information, the transfer of data from the WANG system to the new system will only be possible by re-entering the data manually. The experiments will then have to be examined case by case, so as to decide which data should be entered. Several situations could arise:

- ♦ continue using the WANG with no transfer for the experiments soon to be halted and for which it is not worth keeping raw data.
- ♦ enter the old annual data per tree if necessary for the experiments under way which are due to be continued for several more years.
- ♦ enter the old plot data for the experiments under way for which it is not necessary to keep data per tree, especially in the case of agronomy trials; recording data for coming years can be done normally per tree and the grouping of data per plot will be done by the STATPALM software.
- ♦ for recent experiments, enter all the data in the new system.

For certain experiments, it will be necessary to input the data figuring in the harvesting records or sheets, since the corresponding files may no longer be on the WANG disks.

New experiments should not be stored on the WANG. The STATPALM software should be used in this case.

2. STATISTICAL PROBLEMS RELATIVE TO EXPERIMENTS

The problems relative to experiments and biometry were discussed at a meeting attended by representatives from all the departments. The subjects discussed are summarized below.

2.1. Power of an experiment and the number of replications

The level of a test, called α , usually fixed arbitrarily at 5% or 1%, is the probability that the statistical analysis declares significant an effect which does not exist in reality: in this case a first degree error is committed. Whilst this probability is well known, the same cannot be said of the second degree

error, called β , which is committed when the statistical analysis does not manage to detect an effect that really exists. The power of an experiment is equal to $1-\beta$; it is therefore the probability of detecting the effect of a factor during statistical analysis. The greater the power of the experiment the better it is. Unlike the level, this probability depends on the intensity of the effect studied. It also depends on other parameters that can be specified using the following simple example.

Let us take an experiment in complete randomized blocks with n replications (hence n blocks), intended to compare 2 treatments. It is possible to demonstrate the following relationship (in the case of a two-tailed test):

$$\delta = \sigma \sqrt{\frac{2}{n}} [t(n-1, 1-\alpha/2) + t(n-1, 1-\beta)]$$

n : number of replications
 δ : difference to be detected for the variable studied
 σ : residual standard deviation of the variable studied
 α : test level = first degree risk
 β : second degree risk for the variable studied
 t : Student t value; the parameters in brackets indicate the number of degrees of freedom and the threshold respectively.

Diagrams exist which make it possible to show this relationship graphically (1).

It should be noted that, contrary to the number of replications, which is an experimental constant, the parameters δ , σ and β depend on the variable studied. The level, α , is chosen by the experimenter and is not necessarily the same for all the variables.

This formula makes it possible to calculate one of the 5 parameters, when the other 4 are known. It shows in particular that the power increases with the difference δ to be detected and the number of replications n , whereas it falls as the standard deviation σ increases. The experimenter can therefore increase the power of the experiment in 2 ways: increase n , the number of experiments and reduce σ , the residual standard deviation.

Application of the previous formula makes it possible to calculate the required number of replications when the experiment protocol is being drawn up and before the experiment is set up. Although this protocol is designed to compare 2 treatments, it provides a good approximation when this number is greater. Its

(1) C. DANIEL, F. BONNOT - Setting up experiments in oil palm or coconut plantations. *Oléagineux*, Vol. 42, No. 5, May 1987.

application requires prior knowledge of the residual standard deviation, σ , which can be estimated using the information obtained from neighbouring experiments.

2.2. Methods of reducing standard deviation σ

Standard deviation σ can be reduced in two ways:

- Before setting up the experiment, by choosing an experimental design making it possible to reduce block size, if necessary and choosing their location after field reconnaissance.
- After setting up the experiment, through strict adherence to the treatments and very strict data checks. Such checks should be made at all levels, from the field observations - especially production data - up to computer acquisition. It is important to emphasize the fact that a simple, well conducted experiment will give better results and enable more conclusions to be drawn than a sophisticated experiment with less reliable data.

Out of the experimental designs enabling block size to be reduced, the most widely used are:

- ♦ balanced incomplete blocks, particularly used for selection, and whose lattices are a special case.
- ♦ factorial designs with confounding, particularly used in agronomy.

The use of these designs instead of complete block designs always leads to a loss in degrees of freedom of the residual standard deviation and, in the event of confounding, to the impossibility of estimating certain higher interactions. Nonetheless, these drawbacks are negligible compared to the gain in accuracy obtained by reducing block size.

When randomly allocating treatments to plots in this type of design, it is very important not to reverse two treatments belonging to different blocks: if this is done, simple analysis of the experiment will no longer be possible.

2.3. Study in the case of a split-split-plot trial

A split-split-plot trial was set up by PPM in accordance with the following experiment protocol:

- Main factor : K at 2 levels K0 and K1
- First subdivision : P at 3 levels P0, P1, P2
- Second subdivision : N at 4 levels N0, N1, N2, N3
- 3 replications

The total number of plots is therefore $4 \times 3 \times 2 \times 3 = 72$.

The breakdown of the degrees of freedom is then as follows:

SOURCE OF VARIATION	DEGREES OF FREEDOM
Main plots	5
Replicates	2
K	1
Error 1	2
Sub Plots	12
P	2
P * K	2
Error 2	8
Sub sub plots	54
N	3
N * K	3
N * P	6
N * P * K	6
Error 3	36
TOTAL	71 (5+12+54)

An analysis of variance, carried out in a trial on the FFB variable with SAS software (annex III) reveals no significant effect. However, examination of the number of degrees of freedom for first and second degree errors (equal to 2 and 8 respectively) suggests that the experiment such as it is designed has very little chance of revealing any effect, even if it exists.

Power calculations confirm this hypothesis, as shown in the following table:

		Residual standard deviation observed	Calculated power
Main plots	(K)	103.29	5.3 %
Sub-plots	(P)	37.73	10.2 %
Sub-sub-plots	(N)	17.12	84.8 %

These calculations assume a level α of 5 % and a difference δ to be detected between treatments of 20 kg of FFB per tree, i.e. 10% of the overall mean. They indicate that a difference of 20 kg due to K rates could only be detected in 5.3% of cases and the same difference due to P rates in 10.2% of cases: the power for the factors K and P is therefore clearly inadequate. On the other hand, the power of the experiment for the N factor is satisfactory since a difference of 20 kg between the N levels would be revealed in 84.8% of cases.

This example shows the importance of the power study before setting up the experiment. In this specific case, the number of plots being equal, it would have been preferable to use a 4 x 3 x 2 factorial design with 3 replications. The residual standard deviation common to the 3 factors would have been estimated with 46 degrees of freedom. The respective powers on the 3 factors would then have been more balanced, with the power reduction for the N factor being compensated for by a power increase for P and K.

The following table indicates the breakdown of the degrees of freedom for such a factorial trial in the case of complete randomized blocks and in the case of 3 blocks of 12 plots per replication with confounding of one degree of freedom for the $K \times N$ interaction and of two degrees of freedom for the $K \times P \times N$ interaction.

SOURCE OF VARIATION	DEGREES OF FREEDOM	
	No confounding	Confounding (2 Blocks/Rep)
Replicates	2	2
Blocks within replicates	-	3 <---
K	3	3
P	2	2
N	1	1
$K \times P$	6	6
$K \times N$	3	2 (+1)
$P \times N$	2	2
$K \times P \times N$	6	4 (+2)
Error	46	46
Total	71	71

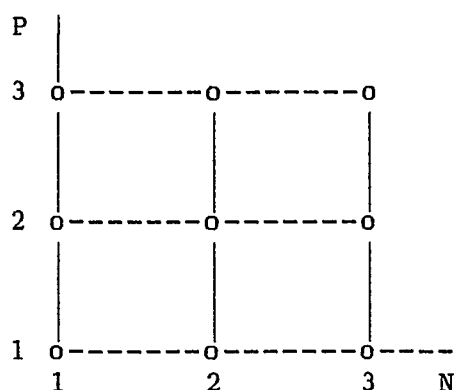
2.4. Response curves and surfaces

The value of an agronomic variable (yields, growth measurements, etc.) depending on fertilizer rates can be graphically represented by a curve called the response curve specific to the fertilizer. In the case of two fertilizers, graphic representation is a response surface. If forecasts are to be made, it is sometimes useful to model the response to one or more fertilizers using a simple equation making it possible to estimate the value of the response for any amount of fertilizer located within the range studied. One of the most widely used

models is the second degree polynomial function. The following table summarizes the equations corresponding to 1, 2 or 3 fertilizers, assumed to be N, P and K.

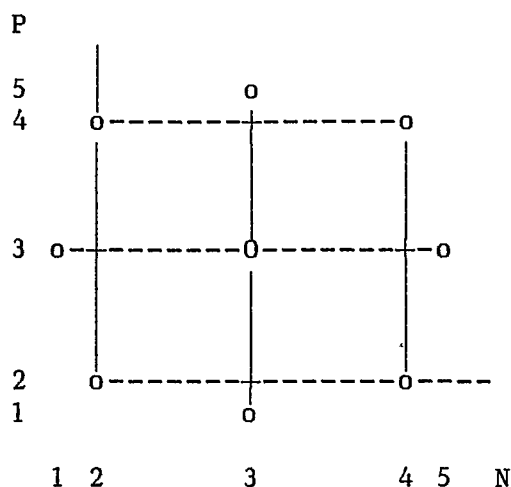
Number of fertilizers	Equation
1	$Y = a_{11} N^2 + a_1 N + a_0$
2	$Y = a_{11} N^2 + a_{22} P^2 + a_{12} NP + a_1 N + a_2 P + a_0$
3	$Y = a_{11} N^2 + a_{22} P^2 + a_{33} K^2 + a_{12} NP + a_{13} NK + a_{23} PK + a_1 N + a_2 P + a_3 K + a_0$

The parameters can be estimated by multiple linear regression from the results of agronomy experiments. In the case of 2 fertilizers, a conventional 3 x 3 factorial experiment could be used, for example, with the 9 treatments arranged as follows:



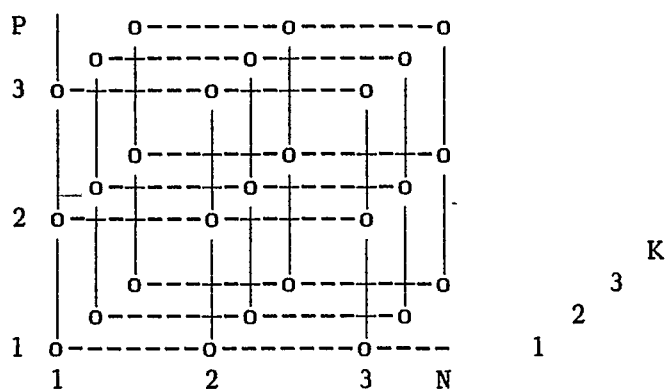
In this design, all the points are not at an equal distance from the centre and do not therefore play an identical role in determining the response surface. Furthermore, the central point, which nonetheless plays a determining role in estimating the curve, is no more represented than the others. The consequence of these two statements is that estimation accuracy is not the same over the entire surface.

In order to overcome this drawback, authors imagined designs giving all the peripheral points an identical role, through a circular rather than square layout, thereby favouring the central point, repeated several times (5 times in the example shown below):

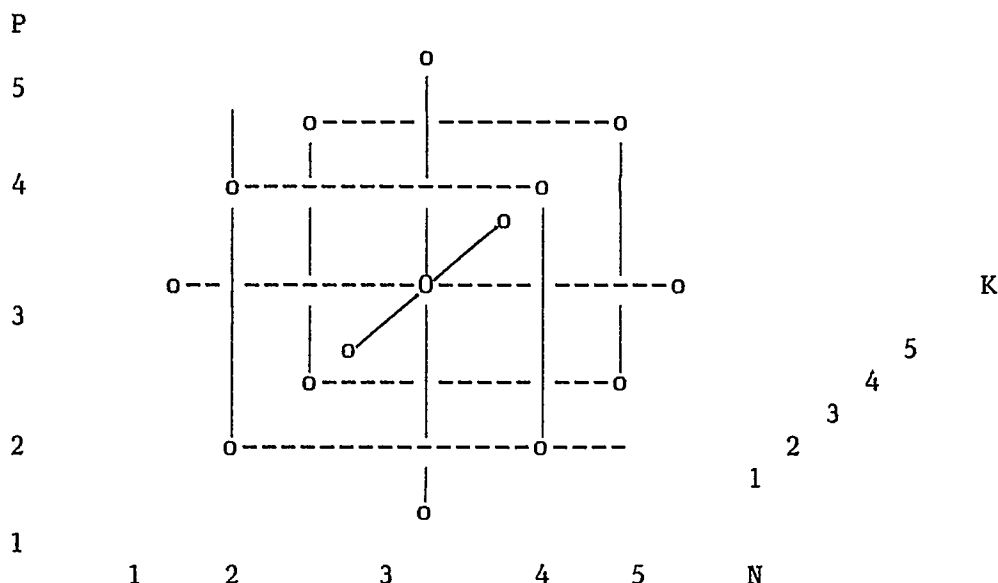


As can be seen, each factor exists at 5 levels, irregularly spaced out, rather than 3, even though the number of combinations tested (9 here) has not been increased compared to the 3 x 3 factorial experiment. This type of design is known in the literature as a central composite rotatable design.

In the case of 3 fertilizers, the treatment layout in a conventional 3 x 3 x 3 factorial experiment is as follows:



The corresponding central composite rotatable type design, sometimes called "octahedron design", contains 14 peripheral points and a central point repeated 6 times, i.e. 20 points in all:



What is to be thought of central composite rotatable designs? Their main advantage is that they provide identical accuracy at any point on the response surface. Nevertheless, it would be wrong to think that their accuracy is better overall than that of a factorial experiment; in the case of 3 factors, one replication of the factorial experiment consists of 27 plots and therefore supplies a greater amount of information than one replication in the octahedron design which only contains 20. Moreover, the latter design involves two not inconsiderable disadvantages:

- ♦ application of the treatments in the field is complicated since each factor is represented at 5 levels spaced out irregularly (which nonetheless provides for 15 combinations).
- ♦ the results can only be exploited through response curve calculation and not by conventional analysis of variance, whereas the factorial experiment can be analyzed using both techniques.

An experimental design making it possible to retain the simplicity of a conventional $3 \times 3 \times 3$ factorial design, whilst improving the accuracy of the response surface curvature estimate can be obtained by adding several replications of the central point to the basic 27-plot design.

For example, calculation of the response surface equation from the simulated data of a 3 x 3 x 3 factorial design with 27 plots was carried out using the RSREG function of the SAS software. The data and results are given in annex IV.

3. COMPUTERIZATION OF LEAF ANALYSIS AND FERTILIZER DATA

The leaf analysis method is used by the Marihat agronomists to monitor mineral nutrition on the commercial oil palm plantations and for recommending the fertilizers to be applied. At the moment, the data are stored on paper and calculations done manually. These operations will be greatly simplified by using the DF software developed by IRHO and designed specifically to manage and exploit this type of data on microcomputer.

The DF software makes it possible to enter, display, modify, store and print the following data:

- ♦ names and characteristics of the different level units making up the plantation,
- ♦ results of leaf analyses per LSU (Leaf Standard Unit),
- ♦ quantities of fertilizer applied per LSU,
- ♦ production data per planting block.

Automatic calculation of fertilizer recommendations can be done in accordance with the results of leaf analyses and LSU characteristics (age, planting material, soil type, cover crop). These calculations are based on tables drawn up by agronomists from the results of reference experiments, which can be introduced and altered at will. Recommendation of fertilizers containing phosphorus takes into account deficiency zones determined by the N/P balance and defined by a second degree equation whose parameters can be introduced into the software.

Several demonstrations were made of the DF software at Marihat. The Bah Jambi estate structure was entered using the software, along with the data from several LSUs and the fertilizer recommendation tables used for this estate. The automatic recommendations could then be easily drawn up by the software.

These demonstrations were made using the French version of the software. The software and technical literature are currently being translated into English.

4. MISCELLANEOUS

4.1. Organization of data processing at Marihat

There are currently 6 microcomputers in the computer room and 1 microcomputer each in the Agronomy and Breeding departments. A microcomputer is also given over to the vegetative propagation laboratory and another to the Finance department.

The different tasks need to be clearly divided up among the available machines.

Two microcomputers in the computer room, each equipped with a hard disk and a printer, are to be allocated to the acquisition and exploitation of experimental data per tree. The first will be assigned to breeding experiments, in view of their considerable volume. The second will be used for agronomy experiments, but may also be used for processing certain breeding experiments if needs be; in this case, we recommend separating the two types of experiments into 2 distinct directories. Whatever the division adopted, it is preferable for the data from the same experiment always to be entered and exploited on the same microcomputer. In order to avoid any keyboard errors likely to lead to data deletion, these microcomputers will be reserved for use by computer personnel and data processing operators for experimental data management and use of statistics software packages (SAS, etc.).

Free access can be given to the other microcomputers in the computer room, depending on needs.

The microcomputers installed in the Breeding and Agronomy departments will not be used for data acquisition, nor for routine analyses, but will be used for carrying out certain specific studies and may, accordingly, receive a copy of the data entered on the microcomputers in the computer room.

Service operations such as secretarial work, word processing and exploitation of administrative data should be clearly separated off from research activities, e.g. by installing dedicated microcomputers for this purpose in a secretariat independent of the research and experimentation services.

4.2. Software technical manuals

Technical manuals are an essential tool for ensuring correct use of any software. Numerous software packages are available at Marihat; unfortunately the corresponding manuals are often missing. Full sets of technical manuals for the most widely used software packages should be procured.

4.3. SAS Software

The Marihat station was temporarily able to obtain the SAS software (Statistical Analysis Software), which was installed in one of the microcomputers up to the end of September. It would be worth regularly renewing the licence for its utilization. Indeed, this software is a powerful statistical data processing tool which was successfully tested at Marihat on the following calculations:

- ♦ analysis of variance of specific balanced designs, such as split-split-plot designs (ANOVA procedure)
- ♦ analysis of variance of non-balanced designs, or designs with missing data (GLM procedure)
- ♦ determination of a response surface using data from a factorial experiment (RSREG procedure).

4.4. Setting printer parameters

A small program called IMP, making it possible to set printer parameters to normal or condensed characters has been installed in several of the microcomputers. It should be called up whenever using a software requiring a 132 column printer (such as STATPALM) if a short carriage printer is connected to the computer.

— .

A N N E X E S

I. INTRODUCTION

A. Purpose of the software

The STATPALM software is designed to ensure total management of oil palm experimental data on microcomputer, along with their statistical analysis using conventional methods. It contains the modules required for the following operations:

- ♦ Management of trial protocols,
- ♦ Management of experimental plot data from trials,
- ♦ Management of tree data from trials, including monthly production data,
- ♦ Calculation of plot means from tree data, grouping them together over several campaigns,
- ♦ Statistical analysis of plot means.

B. Hardware required

The STATPALM software can operate on any microcomputer using the MS/DOS system, equipped with a 640 K central memory. A hard disk is required for the routine operations of an experimental network, but one-off work can be carried out on hardware equipped with a disk drive only.

C. Software composition

The software installed at Marihat includes a version for Selection and a version for Agronomy. The files making up each version are given in the following table:

Version for Selection	Version for Agronomy	Utility
SPSEL.EXE	SPAGRO.EXE	Data management module
BLOCX.EXE LATX.EXE	BLOCX.EXE	Statistical analysis programs

D. Installation

To install the software on hard disk, proceed as follows:

- a) create a directory (e.g. STATPALM) in the hard disk root;
- b) copy the files in the previous table corresponding to the version chosen into this directory.

If both versions are used on the same computer, it is wise to create 2 separate directories for Selection and Agronomy.

It is possible to use the data management module separately on diskette. Merely copy the SPSEL.EXE or SPAGRO.EXE file onto the diskette, depending on the version chosen.

E. Start-up

To initiate the software, proceed as follows:

- a) Access the unit and the directory containing the software,
- b) Enter the name of the data management program: SPSEL or SPAGRO depending on the version chosen.

The programs BLOCX.EXE and LATX.EXE are called up automatically by the data management module and should not be run separately.

F. Data files

Each experiment consists of several data files whose names are based on the following model:

XXXXNNN[M].PPP

XXXXNNN : 4 letters + 3 figures make up the name of the experiment,

[M] : the letter M if it is a monthly data file,

PPP : suffix of 3 figures indicating the type of file.

These files are created automatically by the software when the experiment in question is entered for the first time, or when later entries are made for monthly data files (whose names include the letter M).

The software also uses a small file called HDREXP containing the name of the last experiment to be exploited. The existence of this file in the directory is optional.

The software also uses a small file called HDREXP containing the name of the last experiment to be exploited. The existence of this file in the directory is optional.

G. Data back-up and restore

It is necessary to carry out several data back-ups frequently onto diskettes. To back up the data relative to experiment XXXXNNN located on disk C, proceed as follows:

- a) insert a diskette in the disk drive (e.g. A),
- b) enter the command:

copy c:XXXXNNN*.* a:

If there is a problem with the hard disk C, the backed up data can be restored by inserting the back-up diskette in disk drive A and entering the command:

copy a:XXXXNNN*.* c:

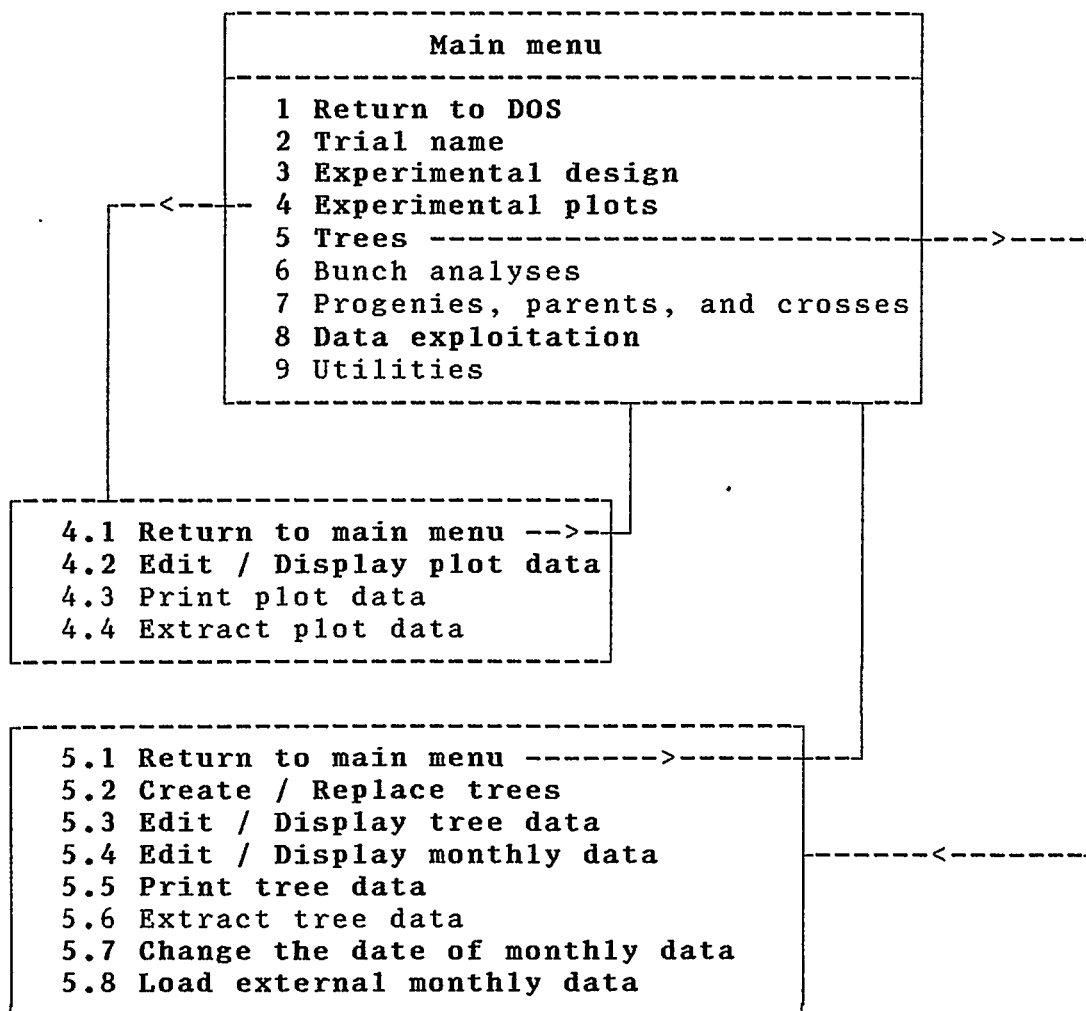
All the files concerning the same experiment are inseparable and must be copied at the same time. Using files originating from different software utilization sessions will lead to errors.

II SOFTWARE ORGANIZATION AND DETAILED DESCRIPTION OF FUNCTIONS

After initiating the software and entering the name of the experiment to be exploited, the main menu appears on the screen. This menu provides access to functions or sub-menus which can be called up using function keys F1 to F9. The sub-menus provide access to other functions or sub-menus. Each function or sub-menu is identified by one or more figures separated by full stops, which appear in the top left of the screen when the function or sub-menu is called up.

—

The general organization of the software can be schematized as in the figure below. The main menu and the main two sub-menus are shown in boxes. The function or sub-menu is preceded by a number. **Bold face** type indicates the functions or sub-menus that are operational in the current version. These functions are described in the following sections.



* Function 1 : Return to DOS

This function triggers a normal exit from the software and a return to MS/DOS. It is essential that it be used to ensure that files are closed properly; any other way of exiting the software (power cut, premature switching off or restart of the equipment, etc) is likely to lead to data loss.

* Function 2 : Trial name

This function is used to define or change the name of an experiment under way. It is called up automatically when the program is initialized and can be called up at any time thereafter from the main menu.

If the files belonging to the experiment are on the current disk, the software checks their coherence. If successful, the experiment indicated becomes the new experiment under way and the main menu is displayed on the screen. If unsuccessful, the files cannot be used and a message appears to indicate that they have to be restored.

If the files belonging to the experiment are not on the current disk, a message appears:

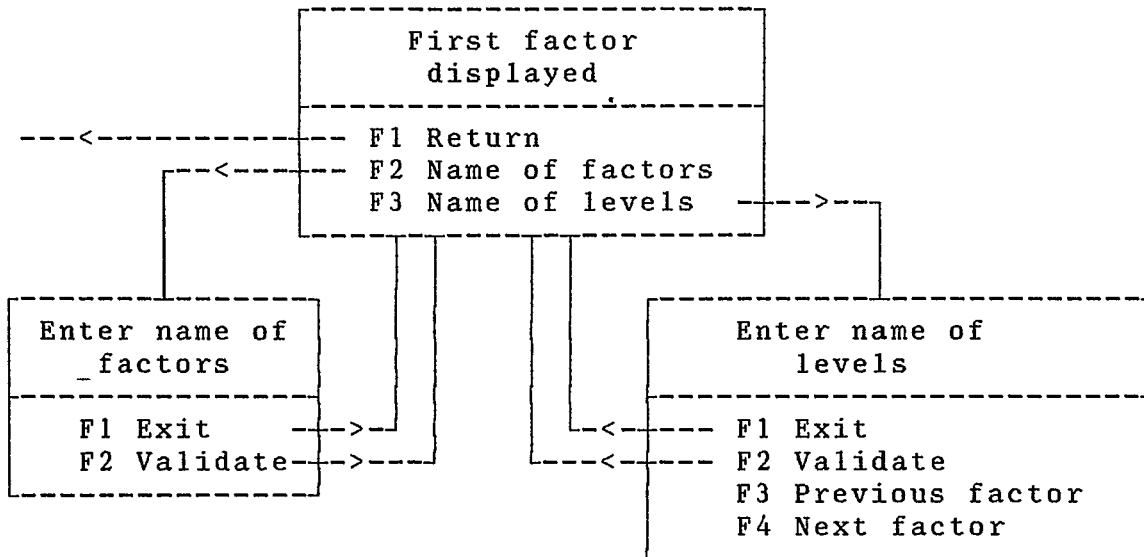
- ♦ if it is not wished to create the experiment files, key F1 is used to return to the main menu,
- ♦ if this is a new experiment to be catalogued, the following 3 fields have to be filled in:
 - **planting year and planting month:** the date the experiment was planted; if all the trees were not planted at the same time, the planting date of the oldest trees should be entered.
 - **month campaign began:** defines the start of each 12-month campaign; enter 1 (January) for Marihat.

Warning: No subsequent modifications can be made to these data.

After validation, the files are created and the experiment becomes the new experiment under way.

* Function 3 : Experimental design

The data concerning the first factor are displayed and the operations succeed each other as follows:



The software can take 5 factors into account. In experiments studying the growth factor, it should preferably be entered as factor 1, which plays a particular role:

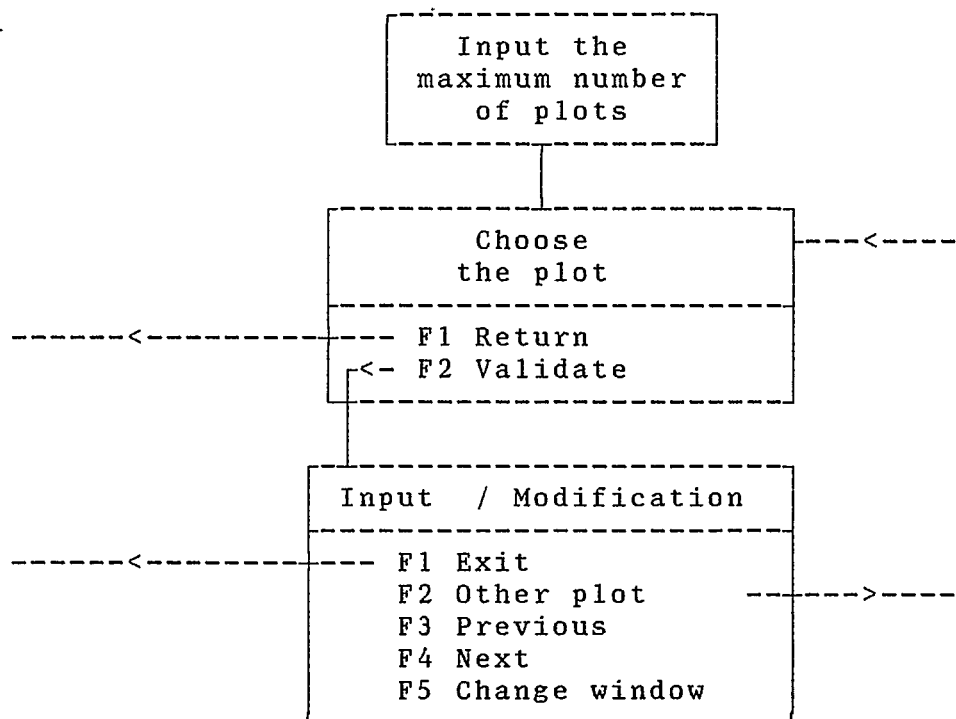
- ♦ the name of its levels can be up to 32 characters long, whereas that of the other levels is limited to 12 characters,
- ♦ the first column on the screen is linked to factor 1; it is intended to receive the names of the progenies of each of the levels of this factor; these progeny names will then be automatically carried over to the experimental plots in accordance with the level taken by factor 1 in each plot.

Note: depending on the design and the program used for statistical analysis, level 0 of the different factors should or should not be used; for complete balanced blocks, factor 1 should begin at level 1, with level 0 remaining unused.____

* Function 4.2 : Edit / Display plot data

This function is used to input the characteristics of each plot: replication, block, level of each factor, progeny. If the progenies corresponding to the levels of factor 1 have been entered in the protocol using function 3, they are automatically carried over into the plots in accordance with the level of factor one of each plot. The data observed in each plot appear on the screen. They can be inputted or modified, but, in principle, they are automatically calculated from data per tree.

Once the function has been entered, the operations succeed each other as follows:



The maximum number of plots in the experiment requested by the software at the beginning of the function serves only to limit plot number scrolling and can be changed at will.

The data input/modification screen appears as follows:

STATPALM Software

Trial

: BJ-SS 8

- 4.2 -		Edit / Display plot data						
Plot	:	1	Year	Nb Norm	Nb Bun	FPB	MW Bun	% PO TO/Ha
Replicate	:	1	1982					
Block	:	1	1983					
			1984					
Factor 1	:	17	1985					
Factor 2	:		1986	16	24.43	139.93	5.72	
Factor 3	:		1987					
Factor 4	:		1988					
Factor 5	:		1989					
			1990					
Progeny	:	BO 78/80	1991					
			1992					
Density	:		1993					
			1994					
Nb of Trees:		16	1995					

F1 EXIT F2 OTHER PLOT F3 PREVIOUS F4 NEXT F5 CHANGE WINDOW

Key F5 is used to move from the left-hand window to the right-hand window. The TAB and BACK TAB keys are used to move the cursor from column to column in the right-hand window.

* Function 5.2 : Create / Replace trees

This function is used to create or replace trees and to input their characteristics: planting year and month, experimental plot number, progeny, parent number, variety, state (producing or not). If the progeny number figures in the experimental plot to which the created or modified tree belongs, it is automatically carried over into the tree's corresponding field.

Tree creation or modification is by group of trees.

In the SPSEL program, a group of trees corresponds to a rectangular zone defined by the intersection of two strips, the first delimited by two row numbers, the second by two tree numbers. In practice, a zone will often correspond to an experimental plot, though this division is by no means obligatory.

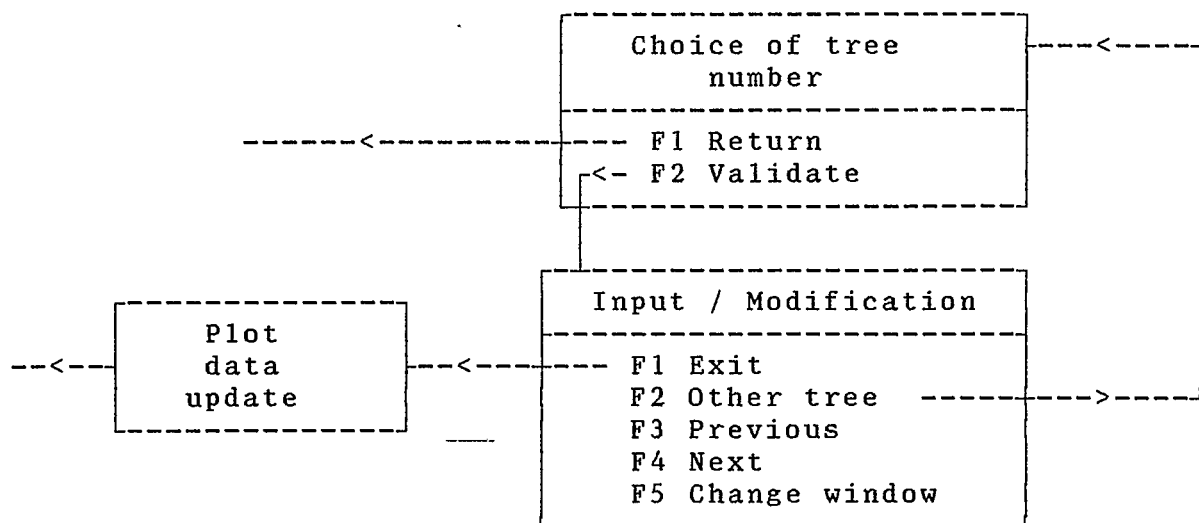
In the SPAGRO program, a group of trees corresponds to an experimental plot or a fragment of experimental plot. Each group is defined by the experimental plot number and the numbers in the plot of the first and last trees in the group.

Once a group of trees has been defined (rectangular zone, experimental plot or fraction of experimental plot) and the corresponding characteristics inputted, the trees in the group which did not exist are created and those which existed already are modified. The number of created and modified trees, along with the total number of existing trees, are displayed on the screen after each operation.

* Function 5.3 : Edit / Display tree data

This function is used to display and modify, tree by tree, the characteristics of the trees created by function 5.2. If the progeny number figures in the experimental plot to which a modified tree belongs, it is automatically carried over into the corresponding tree field. The data observed for each tree appear on the screen and can be inputted or modified. In principle, production data are calculated from the monthly data inputted row by row.

Once this function is triggered, the operations succeed each other as follows:



The data input / modification screen appears as follows:

STATPALM Software

Trial : BJ-SS 8

- 5.3 -		Edit / Display tree data		
Planting block :	21	* Nb Bun	FFB	1987 NB WB
Row :	48			
Tree :	1	1982		Jan 1 7
		1983		Feb
Planting year :	1982	1984		Mar
Planting month :	7	1985		Apr
		1986	22 118	May
Experimental_plot :	1	1987	1 7	Jun
Progeny :	BO 78/80	1988		Jul
Parent number :		1989		Aug
Variety :		1990		Sep
		1991		Oct
State :		1992		Nov
		1993		Dec

F1 EXIT F2 OTHER TREE F3 PREVIOUS F4 NEXT F5 CHANGE WINDOW

The right-hand window contains the monthly production data for the campaign under way.

Key F5 is used to move from one of the 3 windows to another. The TAB and BACK TAB keys are used to move the cursor from column to column in the two right-hand windows.

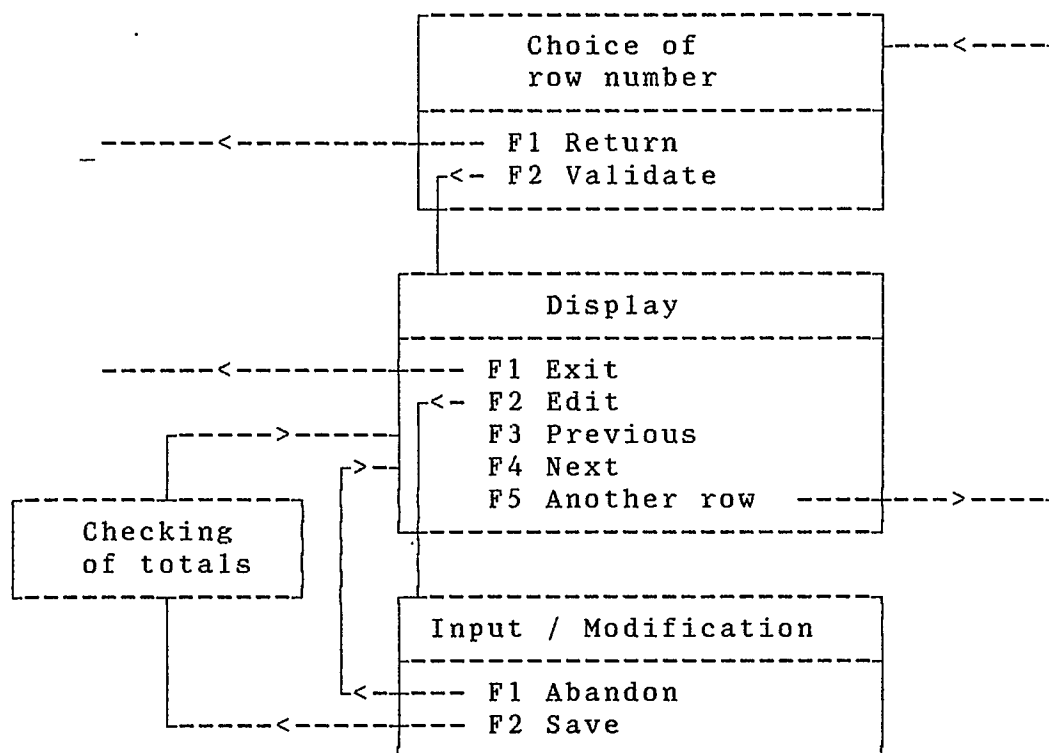
A tree can be temporarily eliminated for one or more years by placing the character "*" opposite the years in question in the first column of the central window.

Exiting the input / modification screen using key F1 causes plot data to be updated using annual data per tree. Only the periods not preceded by the character "*" are taken into account.

* Fonction 5.4 : Edit / Display monthly data

This function is used to input monthly production data per tree for the campaign under way. The data are inputted row by row into the SPSEL program and plot by plot into the SPAGRO program. When all the data of a row or a plot have been entered, the program asks for the sum of the number and bunches and of FFB which had to be calculated manually before. The data can only be recorded if the sums are correct.

Once this function has been triggered, the operations succeed each other as follows (case of SPSEL program):



The input / modification of data screen appears as follows:

STATPALM Software

Trial : BJ-SS 8

- 2,4 - Input monthly data											
Block: 21			Row : 48			Year: 1987			Month: 1		
Tree	N.Bun	FPB	Tree	N.Bun	FPB	Tree	N.Bun	FPB	Tree	N.Bun	FPB
1	1	7	13	1	4	25			37		
2	3	18	14	1	4	26	3	18	38		
3	3	18	15			27	3	21	39		
4	1	12	16	2	13	28	1	8	40		
5			17	2	19	29			41		
6			18	3	15	30	1	6	42		
7	2	10	19	2	15	31	1	5	43		
8	1	3	20	1	8	32	2	11	44		
9	1	7	21			33			45		
10	2	12	22	1	5	34	1	6	46		
11	2	14	23	3	22	35			47		
12	2	10	24			36	2	11	48		

F1 EXIT

F2 EDIT

F3 PREVIOUS

F4 NEXT

F5 CHANGE

The input table always shows 48 trees, but the numbers of existing trees are distinguished from the others by a different colour.

* Function 5.5 : Print tree data

For the time being, only tree characteristics can be printed. Once function 5.5 has been called up from the main menu, press key F2 twice to obtain the output, which will pass via the editor incorporated into the program.

* Fonction 5.7 : Change the date of monthly data

This function is used to define or change the year and month under way for production data. This date implicitly defines the campaign under way. The monthly data then used by the different functions (especially function 5.4) will be relative to the month or the campaign specified using function 5.7, until the date is changed again.

The monthly data for the campaign under way occupy a special place in the files, whereas the monthly data of the other campaigns are backed up in specific files. When a date change implies a campaign change, the monthly data of the old campaign are stored in their specific file and are replaced by the new campaign data, if they exist.

* Function 5.8 : Load external monthly data

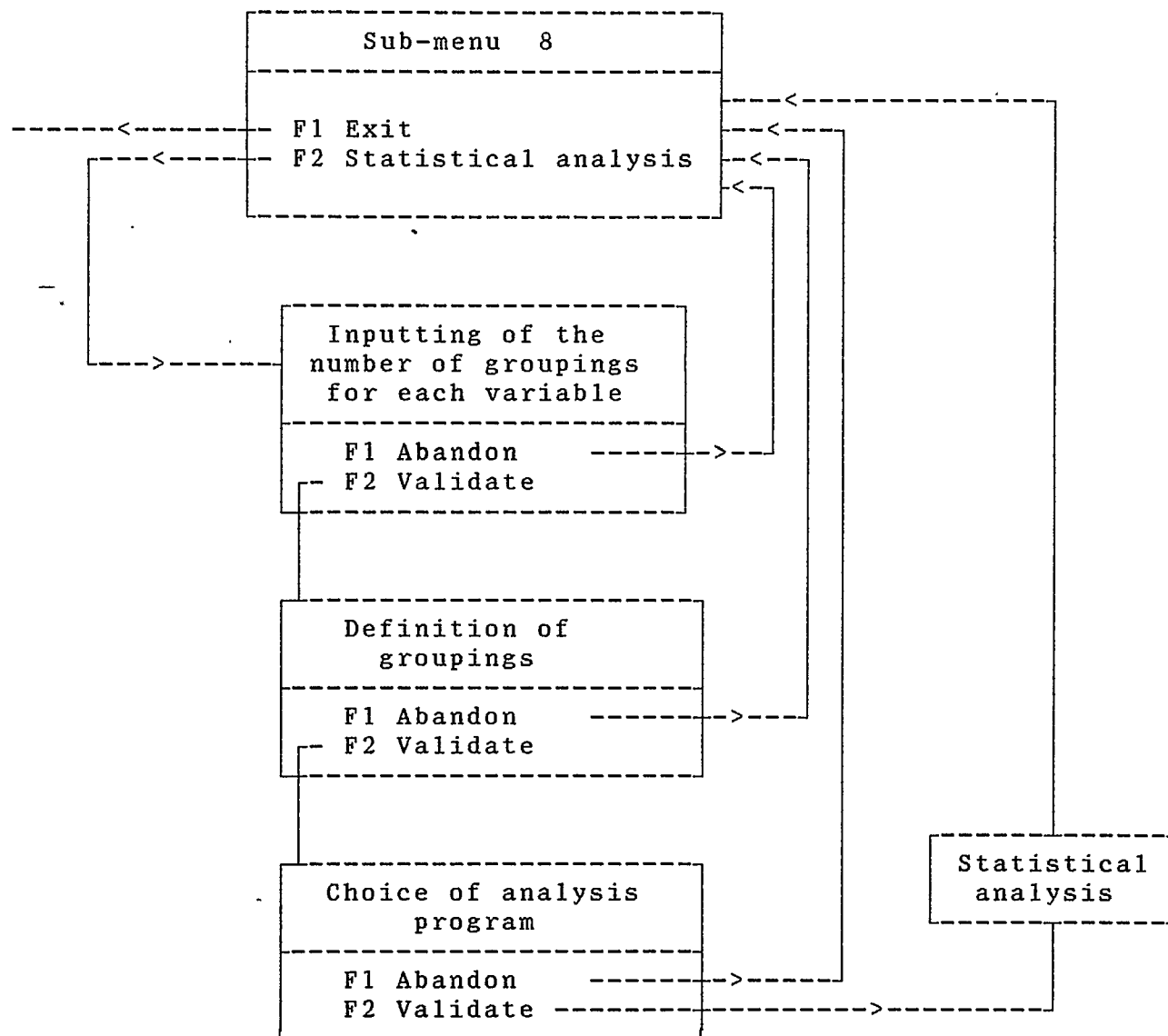
This function is used to load monthly data from a text type ASCII file containing the data per tree for a complete campaign. The file name is requested by the program once the function has been called up. The file format should be as follows:

Columns	Contents	
1- 2	Year campaign began	
3- 4	Month campaign began	
5- 6	Station number (optional)	
7- 9	Planting block	
10-11	Row number (SPSEL) or plot number (SPAGRO)	
12-13	Tree number	
14-15	Number of bunches	For the 1st month
16-18	FFB	of the campaign
69-70	Number of bunches	For the 12th month
71-73	FFB	of the campaign,
74-77	Total number of bunches	For the
78-82	Total FFB	campaign

The campaign indicated in the external file in columns 1-2 must correspond to the experimental campaign under way, defined using function 5.7.

* Function 8 : Data exploitation

Function 8 of the main menu displays a sub-menu in which only function 8.2 is currently available and enables statistical analysis to be undertaken.



The definition of groupings is carried out in a table displayed on the screen, based on the following model.

	1	1	1	1	1	2	2	2	2	2
	9	9	9	9	9	0	0	0	0	0
	7	8	8	9	9	0	0	1	1	2
	6	1	6	1	6	1	6	1	6	1
Nb Bun	++++									
FFB	++ +									

Merely use the "+" sign to specify for each variable all the campaigns for which the program is to calculate the means.

TREATMENT			REPLICATION 1	REPLICATION 2	REPLICATION 3	REPLICATION 4	REPLICATION 5	REPLICATION 6
1.	BO 10/80	BO 535D X BO 220P	24.180	18.120	19.060	22.000	21.310	19.000
2.	BO 5/80	BO 301D X BO 220P	23.500	24.750	21.810	23.060	24.500	21.250
3.	BO 14/80	BO 476D X BO 220P	22.120	24.060	21.560	24.250	18.680	22.310
4.	BO 20/80	BO 420D X BO 220P	25.560	25.750	25.620	26.500	22.310	25.620
5.	BO 50/80	BO 122D X BO 249P	16.430	12.180	19.370	15.620	16.370	12.560
6.	BO 53/80	BO 874D X BO 249P	26.500	28.870	25.310	21.930	25.310	19.180
7.	BO 45/80	BO 489D X BO 249P	19.680	21.060	22.680	22.620	17.750	18.060
8.	BO 29/80	BO 457D X BO 234P	25.810	29.180	20.870	26.870	24.680	24.250
9.	BO 34/80	BO 259D X BO 234P	29.120	25.560	28.370	29.560	26.930	30.870
10.	BO 84/80	BO 029D X BO 242P	21.370	16.370	13.750	18.120	19.370	17.870
11.	BO 81/80	BO 155D X BO 242P	24.620	27.430	22.930	23.120	25.500	19.750
12.	BO 85/80	BO 80D X BO 242P	31.620	32.930	31.930	28.120	30.060	32.180
13.	BO 87/80	BO 476D X BO 242P	26.370	22.370	25.000	23.810	27.310	24.370
14.	BO 91/80	BO 271D X BO 242P	26.500	25.500	23.560	25.500	26.430	28.000
15.	BO 102/80	BO 892D X BO 242P	22.430	22.870	25.060	32.370	24.500	25.810
16.	BO 70/80	BO 873D X BO 183P	24.620	17.560	22.310	23.930	21.060	18.620
17.	BO 78/80	BO 278D X BO 183P	24.430	23.870	22.680	22.680	24.060	23.000
18.	BO 80/80	BO 258D X BO 183P	20.750	21.310	22.060	22.000	23.560	22.060

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	D.F.	SUM OF SQUARES	MEAN SQUARE	COMPUTED F	F PROBABILITY
TOTAL	107	1824.3029			
REPLICATIONS	5	36.2664	7.2533	1.43	0.2196
TREATMENTS	17	1358.0053	79.8827	15.79 **	0.0000
ERROR	85	430.0313	5.0592		

C. V. 9.62 %

COMPLETE RANDOMIZED BLOCKS. Trial : BJ-SS 8
Variable : Number of bunches. Age : - 4 years.

TREATMENT	MEAN	HOMOGENEOUS GROUPS ACCORDING TO DUNCAN TEST (5%)
12. BO 85/80 BO 80D X BO 242P	31.140	!
9. BO 34/80 BO 259D X BO 234P	28.402	!
14. BO 91/80 BO 271D X BO 242P	25.915	! !
15. BO 102/80 BO 892D X BO 242P	25.507	!
8. BO 29/80 BO 457D X BO 234P	25.277	!
4. BO 20/80 BO 420D X BO 220P	25.227	!
13. BO 87/80 BO 476D X BO 242P	24.872	! !
6. BO 53/80 BO 874D X BO 249P	24.517	! !
11. BO 81/80 BO 155D X BO 242P	23.892	! ! !
17. BO 78/80 BO 278D X BO 183P	23.453	! ! ! !
2. BO 5/80 BO 301D X BO 220P	23.145	! ! ! ! !
3. BO 14/80 BO 476D X BO 220P	22.163	! ! ! !
18. BO 80/80 BO 258D X BO 183P	21.957	! ! ! !
16. BO 70/80 BO 873D X BO 183P	21.350	! ! !
1. BO 10/80 BO 535D X BO 220P	20.612	! !
7. BO 45/80 BO 489D X BO 249P	20.308	! !
10. BO 84/80 BO 029D X BO 242P	17.808	! ! !
5. BO 50/80 BO 122D X BO 249P	15.422	!
GENERAL MEAN	23.387	

COMPLETE RANDOMIZED BLOCKS. Trial : MA-PM 1
Variable : Number of bunches. Age : - 8 years- 9 years.

TREATMENT	REPLICATION 1	REPLICATION 2	REPLICATION 3
1. K0	10.150	12.245	12.995
2. K1	12.065	11.370	10.210
3. K2	11.995	13.055	12.715
4. K3	12.060	12.795	15.630
5. K4	12.685	12.865	9.560
6. K5	10.815	13.370	12.435

COMPLETE RANDOMIZED BLOCKS. Trial : MA-PM 1
Variable : Number of bunches. Age : - 8 years- 9 years.

ANALYSIS OF VARIANCE

SOURCE OF VARIATION	D.F.	SUM OF SQUARES	MEAN SQUARE	COMPUTED F	F PROBABILITY
TOTAL	17	33.6510			
REPLICATIONS	2	3.0033	1.5017	0.71	0.5167
TREATMENTS	5	9.6033	1.9207	0.91	0.5110
ERROR	10	21.0443	2.1044		

C. V. 11.92 %

COMPLETE RANDOMIZED BLOCKS. Trial : MA-PM 1
Variable : Number of bunches. Age : - 8 years- 9 years.

TREATMENT	MEAN	HOMOGENEOUS GROUPS ACCORDING TO DUNCAN TEST (5%)
4. K3	13.495	!
3. K2	12.588	!
6. K5	12.207	!
1. K0	11.797	!
5. K4	11.703	!
2. K1	11.215	!
GENERAL MEAN	12.168	

SAS 12:31 Wednesday, September 20, 1989

Analysis of Variance Procedure

Dependent Variable: FFB

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	35	66812.97492	1908.94214	6.52	0.0001
Error	36	10547.86518	292.99626		
Corrected Total	71	77360.84010			
R-Square		C.V.	Root MSE	FFB Mean	
0.863654		8.452911	17.11713	202.499861	

SAS 12:31 Wednesday, September 20, 1989

Analysis of Variance Procedure

Dependent Variable: FFB

Source	DF	Anova SS	Mean Square	F Value	Pr > F
REP	2	21948.37054	10974.18527	37.46	0.0001
K	1	2402.05357	2402.05357	8.20	0.0069
P	2	1939.99641	969.99821	3.31	0.0479
K*P	2	2626.25018	1313.12509	4.48	0.0183
N	3	2099.57984	699.85995	2.39	0.0849
K*N	3	782.78556	260.92852	0.89	0.4553
P*N	6	894.40703	149.06784	0.51	0.7976
K*P*N	6	1393.95971	232.32662	0.79	0.5815
REP*K	2	21338.09445	10669.04723	36.41	0.0001
REP*P(K)	8	11387.47763	1423.43470	4.86	0.0004

SAS 12:31 Wednesday, September 20, 1989

Analysis of Variance Procedure

Dependent Variable: FFB

Tests of Hypotheses using the Anova MS for REP*K as an error term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
REP	2	21948.37054	10974.18527	1.03	0.4930
K	1	2402.05357	2402.05357	0.23	0.6819

Tests of Hypotheses using the Anova MS for REP*P(K) as an error term

Source	DF	Anova SS	Mean Square	F Value	Pr > F
P	2	1939.996411	969.998206	0.68	0.5330
K*P	2	2626.250178	1313.125089	0.92	0.4360

SAS

7:41 Thursday, September 21, 1989

OBS	N	P	K	Y
1	0	0	0	101.805
2	0	0	1	126.920
3	0	0	2	148.397
4	0	1	0	116.917
5	0	1	1	141.238
6	0	1	2	153.376
7	0	2	0	132.514
8	0	2	1	146.913
9	0	2	2	155.406
10	1	0	0	109.032
11	1	0	1	130.262
12	1	0	2	146.750
13	1	1	0	123.685
14	1	1	1	138.196
15	1	1	2	148.256
16	1	2	0	132.204
17	1	2	1	143.341
18	1	2	2	146.699
19	2	0	0	114.650
20	2	0	1	133.433
21	2	0	2	145.306
22	2	1	0	127.425
23	2	1	1	136.584
24	2	1	2	143.614
25	2	2	0	130.942
26	2	2	1	136.052
27	2	2	2	136.954

SAS

7:41 Thursday, September 21, 198

Response Surface for Variable Y

Response Mean	135.069266
Root MSE	1.026335
R-Square	0.9962
Coef. of Variation	0.7599

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	3	3708.705157	0.7965	1173.6	0.0000
Quadratic	3	82.414592	0.0177	26.080	0.0000
Crossproduct	3	846.946361	0.1819	268.0	0.0000
Total Regress	9	4638.066110	0.9962	489.2	0.0000

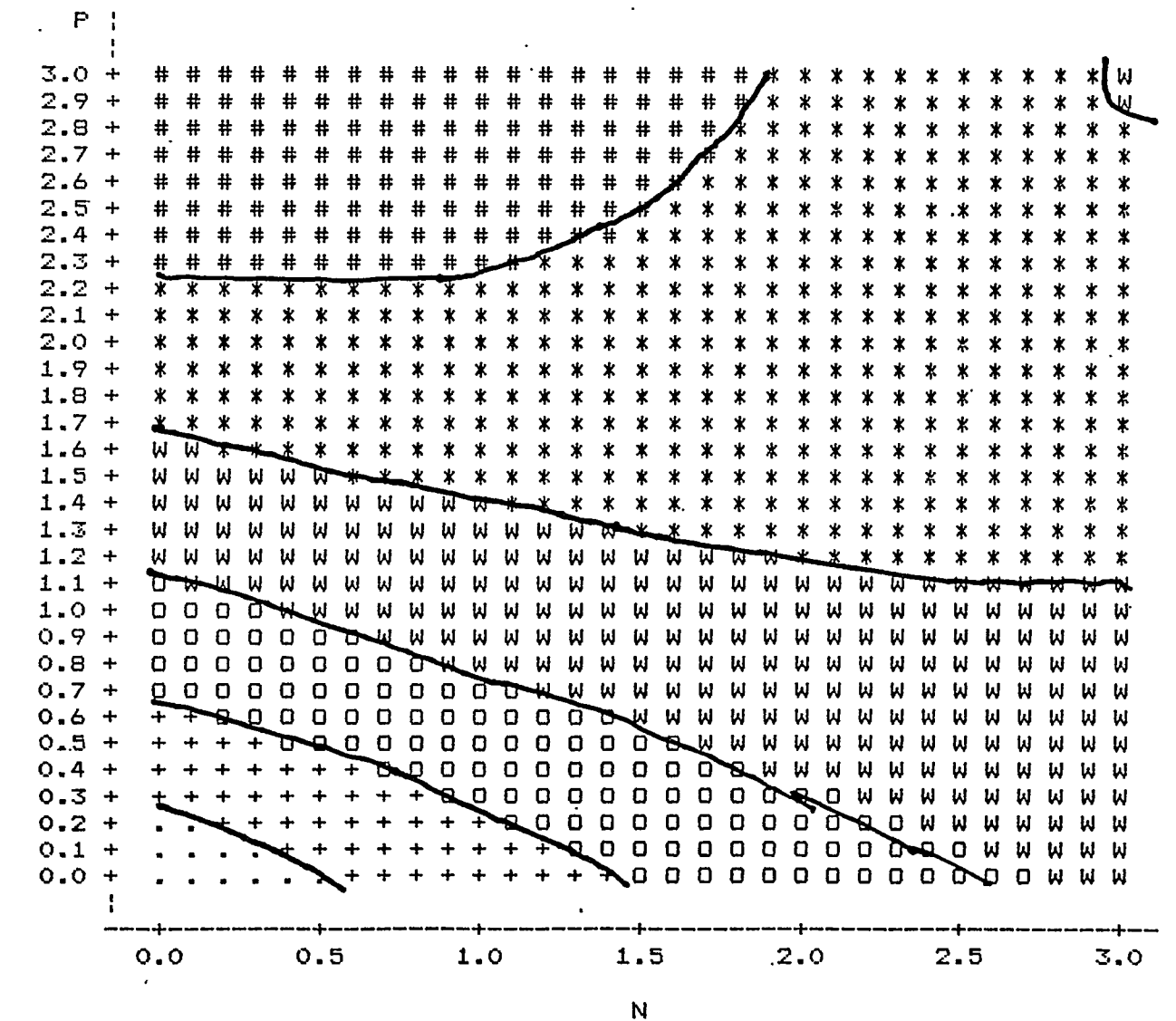
Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	17	17.907188	1.053364

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEPT	1	101.012442	0.732417	137.9	0.0000
N	1	8.258026	0.967638	8.534	0.0000
P	1	20.253087	0.967638	20.930	0.0000
K	1	29.247635	0.967638	30.226	0.0000
N*N	1	-0.466970	0.419000	-1.114	0.2806
P*N	1	-3.929364	0.296277	-13.262	0.0000
P*P	1	-2.277937	0.419000	-5.437	0.0000
K*N	1	-4.423901	0.296277	-14.932	0.0000
K*P	1	-5.963896	0.296277	-20.129	0.0000
K*K	1	-2.885950	0.419000	-6.888	0.0000

Parameter	Parameter Estimate from Coded Data
INTERCEPT	138.823171
N	-1.029179
P	5.803952
K	13.087937
N*N	-0.466970
P*N	-3.929364
P*P	-2.277937
K*N	-4.423901
K*P	-5.963896
K*K	-2.885950

SAS 7:52 Thursday, September 21, 198

Contour plot of P*N.



Symbol	Y	Symbol	Y	Symbol	Y
.....	97.5 - 105.0	00000	112.5 - 120.0	*****	127.5 - 135.0
+++++	105.0 - 112.5	WWWWW	120.0 - 127.5	#####	135.0 - 142.5