# Development of Data Handling and Management System Using Microcomputers for Tropical Soils and Crops

By HIDEO IMAI

Eco-Physiology Research Division, Tropical Agriculture Research Center
(Tsukuba, Ibaraki, 305 Japan)

## Introduction

In response to the increased requests by research institutions of the developing countries in the tropics for transferring production technology in agriculture, a great number of agricultural scientists have been sent from Japan to those countries for the last two decades. During this period, voluminous data concerning soils, crops and cultural practices in the tropics have been accumulated. Enormous volume of relevant data will continue to be piled up from now on. To utilize these data more efficiently and effectively, adequate systems for data management and analysis will have to be established urgently. Unfortunately, however, most of the Japanese agricultural scientists have paid little attention to microcomputer-based data analysis and data storage systems, while working mostly on the mainframe-based ones.

Until recently, it has been beyond the capability of the mainframe or minicomputers to analyze a large quantity of data in a short period of time and a sophisticated manner. However, the microcomputers which have been well developed in recent years are also sufficiently powerful for the management of large database as well as for the analyses on various simulation models. A number of software packages that previously were available only for a larger type of computers are now made available for microcomputers. In addition, a microcomputer is portable and data file diskettes can be easily distributed to collabolators worldwide. Thus, the recent advancement in information technology offers the hope that microcomputers could provide agricultural scientists with an efficient and user-friendly data base management system without a needed access to a mainframe or minicomputer.

A project is now being undertaken with an objective of developing simple systems for fast data-handling and -management with a use of microcomputers for research on tropical crops and soils. In this paper, a proposed data management system is discussed, including *inter alia* subjects relevant to data flow and related file conversion. In addition, the applicability of the proposed system is assessed to confirm to what extent it could adequately meet scientists' needs. The assessment is made on the basis of an example of nutrient uptake data taken in the project on "Developing improved cropping system for vegetable and legumes in the tropics"[4,8,9].

## Materials and method

System requirement: Minimum requirements in the use of hardwares for the proposed system are as follows:

(1) IBM PC/XT/AT or its compatible machine;
(2) Color or B/W monitor;
(3) 20 or 30 MB hard disk;
(4) Co-processor (i8087 or i80287);
(5) Floppy disk driver, 2 sets (2D or 2HD type); and
(6) Plotter, highly desirable.

Softwares tested for the system construction include the following items:

(1) Data entry and editing: Symphony[12], Lotus 1–2–3[5], Multiplan, VP-planner[13];
(2) Database: dBASE III Plus[3], R: Base 5000[6];
(3) Graphics: Statgraphics[11], Businessgraph[2], Dgraph; and
(4) Statistics: SPSS[10], BMDP[1], SAS[7].

All the softwares listed above were tested regarding their applicability to a new system for data handling and management. One software each for the above four types of work was finally chosen. Their functions are: (i) data entry and editing, (ii) drawing graphs, (iii) statistical analysis, and (iv) database management. The relevant four programs selected were combined into one package to transfer experimental data back and forth freely among them.

## Results and discussion

### *1) System construction*

A data flow in the proposed system for data handling and management is presented in Fig. 1. The transferability of data among the different programs is most important in the management, because the same data must be keyed in repeatedly whenever a different program is loaded in. In this system, original data once keyed in a spread sheet are transferred to the database or statistical programs in a minute. The original data of the experiment, first keyed in a Symphony spread sheet, are then transferred to other programs directly or after convertion to ASCII or DIF format using Symphony Utility. The converted data are edited by dBASE III by combining or breaking down files, sorting, picking up specified data, followed by statistical analysis and graphic presentation, as necessary. To run a BMDP statistic program, BMDP instructions which are written in language similar to FORTRAN, are required. In the system, therefore, the BMDP instructions are added to the ASCII-converted data and edited, where necessary, using such a text editor as WORDSTAR.

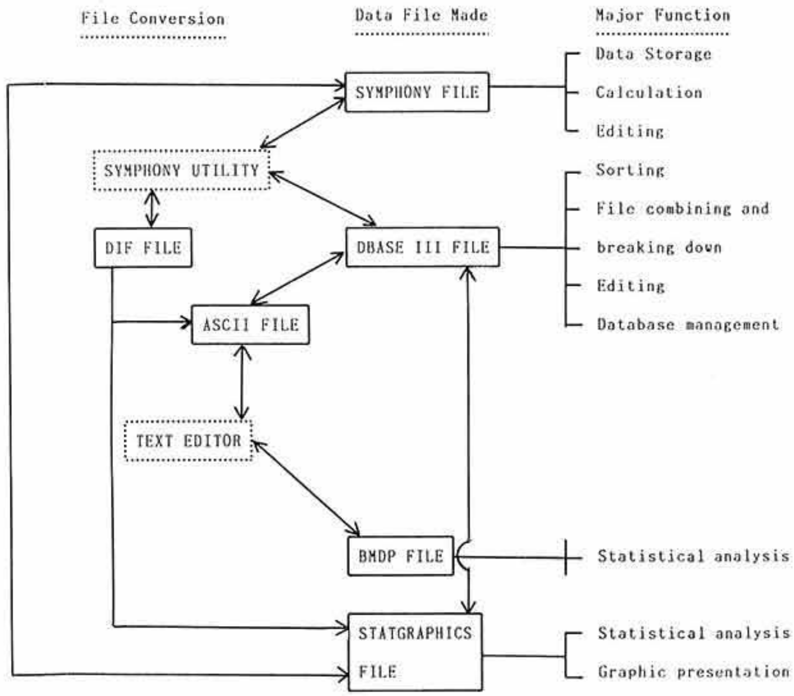


Fig. 1. A data flow under the proposed system for data handling and management

### 2) *Analysis and management of experimental data under the proposed system*

First, the original data on nutrient uptake collected from the experiment were stored in Symphony spread sheets under a specified file name (e.g., SP85UPT, which implies the data file concerning nutrient uptake by rotation crops planted in the spring season of 1985), as presented in Table 1. Based on the spread sheets, calculation and data editing are easily conducted. The edited data are transferred to other programs through ASCII, DIF or occasionally direct conversions. In the same way, data files of SU85UPT and AU85UPT were stored for the summer planting and the autumn planting in 1985, respectively. The three data sets from the three planting seasons were thus combined into one data file under a new file name of ROT85UPT. The management system of data files is summarized in Fig. 2.

The combined data file for each year was further assembled with other related files.

**Table 1. A part of data stored in Symphony spread sheet**

| No. | CROP[1)] | PARTS | DRYWT[2)] | NUPT[3)] | PUPT | KUPT | CAUPT | MGUPT |
|---|---|---|---|---|---|---|---|---|
| 1 | RC | ROOT | 1, 156 | 4. 5 | 0. 9 | 5. 3 | 0. 1 | 0. 6 |
| 1 | RC | STEM | 3, 808 | 24. 8 | 6. 0 | 59. 0 | 12. 6 | 5. 8 |
| 1 | RC | SEED | 4, 225 | 47. 9 | 12. 4 | 8. 7 | 0. 4 | 5. 2 |
| 1 | RC | TOTAL | 9, 162 | 77. 2 | 19. 3 | 72. 9 | 13. 2 | 11. 6 |
| 2 | MB | ROOT | 39 | 0. 6 | 0. 1 | 0. 3 | 0. 3 | 0. 2 |
| 2 | MB | STEM | 142 | 1. 8 | 0. 3 | 1. 9 | 1. 5 | 1. 0 |
| 2 | MB | LEAF | 143 | 3. 5 | 0. 3 | 1. 0 | 4. 9 | 1. 1 |
| 2 | MB | POD | 117 | 1. 0 | 0. 1 | 1. 5 | 0. 8 | 0. 8 |
| 2 | MB | SEED | 365 | 14. 8 | 1. 6 | 3. 1 | 0. 2 | 0. 6 |
| 2 | MB | TOTAL | 807 | 21. 7 | 2. 3 | 7. 9 | 7. 6 | 3. 7 |
| 3 | SB | ROOT | 321 | 2. 5 | 0. 4 | 0. 6 | 0. 8 | 0. 5 |
| 3 | SB | STEM | 1, 762 | 12. 5 | 4. 5 | 7. 2 | 32. 9 | 12. 5 |
| 3 | SB | LEAF | 1, 307 | 22. 2 | 2. 8 | 3. 8 | 49. 0 | 8. 7 |
| 3 | SB | POD | 1, 333 | 11. 6 | 1. 6 | 17. 3 | 9. 7 | 6. 6 |
| 3 | SB | SEED | 3, 850 | 248. 1 | 23. 6 | 41. 4 | 8. 5 | 10. 3 |
| 3 | SB | TOTAL | 8, 573 | 296. 8 | 32. 9 | 70. 3 | 100. 9 | 38. 5 |
| 4 | TT | ROOT | 87 | 1. 0 | 0. 1 | 0. 6 | 1. 5 | 0. 3 |
| 4 | TT | STEM | 1, 257 | 13. 6 | 1. 5 | 7. 6 | 40. 1 | 11. 2 |
| 4 | TT | LEAF | 1, 376 | 27. 4 | 3. 9 | 4. 9 | 76. 5 | 10. 2 |
| 4 | TT | FRUIT | 3, 687 | 80. 0 | 12. 2 | 61. 3 | 12. 6 | 5. 7 |
| 4 | TT | TOTAL | 6, 408 | 122. 0 | 17. 8 | 74. 5 | 130. 8 | 27. 3 |
| 5 | SB | ROOT | 374 | 3. 1 | 0. 6 | 1. 0 | 1. 2 | 0. 6 |
| 5 | SB | STEM | 2, 197 | 16. 5 | 5. 2 | 7. 5 | 37. 4 | 14. 0 |
| 5 | SB | LEAF | 1, 676 | 39. 2 | 2. 5 | 4. 6 | 66. 4 | 11. 0 |
| 5 | SB | POD | 1, 390 | 12. 3 | 1. 5 | 18. 1 | 9. 9 | 6. 9 |
| 5 | SB | SEED | 3, 725 | 238. 5 | 22. 5 | 40. 3 | 5. 2 | 9. 5 |
| 5 | SB | TOTAL | 9, 362 | 310. 0 | 32. 2 | 71. 4 | 120. 1 | 42. 0 |
| 6 | CC | ROOT | 43 | 1. 2 | 0. 3 | 0. 7 | 0. 2 | 0. 1 |
| 6 | CC | IL | 775 | 28. 6 | 4. 4 | 12. 8 | 2. 9 | 1. 2 |
| 6 | CC | OL | 1, 289 | 39. 2 | 5. 2 | 20. 2 | 36. 8 | 3. 7 |
| 6 | CC | TOTAL | 2, 107 | 69. 0 | 9. 9 | 33. 7 | 39. 9 | 4. 9 |

1) : Crop names are abbreviated as follows: RC=Rice; MB=Mungbean; SB=Soybean, TT=Tomato; CC=Chinese cabbage.

2) : DRYWT denotes Dry weight (kg/ha).

3) : NUPT denotes Nitrogen uptake (kg/ha), similarly, PUPT=Phosphate uptake; KUPT=Potassium uptake; CAUPT=Calcium uptake; MGUPT=Magnesium uptake.

The data were collected from the spring planting, 1985, of the experiments on "Developing Improved Cropping System", at AVRDC.

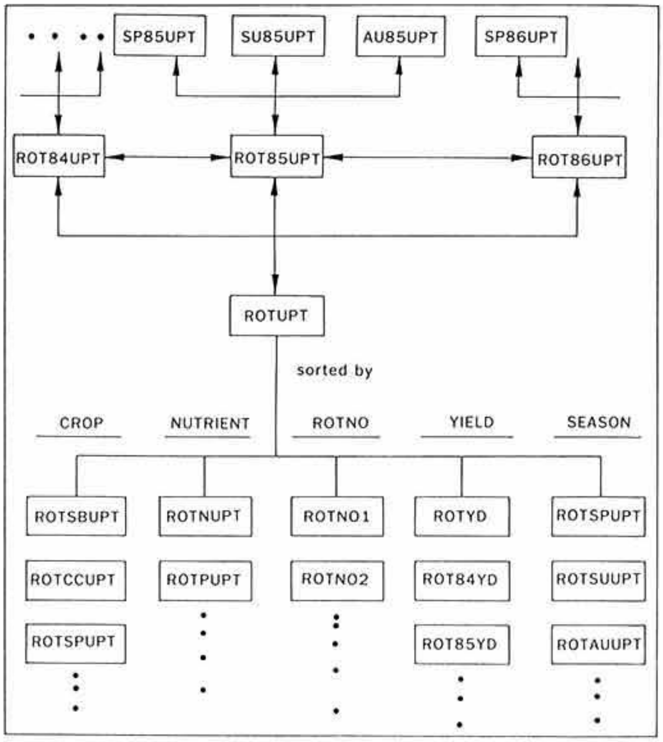


Fig. 2. File management system for nutrient uptake of rotated crops

A final file containing overall integrated data is stored under an adequate name for easy reference on the file contents. A term of ROTUPT, for example, implies the inclusion of all the data concerning nutrient uptake by rotated crops during the period of the experiment from 1983 to 1987. The ROTUPT data file can be easily sorted out by or broken down to crops, nutrients, cropping systems, yields, seasons and other relevant items. Therefore, once original data collected from each experiment are keyed in the spread sheets, various types of data files are easily produced by combining and/or breaking down the related files. Thus, a great quantity of information could emerge from these data files in a few minutes. On the basis of these data files, agronomic parameters such as amount of fertilizers to be applied to specified crops under different cropping sequences, nutrient balance in each experimental plot when plant residues are incorporated, and seasonal variations of nutrient uptake by crops could be easily estimated.

The combined or integrated data files are transferred to the Statgraphics or BMDP programs for statistical analysis and/or graphic presentation. In Fig. 3, the dry weight of each plant part for six different rotation crops over three seasons is plotted against the amounts of phosphate accumulated in the corresponding part (Data file: ROT85UPT). The dry weight increased linearly with increasing phosphate uptake, and all the experimental points except in soybean (indicated as No. 1) fell on a straight line. Regardless of the kind of crops increased phosphate uptake raised the dry weight of crops.

The P uptake by each plant part is plotted against N uptake in Fig. 4. The regression line of soybean veered off from those of other crops and it has a less steep slope than the others at the level below 220 kg of N fertilizer. Soybean, therefore, absorbed a larger amount of both N and P, having resulted in

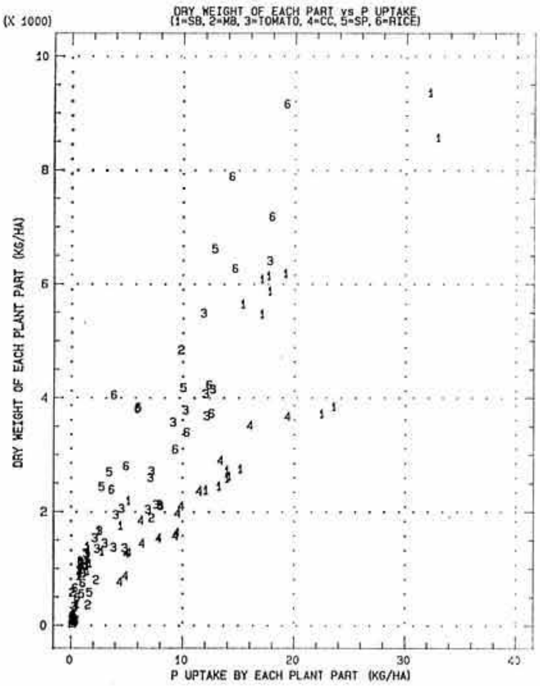


Fig. 3. Plotting of the dry weight of each plant part vs the amount of P accumulated in the corresponiding part
Data taken from data file: ROT85UPT.
Numerals in the graph indicate:
1=Soybean, 2=Mungbean, 3=Tomato, 4=Chinese cabbage, 5=Sweet potato, 6=Rice.

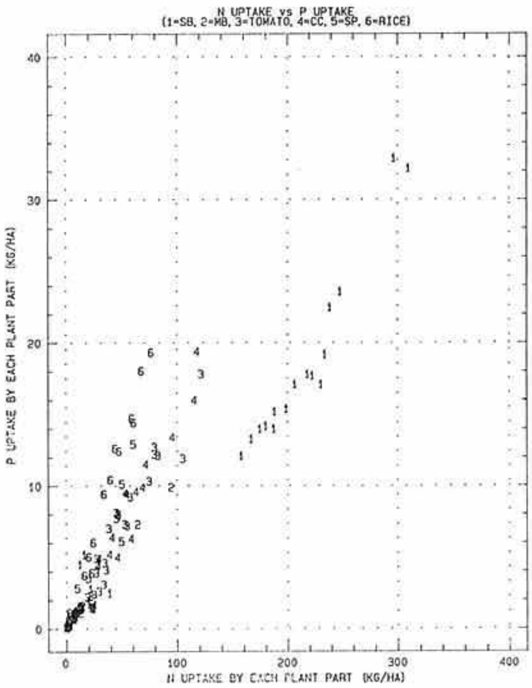


Fig. 4. Plotting of P accumulated in each plant part vs N accumulated in the correspon-ding part
Numerals in the graph indicate rotated crops.
Refer to Fig. 3.

significantly higher N/P uptake ratios, than the other crops studied up to the 220 kg/ha N-uptake level.

Another example of the figure presentation derived from the file of ROT85UPT relates to the data on reproductive organs (seed or fruit). The data on dry weight of these parts are plotted against the amount of P accumulated in the corresponding plant organ (Fig. 5). There are two distinctive lines; one line for tomato and rice, while the other for soybean. Soybean accumulates more P in its seed than the other crops. In the same manner, the data on dry weight of seed or fruit against N uptake were plotted (Fig. 6). The plots were categorized into three groups; the first for soybean, the second for mungbean and tomato, and the third for rice. The figure indicates that in producing 4 t/ha of dry seed, an amount of 250 kg/ha of N is accumulated in soybean, while only 50 kg/ha of N in rice.

In presenting summarized results in graphical displays, a Box-and-Whisker plotting method was adopted for N and P concentrations in seed or fruit of several crops (Figs. 7 and 8). The central box of each Box-and-Whisker covers the middle 50% of the data values between the lower and upper quartiles. The "Whiskers" extend out to the extremes (minimum and maximum values), while the central line is at the median. The N and P concentrations in the soybean seeds are extremely higher than those of other crops.

Finally, the factors affecting dry matter production of reproductive organs are subjected to a multiple regression analysis, the result of which is presented in Table 2.

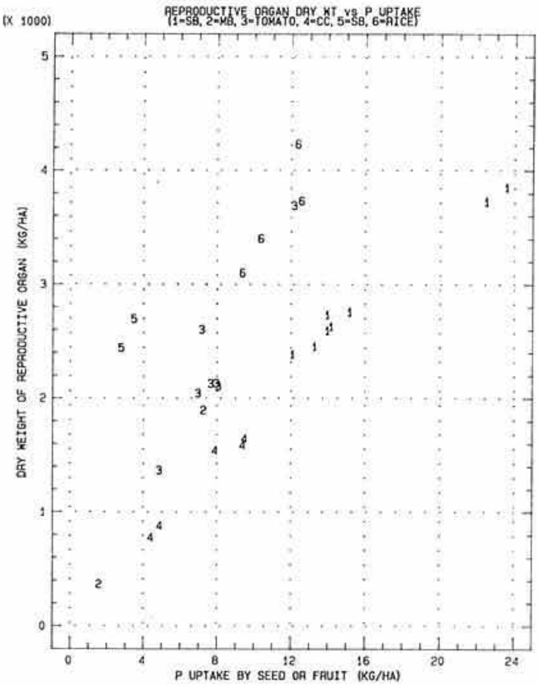


Fig. 5. Plotting of dry weight of reproductive organs of rotated crops vs P accumulated in the corresponding organs.
Numerals in the graph indicate rotated crops.
Refer to Fig. 3.

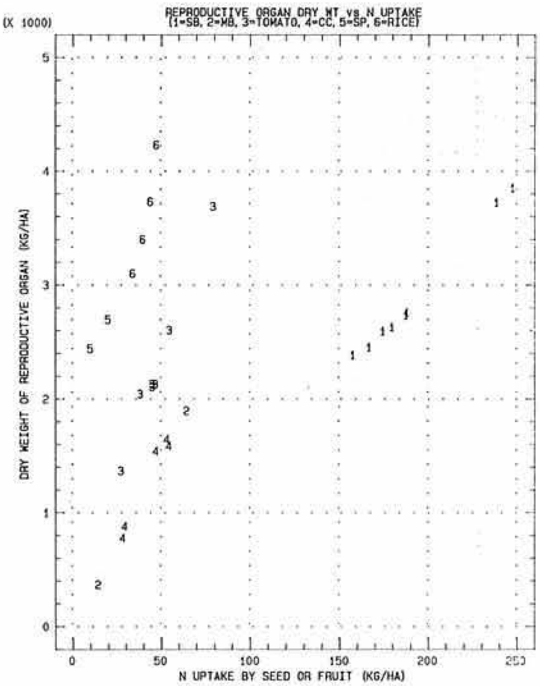


Fig. 6. Plotting of dry weight of reproductive organs of rotated crops vs N accumulated in the corresponding organs
Numerals in the graph indicate rotated crops.
Refer to Fig. 3.

Nearly 90% of its dry matter production is explained by three factors, i.e., P, N and Mg uptake.

In concluding, under the proposed system, original experimental data stored in Symphony spread sheets could be efficiently and effectively combined, broken down, reshaped and converted to other formats in a minute, in accordance with the specified purposes of analyses, through the proposed file conversion procedure (Fig. 2). The statistical analyses and graphic presentations could be completed in a few minutes through the specified data files. Statistical programs of the system cover almost all the statistical techniques necessary for data processing in agricultural research. They are expected to be sufficient for various types of analyses of the data collected from field experiments as well as from laboratory works. Database and graphic softwares contained in the system are also abundant enough and comparable to those avaialble for the mainframes.

The brief explanations above regarding the proposed system for data handling and management were based on the nutrient uptake data in 1985 sampled from the project on "Developing improved cropping system". Although only a part of the works which adopted the proposed system was demonstrated as above, it proved to be much more efficient than the existing integrated programs available in terms of speediness of data processing, flexibility and efficiency in statistical analysis. It may also be added that the data entry into Symphony spread sheets is 2 to 3 times as fast as other data entry formats such as DATA statement in BASIC, and less laborious. Correction of mistyped data, data conversion (e.g., X to log X, or Sin X etc.), insert or delete of variables and reshaping of data files can be easily made at any time as

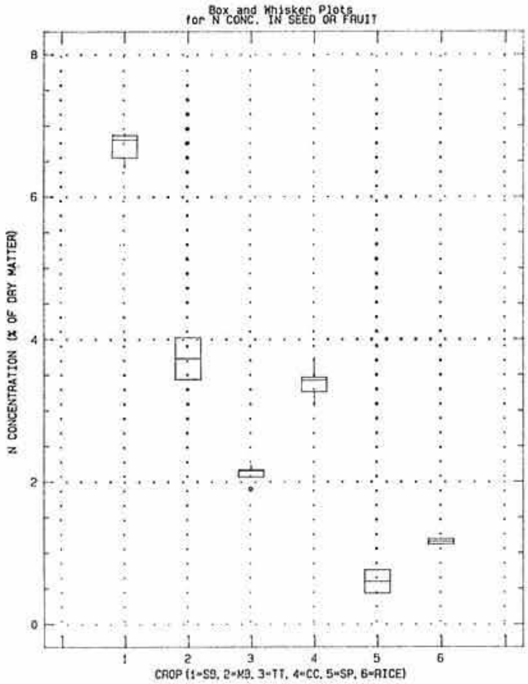


Fig. 7. Box-and-Whisker plot for N concentration in reproductive organs of rotated crops
Numerals on the X axis indicate rotated crops.
Refer to Fig. 3.

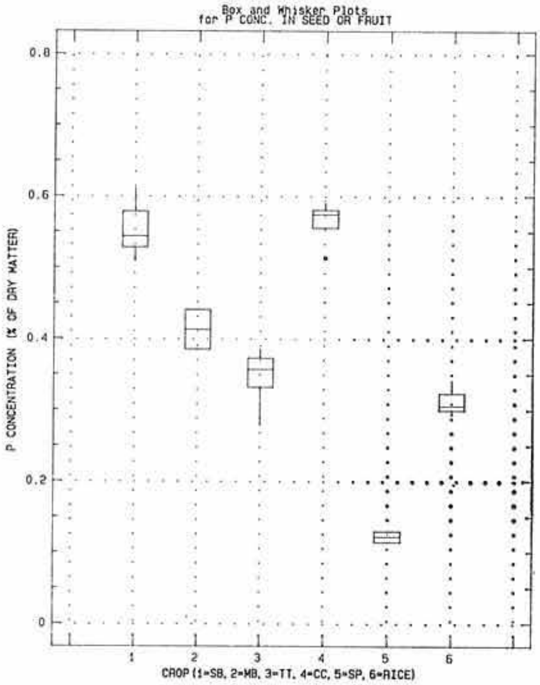


Fig. 8. Box-and-Whisker plot for P concentration in reproductive organs of rotated crops
Numerals on the X axis indicate rotated crops.
Refer to Fig. 3.

**Table 2. Multiple regression analysis for factor affecting dry matter production of reproductive organs**

| Stepwise regression | | | | | |
|---|---|---|---|---|---|
| Selection: Forward<br>F-to-enter=4.0 | | Max steps=50<br>Step 3 | | Control: Automatic<br>F-to-remove=4.0 | |
| R-Squared=0.902237<br>R-Squared (ADJ.)=0.894518 | | | MSE=91348.3 with 38 D.F. | | |
| Variables currently in model | | | Variables currently not in model | | |
| Variable | Coefficient | F-remove | Variable | Partial corr. | F-enter |
| 1. NUPT | −18.18943 | 131.1872 | 3. KUPT | −.0688 | .1762 |
| 2. PUPT | 246.68144 | 144.1315 | 4. CAUPT | −.1868 | 1.3379 |
| 5. MGUPT | 248.00882 | 25.0060 | | | |

Model fitting results

| Variable | Coefficient | Stnd. error | T-Value | Prob (>∼T∼) |
|---|---|---|---|---|
| Constant | 318.111464 | 138.405552 | 2.2984 | .0271 |
| NUPT | −18.189434 | 1.588084 | −11.4537 | .0000 |
| PUPT | 246.681443 | 20.547406 | 12.0055 | .0000 |
| MGUPT | 248.008817 | 49.595799 | 5.0006 | .0000 |

needed. Furthermore, original data collected from various locations, experiments and even from a wide range of countries are easily accessible in this system, providing that they are stored in the specified format for the system. Experimental data, thereby, could be exchanged directly among the collaborators worldwide by sending data file diskettes or communicating with other computers over telephone lines.

## Summary

A simple system for rapid data handling and management using microcomputers is proposed. The system covers the following three types of works: a) database management, b) statistical analysis, c) graphic presentation. Each work could be completed in a few minutes. Original experimental data keyed in Symphony spread sheets could be transferred directly or after convertion to ASCII or DIF format, to dBASE III for further editing and modification of data file structure which is suitable for various kinds of statistical analysis and graphic presentation.

The system proves to be much more effective than the other integrated programs presently available in its simplicity, speediness, flexibility, expansivity and efficiency in statistical analysis. Experimental data stored in the spread sheets could be made directly available by the collaborators by sending floppy diskettes or communicating with other computers over telephone lines.

## References

1) BMDP: Guide to using BMDP on the IBM PC. BMDP Statistical Software, Inc. Calif., U.S.A. (1985).
2) BPS business graphics for the IBM personal computer. Business & Professional Software, Inc., U.S.A. (1984).
3) dBASE III PLUS user's and reference manual. ASHTON-TATE, Inc., U.S.A. (1986).
4) Imai, H.: Data handling and management system for relevant informations on tropical crops, microbes, and soils. *Jpn. J. Trop. Agr.*, 32, 245–253 (1988) [In Japanese with English summary].
5) Kling, B.: The ABCs of LOTUS 1–2–3. Scott, Foresman and Co., Illinois, U.S.A. (1985).
6) RBASE 5000 user's and reference manual. Microrim, WA, U.S.A. (1985).
7) SAS procedure guide for PC, version 6 edition. SAS Institute, Inc., NC, U.S.A. (1985).
8) Soil Science: AVRDC progress report for 1985, 321–328 (1988).
9) Soil Science: AVRDC progress report for 1986, 375–385 (1989).
10) SPSS/PC+: SPSS for the IBM PC/XT/AT. SPSS, Inc., Chicago, U.S.A. (1986).
11) STATGRAPHICS user's guide and reference. PLUS*WARE Product. STSC, Inc., Maryland, U.S.A. (1988).
12) Symphony reference and user's manual (V.2.0). Lotus Development Corp., Calif., U.S.A. (1988).
13) VP-planner, software for the IBM PC. Paperback, Software International, Calif., U.S.A. (1985).

(Received for publication, Aug. 30, 1989)