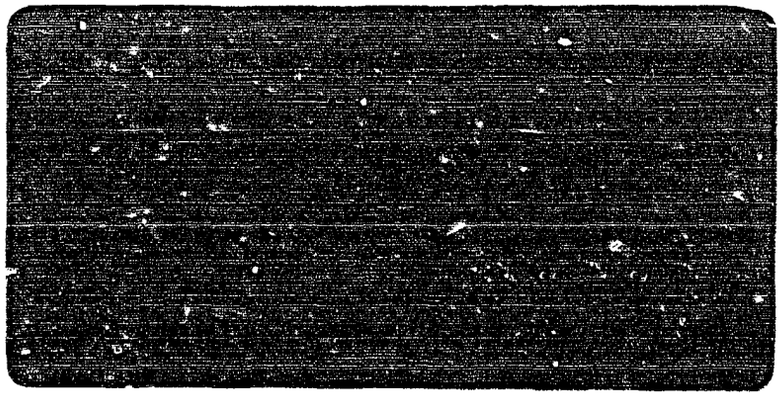


AGRICULTURAL DEVELOPMENT SUPPORT II
HAITI

University of Arkansas,
Fayetteville

GEOGRAPHICALLY REFERENCED CROP TRIAL RECORDS: A CASE STUDY

Report #35

Geographically Referenced Crop Trial Records: A Case Study

by

Douglas Brown
Michael Bertelsen
Amal Chatterjee*

The authors are, respectively, Geographer (University of Arkansas), National Survey Economist (Winrock International), and Agronomist (Winrock International) for the Haitian Agricultural Development Support Project II. This project is funded by the United States Agency for International Development under contract No. 521-0092.

#37

Geographically Referenced Crop Trial Records: A Case Study

Douglas Brown
Michael Bertelsen
Amal Chatterjee

ABSTRACT

Maps are extremely useful tools in analyzing phenomena that exhibit spatial variation. Unfortunately, making a map is a time consuming and technically difficult task. As a result, maps are much less widely used than they might be. The ADS-II project has forged linkages between software packages, that, to a large extent, automate the mapping process. Since many of the time and skill requirements have been eliminated by the system, mapping has become much more widely available as an analytic tool.

The heart of the system is a combination of the Lotus 123 spreadsheet program from Lotus Development Corporation and the geographic information system (GIS) from the Comprehensive Resource Inventory and Evaluation System (CRIES) of Michigan State University. A Lotus spreadsheet is used to establish a data base that is referenced to the spatial record keeping units on the map. The linkage mechanism allows data to be transferred between the two programs. The analytic and report facilities of either program are available for use with the data. The choice of the particular analysis facility used under any given set of circumstances depends on the nature of the data and the analysis to be performed. The real power of the combination comes from the ability to use the graphical output capabilities of the two systems. The time consuming tasks of constructing maps and graphs can be mostly relegated to the computer.

This paper describes the ADS-II integrated data management system in detail. A data base recently established to maintain rice trial records in a farming systems research zone of intervention in the Les Cayes Plain of Haiti is used to illustrate the utility of the system. We first introduce the crop trial data base. Discussion includes the establishment of the data base, training in its use, data entry, data analysis, and production of output products. After discussing the specific data base example, the paper moves on to a more generalized presentation of the organizational infrastructure that the ADS-II project is developing to improve the access that decision makers within the Haitian Ministry of Agriculture will have to the analytic capabilities of the system. The paper concludes with some speculation on the utility of similar data bases in other situations.

Introduction

Maps are an essential tool when analyzing phenomena which exhibit spatial variation. Since maps are time consuming and technically difficult to produce they are not as widely used as would be desirable. In an effort to reduce the time and skill required for map production the Agricultural Development Support (ADS-II) project of the

Haitian Ministry of Agriculture and Natural Resources (MARNDR) has developed an integrated data management system which, to a large extent, automates the map production process. Since the system is based on widely available IBM PC compatible hardware, we expect that this automated mapping capacity will be useful for a wide variety of projects within the MARNDR in the future.

This paper describes the integrated data management system and the organizational structure that ADS-II is developing to support the use of the system within the MARNDR. A recently established data base designed to analyze rice trials in a farming systems research zone of intervention in the Les Cayes plain of Haiti is used to illustrate the system. The paper begins by describing the analytic needs of the rice trials. These needs are used as a framework from which the structure of the data management system is derived. The paper then presents a sample analysis which demonstrates the capacities of the system. From the specific example of the rice crop trials the we move into a more general discussion of the organizational structure that ADS-II is developing to make the benefits of the system more widely available to decision makers in the MARNDR. The paper ends with some speculation about possible directions for further applications and development of the system.

The Rice Crop Trials Data Base

In many cases, the users of spatially referenced data will have a fairly well defined task that needs to be repeated at intervals, with few changes in operating procedure. Although the analytic needs of every project will be different, the rice crop trial data base is relatively uncomplicated and thus provides an excellent example of the application of essential elements of the data base system. The goal in establishing this data base was to record and analyze the performance of a specific variety of rice under different management and physical conditions. In addition to the numerical manipulation tools provided by a data base management program, the researcher conducting the trials is interested in having the capacity to produce two different types of map products. The first product is a map showing the owners of the different test plots. This map must be frequently updated as more plots are added to the experiment. The second product is a series of maps by test plot which represent the locations and outcomes of a wide variety of treatments. This second series of maps will be produced at least once each crop season.

The data base consists of five input variables: 1) ownership of test plots, 2) location of test plots, 3) area of test plots, 4) inputs to the test plots, and 5) yields from the test plots. The first three of these variables consist of single fields, while the last two will have multiple fields. The analytic requirements of the systems include the ability to: 1) add new test plots to the data base, 2) calculate relationships between the variables in the data base, and 3) produce tabular records, graphs, and maps of the data.

With the exception of producing maps, all of the data base functions listed above can be accomplished by all standard spreadsheet programs. We chose to use Lotus 123 (R) for this particular project. To supply the mapping capacity

we chose the Geographical Information System (GIS) developed by the Comprehensive Resource Inventory and Evaluation Systems (CRIES). This system was chosen because it is inexpensive and operates on widely available IBM PC compatible computers.

Two data structures are used in this system: 1) a spreadsheet for the numerical data, and 2) a GIS file to record the spatial locations of the data. A portion of the spreadsheet created to illustrate this paper is shown in table 1. In the interest of saving space, only the first 14 records of 160 are shown in the table. In the table, the spreadsheet is sorted in order of a code assigned to each farmer. This code appears in the first column of the table. Each of the test plots is also assigned a code, which is found in the column of the table labeled Parcel Number. This is the link between the spreadsheet and the GIS. This code is the number which is used in the GIS file to identify the location of the test plot, which is the basic unit of data collection and manipulation.

Table 1:

SEED MULTIPLICATION PROJECT
LES CAYES, HAITI
1987

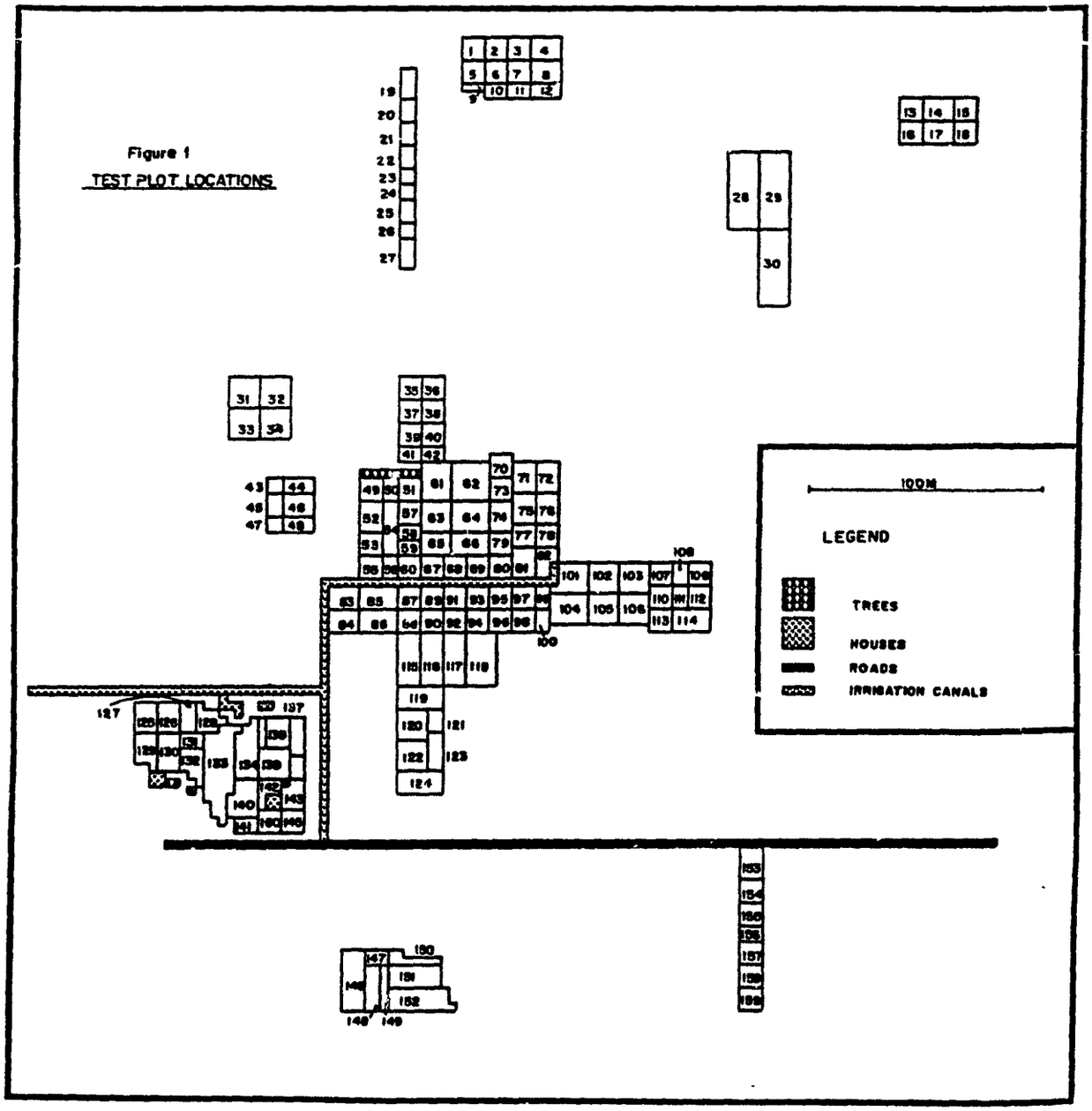
Farmer Number	Parcel Number	Area (Ha)	Total Production	Yield/ Hectare
1	125	0.0224	79.35	3542.34
1	126	0.0224	86.08	3842.86
1	127	0.0192	65.39	3405.78
1	128	0.0196	61.95	3160.77
1	129	0.0284	53.61	1887.65
1	130	0.0336	92.73	2759.87
2	83	0.0252	93.88	3725.46
2	84	0.0288	94.62	3285.25
2	85	0.028	102.05	3644.54
3	49	0.0192	32.27	1680.59
3	50	0.016	25.81	1613.43
3	52	0.0168	36.23	2156.27
4	51	0.0256	61.62	2407.09
4	57	0.0192	41.09	2140.22

The spatial locations of each of the test plots was digitized from a map of test plot locations compiled by the researcher. The GIS file created by this procedure can be updated to reflect future changes in the data base either by additional digitizing, or by a file editing facility in the GIS. Figure 1 is a map of the data in this GIS file.

The first of the requisite map products can be produced using the data in the first two columns of the spreadsheet in conjunction with the GIS test plots file. The procedure consists of creating a new GIS file in which each of the test plot codes in the GIS test plot file is transformed into codes representing the owners. In the example, since parcel numbers 125 through 130 belong to farmer 1, the codes 125 through 130 all need to be changed to 1 in the new file, codes 83 to 85 changed to 2, etc. This transformation is accomplished by using the GROUP function of the GIS. The inputs to this function are two files: 1) the original GIS file, and 2) the groupings file. The output is a new GIS

Figure 1

TEST PLOT LOCATIONS



file with areas assigned the new farm codes. In our example, the output GIS file will contain the ownership map. The groupings file is an ASCII file produced using the spreadsheet, consisting of three records formatted as follows:

<start of old range> <end of old range> <new code>

which are repeated for each new code to be placed in the output file. Table 2 shows the contents of the grouping file that was used to produce the farm ownership file from the data in Table 1. Standard spreadsheet data manipulations can easily be used to create files of this type from the original data.

Table 2:
Grouping File for Farm Ownership Map

```
125 130 1
 83  85 2
 49  52 3
 51  51 4
 57  57 4
  .
  .
  .
```

Once the new GIS file is created containing the codes assigned to each farmer, the final step is to print the data in this file using the standard GIS map printing function. The resulting map, with the title, scale and legend added by hand, is shown in figure 2.

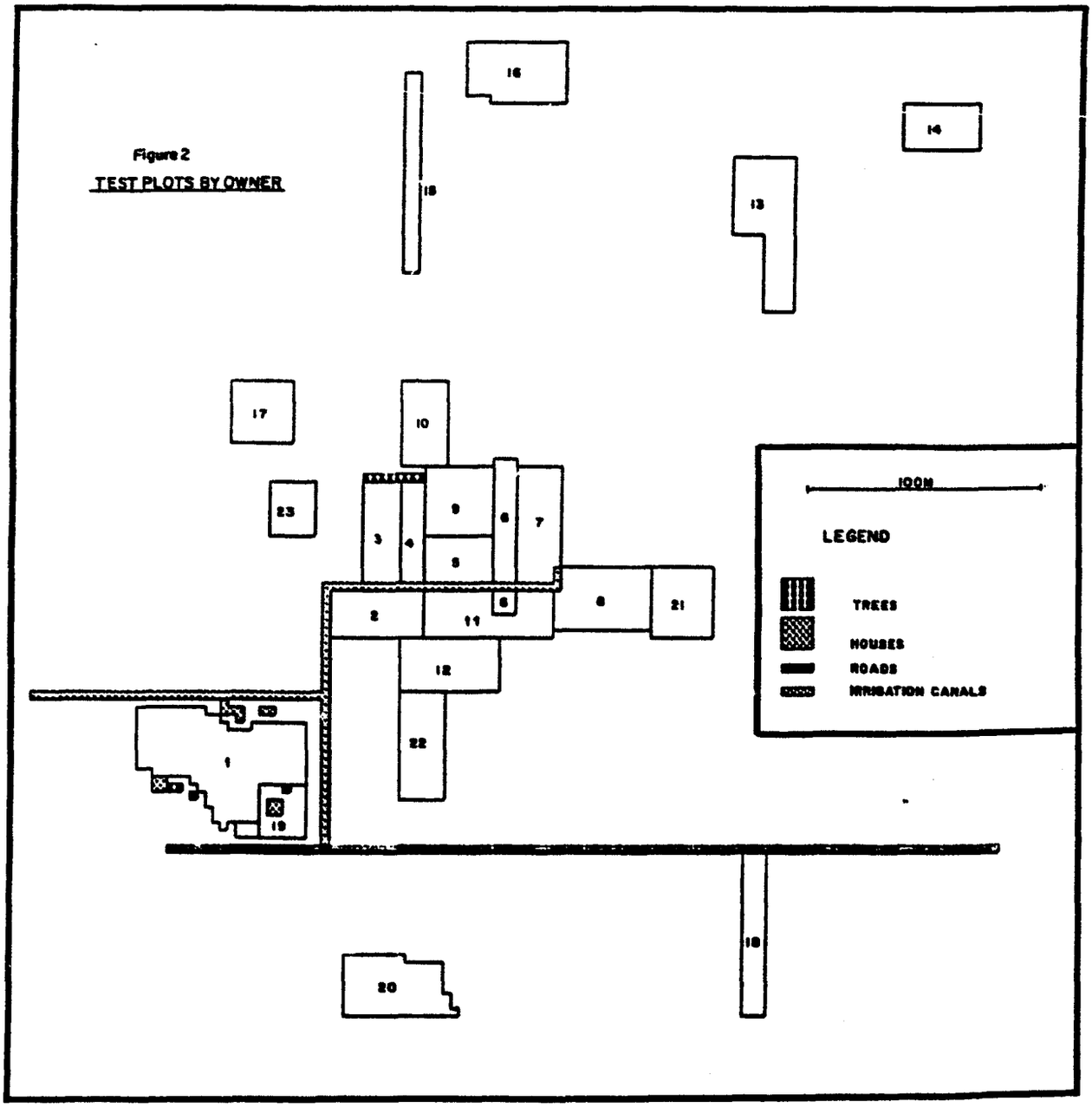
The third column in the spreadsheet (see table 1) is generated by a standard GIS function which reports the area of different features in the GIS file. This information is placed in an ASCII format output file which, after some modification with a word processor, can be imported directly into the spreadsheet. This information is needed to standardize yield measurements on a per hectare basis.

The process of producing the second type of map, maps of the outcomes of various treatments, makes more use of the analytic abilities of the spreadsheet, but otherwise, follows exactly the same procedure as mapping the test plot ownership data. In our example the column labeled total production (see table 1) represents data collected by the researcher and entered into the spreadsheet. Using the spreadsheet functions the raw production data was standardized into yield per hectare for purposes of comparison. The results of this transformation are located in the column of table 1 labeled YIELD/HECTARE and represent the type of data that must be mapped.

Since it is clearly impossible to intelligibly represent 608 different values on a map, the first step was to aggregate the individual yield values into ranges. Although any one of a number of different aggregation ranges could have been chosen, we chose to group our data into eight classes. These classes are shown in Table 3. It should be noted that one of the great benefits derived from

Figure 2

TEST PLOTS BY OWNER



automating the mapping process is that the researcher gains the option of testing several different aggregation schemes on a trial and error basis to find the one that best illustrates the information to be presented. Without automation, the time and effort involved in producing each map precludes this possibility.

Table 3:
Data Classes for Yield Map

Class	Yield (T./ha.)
1	< 1
2	1 - 1.5
3	1.5 - 2
4	2 - 2.5
5	2.5 - 3
6	3 - 3.5
7	3.5 - 4
8	> 4

Once the data are grouped into yield classes, the next step is to create a grouping file similar to the one used to produce the ownership map. In this case each of the codes representing test plots in the GIS file must be transformed into codes representing the data range into which the yield for that plot falls. These are the numbers in the column labeled Class in Table 3. In our example (see Table 1), the yield for plots 125 (3.54 T./ha) and 126 (3.84 T./ha.) fall into class number 7 (3.5 - 4 T./ha.) and should be recoded to a code 7 in the new GIS yield file. Table 4 shows the contents of the grouping file needed to produce a yield map of the data in Table 1. The format is exactly the same as the groupings file used to produce the ownership map. Once again this file can be easily created using the standard functions of the spreadsheet.

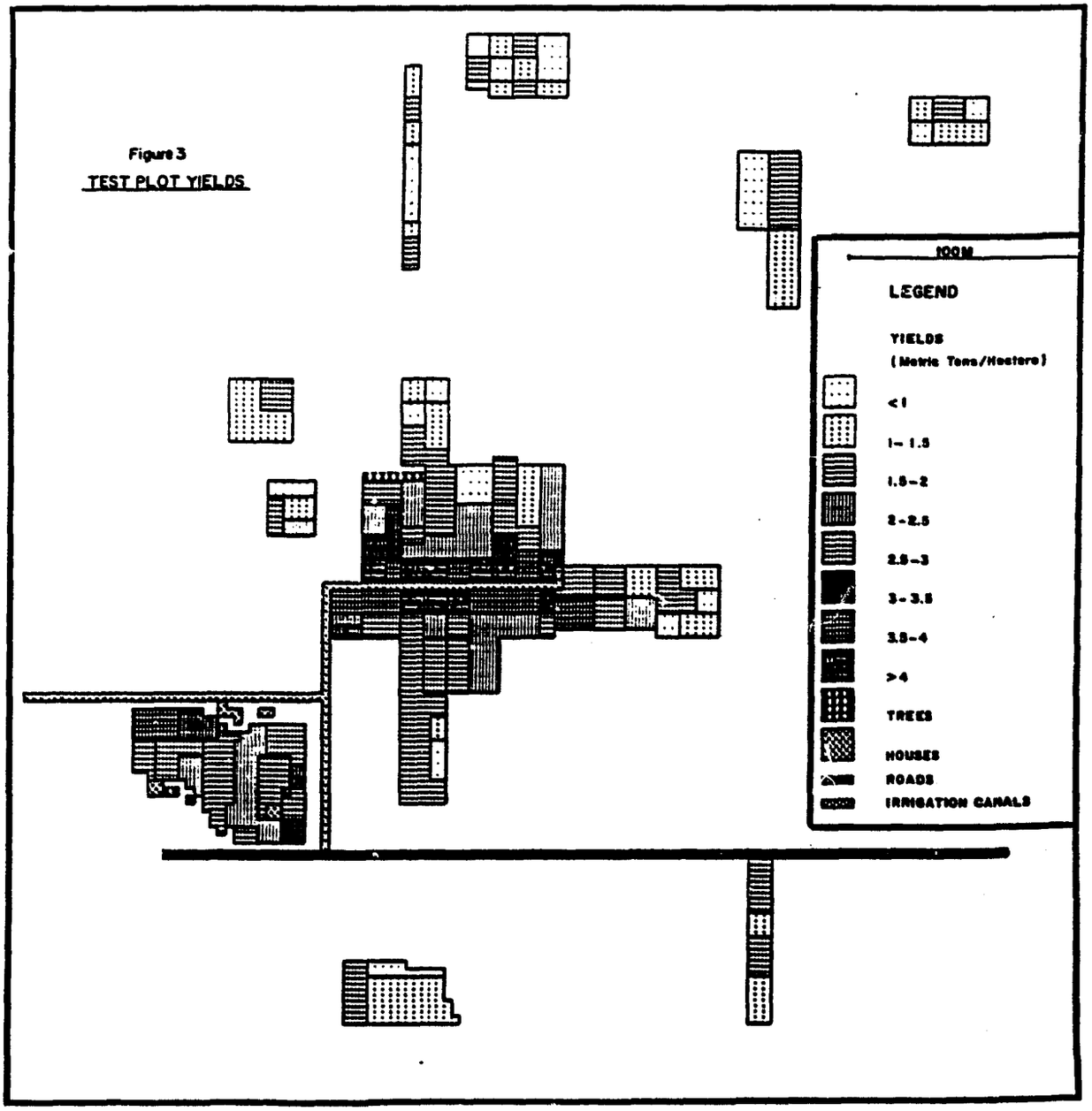
Table 4:
Grouping File for Yield Map

125	126	7
127	128	6
129	129	3
130	130	5
83	83	7
84	84	6
85	85	7
49	50	3
51	52	4
57	57	4
	.	
	.	
	.	

As in the ownership map, this file is used in conjunction with the GIS test plots file as input into the GROUP function of the GIS to create a new GIS file

Figure 3

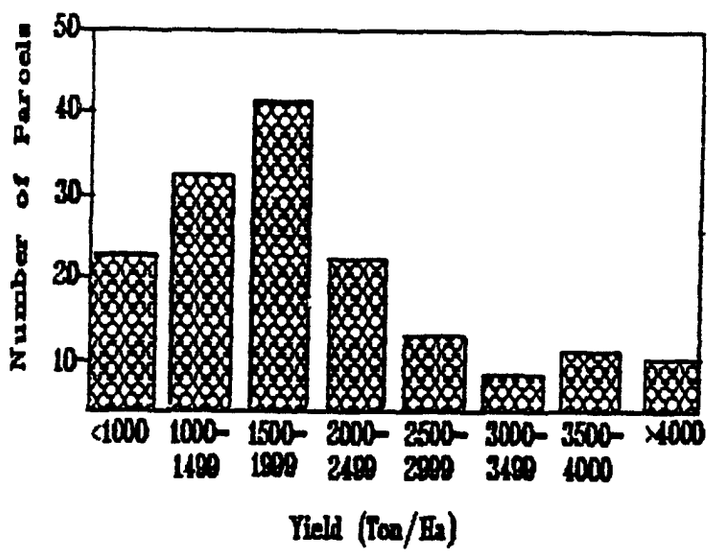
TEST PLOT YIELDS



containing yield information by location. All that remains is to print a map of this file using the standard GIS map printing function. The resulting yield map is shown in Figure 3.

The value of incorporating a mapping system into the data base management system is demonstrated by comparing Figure 4, a histogram of the yield data produced by the spreadsheet, with the yield map in figure 3. The histogram shows that the data is approximately normally distributed, with a slight skew towards the lower end of the curve. Nothing very revealing in that. It isn't until the same data is placed on a map that it becomes apparent that the higher yields are being obtained from the plots closest to the irrigation canal. If the researcher were dependant on manual mapping techniques, it is very unlikely that the data would ever be mapped because of the cost and difficulty involved. As a result, it is very probable that this relationship would be overlooked.

Figure 4: Yields of Parcels



Organizational Structure

The ADS-II project is in the process of establishing a GIS resource unit to make the analytic capacities of the GIS more widely available to potential users within MARNDR. The objective of this unit is not merely to provide GIS output products to users, but identify tasks in the Ministry which can have GIS functions usefully integrated into their overall information management systems, and create "satellite" GIS systems in the offices responsible for carrying out these tasks. The rice crop trials data base is this satellite GIS system concept. The GIS resource unit will provide the following services to the satellites :

1) assistance in establishing new data bases
2) training
3) a repository of seldom used hardware
4) research into new GIS applications and techniques
5) linkage and coordination of data exchange between satellites
6) management of national data bases

The necessity for these services is well illustrated by the rice trials data base. At the outset, integration of the GIS into the data base required consultation between the GIS resource unit and the researcher, an agronomist who is untrained in cartography. Once a data base is planned, it needs to be created. In any data base management project data base creation is inevitably the most costly and time consuming stage. A small project, like the one in our example, would have trouble justifying the high up front costs involved. The GIS resource unit provides experienced personnel and specialized hardware that can greatly reduce the cost and time required for this process.

The need for training is obvious. Our example, which is relatively uncomplicated, should require no more than 5 or 6 hours of training to perform the functions described above. More complicated data bases or use of more of the GIS analytic capacities would require increased training levels. Certainly, training must be provided at some level to establish new data bases, and will be one of the most important functions of the GIS resource unit.

All of the functions described in our example can be performed on an IBM PC compatible computer with an Epson FX series compatible printer. This configuration is very inexpensive and already widely available in offices throughout the MARNDR. Other functions of the GIS require more elaborate and expensive equipment, such as a digitizing tablet or a color ink jet printer. While the optimum hardware configuration for any particular data base will vary according to the needs of its users, it is essential to have a repository of equipment that users need only occasionally. For example, the rice trials data base could not possibly justify the purchase of a $5,000 to $10,000 digitizing tablet needed to create the data base. The GIS resource unit will be able to spread the costs of specialized hardware over a number of projects.

The research function of the GIS resource unit is critical. GIS represents an entirely new tool for managers of natural resources in third world settings. A great deal of research is needed to find the best ways to integrate this tool into the normal decision making process. This means that there remain not only unanswered technical questions about how to best present the data, but perhaps even more important questions about the optimum organization of the administrative infrastructure to insure the flow of information from the system to decision makers. Supplying the answers to these questions will continue to be an important function of the GIS resource unit.

One of the major advantages expected from the widespread use of the GIS is to increase the access of decision makers in all of the MARNDR departments to information collected and stored by other departments. The GIS resource unit will provide a central clearing house for access to all of the satellite GIS data bases. The unit

will not only assist decision makers in finding existing information collected and stored by the satellite GIS systems, but will also attempt to establish formal information channels to those data bases.

Finally, the satellite systems will maintain specialized data bases that meet the functions of the particular department the system serves. The GIS resource unit is needed to manage the general purpose national data base that has already been created. This data base will be essential to give a national context to all of the specialized satellite data bases. However to remain useful, it will have to be updated and enlarged as new data becomes available. Central storage of this national data base is also essential to insure that all of the satellite systems have access to it.

Conclusion

In its role of managing the agricultural and other renewable resources of Haiti, the MARNDR has an enormous need for spatial analysis tools. Digital geographical information systems provide a valuable addition to the range of spatial analysis tools available to decision makers that promises to not only increase the efficiency of existing mapping operations, but also expand the use of maps into new fields. The software and hardware components, as exemplified by the rice trials data base are already in place. The integration of the GIS into the decision making process of the MARNDR remains to be accomplished. This integration will be the primary role of the GIS resource unit.

At the time this is written, three satellite systems have been established: in the farming systems research project conducting the rice trials described above, in the National Agricultural Survey, and in a farming systems research project in Jacmel, Haiti. All of these projects are administered by ADS-II; the satellite systems have been experiments to learn the problems associated with their establishment. Work has recently started on establishing a satellite system in the Technical Secretariat For Watershed Management (STAB) of the Department of Natural Resources, which will be the first system outside ADS-II. We are investigating the possible future establishment of satellite systems in the National Water Resource Service (SNRE), the Targeted Watershed Management Program, and several other MARNDR agencies. We are also investigating the possibility of cooperating with several agencies outside MARNDR, such as the Pan American Development Fund, the National Malaria Eradication Service, and the Department of Public Works, all of which have needs for spatial analysis capabilities.

GIS is a powerful tool that potentially can greatly improve the quality of information and resulting quality of decisions. We believe that the GIS resource unit supporting the satellite systems concept will be successful in ensuring that the GIS will achieve its potential.