D A T A S E R V I C E S U N I T

C I A T

Role, users, functions, resources and products

(1985 - 1989)

CIAT, 13 of June, 1989

DATA SERVICES UNIT
PERSONNEL 1990

a) **Personnel budgeted to the Data Services Unit:**

POSITION	NAME	DEGREE
- Unit Head	Maria Cristina Amezquita (Acting)	Ms. and Dipl. in Mathematical Statistics
- Secretary	Maria Eugenia Echeverri	
- BIOMETRY SECTION		
. Head	Maria Cristina Amezquita	Ms. and Dipl. in Mathemathical Statistics
. Statistical Consultants	Eduardo Granados	Ms. Mathematical Statistics
	James A. García	Ms. Industrial Eng. and Systems
	Myriam Cristina Duque	Bs. Mathematics
	Germán Lema	Bs. Industrial Eng.
. Statistical Programmer:	Rosalba López	Programming Technology
- DATABASES GROUP		
. Analysts	Germán Serrano	Bs. Systems Eng.
	Fernando Rojas	Bs. Systems Eng.
	Carlos E. Erazo	Bs. Systems Eng.
. Programmers	Norbey Marín	Systems Technology
	Carlos Saa	Systems Technology
- IBM 4361 OPERATION		
. System's Programmer	Hugo Macías	Ms. Systems Eng.
. Operators/Transcriptors	Jairo Ramírez	
	Carlos López	
	Luz Mary Barona	
	Rita Maritza Escamilla	
	Amparo Rivadeneira	
	Fernando Arango	

b) **Personnel attached to other Programs but technically responsible to the Data Services Unit:**

- Tropical Pastures	Manuel Arturo Franco	Ms. Industrial Eng.
	Eloina Mesa	Ms. Statistics
	Gerardo Ramírez	Bs. Statistics and Mathematics
- Beans	Javier Crespo	Systems Technology
- Rice	Hector Fabio Ramírez	Bs. Statistics
- Agroecology	Hernán Trejos	Bs. Systems Eng.
	N.N.	
- TCSP	Marco A. Rodríguez	Bs. Civil Eng.

CONTENTS

D A T A S E R V I C E S U N I T,
C I A T

1. OVERALL VIEW: Role, functions, users, resources, products.

1.1 ROLE

The Data Services Unit (DSU) is a support-to-research Unit that reports to one of the two Deputy Directors. Its overall responsibility is to act as a central group for support and advise to CIAT research Programs/Units in aspects of mathematical, statistical and computing techniques that can be applied to the betterment of agricultural research. In particular, the Unit provides support in Biometry and data analysis, Information management with emphasis on a data base concept for information storage , retrieval and transference, and finally, the Unit is responsible for the provision and maintenance of appropriate computer hardware and software to serve the scientific programs needs, through the CIAT mainframe scientific computer: IBM 4361, with its 60 terminals and connected microcomputers. These three distinct support functions are fulfilled by the three existing groups in the Unit: a) Biometry group, composed by Statistical Consultants and Statistical Programmers; b) Information Management group, composed by Analysts and Programmers; and c) IBM 4361 hardware/software support group, composed by the System's Programmer, computer operators and data transcriptors.

1.2 FUNCTIONS

Biometry Group Functions

1. To provide conceptual advise on research design, use of mathematical/statistical methods and interpretation.
2. To provide statistical programming and data analysis support.
3. To participate with the scientists in collaborative "Methodological studies". These are carried out to evaluate in-use experimental procedures or techniques, to evaluate the effectiveness of a given strategy, or to provide answers to relevant questions of research, through the analysis of accumulated information.

Databases group function:
In a very close collaboration with the Programs and in response to their needs, the function of this group is to conceptualize, design, implement and maintain commodity-oriented and general-use databases, that will permit CIAT to keep and provide interactive access to historical records of research and environmental descriptors of its area of impact.

IBM 4361 hardware/software support group functions:

1. To provide and have maintained appropriate computer hardware to serve the IBM 4361 users needs, in consultation to and approval from the Electronic Data Processing Committee (EDP).
2. To provide and have maintained appropriate computer software to satisfy the Programs/Units needs, including operating system, compilers, packages, and application programs. Software - as hardware - acquisition is consulted and approved by the EDP Committee.
3. To provide adequate documentation on the various hardware and software products.
4. To provide a data transcription service to IBM 4361 users.

An overall-Unit function is to provide **training courses** in the use and interpretation of quantitative techniques for data analysis via mainframe and microcomputer software. The Unit provides about six 1-week- courses per year to NARDS researchers, participants in crop-oriented training courses that are annually organized by the TCSP. Also, internal training courses are provided to CIAT-research/associates/assistants from head quarters and from regional projects outside CIAT Palmira.

1.3 USERS

The DSU clients are of two kinds:
a) Scientists from CIAT HQ (research Programs/Units, Special projects, and projects from collaborating Institutions at CIAT) and sporadically from CIAT regional projects in Africa, Asia and Latinamerica.
b) CIAT-NARDS germplasm evaluation Networks.

The 1987 DSU user's survey shows that 77% of the total Senior scientists at CIAT-HQ are users of the biometry services, and 87% of them are users of the computing services offered by the Unit. Although these figures have remain practically the same up to now, the nature of the user's demand has changed into a more conceptual and methodological need and less routine data processing.
With respect to support provided to International or Regional Networks in which CIAT research programs collaborate, the DSU is responsible for data processing, statistical analysis, data storage and production of computer - generated reports to collaborators. Specific computer programs using the existing IBM 4361 software, have been developed by the Unit to satisfy each Network specific needs. This, under a very close discussion with the program scientists responsible.

Although not formally stated, the DSU does provide informal support to non-research areas of CIAT. For example, Stations Operations, Analytical Services Laboratory and Human Resources have received support on some of their needs: survey analysis, analysis of input trends, analysis of salary curves/management of personnel evaluation information, etc.

Support provided by the DSU to the TCSP (Training and Communications Support Program) includes: a) advise in the conceptualization, design and implementation of the "Trainees Information System", a SAS/FSP (SAS Full Screen Product) application. This permits interactive data storage/modification/ retrieval and production of computer-generated reports on information related to about 4000 trainees that have received training at CIAT from 1972 to 1989. b) Maintenance support to UNESCO's CDF/ISIS (Computerized Documentation Facility/ Integrated Set of Information Systems), a bibliographic database software developed by UNESCO and adapted by IICA to satisfy the CIAT library needs in the management of bibliographic information. c) Permanent support with training courses on the use and interpretation of quantitative methods via microcomputer software. These 1 or 2-weeks courses, are included as a component of the various CIAT crop-oriented training courses attended by NARDS professionals.

Tables 1 to 7, at the end of this document, show the DSU users, their demand and level of satisfaction from DSU services (Source: DSU User's Survey 1987).

1.4 **RESOURCES:** Hardware, Software, Personnel

HARDWARE RESOURCES

IBM 4361

- 12Mb memory
- 400Mb disk
- 66 ports
  - 37 Terminals
  - 23 Microcomputers
  - 3 Printers
  - 3 available ports

MICROCOMPUTER LABORATORY

- 7 IBM PC with 649K of memory 20-30Mb of disk
- 2 Printers

SOFTWARE RESOURCES

IBM 4361 SOFTWARE

1.	Compilers	: FORTRAN, COBOL, PL/1
2.	Statistical Packages	: SAS/BASICS SAS/STATISTICS GENSTAT
3.	Econometric and Time Series	: SAS/ETS
4.	Mathematical Programming	: MINOS
5.	Data storage, retrieval, report generation	: SAS/FSP
6.	Database Management	: IDMS/R
7.	Graphics	: CALCOMP Software SAS/GRAPH

MICROCOMPUTER SOFTWARE

1.	Statistical Packages	: SAS/PC, MSTAT
2.	Spreadsheet	: LOTUS 1-2-3
3.	Database Management	: DBASE IV
4.	Graphics	: SAS/GRAPH
5.	Slides, report production	: PC STORYBOARD

- **Personnel resources**: The total number of positions budgeted to the DSU is 21. Out of them, 2 are Principal Staff positions (Unit Head and Biometry Head), 9 national staff professionals and 10 clerical staff. Apart from these, there is a group of 7 professionals that are administratively attached to other Programs but technically responsible to the Unit. The table below shows the distribution of personnel within the three existing groups in the Unit: Biometry group, Information management group and IBM 4361 hardware/ software support group.

Table 10: **P E R S O N N E L 1989**

	Financed by DSU	Financed by other Programs but Technically responsible to DSU	TOTAL
UNIT HEAD	1		
Secretary	1		
Biometry			
- Head	1		
- Statistical Consultants	4	1	
- Statistical Programmer	1	3	
Information Management			
- DB Administrator	1		
- Analysts	3	3	
- Programmers	2		
IBM 4361 Hardware/Software support			
- System's Programmer	1		
- Computer Operators/control	3		
- Transcriptors	3		
TOTAL	21	7	28
Professionals	11	7	18
Clerical Staff	10		10

- Statistical Consultants: Their responsibility includes Statistical advise, data analysis support, involvement in collaborative methodological studies and training courses.

- Statistical Programmers: Their responsibility includes the preparation of computer programs for data analysis under the supervision of the Statistical Consultants. They also collaborate in training.

- Analysts: They are responsible for the design, implementation and maintenance of the various database applications.

- Programmers: Under the Analysts supervision, they are responsible for the preparation of computer programs to satisfy the databases applications needs.

- System's Programmer: Responsible for the maintenance and update of the various software products installed in the IBM 4361. Also for the coordination of IBM 4361 hardware maintenance contracts.

- Computer Operators/Control: Responsible for the day-to-day operation of the IBM 4361.

- Data Transcriptors: Responsible for the data entry job for IBM 4361 users.

1.5 PRODUCTS

The DSU "products" can be grouped in three categories:

a) Services to research: Statistical/mathematical advise, data analysis, and training courses.
b) Software development: Commodity-oriented and general-use databases.
c) Contributions to research: Methodological Studies.

Our available statistical/mathematical software allows us to apply any kind of statistical methods to applied agricultural research problems: analysis of experimental designs (conventional designs, incomplete factorials, non-orthogonal

designs, cross-over designs); Linear models techniques (all range of univariate and multivariate statistical techniques); Non-linear techniques; Categorical data analysis techniques; Non-parametric techniques; Sample surveys design and analysis.

With respect to database software development, section number 2 of this report presents a summary of the philosophy, present state and potential utilization of existing database applications.

With respect to contributions to research through methodological studies, section 3 of this report summarizes 8 selected examples. They represent collaborative studies carried-out between scientists and biometricians through the analysis of accumulated research information: a) to provide answers to relevant questions formulated by program scientists, or b) to evaluate in-use experimental procedures.

2. **DATABASES APPLICATIONS:** Philosophy, present state and utilization.

2.1 **COMMODITY-ORIENTED databases:**

PHILOSOPHY: To store - in an organized way, with minimum redundancy and with easy recovery capabilities - the **historical experimental and environmental information** generated by CIAT Programs and collaborating NARDS across the years of CIAT research. Commodity-oriented databases have been conceptualized, designed and implemented in a very close discussion with the concerned Program and individual scientists, to satisfy what they believe should be electronically stored.

There are four commodity-oriented databases that have been developed using IDMS/R software: TROPICAL PASTURES, BEANS, CASSAVA and RICE databases. At present, some are more complete than others. The four databases follow a similar pattern of information storage. They are designed to:

a) Keep records of all the germplasm bank accessions and their characteristics.
b) Keep records of the accessions performance in individual experiments carried-out in each selection stage. Such selection stages include: Characterization evaluations, preliminary evaluations, advanced evaluations (on-station and on-farm), mutilocational trials for germplasm adaptability across environments, and finally, technology adoption results.
c) Keep records of each accession successive evaluations and performance **across** selection stages.
d) Keep records of complementary experiments: that is, experiments that are support to those related to germplasm-evaluation. Such as: pathology, entomology, microbiology, biotechnology, ecophysiology, soil fertility, plant nutrition, management practices and seed production experiments, as well as socio-economic statistics.

Although similar in their general philosophy, each database is designed in accordance with the Program organization - in terms of its research disciplines - and following its specific germplasm evaluation scheme. For example, the Tropical pastures database is organized in a modular way - by section - (Germplasm, Microbiology, Pathology, Agronomy Carimagua, Agronomy Cerrados, Pasture Quality....Production Systems,

Economics), and the information within each section is organized by experimental project. Also, all the Tropical Pastures Program project proposals and follow-up, between 1978-1989, are computerized and stored in the database. The other three commodity-oriented databases: BEANS, CASSAVA and RICE, are more breeding-oriented. They store the identification and research evaluations of germplasm bank accessions, crosses and their genealogy, early generations, (F2-F5 in the case of Rice only), advanced lines (F6, F7), and finally, multilocational evaluations for promising varieties and International Networks information.

Type of data stored includes:

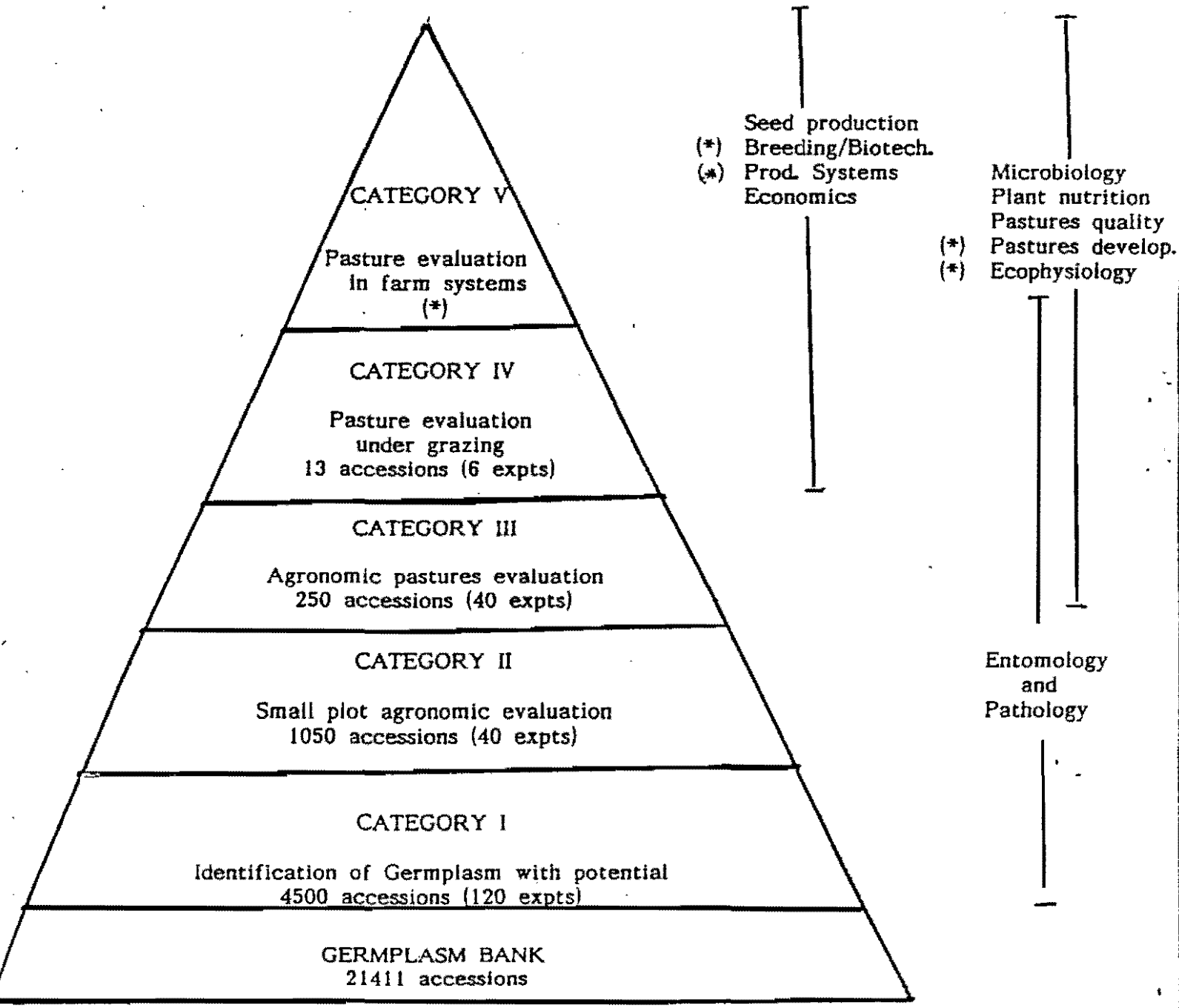
- General description of each project (or experimental):
  - Site description (coordinates, climatic and soil parameters, accessed from the Agroecological database)
  - List of accessions evaluated.
  - Response variables and their evaluation scale, code or units.
- Means per accession in each project (or experiment), for all response variables.
- Precision statistics per project (or experiment): Overall mean, standard error, CV.

A large part of the information stored in the databases has been previously statistically analyzed by the DSU biometricians. The scientist responsible for the information defines whether or not a particular set of experimental results should be stored in the Program database. In this way database information is correct, biologically valid and can be shared with other scientists.

PRESENT STATE AND UTILIZATION OF COMMODITY ORIENTED DATABASES.

- TROPICAL PASTURES DATABASE: is the more complete one in terms of information stored at present. The DSU and the Tropical Pastures Program in very close collaboration, have been working on this database since 1982. Information stored at present - expressed in number of accessions and their evaluations - in each selection stage (or CATEGORY), is presented in the next figure. Tables T1, T2, T3 and T4 describe in more detail the amount and type of experimental information

INFORMATION STORED IN THE TPP DATABASE _CENTRAL PATH_



(*) to be included

Table T1: INFORMATION STORED IN THE TPP DATABASE -BY SECTION-

GERMPLASM UNIT

SECTION	Type of information	No. of accessions	No. of Expts. or years
GERMPLASM	. Passport data	21411	-
	. Seed inventory	18042	-
	. Initial multiplication	3900	7 years
	. Characterization	4500	120
BREEDING	. Evaluation of Centrosema crosses	4313	1
AGRONOMY (Carimagua, Cerrado)	. Under cutting evaluations	1050	30
	. Grazing evaluations	250	40
ENTOMOLOGY	. Characterization evaluations		10
PATHOLOGY	. Characterization evaluations		50
	. Fungus, bacteria	4500/200	-

Table T2:

INFORMATION STORED
IN THE TPP DATABASE
-BY SECTION-

-PASTURE EVALUATION UNIT-

SECTION	Type of information	No. of accessions	No. of Expts.
MICROBIOLOGY	. Rizhobium Strains evaluations	150 Strains	50
	. Rizhobium Bank	4075 Strains	-
SOIL/PLANT NUTRITION	. Fertilizer adjustment evaluations		30
PASTURE QUALITY and PRODUCTION	. Nutritional characterization	40	8
	. Palability		1
	. Pasture productivity	13	6

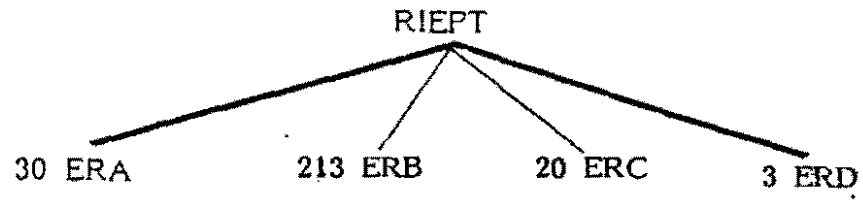
Table T3:

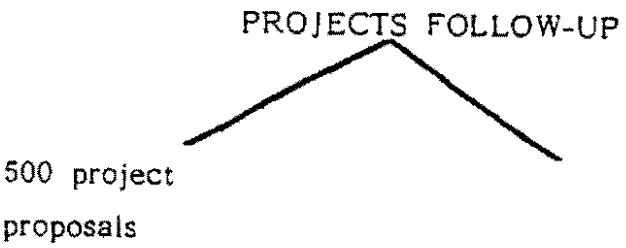
INFORMATION STORED
IN THE TTP DATABASE
-BY SECTION-

SYSTEMS UNIT

SECTION	Type of information	No. of Accessions	No. of Expts.
PRODUCTION	. Reproductive performance		3
SYSTEMS	. Early Wearing		1
ECONOMICS	. Prices of inputs and products. RIEPT sites.		40 sites
	. FAO data on production and trade on 25 products		25 countries

Table T4: INFORMATION STORED IN THE TTP DATABASE





presently stored, by section in each of the three Tropical Pastures Program Units: Germplasm Unit, Pastures Evaluation Unit and Systems Unit, plus the corresponding information on the International Network (RIEPT).

Alternatives for information recovery, include, for example:

- given an accession, find its performance across evaluation stages.
- given an experiment, find the list of material evaluated and their performance
- given an experimental site, find the list of experiments, accessions tested and their evaluation score.

PRESENT UTILIZATION OF THE TROPICAL PASTURES DATABASE

The following catalogues and periodic publications have been produced utilizing information from the database:

CATALOGUES/PERIODIC PUBLICATIONS

1. "Catálogo de germoplasma de especies forrajeras tropicales". Tomos I, II y III 1987. (Schultze-Kraft,R.; Arenas,J.A.; Franco,M.A.; Belalcazar,J. y Ortiz, J.)
   - Passport information on approximately 20.000 grasses and legumes that conform the Germplasm Bank.
2. "La colección de Centrosema de CIAT", 1986. (Schultze-Kraft,R.; Arenas,J.A.; Franco,M.A.; Belalcazar,J. y Ortiz, J.)
   - Information on origin and passport on approximately 1.500 accessions of 37 species.
3. "Catálogo mundial de Centrosema". 1985, 1986, 1987, 1988 (in progress). (Schultze-Kraft,R.; Arenas,J.A.; Franco,M.A.)
   - Name and Identification of each of approximately 2.000 accessions in different Research Institutions.
4. "Catalogue of rhizobium strains for tropical forage legumes". 1987. (Franco,M.A.; Mosquera, D.; Campuzano, F.; Bradley, R.)
   - General information on 4075 rhizobium strains (host, origin, collector, synonyms, acidity, etc.)
   - Results from laboratory and field evaluations on 150 strains from 50 experiments (1980-1987).

RIEPT

- "Resultados 1979-1982", II Reunión RIEPT. (E.A. Pizarro,ed).
  - Analysis by location and by ecosystem. 47 regional trials (10 ERA, 37 ERB).
- "Resultados 1982-1985" III Reunión RIEPT. (E.A. Pizarro,ed).
  - Analysis by location and ecosystem. 121 regional trials (24 ERA, 80 ERB, 10 ERC, 7 ERD).
- "Evaluación del comportamiento de ecotipos dentro y a través de ecosistemas" 1985. (J.M. Toledo; M.C. Amézquita; E.A. Pizarro).
- Análisis de precios de productos e insumos ganaderos en localidades de la RIEPT". 1984, 1985, 1986, 1987. (L. Rivas; C. Seré).

- "Trends in CIAT Commodities", 1982,...,1988. Contribution: management of FAO data on production and trade of 25 food products in 22 countries in Latinamerica and Caribe.

- BEANS DATABASE: Information stored at present, expressed in number of material evaluated and number of experiments from where these evaluations come, include:

	No. of material evaluated
- Germplasm Bank	40000
- Crosses	27631
- Advanced Lines (F6,F7)	11000

Although the database is designed to store every type of Nursery - International or Breeding Nursery - conducted by the Bean Program scientists, the information presently stored in the database includes: Observational Nurseries (VEF, from 1978-1987), Preliminary Yield Trials (EP, from 1979-1987) and International Bean Yield and Adaptation Nursery (IBYAN 1976-1987).

Possibilities for information recovery, include:
- by accession, by cross, by species information.
- Number and identification, of crosses generated from a given set of parental material.

- Lines produced by a given parent, by a given cross, under a given selection criteria, or by a given breeder.
- Identification of materials evaluated in a given experiment or in a given Nursery.
- Evaluations of a given material, in a given experiment or Nursery.
- Genealogy/pedigree of an advanced line.

- CASSAVA DATABASE: Information stored, at present, expressed in number of experiments and materials evaluated, include:

	No. of material
GERMPLASM BANK	
. Passport data	3000
. Seed inventory	800
. Seed interchange (sexual seed only)	471
BREEDING DATA	No. of loaded trials
. Crosses	85
. Selections	70
. Observational trials	59
. Preliminary yield trials	48
. Yield Trials	77
. Regional Trials (54 sites in 5 ecosystems, 1080 varieties)	

- RICE DATABASE: Information presently stored, include:

	No. of accessions
- Germplasm Bank	1620
- Crosses	10000
- Segregating Populations	-
- Observational trials	-
- Advanced lines	-
- Yield trials	-
- International trials (IRTP)	1800 1/

1/ 1800 Lines, 15 types of nurseries, 122 trials conducted in 25 countries. (12 years of results)

Team work on the Rice Program Database started about two years ago. A group of Rice Program

scientists and database analysts have been closely working on the conceptualization and definition, implementation of this database. From the initial conceptualization of the information to be stored to the type of recovery capabilities to be given to this tool, every detail has been discussed. A learning experience from the other databases has been very positive.

2.2 GENERAL-USE DATABASES:

- **Agroecological databases:** This database, developed by the Agroecological Studies Unit with DSU support, consists of 7 sections:
  1. Climate
  2. Land Systems data
  3. Mapping
  4. Cassava microregions
  5. Bean microregions
  6. Rice microregions
  7. Pastures microregions

  Data loaded, at present, includes: Climate data from 10525 meteorological stations; geographic data and coordinates for 4900 municipalities and 310 locations from 225 countries; land systems data from 840 land systems in diverse regions. It is linked with the commodity databases, providing them with climate and soil information of their various experimental sites.

- **CIAT trainees Information System:** Its purpose is to store, provide interactive access and produce a wide variety of computer-generated reports on training information of CIAT-trainees. The system, developed in SAS/FSP software, stores at present personal information, institutional information and type of training received, by 6834 professionals - CIAT trainees - between 1972 and 1989: 3934 trained at CIAT-Palmira and 2900 trained at their own country.

- **FAO information on Trade and Production Statistics:** This database contains 25-years FAO statistics from 25 countries of Tropical America and Caribe, on 23 agricultural products of CIAT interest. It represents a very useful tool for CIAT economists.

- **UNESCO CDF/ISIS (Computerized Documentation Facility/Integrated Set of Information Systems).** It is a bibliographic database, written in PL/1 and VSAM (Virtual Storage Access Method), developed and donated to CIAT by UNESCO and adapted by IICA to satisfy specific needs of the CIAT library. The DSU provides disk space and maintenance support to this database, on the IBM 4361 computer. This tool is fully feeded and utilized by CIAT Library personnel. It holds, at present, 30000 bibliographic summaries of publications on Tropical Pastures (Spanish only), Beans (English and Spanish) and Cassava (English and Spanish).

3. CONTRIBUTIONS TO RESEARCH through Methodological Studies:
Selected examples.

A. **Utilization of information generated by International/Regional Networks for germplasm evaluation.**

Traditionally, the Unit has been responsible for all the data processing and statistical analysis needed to support the various International or Regional Networks in which CIAT research programs collaborate. Information analyzed by the Unit up to May 1989 includes, for example:

- The IBYAN network (International Bean Yield and Adaptation Nursery)
  12 Years of results (1976-1988)
  1267 trials
  65 countries
  20 types of Nurseries
  1030 varieties tested

- The IRTP Network (International Rice Testing Program)
  12 years of results (1976-1988)
  122 trials
  25 countries
  15 types of nurseries
  1800 lines tested

- The RIEPT Network (International Network of Tropical Pastures Trials)
  10 years of results (1979-1989)
  4 types of trials
  208 locations
  20 countries
  730 ecotypes including forage grasses 247 and legumes (483)

- Cassava Regional Trials in Colombia (1983-1986)
  4 years of results
  69 trials, in 34 sites
  5 ecozones
  261 varieties tested

There are some other International Networks whose data is handled by the DSU, such as the International Bean Rust Nursery, (IBRN), International Bean Angular Leaf Spot Nursery and the Rhizobium Network for forage legumes.

Apart from the routine statistical analysis published in the various "reports to collaborators" - involving analysis per location and some multilocational analyses by ecosystem or region, carried out to identify widely adapted and specifically adapted germplasm - we also utilize the information generated by these international networks to answer relevant research questions.

Five examples have been selected of these types of studies, which we would like to summarize in this document.

Example # 1:

ADAPTABILITY ANALYSIS WITH UNBALANCED SETS

Case: Cassava Yield trials in 5 ecozones in Colombia

E. Granados, C. Hershey (in progress)

34 varieties
35 regional trials
(1979-1986)

The adaptability of a genotype is defined as "its physiological response to improvement in environmental quality". For environmental quality we understand the combination of soil conditions, climate, pests, diseases, weeds, and aspects of the management of vegetable material (establishment, evaluation techniques, sampling errors, among others). To quantify "environmental quality", several alternatives have been proposed. The most accepted is to express it by means of the overall mean of crop yield in that particular environment. The yield--as a resultant factor of the interaction of soil-climate and biotics-plant factors--expresses the potential quality of that environment for the growth of a given genotype of the crop.

In general, statistical methods to analyze the adaptability of genotypes through a wide range of environmental conditions, assume that the entire set of genotypes is evaluated in the same set of locations during the same years. That is, these methods are valid for balanced datasets. Such methods are: a) Regression methods (Yates and Cochran, 1938; Finlay and Wilkinson, 1963; Eberhart and Russel, 1966); b) Multivariate methods, like Principal Component Analysis - a reduction of dimensionality technique - (Pearson, 1901; Hotelling, 1933) and Cluster Analysis - a

classification technique - (Abou-el-Fittough, 1967; Rawling and Miller, 1969; Mungomery et al, 1974; Byth et al, 1976; Fox and Rosiete, 1982); and c) Geometric methods like Principal Coordinates Analysis (Schoenberg, 1935), Multidimensional Scaling and Correspondence Analysis.

The british statistician, P.G.N. Digby, (Unit of Statistics, Agricultural Research Council, University of Edinburg), developed a method for adaptability analysis for unbalanced datasets: "Modified regression analysis for incomplete variety x environment data" - (J. of Agricultural Sciences, Cambridge, 1979, q3, p. 81-83). We applied this method for the adaptability analysis of 34 cassava varieties evaluated in 53 cassava yield trials conducted between 1979-1986 in five ecozones of Colombia.

The method is based on the model (non-linear in the parameters)

$$Y_{ij} = \mu_i + b_i A_j + \epsilon_{ij} \qquad \sum_{j=1}^{53} A_j = 0$$

Y_{ij}: yield of variety i in environment j

μ_i: expected mean for variety i in an "average" environment

b_i: Adaptability coeff. for variety i

A_j: Relative environmental performance w.r.t. an "average" environment.

It permits statistical comparisons of **varieties** tested in different sets of locations and years (as μ_i estimates the excepted mean of variety i in an average environment). Also the method allows the comparison of **sites** where not the same set of varieties were tested (as A_j estimates an "environmental index" expressed as the environment yield potential with respect to the average environment). Based on this method, adaptability indexes were estimated and a classification of the 34 cassava varieties was made, using as classification criteria: (μ_i/st. error of μ_i) and (b_i/st. error of b_i).

Figure 1 shows 10 of the varieties with their adaptability index (horizontal axis) and environmental range where they were tested (vertical axis). Figure 2 shows the 8 variety groups obtained (Ward's minimum distance Cluster Analysis).

Fig. 1: ENVIRONMENT RANGE FOR VARIETAL EVALUATION

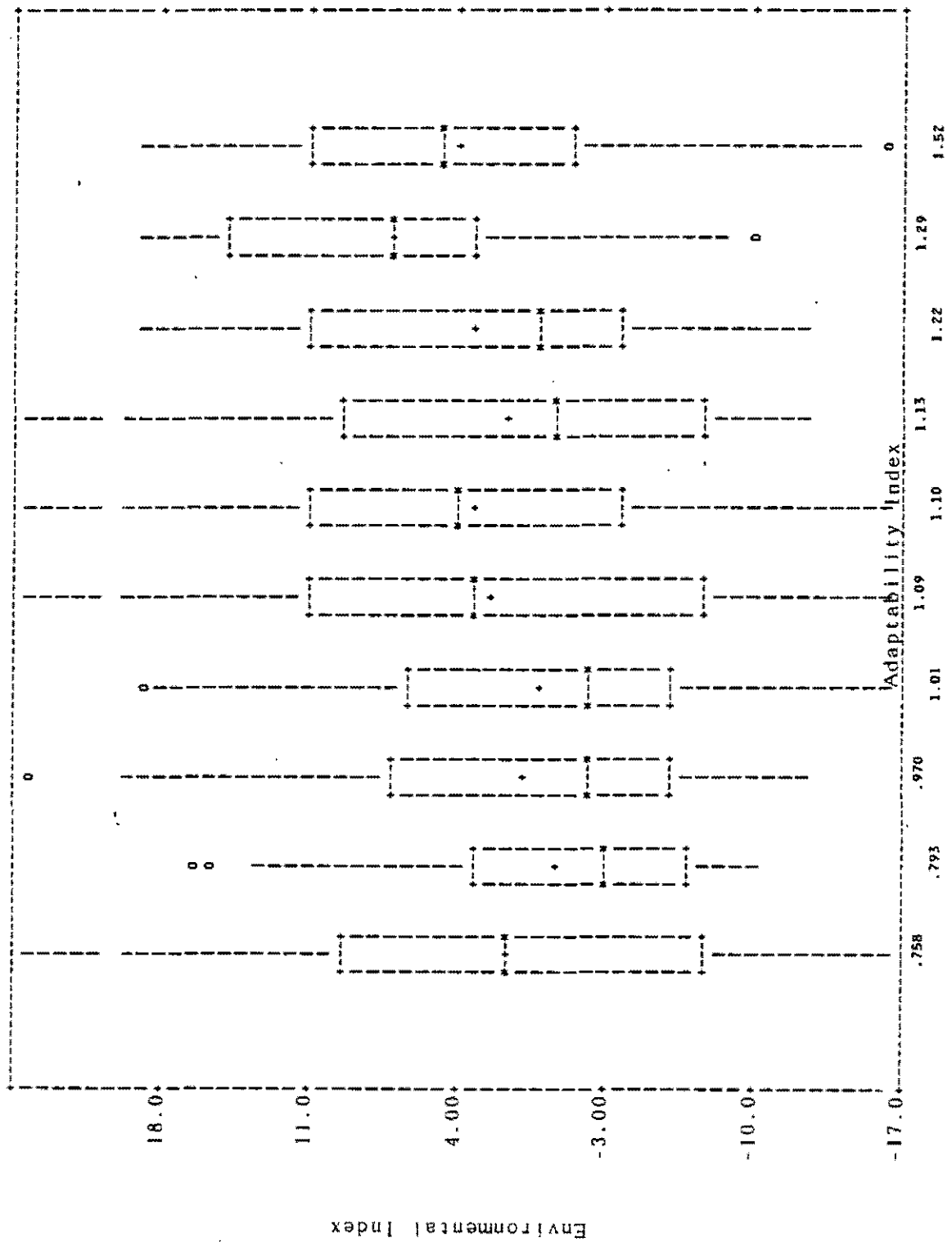
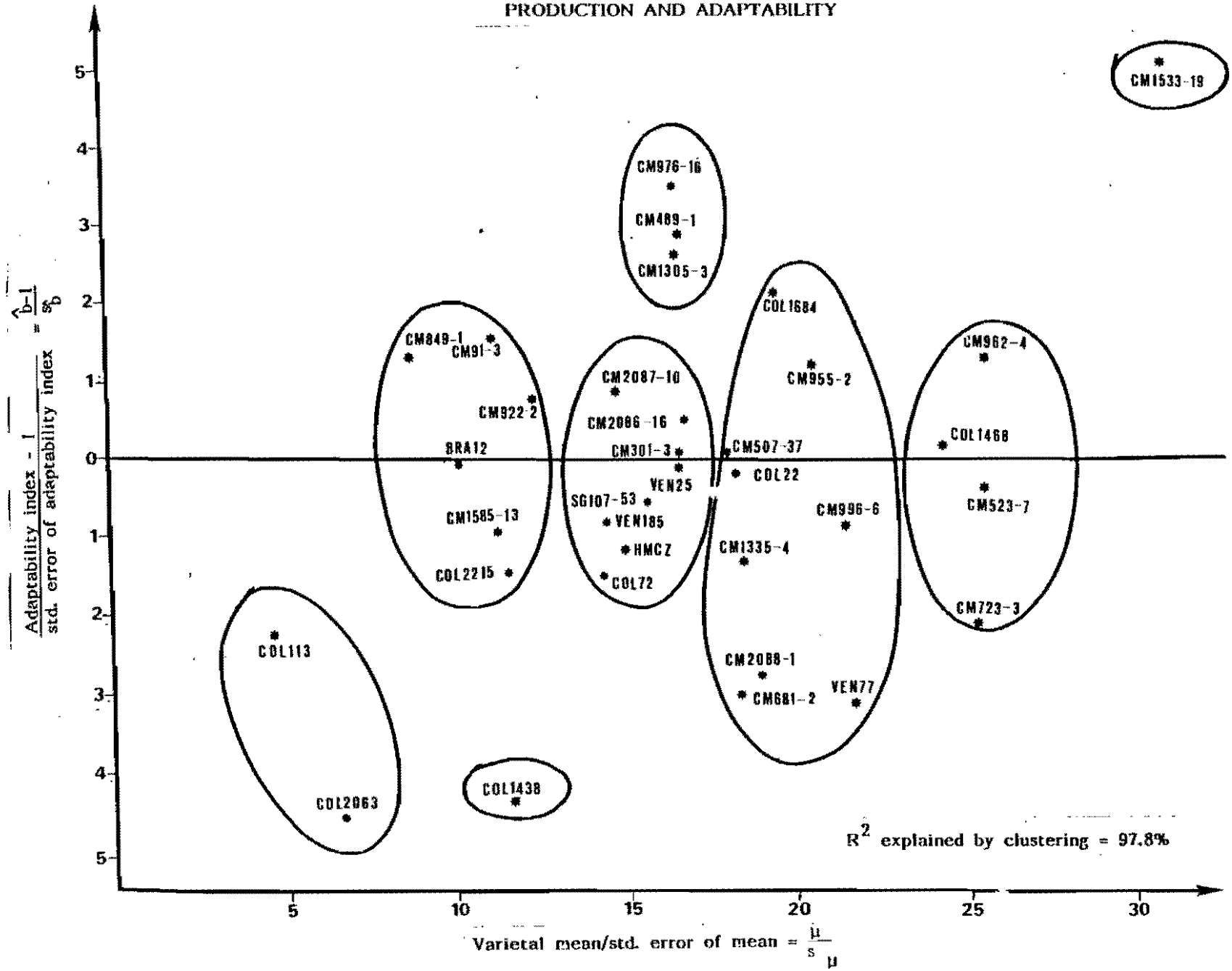


Fig. 2: CLASSIFICATION OF CASSAVA VARIETIES ACCORDING TO PRODUCTION AND ADAPTABILITY



Example # 2

CLASSIFICATION OF ENVIRONMENTS ACCORDING TO PLANT RESPONSE

Case: Andropogon Gayanus in Tropical America.

M.C.Amézquita-E.Pizarro-J.M.Toledo (1986)

The Tropical Pastures Program works on five different major ecosystems which have been identified on the bases of "Wet season potential evaporatranspiration"-a climatic indicator-. However there is a need for a more homogeneous classification of sub-ecosystems within the major ecosystems to better extrapolate the research results. We used the plant as a good indicator of site differences. And we used plant response parameters to classify environments. The study was done with Andropogon gayanus. 44 Regional Trials B, carried out between 1979 and 1986, located in the Savanna and Tropical forest ecosystems was the data source.

Four plant response parameters were selected to classify environments via Cluster Analysis (Ward's method):

- % coverage during establishment
- plant height during establishment
- dry matter production rate in the dry season
- dry matter production rate in the rainy season

The number of soil parameters, was reduced via Principal Components Analysis and an Index of water availability was defined as a function of daily temperature, rainfall, altitude and latitude of the site. Finally, soil and water indexes explaining environments groups were identified.

The three Principal Components obtained out of the original nine soil parameters, are shown in table 8. Table 9 shows the description of the five resulting groups of environments.

Group 2	High dry matter production rate during rainy and dry seasons
Group 1	High dry matter production rate during the rainy season but low in dry season
Group 4	Medium dry matter production rate in both seasons
Group 3	Medium dry matter production rate during the rainy season, low in dry season
Group 5	Low dry matter production rate in both seasons

Table 8. **PRINCIPAL COMPONENTS ON SOIL PARAMETERS**

Soil Parameter	"Fertility index" (31%)	"Texture and acidity index" (25%)	"P content and Ca-K unbalance" (15%)
	S C O R E S		
Sand	-0.27	0.56	0.06
Lime	0.32	-0.23	-0.14
Clay	0.15	-0.56	-0.04
pH	0.36	0.41	-0.18
P	-0.02	0.16	0.52
Ca	0.36	-0.07	0.58
Mg	0.48	0.01	-0.01
K	0.40	0.19	-0.49
Al Saturation	-0.39	-0.28	-0.31

Table 9. **GROUPS OF ENVIRONMENTS**

	1 (n=7)	2 (n=16)	3 (n=7)	4 (n=5)	5 (n=9)	
Plant Parameters						
- Plant height (cm)	83	88	65	77	53	
- % of coverage	57	72	63	57	53	
- Dry matter production rate						
. rainy season (kg/ha/week)	759	670	511	349	250	
. dry season (kg/ha/week)	173	615	129	388	119	
Index of water availability (mm)						
- during rainy season	72.6_{ab}	59.1_{a}	49.4_{b}	45.2_{a}	39.3_{b}	ns [1]
- during dry season	19.1^{ab}	37.1^{a}	15.2^{b}	31.5^{a}	7.4^{b}	*
Principal Components of Soil parameters						
- fertility index	0.99a	0.05ab	0.22ab	-0.50b	-0.79b	**
- Texture and acidity index	0.21	-0.21	-0.25	0.49	0.21	ns
% of sand	43	40	36	65	60	**
% of clay	29	37	37	14	23	**
- P content and Ca-K unbalance	0.18	0.16	-0.12	-0.20	-0.26	ns

1/** Significance of Tukey test for group means comparison, at $p \leq 0.05$
* Significance of Tukey test for group means comparison, at $0.05 < p \leq 0$

Conclusions: a) The groups were explained by "fertility index" and soil texture - associated with dry matter production during the rainy season - and by the "index of water availability during the dry season" - associated with dry matter production during the dry season -. b) Andropogon gayanus presented excellent adaptation to acid and infertile soils, showing however good response to better fertility and heavier soil texture. This result supports the Tropical Pastures Program decision to move to Central America, with moderate acid soils.

Example # 3:

AGRONOMIC PERFORMANCE OF THREE CULTIVARS RELEASED IN COLOMBIA

"Pasto Llanero" (B. dictyoneura 6133)
"Pasto Carimagua" (A. gayanus 621)
"Cultivar Capica" (S. capitata 10280)

E. Mesa, M.C. Amézquita, J.M. Toledo (1988)

The purpose of this study was to provide a quantitative description of the agronomic performance - in Colombia - of these forage cultivars recently released by ICA (Instituto Colombiano Agropecuario).

The study identifies contrasting zones in Colombia based on environmental parameters (altitude, precipitation, Index of bases content, Index of soil texture and organic matter content. The two latter ones corresponding to the first two Principal Components on soil parameters). Then describes performance of each cultivar in each zone making statistical comparisons between zones.
Data source: 22 Regional Trials B conducted in Colombia from 1979-1987.

The results show (see tables 10, 11, 12,) the wide adaptability of both grasses - B. dictyoneura and A. gayanus and their high potential for productivity under the more humid environments with altitudes below 1500 m.a.s.l. (including coffee region and the Amazone). Their performance in the well-drained Llanos ecosystem is similar, with a drastic reduction in productivity during the dry season. Cultivar Capica, on the contrary, is more exclusively adapted to the well-drained sandier Llanos ecosystem, with some potential for dry season production in the coffee zone and the Amazone (Leticia zone). -

Cuadro 10: Comportamiento agronómico del Brachiaria dictyoneura 6133 (Pasto Llanero) en Colombia.

ZONA	Cobertura del suelo (%)	Reacción al salivazo (escala 0-4)	Producción de Materia Seca (kg/ha a 12 semanas de rebrote) Max. Precip.	Min. Precip.
ZONA CAFETERA (n=2 ensayos)	20 b	0.0 c	6284 a	5270 b
QUILICHAO (n=3 ensayos)	17 b	0.3 bc	6089 a	1563
LETICIA (n=2 ensayos)	34 a	2.0 a	3419 b	7450 a
LLANOS ARENOSOS (n=4 ensayos)	18 b	1.3 ab	2595 bc	676 d
LLANOS, MODERADAMENTE ARENOSOS (n=2 ensayos)	10 b	1.5 ab	2246 bc	1358 d
BAJOS INUNDABLES (n=3 ensayos)	36 a	2.0 a	1650 c	3347 c
Media	21	1.1	3442	2586
$\sqrt{CMError}$	1.1 1/	.24 1/	1300	841
CV (%)	25 1/	18 1/	38	32

1/ Correspondiente al análisis con datos transformados ($\sqrt{X+1}$)

Cuadro 11: Comportamiento agronómico del Andropogon gayanus 621 (Pasto Carimagua) en Colombia.

ZONA	Cobertura del suelo (%)	Producción de Materia Seca (kg/ha a 12 semanas de rebrote) Max. Precip.	Min. Precip.
ZONA CAFETERA (n=5 ensayos)	31 b	9566 a	5082 b
QUILICHAO (n=3 ensayos)	10 d	4633 b	3293 c
LETICIA (n=2 ensayos)	62 a	4068 b	6569 a
LLANOS, ARENOSOS (n=1 ensayo)	30 b	3573 b	108 d
LLANOS, MODERADAMENTE ARENOSOS (n=2 ensayos)	19 c	2899 b	1101 d
Media	26	6058	3809
$\sqrt{CMError}$	.78 1/	2416	1013
CV (%)	16 1/	40	26

1/ Correspondiente al análisis con datos transformados ($\sqrt{X+1}$)

Cuadro 12: Comportamiento agronómico del Stylosanthes capitata 10280 (Cultivar Capica) en Colombia.

ZONA	Cobertura del suelo (%)	Reacción a insectos chupadores (escala 0-4)	Producción de Materia Seca (kg/ha a 12 semanas de rebrote) Max. Precip.	Min. Precip.
LLANOS ARENOSOS (n=3 ensayos)	10 bc	3.7 a	2862 a	630 b
QUILICHAO (n=2 ensayos)	5 c	2.5 b	2509 ab	507 b
LLANOS, MODERADAMENTE ARENOSOS (n=2 ensayos)	15 b	1.5 b	1692 c	-
ZONA CAFETERA (n=1 ensayo)	3 c	1.0 b	942 cd	2405 a
BAJOS INUNDABLES (n=2 ensayos)	17 ab	2.5 b	701 cd	-
LETICIA (n=2 ensayos)	27 a	2.0 b	391 d	3152 a
Media	13	2.4	1727	1538
$\sqrt{CMError}$	0.91 1/	.16 1/	784	598
CV (%)	26 1/	9 1/	45	39

1/ Correspondiente al análisis con datos transformados ($\sqrt{X+1}$)

Example # 4:

PROGRESS IN DISEASE RESISTANCE AND YIELD IN *Phaseolus vulgaris*: AN ANALYSIS OF OBSERVATIONAL NURSERIES (VEF 1981-1986), PRELIMINARY YIELD TRIALS (EP 1984-1986) AND INTERNATIONAL YIELD TRIALS (IBYAN 1976-1986)

O. Voysest and J. García, 1987
M.C. Amézquita and O. Voysest, 1987

Three stages are considered under the CIAT Bean Program scheme for evaluation of advanced lines. They are:

1- VEF Nursery, an observational non-replicated nursery, in which lines are evaluated under artificially - controlled high disease-pressure environments, by their reaction to five limiting diseases: rust, anthracnose, angular leaf spot, common bacterial blight and Bean Common Mosaic virus.
2- EP Nursery, a preliminary yield replicated trial, where material selected from VEF are evaluated by their yield and again by their reaction to diseases .
3- IBYAN Nursery, the International Bean Yield and Adaptation Nursery, where promising material are distributed and evaluated in many sites in the world.

The purpose of these two studies (presented and published in the Memories of the "Bean International Trials Workshop", CIAT, Cali, Colombia, October 12-16, 1987), was to describe the progress of the CIAT Bean Program, during 1976-1986, in the production of disease-resistant lines, in the attainment of high-yielding lines with disease-resistant attributes, and in the total yield increment. Table 13, shows that 637 lines (10% of total) are resistant to all five diseases. Specific regional needs that they can satisfy are shown in table 14. Table 15 shows that out of the 317 higher-yielding lines, a high proportion present combined-disease-resistance attributes. The results from a regression analysis of yield vs. time, carried-out in black seeded beans, compares CIAT bean lines with the check variety JAMAPA, and shows that yield increment over the check variety, in a 10-years period, has been of 17.4%.

Table 13: CONTRIBUTION IN SOURCES OF COMBINED RESISTANCE — ALL FIVE DISEASES — (1981 - 1986)

BEAN	TYPES	Total no. of lines produced	no. of "R-I" lines (and %)		% from Total "R-I"
Habit	**Type**				
bush	Red Mottled	1535	301	(20)	50
bush	Brazilian Types	675	65	(10)	10
bush	Mexican Types	498	60	(37)	9
climbing	Low Temperatures	1046	50	(5)	8
bush	Small Red	807	46	(6)	7
bush	Small Black	536	37	(7)	6
climbing	High Temperature	452	32	(7)	5
bush	Small White	475	26	(6)	4
bush	Medium White	100	10	(10)	2
bush	Pacific Coasts	165	10	(6)	2
Total		6289	637	(10%)	100%

Table 14:

— NEEDS BY REGION[1] —
Available Sources of disease resistance
(some examples)

ARGENTINA

		"R-I"
Small black	BB	118
	Ant - ALS	209
Large - Medium White	BB	28
	Ant - ALS	40

BRAZIL

		"R-I"
Small Black	Rust - BB	71
	Ant - ALS	209
Brazilian Types	Rust - BB	106
	Ant - ALS	345

MEXICO

		"R-I"
Small Black	Rust-BB	71
	Ant-ALS	209
Mexican Types	Rust - BB	103
	Ant - ALS	166
Pacific Coasts	Rust - BB	15
	Ant - ALS	50

COLOMBIA — ECUADOR (Sierras)

		"R-I"
Large Red Red Mottled	Rust-BB Ant-ALS	457 507

CUBA

		"R-I"
Small Red	Rust	523

BOLIVIA

		"R-I"
Small Black	Rust - ALS	267

[1] Source: Voysest, O. 1983. "Variedades de frijol en America Latina y su origen". Pastor Corrales, M.A./Morales, F. (personal communication)

- Resistance to BCMV needed in all cases

Table 15: DISEASE RESISTANCE ATTRIBUTES OF HIGH YIELD POTENTIAL LINES

ATTRIBUTE		No. of lines	%
High	Yield potential	317	100
R to	CMV	269	85
R-I to	Anthracnose	229	72
	Rust	189	60
	ALS	167	53
	R-Ant-CMV	125	39
	Ant-ALS-CMV	108	34
	R-ALS-CMV	102	32
	BB	68	21
	R-BB-CMV	36	11
	R-BB-ALS	36	11
	Ant-BB-ALS	28	9
	All diseases	19	6

Example # 5:

IDENTIFICATION OF A SELECTION SITE FOR THE CENTRAL AMERICA - MEXICO RAINFED RICE REGION

F.Cuevas-M.C.Amézquita-M.Rosero (1987)

. 67 IRTP trials (1978 - 1984)
. 112 lines
. 2 international checks(CICA 4, CICA 8)

The Rice Program decision to move to rainfed rice research in the Central America - Mexico area, raised the need for the Program to identify a major selection site for the area. This study provided quantitative confirmation for the choice of the major selection site. We used 67 trials, corresponding to 7 years of results of IRTP, in which 112 lines had been tested, including 2 common checks in all locations and years.

The desired selection site, apart from being representative of climatic and other environmental conditions of the area, should provide the best generalization capability for yield and disease reaction in the area.

So, we wanted a site that could:

- Maximize the probability of adoption of lines selected by the site, that is, maximize
  P_c = Prob (select in the region|Selected by the site)
- and minimize the probability for the site to reject good lines for the region, that is, minimize
  P_d = Prob(select in the region|Discarded by the site)

We have called these two probabilities:
Pc = Prob. of coincidence in selection, and
Pd = Prob. of divergence in rejection.

- A line was considered "selected" in terms of yield if it yielded equal to or more than the best check in the location.

- A line was considered "selected" based on its disease reaction, if its disease index (see table 16) was less or equal to 4 in a 0-9 scale.

Table 16. **Disease Index for rainfed rice in Central America and Mexico**

Rice disease	Weight
Blast (Pyricularia oryzae)	
- Neck	0.35
- Leaf	0.25
Brown Spot (Helmintosporium oryzae)	0.25
Leaf Scald (Rhynchosporium oryzae)	0.15
Total	1.00

Table 17 shows the mean proportion of coincidence in selection and divergence in rejection for five possible locations candidates to Main selection site. Their generalization capability was studied with 60 sites of the region (in terms of yield selection) and 35 (in terms of disease resistance selection)

Table 17. **Mean proportion of coincidence in selection (p_c) and divergence in rejection (p_d) for five possible selection sites**

Candidates to Selection Site	Selection for yield			Selection for disease resistance	
	% of Selection	pc	pd	% of Selection	pc
Alanje, Panamá	75	0.84	0.74	30	0.72
Arce, El Salvador	70	0.80	0.77	45	0.60
Cañas, Costa Rica	75	0.81	0.71	37	0.58
Cuyuta, Guatemala	57	0.85	0.81	60	0.57
Guaymas, Honduras	73	0.83	0.70	70	0.59
No. of participating locations		60	60		35
Pooled S.D.		0.19	0.26		0.27
LSD 0.05		0.07	0.09		0.11

In terms of yield selection, it is difficult to identify the best candidate as, although Cañas and Guaymas minimize the probability of rejection of high-yielding lines for the region, the figure (70%) is rather high. That is, they will discard 70% of lines that could yield well in other locations in the area.

In terms of disease resistance selection, the results are better, as Alanje, being a high disease pressure site, shows the highest probability of success in selecting resistant lines (72% of lines selected by Alanje will be resistant at other locations of the area).

Now, when we study the percentage of high yielding lines within the group of disease-selected lines (see table 18), we observe that Alanje and Guaymas maximize this figure - that is, 85% of resistant lines are high-yielding lines - but Guaymas has the advantage of minimizing the percentage of rejection of high yielding lines, (45%).

Table 18. **Percentage of lines selected for yield, within disease selected and disease discarded groups**

LOCATION	Disease Selected	Disease Discarded	X^2	
	% of lines selected for yield			
Alanje, Panamá	85	71	2.6	NS
Arce, El Salvador	77	50	7.3	**
Cañas, Costa Rica	56	57	0.02	NS
Cuyuta, Guatemala	56	49	0.6	NS
Guaymas, Honduras	86	45	19.9	**

CONCLUSIONS: a) The method represents a practical way to evaluate a location's generalization ability. b) Best alternative: To evaluate the two selection criteria in two different sites and combine information latter on. For disease-resistance selection: Alanje, Panamá. For yield selection: Guaymas, Honduras

B. EVALUATION OF IN-USE EXPERIMENTAL PROCEDURES

Example # 6:

SOURCES OF VARIABILITY IN GRAZING EXPERIMENTS A historical study (1972-1983)

M.C.Amézquita (1986)

This study represents the answer given by the Biometry Section of the DSU, to the Tropical Pastures Program in terms of some methodological recommendations for the evaluation of pastures under grazing, specifically for Regional Trial D. Data on 12 long-term grazing experiments representing 11 years of research at Carimagua Experimental Station (1972-1983), were used for this study.

Five major sources of variability were studied: Soil, climate, pasture, animal, and measurement techniques. According to their relative importance, recommendations were made, in terms of:

- No. of replications in space
- the need or not to use replications in time.
- desired duration of a grazing experiment
- frequency of evaluations in time
- variability in forage productivity associated with pasture nature, considering erect vs decumbent species.
- type of animal: - breed, age and weight to be used as a measurement tool for pasture productivity evaluations.

Two of the results are shown here: The first one (see table 19) refers to **climate** effect. It shows seasonal effect on weigh gain of steers - using data from 7 experiments. - Its results show that dry season effect is one of the most important sources of variability in animal response, not only affecting the mean weight gain but also increasing variability between animals; implying that, if the Program wants to continue measuring dry season effect with confidence, the initial number of animals have to be increased.

Table 19. **Seasonal and annual effect on variability of weight gain of steers. (7 experiments, Carimagua, 1979-1983)**

Factor	No. of Steers	Weight gain (gr/an/day)	Pooled MSerror (gr/an/day)	CV (%)
Season				
Wet	283	485a	80	16.5
Dry	283	175 b	178	101.7
Year				
1979	170	376ab	182	48.4
1980	146	409a	173	42.3
1981	170	383ab	126	32.8
1982	80	348 b	149	42.7

Data source: O. Paladines L. Tergas

The second result (see table 20) shows the effect of **type of animal** on the evaluation of pasture productivity.

Table 20. **ANIMAL VARIABILITY DUE TO BREED**

	PARAMETER	N	MEAN WEIGHT(kg)	CV 3/ (%)
CEBU 1/				
	Weaning weight-steers (kg)	1201	226.7	13
	18 months weight steers (kg)	171	332.1	17
	Interval between parturitions (months)	2226	14.0	18
CEBU x CRIOLLO 2/				
	Weaning weight-steers (kg)	524	122.8	18
	18 months weight steers (kg)	428	169.4	21
	Interval between parturitions (months)	527	18.0	18

1/ CEBU - highly selected pedigree: finca Montenegro, Antioquia (animals descendents from 81 different sites)
2/ Commercial Cebu: from ETES farms (Matto Grosso, Brazil)
3/ CV free from "farm" and "season" effects.

It shows that, Cebu - highly selected pedigree provides higher estimates of the pasture productivity than commercial-type of Cebu. However, animal- variability estimates provided by the two animal types are similar. So the study recommended the use of a standard Cebu x Criollo type of animal for the evaluation of pasture productivity under grazing; standard in terms of breed, age, sex and weight.

This study was published (in Spanish and proximately in English) as one chapter of the book "Methodological Alternatives for grazing experiments". Its recommendations are being adopted by the Tropical Pastures Program of CIAT and by other National Institutions.

Example # 7:

EFFECTIVENESS OF SELECTION FOR white Belly in rice

C. Martinez, F. Cuevas, M.C. Amézquita
H.F. Ramirez (in progress)

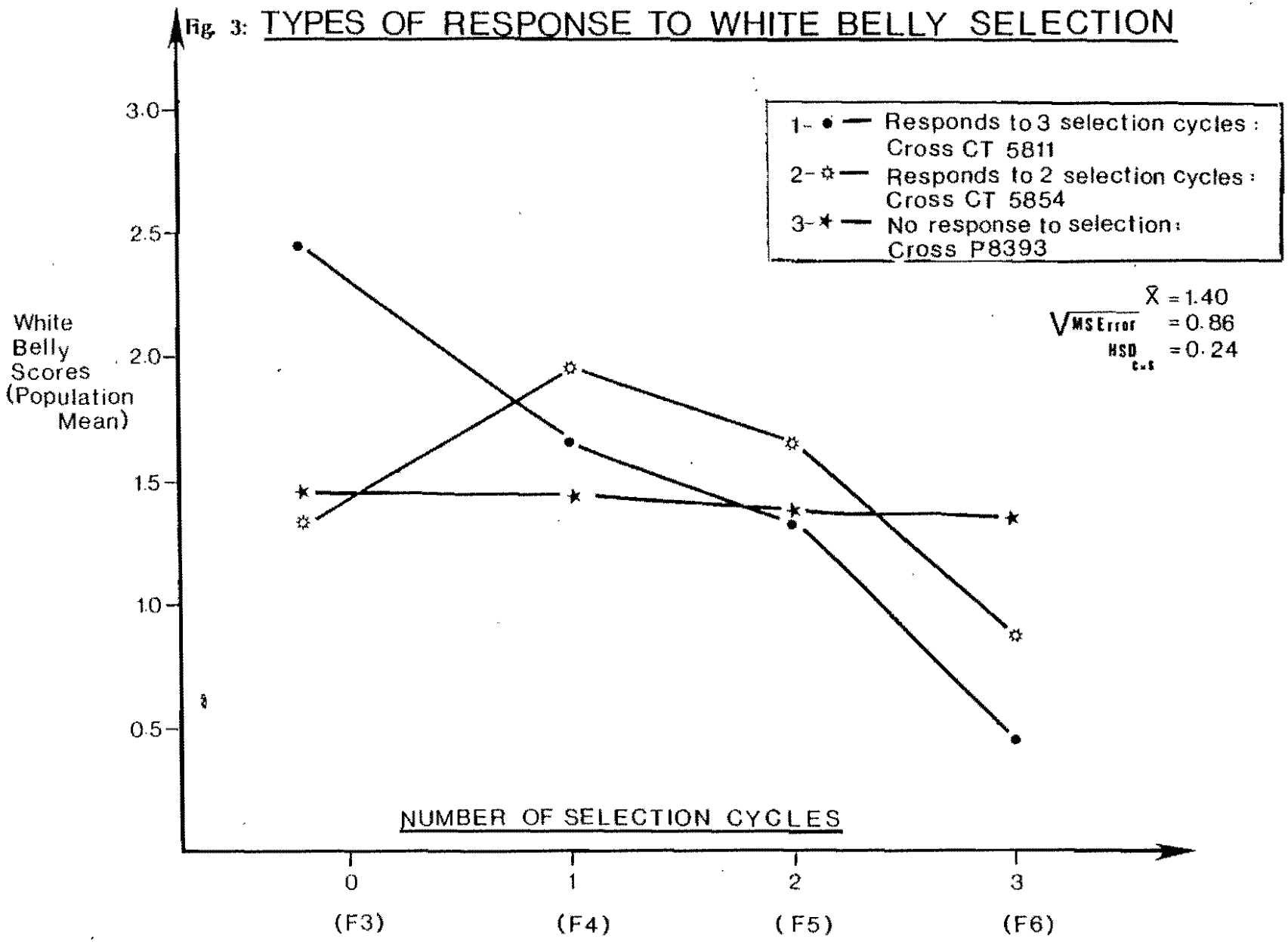
11 Crosses, evaluated in
F3, F4, F5 and F6

The purpose of this study was: a) To study the effectiveness of selection strategies of White Belly in rice; and b) To understand existing variability in White Belly scores in rice parental material. Under these objectives, an experiment was designed to compare response to selection of 11 crosses - resulting from 22 parental material - when selection started in different generations: F3, F4 and F5. Parental material was evaluated through the 4 semesters of duration of the experiment.

Preliminary results, product of mean comparisons of White Belly scores on four F6 - populations (selected from F3, selected from F4, selected from F5 and non-selected), show that there are very different trends in response to selection among crosses. Three main types of responses were identified: (1) crosses that respond to very early selection (selection from F3): CT5811 (2) crosses that respond to selection from F4: CT5854; and (3) crosses that do not respond to selection: P8393. (See figure 3)
Trying to explain this interaction, an analysis of variability in White Belly scores in parental material was carried-out to determine whether it was associated with the differences in response to selection. Table 21 shows a classification of the 22 parents according to their mean and genetic variance in White Belly scores.

Fig 3

Fig. 3: TYPES OF RESPONSE TO WHITE BELLY SELECTION



The study is in progress. Once the "cross x selection stage" interaction can be explained, through an understanding of variability within parental material, proper estimations of heritability coefficients can be attained.

Table # 21: **PRELIMINARY RESULTS**

Classification of 22 rice lines based on White belly mean and variance using Cluster Analysis 1/

Cluster	No. Parental Lines	White Belly Score Mean	Variance	Designation 2/
1	10	0.88	2.22	LL
2	6	2.20	10.75	II
3	1	2.37	28.91	IH
4	5	3.43	4.81	HL

1/ data from 4 seasons. R^2 explained by Clustering = 85%

2/ mean and variance, where L = low, I = intermediate and H = high.

Example # 8:

SCREENING FOR RICE VARIETAL TOLERANCE TO DELAYED HARVESTING: AN APPLICATION OF DISCRIMINANT ANALYSIS

F. Cuevas, E. Granados, L.E. Berrio
(in progress)

This study is being carried-out to develop a practical method for predicting - through laboratory evaluations of the grain - the tolerance of rice varieties to delayed harvesting. Delayed harvesting and its associated problems cause a decrease in rice milling quality and subsequently in its market price.

The methodology followed was: (1) To identify environmental factors causing a decrease in milling quality. (2) To identify rice grain variables that could help to better

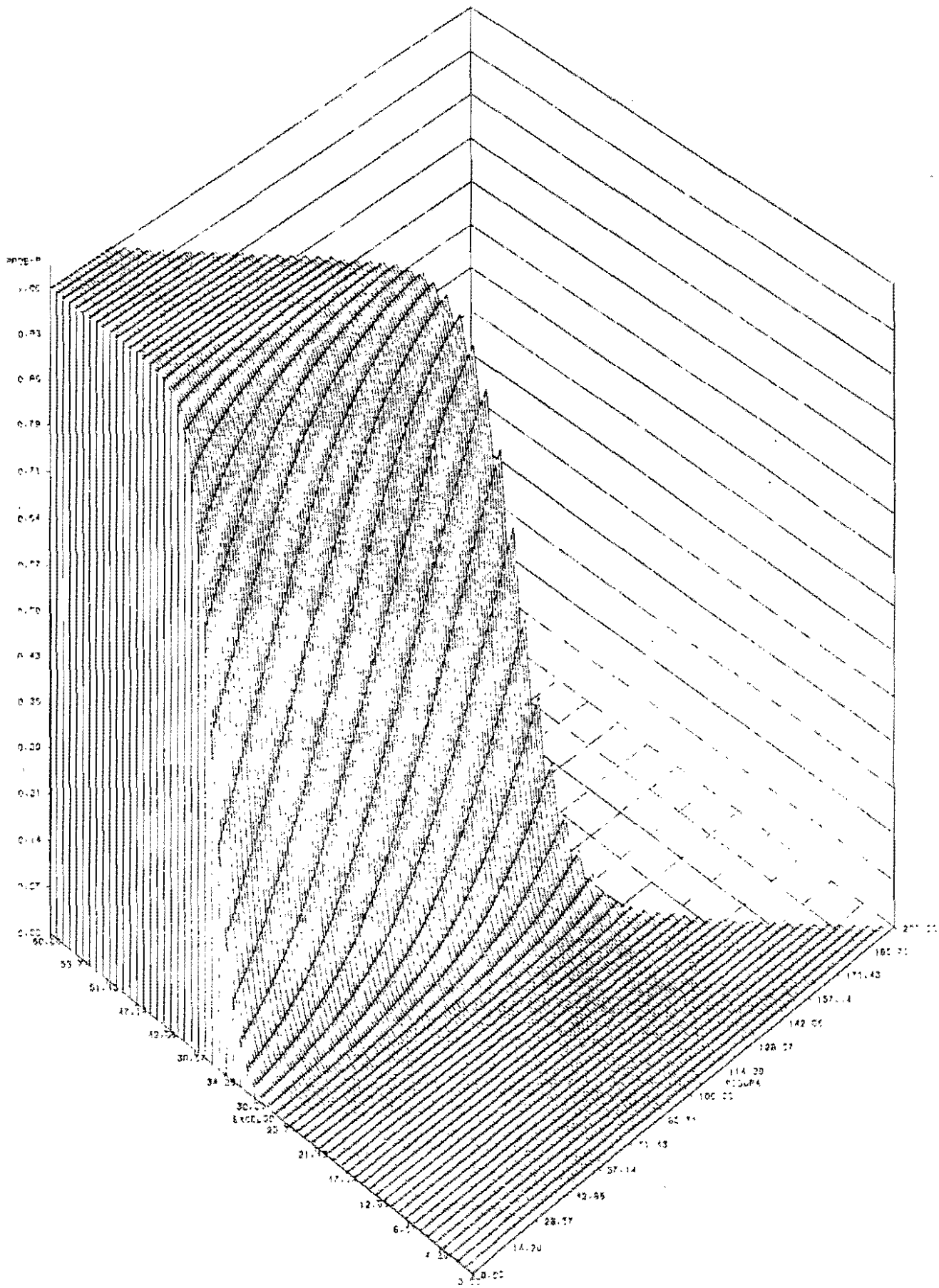
discriminate between resistant (R), medium-resistant (MR), medium- susceptible (MS) or susceptible (S) varieties to delayed harvesting. (3) Construct discriminant functions based on the previously identified variables, that will be able to classify rice varieties in one of the 4 quality classes. (4) Validate the discriminant function with check varieties with known tolerance; and finally, (5) use the discriminant functions to classify new rice varieties, of unknown tolerance levels, in the 4 mentioned classes.

The results of the study show that the most important environmental factor causing a decrease in milling quality was humidity. Two rice-grain variables were identified (through Stepwise Discriminant Analysis) as the best predictors of milling quality: a) **whole grain weight** before milling (sample of 125 gr.); and b) **Number of grains with fissures** -out of 200-. Based on these two variables, the four **discriminant functions** were constructed to classify any given variety in one of the four quality classes (R, MR, MS, S). Six check rice varieties were used to calibrate them.

Figure 4 shows the probability function to classify a variety in class "R". Figure 5 shows the classification region based on the two grain variables: whole grain weight and no. of grains with fissures.

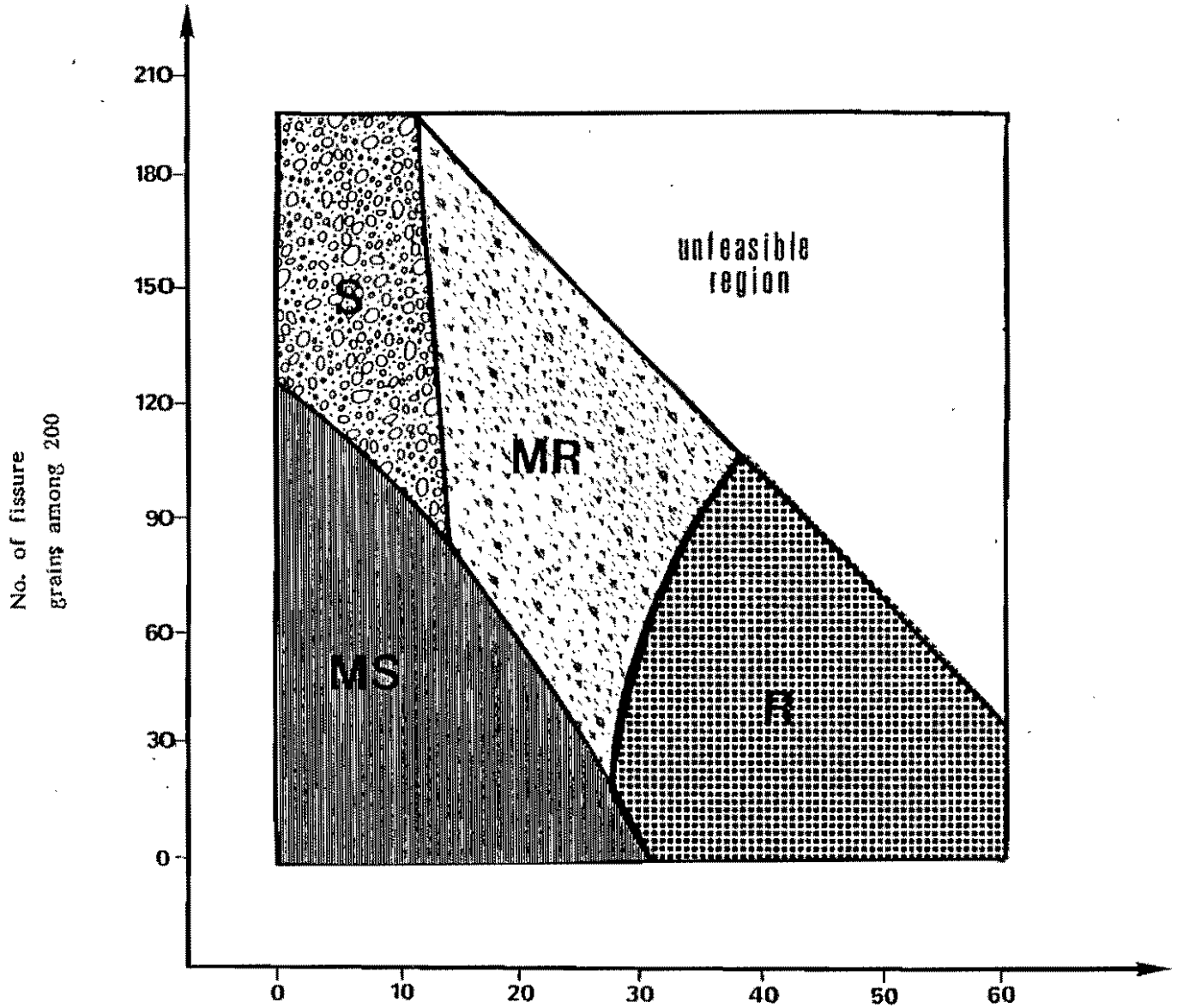
Fig.4: TOLERANCIA DEL ARROZ A RETRASO EN LA COSECHA

CLASIFICACION SEGUN PGE DE ARROZ EXCELSO Y NRO DE GRANOS CON FISURAS



PROBABILIDAD DE CLASIFICAR UN MATERIAL
[illegible] RETRASO EN COSECHA

Fig. 5: CLASSIFICATION REGIONS IN 4 QUALITY GROUPS



Rice "excelso": (Whole grain weight. sample = 125 gr)

4. FUTURE

We believe that the Unit should continue improving its work on the same line:

1. Providing more and better support to the research programs methodological needs. The biometricians will hopefully be able to concentrate more on the use of massive information accumulated by the Programs to continue answering relevant research questions formulated by the scientists and providing support to their methodological needs.
2. Offering statistical consultancy, with emphasis on the appropriate use of statistical techniques, and on the definition of a statistical analysis strategy that can provide the most objective evidence in support to a given hypothesis.
3. Through a very close interaction with the Programs, and in response to their needs, the Unit will continue the development of commodity-oriented and general-use databases, incorporating higher levels of data aggregation. This will make possible the use of the databases as decision-making tools at various institutional levels.
4. Satisfy the increasing demand for training in the use of quantitative methods for research data analysis and interpretation, via microcomputers,
   - for CIAT research associates/assistants
   - for National Institutions biometricians and agricultural researchers.

This is to us a very interesting challenge.

Thank very much

MARIA CRISTINA AMEZQUITA

Table 1: **USERS OF THE DSU (at CIAT HQ) (DSU USER's SURVEY 1987)**

	COMPUTING N	%	BIOMETRY N	%
Users	46	87	41	77
Total no. of responses	53		53	

Response rate = 53/69 = 76.8

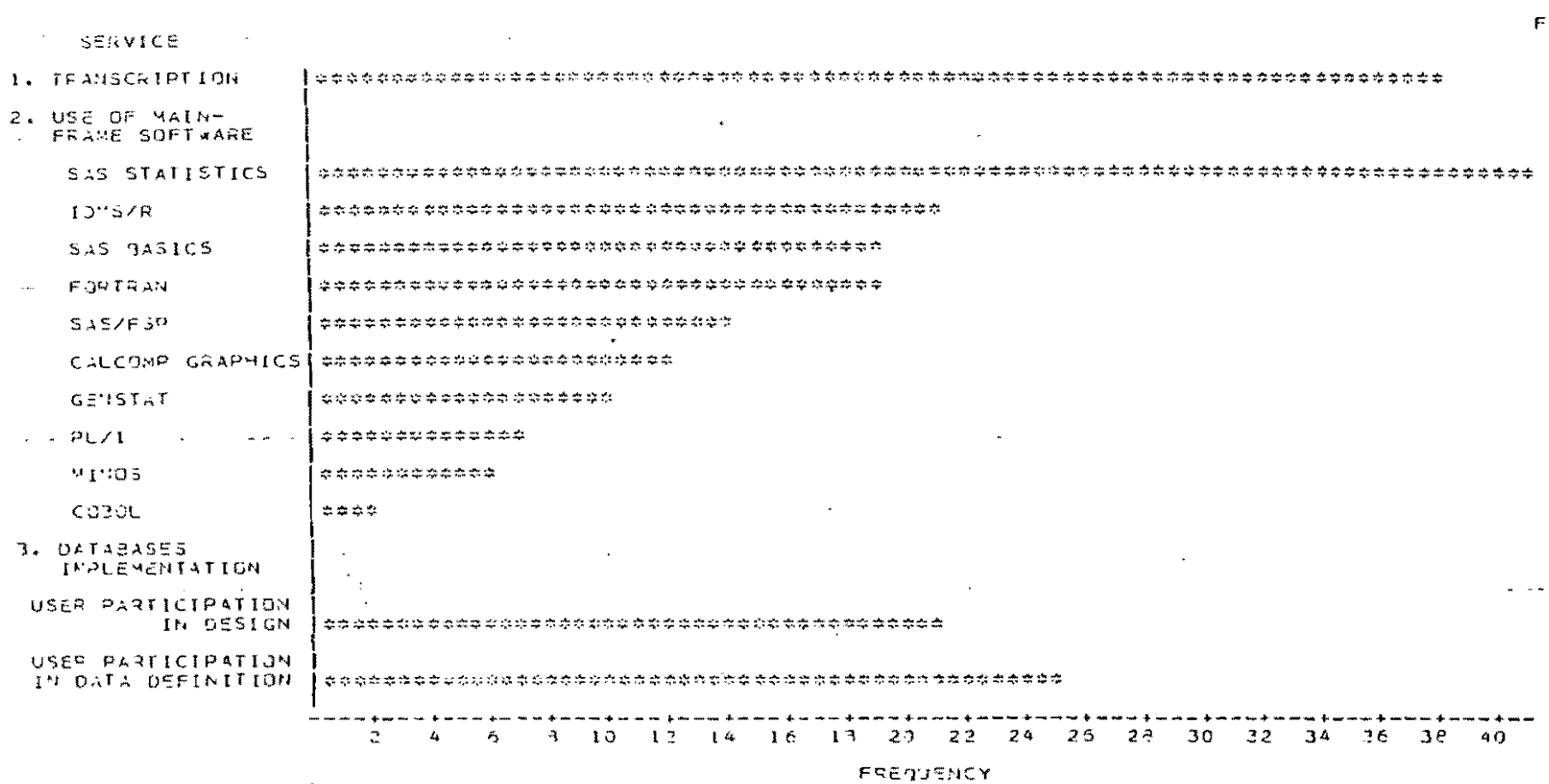
Table 2: **USERS OF THE DSU (at CIAT HQ) DISCRIMINATED BY PROGRAM**

	No. of formulaires Sent	Received	No. of Users	%
Beans	15	8	8	100
Cassava	11	9	8	89
Pastures	16	14	13	88
Rice	6	6	6	100
Support Units + Projects	21	16	11	69
TOTAL	69	53	46	87

Table 3: **USERS RATING OF THE COMPUTER SYSTEM IN GENERAL**

Rating	N	%
Satisfactory	24	52
Inadequate	5	11
No response	17	37
Total no. of responses	46	100

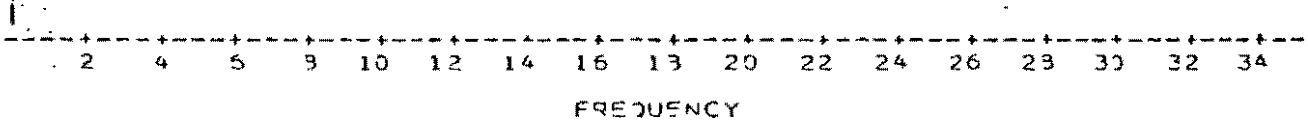
TABLE 4 : DEMAND FOR COMPUTING SERVICES



Total number of Computing Users = 46
Total number of responses = 53
% of Users = 87%

TABLE 5: DEMAND FOR BIOMETRY SERVICES

SERVICE	FREQ	%
1. ADVISE		
PLANNING, DESIGN	33	80
STATISTICAL METHODOLOGY	35	85
DATA PROCESSING/ ANALYSIS	28	68
2. DATA PROCESSING AND ANALYSIS SUPPORT	21	51
3. COLLABORATIVE STUDIES WITH BIOMETRICIANS	15	37
4. TRAINING IN STATISTICAL METHODS AND DATA ANALYSIS		
-IN-HOUSE COURSES	26	63
-EXTERNAL COURSES	19	46
5. PARTICIPATION OF BIOMETRICIANS IN CIAT'S PROGRAMS WORKSHOPS	21	51



Total number of Biometry Users = 41
Total number of responses = 53
% of Users = 77%

Table 6: **HOW DO USERS RATE STATISTICAL CONSULTANCY AND STATISTICAL PROGRAMMING SUPPORT**

	SATISFACTORY N	%	INADEQUATE N	%	NO RESPONSE N	%
Statistical Consultancy						
- Quality of Consultants	39	95	2	5	-	-
- Availability of Consultants	29	71	12	29	-	-
Statistical Programming						
- Quality of Programmers	31	76	9	22	1	2
- Availability of Programmers	9	22	-	-	32	78

Total no. of users = 41

Table 7: **PREFERRED MODE OF OPERATION WITH BIOMETRY**

	N	%
1. Advise only	7	17
2. Advise + data processing and analysis in complex jobs	20	49
3. Advise + data processing and analysis in all jobs	12	29
4. No response	2	5
Total no. of users	41	100