

BIOMETRY UNIT
ANNUAL REPORT 1992

**For internal circulation
and discussion only**

CIAT Internal Program Review 1992

November 30 - December 4, 1992

BIOMETRY UNIT ANNUAL REPORT 1992

- EXECUTIVE SUMMARY -

1. BIOMETRY UNIT: Role, functions and resources

The Biometry Unit provides advice and support to CIAT research Programs/Units in two main areas:

- a) Biometry/data analysis methods.
- b) Conceptualization and implementation of databases of experimental results.

During this transition period to a distributed processing environment at CIAT and the possible creation of a new Unit responsible for providing CIAT-wide computer services, the Biometry Unit (previously called Data Services Unit) is still responsible for the provision and maintenance of appropriate computer hardware and software to serve the scientific program's needs through the CIAT mainframe computer, IBM 4361, with its network of 37 terminals and 41 connected microcomputers.

Biometry support: Role definition:

Biometry collaboration to CIAT research Programs/Units includes the following functions:

- a) Statistical/mathematical advice to researchers in project design, data analysis methodology, interpretation of results, their generalization capacity, and final presentation.
- b) Involvement in collaborative methodological studies and specific data analysis projects with researchers.
- c) Training in statistical/mathematical methods and data analysis to colleagues in other disciplines (both internal and external).
- d) The biometrician participates as member of database teams --databases containing CIAT experimental results on crops and forages-- by providing selected input to the databases: appropriate statistical summaries with which to characterize a given research process.

An important activity of the biometrician is his/her involvement in **collaborative data analysis projects** with researchers, aimed at responding relevant questions of research. These projects utilize data generated by a given research project through the years, combine experimental results of a given research discipline, or combine data generated by various disciplines within a Program. The results of some of these projects have appeared as chapters of CIAT Programs publications, some as contributions to International Networks reports, some have been published as joint papers with the scientists, and some others are in progress. A brief summary of selected case studies are included in this report.

Another important activity is the biometrician involvement in **collaborative methodological studies** with researchers. The biometrician contribution in this context is to identify sources of variation relevant for research planning; evaluate and recommend appropriate experimental designs for a given research; evaluate and recommended appropriate statistical analysis methodologies: their efficiency, accuracy and applicability for a given research problem.

Basic **training** in statistical methods and data analysis is provided to CIAT research associates/assistants and to National Institutions researchers from Latinamerican and African institutions, CIAT collaborators. The Microcomputer Training Laboratory is used for this purpose. During the five years of existence of the Laboratory, Biometry has offered a total of 35 one to two-week training courses, with a total number of 344 National Institutions researchers trained from Latinamerica (260), Asia (24) and Africa (50). An approximate number of 105 participations from CIAT research associates/assistants have benefit from this effort, during a 5-year period.

In the light of the new CIAT, new areas of biometrical expertise in which invited Biometrician Consultants can add useful contributions are foreseen. For example: a) Design and analysis of intercropping experimentation, combining multiple short-cycle crops or combining perennial and short-cycle crops. b) Design and analysis of agro-silvo-pastoral systems. c) Quantitative genetics/population dynamics. d) Econometric techniques in response to a new expected demand from the economists/social scientists.

Statistical/data analysis software for the mainframe computer include: SAS/BASICS, SAS/STATS, SAS/GRAPH, SAS/FSP, SAS/ETS, SAS/IML and SAS/OR from SAS Institute Inc. Raleigh, North Carolina, USA; GENSTAT, from the NAG Algorithm Group, London, England. Microcomputer statistical/data analysis software include MSTAT, from Michigan State University; GLMM, from Louisiana State University; SYSTAT, from SYSTAT Inc. Chicago, Illinois, AGROBASE/4, from Agronomix Software, Manitoba, Canada; MATMODEL from Soil, Crop and Atmospheric Sciences, Cornell University, Ithaca, New York; Lotus 1-2-3 and Dbase III.

Databases of experimental results:

The conceptualization, design and implementation of databases to store crops and pasture research results, require from the "database teams" a clear understanding of the biological nature of the crop and its multiple components. Members of our "database teams" include: a) the System's Analyst, who is the software expert, the designer of the how to efficiently store the datafiles with minimum redundancy, and how to provide interactive access to the data in the most effective manner; b) the Researcher(s), who have a clear understanding of the problem and the purpose of the database; and c) the Biometrician, who has a clear understanding of the data and of the most suitable statistical summaries with which to characterize a given research process.

In terms of database management software technology an important decision was reached in September 1991: that of moving from the 'network' database technology, represented by our previous database management software IDMS/R from Computer Associates Inc., to the 'relational' database technology, represented by ORACLE, from Oracle Corporation. Starting October 1, 1991, ORACLE was acquired as the standard database management software for CIAT's mainframe and microcomputer environment, as a replacement of IDMS/R. Previous problems in design, implementation and utilization of research databases developed on IDMS/R during the past 10 years, such as lack of flexibility for modifying a database design, extremely long data loading times, lack of a user-friendly query tool, lack of a flexible and powerful development tool and lack of micro-mainframe interface, have been greatly solved. Between October 1991 and December 1992, four important ORACLE database applications have been designed, implemented and released to the end-user:

- a) **The Genetic Resources Database**, including passport and characterization data on all CIAT's germplasm collections: beans (40,000 accessions approx.), cassava (4,700 varieties approx.), tropical forages-legumes and grasses (25,000 accessions approx.).
- b) **The Cassava Breeding Database**, including information on parents and crosses, F₁ evaluations, statistical summaries of preliminary yield trials, advanced yield trials and regional trials. This database contains at present research results between 1974 and 1992.
- c) **The Bean Breeding Database**, linked with the Bean Genetic Resources database, includes information on crosses genealogy, advanced lines genealogy, advanced lines evaluations, VEF - EP trials from 1977 - 1991 and, IBYAN trials from 1976 - 1990. Results and statistical summaries from other international nurseries conducted by Bean Scientists are planned to be added in 1993.
- d) **The Tropical Forages Evaluation Database**, linked with the Tropical Forages genetic resources database, contains results from the characterization and early evaluation of grasses and legumes carried-out by Program Scientists since 1978 up to present. Most of this application was implemented in IDMS/R. Its re-design and implementation in ORACLE of a large part of its components, was carried-out in 1992 in a very close collaboration with the Program Leader and Scientists. This effort will continue during 1993.

Details on these four major ORACLE databases are described in this report.

Training courses for end-users in the use of ORACLE databases: The Unit has provided 4-day training courses to users from the following programs: Bean Program breeding sections and Cassava breeding Section (February 92), Genetic Resources Unit (Dec/91 and November/92) and Tropical Forages support personnel (November/92). These short courses were intended to explain content of the databases to the direct users and teach them how to do easy queries to each database.

PERSONNEL RESOURCES. Composition of Biometry Unit teams to support CIAT research Programs/Units

The list of our 1992 personnel is presented at the beginning of this report. In order to provide a multi-skilled support to each CIAT research Program/Unit in the two areas of the Unit responsibility --Biometry and Databases of experimental results-- teams of biometrician/analyst/programmer are assigned to serve the needs of projects within each Program/Unit.

Through this approach, each Program/Unit scientist has access to the range of technical skills represented by personnel within the Biometry Unit. The attached table shows the composition of these teams.

Biometry Unit

PERSONNEL 1992

B
DBA
P

NAME	POSITION	DEGREE
Unit Head	→ Maria Cristina Amézquita	Ms. and Dipl. in Mathematical Statistics
Secretaries	Maria Eugenia Echeverri → left ciat Marta Elena Carvajal	
BIOMETRY		
Statistical Consultants:	Eduardo Granados → leaving ciat James A. García → IMNS	Ms. Mathematical Statistics Ms. Industrial and Systems Eng.
	→ Eloina Mesa	Ms. Statistics
	→ Myriam Cristina Duque	Bs. Mathematics
	→ Germán Lema	Bs. Industrial Eng.
Statistical Programmer:	Norbey Marín → Cassave Rosalba López → is leaving	Statistics (Student)
DATABASES OF EXPERIMENTAL RESULTS		
Analysts:	Arturo Franco → IMNS Germán Serrano → Langhuse Fernando Rojas → IMNS	Ms. Industrial and Systems Eng. Bs. Systems Eng. Bs. Systems Eng.
Programmers:	Carlos Saa → Savainny Fernando Arango → IMNS	Industrial Eng. (Student) Programming Technology (Student)
IBM 4361 OPERATION:		
System's Programmer:	Hugo Macías	Ms. Systems Eng.
Operators:	Jairo Ramírez Carlos López	
Data transcriptors:	Elizabeth González Amparo Rivadeneira	

Personnel budgeted to research Programs but technically responsible to the Unit:

FORAGES	Gerardo Ramirez	Bs. Statistics and Mathematics
	Carlos A. Hernandez	Systems Technology
RICE	Hector Fabio Ramirez	Bs. Statistics

CONTENT

Pages No.

Biometry Unit personnel 1992

1.	BIOMETRY UNIT: Role, functions and resources.	1 - 4
2.	CONTRIBUTIONS TO BEANS	
2.1	Biometry contributions: <u>Methodological and data analysis studies.</u>	5 - 8
2.2	The Bean Breeding Database. Work progress on other discipline-oriented databases of experimental results.	9
3.	CONTRIBUTIONS TO CASSAVA	
3.1	Biometry contributions: <u>Methodological and data analysis studies.</u>	10 - 14
3.2	The Cassava Breeding Database. Its extension to include research results from other disciplines.	14 - 16
4.	CONTRIBUTIONS TO TROPICAL FORAGES	
4.1	Biometry contributions: <u>Methodological and data analysis studies.</u>	17 - 23
4.2	The Tropical Forages Database of experimental results. The RIEPT Database.	23 - 25
5.	CONTRIBUTIONS TO RICE	
5.1	Biometry contributions: <u>Methodological and data analysis studies.</u>	26 - 37
5.2	Evaluation study. Performance of SAS and ORACLE software in the Rice Local Area Network (LAN).	38
6.	CONTRIBUTIONS TO GENETIC RESOURCES Case: The CIAT Genetic Resources Database.	39 - 40

Annex: "CIAT Biometry needs: Demand and Supply": results of meetings, discussions and survey conducted among scientists from research programs/Units, in October-Nov. 1992.

BIOMETRY UNIT ANNUAL REPORT 1992

1. BIOMETRY UNIT: Role, functions and resources

The Biometry Unit provides advise and support to CIAT research Programs/Units in two main areas:

- a) Biometry/data analysis methods.
- b) Conceptualization and implementation of databases of experimental results.

During this transition period to a distributed processing environment at CIAT and the possible creation of a new Unit responsible for providing CIAT-wide computer services, the Biometry Unit (previously called Data Services Unit) is still responsible for the provision and maintenance of appropriate computer hardware and software to serve the scientific program's needs through the CIAT mainframe computer, IBM 4361, with its network of 37 terminals and 41 connected microcomputers.

Biometry support: Role definition:

Biometry collaboration to CIAT research Programs/Units includes the following functions:

- a) Statistical/mathematical advice to researchers in project design, data analysis methodology, interpretation of results, their generalization capacity, and final presentation.
- b) Involvement in collaborative methodological studies and specific data analysis projects with researchers.
- c) Training in statistical/mathematical methods and data analysis to colleagues in other disciplines (both internal and external).
- d) The biometrician participates as member of database teams --databases containing CIAT experimental results on crops and forages-- by providing selected input to the databases: appropriate statistical summaries with which to characterize a given research process.

An important activity of the biometrician is his/her involvement in **collaborative data analysis projects** with researchers, aimed at responding relevant questions of research. These projects utilize data generated by a given research project through the years, combine experimental results of a given research discipline, or combine data generated by various disciplines within a Program. The results of some of these projects have appeared as chapters of CIAT Programs publications, some as contributions to International Networks reports, some have been published as joint papers with the scientists, and some others are in progress. A brief summary of selected case studies are included in this report.

Another important activity is the biometrician involvement in **collaborative methodological studies** with researchers. The biometrician contribution in this context is to identify sources of variation relevant for research planning; evaluate and recommend appropriate experimental designs for a given research; evaluate and recommended appropriate statistical analysis methodologies: their efficiency, accuracy and applicability for a given research problem.

Basic **training** in statistical methods and data analysis is provided to CIAT research associates/assistants and to National Institutions researchers from Latinamerican and African institutions, CIAT collaborators. The Microcomputer Training Laboratory is used for this purpose. During the five years of existence of the Laboratory, Biometry has offered a total of 35 one to two-week training courses, with a total number of 344 National Institutions researchers trained from Latinamerica (260), Asia (24) and Africa (50). An approximate number of 105 participations from CIAT research associates/assistants have benefit from this effort, during a 5-year period.

In the light of the new CIAT, new areas of biometrical expertise in which invited Biometrician Consultants can add useful contributions are foreseen. For example: a) Design and analysis of intercropping experimentation, combining multiple short-cycle crops or combining perennial and short-cycle crops. b) Design and analysis of agro-silvo-pastoral systems. c) Quantitative genetics/population dynamics. d) Econometric techniques in response to a new expected demand from the economists/social scientists.

Statistical/data analysis software for the mainframe computer include: SAS/BASICS, SAS/STATS, SAS/GRAPH, SAS/FSP, SAS/ETS, SAS/IML and SAS/OR from SAS Institute Inc. Raleigh, North Carolina, USA; GENSTAT, from the NAG Algorithm Group, London, England. Microcomputer statistical/data analysis software include MSTAT, from Michigan State University; GLMM, from Louisiana State University; SYSTAT, from SYSTAT Inc. Chicago, Illinois, AGROBASE/4, from Agronomix Software, Manitoba, Canada; MATMODEL from Soil, Crop and Atmospheric Sciences, Cornell University, Ithaca, New York; Lotus 1-2-3 and Dbase III.

Databases of experimental results:

The conceptualization, design and implementation of databases to store crops and pasture research results, require from the "database teams" a clear understanding of the biological nature of the crop and its multiple components. Members of our "database teams" include: a) the System's Analyst, who is the software expert, the designer of the how to efficiently store the datafiles with minimum redundancy, and how to provide interactive access to the data in the most effective manner; b) the Researcher(s), who have a clear understanding of the problem and the purpose of the database; and c) the Biometrician, who has a clear understanding of the data and of the most suitable statistical summaries with which to characterize a given research process.

In terms of database management software technology an important decision was reached in September 1991: that of moving from the 'network' database technology, represented by our previous database management software IDMS/R from Computer Associates Inc., to the 'relational' database technology, represented by ORACLE, from Oracle Corporation. Starting October 1, 1991, ORACLE was acquired as the standard database management software for CIAT's mainframe and microcomputer environment, as a replacement of IDMS/R. Previous problems in design, implementation and utilization of research databases developed on IDMS/R during the past 10 years, such as lack of flexibility for modifying a database design, extremely long data loading times, lack of a user-friendly query tool, lack of a flexible and powerful development tool and lack of micro-mainframe interface, have been greatly solved. Between October 1991 and December 1992, four important ORACLE database applications have been designed, implemented and released to the end-user:

- a) The **Genetic Resources Database**, including passport and characterization data on all CIAT's germplasm collections: beans (40,000 accessions approx.), cassava (4,700 varieties approx.), tropical forages-legumes and grasses (25,000 accessions approx.).
- b) The **Cassava Breeding Database**, including information on parents and crosses, F₁ evaluations, statistical summaries of preliminary yield trials, advanced yield trials and regional trials. This database contains at present research results between 1974 and 1992.
- c) The **Bean Breeding Database**, linked with the Bean Genetic Resources database, includes information on crosses genealogy, advanced lines genealogy, advanced lines evaluations, VEF - EP trials from 1977 - 1991 and, IBYAN trials from 1976 - 1990. Results and statistical summaries from other international nurseries conducted by Bean Scientists are planned to be added in 1993.
- d) The **Tropical Forages Evaluation Database**, linked with the Tropical Forages genetic resources database, contains results from the characterization and early evaluation of grasses and legumes carried-out by Program Scientists since 1978 up to present. Most of this application was implemented in IDMS/R. Its re-design and implementation in ORACLE of a large part of its components, was carried-out in 1992 in a very close collaboration with the Program Leader and Scientists. This effort will continue during 1993.

Details on these four major ORACLE databases are described in this report.

Training courses for end-users in the use of ORACLE databases: The Unit has provided 4-day training courses to users from the following programs: Bean Program breeding sections and Cassava breeding Section (February 92), Genetic Resources Unit (Dec/91 and November/92) and Tropical Forages support personnel (November/92). These short courses were intended to explain content of the databases to the direct users and teach them how to do easy queries to each database.

PERSONNEL RESOURCES. Composition of Biometry Unit teams to support CIAT research Programs/Units

The list of our 1992 personnel is presented at the beginning of this report. In order to provide a multi-skilled support to each CIAT research Program/Unit in the two areas of the Unit responsibility --Biometry and Databases of experimental results-- teams of biometrician/analyst/programmer are assigned to serve the needs of projects within each Program/Unit.

Through this approach, each Program/Unit scientist has access to the range of technical skills represented by personnel within the Biometry Unit. The attached table shows the composition of these teams.

Composition of Unit teams in support to CIAT research programs - 1992

CIAT Program/Unit	Biometrician(s)	TEAM MEMBERS	
		Analyst	Programmer(s)
1. Beans ¹	James García	Germán Serrano	-
2. Cassava ¹	Eduardo Granados Eloina Mesa	Fernando Rojas	Carlos Saa
3. Forages ¹	German Lema Gerardo Ramírez	Arturo Franco	Carlos A. Hernández Rosalba López
4. Rice ¹	Eduardo Granados Germán Lema	Fernando Rojas	Hector F. Ramirez
5. Biotechnology Genetic Resources Virology	Myriam Cristina Duque James García	Arturo Franco Germán Serrano Fernando Rojas	Carlos A. Hernández Carlos Saa
6. New Programs (Savannas, Hillsides, Forest Margins)	Maria Cristina Amézquita Eloina Mesa	-	-
7. Socio-Economics	Myriam C. Duque	Arturo Franco	-
8. Entomologies/IPM	Myriam C. Duque	-	-
9. Sister-institutions Projects (CYMMIT, IBPGR,INTSORMIL).	Eduardo Granados	-	Rosalba López

¹ Excluding Entomology/IPM and Socio-Economics

* The Unit Head provides Biometrical advice and participates in collaborative methodological studies/data analysis projects across germplasm Programs.

2. CONTRIBUTIONS TO BEANS

2.1 Biometry: Methodological and data analysis studies.

Case Study 1:

Statistical methods applied to the analysis of Genotype x environment interaction. Case: Analysis of Bean International Nurseries

J. García, O. Voysest (1992)
J. García, B. Smithson (1991)

Statistical analysis methods for data generated by International Nurseries very commonly involve an **adaptability analysis**, the application of reduction of dimensionality and ordination techniques known as **Pattern Analysis**, such as **Principal Components Analysis** (for continuous or for qualitative variables), **Correspondence Analysis** or **Principal Coordinates Analysis**; and finally, a **Cluster Analysis** for grouping varieties and/or locations. The first methodology is applied to characterize varieties according to their yield and their physiological response to improvement in environmental quality; the second group of methods are applied to reduce the dimensionality of the problem, and place varieties and environments in a reduced space; and the clustering methods are applied to group varieties with similar performance across locations, or to group locations according to their similarity in varietal performance. These methods have proved useful in the identification of good genetic material.

The Biometry Unit has collaborated very closely with the Bean Program in unilocation and multilocation data analysis of two important international nurseries: a) IBYAN (International Bean Yield and Adaptation Nursery), since its creation in 1976, with a total number of 1310 trials performed in 63 countries in the world, through which, a total number of 1399 bean lines produced by the CIAT Bean Program breeders have been evaluated; and b) AFBYAN (African Bean Yield and Adaptation Nursery), with 14 trials reported at present.

The above mentioned statistical techniques have been used in data analysis studies of international nurseries data to better understand the genotype x environment interaction (results appear in the Bean Program Annual Reports 1991 and 1992).

IBYAN data on varietal performance per location as well as relevant statistical summaries by trial, by variety across locations and by superior varieties vs. local checks in each location are stored in the IBYAN database that has been incorporated to the Bean Breeding Database implemented in ORACLE software, and described in section 2.2 of this report.

Case Study 2:

**The application of correspondence analysis
to international nursery data,
Case: Analysis of AFBYAN data
(The African Bean Yield and Adaptation Nursery)**

J. García and B. Smithson (1991)

The Correspondence Analysis technique is a weighted Principal Components Analysis on binary (0,1) or categorical response variables. Then, if the varietal response variables for each location are (0,1) responses (for example, the variety i "was selected or not" at the location), this technique is applicable. It reduces the number of response variables to a reduced-set of independent ones (the "Principal Components") and then, through graphical representation of the locations in terms of the new axis, groups of locations that would select the same variety(ies) are identified.

The purpose of this study was to evaluate the usefulness of the Correspondence Analysis in the identification of promising varieties, by comparing its results with those obtained through Adaptability Analysis and Clustering. The results of 14 AFBYAN trials in which 21 bean varieties were evaluated in terms of yield, were utilized as source of information.

Data analysis methodology used covered four stages:

- a) Adaptability Analysis for the 21 varieties across 14 locations.
- b) Cluster analysis to group varieties according to their yield at each location.
- c) Correspondence Analysis applied to a newly generated data set, with location as row variable, and binary response variables "the variety i was selected or not at the location". For this purpose, a variety was considered **selected** within a given location when its yield was greater or equal to the 85% of the maximum location yield.
- d) Cluster Analysis to group locations based on the 3 first Principal Components resulting from Correspondence Analysis.

Results: The three types of analysis used identified the same group of varieties as the best ones. Groups of locations that would select the same varieties were identified by using the Correspondence Analysis. This study shows that the analysis of binary or categorical responses in multilocal trials is feasible.

Case Study 3:

**Evaluación de parcelas testigos y promedios móviles
para control del error experimental en ensayos de frijol
Proyecto Colaborativo con la sección de Mejoramiento**

El uso de parcelas testigo distribuidas sistemáticamente a través de ensayos donde se prueban gran número de líneas, es una técnica ampliamente usada por los investigadores con el objeto de que los valores expresados por las líneas de prueba sean corregidos por las variaciones microambientales existentes en un ensayo en particular. Dado que el uso de parcelas testigo aumentan el tamaño de los ensayos con consecuencias en aumentos de costos experimentales, el objetivo de éste estudio **es evaluar el uso del método de promedios móviles**, como alternativa al uso de parcelas testigo, **donde el promedio de un número de parcelas adyacentes excluyendo la parcela en consideración es utilizado como indicador del ambiente.**

Para tal fin, se utilizó la información generada por 8 ensayos independientes por tipo de hábito de crecimiento (Hábitos I, II, III y IV) sembrados en condiciones de stress y sin stress por fósforo. El número de líneas por ensayo era alto, oscilando entre 81 y 144. Cada ensayo fue sembrado en Diseños en Lattice, con 2 replicaciones. Cada línea ocupaba un surco. Cada 6 surcos había 1 surco testigo.

La metodología de análisis consiste en expresar el rendimiento original de cada línea como:

- 1- Rendimiento expresado como porcentaje del promedio móvil
- 2- Rendimiento expresado como porcentaje del promedio de las parcelas testigo cercanas
- 3- Rendimiento ajustado usando análisis de covarianza con el promedio móvil como covariable
- 4- Rendimiento ajustado usando análisis de covarianza con el valor promedio de las parcelas testigo más cercanas utilizada como covariable.

y analizar cual de las 4 alternativas permitía una comparación más precisa de las variedades. El criterio de "eficiencia" utilizado fue la **reducción en el error experimental**, del Lattice vs. BCA.

Los resultados preliminares muestran que el uso del promedio móvil utilizado como covariable es más eficiente que las otras alternativas propuestas. El promedio móvil conformado por las ocho parcelas adyacentes que rodean a una parcela en particular, bajo condiciones donde pueda manifestarse mayor variabilidad como es el caso de los ensayos de stress por fósforo mostró reducción en el error experimental del orden del 14% para Hábitos III y IV al 30% para hábitos I y II considerando como base el error experimental obtenido en el análisis del ensayo analizándolo bajo el modelo de un diseño de **Bloques Completos al Azar**.

La reducción en el error experimental considerando el modelo de un diseño Lattice fue más o menos similar al logrado con el análisis de covarianza bajo modelo BCA. El empleo del promedio móvil como covariable y utilizando el modelo del diseño Lattice produjo reducciones adicionales en el error que variaron entre el 0% para materiales de hábito voluble y del 3% al 11% para materiales de tipo arbustivo.

Esta metodología puede ser de gran utilidad en el análisis de cualquier tipo de experimentación, pero principalmente en el análisis de ensayos de evaluación de materiales o líneas en generaciones tempranas donde el **gran número de materiales probados y la poca cantidad de semilla disponible exige realizar comparaciones más precisas** para disminuir la posibilidad de descartar materiales que en un futuro pudieran ser promisorios. Por otro lado, el cuantificar las variaciones microambientales puede permitir disminuir el número de parcelas con materiales testigo lo que implica una reducción en todo tipo de recursos empleados en los ensayos.

Case Study 4:

Diversidad genética en variedades de frijol de grano negro y rojo estimada con base en el coeficiente de parentesco¹

Oswaldo Voysest, María Cristina Amézquita,
María Clara Valencia²

En el proceso de mejoramiento genético siempre está latente el peligro de que haya una pérdida en diversidad genética sea porque unas pocas variedades mejoradas podrían desplazar a las muchas variedades locales o bien porque la base genética del proceso es muy restringida.

El mejoramiento de frijol en América Latina se remonta a la década de 1930; en el lapso que marca esa fecha hasta la actualidad los mejoradores han producido por selección o hibridación, alrededor de 500 variedades, muchas de las cuales han alcanzado gran difusión tanto en el país de origen como en otros. Resultaría aventurado, sin embargo, hacer cualquier inferencia respecto al grado de uniformidad genética en el frijol, sin analizar la situación tanto por país o regiones como por color de grano.

El presente trabajo tiene por objeto analizar la base genética de 2 tipos de frijol representativos de los grandes grupos de germoplasma, el Mesoamericano y el Andino. Se trata de las variedades de origen híbrido de grano negro y grano rojo grande que se cultivan en América Latina. El material de grano negro abarca variedades de México, Guatemala, Costa Rica, Colombia, Argentina y Brasil mientras que el material de grano rojo lo compone exclusivamente variedades desarrolladas por el ICA (Colombia). Se trata de determinar la contribución genética de los progenitores y el grado de participación de los diversos acervos (pool) de genes en el mejoramiento de estas dos clases de frijol.

La similitud genética fue estimada haciendo uso de coeficiente de parentesco calculado con base en el examen de los pedigrís. El análisis de 43 variedades de grano negro muestra que en el desarrollo de todas estas variedades, intervinieron 63 materiales genéticos a los cuales nos referiremos como antecesores. Esto que aparentemente podría ser un indicio de una amplia base genética en las variedades de grano negro de origen híbrido, en realidad no lo es, pues la gran mayoría de las líneas empleadas como progenitores pertenecen a las razas de origen mesoamericano sumando un aporte cercano al 90% de los genes. La variedad Jamapa es entre todas, la que mayor contribución genética ha hecho (9%).

Para el desarrollo de las 16 variedades que integran el grupo de los frijoles andinos de grano rojo, se usó como progenitores en el transcurso de 35 años, 19 genotipos; 9 de ellos han contribuido aproximadamente con el 80% de genes. Ocho de los nueve progenitores que han hecho el 80% de contribución genética, con excepción de Italia 5 del cual no se tiene mayores referencias, son frijoles del grupo de origen andino y todos pertenecen a la raza Nueva Granada.

En conclusión, un examen de los pedigrís de las 43 variedades de origen híbrido tipo mesoamericano y 16 de tipo andino confirman apreciaciones acerca de la estrecha base genética sobre la cual ha operado el mejoramiento genético del frijol. En ambos casos, el mejoramiento genético de los frijoles negros y rojos se ha confinado a hibridaciones dentro de los grupos específicos a los que estos frijoles pertenecen, mesoamericano y andino, respectivamente.

¹ Trabajo presentado en RELEZA III, Cochabamba, Bolivia, 17-20 junio, 1992

² Investigador, biometrista y asistente estadístico, CIAT, A.A. 6713, Cali, Colombia

2.2 The Bean Breeding Database

Present state of the Bean Breeding Database:

This database represents a continuous and collaborative effort between the Bean Program breeders and genotype x environment specialists, who decide on the key information that needs to be stored and the Biometry Unit. At present, the "Bean Breeding Database" contains research results from 1976 to 1991, and includes:

- a) **Crosses:** 394 with their code (which includes crossing criteria) and their parent's names.
- b) **Advanced Lines:** Includes data on 14445 advanced lines generated by the Bean Program breeders between 1978 and 1990.

Descriptors include: line code and genealogy.

- c) **Evaluation of advanced lines in research stations**
 - VEF Data on 12901 advanced lines tested by the Bean team between 1978 and 1990.
 - EP Data on 4571 advanced lines tested in Palmira, Popayán and La Selva chosen as representative ecosystems sites.

Includes statistical summaries of agronomic and disease-resistance reaction to rust, antracnosis, angular leaf spot, common bacterial blight, and Ascochyta.

- d) **International nurseries**
 - IBYAN Data stored between 1976 and 1990 coming from 62 countries around the world. Includes experimental results from 1310 trials in which 3999 CIAT's advanced lines have been evaluated.

The "Bean Breeding database" is now fully operational. On-line applications have been developed to query and update the database. End-users training courses have been given to teach them how to update and make easy queries.

Pathology evaluation nurseries

At present, a full design of the database has been done to store data on international nurseries of Rust, Antracnosis, Angular Leaf Spot and Common Bacterial. Data analysis and data loading phases will start during 1993.

3. CONTRIBUTIONS TO CASSAVA

3.1 Biometry contributions: Collaborative methodological and data analysis studies

Case Study 1:

Sensibilidad a cambios ambientales de 15 genotipos de yuca en 14 ambientes: Estudio colaborativo con la Sección de Genética/interacción genotipo x ambiente

C. Iglesias, E. Mesa

Objetivos: a) Determinar diferencias en la sensibilidad del rendimiento de raíces frente a cambios ambientales. b) Determinar importancia de la variación temporal y espacial cuando se seleccionan sitios de evaluación para Mejoramiento de Yuca.

Información considerada: Se consideraron 7 clones elites y 8 accesiones de ensayos de rendimiento conducidos en Colombia en las localidades de Palmira, Media Luna y Carmen de Bolívar de 1986 a 1990.

Las variables consideradas en el estudio fueron: Producción de raíces (ton/ha), número de raíces comerciales por planta, índice de cosecha, altura de planta (cms), longitud de tallo (cms), índice de ramificación, contenido de materia seca (%) y contenido de cianuro (escala de 1 a 5).

Metodología:

1. Para determinar la sensibilidad de los materiales a cambios ambientales, se utilizó el método de regresión conjunta modificada ("Modified Joint Regression Method") propuesto por Digby (1979), en el cual se propone el modelo:

$$Y(i,j) = T(i) + B(i)*THETA(j) + E(i,j)$$

$i = 1,2,...,V$ genotipos

$j = 1,2,...,L$ ambientes

donde:

$Y(i,j)$: es el rendimiento del genotipo i en el ambiente j
 $T(i)$: media esperada para el genotipo i
 $THETA(j)$: efecto del ambiente j
 $E(i,j)$: termino de error

2. La caracterización de ambientes se hizo con base en el análisis de componentes principales considerando el promedio del ensayo para cada una de las variables de respuesta y así reducir el número de variables que posteriormente entran para la agrupación de ambientes.

Resultados:

Los clones CG 915-1, CM3306-4 y BRA12 se adaptan bien a ambientes favorables, mientras los clones CM3372 y CM355-6 pueden ser recomendados para ambientes de bajo potencial.

Los clones CG1141-1, CM3299-4, CM3306-9 y BRA191 tienen altos rendimientos medios y no presentan fuertes cambios a variaciones ambientales.

Para HCN la combinación deseable es rendimiento medio-bajo y baja sensibilidad como BRA191, COL1505, COL2215, CM3306-4 y CM3772-4.

En la caracterización de ambientes se consideraron tres componentes que explicaron el 82% de la variabilidad total:

Comp. # 1: Refleja relación general entre variables con baja importancia para índices de cosecha y ramificación.

Comp. # 2: Valores altos de raíces comerciales e índice de cosecha y bajos índices de ramificación.

Comp. # 3: Valores altos de índice ramificación.

Con base en estas tres componentes se formaron tres grupos que explicaron el 75% de variación.

Grupo # 1: Palmira 1988
Palmira 1990
Media Luna 1987
Carmen de Bolívar 1987

Grupo # 2: Carmen de Bolívar 1986
Carmen de Bolívar 1988
Carmen de Bolívar 1990

Grupo # 3: Media Luna 1986
Media Luna 1988
Media Luna 1989
Media Luna 1990

Grupo # 4: Palmira 1987
Palmira 1989

El grupo 4 lo conforman ambientes con alto potencial en producción de materia seca; los otros grupos corresponden a ambientes de bajo potencial y sus diferencias radican en el número de raíces comerciales, altura de planta e índice de ramificación.

El grupo 1 lo conforman ambientes de menor rendimiento, menores índices de cosecha, menores alturas y menor número de raíces comerciales. Los grupos muestran que hay una mayor variación entre sitios que entre años.

El Informe Anual 1992 del Programa de Yuca explica con mayor detalle estos resultados.

Case Study 2:

Indice para selección de variedades de yuca promisorias: Un estudio colaborativo con la Sección Genética/interacción Genotipo x ambiente

C. Iglesias, E. Mesa (en progreso)

Objetivo: Establecer índices de selección para cuatro sitios de evaluación: Palmira, Costa Atlántica, Popayán y Llanos Orientales.

Información considerada: Se consideró la información correspondiente a ensayos de campo de observación y de rendimiento para Palmira, Media Luna, Popayán, Carimagua y La Libertad de 1980 a 1990.

Las variables consideradas en el estudio fueron: Producción de raíces (ton/ha), número de raíces comerciales por planta, índice de cosecha, altura de planta (cms), longitud de tallo (cms), índice de ramificación, contenido de materia seca (%) y contenido de cianuro (escala de 1 a 5).

Metodología: a) Análisis de regresión por pasos (stepwise) de rendimiento en función de las restantes variables, con el fin de determinar las variables que entran en cada caso. Este análisis se hace para cada localidad y cada año. b) Agrupación de ambientes con base en variables ambientales (precipitación y variables de suelo), para luego determinar el peso de cada variable. c) Análisis de factores con el fin de agrupar variables en conceptos de tal forma que el promedio de coeficientes estandarizados determine el peso o ponderación de cada variable.

Resultados: (en progreso)

Case Study 3:

Evaluación de calidad de la Colección Central de Yuca en CIAT

C. Wheatley, J. Orrego, T. Sanchez, E. Granados

Objetivo: Clasificar la Colección Central de Yuca según su calidad.

Metodología: La Colección Central de Yuca está compuesta por 630 clones escogidos por su origen geográfico, diversidad morfológica, actividad de banda de esterasa, inclusión en estudios de Biotecnología, cultivares nativos más comunes y por híbridos éliticos. Está sembrada toda la colección en parcelas de 6 plantas por clon en una distancia de 1x1 metro. Debido a la disponibilidad se cosecharon 565 clones con muestras de aproximadamente 3 kgs. En el laboratorio de Calidad después de pelar las raíces se determinó para cáscara y parénquima separadamente la materia seca (%) y el contenido de cianuro en base seca. El análisis de Amilosa (%) se realizó a 503 clones.

Para los análisis estadísticos se transfirieron a escala logaritmo natural los contenidos de cianuro debido a la composición de la colección que tiene una gran cantidad de clones con bajo cianuro, lo que hace la distribución altamente asimétrica.

Previo a la clasificación se redujo el número de variables (5) a un número menor (3) mediante la

técnica de análisis de factores con la opción de componentes principales. Esta reducción es debido a la alta correlación presentada entre los análisis en cáscara y parénquima. El factor 1 correlacionó positivamente con el logaritmo de cianuro en parénquima y logaritmo de cianuro en cáscara. El factor 2 correlacionó con materia seca (%) en parénquima y cáscara. El factor 3 explicó bien amilosa (%).

El análisis de conglomerados sobre los 3 factores (no redundantes en información) mediante agrupamiento jerárquico con el criterio de WARD facilitó la clasificación en 10 grupos de diversas características.

Para ampliación de los resultados remítase a Informe Anual Programa de Yuca - Sección Utilización.

Case Study 4:

A methodology for the statistical analysis of electrophoretic patterns
Case study: Biochemical differentiation of populations of the mite Amblyseius limonicus Garman and Mc. Gregor (Acarina: Phytoseiidae)

M.C. Duque, Ma. E. Cuellar and Ann Braun

This study started during 1991 and was completed this year.

In order to determine an effective strategy for the biological control of a serious cassava pest -the mite *Mononychellus tanajoa* (Bondar) (Acarina: Tetranychidae) ("acaro verde de la yuca")- it is necessary to clearly characterize its natural enemies, both in terms of their ecologic and biological behavior. Among them, the mite *Amblyseius limonicus* Garman and Mc. Gregor (Acarina: Phytoseiidae) is known as its most important predator.

The present study was carried-out to make a biochemical differentiation of populations of the mite *A. limonicus* and to test the hypothesis that variability observed between populations of distinct geographic origin may be associated with differences biochemical patterns between them.

222 samples of *A. limonicus* collected in 16 distinct sites of Tropical America were submitted to electrophoretic analysis utilizing the isoenzymes GOT and MDH. The presence or absence of 70 electrophoretic bands (representing 70 distinct proteins or protein fractions in the *A. limonicus* DNA) were recorded for each one of the samples. In this way, the resulting data set was constituted by 222 rows (samples) and 70 binary (0,1) response variables.

For the statistical analysis of the electrophoretic binary results, a Correspondence Analysis was applied. This technique, a reduction-of-dimensionality technique for categorical variables, similar to the Principal Components Analysis, finds a low-dimensional graphical representation of the 222 samples. In this way, visual groups of samples are formed, being these groups interpreted as possible distinct populations of the mite *A. limonicus*.

As a result, six distinct groups were identified in a 3-dimensional graphical representation (a reduction of the 70 initial binary response variables) as illustrated in figure 1. The six groups corresponded to samples of *A. limonicus* of similar geographic origin. The hypothesis of

association between geographic origin of *A. limonicus* and their distinct biochemical composition was then accepted.

In order to verify the results, two experiments were conducted: a) the first to evaluate whether the dietary composition of the various populations of *A. limonicus* was similar, and b) the second one, to evaluate whether their reproductive performance.

3.2 The Cassava Breeding Database. Its extension to include research results from other disciplines

F. Rojas, E. Granados, N. Marín, from the Biometry Unit
C. Hershey, C. Iglesias and team, from the Cassava Program

The purpose of the Cassava Breeding Database is to store in an organized way, maintain and make interactively available to cassava researchers, all the information related to the collection, generation, and testing of cassava germplasm, including research results attained by the Cassava Breeding Section and the CIAT's Genetic Resources Unit during a 12-years period: 1978-1991.

The Cassava Breeding Database includes at present the following information:

1. Germplasm Bank: 4,650 varieties with
 - . passport data (collection site descriptors, collection date, origin, local names, etc.)
 - . morpho-agronomic characteristics (19 descriptors on 4081 varieties)
 - . electrophoretic characterization (presence or absence of electrophoretic bands 4251 varieties).
2. Crosses: 14,142 crosses with their names and their parent's names.
3. Progeny Evaluation trials: F₁ evaluations (non-replicated trials. Results correspond to original observations per material).
4. Advanced trials: Data stored correspond to statistical summaries of replicated experiments for the evaluation of advanced cassava material, conducted by the Cassava Breeding Section between 1978 and 1992, with an average of 60 experiments/year, as follows:
 - . Observational trials: 124 experiments
 - . Preliminary yield trials: 91 experiments
 - . Advanced yield trials: 143 experimentsFor each one of these trials, the following information is stored:
 - . experimental site descriptors (10 variables)
 - . morphological response variables (20 variables)
 - . agronomic response variables (6 variables)
 - . pests/diseases scores (variable number of response depending on the trial)
5. Regional Trials: Statistical Summaries of 421 replicated trials conducted in 5 cassava ecozones in Colombia between 1978 and 1990. For each regional trial the following information is stored in the database:
 - . experimental site descriptors
 - . agronomic responses per variety
 - . pest/diseases scores per variety
 - . Statistical Summaries of the trial for the most important response variables.

6. Elite Clones: 284 clones are stored, with the following descriptors:
 - . clone code
 - . parent's code
 - . principal (and second principal) adaptation zone
 - . yield potential (mean yield in the principal adaptation zone)
 - . quality descriptors
 - . resistance to diseases score (5 diseases)
 - . morphological descriptors (6 descriptors)
7. Seed inventory and international shipments according to seed type: (50 shipments/year approx.)

a) dry stakes	b) immature stakes
c) in-vitro	d) sexual seed

Now, the Cassava Breeding Database is complete, fully operational and ready to be made available to the Cassava Breeding Section and other cassava researchers by the end of November 1992. It can be accessed through a terminal (or connected micro) of the IBM 4361 mainframe computer or through a PC with 4Mb of memory/40Mb hard disk and equipped with ORACLE tools.

Applications development:

Objetivos: a) Facilitar a los usuarios el acceso a la información del Programa de Yuca y de Recursos Genéticos de Yuca, almacenada en la Base de Datos ORACLE; b) automatizar procedimientos que actualmente se efectúan manualmente, agilizando las labores del personal encargado; c) Mantener el manejo integrado de toda la información del Programa.

Usuarios: Investigadores del Programa de Yuca, Unidad de Recursos Genéticos, Unidad de Biometría, Otros programas o áreas interesadas

Características de las aplicaciones:

- . Módulos completamente integrados que facilitan la consulta, modificación y actualización de la información.
- . Control centralizado de las diferentes clases de acceso y privilegios que tienen los distintos usuarios del sistema.
- . Facilidad de producir reportes totalmente paramétricos en todas las aplicaciones.
- . Capacidad de integrar al sistema reportes o consultas Ad-Hoc, elaboradas por los usuarios finales.
- . Reportes producidos en la impresora del mainframe o en la impresora local del microcomputador, dependiendo del ambiente en que se este ejecutando.
- . Utilización de las diferentes herramientas de ORACLE en un ambiente integrado.
- . En la actualidad el sistema puede correr en 2 ambientes: desde una terminal del mainframe en una máquina virtual con 4MB de memoria; o desde un microcomputador equipado con 4MB de memoria y el software "ORACLE-TOOLS".
- . El sistema puede correr en cualquier plataforma de hardware que este soportado con un manejador de Base de Datos ORACLE (LAN, Workstations, etc....); y bajo cualquier sistema operativo.
- . Flexibilidad y capacidad de crecimiento.

Aplicaciones que componen el sistema:

Las distintas aplicaciones permiten la consulta, actualización y modificación de la información, así como la generación de los distintos reportes. Estas aplicaciones fueron desarrolladas para facilitar la consulta, actualización y modificación de datos de:

- Banco de germoplasma
- Pasaporte
- Envíos de germoplasma
- Clones Elite
- Inventario y Movimiento de semillas
- Banco In-Vitro (Conservación, Indexación, Ubicación)
- Cruzamientos e hibridaciones
- Selecciones (Ensayos F1, F1/C1)

En 1993 se continuará desarrollando la segunda parte de aplicaciones que comprenden Ensayos Avanzados (Campos de observación, Ensayos preliminares de rendimiento, Ensayos de rendimiento y Pruebas regionales).

4. CONTRIBUTIONS TO TROPICAL PASTURES

4.1 Biometry: Methodological contributions to Tropical Forages Research

The special nature of tropical pastures research needs to be recognized. When compared with classic agricultural research carried-out with short-cycle crops whose final product is obtained at harvest in small plots under standard experimental designs, tropical pastures research presents many additional complexities. Tropical pastures research is of long-term nature as it deals with a perennial crop. Since the final products of a pasture are milk, meat, wool or other animal products, the pasture researcher has to recognize that small-plot clipping trials and medium and large-scale grazing experiments are complementary.

Under the evaluation scheme used by the CIAT's Tropical Pastures Program, a grass or legume accession is first submitted to small-plot agronomic trials to evaluate its adaptation to soil, climate and biotic conditions and its biomass production potential; then, selected grass-legume associations are submitted to agronomic evaluations under grazing to study their compatibility and persistence under the animal influence; advanced materials are then submitted to large-scale grazing trials to measure animal productivity, the latter expressed in terms of weight gain of young steers, milk production capacity of a dairy herd, reproductive performance of breeding herds, or mixed beef and milk production under double-purpose production systems at farm level.

In small-plot agronomic experiments, standard experimental designs are utilized, in which the effect of one or more experimental factors at various levels can be studied under replicated factorials for example. However, response variables need to be analyzed as repeated measurements within season, and the statistical analysis may involve a response curve fitting by season and multivariate comparisons of regression parameters among treatments; or they could be expressed as summary indicators over the experimental period.

In pasture evaluation experiments under grazing, experimental designs tend to be simple, but additional sources of variation on the pasture response need to be considered for data analysis purposes. Besides 'soil', 'year', 'season within year' and 'pasture quality across time', 'animal variability' (sex, age, origin, condition) is of great importance.

Milk production trials, given the high cost of experimental animals, represent a very interesting research area for the utilization of change-over designs, which require less experimental units to attain similar levels of significance, when compared to standard continuous designs.

On the other extreme, large-scale pasture evaluation experiments with commercial breeding herds, conducted to measure reproductive efficiency in beef cattle, although of simple design (as RCB or CR) and oriented towards a direct adoption by producers, require the use of sophisticated and often complex statistical methods for efficient data manipulation and analysis. These facts make tropical pastures research an extremely challenging field of work for biometricians.

Now, under the newly created Tropical Forages Program, new challenges will appear. They will be mainly associated with the Program need to evaluate new germplasm alternatives in new agro-ecosystems. Eg.: tree legumes, alternative uses for herbaceous legumes, associations between forages and short-cycle crops. New challenges will include new experimental design and data analysis methodologies.

Support to RIEPT in the management and statistical analysis of its information:

Since its creation in 1979, the RIEPT* assigned the CIAT Tropical Pastures Program the responsibility to centralize and make available to network members all the information generated by the network. Since then, the CIAT Biometry Unit has collaborated very closely with the Tropical Pastures Program and now with the Tropical Forages Program, in the organization, storage and statistical analysis (by site, by country, by ecosystem or across-ecosystem data analysis) of RIEPT-generated research results. Up to now, 251 agronomic-trials (ERA and ERB) and some 10 grazing trials have been statistically analyzed and their results stored in the RIEPT database. Multilocal analysis to identify promising germplasm by agro-ecosystem have been performed using RIEPT information generated between 1979 and 1992. Recent studies include: "The analysis of germplasm evaluated in the Humid Tropics" presented at the 1990 RIEPT meeting in Pucallpa, Perú in November 1990 and published in its memories; "The analysis of forage germplasm for Central America" published in the document " "; and "The analysis of germplasm evaluated in the savanna ecosystem" presented at the 1992 RIEPT meeting in Brasília, Brazil (Nov. 22-26). During this meeting, the CIAT Biometry Unit presented an overview talk "Use of the information generated by RIEPT (1979-1992)" and offered to RIEPT members the microcomputer version of the RIEPT database, (for a more detailed description refer to the section entitled the RIEPT database in this report.

Some selected examples of data analysis studies carried-out between the Biometry Unit and the Tropical Forages Program are summarized in this report.

Case Study 1:

Range of adaptation of *Stylosanthes guianensis*, cv. Pucallpa in the american tropical rain forest ecosystem.

M.C. Amézquita, J.M. Toledo, and G. Keller-Grein
(Published by Tropical Grasslands, Sept. 1991).

The purpose of this study was to define the range of adaptation of *Stylosanthes guianensis* CIAT 184, released in 1985 as cv. Pucallpa by IVITA (Instituto Veterinario de Investigaciones Tropicales y de Altura) and INIPA (Instituto Nacional de Investigación y Promoción Agropecuaria) in Perú. Data from 32 RIEPT (International Network for Tropical Pastures Evaluation) type B trials conducted in the American Humid tropics between Mexico and Bolivia were used for this study (table 1, fig.1). Statistical methodology covered four stages:

- a) The definition of agronomic indicators of rapidity in establishment and biomass productivity.
- b) The identification of environmental parameters that would affect establishment and production of the cultivars. Stepwise regressions were carried-out with the agronomic indicator as the dependent variable in each regression, and a reduced-set of non-correlated environmental parameters (soil, climate and location) as independent variables. Those environmental parameters found significant in the regression were considered important sources of variability on the agronomic performance of cv. Pucallpa.
- c) The identification of groups of sites with similar environmental conditions for the cultivar performance. A hierarchical Cluster Analysis technique with Ward's minimum variance method was used for this purpose.
- d) The description of agronomic performance of Cultivar Pucallpa in each group of environments.

This study shows that cv. Pucallpa is tolerant to anthracnose under a wide range of soil, climate and locations; the cultivar is better adapted to low altitudes (<850 m.a.s.l.), on soils that are acid (pH 5.0), which have low levels of organic matter (<3.4%), are moderately sandy (18-56% sand), and which have rainfall accumulated in 12 weeks > 800mm; at higher altitudes (>1000 m.a.s.l.), the cultivar appears to respond to higher levels of organic matter (tables 2,3 and 4).

Case Study 2:

A data analysis methodology for the evaluation of large germplasm collections

Case study: Evaluation of the CIAT *Brachiaria* collection in Brazil

Cacilda do Valle*, and M.C. Amézquita

(To be presented to the XVII International Grassland Congress in New Zealand)

This study, initiated during 1991, concluded this year. It used as data source 3-year experimental results of the agronomic evaluation of 194 accessions of *Brachiaria* species, carried-out by EMBRAPA, in Campo Grande, Brazil in small plots, under a split-plot design. During these 3 years 18 evaluations were performed: 14 during the rainy season and 4 during the dry season. Let us present a short summary.

The agronomic evaluation of forage germplasm collections in the Tropics involves periodic measurements of plant responses that cover the most contrasting seasonal periods of the region of interest. In order to characterize an accession, summary indicators by season or dry-rainy season relations need to be computed. As the resulting number of plant response indicators is normally very large and significant correlations between them may exist, reduction-of-dimensionality techniques need to be applied to reduce them to a minimum number of non-correlated ones. The present study illustrates these aspects. It presents a methodology for data analysis of the agronomic evaluation of a large germplasm collection. Biomass production (total, leaf, stem) and regrowth capacity were periodically measured. Additionally, observations on resistance to insects Spittle bug, diseases, and plant vigor were made periodically. Early flowering capacity was recorded only once during the experimental period.

Methodology: A set of eleven highest priority summary indicators were computed as functions of the original measurements. They were: 1.

1. Annual accumulated total dry matter (kg/ha/year) (ATDM).
2. Accumulated total dry matter during the dry season, expressed as percentage of annual total dry matter $((TDM_{dry}/ATDM) \times 100)$.
3. Annual accumulated leaf dry matter (kg/ha/year) (ALDM).
4. Accumulated leaf dry matter during the dry season expressed as percentage of annual leaf dry matter. $((LDM_{dry}/ALDM) \times 100)$.

* EMBRAPA researcher. CIAT Visiting Scientists during 1991

- 5,6. Percentage of leaf dry matter from total dry matter
 - . during the dry season ($PLDM_{dry}$)
 - . during the rainy season ($PLDM_{rainy}$)
- 7,8. Leaf-stem relation, based on dry matter
 - . during the dry season ($LDM_{dry}/SDM_{dry} \times 100$)
 - . during the rainy season ($LDM_{rainy}/SDM_{rainy} \times 100$)
- 9,10. Regrowth capacity (ordinal 0-6 scale)
 - . during the dry season (RC_{dry})
 - . during the rainy season (RC_{rainy})
- 11. "Index of Spittle bug resistance", calculated as the percentage of a score ((0='the plant was resistant') 1='the plant was not resistant') assigned to a given accession among the 14 rainy season scores.

A Factor Analysis, with varimax rotation method, was applied to these 11 indicators. Based on the resulting reduced number of factors, a Ward's minimum variance Cluster Analysis was performed to classify accessions with similar agronomic characteristics within species.

Results: As a result, the three first factors -explaining 87.8% of the total variation- were selected as a reduced set of non-correlated groups of indicators. One indicator from each one of the factors, was chosen to represent the factor. These were: a) Annual accumulated leaf dry matter (kg/ha/year); b) Leaf-Stem relation, during the dry season (%); and c) Index of resistance to spittle bug (expressed as a % of zeros among 14 evaluations).

The Cluster Analysis helped identify 22 promising accessions, out of which 9 were selected to advance for grazing studies: 6 from *B. brizantha*, superior to the standard cultivar cv. "Marandú"; 1 from *B. decumbens*, superior to cv. "Basilisk"; 1 from *B. humidicola*, and 1 from *B. jubata*. (See Tables 1 and 2).

Table 1: *Brachiaria* species evaluated in Campo Grande, Brazil

OVERALL DESCRIPTIVE STATISTICS

Specie	No. of accessions	Accumulated Leaf Dry Matter (kg/ha/year)	Leaf-Stem relation during the dry season	Index of resistance to Spittle bug (% of zero score among 14 wet season eval.)
<i>B. brizantha</i>	96	9324	1.44	63.2
<i>B. decumbens</i>	35	4229	0.81	67.9
<i>B. humidicola</i>	21	5843	0.93	76.9
<i>B. jubata</i>	11	4085	1.19	67.5
<i>B. ruzisiensis</i>	20	3809	1.43	55.0
<i>B. arrecta</i>	6	2096	0.58	77.4
<i>B. dictioneura</i>	2	9391	-	82.2
<i>B. negropedata</i>	1	4004	-	78.6
<i>B. adspersa</i>	1	2743	0.7	100.0
Total	193	5058	1.0	74.3

Table 2: Multivariate evaluation of the CIAT *Brachiaria* collection (193 accessions) in Campo Grande, Brazil for a 3-years period.

PROMISING ACCESSIONS ³				
Accession Identification		Accumulated Leaf Dry Matter	Leaf-Stem relation during the dry season	Index of resistance to Spittle bug
CIAT #	EMBRAPA #	(Kg/Ha/year)		(% of zero score among 14 wet season evaluation)
<i>B. brizantha</i>				
16288	B132*	19234	1.36	85.7
16467	B166*	17021	1.18	71.4
16306	B138*	16542	1.10	85.7
16316	B144*	14971	1.48	71.4
16473	B89 *	14268	1.06	71.4
	B163	13864	1.19	57.1
	B73	13823	1.41	57.1
	B65	11838	1.65	50.0
	B52	10648	1.51	42.9
	B51	10351	1.58	42.9
	B137	10289	1.41	42.9
	B59	10252	1.30	35.7
	B136	10127	1.35	42.9
<i>B. decumbens</i>				
16488	D1*	12892	1.02	57.1
606	D62	9157	1.32	50.0
6699	D70	8206	1.46	71.4
<i>B. humidicola</i>				
16886	H13	8226	1.40	78.6
26155	H18*	8027	1.19	85.7
	H25	7318	1.18	85.7
<i>B. juvata</i>				
26237	J13*	7325	1.10	64.3
	J3	4748	1.51	57.1
<i>B. ruzisiensis</i>				
	R103	5105	2.67	42.9

³

High Annual Leaf Dry Matter, High Leaf-Stems relation during the dry season and low or medium incidence of spittle bug.

* Out of these 22 accessions, the 8 accessions with an * were identified to advance for grazing studies.

Case Study 3:

Implementación de Modelos de Evaluación Económica - Proyecto colaborativo con Economía de Forrajes -

Objetivo: Desarrollar instrumentos metodológicos para evaluar el impacto económico del uso de Nueva Tecnología.

Resultados: Se diseñó e implementó un modelo en Lotus 1-2-3 llamado MODEXC (Modelo de excedentes económicos) que permite el cálculo de los excedentes y la rentabilidad social que generan las Nuevas Tecnologías.

Utilización del Modelo:

- Plan Estratégico CIAT: véase: Janssen W. et al (1991). "CIAT's Commodity Portfolio Revisited: Indicators of Present and Future Importance" (Pag. 21-59).
- Rivas, L., Toledo, J.M. y Sanint, L.R. (1991). "Potential Impact of the use of Pastures Associated with Crops in the Tropical Savannas of Latin America". Pag. 78-91 in: Trends in CIAT Commodities 1991: Adoption and Impact.
- Curso de Adiestramiento en Economía, Octubre-Noviembre, 1991. Sección de Evaluación Económica Ex-Ante.

Usuarios de MODEXC fuera del CIAT:

- INTA, Argentina, Centro Regional Chaco-Formosa, Area de Recursos Humanos, Control y Evaluación. Ing. José L. Russo.
- Jarvis L. y Seré C. (1991). The Role of Risk and Dynamic Supply and Demand in Ex-Ante Estimates of the Return to Agricultural Research: Improved Pastures for the Latin American Tropics. Submitted to the XXI International Conference of Agricultural Economists. August 22-29. Tokyo, Japan.
- CENICAÑA, Cali, Colombia, Sección de Socio-Economía.
- Universidad Nacional de Colombia, Seccional Palmira.

4.2 The Tropical Forages Database. The RIEPT database.

A. Franco, E. Mesa, M.C. Amézquita and C.A. Hernández, from the Biometry Unit
and
Luis H. Franco and Tropical Pastures Program Leader and Scientists, from the Forages Program

Genetic Resources Database:

The purpose of the genetic resources database is to store in an organized way, maintain and make interactively available, all the information related to tropical pastures genetic resources that have

been generated, collected and handled by the CIAT Genetic Resources Unit or by the Germplasm Evaluation Section of the CIAT Tropical Pastures Program between 1978 and 1991.

The tropical pastures genetic resources database includes information on 22,818 accessions of grasses and legumes. The database is organized in seven sub-systems, according to the type of information being stored, as follows:

1. Passport data (22,818 accessions)
2. Short-term seed inventory (20,220 accessions)
3. Long-term seed inventory (3,872 accessions)
4. Seed international shipments (200 shipments to Latinamerican, Asian and African countries)
5. Seed multiplication, at greenhouse and at field level (historical records of all material submitted to seed multiplication: 4,815 multiplication events)
6. Inventory of materials stored as "herbario" (9,699 accessions)

Germplasm evaluation Database:

The purpose of this database is to store, maintain and make interactively available to tropical pastures scientists, result summaries of the different research projects conducted between 1978-1991 on forage germplasm evaluation by the CIAT Tropical Pastures Program. These projects cover research work carried-out by the following groups of sections: a) Germplasm Morpho-agronomic characterization of germplasm (6,566 accessions, evaluated through 128 experimental projects); b) Agronomy: small-plot agronomic evaluations (30 projects concerning 1050 accessions) and agronomic evaluation under grazing (40 projects); c) Plant-protection experiments: phytopathology (50 projects); fungal and bacterial collection (descriptive characteristics on 4,500 fungus and 200 bacteria); d) Soil-plant and soil microbiology research projects (30 soil-plant nutrition experiments, 50 microbiology research projects and the Rhizobium strains collection containing information on 5,037 strains); e) Pasture quality evaluations (8 projects on pasture nutritional characteristics and 8 palability trials at Quilichao Research Station); f) Pasture productivity evaluations in terms of animal production parameters (6 long-term projects conducted in Carimagua and Quilichao; g) Production systems: 3 long-term experiments on reproductive efficiency of beef cattle conducted in Carimagua (1972-1988); 1 experiment on early weaning and 1 on methodological research concerning animal categories.

The germplasm evaluation database also contains the information generated by three international networks in which the CIAT Tropical Pastures Program participates as an active member. These are: RIEPT (International Network for Tropical Pastures evaluation, 1979-1992), Centrosema International Network (1989-1992), and WECAFNET (West and Central Africa Forage evaluation Network, which started on October 1990).

Additionally, the Germplasm Evaluation Database contains a sub-system called "Research proposals follow-up", that includes summaries of 720 research proposals and their follow-up, carried-out by Tropical Pastures Program researchers between 1978 and 1991. This very valuable historical information provides a feed-back to the Tropical Pastures Program Leader on the type of research topics being addressed.

The RIEPT Database (1979-1992):

The RIEPT database includes environmental and experimental information on 241 trials conducted (and reported to CIAT) by RIEPT between 1979 and 1992 in 18 countries of Tropical America between Mexico and Bolivia including Caribbean countries. The 241 trials are composed by 45 small-plot adaptability trials (ERA) and 196 small-plot agronomic trials (ERB). Information concerning grazing trials (ERC and ERD) is being collected and organized to be stored in the future. Additionally, the RIEPT database contains information on prices on beef production inputs/products at all RIEPT sites. The RIEPT database, resident in the CIAT IBM 4361 mainframe computer, is being now distributed to Latinamerican National Programs in diskettes to be consulted via DBASE III.

A microcomputer version of this database, together with a training video, has been developed between the Tropical Forages Program and the Biometry Unit. The microcomputer version, requiring a minimum hardware capacity of 512 K of memory and no hard disk, was made available to RIEPT member institutions at the 1992 RIEPT meeting, held in Brasilia, during November 22 and 26, 1992.

Publications based on Tropical Pastures databases:

- "Catálogo de germoplasma de especies forrajeras", (1987) (3 Volumes)
- "Colección de Centrosema del CIAT", (1986).
- "Catálogo mundial de germoplasma de Centrosema", (1989).
- "Catalogue of Rhizobium strains for tropical forage legumes", (1985, 1986, 1987 and 1988)
- "RIEPT - Resultados 1979-1982", II Reunión.
- "RIEPT - Resultados 1982-1985", III Reunión.
- "RIEPT - Resultados RIEPT-Amazonía", (1990)
- "RIEPT - Análisis sobre localidades y evaluaciones de germoplasma en el Trópico Húmedo", (1990).
- "RIEPT - Análisis sobre localidades y evaluaciones de germoplasma en Centroamérica y Caribe" (1991).
- "RIEPT - Recursos disponibles, demanda de servicios y logros en la RIEPT". Contribución de las pasturas mejoradas a la producción animal en el trópico", (1989).
- "RIEPT - Evaluación del comportamiento de ecotipos dentro y a través de ecosistemas", (1985).
- "RIEPT - Análisis de precios de productos e insumos ganaderos en localidades de la RIEPT", (1984, 1985, 1986, 1987, 1988, 1990).
- "Trends in CIAT commodities", (1982,..., 1991).
- "RIEPT - Base de datos estadística. Información y opciones para su utilización", (1987).
- "La colección de forrajeras tropicales del CIAT", (1991) (3 volumes).
 - I. Catálogo de germoplasma de Asia Suroriental
 - II. Catálogo de germoplasma de Venezuela
 - III. Catálogo de germoplasma de Centroamérica, México y el Caribe
- "Utilización de información de ensayos multilocacionales de evaluación de germoplasma. Organización de bases de datos", (1988).
- "RIEPT - Análisis sobre localidades y evaluaciones de germoplasma en Sabana" (1992).
- Base de datos estadística RIEPT versión microcomputadores (1992).
- Red internacional de evaluación de pastos tropicales RIEPT. Primera reunión Sabanas (1992).

5. CONTRIBUTIONS TO RICE

5.1 Biometry contributions: Methodological and data analysis studies

Case Study 1:

A methodology to determine the minimum evaluation period for disease-resistance characterization in rice

E. Guimaraes, M.C. Amézquita, G. Lema and F. Correa

This study which started during 1991 was completed this year.

Santa Rosa Experimental Station, located at the eastern Colombian savannas (at 333 m.a.s.l., 25°C, 66-87% relative humidity) is used by the CIAT Rice Program as a hot spot site for screening breeding lines for the prevalent diseases in Latin America. Given the high variability in disease pressure, even at this hot spot, varietal characterization scores may vary from one semester to the next. An objective criteria to decide on the minimum evaluation period required to characterize rice varieties by their disease reaction in Santa Rosa supports an efficient use of research resources and represents a methodological contribution to partner institutions.

Results on disease-evaluation trials conducted at Santa Rosa Station during a 4-year period were used to accomplish this objective. Data source selected for this study corresponds to disease-reaction scores on 70 varieties commercially grown in Latinamerica, evaluated through 7 consecutive semesters (4 semesters "A", under high rainfall (242 to 460mm/month) and 3 semesters "B", under lower rainfall (25 to 36mm/month)) between 1987 and 1990. Disease evaluations include: 1) leaf blast (LBI), at 42 days after sowing; 2) leaf scald (LSc), at flowering time; 3) neck blast (NBI), 30 days after flowering and 4) grain discoloration (GD), 30 days after flowering. Disease reaction was recorded using the 0-9 ordinal scale from the "Standard Evaluation System for Rice".

Data analysis methodology

The analysis has two main objectives:

- a) Assuming seven seasons of continuous evaluations to be the most reliable experimental period length to characterize and select rice varieties for their stable resistance to rice leaf blast, the analysis aims to find out whether a **shorter** period of continuous evaluations would produce the same set of selected material, or would at least exhibit a high percentage of coincidence in selection and a low number of misclassified entries.
- b) To illustrate that different (and possible wrong) conclusions might be reached when treating the ordinal scale as a continuous variable.

For data analysis purposes in phase a, the disease-reactions on the 0-9 scale, were converted into 'disease-severity' according to the Standard Evaluation System for Rice (IRRI, 1988). Disease-pressure at SREE was estimated for each growing season (or semester), as the mean 'disease-severity' over the 70 varieties tested. The 70 varieties were characterized by their mean disease-severity (M) and by their response to increased levels of disease-pressure (b) using the 7-semester evaluations. Statistical comparison of varietal means and a test of homogeneity of slopes (b') were performed using the model illustrated in table 2. In order to correct for lack of normality, 'disease-severity' (Y) was transformed into Y^T using the Box and Cox transformation for ratios (Johnson and Wichern, 1982).

$$Y_i^T = \frac{Y_i^\lambda - 1}{\lambda}, \text{ if } \lambda \neq 0$$

$$Y_i^T = \frac{\lambda}{\ln Y_i}, \text{ if } \lambda = 0$$

where λ value for this particular problem was estimated as $\lambda = 0.4$

Based on the 7-semester evaluations, a group of varieties was selected for their stable resistance. This group, the 'ideal' selections, included varieties with low mean disease-reaction and lack of response to increased levels of disease-pressure. That is, varieties whose M was not statistically different from that of the top variety, using the Waller-Duncan LSD test for mean comparisons, and whose b was not statistically different from 0. The same analytical procedure was applied to data sets simulating shorter continuous experimental period lengths. Two 6-semester periods, three 5-semester periods, four 4-semester periods and five 3-semester periods were simulated. For each one of the fourteen cases, a set of selected varieties was produced. Each set was compared to the 'ideal' set of selections produced by the 7-semester data analysis. The decision on the minimum number of continuous evaluations required to select promising varieties for their stable resistance was achieved based on the percentage of coincidence in selection, and the number of misclassified entries, when compared to the 'ideal' set.

In order to achieve objective b, the same data analysis methodology to identify the 'ideal set of selections' previously applied to 'disease-severity' was applied to the 0-9 scores using as dataset the 7-semester period length. Shorter simulated periods were not analyzed using 0-9 scores. The resulting set of selected varieties was compared with the previously identified 'ideal set of selections' when using 'disease-severity' as the response variable.

Results:

Table 1 confirms the high, but variable levels of disease-pressure at SREE during the 4-year period considered. This supports the use of this 'hot spot' as experimental site to characterize and select rice varieties for their stable resistance to leaf blast, as was pointed out by Correa-Victoria and Zeigler (1992b).

As a result from the analysis on 'disease-severity', a group of 18 varieties, out of the 70 varieties tested, was selected as promising parental material in terms of their stable resistance to leaf blast disease (table 2). They are characterized by low mean disease-severity (M)—ranging between 4.71 and 20.43, when the overall mean was 40.8 and the maximum was 100.0—, and lack of response to increased levels of disease-pressure (b), not different from 0).

Selections resulting from shorter experimental period lengths were compared with this 'ideal' set of 18 promising material presented in table 2. Criteria used for comparison were coincidence in selection and number of misclassified entries. Resulting values are presented in table 3. These results indicate that 6 or even 5 continuous growing seasons would produce the same selections. However, when shorter periods are considered, the number of misclassified entries is high. Correlations between M 's and between b 's, also support the decision of 5 semesters being an appropriate period length.

Selections resulting from the analysis of disease-reaction scores, expressed in a 0-9 ordinal scale, show only an 33.3% coincidence and a high number of misclassified entries, when compared to the set of 18 'ideal' selections resulting from the analysis of 'disease-severity'. This indicates that misleading results and conclusions can be drawn when appropriate statistical methodology is not applied.

Conclusions: The study allows the following conclusions:

- a) The minimum evaluation period length to characterize and select rice varieties by their **stable resistance to leaf blast** is 5 semesters. This guarantees confidence in a proper selection of promising material with a more efficient use of research resources.
- b) Very different, and possibly misleading results (selections) were attained when analyzing disease-reaction scores (in the 0-9 ordinal scale) as a continuous variable.

Table 1: **Disease-pressure levels for leaf blast evaluation at Santa Rosa Experimental Station, Colombia, during a four-year period (1987 - 1990)**

Year/Semester	Leaf Blast pressure (Mean disease severity)^{1/}	Range in varietal disease severity (min, Max)
1987 A	51.5	(6.0, 94.0)
1987 B	48.8	(6.0, 94.0)
1988 A	58.7	(0.0, 100.0)
1989 A	42.0	(6.0, 87.0)
1989 B	24.8	(3.0, 100.0)
1990 A	48.9	(3.0, 87.0)
1990 B	7.6	(3.0, 50.0)
Mean	40.8	
Standard deviation	32.1	
Mean standard error	1.4	
CV (%)	78.7	

^{1/} Expressed as the mean 'disease-severity' per semester over the 70 varieties tested.

Table 2: Group of selected varieties by their stable resistance to leaf blast, based on a 7-semester evaluation period (the most reliable)

Variety	Mean disease severity (M) ^{1/}	Response to disease-pressure (b)	Standard error of b (S _b)	Prob of significance of T statistic (for H ₀ : b = 0)
1. Amistad 82	4.71	0.061	0.4	0.879
2. Ceysvoni	4.71	0.093	0.4	0.879
3. Panamá 1537	5.50	0.100	0.4	0.821
4. Araure 2	6.00	0.100	0.4	0.801
5. IR 58	6.86	0.093	0.4	0.816
6. Panamá 1048	8.67	0.308	0.4	0.452
7. Centa A1	9.00	0.169	0.4	0.673
8. Dawn	9.67	0.313	0.4	0.436
9. Eloni	9.67	0.313	0.4	0.436
10. Colombia 1	10.43	0.349	0.4	0.384
11. Juma 58	11.33	0.363	0.4	0.376
12. Iniap 415	11.43	0.229	0.4	0.568
13. Tanaioka	12.71	0.358	0.4	0.373
14. Juma 62	13.57	0.387	0.4	0.334
15. IR 43	15.00	0.346	0.4	0.388
16. Iniap 7	16.29	0.315	0.4	0.432
17. Ciwini	16.50	0.521	0.4	0.193
18. Araure 4	20.43	0.596	0.4	0.138

^{1/} LSD Walter Duncan value on the transformed variable (=4.34), indicates no significant differences between M's for this group of varieties

Table 3: Coincidence in varietal selection, between the 7-semester evaluation period (the most reliable and shorter periods (using 'disease severity' as response variable

Period length/combination	Number of selected varieties	Coincidence in selection (%)	Misclassified entries	M's range for selected varieties (Min.) (Max.)		b's range on selected varieties (Min.) (Max.)	
7-semester	18	-	-	4.7	20.4	0.06	0.59
6-semester/combination 1	17	93	1	5.0	19.2	0.04	0.71
/combination 2	23	100	5	4.5	18.6	0.03	0.99
5-semester/combination 1	19	100	1	5.4	28.6	0.07	0.93
/combination 2	17	94	1	4.8	18.0	0.01	0.58
/combination 3	19	100	1	4.2	13.6	0.02	0.43
4-semester/combination 1	27	100	9	6.0	53.0	0.02	2.43
/combination 2	19	100	1	5.3	26.0	0.01	0.86
/combination 3	19	100	1	4.5	16.3	0.01	0.60
/combination 4	28	100	10	3.0	17.8	0.02	0.64
3-semester/combination 1	36	100	18	6.0	66.0	-3.28	3.43
/combination 2	22	100	4	6.0	43.6	-2.47	2.22
/combination 3	22	100	8	5.0	35.0	0.03	1.25
/combination 4	26	100	13	3.0	19.7	-0.10	0.54
/combination 5	31	100		3.0	34.3	0.00	0.92

Case Study 2:

Efecto ambiental en la calidad de grano de 8 variedades de arroz

E. Granados, C. Martinez,
J. Holguín, H.F. Ramírez

Objetivo: Determinar la influencia de las condiciones ambientales en la calidad del grano de arroz.

Metodología: Se consideraron 8 variedades de arroz en 12 siembras sucesivas en la estación experimental del CIAT. Se analizó para cada variedad y siembra la producción total de arroz, rendimiento excelso y la cristalinidad de grano (%). Mediante análisis combinados de varianza se determinó la existencia de interacción genotipo x siembra. Como las condiciones climáticas (humedad relativa, brillo solar, temperaturas, etc.) son los principales aspectos cambiantes entre una siembra y otra, se espera explicar la variabilidad debida a la interacción por medio de las variables climáticas. Para ello una vez coleccionada la información climática, según los ciclos de desarrollo de la planta se consideró el modelo compuesto de análisis de factores simultáneo tanto en la matriz de interacción como en la matriz de clima. Identificando el número ideal de factores que expliquen simultáneamente en las dos matrices la mayor variabilidad y efectuando las rotaciones de ejes apropiadas, se puede interpretar de acuerdo a los pesos en la matriz de interacción qué variedades se afectan más por algunos componentes del clima, dejando en las matrices residuales la variabilidad que no es explicable por los factores ambientales.

Modelo Estadístico: Supóngase que se tiene la matriz X_{ikK} de los promedios de cada variedad ($k=1,2,...,K$) en cada una de las siembras ($i=1,2,...,I$) y la matriz de información ambiental U_{ijJ} estandarizada para las características ambientales ($j=1,2,...,J$) a través de las I siembras. Si de la matriz X se elimina la variación debida a cada una de las variedades y siembras, se obtiene una matriz Z con únicamente la contribución a la interacción, variedad por siembra.

Se establece entonces el modelo de análisis de factores:

$$Z_{ikK} = T_{ikA} Q_{Akk} + F_{ikK}$$

$$U_{ijJ} = T_{ijA} W_{Ajj} + E_{ijJ}$$

donde T_{ikA} es la matriz de puntajes con A factores principales que descompone simultáneamente las matrices Z y U ; Q_{Akk} y W_{Ajj} son las matrices de los pesos de los factores en las correspondientes variedades y características ambientales; F_{ikK} y E_{ijJ} son las correspondientes matrices residuales que representan la discrepancia en la reproducción de la matriz de información original en el modelo con A componentes o factores.

Implementación del Algoritmo de Solución: Debido a que en los paquetes estadísticos disponibles no existe un procedimiento que permita encontrar la solución al enfoque de análisis, se ha desarrollado una MACRO SAS que combina SAS-IML y Factor con el algoritmo denominado Regresión Parcial de Mínimos Cuadrados. Esta macro desarrolla el algoritmo propuesto por A. Halvor Aastveit y Harald Martens (Biometrics 42, 829-844, Dic. 1986).

Resultados preliminares para porcentaje de granos cristalinos: La tabla 1 muestra los promedios del porcentaje de granos cristalinos en 12 siembras de las variedades Blubonnet 50, IR-22, CICA4, IR-8, CICA8, ORYZICA 1, ORYZICA 2 y ORIZYCA 3.

La matriz de información ambiental se conformó según el clima durante el desarrollo del cultivo en 2 etapas comprendidas entre floración-maduración-cosecha, en lo referente a máxima diferencia diaria de temperatura (DELT) en el período, máxima diferencia diaria en humedad relativa (DELH), brillo solar total (BRILLO), precipitación total (PRECI) y radiación solar (RADI) (tabla No. 2).

Una vez aplicado el algoritmo de solución se analizó el porcentaje de varianza residual en modelos con cero factores hasta ocho factores y se escogió el modelo con cuatro factores que arroja 60% y 23% de varianza residual en la reproducción de la información de interacción Variedad x Siembra y de la información climática estandarizada.

Después de hacer rotación de ejes por el método VARIMAX en todos los factores, se determinó qué factores están mejor correlacionados con qué características climáticas y variedades; observando la correspondencia entre las ponderaciones en un mismo factor (tabla 3), se determinó la posible causa de características ambientales en la interacción variedad x siembra. En gráficas que combinen diferentes escalas en los mismos ejes se puede interpretar la correspondencia existente entre las variedades y las características climáticas. Es importante por una parte, observar en el gráfico la posición y separación entre las variedades en el sistema de ejes (factores). Análogamente se debe observar con igual criterio la posición y separación entre las características climáticas. En cambio la distancia entre una variedad y una variable de clima no se interpreta por tener ellas distinta magnitud pero sí se infiere a partir de la posición relativa que ocupan dentro del sistema de ejes. Las gráficas 1(a) y 1(b) representan la ubicación relativa de las ponderaciones para los cuatro factores tanto de variedades como de clima (multiplicada por 3).

Conclusiones

- a) Las variedades Oryzica 1, Oryzica 3 e IR-22 presentaron el mismo patrón de interacción y con un patrón contrario se presentó IR8. Las primeras se afectaron positivamente por el brillo solar y la radiación solar en el período entre maduración y cosecha ocurriendo lo contrario (pérdida de cristalinidad) en la variedad IR-8. Esta última también se afectó negativamente por la máxima diferencia diaria de humedad relativa entre floración y maduración y por la precipitación en los dos últimos períodos del cultivo.
- b) Con otro patrón, las variedades Oryzica 2 e inversamente CICA 8, reaccionaron al brillo solar y radiación solar en el período floración-maduración. La máxima diferencia en temperatura de este período también influyó positivamente en Oryzica 2.
- c) Para las variedades CICA 4 y Blubonnet 50 no se explicó su interacción por medio de los factores aplicados sobre las variables climáticas escogidas.
- d) Existe alguna dificultad en la escogencia apropiada de las características ambientales. Un examen de los ciclos del cultivo y la respuesta varietal analizada es indispensable para plantear correctamente el modelo.

Tabla 1. Promedios de grano cristalino (%) de 8 variedades de arroz observados en 12 siembras sucesivas en CIAT - Palmira.

SIEMBRA	BLUBONET 50	IR-22	CICA 4	IR 8	CICA 8	ORYZICA 1	ORYZICA 2	ORYZICA 3	MEDIA	STD.
S1	88.4	95.2	88.0	51.0	77.4	97.0	65.0	93.3	82.0	15.2
S2	93.8	99.4	92.0	70.3	91.0	96.0	61.0	98.0	88.0	13.2
S3	93.5	99.2	90.0	58.4	84.0	99.0	61.0	98.4	85.3	16.0
S4	91.0	98.0	87.3	46.0	76.4	98.0	75.0	98.2	84.0	17.0
S5	80.5	96.0	83.0	36.	81.0	96.0	67.1	96.0	79.2	19.0
S6	96.2	98.0	85.0	55.0	90.0	97.4	78.5	97.0	87.0	14.0
S7	90.0	96.1	88.0	45.1	83.0	94.2	64.0	95.2	82.0	17.0
S8	83.4	90.3	75.4	57.0	88.0	93.0	65.0	96.0	81.0	13.0
S9	82.5	98.4	86.0	58.0	82.1	97.0	66.0	97.0	83.2	14.0
S10	77.1	97.3	80.0	23.0	77.0	95.2	49.2	93.0	74.0	24.1
S11	74.9	96.1	88.0	36.2	56.0	95.0	60.0	96.0	75.2	21.0
S12	82.5	92.0	74.0	38.1	73.0	96.0	60.4	97.2	76.5	19.0
MEDIA	86.2	96.2	85.0	48.0	80.0	96.0	64.2	96.2	81.3	
STD	6.7	2.6	5.3	13.0	9.0	1.5	7.1	2.0		

Tabla 2: Características climáticas consideradas en el análisis de factores conjunto con la interacción de 8 variedades de arroz en 12 siembras sucesivas en CIAT-Palmira.

Siembra	CARACTERISTICAS DE CLIMA									
	Máxima diferencia diaria presentada									
	Temperatura		Humedad relativa		Brillo solar		Precipitación total		Radiación total	
	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)	(3)	(4)
S1	13.5	12.1	44.3	38.3	158.3	137.6	54.4	173.9	503.5	491.7
S2	12.1	12.1	38.3	34.6	145.5	148.1	175.1	14.7	501.8	472.4
S3	12.6	15.5	33.7	40.3	147.1	173.3	10.8	39.8	469.8	497.5
S4	15.5	14.6	40.3	40.0	170.7	176.5	39.8	29.5	495.8	523.3
S5	15.9	17.0	40.0	41.3	175.6	208.3	29.5	6.7	521.4	531.6
S6	17.0	15.2	41.3	37.7	208.3	187.1	6.7	5.2	531.6	553.3
S7	15.2	14.0	37.7	30.3	188.4	147.6	5.7	167.5	555.6	493.2
S8	14.0	12.0	30.3	34.7	147.6	175.4	167.5	89.6	493.2	527.5
S9	12.0	12.0	34.7	30.7	182.2	158.7	89.7	57.7	525.6	486.3
S10	12.0	13.6	32.0	33.3	158.1	234.1	58.3	55.4	486.1	555.1
S11	13.6	13.6	33.3	34.3	234.1	214.1	55.4	27.1	555.1	544.7
S12	13.6	13.5	34.3	33.7	212.9	171.9	44.4	135.4	545.3	535.9
Media	13.9	13.8	36.7	35.8	177.4	177.7	61.4	66.9	515.4	517.7

(3) Período comprendido entre floración y maduración

(4) Período comprendido entre maduración y cosecha

Tabla 3. Ponderaciones de las variedades y características climáticas y puntajes de las 12 siembras de arroz en el modelo con 4 factores del análisis de factores conjunto

		FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
VARIEDAD (1)	BB50	1.16	-0.63	-0.25	1.02
	IR22	-1.50	1.20	0.17	1.31
	CICA4	-0.88	-0.19	0.65	1.48
	IR8	0.80	-3.50	-0.48	-4.20
	CICA8	0.87	-0.09	-2.80	-0.57
	ORY1	-1.50	1.24	0.32	1.46
	ORY2	2.51	0.65	1.85	-0.85
	ORY3	-1.50	1.31	0.53	0.38
CLIMA (1)	DELT3	0.75	0.02	0.06	-0.18
	DELT4	0.33	0.31	-0.15	0.17
	DELH3	0.31	-0.30	0.06	0.56
	DELH4	0.33	-0.02	-0.05	0.06
	BRILLO3	-0.03	0.06	0.77	-0.06
	BRILLO4	-0.21	0.64	-0.02	0.13
	PRECI3	-0.09	-0.25	-0.11	-0.56
	PRECI4	-0.22	-0.04	-0.08	0.51
	RADI3	-0.03	-0.05	0.59	0.07
	RADI4	0.15	0.57	0.05	-0.14
SIEMBRA (2)	S1	0.43	-2.60	-0.43	1.44
	S2	-0.36	-2.70	-1.30	-2.00
	S3	0.02	-0.03	-1.70	0.91
	S4	1.74	0.13	-0.31	-0.02
	S5	2.03	1.60	0.21	0.56
	S6	2.30	1.55	1.50	-0.22
	S7	0.06	-0.93	1.22	1.05
	S8	-0.73	-0.36	-1.50	-1.90
	S9	-1.50	-1.40	0.15	-0.82
	S10	-2.10	2.12	-1.20	1.12
	S11	-0.96	2.04	2.05	-0.55
	S12	-0.87	0.59	1.38	0.41

(1) Ponderaciones. (2) puntajes

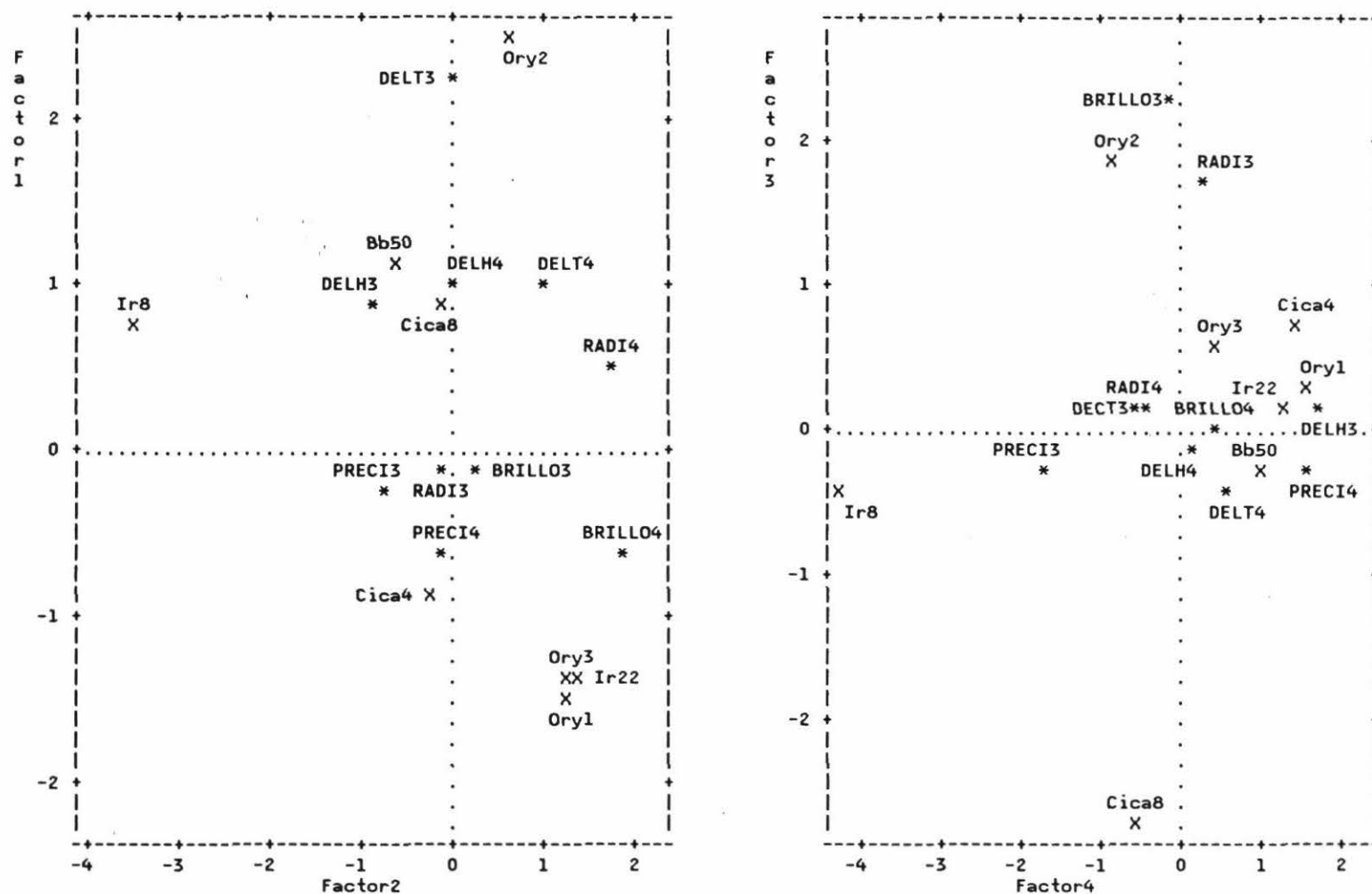


Figura 1. Representación de los puntajes de las 8 variedades de arroz y las características climáticas (multiplicadas por 3), obtenidos en el Análisis de factores conjunto.

5.2 Evaluation Study: performance of SAS and ORACLE in the Rice Program Local Area Network (LAN)

In December 1991, a Local Area Network (LAN) with a 386/16MB memory/650MB disk server with Novell software, and 12 DOS microcomputers as clients, was installed to serve as a pilot project for CIAT to evaluate the advantages of a computer network environment software in the LAN included: WORDPERFECT LAN, OFFICE FOR LAN (MAIL), HARDVARD GRAPHICS LAN, FOX-PRO2 LAN and ISIS LAN. However, standard CIAT software for scientific data processing, statistical analysis and database management --as SAS and ORACLE software products-- were not included in the Server.

The purpose of this short-term evaluation study was to quantify performance of SAS/PC version 6.04 and ORACLE version 5 software in the LAN considering "response time" as the performance criteria, in comparison to "response time" in the mainframe using SAS 6.07 and ORACLE version 5. By "response time" it is understood the time (in seconds) taken to successfully complete a given SAS job and provide an interactive answer to the user.

SAS evaluation: Four factors affecting "response time" were considered for this evaluation:

- a) Complexity of the SAS job (at 2 levels: basic descriptive statistics --MEANS, PLOT, FREQ, CHART and inferential statistics/multivariate techniques-- ANOVA, GLM, VARCOMP, CLUSTER, TREE);
- b) Size of datafile at 2 levels: small (180 registers) and large (1500 registers); c) Balance or unbalance of data files and d) Present load work in the LAN. All SAS jobs were previously run in the mainframe and in a 4MB memory independent micro.

Results show that: a) Response time and SAS performance in the LAN was similar to that on an independent PC with 4MB memory; b) for simple jobs with small balanced datafiles, response time in the LAN was half of that in the mainframe; c) complex jobs with balanced datafiles took much longer time to run in the LAN than in the mainframe and d) complex jobs with unbalanced datasets did not run in the LAN. As conclusion, in order to successfully run in order SAS/STATS more complex procedures in the LAN, a more capable Server would be required.

ORACLE evaluation: The evaluation project included software installation; implementation of a pilot database in the LAN Server (the Cassava breeding database was chosen), implementation of existing applications for update, modification query and report generation; and finally quantification of response time to a wide range of queries. These response times were compared with those attained on the IBM 4361 mainframe computer.

Results show: a) Complete portability of ORACLE database design, data structures and applications between VM/CMS, DOS 5.0 and Netware 3.11; b) Response time in the LAN was much better than in the mainframe; c) ORACLE use showed a substantial increase in the LAN load, causing delays in response time to other LAN users. As conclusion, in order to successfully run ORACLE in a Novell LAN, a more capable SERVER would be required. A 486 Server would be better.

6. CONTRIBUTIONS TO THE GENETIC RESOURCES UNIT

Case: The Genetic Resources Database

F. Rojas, G. Serrano, A. Franco, C. Saa and M.C. Amézquita, from DSU
M. Iwanaga, B. Maas, R. Hidalgo and team, from GRU

The purpose of the Genetic Resources Database is to store in an organized way, maintain and make interactively available to the GRU personnel, to CIAT Scientists and to NARI's partners, all the information related to the **collection, characterization, storage and distribution** of germplasm collections handled by CIAT between 1972 and 1990.

The Genetic Resources Database is composed at present by three sub-systems corresponding to:

a) The "Tropical Pastures Genetic Resources Database"

The purpose of the genetic resources database is to store in an organized way, maintain and make interactively available, all the information related to tropical pastures genetic resources that have been generated, collected and handled by the CIAT Genetic Resources Unit or by the Germplasm Evaluation Section of the CIAT Tropical Pastures Program between 1978 and 1991.

The tropical pastures genetic resources database includes information on 22,818 accessions of grasses and legumes. The database is organized in seven sub-systems, according to the type of information being stored, as follows:

1. Passport data (22,818 accessions)
2. Short-term seed inventory (20,220 accessions)
3. Long-term seed inventory (3,872 accessions)
4. Seed international shipments (200 shipments to Latinamerican, Asian and African countries)
5. Seed multiplication, at greenhouse and at field level (historical records of all material submitted to seed multiplication: 4,815 multiplication events)
6. Inventory of materials stored as "herbario" (9,699 accessions)

b) The "Cassava Genetic Resources Database"

This database includes at present the following information:

Germplasm Bank: 4,650 varieties with

- . passport data (collection site descriptors, collection date, origin, local names, etc.)
- . morpho-agronomic characteristics (19 descriptors on 4081 varieties)
- . electrophoretic characterization (presence or absence of electrophoretic bands 4251 varieties).

c) The "Bean Genetic Resources Database"

The Bean Genetic Resources Database includes at present the following information:

- 1) **Bean Germplasm Bank**: 25,000 *phaseolus* accessions belonging to *vulgaris*, *acutifolius*, *coccineus*, and *lunatus* species. Out of the 32,500 accessions in stock (with "S" identification code), these 25,000 have passed through quarantine, have been characterized and are given the "G" code. They constitute the Germplasm Bank.

Information on each "G" accession includes:

- . passport data (collection site descriptors, collection data, origin, local names, etc.)

- . morpho-agronomic descriptors (60 descriptors on 25,000 accessions, product of experimental evaluations carried-out by the GRU and the CIAT's Bean Program).
- 2) **Bean Germplasm in Stock:** (accessions with "S" code but not in the germplasm bank): 15,400 phaseolus accessions with passport data.
- 3) **Seed inventory and international shipments:** 24,000 out of the 25,000 germplasm bank accession have seed inventory information. Also, information related to seed distribution to research institutions and universities all around the world is included.
- d) The "Rice Genetic Resource database" is at present handled directly by the CIAT Rice Program and is not included in this central database.

Between September 1991 and the end of November 1992, the "Genetic Resources Database" has been re-designed and implemented in ORACLE software. It is fully operational now, and can be accessed through IBM 4361 terminals or through a PC with 4Mb of memory/40Mb hard disk and equipped with ORACLE tools. Online applications have been developed by the Biometry Unit to query and update the bean germplasm database: Training courses have been offered to end-users to teach them how to use these applications.

More detailed information on this database appears in the 1991 and 1992 GRU Annual Reports.

Annex: "CIAT Biometry needs: Demand and Supply": results of meetings, discussions and survey conducted among scientists from research programs/Units, in October-Nov. 1992.

Study Results

Purpose of the study:

To identify needs of the new CIAT in terms of biometrical support and prepare our Biometry Unit to meet them.

Methodology:

In order to accomplish this objective, the Head of the Biometry Unit had a series of meetings with each research Program/Unit, in which the Scientists needs were analyzed in terms of their requirements for technical support in a wide range of biometrical techniques. In these meetings the present capacity of the Unit was explained, both in terms of personnel resources as in terms of their present technical expertise. A list of biometrical areas of technical support was presented to the Scientists who were then asked to express their short and medium term needs for technical support in each area, by filling-in a pre-designed format. Areas of technical support included:

- a) **Biometrical methodologies:** 1) Design of crop experiments; 2) design of agro-silvo-pastoral experiments; 3) sample survey design and analysis; 4) exploratory data analysis; 5) general Linear Models techniques (univariate and multivariate); 6) Non-linear models; 7-8) multivariate techniques (reduction of dimensionality, classification and ordination technique); 9) categorical data analysis technique; 10) Sampling techniques; 11) Time Series; 12) Analysis of non-normal response variables. Additionally, other quantitative methodological areas were included as quantitative genetics, Spatial variability techniques, operations research and simulation.
- b) **Use of databases of experimental results:** It was intended to identify the scientists needs to store statistical summaries resulting from multilocal trials, across years experimental results or across disciplines experimental results. Also to identify their needs in terms of collaboration from the Biometry Unit in long-term data analysis projects using the existing historical results.
- c) **Training** in statistical methods and use of statistical software for data analysis purposes (to internal personnel or to NARS partners).

Scientific disciplines within Programs were grouped by "data domain" with similar needs. Summarized answers by "data domain" within Program/Unit are included in this annex (see attached summary formats).

Results: 42 CIAT scientists were consulted and 34 formally expressed their needs by filling-in the format. In general CIAT scientists expect support - both in terms of advice and participation in collaborative data analysis projects - from the biometricians. Biometrical areas of technical support classified according to their **demand**, are as follows:

TECHNICAL AREAS

[illegible]

Most scientists suggested that Simulation/Mathematics Modelling should not be included as part of the Biometry Unit responsibility, but as specialized projects within programs.

In terms of training, more emphasis was placed in training their own staff than offering training to NARS researchers. Some scientists have suggested the importance of collaboration between their program and the Biometry Unit in offering on-site consultancies to specific NARS. Econometric techniques in response to a new expected demand from the economists/social scientists.