Quantitative Analysis of Data from Participatory Methods in Plant Breeding

Mauricio R. Bellon and Jane Reeves, Editors

Quantitative Analysis of Data from Participatory Methods in Plant Breeding

Mauricio R. Bellon and Jane Reeves, Editors

CIMMYT
International Maize and Wheat
Improvement Center

PRGA
Participatory Research and Gender Analysis

IRRI
International Rice Research Institute

CIMMYT® (www.cimmyt.org) is an internationally funded, nonprofit, scientific research and training organization. Headquartered in Mexico, CIMMYT works with agricultural research institutions worldwide to improve the productivity, profitability, and sustainability of maize and wheat systems for poor farmers in developing countries. It is one of 16 food and environmental organizations known as the Future Harvest Centers. Located around the world, the Future Harvest Centers conduct research in partnership with farmers, scientists, and policymakers to help alleviate poverty and increase food security while protecting natural resources. The centers are supported by the Consultative Group on International Agricultural Research (CGIAR) (www.cgiar.org), whose members include nearly 60 countries, private foundations, and regional and international organizations. Financial support for CIMMYT's research agenda also comes from many other sources, including foundations, development banks, and public and private agencies.

Future Harvest® builds awareness and support for food and environmental research for a world with less poverty, a healthier human family, well-nourished children, and a better environment. It supports research, promotes partnerships, and sponsors projects that bring the results of research to rural communities, farmers, and families in Africa, Asia, and Latin America (www.futureharvest.org).

Correct citation: Bellon, M.R., and J. Reeves (eds.). 2002. *Quantitative Analysis of Data from Participatory Methods in Plant Breeding.* Mexico, DF: CIMMYT.

Abstract: Although participatory plant breeding (PPB) is gaining greater acceptance worldwide, the techniques needed to analyze the data from participatory methodologies in the context of plant breeding are still not well known or understood. Scientists from different disciplines and cropping backgrounds, working in international research centers and universities, discussed and exchanged methods and ideas at a workshop on the quantitative analysis of data from participatory methods in plant breeding. The papers in this volume address the three themes of the workshop: designing and analyzing joint experiments involving variety evaluation by farmers; identifying and analyzing farmers' evaluations of crop characteristics and varieties; and dealing with social heterogeneity and other research issues. Topics covered included different statistical methodologies for analyzing data from on-farm trials; the mother-baby trial system, which is designed to incorporate farmer participation into research; the identification and evaluation of maize landraces by small-scale farmers; and a PPB process that aims to address the difficulties of setting breeding goals and choosing parents in diversity research studies. Summaries of the discussion, as well as the participatory breeding work currently conducted by the participants, are provided.

ISBN: 970-648-096-X
AGROVOC descriptors: Maize; Plant breeding; Agronomic characters; Research; Methods; Quantitative analysis; Data analysis; Statistics; Evaluation; Scientists; Small farms; Research institutions
AGRIS category codes: U10 Mathematical and Statistical Methods;
F30 Plant Genetics and Breeding
Additional keywords: CIMMYT, participatory plant breeding
Dewey decimal classification: 633.1553
Printed in Mexico.

Contents

Preface

The papers in this volume were presented at a workshop “Quantitative Analysis of Data from Participatory Methods in Plant Breeding”, held at the Castle of Rauischholzhausen Conference Center of the Justus Liebig University, Germany, 23-25 August 2001. The workshop was initiated by scientists within the Consultative Group on International Agricultural Research (CGIAR) who wished to review and discuss different quantitative techniques for analyzing data generated through participatory methodologies in the context of plant breeding. Participatory plant breeding (PPB) is gaining wider acceptance worldwide—it is increasingly being used within the CGIAR—and its merits and limitations are beginning to be better understood. Many scientists involved in these efforts, however, have realized that the quantitative techniques needed to analyze the data from the participatory methodologies used in PPB are still not well known or understood by many practitioners. Further discussion and exchange of methods and ideas are needed.

The workshop was organized by the International Maize and Wheat Improvement Center (CIMMYT) and the Justus Liebig University. It was sponsored by CIMMYT, the International Rice Research Institute (IRRI), the CGIAR Program on Participatory Research and Gender Analysis for Technology Development and Institutional Innovation, and other participating CGIAR Centers. Experts from outside the CGIAR were also involved.

Scientists from different disciplines (breeders, social scientists, biometricians, and agronomists) and cropping backgrounds (maize, rice, potatoes, cassava, sorghum, barley, and agroforestry) were brought together for the workshop. All participants were experienced in participatory plant breeding and had also worked in interdisciplinary teams. A total of 21 scientists took part, representing 10 CGIAR Centers: CIMMYT, IRRI, CIAT (the International Center for Tropical Agriculture), CIP (the International Potato Center), the International Centre for Research in Agroforestry (ICRAF), the International Center for Research in the Dry Areas (ICARDA), the International Institute of Tropical Agriculture (IITA), the West Africa Rice Development Association (WARDA), the International Plant Genetic Resource Institute (IPGRI), and the International Center for Research in the Semi-Arid Tropics (ICRISAT). In addition there were participants from Justus Liebig University, University of Wales, and Michigan State University.

The workshop was organized around three themes:

- Designing and analyzing joint experiments involving variety evaluation by farmers
- Identifying and analyzing farmers’ evaluations of crop characteristics and varieties
- Dealing with social heterogeneity and other research issues

Lectures and case studies, followed by discussion, were devoted to each theme. This format facilitated the learning of specific methodologies, the sharing of well-defined examples, and a free exchange of ideas and experiences among participants.

This publication presents both the papers from the workshop and the ensuing discussions. There are two types of papers: one focusing on methodologies, the other on case studies. In the first methodology paper, Steven Franzel and Richard Coe discuss the suitability of different types of trials for participatory on-farm technology testing. Richard Coe's two papers deal with statistical methodologies. His first paper focuses on the analysis of data from on-farm trials, while the second analyses the ratings and rankings commonly used in PPB to assess the value of traits and the performance of different varieties from the farmer's perspective. Sieglinde Snapp describes a methodology called the mother-baby trial system, which is designed to incorporate farmer participation in research. Although the methodology presented is in the context of soil fertility management, it is still relevant to PPB, as shown in the papers by Gary Atlin and colleagues and Hugo de Groote and colleagues. The paper by José Crossa and colleagues shows a statistical methodology for grouping farmers into homogenous groups. Among the case study papers, Gary Atlin and colleagues discuss the analysis of the mother-baby trial scheme in participatory varietal selection and its implications for rice breeding. Mauricio Bellon details a case study on the identification and evaluation of maize landraces by small-scale farmers in Mexico. Hugo de Groote and colleagues present the results of a study that aimed to identify farmers' criteria for assessing and evaluating maize varieties in Eastern Africa. Bhuwon Sthapit describes the preliminary results of participatory plant breeding processes that address the difficulty of setting breeding goals and choosing parents in diversity research studies.

The resulting discussions are summarized at the end of each paper. Also, all participants provided a one-page summary highlighting the participatory breeding work currently being conducted by their respective research centers. These summaries are presented in Appendix 1.

I believe that these proceedings will be useful to all practitioners of participatory plant breeding.

Mauricio R. Bellon
CIMMYT Economics Program

Acknowledgements

Funding for the workshop was provided by the International Maize and Wheat Improvement Center, the International Rice Research Institute, and the CGIAR Program on Participatory Research and Gender Analysis for Technology Development and Institutional Innovation, which also funded the production of these proceedings. We also express our thanks to Prof. Ernst August Nuppenau from the Institut für Agrarpolitik und Marktforschung, Justus-Liebig University, for his invaluable support in hosting the workshop at the Castle of Rauischholzhausen Conference Center, and to Monika Zurek for her logistical help in organizing the workshop. Finally, we thank Marcelo Ortiz and Miguel Mellado for the design and layout.

Acronyms and Abbreviations

CBR	Community biodiversity register
CGIAR	Consultative Group on International Agricultural Research
CIAT	International Center for Tropical Agriculture
CIMMYT	International Maize and Wheat Improvement Center
CIP	International Potato Center
CRURRS	Central Upland Rice Research Station
DFID	Department for International Development, UK
DGIS	Directorate-General for International Cooperation, the Netherlands
FAO	Food and Agriculture Organization of the United Nations
GEI	Genotype x environment interaction
ICAR	Indian Council of Agricultural Research
ICARDA	International Center for Research in the Dry Areas
ICRAF	International Centre for Research in Agroforestry
ICRISAT	International Center for Research in the Semi-Arid Tropics
IDRC	International Development Research Centre, Canada
IFAD	International Fund for Agricultural Development
IITA	International Institute of Tropical Agriculture
IM	Independent mixture
IPGRI	International Plant Genetic Resources Institute
IRRI	International Rice Research Institute
LM	Location model
masl	Meters above sea level
ML	Maximum likelihood
MLM	Modified location model
MVN	Multivariate normal
NARC	Nepal Agricultural Research Council
NGO	Nongovernmental organization
PPB	Participatory plant breeding
PRA	Participatory rural appraisal
PRGA	Participatory Research and Gender Analysis Program
PVS	Participatory varietal selection
RCB	Randomized complete block
RUA	Rajendra Agricultural University
SARRNET	Southern Africa Root Crops Research Network
TE	Target environment
UPGMA	Unweighted pair group arithmetic averaging
WARDA	West Africa Rice Development Association

Participatory On-Farm Technology Testing: The Suitability of Different Types of Trials for Different Objectives

STEVEN FRANZEL AND RICHARD COE

Abstract

This paper outlines objectives for conducting on-farm trials and presents a typology for classifying on-farm trials, focusing on how different types of trials may be used to meet different objectives. It presents the rationale for conducting on-farm trials, the main elements of participatory technology development, and a classification for on-farm experiments based on the degree of control over the experiment by scientists and farmers. The classification recognizes three types of trials, depending on the objectives of the trial, who designs it, and who manages it. Type 1 trials are researcher designed and managed and their objective is to assess biophysical responses. Type 2 trials are researcher designed and farmer managed, e.g., farmers agree to implement a common design. It is useful to get farmer feedback on specific prototypes or for conducting economic analyses. Type 3 trials are farmer designed and managed where farmers can experiment on their own. The objective of this type of trial is to assess farmer innovation and acceptability.

Introduction

In participatory on-farm evaluation, farmers take a lead role in the design, implementation, and evaluation of technology. This paper outlines objectives for conducting on-farm trials and presents a typology for classifying on-farm trials, focusing on how different types of trials may be used to meet different objectives. Some main issues in the management of different types of trials are also discussed. This paper draws heavily on Franzel et al. (2002a).

Objectives of On-Farm Experimentation

On-farm experimentation has several different objectives. First, it permits farmers and researchers to work as partners in the technology development process. The more often and the earlier that farmers are involved in the technology development process, the greater the probability that the practice will be adopted. On-farm trials are important for ascertaining farmers’ assessments of a practice and their ideas on how it may be modified and for observing their innovations. Assessments are likely to vary and may be associated with particular biophysical (e.g., soil type) or socioeconomic (e.g., wealth status) circumstances. Farmers' innovations often serve as a basis for new research or for modifying recommendations (Stroud 1993; van Veldhuizen et al. 1997).

Secondly, on-farm testing is useful for evaluating the biophysical performance of a practice under a wider range of conditions than is available on-station. This is especially important because soil type, flora, and fauna on research stations are often not representative of those found on farms in the surrounding community.

Thirdly, on-farm trials are important for obtaining realistic input-output data for financial analysis. Financial analyses conducted on on-station experiments differ from those conducted on farm trials because (1) yield response is often biased upward, (2) estimates of labor used by station laborers on small plots are unrepresentative of the farming community, and (3) operations often differ, e.g., when tractors instead of oxen or hoes are used for preparing land.

And finally, on-farm testing provides important diagnostic information about farmers' problems. Even if diagnostic surveys and appraisals have already been conducted, researchers can still learn a great deal about farmers' problems, preferences, and livelihood strategies from interacting with them in on-farm trials. Trials have important advantages over surveys in that they are based on what farmers do rather than on what they say.

Types of On-Farm Trials

On-farm trials can provide critical information for determining the biophysical performance, profitability, and acceptability of agroforestry, i.e., its adoption potential. However, the design of a trial depends on its specific objectives.

Assessment of biophysical performance requires biophysical data on the products and services that the technology is planned to produce. These are likely to change with different adaptations of the technology as might occur if farmers were asked to manage them. To prevent such possible variation, trials designed to assess biophysical performance should be controlled to replicate specific technology designs. The trials should also be implemented in a way that farmers' willingness and ability to establish and maintain the trials does not affect the outcome. Thus trials to assess biophysical performance need a high degree of researcher control in both design and implementation.

The assessment of profitability requires biophysical data (to estimate returns) that must be generated from standardized experiments. However, the financial analysis also requires realistic input estimates, of which labor poses most difficulties. Realistic data can only be obtained if farmers manage the trials to their own standards. Thus profitability objectives require trials in which researchers have considerable input into the design but farmers are responsible for implementation. The objectives of assessing feasibility and acceptability require data on farmers' assessments and adaptations of the technology. These can only be assessed if farmers are left to experiment with little researcher involvement.

There are many different ways of classifying on-farm trials (Okali et al. 1994). The differing requirements of the objectives of biophysical performance, profitability, and acceptability mean it is helpful to classify trials according to the balance of researcher and farmer involvement in their design and implementation. The classification described in this paper involves three types of trials and draws upon Biggs (1989).

Type 1: Trials designed and managed by researchers

Type 1 trials are simply on-station trials transferred to farmers' fields. They are useful for evaluating biophysical performance under farmers' conditions, and require the same design rigor as on-station research with regard to treatment and control choice, plot size, replication, and statistical design. In the design stage, researchers need to consult the farmer on the site's homogeneity and history. If possible they should observe a crop in the field before establishing a trial.

Because type 1 trials take place on farmers' fields, trial results are generally more representative of farmers' biophysical conditions than on-station trials (Shepherd et al. 1994). More accurate information may be obtained on interactions between the biophysical environment and management, e.g., how different species in an improved fallow trial compare on different soil types.

Type 1 trials are usually more expensive and more difficult to manage than on-station trials; they often involve renting land from farmers and using laborers from the station to implement them. Farmers' assessments are an important objective of type 1 trials; as with on-station trials, it is useful to get farmers' feedback on the different treatments (Sperling et al. 1993; Franzel et al. 1995).

Type 2: Trials designed by researchers and managed by farmers

Here, farmers and researchers collaborate in the design and implementation of the trial. The trial is labeled "researcher designed" because it follows the conventional scientific approach to conducting an experiment: one or more test treatments are laid out in adjacent plots and compared to one or more control treatments. Researchers consult farmers on the design of the trial and each farmer agrees to follow the same prototype (or chooses one of several possible prototypes), so that results may be compared across farms. Farmers are responsible for conducting all of the operations in the trial.

In type 2 trials, reliable biophysical data over a broad range of farm types and circumstances are sought. The trials also facilitate the analysis of costs and returns; inputs, such as labor, and outputs, such as crop yields, are relatively easy to measure because plot size is uniform and known. The trials are also useful for assessing farmers' reaction to a specific practice and its suitability to their circumstances. Farmers are encouraged to visit each other's trials and to conduct group field days to assess the practice at different growth stages.

Type 3: Trials designed and managed by farmers

In type 3 trials, farmers are briefed about new practices through visits to field stations or on-farm trials. They then plant and experiment with the new practices as they wish. They are not obliged to plant in plots or to include control plots. Researchers monitor the farmers' experiments, or a subsample of them, focusing in particular on their assessment of the new practice and their innovations. In addition farmer-to-farmer visits and meetings are useful for farmers to compare their experiences and assessments with others. Any farmers experimenting with a new practice could be said to have a type 3 trial, regardless of whether they obtained planting material and information from researchers, other facilitators, or other farmers. This hands off approach, which assumes that farmers best know how to test a new practice on their own farms, is supported by some in the literature (Lightfoot 1987). Others emphasize training farmers to conduct trials following scientific principles, such as replication and non-confounding of treatments (Ashby et al. 1995).

Suitability of Trial Types for Meeting Objectives

The suitability of the different trial types for differing objectives is summarized in Table 1. Suitability involves both the appropriateness of the trial for collecting the information and the ease with which it can be collected. Different types of trials are suited to different types of analyses. Biophysical measurements are most meaningful in types 1 and 2 trials; they are less useful in type 3 trials because each farmer may manage the practice in a different manner. Type 2 trials are well suited for collecting parameters (e.g., labor use) for financial analysis; such data are difficult to collect in type 3 trials because plot size and management vary. The data can be collected in type 1 trials but will be less relevant to farmer circumstances; yield response to new practices tends to be biased upward, and labor use, measured using laborers hired by researchers and working on small plots, is unrepresentative of farmers' labor use.

Farmers' assessments are more accurate in type 3 trials for several reasons. Because farmers control the experimental process, they are likely to have more interest and information on the practice. Furthermore, because farmers in type 3 trials usually have less contact with researchers than farmers in other types of trials, their views of a technology are less influenced by researchers' views. Finally, whereas it is often necessary to provide inputs to farmers in type 2 trials to ensure that results are comparable across farmers, no inputs, with the possible exception of planting material, are provided in type 3 trials. Thus farmers' views in type 3 trials are more likely to be sincere than in type 2 trials, where positive assessments may simply reflect the farmers' interest and satisfaction in obtaining free inputs. For example, in a hedgerow intercropping trial in western Kenya (Swinkels and Franzel 1997), 50% of the farmers claimed that hedges increased crop yields, whereas technicians noted yield increases on only 30% of farms—the technicians claimed that the difference was due to farmers trying to please researchers.

Finally, all three types of trials play a potentially important role in defining the boundary conditions for the technology, i.e., the biophysical and socioeconomic conditions under which the practice is likely to be adopted by farmers. Which type of trial is best depends on the objectives and particular circumstances of the participants (facilitators and farmers).

Continuum and sequencing of trial types

The different types of trials are not strictly defined, rather they are best seen as points along a continuum. For example, it is common for a trial to fit somewhere between type 2 and type 3, as in the case where farmers agree to test a specific protocol (type 2), but over time, individuals modify their management of the trial (type 3). For example, in the hedgerow intercropping trial in western Kenya mentioned above, farmers planted trials in a similar manner but most later modified such variables as the intercrop, pruning height, and pruning frequency.

The types of trials are not necessarily undertaken sequentially; researchers and farmers may decide to begin with a type 3 trial, or to simultaneously conduct two types of trials. For example, in the case of upper-storey tree trials in western Kenya (Franzel et al. 2002b), no type 1 or type 2 trials were needed because much was already known about the growth of the trees in the area. Rather, farmers planted type 3 trials in order to assess the performance of the species on their farms. In Zambia, many farmers planted type 2 and type 3 improved fallow trials in the same year (Kwesiga et al. 1999). They tested a particular set of practices in their type 2 trials and used type 3 trials either to extend their plantings or to test a modification of the practice. Researchers wished to assess biophysical response in the type 2 trials and to monitor farmers' innovations in the type 3 trials. Types 2 and 3 trials often generate questions or sharpen hypotheses about biophysical factors, which can then be best evaluated through type 1 on-farm or on-station trials. In western Kenya, several researcher-managed trials to explore

Table 1. The suitability of types 1, 2, and 3 trials for meeting specific objectives.

Information types	Type 1	Trials[†] Type 2	Type 3
Biophysical response	H[‡]	M	L
Profitability	L	H	L
Acceptability			
Feasibility	L	M	H
Farmers assessment of a particular prototype[§]	L	H	M
Farmers assessment of a practice	L	M	H
Other			
Identifying farmer innovations	0	L	H
Determining boundary conditions	H	H	H

[†] Type 1 = researcher designed, researcher managed; Type 2 = researcher designed, farmer managed; Type 3 = farmer designed, farmer managed.

[‡] H = high, M = medium or variable, L = low, 0 = none. The suitability involves both the appropriateness of the trial for collecting the information and the ease with which the information can be collected.

[§] By particular prototype, we mean a practice that is carefully defined. For example, a prototype of improved fallows would include specific management options such as species, time of planting, spacing, etc.

specific aspects of improved fallow function and design were set up following farmer-managed trials (Swinkels et al. 1997).

Handling complexity

Complexity is determined by the number and diversity of components (intercropping trees and crops, as opposed to trees or crops in pure stand), the length of the cycle of the technology (3+ seasons as opposed to single-season cycles), and the size of the trial (whether it takes up more than 10% of a farmer's cultivated area). In a trial comparing annual crop varieties, it is often possible to combine biophysical and socioeconomic objectives because according to the above definition, the trial is not complex. However, most agroforestry trials are complex and thus different trial types are needed to meet the differing objectives.

Promoting farmer innovation

Promoting farmer innovation is an often-mentioned objective of on-farm trials, yet little is written on how to achieve it. Type 2 trials require the standardizing of practices across farms, thus actually reducing farmers' motivation to innovate. Only in type 3 trials, where farmers completely control the experimental process, are farmer innovations likely to emerge and be captured. In type 3 trials on improved tree fallows in eastern Zambia, two of the main technological components being extended to farmers emerged from farmer innovations in type 3 trials (Kwesiga et al. 1999; Franzel et al. 2002c). In the first example, farmers were given potted seedlings, raised at farmer training centers, for planning improved fallows on their farms. In order to reduce the cost of transporting them to the farms, a farmer removed the seedlings from the pots and carried them bare-rooted in basins. When farmers' plantings of these seedlings proved successful, researchers conducted type 1 trials to compare the performance of bare-rooted seedlings grown in raised seedbeds with potted seedlings. They found no significant difference in performance and, as potted seedlings were much more costly to produce, they were phased out (Kwesiga et al. 1999).

Farmers' second main innovation, intercropping trees with maize during the year of tree establishment, was also later tested in on-farm trials. The trials found that intercropping reduces maize yields and tree growth during the year of establishment, but most farmers prefer it because it economizes on land and labor use relative to planting in pure tree stands.

Conclusions

The type 1-2-3 classification system is useful for highlighting the different objectives for conducting on-farm trials and for illustrating the suitability of different types of trials for particular types of assessments. It is tempting for researchers to use the same on-farm trial to collect information on biophysical responses and farmer assessment, however these objectives are often conflicting. A high degree of control is needed to collect accurate biophysical data, whereas farmer assessment is most valid when individual farmers are allowed to use the practice in the manner they see fit. Researchers and farmers interested in biophysical and socioeconomic data may be better off conducting type 1 trials for biophysical data and type 3 trials for socioeconomic assessment, rather than a single type 2 trial that tries to do both. The more complex the trial or technology, the less effective a type 2 approach is likely to be for both biophysical and socioeconomic assessments.

References

Ashby, J., T. Gracia, M. Guerrero, C. Quiros, J. Roa, and J. Beltran. 1995. *Institutionalizing farmer participation in adaptive technology testing with the 'CIAL'.* Overseas Development Institute Network Paper 57. London: ODI.

Biggs, S. 1989. Resource-poor farmer participation in research. The Hague: ISNAR.

Franzel, S., L. Hitimana, and E. Akyeampong. 1995. Farmer participation in on-station tree species selection for agroforestry: a case study from Burundi. *Experimental Agriculture* 31:27-38.

Franzel, S., R. Coe, P. Cooper, F. Place, and S.J. Scherr. 2002a. Methods for assessing agroforestry adoption potential. In S. Franzel and S.J. Scherr (eds.), *Trees on the farm: assessing the adoption potential of agroforestry in Africa.* Wallingford, U.K.: CAB International.

Franzel, S., J.K. Ndufa, C.O. Obonyo, T. Bekele, and R. Coe. 2002b. Farmer-designed agroforestry tree trials: Farmers' experiences in Western Kenya. In S. Franzel and S.J. Scherr (eds.), *Trees on the farm: assessing the adoption potential of agroforestry practices in Africa.* Wallingford, U.K.: CAB International.

Franzel, S., D. Phiri, and F.R. Kwesiga. 2002c. Assessing the adoption potential of improved tree fallows in Eastern Zambia. In S. Franzel and S.J. Scherr (eds.), *Trees on the farm: assessing the adoption potential of agroforestry practices in Africa.* Wallingford, U.K.: CAB International.

Gittenger, J.P. 1982. *Economic analysis of agricultural projects.* Baltimore: Johns Hopkins University Press.

Kwesiga, F.R., S. Franzel, F. Place, D. Phiri, and C.P. Simwanza. 1999. *Sesbania sesban* improved fallows in eastern Zambia: their inception, development, and farmer enthusiasm. *Agroforestry Systems* 47:49-66.

Lightfoot, C. 1987. Indigenous research and on-farm trials. *Agricultural Administration* 24:1-11.

Okali, C., J. Sumberg, and J. Farrington. 1994. *Farmer participatory research. Rhetoric and reality.* London: Intermediate Technology Publications.

Shepherd, K., R. Swinkels, and B. Jama. 1994. A question of management: the pros and cons of farmer- and researcher-managed trials. *Agroforestry Today* 6:4:3-7.

Sperling, L., M.E. Loevinsohn, and B. Ntabomvura. 1993. Rethinking the farmers' role in plant breeding: Local bean experts and on-station selection in Rwanda. *Experimental Agriculture* 29(4):509-518.

Stroud, A. 1993. *Conducting on-farm experiments.* Cali, Columbia: CIAT.

Swinkels, R., and S. Franzel. 1997. Adoption potential of hedgerow intercropping in the maize-based cropping systems in the highlands of Western Kenya. Part II: Economic and farmers' evaluation. *Experimental Agriculture* 33:211-223.

Swinkels, R., S. Franzel, K. Shepherd, E. Ohlsson, and J. Ndufa. 1997. The economics of short rotation improved fallows: evidence from areas of high population density in western Kenya. *Agricultural Systems* 55:99-121.

van Veldhuizen, L., A. Waters-Bayer, and H. de Zeeuw. 1997. *Developing technologies with farmers: A trainer's guide for participatory learning.* London: Zed Books.

Discussion Summary

The discussion focussed on the need to make the type 1-2-3 trial classification relevant in terms of participatory plant breeding. A key question was why do we need participatory on-farm trials in breeding? One answer was that participatory on-farm trials can generate information from many different environments at a lower cost. This was a recurrent topic throughout the workshop. The data generated from certain types of participatory trials (particularly type 1) can be useful for variety release committees, but this may require convincing and educating these committees on the usefulness and validity of data generated through this type of trial. Participatory trials can be a tool to gauge farmers' acceptability of different varieties and to provide data that cannot be generated through more conventional methods, such as performance under farmer management. These trials also can help to identify varieties that are appropriate for different niches. Moreover, they can be particularly useful for generating testable hypotheses.

Further discussion centered on the value and limitations of different types of trials, particularly type 3. It was argued that type 1 trials may be good for release proposals and type 3 for biophysical performance. While some felt that type 3 trials could not provide good biophysical data, it was pointed out that tapping into farmers' knowledge and their ability to recognize and characterize the niches where they farm could overcome the problems of characterization. In fact the need for proper characterization emerged as one of the most important topics for participatory trials. Characterization is important for controlling variability and provides the covariates that would allow for better interpretation of data generated in these trials.

Another important issue discussed was the need for and the limitation of measuring yield in types 2 and 3 trials. It is rare to obtain good quality yield measurements under the conditions prevalent in these types of trials. A question raised was whether direct yield measurements can be substituted with rating or ranking data from farmers involved in the trial. This topic generated much interest during the workshop. There is also a need for simple evaluation and measurement protocols that can be handled by farmers. Simpler protocols can enhance farmer participation.

A useful approach to stimulate type 3 experiments is to give farmers seed of particular varieties and then monitor the results. This could provide important information on acceptability and potential adoption. However, monitoring should be carried out for more than two years, since it may not be until the third year that farmers feel more confident and begin to appreciate the varieties.

Quantifying Farmer Evaluation of Technologies: The Mother and Baby Trial Design

SIEGLINDE SNAPP

Abstract

This paper presents five years of experience in Malawi, experimenting with a novel mother and baby trial design to systematically connect assessment of technologies by farmers with biological performance. This design consists of two types of trials. The "mother" trial is replicated within-site to test a range of technologies and research hypotheses under researcher management. This trial is either located on a research station or on-farm, e.g., at a central location in the village. The "baby" trial comprises a number of satellite trials (each trial is one replicate) of large plots under farmer management and farm resources. Each trial compares one to four technologies (usually a subset of those tested in the mother trial chosen by the farmer or researcher) with farmers' technologies/cropping systems. Researchers indicate the recommended management for each technology, then monitor actual farmer practice, and document farmer perceptions and ranking. Researchers test complex questions (e.g., variety response to inputs) at the central mother trial, while farmers gain experience with the subset of technologies. Farmer perceptions are systematically monitored together with biological performance of the technologies. Farmer participation in the design of baby trials can vary from limited to high, depending on the research objectives. This linked trial process provides quantitative feedback to researchers for improving the design of future technologies. In this study, different analytical tools were used in conjunction with the trial design. Biological performance data included "adaptability analysis", which consists of regressing yield or other data against an environmental index (average yield, soil factor, or others), analysis of variance and marginal rate of returns economic analysis, and evaluating benefits and risk aversion. Survey data included the use of descriptive statistics for different farmer group answers, analysis of the ranking of technologies by farmers, grouping answers from open-ended questions, and expressing each answer as a percentage of all answers.

Introduction

There has been limited adoption of improved seed and farming technologies by smallholder farmers in many regions of the world. According to the participatory research literature, one of the major barriers to uptake has been insufficient attention to understanding farmer priorities and perceptions (Chambers et al. 1989; Ashby and Sperling 1995). Researchers and extension staff are frequently aware that farmers need to be consulted and that indigenous knowledge should be documented, however the time and resources required for participatory research are seen as onerous (Snapp et al. 2002a). Rigorous and practical tools are urgently required to improve the process of participatory variety selection and technology development (Bellon 1998). This workshop was convened to address the need for quantitative methodology and statistical approaches to document farmer criteria and perceptions.

In this paper we discuss five years of experience in Malawi, experimenting with a novel mother and baby trial design to systematically connect farmer assessment of technologies with biological performance (Snapp 1999). Methodical cross-checking of performance evaluation by researchers and farmers provides complementary rather than competing information from conventional research and participatory processes (van Eeuwijk et al. 2001). We investigated the biological performance of intensified legume use within a maize-based system, and invented the mother and baby trial concept to test the potential for widespread adoption of these technologies by smallholder farmers in southern Africa (Snapp et al. 2002b).

The lessons regarding on-farm trial design and documenting farmer perceptions appear to have wider application than Malawi—the mother and baby trial design is meeting acceptance by many researchers in the region. Scientists from the International Maize and Wheat Improvement Center (CIMMYT) have recently adapted the trial design, using an incomplete lattice design for baby trials, to conduct hundreds of linked mother and baby trials in southern and eastern Africa (Bänziger and Diallo 2001). A survey of 30 participatory research scientists conducted in 2001 found that 11 were using the mother and baby trial design or were in the process of adopting it, which frequently included adapting it to local circumstances (Morrone and Snapp 2001). The primary reason cited for interest in the approach was the ability to systematically involve many farmers and to rapidly elicit evaluation of technologies and varieties.

On-Farm Trial Methodology

It is now over 20 years since the farming-system approach was initiated in southern Africa, and now research is primarily conducted on-farm (Heisey and Waddington 1993). Methods to document the biological performance and yield potential of varieties and technologies are widely known. For example, it is highly recommended that on-farm trials be conducted at representative, well characterized sites, so that results can be extrapolated to recommendation domains. In some cases researchers use trial designs on-farm similar to those conducted at research stations, with four or five replicated plots per treatment and a randomized complete block or similar design. Generally farmers are treated in a contractual manner, and this trial design can be an effective means for evaluating technology performance under edaphic conditions typical of a farming community.

Another widely used approach is to conduct a large number of on-farm trials to evaluate technology performance across a spectrum of environments (Fielding and Riley 1998; Mutsaers et al. 1997). This takes into account the variability of the heterogeneous environment that characterizes many smallholder regions. A trial design where each site acts as a replicate is one approach that allows many environments to be sampled (Mutsaers et al. 1997). Adaptability analysis and related statistical tools can use data from the many sites to evaluate technology performance across different environments. This may make it possible to detect which varieties perform best in a weedy environment or on acid soils, for example (Hildebrand and Russell 1996). Another recently developed tool for multi-environment trial data is multiplicative mixed models, which can be used to model genetic variances and covariances. These statistical approaches are illustrated by van Eeuwijk and colleagues (2001) for participatory breeding and variety selection in barley.

Despite the extensive on-farm experience of many research programs, there is still widespread inability to understand or take account of farmers' priorities. Farmers' production priorities are often assumed to focus on maximizing yields or financial returns, while in reality they may be concentrate on gaining the best return from a very small cash investment, or on maximizing food security (Snapp et al. 2002a). Tools to evaluate potential profitability of technologies from trial data are documented, such as partial budgeting to estimate economic returns (CIMMYT 1988).

In contrast to economic budgets, there is limited documentation of methodology that systematically involves farmers in technology evaluation. There are a few outstanding examples, however, such as the use of expert farmer panels to document farmer criteria and improve variety selection in West and East Africa (Sperling et al. 1993; Kitch et al. 1998). Other methods are described in newsletters, working papers, and other publications that are important, but can be difficult to access (Bellon 1998; Kamangira 1997). Here we describe an approach that facilitates and documents the hands on experience of farmers. This provides a relatively rapid and rigorous approach to systematically involving farmers in the development of best bet technologies or varieties. Researchers assess input from farmers through surveys, farmer ranking of technologies, and by monitoring farmer adaptations and spontaneous experimentation (Snapp et al. 2002a). Through the mother and baby trial design we catalyze and improve on the ongoing experimentation by farmers through a systematic process.

Mother and Baby Trial Case Study

The sites

Four agroecosystems for participatory research were chosen in Central and southern Malawi, where about 70% of the country's smallholder agriculture is practiced. The agroecosystems, with the study sites in parentheses, are:

1. Central Malawi: subhumid, mid altitude plain (Chisepo, Mitundu, and Mpingu)

2. Central Malawi: subhumid, high altitude hills (Bembeke)
3. Malawi lakeshore: semi-arid zone (Chitala and Mangochi)
4. Southern Malawi: subhumid, mid altitude plateau (Songani)

Mother and baby trial design

The "mother and baby" trial was named by one of the farmers involved in the trials. The "mother" trials test many different technologies, while the "baby" trials test a subset of three or fewer technologies, plus one control (Snapp 1999). The design makes it possible to collect quantitative data from mother trials managed by researchers, and to systematically crosscheck them with baby trials on a similar theme that are managed by farmers (Figure 1). The design is flexible: the mother trials described here were located on-farm at central locations in villages, but they can be located at nearby research stations (Snapp 1999). The level of farmer participation in baby trial design and implementation can vary from consultative to collaborative. We discuss here a consultative process where researchers lead the implementation of baby trials, however the role of farmer participation in baby trials can be much higher. For example, at the Bembeke site, the nongovernmental organization (NGO) Concern Universal has catalyzed greater farmer involvement, including in baby trial design (Figure 2).

This study started in 1996, when soil scientists and agronomists from the University of Malawi and the Malawian Department of Agriculture and Irrigation met to synthesize published information and results from years of on-farm research (Figure 3a). Reconnaissance surveys and village meetings helped to form the hypotheses that smallholder farmers have limited resources, use small amounts of mineral fertilizer, and experiment with alternative nutrient sources such as legume residues (Kanyama-Phiri et al. 2000). Researchers designed best bet technologies to improve soil productivity that required minimal cash

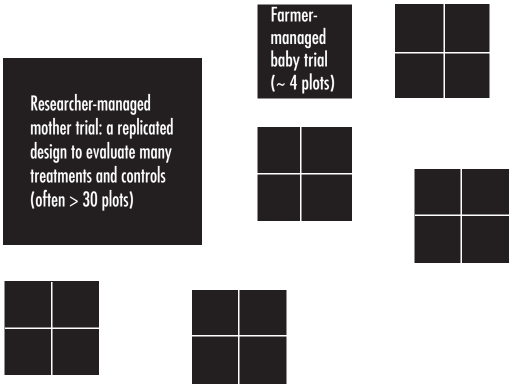
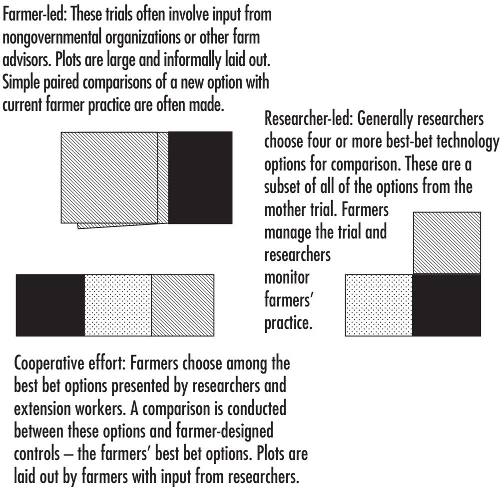


Figure 1. Mother and baby trial design layout. A mother trial is centrally located in a village or at a nearby research station and replicated on-site. Baby trials are located in farmers' fields and compare a subset of technologies or varieties from the mother trial. Each baby trial site is a replicate.



Farmer-led: These trials often involve input from nongovernmental organizations or other farm advisors. Plots are large and informally laid out. Simple paired comparisons of a new option with current farmer practice are often made.

Researcher-led: Generally researchers choose four or more best-bet technology options for comparison. These are a subset of all of the options from the mother trial. Farmers manage the trial and researchers monitor farmers' practice.

Cooperative effort: Farmers choose among the best bet options presented by researchers and extension workers. A comparison is conducted between these options and farmer-designed controls – the farmers' best bet options. Plots are laid out by farmers with input from researchers.

Figure 2. Different levels of farmer and researcher participation in the design and implementation of baby trials.

and labor. Representative villages in key agroecosystems were chosen on the basis of information from community meetings, consultations with extension staff, and government statistics on population density and agroclimatic data (Snapp et al. 2002a). The selected villages had to be representative of four major agroecozones and in terms of population density and access to markets.

The researchers involved in the mother and baby trials selected the "test" farmers in collaboration with community members at a meeting. They asked for volunteers and stressed the need to include both well-off farmers and those with few resources, as well as households headed by women. The trial design was geared to meet both farmers' and researchers' objectives, which by no means are identical. Relatively simple "one-farmer, one-replica" trials were managed by farmers as satellites or baby trials to a central mother trial, which was managed by researchers and had within-site replications (Figure 1). A trial design with a maximum of four plots and no replication within the farmer's field fits a limited field size, simplifies the design, and makes it easier for farmers to evaluate technologies.

Many replicates across sites make it possible to sample wider variations in farm management and environment (Fielding and Riley 1998; Mutsaers et al. 1997). However, replication within a site and intensive, uniform management improve research on biological processes. The mother and baby trial design is the first attempt we are aware of that methodically links "replicated within a site" researcher-led trials with "one site, one replica" farmer-led trials (Figure 1). Van Eeuwijk and colleagues (2001) advocate using both types of trials, but do not explore the deliberate, simultaneous use of the trials in a design that systematically links the two.

Technology evaluation in the mother-baby trial approach

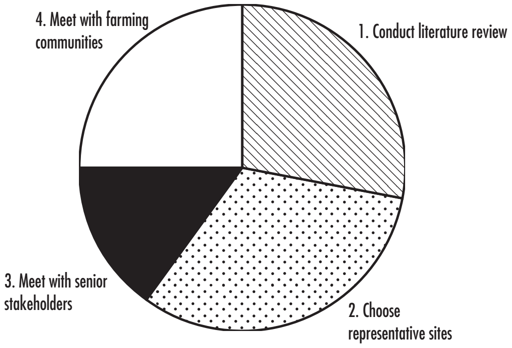
Farmers initially chose their test technologies on the basis of information given in introductory community meetings (Figure 3a). Descriptions of promising technology options were presented, and visits to research station trials were arranged where possible. Researchers and assistants provided supervision and interaction through monthly visits to sites. Enumerators were based at each site to assist with trial setup and measurements, in collaboration with local extension or NGO staff and farmers (Figure 3b). Training in participatory approaches and survey techniques to reduce bias was conducted at annual project meetings.

Plot size for mother and baby trials was approximately 8 m by 8 m. Ridges were prepared by hoe and placed about 0.9 m apart, following conventional practice. A wide range of cropping system technologies was compared to current farmer practice, as described in Snapp et al. (2002a). The mother trials were planted by extension staff with assistance from enumerators within 10 days of the arrival of the rainy season. It was interesting to note that farmers were very timely in planting their baby trials—in many cases they were planted before the mother trials.

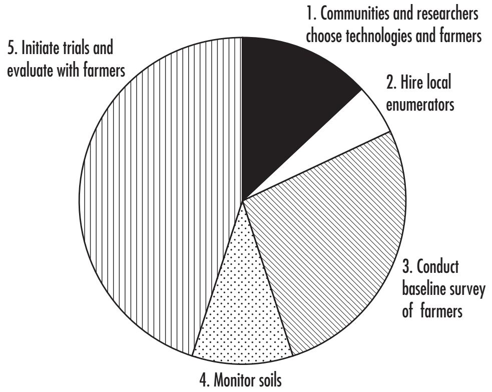
Data collected from trials included: plot size measurements, planting date, emergence date, population density at emergence, early weed cover, dates when plot was weeded (plots were weeded twice, approximately 5 and 10 weeks after planting), aboveground biomass of a subsample of legumes measured at flowering, harvest plant population, and grain yield at harvest. Fresh weight measurements were conducted in the field, and subsamples of 5-15 kg were collected to determine grain moisture content and dry weight to fresh weight conversions. Soil samples from the topsoil were collected at all sites, and soil pH, organic carbon, inorganic nitrogen, and texture analyses were conducted (Snapp et al. 2002a).

The farmers provided quantitative feedback on their evaluation of technologies to researchers through surveys, paired matrix ranking, and by rating technologies. Examples of the type of short survey and rating exercises used are presented by Bellon (1998). Qualitative feedback was obtained from meetings between farmers and researchers and comments recorded at field days. The mother trials were evaluated more informally during discussions held at field days. This made it possible to integrate the farmers' assessment and improve research priority setting (Figure 3c). Meetings were also held with senior stakeholders, conducted as part of an iterative process to maintain support and inform priority setting at every level. This included policymakers, supervisors of extension and NGO staff, senior researchers, and industry representatives (Figure 3c).

a) Year 1 mother/baby trials



b) Year 2 mother/baby trials



c) Year 3 mother/baby trials

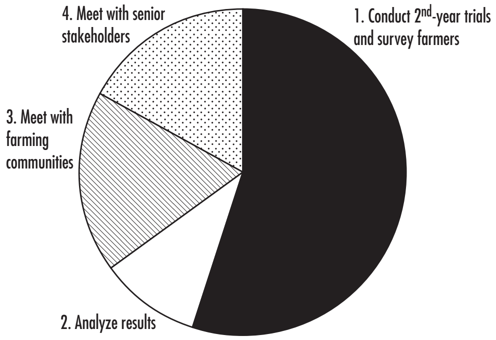


Figure 3. The sequence of steps in designing and implementing mother and baby trial methodology. Approximate time allocation for activities in a) year one, b) year two, and c) year three of the mother and baby trial approach.

Statistical analysis

Adaptability analysis was used for an initial review of all the data combined from mother and baby trials (Hildebrand and Russell 1996). This regression approach allows performance of technologies to be compared across a range of environments, where average yield or edaphic factors are used as an environmental index. Yield potential of varieties under stressed conditions can be reviewed through adaptability analysis, providing insight into the risk associated with different technologies. A more rigorous approach is provided by mixed models, such as factor-analytic models for modeling variance and covariance from multi-environment trial data (van Eeuwijk 2001). An incomplete lattice design for the baby trials allows CIMMYT scientists to systematically evaluate new stress-tolerant varieties of maize (Bänziger and Diallo 2001).

Our statistical analyses relied on the analyses of variance module of a statistical package (Statsoft 1995). The response of maize yield gain in year two of mother trials was evaluated through a two-way analysis of variance conducted for technology and location. Where technology effects were significant in the analysis of variance, a planned non-orthogonal comparison was used to evaluate mean technology effects compared to the control (continuous maize without nutrient inputs). A separate analysis of variance was conducted for baby trials, where a one-way analysis of variance was conducted to evaluate the effect of technologies. Descriptive statistics were conducted for farmer rating data, and means compared using paired t-tests (Taplin 1997).

Economic analysis

Economic analysis of net benefits was conducted over two years. This allowed comparison of best bet technologies that involved intercrop systems and rotation treatments requiring a two year evaluation period. The difference was computed between the value of maize and legume grain yields (total price benefits) accruing from fertilizer and legume seed inputs and costs (CIMMYT 1988).

Conclusion

By facilitating hands on experience for farmers, the mother and baby trials provide a relatively rapid approach to developing improved varieties and soil management technologies. In contrast to some approaches which merge objectives, such as research validation of technologies and farmer experimentation, the goal of the mother and baby trial approach is to facilitate communication across different approaches to experimentation and information flow among the partners. The linked trial design provides researchers with tools for quantifying feedback from farmers. Farmer input generated new insights, such as the need to broaden the research focus beyond soil fertility or variety selection to include system-wide benefits such as weed suppression. Some Malawi extension staff and researchers have expressed reservations about the time requirements for participatory approaches; however, the success of the approach is reflected in the uptake of the mother and baby trial design by researchers in ten neighboring countries.

References

Ashby, J.A., and L. Sperling. 1995. Institutionalizing participatory, client-driven research and technology development in agriculture. *Development and Change* 26:753-770.

Bänziger, M., and A.O. Diallo. 2001. Stress-tolerant maize for farmers in sub-Saharan Africa. In *Maize Research Highlights 1999-2000, CIMMYT*. www.cimmyt.org/Research/Maize/.

Bellon, M.R. 1998. *Farmer participatory methods for soil fertility research*. Soil FertNet Methods Working Paper No. 4. Harare, Zimbabwe: CIMMYT. 25 pp.

Chambers, R., A. Pacey, and L.A. Thrupp. 1989. *Farmers first: farmer innovation and agricultural research*. London, UK: Intermediate Technology Publications.

CIMMYT (International Maize and Wheat Improvement Center). 1988. *From agronomic data to farmer recommendations: an economic training manual*. Completely revised edition. Mexico, D.F.: CIMMYT. Pp. 78.

Fielding, W.J., and J. Riley. 1998. Aspects of design of on-farm fertilizer trials. *Experimental Agriculture* 34:219-230.

Heisey, P., and S. Waddington (eds.). 1993. *Impacts of on-farm research: Proceedings of a networkshop on the impacts of on-farm research in eastern and southern Africa.* CIMMYT Network Report No. 24. Harare, Zimbabwe: CIMMYT. 475 pp.

Hildebrand, P.E., and J.T. Russell. 1996. *Adaptability analysis: A method for the design, analysis and interpretation of on-farm research and extension.* Ames, I.A.: Iowa State University. 189 pp.

Kamangira, J.B. 1997. Assessment of soil fertility status using conventional and participatory methods. MSc thesis, Department of Crop Science, Bunda College of Agriculture, University of Malawi. 107 pp.

Kanyama-Phiri, G., S. Snapp, B. Kamanga, and K. Wellard. 2000. *Towards integrated soil fertility management in Malawi: Incorporating participatory approaches in agricultural research.* Managing Africa's Soils, Working Paper No. 11. London, U.K.: IIED. 28 pp.

Kitch, L.W., O. Boukar, C. Endondo, and L.L. Murdock. 1998. Farmer acceptability criteria in breeding cowpea. *Experimental Agriculture* 34: 475-486.

Morrone, V.L., and S.S. Snapp. 2001. Uptake of a new on-farm trial design that includes the small-holder farmer. *HortScience* 36:477 (abstract).

Mutsaers, H.J.W., G.K. Weber, P. Walker, and N.M. Fisher. 1997. *A field guide for on-farm experimentation*. Ibadan, Nigeria: IITA/CTA/ISNAR. 235 pp.

Snapp, S.S. 1999. Mother and baby trials: A novel trial design being tried out in Malawi. Target Newsletter of the Southern Africa Soil Fertility Network Vol. 17, pp. 8.

Snapp, S.S., G. Kanyama-Phiri, B. Kamanga, R. Gilbert, and K. Wellard. 2002a. Farmer and researcher partnerships in Malawi: Developing soil fertility technologies for the near-term and far-term. *Experimental Agriculture* (in press).

Snapp, S.S., D.D. Rohrbach, F. Simtowe, and H.A. Freeman. 2002b. Sustainable soil management options for Malawi: Can smallholder farmers grow more legumes. *Agriculture, Ecosystems and Environment* (in press).

Sperling, L., M.E. Loevinsohn, and B. Ntabomvura. 1993. Rethinking the farmer's role in plant breeding: Local bean experts and on-station selection in Rwanda. *Experimental Agriculture* 29:509-519.

Statsoft. 1995. Statistica for Windows. 1995. Tulsa, Oklahoma.

Taplin, R.H. 1997. The statistical analysis of preference data. *Applied Statistics* 46:493-512.

van Eeuwijk, F.A., M. Cooper, I.H. DeLacy, S. Ceccarelli, and S. Grando. 2001. Some vocabulary and grammar for the analysis of multi-environment trials, as applied to the analysis of FPB and PPB trials. *Euphytica* 122:477-490.

Discussion Summary

The discussion following the presentation dealt with the use of the mother-baby trial system in the context of participatory plant breeding, and was divided into two themes: (1) the technical advantages and disadvantages of mother-baby trials for selection and breeding, and (2) the role of mother-baby trials in formal research systems.

Discussing the first theme, it was pointed out that one of the main advantages of this system, particularly of the baby trials, is the number of trials that can be evaluated. Selection intensity relates directly to genetic advance, and the objective is to obtain high precision ranking. The choice is to evaluate a small number of plots more intensely or a larger number less intensely. It was pointed out that the use of incomplete multilocation trials may sacrifice some precision, but this is circumvented by having access to the appropriate environments. Furthermore, many environments can be sampled at a lower cost, although it was pointed out that in India, on-farm trials were more precise than station trials, and heritabilities can be higher. Farmers' knowledge of their fields and their heterogeneity can be used to design the baby trial to increase heritabilities. An issue that reappeared throughout the workshop was the appropriateness of having many unreplicated trials in many different locations versus having fewer replicated trials and therefore fewer locations. There appeared to be general agreement that the former option may be better because it may generate more useful information at a lower cost. Replication within a site may yield less information than sampling numerous sites, particularly from a cost-effective viewpoint.

The other point discussed was the use of mother-baby trials in the formal research system in relation to national agricultural research systems (NARS) and variety releases. A challenge is to get NARS to assess the value of these new tools and to incorporate this type of trial system, particularly in conjunction with participatory varietal selection (PVS), especially with PVS for variety release systems. Involving national programs in PVS may be the most straightforward way to link PVS with regulatory systems; PVS by the Centers of the Consultative Group on International Agricultural Research should not be done in isolation. The linking of PVS, innovative trial systems, and regulatory agencies is already underway in Nepal and Kenya. It was pointed out that regulatory agencies are more closely linked to formal seed systems than informal farmer-based systems in which PVS may take place. Regulatory committees may disfavor systems that are perceived as threatening, and hence lobbying is necessary to make the system more active. However, it is necessary that these committees do not perceive these new approaches as substitutes, but more as cost-effective complements to their work. There may be some resistance.

Analyzing Data from Participatory On-Farm Trials

Richard Coe

Abstract

Researchers conducting participatory on-farm trials, particularly variety selection trials, often have difficulty analyzing the resulting data. The irregularity of trial designs means that some of the standard tools based on analysis of variance are not appropriate. In this paper some simple extensions to analysis of variance, using general linear models and linear mixed models, are shown to facilitate insightful analysis of these awkward designs.

Introduction

Data from on-farm trials take many forms, from crop yields measured on individual plots to the reported consensus of participants at a group meeting. Any set of data comprising multiple observations that are not all identical will require some sort of statistical analysis to summarize the common patterns. Choice of appropriate analysis methods depends on:

1. The objectives of the analysis
2. The design (who compared what treatments or varieties under which conditions)
3. The type of measurements taken

In the second section of this paper I discuss different styles and objectives of analysis. A formal approach, similar to that commonly conducted, for example, on crop yields measured in a classical variety trial using analysis of variance and reporting variety means, has a role in the analysis of some participatory trials. The irregularity of designs often means that the well known methods may be inappropriate. In the fourth section I show how some extensions to the usual methods can be used. Many researchers report that results from on-farm trials are highly variable. The fifth section shows how some of this variation may be interpreted to gain further insight, particularly into differing responses in different situations, or genotype by environment interaction (GEI). Examples used to illustrate the methods are introduced in the third section. The methods described in this paper are appropriate for responses measured on a continuous scale, such as crop yields. The analysis of responses recorded as scores or ranks is the subject of a companion paper (see Coe 2, this proceedings).

The methods presented in this paper are neither new nor described in depth. Technical descriptions can be found in numerous publications including Kempton and Fox (1997) and Hildebrand and Russell (1998).

Approaches to Analysis

An assumption of this paper is that participation and the systematic collection, analysis, and interpretation of data are not contradictory activities. Among some practitioners there is a belief that adoption of a participatory paradigm removes the need, or even makes it impossible, for researchers to collect and analyze data. The purpose of participation is seen as empowerment of local people, which is inconsistent with researchers conducting activities that meet their own objectives. However, many researchers recognize that broad conclusions of relevance beyond the immediate participants are still necessary, and that a part of this research must be the collection and interpretation of data. Coe and Franzel (2000) summarize the research design principles that must still be followed if the research is to lead to valid inferences.

A participatory approach does, however, have implications for the collection, analysis, and presentation of data. Data collection is discussed in another section of this paper. Data analysis can be for, and to some extent by, different participants, each of whom will have their own interests and objectives. In the case of participatory crop breeding trials, participants may include farmers, researchers, extension staff, and regional planners. While a farmer is interested in making decisions about varieties to select for his/her farm, a regional planner might be interested in average performances, and a researcher in reasons for heterogeneous responses. Each will require a different type of analysis. As researchers are often also the facilitators of the whole process, it is their responsibility to ensure that each participant has the data they need in a useful format.

It is particularly important for a researcher to make data and results available to farmers. There are at least three reasons for this:

1. Farmers are supposed to be beneficiaries of the activities and can only benefit if information is given back to them.
2. Giving farmers results is a courtesy as they have made the research possible though their involvement.
3. Farmers can provide considerable insight into the analysis and results. It is very common to hear the complaint that data from on-farm trials are very variable. This variation is a reality, and understanding its causes should be an objective of the research. Such an understanding will eventually lead to improved farmer decision making. Farmers understand some of the reasons for the variation, and their insights can often provide a framework or hypotheses for analysis.

When plant breeders conducted classical, on-station experiments, the analysis performed often followed a standard pattern, for example, analysis of variance followed by tabulation of means and application of "means separation procedures". Often little attention was paid to exploratory analysis, designed to detect the main patterns and surprising observations. Nor was much effort made at imaginative presentation of results—researchers knew how to read the tables and they were the intended audience. When participatory approaches gained popularity, analysts made attempts to find interesting and informative presentations of data, but tended to forget about formal analysis, and, hence, sometimes reached invalid conclusions.

Of course both approaches to analysis are needed; they reinforce each other. Graphical and exploratory methods show the important results and reveal odd observations and unexpected patterns. Formal methods allow measures of precision to be attached to results and allow extraction of estimates from complex data structures. We cannot say that either of the approaches is better—both are needed to satisfy different roles. In this document I have concentrated on formal analysis, as requested. It is easier to find general methods and approaches of this type of analysis that can be described and applied in many situations.

Presentation and analysis are not the same. The method of presenting results depends on the nature of the result, the story they are to tell, and the audience. I am not aware of any work that shows that literate farmers find it easier to interpret graphs than numerical information; indeed, it seems likely that a simple numerical table may be more familiar than a quantitative graph.

The steps in analysis of any data set can be summarized as:

1. Define the analysis objectives. These drive the rest of the analysis. It is impossible to carry out a good analysis without clear objectives. Often the key graphs and tables can be defined at this stage, even without the results with which to fill them in.
2. Prepare the data. Data sets will have to be entered and checked, suitable transformations made (e.g., to dry weight per unit area), relevant information from different sources (e.g., farm household data and plot level yields) extracted to the same file, and so on.
3. Exploratory and descriptive analysis. The aim is to summarize the main patterns and notice further patterns that may be relevant.
4. Formal statistical analysis. The aim is to add measures of precision and provide estimates from complex situations.
5. Interpretation and presentation.

Iteration between the steps will be necessary. Training materials by Coe et al. (2001) provide more information on analysis of experiments.

A spreadsheet package such as Excel is good for much of the descriptive analysis. Its flexible facilities for data selection and transformation, tabulation, and graphics are useful. However, dedicated statistical software is needed for the analyses described here—they cannot be done in Excel. There are several packages with almost equivalent facilities. All examples given in this paper use Genstat (2000)—I often find it most convenient and easiest to understand, particularly as methods for different problems can be addressed with a similar set of commands. The key commands used to produce each analysis are included in the text with their output. SPSS is widely used by social scientists but is not particularly useful for the analyses described here.

Examples to Illustrate Analysis Methods

1. Soil fertility under agroforestry in Malawi

This is not a breeding trial but is included because the design is typical of many participatory on-farm trials. Three soil fertility strategies are compared over a number of years:

g	Mixed intercropping of maize and gliricidia
s	Relay planting of maize and sesbania
c	The control of continuous maize

Forty-one farmers each compared the control with one or both of the other treatments. Crop yield is the response of interest. A number of covariates were measured at the plot or farm level to help understand the reasons for variation across farms.

2. Maize varieties in Zimbabwe

This was a "baby" trial.[1] Twelve maize varieties were compared. A total of 146 farmers in 25 different sites took part, each testing 4 of the 12 varieties. The varieties tested were chosen by the researcher. Some household and field covariates were recorded. The actual crop yields obtained were not available for analysis, so the examples here use simulated yield data but the original field design.

Average Treatment Effects

Example 1

The starting point for the analysis should be simple explorations, such as the table of means below (created in Excel) that gives the mean yield for each treatment in the 1998 season, together with the number of observations.

	Data	
trt!	Average of yield98	Count of yield98
c	1.73	31
g	2.47	39
s	2.50	24
Grand total	2.23	94

The formal analysis has two general aims:

1. To improve the estimates. In this case we know that all treatments do not occur on each farm, so some adjustment for farm effects may be needed (see Example 2).
2. To provide measures of precision, i.e., standard errors and confidence intervals.

This is the role of analysis of variance and associated procedures in "regular" designs. The exact same ideas can be used here.

Genstat commands to complete the analysis are:

```
model yield98
fit [p=a;fprob=y] name+trt
predict trt
```

---

[1] The mother-baby trial design comprises a central researcher-managed "mother" trial, which tests all varieties, and farmer-managed "baby" trials, which test a subset of the varieties from the mother trial.

```
***** Regression Analysis *****
*** Accumulated analysis of variance ***

Change   d.f. s.s.     m.s.    v.r.  F pr.

+ name   38   168.6518 4.4382  13.39 <.001
+ trt    2    15.9187  7.9594  24.01 <.001
Residual 53   17.5691  0.3315

Total   93   202.1396 2.1735

Response variate: yield98

     trt       Prediction      S.e.

      c          1.6386       0.1066
      g          2.6235       0.0952
      s          2.3677       0.1240
```

Standard errors of differences (sed) can also be found. They are:

	sed
g-c	0.145
s-c	0.166
g-s	0.160

While this analysis is correct and technically efficient, it is possibly a little opaque! An alternative that is more easily understood is described as follows.

The researcher is interested in the comparison of treatments and in the change in performance (e.g., yield) realizable by changing from one treatment to another. Farmers are also interested in this comparison, though the criteria for comparison may be different. Experiments are designed to assess this change. It is therefore natural to approach analysis of the data by focusing on these changes. The steps are:

1. Choose a treatment pair, the comparison of which is of interest, e.g., g (maize intercropped with gliricidia) and c (monocropped maize).
2. For each farm on which this pair occurs, calculate the difference in response g-c.
3. Summarize this set of differences.

In this trial, 31 farms have yield data for the pair of treatments in 1998. The column of differences is y98g_c.

Summary statistics for y98g_c

```
Number of observations = 31
Number of missing values = 10
Mean = 1.008
Median = 0.841
Minimum = -0.739
Maximum = 2.712
Lower quartile = 0.400
Upper quartile = 1.766
Variance = 0.791
Standard deviation = 0.889
```

The mean difference of 1.008 has a standard error of √(0.791/31) = 0.16. A 95% confidence interval for the mean difference is thus 1.01 ± 2 x 0.16 = (0.69, 1.33). A statistical test of the hypothesis of no difference in mean yield from the two treatments would use the t statistic t = difference/se(difference) = 1.01/0.16 = 6.3. This mean, together with its standard error, is almost identical to that produced by the modeling analysis above. Differences are due to:

1. The modeling analysis uses part of the information from three farmers with sesbania and gliricidia but not the control treatment. [If we can estimate g-s and s-c within farms then we also estimate g-c = (g-s)-(s-c)].
2. All the data is used to estimate the residual variance, not just part of it.

The summary statistics above emphasize that observing the mean difference is only the beginning of the analysis. There is considerable variation in the difference across different farms that needs understanding and interpreting. This is the subject of the fifth section of this paper.

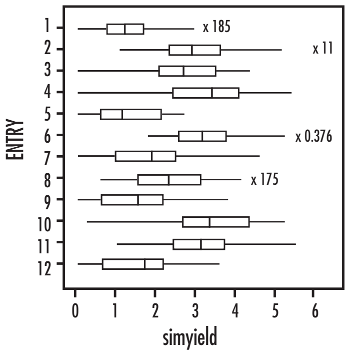
Example 2

The first step must be to check the data for errors and oddities. This is not illustrated. Next, simple summaries—numerical and graphical—are needed. The following table gives the mean, 25%, 50%, and 75% points for each entry, together with the number of plots from which it was calculated. Note: Excel is very good for this type of tabulation but cannot give the % points.

```
tabulate [class= ENTRY;p=nobs,means,quant;
percent=!(25,50,75)] data=simyield
```

```
ENTRY Nobservd  Mean   _25.0%  Median  _75.0%

  1      50    1.276   0.679   1.238   1.699
  2      47    3.077   2.344   2.909   3.639
  3      47    2.713   2.076   2.699   3.521
  4      50    3.305   2.416   3.473   4.083
  5      49    1.323   0.624   1.138   2.124
  6      49    3.371   2.594   3.195   3.792
  7      50    1.760   0.973   1.742   2.499
  8      49    2.429   1.573   2.362   3.143
  9      42    1.436   0.659   1.584   2.202
 10      51    3.448   2.708   3.401   4.380
 11      50    3.099   2.494   3.165   3.761
 12      50    1.597   0.677   1.788   2.206
```

Similar information is presented graphically in a boxplot:



This particular boxplot has highlighted some outlying observations that should be checked for possible errors.

These overall summaries are unlikely to be of interest to farmers in any one location, but the data from their neighborhood should be very relevant. A simple table of farm by entry for each site may be a useful discussion tool for this group of eight farmers, as it highlights both the variation between entries and variation between farmers testing the same things. It is likely that farmers can provide insight into reasons for the variation, which may help to direct formal analysis. For example, if farmers identify that some of the low yields come from plots known to be infertile, some measures of fertility should be built into the formal analysis. Farmers may also be able to tell you something about the tradeoffs between different assessment criteria, for example, expressing satisfaction with a variety that is not the highest yielding, but has some other desirable property. The data may need converting to units that farmers can use and understand.

SITE! 1

Average of simyield	FARM!								
ENTRY!	1	2	3	4	5	6	7	8	Grand total
1			2.03			1.70			1.86
2			3.39	2.43			2.63		2.82
3	1.51				2.66		2.11	1.81	2.02
4			4.97		3.36	4.01			4.11
5		0.28		0.29		1.55		0.74	0.72
6	3.06				2.35			1.96	2.45
7		0.45					1.82		1.13
8		2.00							2.00
9		0.00		1.77					0.89
10			4.47			3.15			3.81
11	2.06			2.40			1.02		1.83
12	1.40				1.79			0.40	1.20
Grand total	2.01	0.68	3.72	1.72	2.54	2.60	1.89	1.23	2.05

These Excel tables can be rearranged to clarify important information, for example, sorting by mean may make the table easier to read:

SITE! 1

Average of simyield	FARM!								
ENTRY!	1	2	3	4	5	6	7	8	Average
4			4.97		3.36	4.01			4.11
10			4.47			3.15			3.81
2			3.39	2.43			2.63		2.82
6	3.06				2.35			1.96	2.45
3	1.51				2.66		2.11	1.81	2.02
8		2.00							2.00
1			2.03			1.70			1.86
11	2.06			2.40			1.02		1.83
12	1.40				1.79			0.40	1.20
7		0.45					1.82		1.13
9		0.00		1.77					0.89
5		0.28		0.29		1.55		0.74	0.72
Average	2.01	0.68	3.72	1.72	2.54	2.60	1.89	1.23	2.05

The formal analysis of this data is needed to give means corrected for site and farm effects, together with correct standard errors of differences. The usual starting point would be an analysis of variance; however, the analysis has to account for expected variation due to differences between farms and sites, and the design used in the trial ended up with a rather irregular distribution of varieties across farms and sites. For example, in site 1 (see Excel table) entries occur between 1 and 5 times. The design is described as unbalanced (differing amounts of information about each treatment comparison) and treatments are non-orthogonal to farms and sites. The latter implies that treatment means adjusted for site and farm effects are more realistic summaries of treatment differences than raw means.

The need for some sort of adjustment is evident from the Excel table for site 1. Entries 5, 7, and 9 have low means; however, they all occur on farm 2, which may be a poor farm, hence depressing the means for these entries. Calculation of these adjusted means is described below. The results, which only include data for site 1, show that the ranking of entries is changed considerably, but the logic of the changes is visible if compared with the data. For example, entry 1 has the lowest adjusted mean. The raw data shows that this entry appeared on just two farms, both of which seem (compared to the performance of other entries) to be good.

Entry	Raw mean	Adjusted mean
4	4.11	3.48
10	3.81	2.94
2	2.82	2.46
6	2.45	2.57
3	2.02	2.16
8	2.00	3.20
1	1.86	1.00
11	1.83	1.88
12	1.20	1.32
7	1.13	1.83
9	0.89	1.48
5	0.72	1.03

The adjusted means are found by fitting a model with farm and entry effects. This model can be used to predict the performance of each entry on each farm, and the adjusted mean is then the average of these predictions across all the farms. The commands to do this in Genstat are simple, the last one being needed to obtain the standard errors of differences between adjusted means. The results below are for the whole data set, not just site 1.

```
model simyield
fit [p=*] FARM+ENTRY
predict ENTRY
rpair !P(ENTRY)
```

```
Response variate: simyield
       ENTRY    Prediction     S.e.

          1        1.234      0.107
          2        2.878      0.111
          3        2.612      0.111
          4        3.328      0.107
          5        1.483      0.108
          6        3.305      0.108
          7        1.834      0.107
          8        2.423      0.108
          9        1.488      0.118
         10        3.409      0.107
         11        3.167      0.107
         12        1.667      0.107

796 rpair !P(ENTRY)
```

```
***** Pairwise Differences *****

***** Regression Analysis *****

Response variate: simyield
         Fitted terms: Constant + FARM + ENTRY


Standard errors of pairwise differences

          1             *
          2          0.1549           *
          3          0.1560        0.1568           *
          4          0.1534        0.1569        0.1561           *
          5          0.1560        0.1561        0.1574        0.1528           *
          6          0.1543        0.1548        0.1564        0.1547        0.1535
          7          0.1535        0.1574        0.1561        0.1511        0.1562
          8          0.1533        0.1587        0.1570        0.1542        0.1565
          9          0.1565        0.1613        0.1608        0.1618        0.1617
         10          0.1524        0.1565        0.1599        0.1490        0.1486
         11          0.1557        0.1531        0.1518        0.1506        0.1518
         12          0.1494        0.1544        0.1541        0.1548        0.1544

                        1             2             3             4             5

          6             *
          7          0.1538           *
          8          0.1550        0.1512           *
          9          0.1621        0.1600        0.1612           *
         10          0.1536        0.1500        0.1494        0.1639           *
         11          0.1516        0.1531        0.1550        0.1643        0.1549
         12          0.1524        0.1523        0.1525        0.1605        0.1543

                        6             7             8             9            10

         11             *
         12          0.1562           *

                       11            12
```

Note that the sed values are not all the same due to the irregularity in the design; however, they are close enough for it to make sense to quote a single sed of 0.16.

If these adjusted means are compared with the raw means, the differences are not as great as when we analyzed just one site. The means are averages over a greater number of farms, so the effects of “good” and “bad” farms on individual means tend to cancel out.

Entry	Raw mean	Adjusted mean
10	3.45	3.41
6	3.37	3.31
4	3.30	3.33
11	3.10	3.17
2	3.08	2.88
3	2.71	2.61
8	2.43	2.42
7	1.76	1.83
12	1.60	1.67
9	1.44	1.49
5	1.32	1.48
1	1.28	1.23

In this case the model could also have been fitted as:

```
model simyield
fit [p=a] SITE/FARM+ENTRY

***** Regression Analysis *****

*** Accumulated analysis of variance ***

Change      d.f.  s.s.      m.s.     v.r.
+ SITE      24    189.0435  7.8768   16.57
+ SITE.FARM 121   327.6509  2.7079   5.70
+ ENTRY     11    289.1360  26.2851  55.28
Residual    427   203.0184  0.4755

Total       583   1008.8488 1.7304
```

This analysis of variance can be interpreted in the usual way, and shows that some of the between-farm variation actually occurs between sites. In other words, farms within a site tend to be more similar than farms on different sites, as expected.

The analysis presented above is valid; however, it does not capture all of the information in the data and hides some of the structure. An alternative approach is to treat sites and farms within sites as if there were a random selection from those available, and to use a model that describes this. REML procedures handle these problems and are easy to use in Genstat.

```
VCOMPONENTS [FIXED=ENTRY] RANDOM=SITE/FARM
REML[PRINT=model,components,waldTests,means;
PSE=differences] simyield
```

The option `FIXED=ENTRY` specifies that we want to estimate separate means for each of the entries. The parameter `RANDOM=SITE/FARM` tells Genstat that there are sites that are expected to vary and there are farms within each site that also vary. Genstat automatically adds the plot level or residual variance, but this could be explicitly put in if the data set had another factor labeled `PLOT` by specifying `RANDOM=SITE/FARM/PLOT`. The output is shown below.

Note that the trial was originally planned with a "replicate" being a set of all of the varieties (spread across three farms) with three replicates per site. However, due to a lack of available land as well as some mistakes, this is not how the design was implemented. Replicates therefore do not correspond to any physical source of variation in the experiment, and thus it does not make much sense to include them in the analysis. On the other hand, both sites and farms correspond to physical layout factors that could reasonably be expect to influence results, so these must be allowed for.

```
***** REML Variance Components Analysis *****

Response Variate  :       simyield

Fixed model      :        Constant+ENTRY
Random model     :        SITE+SITE.FARM

Number of units  :        584

* Residual term has been added to model

*** Estimated Variance Components ***

Random term   Component     S.e.

SITE          0.2516        0.0992
SITE.FARM     0.3535        0.0616
```

```
*** Residual variance model ***

Term          Factor        Model(order)     Parameter     Estimate      S.e.

Residual                    Identity         Sigma2        0.475         0.0325

*** Wald tests for fixed effects ***

Fixed term                  Wald statistic   d.f.          Wald/d.f.     Chi-sq prob

* Sequentially adding terms to fixed model

ENTRY                       663.07           11            60.28         <0.001

* Message: chi-square distribution for Wald tests is an asymptotic approximation (i.e.,
for large samples) and underestimates the probabilities in other cases.

*** Table of predicted means for Constant ***

        2.455    Standard error: 0.1165

*** Table of predicted means for ENTRY ***

   ENTRY       1         2         3         4         5         6         7         8
           1.308     2.984     2.681     3.369     1.495     3.377     1.858     2.478

   ENTRY       9        10        11        12
           1.528     3.469     3.205     1.704

Standard error of differences:
                      Average   0.1510
                      Maximum   0.1585
                      Minimum   0.1457

Average variance of
differences:                   0.02281
```

The first part of the output reports variance components, which are interpreted in the next section.

The Wald test is equivalent to the F-test for treatment effect in the usual anova. The “highly significant” effect says that there are real differences between these 12 variety means.

The table of predicted means gives means for each entry adjusted for farm and site effects. In this case most of the means are close to the unadjusted means, however, this will not always be so. The adjustments allow for the fact that some farms are better (produce higher average yields) than others. Entries that are tested on “good” farms will have their means biased upwards compared with entries tested on “bad” farms. In this design each entry is tested on about 50 farms, so the good and bad farms tend to cancel out; however, if there were fewer farms this would not be the case. The predicted means are those that should be reported and interpreted, not the raw means presented earlier.

The sed values for comparing predicted means are not all equal, so Genstat reports the minimum, maximum, and average. They are not equal because different pairs of means are compared with different precision. For example, counting shows that entries 1 and 2 occur together on the same farm 14 times, whereas entries 9 and 10 occur together on the same farm only 5 times.

We would therefore expect the sed for comparing entries 1 and 2 to be lower than that for comparing 9 and 10. In this case the range in sed values is not large, so we do not go far wrong if the average (or, more conservatively, the maximum) is used.

The output does not contain information that indicates which entries differ from each other; it only shows that there are some overall variety differences. We have not included any information about possible differences between entries in the analysis, so the only possibility would be an analysis based on ignorance, for example, one with letters attached to varieties deemed to be not significantly different from each other. There are both technical and philosophical problems with this approach and it should be avoided.

Suppose that the entries came from three groups, depending on pedigree, as follows:

Group	a	b	c
Entry	1, 5, 7, 9, 12	4, 10, 11	2, 3, 6, 8

Then we can look for differences between and within groups by replacing the fixed model by `FIXED=GROUP/ENTRY`.

```
*** Wald tests for fixed effects ***


Fixed        Wald
term      statistic  d.f. Wald/d.f.  Chi-sq
prob

* Sequentially adding terms to fixed model

GROUP       602.80     2    301.40   <0.001
GROUP.ENTRY 60.27      9     6.70    <0.001

* Message: chi-square distribution for Wald
tests is an asymptotic approximation (i.e.,
for large samples) and underestimates the
probabilities in other cases.
```

```
*** Table of effects for GROUP ***

 GROUP       a          b          c
           0.000      2.061      1.676

Standard error of differences: Average
0.1506
                    Maximum    0.1521
                    Minimum    0.1490
```

The Wald tests show that there is considerable variation between groups of entries, but still some remaining variation between entries within a group. The table of effects for `GROUP` summarizes the difference between groups—entries in group b have mean yields 2.06 higher than those in group a.

Comparing approaches

In Example 1 we based an analysis of the difference between yields of two treatments on either a linear model or the set of difference within each farm. The two methods produced almost identical results. So why not use the difference method illustrated in Example 2? Some of the reasons are discussed below.

Of the three treatments in Example 1, there are three pairs of treatments that could be used to form differences, hence, we might repeat the analysis three times. These analyses are not independent but that does not matter. However, with the 12 treatments in Example 2 there are 12 x 11/2 = 66 pairs that we could choose to make differences. Analysis of all these would not only be tedious, it would involve a lot of repetition (there are only 11 df in 12 treatments). But which subset of pairs should be chosen?

The set of treatments on any farm is small—only 4 out of 12. Thus, for example, treatment 1 occurs on 50 farms and treatment 2 on 47, yet they occur together on only 14. So if we work with the entry 1-entry 2 difference, we would use data from just 14 farms. However there is a lot more information about the two treatments that is reflected in the differing sed values from the two approaches. Modeling gave a sed of 0.155 for entry 1-entry 2 and the difference method gave a sed of 0.180. This difference may seem small but equates to a 42% increase in trial size. Other limitations of the difference methods will be described later.

The difference between the two analyses (i.e., between the analysis that takes farms and sites as fixed and the REML analysis, which takes farms and sites as random) lies in what can be reasonably assumed about farm and site differences. If they are slightly different, but we can make no realistic assumptions about the nature of those differences, then they should be considered fixed. This means that each site or farm has its own characteristic mean, unconnected with any other, which has to be estimated. Information on treatment differences then comes from differences within each farm. However, if sites or farms can be considered a random sample from the set of possible sites or farms, and have effects which roughly follow a normal distribution, then we estimate the variance of that normal distribution. This changes the estimates of the treatment effects because between-farm and between-site information is recovered. The source of this information can be understood as follows: if all farms that had treatment 1 had a high mean, and all those that had treatment 2 had a low mean, it could be concluded that treatment 1 is better than treatment 2. If farms really are a random sample, however, then treatment 1 is unlikely to end up on all of the best farms by chance. Hence some information from the farm effects needs to be added to our evidence that treatment 1 has a higher mean than treatment 2. The REML method combines this information with the within-farm information, which modifies the estimates of treatment effects and sed values compared with the earlier fixed effect analysis. If the assumptions of the random site and farm effects are realistic, then this analysis will always be more efficient.

Understanding Variation and Genotype x Environment Interaction

Example 2

The analysis above has produced estimates of variance components as follows:

Component	Estimate	Std error
SITE	0.2516	0.0992
FARM	0.3535	0.0616
PLOT or residual	0.4750	0.0325

What do these tell you?

The model used to analyze the data, as specified in the `VCOMPONENTS` command, is:

yield = mean + site effect + farm effect + variety effect + residual

The residual is thus the deviation of an individual plot yield from the average for that site, farm, and variety. It encompasses all of the unexplained variation from plot to plot, due to local environmental effects (soil, pests), management, measurement error, and so on. The variance of 0.475 means that the standard deviation of this plot-to-plot variation is √(0.475) = 0.698. If the data have an approximately normal distribution, then most observation lie within 2 sd of the mean. Thus the plot-to-plot variation represents variation of approximately ±1.4 about the mean for a farm growing a uniform variety. This is a typical level of variation in such trials.

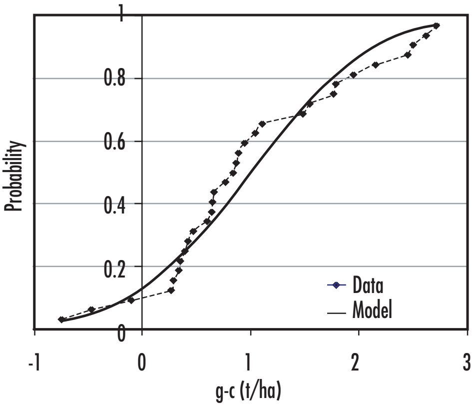
The farm variance can similarly be interpreted. It shows how much the average yield for a very large number of plots varies between farms within the same site.

Explaining variation—interaction and risk

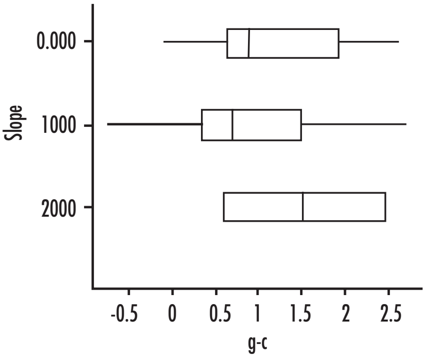
Example 1. In the last section we analyzed Example 1 by taking the 31 differences in yield for g-c and looking at their mean and variation. Here I take this analysis further.

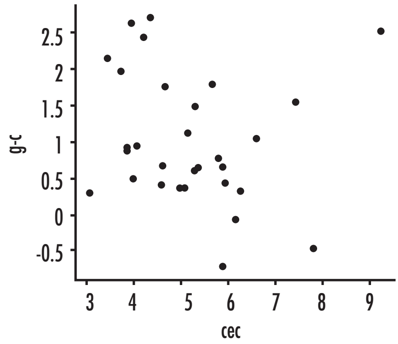
The mean difference of 1.01 t/ha is naturally of interest in some analyses, and this is the quantity most often reported, together with a proud statement that it is "significantly greater than zero". This is not of interest, however, to an individual farmer. A farmer's decision on whether to use g rather than c will depend on many things, whereas the yield component of the decision will be based on the yield increase he/she might achieve on his/her farm. In the absence of any other information, the mean is the best estimate of what this might be, but there is, of course, a lot of variation around the mean. This variation is an indication of the level of risk associated with a mean-based decision. In the figure below, the risk of obtaining a yield increase less than any specified amount is plotted. There is an approximate 10% chance that a farmer will achieve a lower yield based on g rather than c and a 55% chance of achieving an increase of less than 1 t/ha; however, 20% of farmers achieved an increase of more than 2 t/ha. A simple model for the variation is obtained by assuming a normal distribution, also shown on the graph. It is not a particularly good fit but still has some value, which is explained later. Note that if there were many more than 31 farmers in the study, we would expect a better (more precise) estimate of the mean difference between g and c, but no reduction in the variation in this difference across farms. More farms would give a better estimate of the chance of achieving a lower yield with g than c, but it would still be around 10%.

Knowing what distinguishes a +2 ton farmer from a +0 ton farmer is important, both for the farmer's decision making and the researcher's understanding.



An approach to the problem should be clear. We have a set of 31 differences and we want to know what determines them. Hypotheses of possible causes may come from farmers or researchers. The hypotheses are tested by collecting suitable data and statistical analysis. In Coe 2 (this proceedings), slope and cec (cation exchange capacity) are hypothesized causes of the variation in this example, so we can explore evidence for this in the data. Slope, in this case, is a categorical variable. The boxplot below shows little evidence of a consistent difference in the size of g-c for different slope categories. The scatter plot of g-c against cec does not show a clear relationship, but does show some outlying points that could be followed up. For example, the farm in the top right of the scatter used fertilizer, which suggests further ideas for investigation.





Note that what we are doing here is identifying GEI, where the G is the three treatments and E is characterized by slope and cec.

A formal statistical analysis would now use usual regression modeling approaches to quantify any effects. If y_{ij} is the yield on farm i under treatment j, then the differences being analyzed are:

$$d_i = y_{ig} - y_{ic}$$

with variance σ_d^2. This is the variation reflected in the above graph and in the simple risk model.

A regression model to look at the effect of a farm level covariate x would then be:

$$d_i = c + b_{gc}x_i + e_i$$

Here b_{gc} is the regression effect when considering the g-c difference and the residual e_i. The variance of the residual is σ_r^2. This measures the still unexplained variation in d, or the risk still remaining with knowledge of the covariate. Again, if a normal distribution model is acceptable, then the parameters of the regression model with σ_r^2 allow predictions of the risk of yield changes associated with switching from c to g conditional on the value of the covariate.

The usual analysis of variance model for this data, with treatments and farms in the design, would be:

$$y_{ij} = c + f_i + t_j + e_{ij}$$

with the variance of these residuals σ^2. Then the g-c differences are:

$$d_i = t_g - t_c + e_{ig} - e_{ic}$$

The connection between the analysis of variance approach and the analysis of plotwise differences now becomes clear: the variance of the differences $\sigma_d^2 = 2\sigma^2$. The effect of the covariate could be included in the analysis of variance model as:

$$y_{ij} = c + f_i + t_j + b_j x_i + e_{ij}$$

The term $b_j x_i$ describes how the treatment effect is modified on farms of different types (i.e., with different values of the covariate x). It is thus a treatment by farm interaction and is often the basis of the most useful results from an on-farm trial. With information on such interactions we can refine predictions and recommendations and reduce the risk associated with decisions based on the data. The covariates useful for this may be social variables (gender, household size, etc), biophysical variables (soil type, slope, etc), or management variables (weeding, planting time, etc).

Note that a common misunderstanding in experimental design is that farm x treatment interaction cannot be detected if only a single replicate is placed on each farm. The types of farm x treatment interaction that are important are those that are structured to show consistent patterns across farms. These can be explained and predicted in terms of explanatory variables, and can be estimated from designs with no more than one replicate per farm, as shown here, though this does not mean that design is unimportant. Also, more effective designs can be used if it is known which covariates will be of interest before the trial starts.

The analysis above identifies and describes what has always been known by breeders as GEI. The classical approach to this has been a "complete" trial in a number of locations, each representing different environments. Once a variety x location interaction is detected, an attempt is made to find which aspects of the environmental variation are responsible for the interaction. The approach used here allows GEI to be detected and described when only a subset of the genotypes is tested in a large number of locations, each genotype in an unreplicated trial. The approach does require that the locations be characterized by measurement of appropriate covariates. One reason for undertaking participatory breeding trials is that critical GEI is due to varying social or economic environments. For example, it is often hypothesized that men and women will favor different varieties, or that farmers' assessment of genotypes will depend on level of market integration. These types of interaction can be detected and described as long as the design covers sufficient variation, and suitable indicators of the social or economic variables are recorded.

Summary

The key points made in this paper are:

- Analysis of data from participatory trials can and should use a combination of exploratory/descriptive methods and formal statistical modeling.
- The analysis may be complicated by the irregular layout of the experiment and multiple layers of variation introduced by the hierarchical design.
- Approaching the analysis by calculation of treatment contrasts on each farm can simplify many complex problems and lead to new insights into the data; however, it can be inefficient or too repetitive if there are many treatments.
- Approaching the analysis by fitting regression models or their equivalent with multiple error terms allows many designs to be analyzed within a common framework; however, the analysis can be opaque and estimates non-intuitive.

- The two approaches can often be made to equate.
- The most useful analysis is often one that concentrates on finding explanation for variation in treatment effects across farms.
- Variation (at any level in the design) can be interpreted as risk, not just as unexplained noise.

References

Coe, R., and S. Franzel. 2000. Keeping *research* in participatory research. Presented at the 3rd International PRGA Conference, 6-11 November 2000, Nairobi, Kenya.

Coe, R., R. Stern, and E. Allen. 2001. *Analysis of data from experiments. Training materials*. Nairobi: SSC/ICRAF. 250 pp.

Genstat. 2000. *Genstat for Windows*. 5th Ed. Oxford: VSN International.

Hildebrand, P.E., and J.T. Russell. 1998. *Adaptability analysis*. Iowa, U.S.A.: Iowa State University Press.

Kempton, R.A., and P.N. Fox. 1997. *Statistical methods for plant variety evaluation*. London: Chapman and Hall.

Discussion Summary

The discussion following the presentation dealt with questions on data analysis, analysis of genotype x environment interaction (GEI), farmers' involvement in trials, and the statistical packages available to analyze results. In terms of data analysis, a common problem is the variation in the number of times a given entry is included in a trial (e.g., one to four). In other words, if the performance data for a variety was recorded only once, should this information be eliminated? The answer is no: the alpha lattice method (REML) makes an adjustment following assessment of the robustness of different data points, and the resulting adjusted means are more robust. It is also important to include zero as a response if, for example, the plot matured but there was no yield, but not if it did not yield due to external factors. Sensitivity tests can be run to determine the course of action with respect to outlying data points. An analysis can be run with and without these, but if the data point is very influential, the cause needs to be considered, as it may be necessary to repeat the trial. Another question was if participatory varietal selection (PVS) trials consist of two entries (one of which is the local check), could adjustments be made with respect to this control? The answer is no, since this would build uncertainty into the results because the performance of the check is variable. A related question was raised on how to use the differences in performance (yield) between entries and a control? And can these differences be used as a comparison across varieties? Are there guidelines to use the differences? The answer is that it is necessary to ask, "What can I see in the set of differences?" For example, look at the average and the size of the differences, and use graphics that allow the visualization of the results.

The issue of GEI is very important. A complete table of environments, farms, and sites (locating and enabling interpretation of crossover effects) is more appropriate for studying GEI. There are many tools that can be used to address this issue, but first it is important to know what constitutes the environment. This can be done using covariates, which also allow better hypotheses testing and interpretation. It was pointed out that most treatments overfit the data by making each trial a different environment; however, a trial *samples* a population of environments, it is not an environment. Hence, trials should be grouped according to similarities, and the resulting groups used as the environments for the analysis.

Farmer involvement in the interpretation and analysis of trials helps in two ways: it puts the information in context and provides useful explanations of the results. This can be achieved with a farmer focus group, where the results are presented and discussed. An important question is how to present the results to farmers, particularly when the trials are very extensive and located over a large area. This may require the involvement of local extension workers and simple representation of results for analysis. There was discussion on whether to use simple tables, charts, or even a physical representation of yield, e.g., bags. Bags can be cumbersome, and it was noted that tables are usually easier for farmers to interpret than charts. It is very important that farmers understand the purpose of the trial and what is being assessed—some sort of training may be required. Lack of understanding may lead to the generation of inaccurate or unimportant information. Worse still, it may lead to inappropriate actions by farmers which may invalidate the experiment, for example, by spraying one plant to protect it, when the purpose of the experiment was to assess the resistance of two varieties to a pest or pathogen.

The final discussion point centered on the availability of statistical programs and tools for breeders from national agricultural research programs to conduct analyses. Many of the available analysis programs are expensive, although countrywide licensing may be possible. It is important to assist the national programs in accessing affordable software. Further training in the software may also be required.

Sources of Variation in Participatory Varietal Selection Trials with Rainfed Rice: Implications for the Design of Mother-Baby Trial Networks

GARY ATLIN, THELMA PARIS, AND BRIGITTE COURTOIS

Abstract

Little information has been published on the repeatability of participatory varietal selection (PVS) trials. Repeatability estimates, which can be derived from the combined analysis of trials over locations and years, are useful for determining the number of replications and the optimal blocking structure of PVS trials. Variance components were estimated from a series of upland and lowland PVS trials conducted in the states of Jharkand and Bihar in eastern India, and used to estimate the repeatability of means. In both sets of trials the cultivar x site x year variance component was larger than the cultivar x site component, indicating that there was little specific adaptation to sites within the trials series. Participatory varietal trials conducted on-farm under farmer management were quite repeatable; replication over 5 sites was predicted to result in a repeatability of more than 0.5 in both data sets. Simulation indicated that a modest benefit is likely from the use of alpha-lattice designs when among-farm variances are large in experiments conducted using the mother-baby design, which treats farms as incomplete blocks.

Introduction

Breeders of rainfed rice in eastern India recognize the need to introduce participatory methods into their variety testing systems to improve the effectiveness of breeding programs. Performance in farmers’ fields and in the breeder’s nursery can be thought of as correlated traits expressed by a single genotype in separate environments.

Theory developed by Falconer (1989) and extended to the analysis of plant breeding programs by Pederson and Rathjen (1981) and Atlin and Frey (1989; 1990) permits breeding strategies to be evaluated on the basis of the predicted response in the target environment resulting from selection conducted in a breeding nursery. When selection is among pure lines, this response may be modeled using the formula:

$$CR_T = i_S r_G \sqrt{H_S H_T} \sigma_P \qquad (1)$$

where CR_T is the correlated response in the target environment (farmers' fields) to selection in a breeding nursery; i_s is the standardized selection differential applied in the selection nursery; r_G is the genotypic correlation between cultivar yields in the selection and target environments; H_S and H_T are repeatabilities or broad-sense heritabilities in the selection and target environments, respectively; and σ_P is the phenotypic standard deviation in the target environment. When response is being predicted for a particular target environment, H_T and σ_P may be considered constants. Therefore:

$$CR_T \propto i r_G \sqrt{H_S} \qquad (2)$$

Inspection of this relationship indicates three important considerations for designing breeding programs for stress environments:

1. i must be maximized by screening large populations, permitting a high selection intensity to be achieved.
2. r_G (or accuracy) must be maximized by ensuring that performance in the selection environment or screening system is highly predictive of performance in the target stress environments.
3. A high level of H_S (or precision) must be achieved, typically through replicated screening.

One reason for the poor performance characterizing many conventional rainfed rice breeding programs is that the research conditions are not reflective of on-farm conditions; in other words, r_G is low. In participatory varietal selection (PVS) programs, the genetic correlation between performance in the selection and target environments is very high, since selection is conducted in farmers' fields (Atlin et al. 2001). Therefore, the main factor affecting response to PVS in programs of a particular size is H_S. However, the scale and design of PVS schemes needed to achieve acceptable H_S levels is unknown, because no information has been published on the extent of farm to farm variation in cultivar performance in PVS experiments. Variance component estimates from the analysis of PVS trials over locations and years can be used to estimate H of means for grain yield and other agronomic characteristics resulting from a given number of sites and years of testing. These estimates can be used to determine the scale of testing needed to achieve adequate precision from PVS trials and the best method of analysis for PVS programs using the mother-baby model, which treats individual farms as incomplete blocks.

The International Rice Research Institute (IRRI) has conducted PVS trials in rainfed rice over three years in several villages in eastern India. The original objective of these experiments was to compare varietal rankings within and among groups of farmers and breeders (Courtois et al. 2001), but the trials also provide information on the sources of variation for agronomic traits in PVS trials conducted with rainfed rice. This report presents variance components estimated from the combined analysis of on-farm PVS trials over farms and years in two regions in eastern India and their use in estimating the repeatability of means from rainfed rice PVS trials. The implications of these estimates for the design of mother-baby trial networks are considered.

Mother-baby PVS trial networks are now being planned or implemented by several research groups in India. The mother-baby design has two components: the mother trial, in which a complete set of cultivars is evaluated in replicated researcher-managed trials at several locations; and the baby trials, wherein farmers each evaluate a subset of the cultivars tested in the mother trial. Villages and farms within villages may be considered separate blocking strata within a mother-baby trial. Variation in mean yield among farms within villages is expected to be substantial. This variation contributes to the variance of cultivar means when farms are used as incomplete blocks, and can be controlled to some extent by designs that control within-block variation, such as the alpha-lattice design. In establishing these trials, we have found that the lack of easily accessible software for the analysis of alpha-lattice designs is a serious constraint. Sets of baby trials may be analyzed as randomized-complete-block (RCB) design or completely randomized designs, but if among-farm variance is large, losses of precision resulting from selecting on the basis of unadjusted cultivar means are likely to be great. To test this hypothesis, a simulation exercise was also conducted to examine the impact of yield variation among villages and among farms within villages on the relative effectiveness of alpha-lattice and RCB design analyses.

Methods

Variance component estimation in participatory varietal selection trials in rainfed rice. Participatory variety selection trials were conducted under farmer management in three eastern Indian districts in 1997-2000. Upland cultivar trials were conducted in three villages in southern Bihar (now Jharkand) in collaboration with the Central Upland Rice Research Station (CRURRS), Hazaribag. Lowland PVS trials were conducted in collaboration with Rajendra Agricultural University (RAU), Pusa, Bihar. In each set, several varieties were evaluated in unreplicated trials on three or four farms over at least two years. Details of the trials are presented in Table 1; however, they are more completely described by Courtois et al. (2001). Grain yield data were analyzed using the REML algorithm of SAS PROC VARCOMP with a cross-classified model, with cultivars, farms, and years as random factors. Broad-sense heritability or repeatability (H) was estimated as:

$$H = \sigma^2_G / \{\sigma^2_G + (\sigma^2_{GL}/l) + (\sigma^2_{GY}/y) + (\sigma^2_{GLY}/ly)\} \qquad (1)$$

where σ^2_G, σ^2_{GL}, σ^2_{GY}, and σ^2_{GLY} are the genotype, genotype x location, genotype x year, and genotype x location x year variance components, respectively, and l and y are the number of locations and

Table 1. Description of participatory varietal selection trials in eastern India.

Location	Cooperating institution†	Ecosystem	No. of years	No. of locations	No. of genotypes	Mean yield (t/ha)
Hazaribag	CRURRS	Upland	3	3	12	1.96
Pusa	RAU	Lowland	2	3	9	4.21

† CRUURS = Central Upland Rice Research Station; RUA = Rajendra Agricultural University.

the number of years, respectively. It should be noted that when estimated from unreplicated trials, the σ^2_{GLY} component also contains the within-trial plot error or residual variance.

Simulating the predictive power of mother-baby trials analyzed as randomized-complete-block versus alpha-lattice designs

A simulation was conducted using the following model:

$$P_{ijklm} = M + Y_i + V_j + YV_{ij} + F(YV)_{k(ij)} + G_l + GY_{li} + GV_{lj} + GYV_{lij} + e_{ijklm} \qquad (1)$$

where:

P_{ijklm} = the measurement on a plot containing genotype *l* on farm k in village j in year i

M = the overall mean of the trials

Y_i = the effect of year i

V_j = the effect of village j

VY_{ij} = the interaction between year i and village j

$F(VY)_{k(ij)}$= the interaction between year i and village j and farm k

G_l = the effect of genotype l

GY_{li} = the interaction between genotype l and year i

GV_{lj} = the interaction between genotype l and village j

GVY_{lij} = the interaction between genotype l, year i, and village j

e_{ijklm} = the within-village residual

The SAS program was used to simulate values for P, assuming all factors random in the model. An overall mean (M) of 2.2 t/ha was assumed. Effects were generated with the SAS RANNOR function, using the appropriate variance components as function arguments. Variance components used in the simulation were taken from the literature or from analyses of rice variety trial data available at IRRI. Three scenarios were identified regarding the relative magnitudes of the GYV and F(GYV) variances. In one scenario, there was little variation among farms within villages in mean yield, but considerable variation across villages. In another scenario, there was little variation among farms within villages, but substantial mean yield differences among villages. In the third, variance among villages and among farms within villages was approximately equal in magnitude. (It should be noted that other estimates might lead to different simulation results.) The variance components used in the simulation (listed below) are based on estimates derived from the combined analysis of the Philippine Upland Rice National Cultivar Trials for 1997-99:

σ^2_{Y} = 2700
σ^2_{V} = 5000
σ^2_{YV} = 800000 or 500000 or 200000
$\sigma^2_{F(YV)}$ = 200000 or 500000 or 200000
σ^2_{G} = 44600
σ^2_{GY} = 39000
σ^2_{GV} = 5000
σ^2_{GYV} = 300000
σ^2_{e} = 100000

s

Single-replicate PVS trials testing a set of 16 cultivars in 3, 5, or 10 villages were simulated, with 4 cultivars per block. Alpha-lattice designs generated by the Alphagen program were used. Cultivar means over villages and farms were calculated in three ways:

1. Raw means were calculated over all villages and farms.
2. Data were standardized within farms and then means were calculated over farms and villages.
3. Means adjusted for lattice incomplete block effects were calculated using the REML option of SAS PROC MIXED, with genotypes considered fixed and all other effects random.

The simulation for each of the 9 conditions (3 experiment sizes x 3 estimators of variety means) was replicated 10 times. For each run, the correlation between simulated genotypic values and simulated cultivar mean yields was calculated. This correlation, equivalent to the square root of the heritability of cultivar means, is an easily understood measure of the repeatability of cultivar trials, and is more directly related to their predictive power than is the variance of cultivar means.

Results and Discussion

Variance component estimation in participatory varietal selection trials in rainfed rice

Variance components are presented in Table 2. The relative magnitude of these components varied greatly from trial to trial. For the upland target environment (TE), site variance was the largest component, reflecting the large range in soil quality among sites. For the Pusa rainfed lowland TE, year to year variances were large. Cultivar effects were significant in two of the three TEs. Cultivar x year interactions were small for all three TEs. Cultivar x site interactions were also relatively small for all three TEs, indicating that cultivars responded similarly across sites within TEs. The residual error for the combined analysis, which contains both the cultivar x year x site and within-site residuals, was large in all cases, indicating that within-site soil heterogeneity and/or random variation in cultivar ranking among sites and years were the most important sources of noise in the trials.

Using the variance components in Table 2, repeatability estimates were calculated for means estimated from 1, 2, 5, or 10 trials for the 2 trial sets in which genotypic variation for grain yield was significant (Table 3). In both cases, means estimated from a single trial had very low repeatability. Replication over 5 sites increased predicted repeatability to more than 0.5 in both data sets.

In summary, these experiments indicate that specific adaptation to sites within the TEs served by the CRURRS and RAU breeding programs appears to be limited. Site to site and year to year variability among PVS trials was large,

Table 2. Variance component estimates from participatory varietal selection trials in eastern India.

Location	σ^2_Y	σ^2_L	σ^2_{YL}	σ^2_G	σ^2_{GY}	σ^2_{GL}	σ^2_{GLY}
Hazaribag	0.02	1.03	0.00	0.13	0.00	0.04	0.29
Pusa	1.36	0.13	0.08	0.20	0.15	0.01	0.20

but rank changes across sites were limited. Replication of trials over 3-5 sites or farms may be sufficient to achieve useful levels of repeatability in PVS trials.

Simulating the predictive power of mother-baby trials analyzed as randomized-complete-block versus alpha-lattice designs

The results of the simulation are presented in Table 4. For trials comprising 3 village replicates, the correlation between genotype value and cultivar means estimated from lattice-adjusted data ranged from 0.45 to 0.51. The correlation increased to approximately 0.6 when the number of village replicates increased from 3 to 5, but no increase was observed from increasing the number of villages from 5 to 10. If the variances used in this simulation are representative of rainfed rice trials in eastern India, mother-baby networks consisting of as few as 3-5 village replicates may be adequate for progress from PVS to be made. For all three ratios of $\sigma^2_{VY} : \sigma^2_{F(VY)}$ and all trial sizes, the correlation between genotypic value and the means estimated from trials was greater for lattice-adjusted means than for raw means. Standardization within farms did not consistently improve the relationship between phenotypic and genotypic value. The increase in selection response resulting from the use of lattice designs is expected to be approximately $r_{lattice}/r_{raw}$, where $r_{lattice}$ is the correlation between genotypic value and cultivar means estimated with lattice adjustment, and r_{raw} is the correlation for raw means. This ratio is roughly equal to the selection responses that can be expected from lattice adjustment, relative to the analysis of raw means. $r_{lattice}/r_{raw}$ was approximately 1.1-1.3 for all simulations, indicating that lattice adjustment may be advantageous even when the number of village replicates is quite large if there is considerable variation in the mean yields of farms.

Table 3. Predicted repeatability (H) of cultivar means estimated from 1, 2, 5, or 10 unreplicated on-farm trials conducted in a single season in eastern India.

	H			
Location	1 site	2 sites	5 sites	10 sites
Hazaribag	0.28	0.44	0.66	0.80
Pusa	0.36	0.44	0.51	0.54

Table 4. The effect of trial number, method of estimating means, and the ratio $\sigma^2_{VY}:\sigma^2_{F(VY)}$ on the correlation between genotypic and phenotypic values in simulated mother-baby trials, eastern India.

		Estimation method		
No. of trials	$\sigma^2_{VY} : \sigma^2_{F(VY)}$	Raw means r	Standardized means r	Lattice-adjusted means r
3	4:1	0.43	0.39	0.45
	1:1	0.37	0.48	0.48
	1:4	0.38	0.46	0.51
5	4:1	0.57	0.56	0.64
	1:1	0.63	0.69	0.72
	1:4	0.61	0.58	0.63
10	4:1	0.54	0.60	0.64
	1:1	0.59	0.64	0.67
	1:4	0.63	0.62	0.67

Conclusions

Participatory varietal selection trials produce repeatable estimates of rainfed rice cultivar means. In the experience of the authors, the repeatability of grain yield estimates from the farmer managed trials was not markedly lower than for on-station trials. It was also found that rainfed rice PVS trials conducted using the mother-baby model generate estimates of cultivar mean yields with useful precision from testing as few as five farms per cultivar. Random cultivar x site x year interaction was the most important source of genotype x environment interaction (GEI) in eastern Indian rainfed rice. There was no evidence of village-specific adaptation. This is consistent with on-station research on GEI in rainfed rice, which also indicates that cultivar x site x year variances are the largest GEI component. Cultivar x site interactions appear to be rare across sites at similar levels in the toposequence and within geographic regions of the scale served by the CRURRS and RAU breeding programs. The effect of the large cultivar x site x year component of the phenotypic variance can be reduced, and H concomitantly increased, by increasing the number of sites and years of testing. Because small rainfed rice breeding programs often cannot easily increase the number of sites they handle, they should consider replication over years to increase the precision of variety trials. If variance among farms within villages is large, simulation indicates that the alpha-lattice designs can significantly increase repeatability. Standardization within farms was not effective in increasing precision. Freely available, easy to use software for the generation and analysis of alpha-lattice designs is needed by researchers from national agricultural research programs if the mother-baby design is to be widely and effectively adopted.

References

Atlin, G.N., M. Cooper, and Å. Bjørnstad. 2001. A comparison of formal and participatory breeding approaches using selection theory. *Euphytica* 122:463-475.

Atlin, G.N., and K.J. Frey. 1989. Breeding crop varieties for low-input agriculture. *American Journal of Alternative Agriculture* 4:53-57.

Atlin, G.N., and K.J. Frey. 1990. Selecting oat lines for yield in low-productivity environments. *Crop Science* 30:556-561.

Courtois, B., B. Bartholome, D. Chaudhary, G. McLaren, C.H. Mishra, N.P. Mandal, S. Pandey, T. Paris, C. Piggin, K. Prasad, A.T. Roy, R.K. Sahu, V.N. Sahu, S. Sarkarung, S.K. Sharma, A. Singh, H.N. Singh, O.N. Singh, N.K. Singh, R.K. Singh, R.K. Singh, S. Singh, P.K. Sinha, B.V.S. Sisodia, and R.Thakur. 2001. Comparing farmers and breeders rankings in varietal selection for low-input environments: a case study of rainfed rice in eastern India. *Euphytica* 122(3): 537-550.

Falconer, D.S. 1989. *Introduction to quantitative genetics.* 3rd Ed. London: Longman.

Pederson, D.G., and A.J. Rathjen. 1981. Choosing trial sites to maximize selection response for grain yield in spring wheat. *Australian Journal of Agricultural Research* 32:411-424.

Discussion Summary

The discussion dealt with the analysis of incomplete designs and the feasibility of obtaining data and its accuracy in participatory varietal selection (PVS) trials. It was pointed out that there are several incomplete designs and the alpha-lattice design is not always needed. In response to this, the author said that trials are samples of environments; once sampling is effective and frequent enough, data adjustment becomes less important.

Obtaining yield data from baby trials is not trivial. It is easier to obtain yield data from mother trials; however, baby trials are particularly effective for identifying farmers' perceptions and assessments. Researchers in nongovernmental organizations are capable of obtaining appropriate data, but development workers generally are not. It is valuable to know the farm to farm variability for ranking of cultivars and variance components. In the eastern India case study, it was pointed out that special arrangements had to be made to enable station staff to attend the harvest of the baby trials. These arrangements included a car, a moisture meter, and other items. Additional people were hired as well. The logistical difficulties of obtaining yield data from baby trials may limit the ability of some programs to secure quality data.

Analyzing Ranking and Rating Data from Participatory On-Farm Trials

RICHARD COE

Abstract

Responses in participatory on-farm trials are often measured as ratings (scores on an ordered but arbitrary scale) or rankings (respondents are simply asked to order treatments). Usual analysis of variance and linear model based analyses are not appropriate for these data. Alternative analyses based on generalized linear models are described in this paper. These methods can be successfully used when trial designs are irregular—a common characteristic of participatory trials—and when covariates are measured on each plot or farm in order to identify genotype x environment interaction.

Introduction

Participatory methods have been widely adopted by researchers working on applied agricultural problems including crop breeding. This change in paradigm has implications for the design and analysis of methods used. Some of these are summarized in a companion paper by this author (see Coe 1, this proceedings). An assumption of this paper is that formal analysis of quantitative data collected systematically from trials is still an important part of the process. Without this it is difficult to see how the research activity can generate information relevant to anyone other than the small number of farmers directly involved. Breeders that have adopted participatory methods have generally recognized this, but have faced three difficulties:

1. The experimental designs used are often irregular in layout due to farmer participation (e.g., in choosing which varieties to test on their farms) or constraints arising from trials being located in farmers fields.
2. The focus of analysis often shifts from overall varietal selection to understanding the variation in varietal response across farms. This is genotype x environment interaction (GEI), where the E may include social or economic variables in addition to biophysical environments.
3. Much of the quantitative data collected may be ratings and rankings, for which the more usual methods of analysis may not be appropriate.

In this paper, I describe the analysis methods that simultaneously deal with each of these difficulties. I start by summarizing the nature of ranking and rating data and reviewing approaches to analysis. Examples to illustrate the analyses are presented in the third section. In the fourth and fifth sections, I present a detailed discussion of an approach to analysis of ranking and rating data, respectively. The discussion in the final section highlights some outstanding problems and implications of the methods presented.

Data and Analysis Types

The nature of the response variable is one factor which determines the type of analyses that can be conducted, whether formal or informal. It is therefore important to understand exactly how the data were collected and what the numbers represent.

Continuous

Quantities such as crop yield can be measured on a continuous scale, e.g., in kg/m^2. The numbers have the property that "2 really is the average of 1 and 3", making many common statistical procedures appropriate. Such quantities may be on a "ratio" or "interval" scale, the difference depending on whether the scale has a real zero. For example, a yield of 1 t/ha is 50% of a 2 t/ha yield, but a temperature of 10°C is not 50% of 20°C, as the zero for temperature is arbitrary.

Scores or rating

Here I refer to data that is recorded on a scale from poor to excellent, or less than enough to more than enough. The categories used are often given numerical labels, such as 1, 2, 3, 4, 5. These are called scores or ratings and such a scale is also described as ordered categorical. The labels are arbitrary. An observation of 3 is higher than an observation of 2, but we cannot say that it is better by the same amount that an observation of 5 is better than one of 4. An analysis would ideally order the data without using the actual numerical label to ensure that the results are the same regardless of which labels are used.

Binary

Data recorded with just two categories are common, e.g., yes or no, dead or alive, acceptable or not acceptable. Analysis is based on the frequency with which the categories occur. Analysis methods for these data are widely described in the literature but are not discussed further here.

Ranks

In many investigations of preference, data are collected by asking respondents to rank alternatives. The options available are placed in order without any attempt to describe how much one differs from another or whether any of the alternatives are, for example, good or acceptable. We might have variety A ranked above B, which is ranked above C, yet none of the three are considered good. The data would look the same in the case where a respondent placed them in the same order, but one, two, or all three were acceptable.

Other scales may be hybrids of these.

The steps in analyzing any data set can be summarized as:

1. Define the analysis objectives. These drive the rest of the analysis. It is impossible to conduct a good analysis of data without clear objectives. Often the key graphs and tables can be defined at this stage, even without the results with which to fill them in.

2. Prepare the data. Data sets will have to be entered and checked, suitable transformations made (e.g., to dry weight per unit area), relevant information from different sources (e.g., farm household data and plot level yields) extracted to the same file, and so on.
3. Exploratory and descriptive analysis. The aim is to summarize the main patterns and notice further patterns that may be relevant. This step is only covered briefly in this paper, as the methods used will depend on the context in which the analysis is carried out and the audience at which the results are aimed.
4. Formal statistical analysis. The aim is to add measures of precision and provide estimates from complex situations.
5. Interpretation and presentation.

Iteration between the steps will be necessary. Training materials by Coe et al (2001) provide more information on analysis of experiments. Some comments on the roles of these steps in analysis of participatory experiments are given in the companion paper (Coe 1, this proceedings).

A common objective in the analysis of many participatory breeding trials is to understand the nature of variation in responses from different farmers. Many researchers report that participatory on-farm trials give highly variable results, making interpretation difficult. Certainly if a standard analysis aimed at identifying differences in variety means is conducted, the result may well be a very high residual variation with correspondingly large standard error or variety differences, implying only vague knowledge about the relative performance of the entries. However, the variation can often be understood as genotype x environment interaction (GEI). The environment in which a participatory trial takes place is heterogeneous. Though a researcher may have full control over a trial, there will be many sources of hidden variation, including social or economic factors, as well as the more usual biophysical definitions of environment. For example, male and female farmers may assess varieties differently, or ratings may depend on the level of market integration of a farmer. The analyses must therefore be able to identify and describe these GEIs. When this occurs, the results are often the most useful output of the trial because they allow recommendations to be adjusted to particular local conditions

A spreadsheet package such as Excel is good for much of the descriptive analysis. Its flexible facilities for data selection and transformation, tabulation, and graphics are useful. However, dedicated statistical software is needed for the analyses described here—they cannot be performed in Excel. There are several packages with approximately equivalent facilities. All examples given in this paper use Genstat (2000)—I find it often the easiest to understand, particularly since methods for different problems can be addressed using a similar sets of commands. The key commands used to produce each analysis are included in the text with their output. SPSS is widely used by social scientists but is not particularly useful for the analyses described here.

Examples to Illustrate Analysis Methods

1. Soil fertility under agroforestry in Malawi

This is not a breeding trial but is included because the design is typical of many participatory on-farm trials. Three soil fertility strategies are compared over a number of years:

g	Mixed intercropping of maize and gliricidia
s	Relay planting of maize and sesbania
c	The control of continuous maize

Forty-one farmers each compared the control with one or both of the other treatments. Crop yield is the response of interest. A number of covariates were measured at the plot or farm level to help understand the reasons for variation across farms.

2. Maize varieties in Zimbabwe

This was a "baby" trial.[1] Twelve maize varieties were compared. A total of 146 farmers in 25 different sites took part, each farmer testing 4 of the 12 varieties. The varieties tested by each farmer were chosen by the researcher. Some household and field covariates were recorded. The actual crop yields obtained were not available for analysis, so the examples described here use simulated yield data but the original field design.

3. Maize varieties in Kenya

In this baby trial, 18 varieties of maize were compared, 2 of which were local controls. Twenty-nine farmers participated, each planting 2 replicates of all 18 entries. Crop performance was rated on a scale of 1, 2, 3, 4, 5. Gender of the respondent and farm size were also recorded.

Analyzing Ratings or Scores

Example 1

The crop yields in Example 1 were measured in tons per hectare; however, I have converted them to scores to illustrate the method of score analysis. The conversion is "exact" (i.e., the scores are those that farmers would give if asked to assess yield and are without error) so that for illustration purposes the results can be compared with those obtainable from actual yields. Scores were allocated as follows:

Yield	**Score**	**Label**
y < 1	1	Poor
1 <= y < 2	2	OK
2 <= y < 3	3	Good
3 <= y	4	Excellent

We could just tabulate frequencies as:

```
TABULATE [PRINT=nobs;
CLASSIFICATION=trt,score98; MARGINS=no]
score98
```

```
                     Nobservd
score98   poor         ok        good  excellent
 trt
  c         9          13          6          3
  g         5          15         10          9
  s         3           7          9          5
```

---

[1] The mother-baby trial design comprises a central researcher-managed "mother" trial, which tests all varieties, and farmer managed "baby" trials, which test a subset of the varieties from the mother trial.

This is informative. For example, the mode of the distribution for treatment g is deemed "OK". This shifts to "good" for treatment s. For treatment c the mode is also OK but the frequencies of other scores suggest that g is better than c.

This type of analysis has obvious drawbacks:

1. It is difficult to know how to handle more complex patterns, e.g., two modes.
2. It seems to ignore some of the structure in the data. For example, we have not used the fact that each farmer rates two or three treatments.
3. It is not obvious how the analysis could be extended to deal with more complex problems such as identifying and describing the effects of covariates to describe GEI.
4. It is not obvious how to formalize the analysis so that we can give measures of uncertainty, i.e., standard errors, confidence intervals, or statistical hypothesis tests.

A common approach is to treat the scores as quantities measured on a continuous scale. Means can then be calculated (see below) and all methods of analysis of variance, regression, and related modeling can be tried.

```
         Mean     Variance
trt
c        2.097    0.8903
g        2.590    0.9852
s        2.667    0.9275
```

There two reasons to feel uncomfortable about this. The first is that many of the assumptions of analysis of variance or linear regression modeling may well be inappropriate, given the limited range of the observations. A critical assumption is that the variance between observations of the same treatment is constant across treatments. This is commonly not the case, with the extreme entries showing less variation in score than those with a mean of 1 or 2.

The second reason is that the method makes some assumptions about the scores that may not be appropriate. For example, is the average of "poor" and "good" really "OK"? The seriousness of this objection is obvious when it is realized that the scores 1, 2, 3, and 4 are just labels, but the results critically depend on which label is given. If we used 0, 1, 5, 100, for example, the results would look very different, yet logically speaking these labels are equally acceptable.

There are situations in which both these objections are unimportant and so a useful analysis can be made. However, we would prefer an approach that is theoretically more sound and more robust, and applicable in a wider range of cases.

A second approach is to dichotomize the response, i.e., change it from a 4-level to a 2-level scale. For example, we could group poor and OK together and good and excellent together to give a measure with just two possible values. There are well established methods for analyzing such data, including models (e.g., logistic regression) that allow the effects of complex arrangements of covariates to be disentangled, and even methods (generalized linear mixed models) that allow random effects to be incorporated, such as the REML analysis of continuous data (Coe 1, this proceedings). However, this approach is also unsatisfactory. If the variable is originally measured on a 4-point scale and we reduce it to a 2-point scale, then we must be losing information.

Valid methods have been developed which use the all information without making unreasonable assumptions and can model the effect of covariates. In order to understand the model, we look first at the data for just two treatments, g and c, and forget that the observations are paired by farmer. The data are thus the following frequencies:

Treatment	Poor	OK	Good	Excellent
c	9	13	6	3
g	5	15	10	9

If we combine the top three categories, the data is reduced to a 2 x 2 table:

Treatment	Poor	OK + Good + Excellent
c	9	22
g	5	34

From these tables it looks as though g is better than c. A higher proportion of the plots are in the OK + good + excellent category. A common measure of this association is the odds ratio, O, or log odds ratio, log(O) = L, which can be calculated as follows:

$$O = \frac{\text{odds on g high}}{\text{odds on c high}} = \frac{34/5}{22/9} = 2.78$$

L = log(2.78) = 1.02

If we “cut” the categories differently, combining poor and OK, the following data are produced:

Treatment	Poor + OK	Good + Excellent
c	22	9
g	20	19

From this table O = 2.32, and L = 0.84.

A third cut is possible, combining poor, OK, and good to give:

Treatment	Poor + OK + Good	Excellent
c	28	3
g	30	9

From this table, O = 2.80 and L = 1.02.

In this case the O values are similar for each cut. If we make the assumption of such proportional odds with a constant value of O, then its value and standard error can be estimated without choosing any particular cut. In Genstat the calculations are made using the regression modeling commands. Note that the data have to be arranged so that there is a response variable for each possible response category. The variable for each score contains the number of plots that had that score.

```
print treat,s1,s2,s3,s4

     treat     s1      s2      s3      s4
         c  9.000   13.00   6.000   3.000
         g  5.000   15.00  10.000   9.000
```

```
model[dist=multinomial;yrel=cumulative;link=logit]
s1,s2,s3,s4 fit [p=e,a] treat
```

```
***** Regression Analysis *****

*** Estimates of parameters ***

                                         antilog of
             estimate  s.e.   t(*)  estimate
Cut-point 0/1  -0.927   0.367  -2.53  0.3956
Cut-point 1/2   0.948   0.367   2.58   2.581
Cut-point 2/3   2.161   0.438   4.93   8.680
treat g         0.932   0.452   2.06   2.539

* MESSAGE: s.e.s are based on dispersion
parameter with value 1

Parameters for factors are differences
compared with the reference level:
      Factor  Reference level
       treat         c

*** Accumulated analysis of deviance ***

Change                       mean    deviance
         d.f.   deviance  deviance    ratio
+ treat    1    4.37545   4.37545     4.38
Residual   2    0.16035   0.08018

Total      3    4.53580   1.51193

* MESSAGE: ratios are based on dispersion
parameter with value 1
```

The analysis of deviance is interpreted similarly to an analysis of variance, comparing the deviance with a chi-squared distribution to judge the importance of the effect. In this case there seems to be a significant treatment difference.

The parameter estimate `treat g` measures the difference between treatments g and c. The estimate 0.932 is the log odds ratio = log(odds of g being high vs. low/odds of c being high vs. low). Here high and low refer to being above and below a cut point, respectively, in the ordered set of scores (it doesn't matter which cut point, as the model constrains this odds ratio to be the same for any choice of cut point).

The value of 0.932 for the log odds ratio means that the odds ratio is exp(0.932) = 2.539. This is similar to the average of the three odds ratios found directly from the data. The standard error can be used to test the hypothesis of no difference between g and c (log odds ratio of 0) or to give a confidence interval for the log odds ratio or odds ratio. The cut-point parameters listed by Genstat do not have a useful interpretation in this case.

Now we analyze the whole data set using the same ideas and include a term for `farm` to account for the fact that each farmer is evaluating two or three plots. There is a row of data for each plot and a column for each possible score (poor, ok, good, excellent, or 1, 2, 3, 4), which are now assigned the names `s98[1]`, `s98[2]`, `s98[3]`, and `s98[4]`. The data value is again the number of plots that were given that score, but now this value is simply 0 or 1, with a single 1 given in each row. A small part of the data is shown:

```
print name,trt,s98[1...4],score98;10;decimals=0

name     trt s98[1] s98[2] s98[3] s98[4] score98
Chakame  g     0      0      0      1  excellent
Chakame  s     0      0      1      0     good
Chakame  c     1      0      0      0     poor
Thobola  g     0      1      0      0      ok
Thobola  s     1      0      0      0     poor
Thobola  c     0      1      0      0      ok
Adisani  g     0      1      0      0      ok
Adisani  c     1      0      0      0     poor
Majoni   g     0      0      0      1  excellent
Majoni   s     0      0      0      1  excellent
.        .     .
.

model
[dist=multinomial;yrel=cumulative;link=logit]
s98[1...4]
fit [p=*] name
add [p=a] trt

* MESSAGE: Term name cannot be fully
included in the model
because 2 parameters are aliased with
terms already in the model

(name Komwa(died 97)) = 0

(name Lipenga(died 98)) = 0
```

```
***** Regression Analysis *****

*** Estimates of parameters ***

                                        antilog of
              estimate  s.e.   t(*)   estimate
Cut-point 0/1   1.90    2.00   0.95     6.716
Cut-point 1/2   7.23    2.27   3.18     1382.
Cut-point 2/3  10.72    2.43   4.41    45305.
name Belo       4.04    2.56   1.58     56.75
name Bisiwiki   0.00    2.76   0.00     1.000
name Chakame    5.63    2.51   2.24     278.9
name Chimimba   0.97    3.57   0.27     2.638
.
.
name White      5.56    2.51   2.21     259.7
trt g          3.598   0.770   4.67     36.51
trt            2.722   0.786   3.47     15.21

* MESSAGE: s.e.s are based on dispersion
parameter with value 1

Parameters for factors are differences
compared with the reference level:
         Factor     Reference level
         name       Adisani
         trt        c


*** Accumulated analysis of deviance ***

Change                     mean       deviance
         d.f.  deviance   deviance   ratio
+ name   38    115.468    3.039      3.04
+ trt    2     30.218     15.109     15.11
Residual 51    105.977    2.078

Total    91    251.663    2.766

* MESSAGE: ratios are based on dispersion
parameter with value 1
```

The analysis of deviance is interpreted in the usual way using a chi-squared distribution to assess the size of contributions. A deviance of 30.2 with 2 df confirms that treatment is having an obvious effect on the ratings.

The parameter estimates for each farmer are uninteresting since they only reflect the fact that farmers can differ in the mean rating given, whereas the estimates for the treatments are important because they give the quantitative summary of the ratings. In this example the control treatment c is the baseline from which the others are measured. The important results are shown in the table below. For comparison, analysis of the actual yields using a similar method (linear model fitting farmer + treatment effects) is also shown. (For details see Coe 1, this proceedings.) Remember the scales are different. We cannot hope to recover information on actual yield per hectare from data that has simply been recorded as poor, OK, etc. What are important are the differences and similarities between treatments, which are revealed by this analysis.

	Rating		Yields		Scaled yields†	
Treatment	**Log odds ratio**	**se‡**	**Adjusted mean**	**se**	**Adjusted mean**	**se**
g	3.60	0.77	2.62	0.15	3.60	0.53
s	2.72	0.79	2.37	0.17	2.68	0.61
c	0.00	-	1.64	-	0.00	-

† Yield means scaled to match the log odds ratio scale.
‡ se is the standard error of the difference from c.

When the scales are aligned, the results of the analyses are remarkably similar. The se values for the rating data are higher because ratings contain less information than actual yields.

The value of the analysis becomes clear when we start to look at differences between groups of farmers or try to understand the effect of covariates. For example, slope2 is a factor which classifies farms into flat or sloping categories. The variate cec is related to soil fertility. It was hypothesized that g would perform relatively better on flat land and that both g and s would be superior to c when cec is low. These are investigated in the following table:

```
*** Accumulated analysis of deviance ***

Change                          mean   deviance
            d.f.  deviance  deviance   ratio
+ cec         1     1.945     1.945     1.94
+ slope2      1     8.959     8.959     8.96
+ trt         2     6.259     3.129     3.13
+ name       37   133.823     3.617     3.62
+ cec.trt     2     2.087     1.043     1.04
+ slope2.trt  2     2.543     1.271     1.27
Residual     44    90.606     2.059
```

There is no clear evidence for either slope or cec showing an interaction with treatment.

Example 3

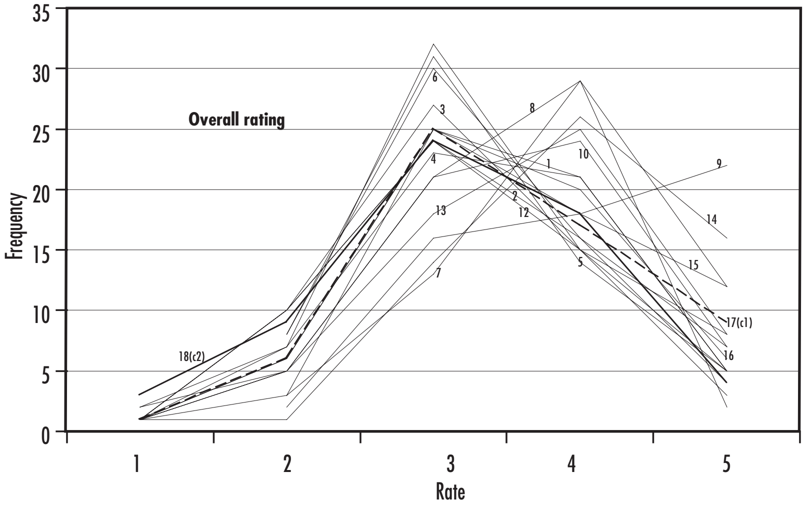
In this example, the performance of 18 varieties was rated on a scale of 1 (poor) to 5 (excellent). Criteria were yield, cob size, cob filling, and overall assessment. The design was straightforward: 29 farmers each evaluated the 18 varieties from 2 plots of each. Simple descriptive statistics can therefore give a useful summary of some of the characteristics. For example, the graph below only shows the frequency of responses for the overall rating of each of the 18 varieties. Varieties 17 and 18 have been highlighted because they are local checks.

The varieties seem to fall into 2 main groups—those with a mode at 3 and those with a mode at 4—with entry 9 standing out by being rated higher than all the others.

There a number of reasons why the modeling analysis is still worthwhile:

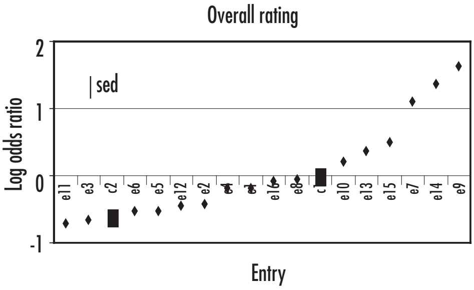
1. It provides simple concise summaries with measures of precision.
2. It makes the inclusion of covariates straightforward. In this case both farm size and gender of the respondent have been recorded.
3. It simplifies the comparison of the ratings under different criteria.

The analysis follows a similar pattern to the previous example. Note that the layout with two replicates per farm can be explicitly included in the analysis if sensible. Here I have assumed the two replicates correspond to two blocks on each farm. Farms are distinguished by the factor IDNO and blocks within farmers by REP.



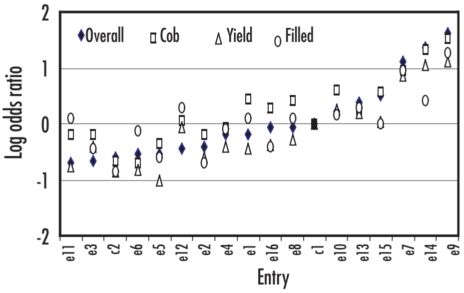
```
model
[dist=multinomial;yrel=cumulative;link=logit]
overall[]
fit [p=*]
add [p=*] IDNO
add [p=*] IDNO.REP
add [p=a,e] ENTRY

*** Accumulated analysis of deviance ***

Change                        mean  deviance
           d.f.  deviance deviance  ratio
+ IDNO      28   342.309   12.225   12.23
+ IDNO.REP  29    82.354    2.840    2.84
+ ENTRY     17   123.623    7.272    7.27
Residual   966  2189.420   2.266

Total    1040  2737.706   2.632

*** Estimates of parameters ***

                                    antilog of
            estimate  s.e    t(*)  estimate
Cut-point 0/1  -7.333  0.609  -12.05 0.0006534
Cut-point 1/2  -4.558  0.541   -8.43  0.01048
Cut-point 2/3  -1.560  0.522   -2.99   0.2102
Cut-point 3/4   0.848  0.520    1.63    2.335
IDNO 2          0.170  0.645    0.26    1.185
.
.
.
IDNO 29        -1.993  0.649   -3.07   0.1363
IDNO 1 .REP 2  -1.365  0.641   -2.13   0.2554
.
.
.
IDNO 29 .REP 2 -0.319  0.652   -0.49   0.7271
ENTRY c2       -0.610  0.361   -1.69   0.5433
ENTRY e1       -0.182  0.359   -0.51   0.8338
ENTRY e2       -0.419  0.360   -1.16   0.6575
ENTRY e3       -0.653  0.361   -1.81   0.5206
ENTRY e4       -0.196  0.359   -0.55   0.8219
ENTRY e5       -0.530  0.361   -1.47   0.5883
ENTRY e6       -0.539  0.361   -1.49   0.5834
ENTRY e7        1.109  0.360    3.08    3.030
ENTRY e8       -0.049  0.359   -0.14   0.9523
ENTRY e9        1.625  0.365    4.45    5.078
ENTRY e10       0.223  0.358    0.62    1.250
ENTRY e11      -0.701  0.361   -1.94   0.4963
ENTRY e12      -0.438  0.360   -1.22   0.6453
ENTRY e13       0.377  0.358    1.05    1.458
ENTRY e14       1.380  0.362    3.81    3.974
ENTRY e15       0.510  0.358    1.43    1.666
ENTRY e16      -0.078  0.359   -0.22   0.9248

* MESSAGE: s.e.s are based on dispersion
parameter with value 1

Parameters for factors are differences
compared with the reference level:
        Factor       Reference level
        IDNO         1
        ENTRY        c1
```

The analysis of deviance suggests that there are large differences between the entries; the parameter estimates summarize these. Remember the estimates are log odds ratios that describe the chance of being in a high response category rather than a low one for each entry compared with the baseline. I set up the data files so that the baseline is the first local check, c1. A simple graph reveals the patterns:



Apart from entries 9, 14, and 7, there is a continuous spread of ratings of the varieties rather than any clear groupings, with one of the local checks occurring towards the lower end of the spread and the other towards the upper end.

The ratings on each criterion can now be compared by repeating the analysis and putting the log odds ratio for each on the same graph. The pattern is much the same for each criterion. Of the three best performing entries, 14 does least well for cob filling.



There are two covariates of interest recorded: the gender of the respondent and farm size. The question of interest is whether males and females rate the entries differently, and whether the relative ratings depend on farm size.

```
model
[dist=multinomial;yrel=cumulative;link=logit]
overall[]
fit [p=*]
add [p=*] IDNO
add [p=*] IDNO.REP
add [p=*] ENTRY
add [p=*] SEX.ENTRY
add [p=a] SIZE.ENTRY

*** Accumulated analysis of deviance ***

Change                              mean   deviance
             d.f.  deviance deviance    ratio
+ IDNO         28   342.309   12.225    12.23
+ IDNO.REP     29    82.354    2.840     2.84
+ ENTRY        17   123.623    7.272     7.27
+ ENTRY.SEX    17    21.404    1.259     1.26
+ SIZE.ENTRY   18    21.701    1.206     1.21
Residual      931  2146.315    2.305

Total      1040  2737.706   2.632

MESSAGE: ratios are based on dispersion
parameter with value 1
```

Neither of the covariates show any interaction with entry; thus, we can conclude that the overall rating of varieties is much the same for males and females and does not have any linear relationship with farm size. The effect of farm size could perhaps be investigated further, for example, by putting farms into a few (three or four) size categories. This approach removes the assumption of a linear relationship between farm size and the log odds ratios.

Analyzing Ranks

At first glance, data from ranks look similar to rating data. Like ratings, the observations are integers from a limited range, and we want to find out the same sort of information, i.e., are there consistencies in the rankings given to different treatments that can allow us to reach conclusions about which treatments consistently ranked high? However, there are some important differences from rating data that will emerge.

Example 1

To illustrate the method I have once again converted yield data from Example 1 to ranks. Each farmer compared two or three treatments. Ranks have been allocated exactly, so that the treatment with the lowest yield on each farm is given rank 1, the next lowest rank 2, and the third (if there is one) rank 3. There were no ties. A small part of the data is shown.

```
    name       yield98     rank     trt
  Adisani       1.449      2.000     g
  Adisani       0.801      1.000     c
   Belo           *          *       c
   Belo         2.071      2.000     s
   Belo         1.246      1.000     g
 Bisiwiki       0.643      1.000     c
 Bisiwiki       1.514      2.000     g
  Chakame       0.761      1.000     c
  Chakame       3.380      3.000     g
  Chakame       2.142      2.000     s
 Chimimba       1.943      1.000     g
 Chimimba         *          *       s
 Chimimba         *          *       c
 Chinzeka       2.356      3.000     g
 Chinzeka       1.477      1.000     s
 Chinzeka       1.713      2.000     c
```

Simple displays of the data can be designed. For example, we can tabulate the number of farmers who rank each treatment as 1, 2, or 3.

```
TABULATE [PRINT=counts;
CLASSIFICATION=trt,rank; MARGINS=no] yield98
```

```
                          Count
       Rank     1.000     2.000     3.000
        trt
          c        24         6         1
          g         9        16        14
          s         6        12         6

Unknown count      14
```

Treatment g is ranked 3 more often than s, which is an indication that it is superior. But a difficulty is immediately clear: g is also ranked 2 and 1 more often than s. The problem arises from the fact that each farmer is only ranking the treatments he/she tests, and these are not the same for each. In the table above, when g = 2 we cannot tell whether g was the best out of 2 treatments or second best out of 3. Changing the ranking method to 1 = best does not help. Some authors suggest converting ranks to scores, but of course the problem cannot be fixed by any conversion that simply changes the ranks 1, 2, 3 to another set of numbers.

A more realistic summary comes from studying each treatment pair. If we take, for example, g and c, we can look at all of the farmers that compared these two, and calculate the proportion that ranked g higher than c.

Pair	No. of comparisons	No. with first of pair ranked higher than second	Proportion with first of pair ranked higher than second
g-c	31	28	0.903
s-c	21	16	0.762
g-s	24	16	0.667

This summary now correctly only relies on the rankings within each farm and is explicit about what is compared with what. Its shortcomings and the reasons for wanting a formal analysis are much the same as for the rating data. We need to put measures of precision on results and would like to extend the analysis to look at the effect of covariates or groupings of respondents. The analysis also seems unsatisfactory when we think of Example 2 with its 12 treatments and, hence, 66 pairs of treatments. However, a table like the one above but with 66 rows to describe performance of 12 varieties would be nothing but opaque!

The modeling approach to this type of data is based on the above table. The idea is to find a score s_i for each treatment such that the probability that treatment i is ranked higher than treatment j, when the two are compared, depends on the difference between the scores s_i-s_j. If the relationship between scores and probability is a logistic function, then the model can be fitted using standard logistic regression software. Hence:

p_{ij} = Prob(i ranked above j) and

$\log(p_{ij}/1-p_{ij}) = s_i-s_j$

Setting up the data to fit the model is slightly messy. There has to be a row for each pair of treatments compared. Thus, a farmer with just g and c will contribute 1 row of data for the pair g-c. A farmer with three treatments, g, s, and c, will contribute three rows of data, g-c, s-c, and g-s. Indicator variables are needed for each treatment, and the response variable contains 0s and 1s. The first few rows of data are shown:

```
Namel    firstl  secondl  c    g    s  compl
Adisini    g       c     -1    1    0    1
Belo       s       c     -1    0    1    *
Belo       s       g      0   -1    1    1
Belo       g       c     -1    1    0    *
Bisiwiki   g       c     -1    1    0    1
Chakame    g       c     -1    1    0    1
Chakame    s       c     -1    0    1    1
Chakame    s       g      0   -1    1    0
Chiminbo   g       c     -1    1    0    *
Chiminbo   s       g      0   -1    1    *
Chiminbo   s       c     -1    0    1    *
Chinzeka   g       c     -1    1    0    1
Chinzeka   s       g      0   -1    1    0
Chinzeka   s       c     -1    1    0
```

The first row of data shows that Adisini compared g and c. Treatment g was ranked higher than c, so when g is the first and c is the second, the response is "success", indicated by a 1 in the last column. Belo compared all three treatments but the observation for c was missing, therefore both the s-c and g-c comparisons are also missing.

The modeling now proceeds similarly to that in other situations.

```
model [dist=b] compl; nbin=1
fit [con=o]g+s+c

*** Summary of analysis ***
                                          mean
                                      deviance
          d.f.  deviance  deviance   ratio
Regression  2       *         *
Residual   74    73.49    0.9931
Total      76       *         *

* MESSAGE: ratios are based on dispersion
parameter with value 1

*** Estimates of parameters ***

                                       antilog of
    estimate     s.e.     t(*)       estimate
g     2.072      0.435    4.76         7.939
s     1.290      0.425    3.04         3.632

* MESSAGE: s.e.s are based on dispersion
parameter with value 1
```

The output looks a little odd because Genstat does not know what to use as a null model when the constant is omitted. It cannot calculate a total deviance and hence also cannot calculate a regression deviance. In this case the sensible null model is one of no preference between treatments, corresponding to $p_{ij} = 0.5$ for all pairs, or $\log(p_{ij}/1-p_{ij}) = 0$. The deviance for this model is given by:

```
model [dist=b] compl; nbin=1
fit [con=o]
```

Now the analysis of deviance can be reconstructed:

```
                                             mean
                                         deviance
          d.f.  deviance   deviance     ratio
Regression  2     31.91      15.96      15.96
Residual   74     73.49      0.9931
Total      76     105.4      1.386
```

The model appears to explain much of the variation, suggesting real difference between the treatments. When interpreting the parameter estimates, remember that the p_{ij} value depends only on the differences s_i-s_j. Hence we only need to estimate two of the treatments and can arbitrarily set the third, in this case c, to zero. The estimates above therefore give an ordering and even magnitude of differences between the treatments. They can be compared with the results from analyzing both actual yields and the scores.

	Ranking		**Rating**		**Yields**		**Scaled yields**[†]	
Treat-ment	**s_i**	**se**[‡]	**Log odds ratio**	**se**	**Adjusted mean**	**se**	**Adjusted mean**	**se**
g	2.07	0.44	3.60	0.77	2.62	0.15	2.07	0.32
s	1.29	0.43	2.72	0.79	2.37	0.17	1.54	0.36
c	0.00	-	0.00	-	1.64	-	0.00	-

[†] Yield means are scaled to match the s scale of the ranking data.
[‡] se is the standard error of the difference from c.

The analysis of ranks has, to within the arbitrary scaling, produced an order and relative difference between treatments which is remarkably similar to that from the actual yield data, yet has larger sed values: the ranks contain less information than actual yields.

Note that the table of pairwise probabilities p_{ij} can be reconstructed from the scores s_i using the following relationship:

$$p_{ij} = \exp(s_i - s_j) / (1 + \exp(s_i - s_j))$$

These are shown in the table below and indicate a reasonable fit of the model.

Pair	No. of comparisons	No. with first of pair ranked higher than second	Proportion with first pair of ranked higher than second	Fitted probabilities p_{ij}
g-c	31	28	0.903	0.888
s-c	21	16	0.762	0.784
g-s	24	16	0.667	0.686

As in other situations, an advantage of using an explicit model to analyze the ranks, rather than relying on more ad hoc methods, is that the effects of covariates can be identified. To illustrate this I have looked at slope, classified into 2 levels (0 = flat, 1 = sloping), since one of the hypotheses was that g would perform relatively less well on sloping land.

```
add [p=a,e] slopel.(g+s+c)
```

```
*** Accumulated analysis of deviance ***

Change                           mean   deviance
          d.f.    deviance deviance   ratio
- Constant
+ g
+ s
+ c         1         *
+ g.slopel
+ s.slopel
+ c.slopel  2      0.778     0.389     0.39
Residual   72     72.715     1.010

Total      75         *

* MESSAGE: ratios are based on dispersion
parameter with value 1
```

The analysis of deviance suggests that there is no consistent difference in the way g, s, and c are ranked on flat and sloping land. This conclusion is also reflected in the parameter estimates.

```
*** Estimates of parameters ***

                                      antilog of
           estimate   s.e.   t(*)   estimate
g            2.117   0.583   3.63    8.305
s            1.598   0.607   2.63    4.944
g.slopel 0  -0.056   0.901  -0.06   0.9454
g.slopel 1   0.000     *      *      1.000
s.slopel 0  -0.632   0.858  -0.74   0.5313
s.slopel 1   0.000     *      *      1.000

* MESSAGE: s.e.s are based on dispersion
parameter with value 1
```

These values can be combined in a table of scores, together with standard errors of the difference between treatments within slope categories.

Treatment	Slope = 0	Slope = 1
g	2.117 - 0.056 = 2.061	2.117
s	1.598 - 0.632 = 0.966	1.598
c	0	0.000

If the standard errors of the interaction effects were smaller, we would say the results were consistent with the hypothesis—that the difference between g and s is greater on flat than on sloping land.

Remember it is impossible to look at the main effect of slope. We cannot determine whether the treatments are generally assessed as better on flat than sloping land. Each participant ranks among the alternatives tested on their farm, and each farm is classed as either sloping or flat. Similarly we cannot compare the two columns in the table above, i.e., compare g on flat and sloping land. There is no information in the data on this comparison, as all rankings are done within farms. The situation would be different if there were farms that had both flat and sloping land.

Example 2

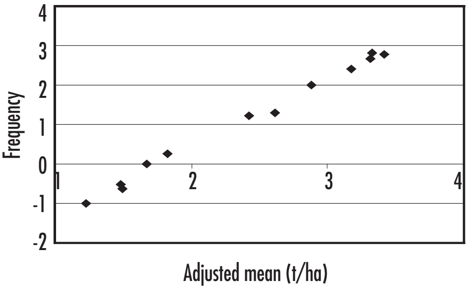
Yields for Example 2 were also converted to ranks for the purpose of illustrating the analysis. Remember this study comprises 12 varieties, with 146 farmers each comparing 4 varieties. It is difficult to think of a useful, simple, descriptive analysis of this rank data that shows the differences between varieties. The design is very unbalanced, so any simple totaling of ranks will give a biased picture. We could look at all 12 x 11/2 = 66 pairwise comparisons, and find the proportion in which one treatment ranks above another; however, it is not easy to view a matrix of 66 values and understand the relative performance of 12 varieties. It has been suggested by Russell (1997) that an overall score be given to each variety by counting the number of times each one ranks above another; however, this requires that each occurs equally often. Some sort of average proportion could be devised; however, the modeling approach is simple once the data file is set up.

The data file structure and modeling proceed as per Example 1. Twelve indicator variables, `e[1]`,...`e[12]`, are needed for the 12 varieties. In the statements below the first `FIT` gives the correct total deviance from which the analysis of deviance table is constructed.

```
model [dist=b] compl; nbin=1
fit [con=o]
fit [con=o] e[1...12]
```

```
                                 mean  deviance
           d.f.  deviance  deviance  ratio
Regression  11     439.1    39.9200  39.92
Residual   865     774.9     0.8959
Total      876    1214.0     1.386

*** Estimates of parameters ***

                                     antilog of
        estimate    s.e      t(*)    estimate
e[1]     -0.996    0.293    -3.40     0.3692
e[2]      1.998    0.300     6.66     7.378
e[3]      1.282    0.284     4.52     3.606
e[4]      2.818    0.317     8.88     16.74
e[5]     -0.523    0.303    -1.73     0.5926
e[6]      2.655    0.312     8.51     14.22
e[7]      0.247    0.277     0.89     1.281
e[8]      1.205    0.281     4.29     3.336
e[9]     -0.637    0.299    -2.13     0.5289
e[10]     2.769    0.318     8.70     15.95
e[11]     2.423    0.303     8.00     11.28

* MESSAGE: s.e.s are based on dispersion
parameter with value 1
```

Genstat has put the score for the last treatment, s_{12}, to zero. The parameter estimates above give scores that show the relative performance for each variety. If these are compared with the results based on actual yields, it can be seen (in the graph below) that the method not only reproduces the ordering of the varieties very closely, but also the relative differences. Of course in this case the ranks were calculated from the yields without error. However, it still seems very surprising that this information about relative performance of the varieties can be recovered from just the four ranks on each farm.



As each score is relative to the score of zero for variety 12, the se values listed with the estimates are for the comparison of that variety with variety 12. Other se values are most easily found using `predict`. For example, the difference between scores for variety 1 and 2 is found by:

```
predict [back=n] e[1...11]; 1,-1,0,0,0,0,
0,0,0,0,0

Prediction  S.e.
-2.995      0.335
```

More complex contrasts between treatments can be similarly calculated. For example, varieties 1, 5, 7, 9, and 12 are in one group, a. Varieties 4, 10, and 11 form group b. We can calculate the difference between the average scores for groups a and b by taking $(s_1+s_5+s_7+s_9+s_{12})/5 - (s_4+s_{10}+s_{11})/3$. Remembering that $s_{12} = 0$, `predict` can be used for this:

```
predict [back=n] e[1...11]; 0.2,0,0,-
0.3333,0.2,0,0.2,0,0.2,-0.3333,-0.3333

Prediction  S.e.
-3.052      0.213
```

Group a is clearly worse than group b.

As in Example 1, it is simple to turn differences in scores into probabilities of one variety being ranked higher than another. For example, the chance that variety 1 is ranked higher than 2 is given by:

```
predict e[1...11]; 1,-1,0,0,0,0,0,0,0,0,0

Prediction  S.e.
0.0477      0.0152
```

Variety 2 is almost certain to be ranked higher than variety 1.

As before, the model can now be extended to look at the extent to which covariates interact with treatment differences. I use two continuous covariates: soil P and sand content. The data file has been set up with a column giving the sand and P value for each pairwise comparison.

```
fit [p=*; con=o] e[1...11]
add [p=*; con=o] sandfl.e[1...11]
add [p=a; con=o] Pfl.e[1...11]
```

The analysis of deviance table can be constructed from this output. Note the total degrees of freedom has changed from that given earlier as there are missing values in the covariates.

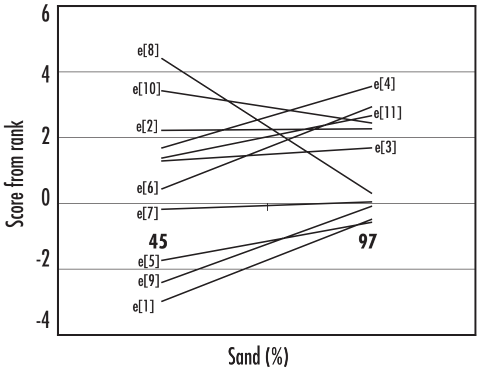
```
                                        mean
                   d.f.   deviance  deviance
e[1...11]           11      309.50    28.140
+sandfl.e[1..11]    11       15.64     1.420
+Pfl.e[1...11]      11       15.16     1.380
Residual           561      483.30    0.8614
Total              594      823.50     1.386
```

The results show that neither P nor sand have a strong interaction with variety. However in order to show the types of results obtainable, the model with sand is refitted and parameter estimates produced.

```
fit [con=o;p=e] e[1...11]+sandfl.e[1...11]
```

```
*** Estimates of parameters ***

                                          antilog of
             estimate   s.e.    t(*)   estimate
e[1]          -5.11     2.83   -1.81   0.006012
e[2]           2.22     3.21    0.69     9.223
e[3]           0.97     2.49    0.39     2.631
e[4]           0.10     3.20    0.03     1.109
e[5]          -2.73     2.61   -1.04    0.06530
e[6]          -1.72     2.75   -0.62     0.1796
e[7]          -0.42     2.38   -0.18     0.6559
e[8]           7.90     3.70    2.14  2699.
e[9]          -4.38     2.64   -1.66    0.01248
e[10]          4.28     3.46    1.24    72.58
e[11]          0.28     4.01    0.07     1.317
e[1].sandfl    0.0475   0.0327  1.45     1.049
e[2].sandfl    0.0004   0.0370  0.01     1.000
e[3].sandfl    0.0075   0.0290  0.26     1.008
e[4].sandfl    0.0357   0.0375  0.95     1.036
e[5].sandfl    0.0223   0.0301  0.74     1.023
e[6].sandfl    0.0478   0.0320  1.49     1.049
e[7].sandfl    0.0049   0.0280  0.17     1.005
e[8].sandfl   -0.0781   0.0417 -1.87     0.9249
e[9].sandfl    0.0444   0.0304  1.46     1.045
e[10].sandfl -0.0187   0.0398 -0.47     0.9814
e[11].sandfl  0.0244   0.0460  0.53     1.025
```

The scores for each variety now depend on the sand content, for example, the score for variety 1 is $s_1 = -5.11 + 0.0475$ sand. These are plotted below for the range of sand contents (45-97%) found in the trial.



Remembering that the scores show the relative performance of varieties, with variety 12 fixed at a score of zero, two main patterns emerge. Several varieties (1, 4, 5, 6, and 9) with increasing sand content rank higher than variety 12. Variety 8 with high sand content ranks distinctly worse.

Discussion

General

The methods for analyzing ranking and rating data described above are not new, but they have not been routinely used in the analysis of agricultural trials. A discussion of the proportional odds model used for rating data can be found in Agresti (1996). The model for ranks is not so widely used, explaining why common statistical software does not make it immediately available. When the observations are just paired comparisons (i.e., each participant is asked to state which of two treatments is superior), the model used is the Bradley-Terry model (Bradley and Terry 1952), which has been widely used, particularly in social science applications. Dittrich et al. (1998) use the method for paired comparisons when there are categorical covariates, and mention that it is possible with continuous covariates. The approach used when more than two treatments are compared is described by Critchlow and Fligner (1991).

Both models make assumptions about the nature of the data; however, this is true of all statistical analyses. It is a necessary part of attempting to reach conclusions about general patterns. Methods for checking the key assumptions are well developed for established linear model methods (e.g., looking for various patterns in residuals), and similar tools need developing for the models discussed. Alternative models may be more appropriate for ranks and rates. The methods presented here appear to be the simplest of those that have proved useful in some common situations. Again this is true of all statistical modeling. For example, linear regression analysis is widely used but not because "nature has to be like that" but because the model has been found to be a useful approximation for many problems.

From the examples in this paper it should be clear that statistical analysis of participatory breeding trials cannot be automatic. When researcher designed trials were run using a very regular design, it was possible (though probably not wise) to run a standard analysis on each data set. Such an approach will not recover most of the useful information from participatory trials.

Further discussion on analysis of ranks

The method described above for analyzing data presented in the form of ranks is appealing and powerful. It is able to produce an overall ordering of treatments and even indicate the relative magnitudes of the differences between treatments. It can handle awkward incomplete sets of data in which each farmer does not rank all treatments. Most importantly it can show how treatment order interacts with covariates. In Example 1, the covariate was a categorical variable, dividing the sample of farmers into groups. Continuous covariates were analyzed in Example 2.

Unlike many other approaches to analysis of ranks, the method uses estimation, not just testing. This distinction is often made when analyzing continuous variates such as crop yield. It is rarely beneficial to simply conclude that mean yields differ significantly between varieties, or even that variety A yields significantly higher than variety B. Useful conclusions can be drawn when we are able to identify by how much A outyields B, and put a confidence interval around this. The same is true when analyzing ranks. It is rarely going to be useful to simply report that treatments differ significantly in their ranks, yet this is all that most statistical procedures for analysis of ranks do. The method presented here shows the relative magnitude of the differences, and these can be interpreted. For example, we may show that A and B are ranked significantly differently. The scores for the varieties can be converted to a probability p_{AB} that A will be ranked higher than B. If $p_{AB} = 0.95$, the interpretation is very different than if $p_{AB} = 0.55$, yet both could be significantly different from the no-preference value of $p_{AB} = 0.5$.

There are a number of questions regarding this analysis, some of which require theoretical statistical investigation.

1. The model makes assumptions about the nature of the data and the effects of treatments and covariates. It is not clear how to check whether they are reasonable or how robust the results are to departures from the assumptions.

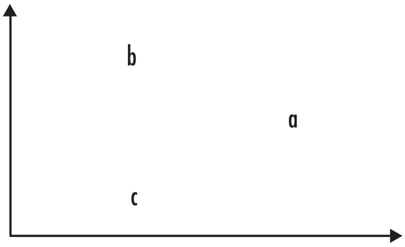
2. The analysis depends on the model which assumes the treatments can be allocated scores such that the probability of one ranking higher than another depends on the difference between the scores. This is the "linearity assumption" made by Taplin (1997). It is helpful to represent this graphically. If farmers consistently rank a>b>c then we could derive scores that would put the treatments on a line:

   It is easy to produce data for which this linearity assumption fails. For example, we can have a>b and a>c equally often, suggesting the ordering should look like the line below.

   However at the same time we can have b>c. This might occur if, for example, different farmers were making the comparisons and using different criteria for each one.
   The problem occurs in other examples of ordination, for example, that used by ecologists to describe species occurrence. One answer is to introduce a second dimension. The distances between point a, b, and c can reflect the rankings if they are arranged as follows:

   With just three treatments (as in Example 1) it is clear how an extra term can be introduced in the model to test whether a non-linear arrangement is superior. (In Example 1 it is not.) However, I do not know how to fit models that follow the usual, useful multivariate approach of gradually increasing the number of dimensions until a suitable fit is obtained.
3. In some ranking procedures ties are allowed, for example, farmers are allowed to state that they have no preference between two or more of the treatments. Dittricht et al. (1998) show how the model can be modified to allow for ties. Extra parameters are included so that for each pair we estimate the probability that they tie, as well as the probability that one is ranked above the other.
4. Coe (1, this proceedings) illustrates the value of being able to describe variation at different levels in the design with random effects. It is not clear if these ideas are useful or could be used here. In principle we can fit the model with random effects using the GLM framework; however, maybe it is not necessary. All of the information in the ranks is at the within farm level; hence, we can look at treatment differences and interactions of these within-farm level (or higher level) covariates. However we cannot look at any between farm effects. It is also not clear if plot level covariates can be incorporated. For example, suppose farmers ranked treatments but also reported whether each plot was normally fertile or not, so that plots within one farm could have differing values of the fertility covariate. A model that uses this data would have to be built on the probability of a>b, depending on both the difference in treatment scores and the difference in fertility.

5. The analysis described here is suitable for one objective—for determining treatment effects and their interaction with covariates when the observations are ranks and the design is incomplete or irregular. If the design is more complete, for example, with each respondent comparing all treatments, then other approaches are possible. Different objectives may be of interest, for example, comparison of the rankings under different criteria, or partitioning the sample of respondents into homogeneous groups, when again different methods will be appropriate. Remember also that if the data are rankings produced at, say, a group meeting, so that a consensus is achieved, then no statistical analysis is necessary. Abeysekera (2001) and Riley and Fielding (2001) describe some of the simple alternatives. Taplin (1997) describes a number of the statistical tests available.

Ranks or rates

Having seen how data from these trials can be analyzed, it is worth looking again at the relative merits of using ranking or rating.

First it should be clear that a response measured on a continuous scale, using an accurate and unbiased instrument, contains more information than the equivalent observation using a rating scale with a few levels or than using ranks. Reasons for not using the continuous variate include:

1. Time, money, and logistics, e.g., we may not be able to measure crop yield because we don't know when farmers will be ready to harvest.
2. Lack of a suitable instrument. If we want to assess taste or opinions, there is not much alternative to rating or ranking.
3. Participation. Collecting ranking and rating data requires participants. Other measurement methods may be alienating.

Methods of collecting rating data have been described (e.g., Ashby 1990) and include tools that can give high quality, repeatable and reliable data. It appears that farmers are able to give scores to a large number of alternatives. There are statistical questions regarding the number of levels to use. There is no point in using too many levels, as small differences in rating will probably not reflect real differences in opinion. Note that we do not make a rating scale into a continuous variate simply by using many levels. The fundamental characteristic of a rating scale is that the numbers represent qualitative labels (very good, poor, etc) and the quantitative analogue is missing. This may not be the case if markers are used to represent the score. There is a lot of theory and practice from the social science literature that is relevant here. Respondents are often reluctant to use the ends of a scale, particularly the lower end. Hence a 5-point scale may in practice be used as a 3-point scale. Note that some degree of consistency in the use of the scale by different participants (particularly in different locations) can be achieved by explaining what poor, excellent, etc, mean. For example, poor might correspond to "I would never consider growing this again", while excellent might mean "I would like this to become a main variety on my farm each year".

Ranking is used when it is thought that participants might find it easier to order alternatives than to give them a score.

One reason for this is clear: participants may have a preference for two alternatives given the same score (e.g., both excellent), and hence be able to give them different ranks. However, a shortcoming is immediately clear: the ranks may be the same if both alternatives are also considered poor. This is an important problem. The information in ranks is all "within respondent", that is, we can identify whether, for example, participants consistently rank A above B, and we can determine whether this is true for both male and female respondents. However, we cannot determine what either group of participants actually thinks of A and B. An important part of any research is to make generalizations and extrapolations; however, this often not possible from rank data. Abeysekera (2001) makes the point that the information in ranks is enhanced considerably if some sort of baseline is also measured. For example, if a local control variety is included in each participant's set of alternatives, then we could find a rating for the local control and rank the others relative to this. It is not clear exactly how such data could be analyzed.

A study by ICRAF (1996: 55) assessed the suitability of 12 tree species for firewood. The researcher thought that women could only realistically compare pairs of species. The participants ranked each pair tested, from which it was possible to produce an overall ordination. However, they were also asked the reasons for preferring one species to the other. An alternative design would have used a pilot study of this type to elicit important criteria and then asked these to be rated for each species tested.

Remember there is no information in ranks on effects of quantities that vary across farms. In Example 1, we were unable to determine whether g was more effective on sloping or flat land.

Overall there seems to be little reason to use rank data unless they are specifically required according to the objectives.

References

Abeysekera, S. 2001. *Analysis approaches in participatory work involving ranks or scores.* DFID Theme Paper (revised). UK: Statistical Services Centre, University of Reading.

Agresti, A. 1996. *An introduction to categorical data analysis.* New York: Wiley.

Ashby, J.A. 1990. *Evaluating technology with farmers: a handbook.* Cali: CIAT.

Bradley, R.A., and M.E. Terry. 1952. Rank analysis of incomplete block designs, 1. The method of paired comparisons. *Biometrika* 63:255-262.

Coe, R., and S. Franzel. 2000. Keeping *research* in participatory research. Presented at the 3rd International PRGA Conference, 6-11 November, Nairobi, Kenya.

Coe, R., R. Stern, and E. Allen. 2001. *Analysis of data from experiments.* N.p.: SSC/ICRAF. 250 pp.

Critchlow, D.E., and M.A. Fligner. 1991. Paired comparisons, triple comparisons and ranking experiments as generalised linear models, and their implementation in GLIM. *Psychometrika* 56:517-533.

Dittrich, R., R. Hatzinger, and W. Katzenbeisser. 1998. Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Applied Statistics* 47:511-525

Genstat. 2000. *Genstat for Windows.* 5th Ed. Oxford: VSN International.

International Centre for Research in Agroforestry (ICRAF). 1996. *Annual Report 1995.* Nairobi: ICRAF.

Riley, J., and W.J. Fielding. 2001. An illustrated review of some farmer participatory research techniques. *Journal of Agricultural, Biological and Environmental Statistics* 6:5-18.

Russell, T. 1997. Pair wise ranking made easy. PLA Notes 28:25-26

Taplin, R.H. 1997. The statistical analysis of preference data. *Applied Statistics* 46:493-512.

Discussion Summary

The discussion centered on three themes: (1) the advantages and disadvantages of ranking versus rating, (2) the relationship between rating and parametric and non-parametric tests, and (3) the use of ranking in the context of group interviews, particularly the popular participatory technique of matrix ranking. In terms of the pros and cons of ranking versus ratings, ranking imposes a more rigid framework than rating and ties are not allowed. It is difficult to rank more than six varieties. For example, in an initial mother trial of 52 varieties it is impossible to rank them all, but in a baby trial with only 4 to 8 varieties, ranking is feasible. It was pointed out that with many varieties, farmers may be able to rank the first few, but the rest probably would be considered equally (poor) and would not be ranked effectively. Another problem with ranking is that it may not be possible to know how good or poor a variety really is. Varieties may be ranked, but may all still be considered poor, or vice versa. Different groups of farmers may rank and score differently. For example, ranks for cassava varieties change for different characters. It was pointed out that another potential disadvantage of the use of ratings is that they can vary from village to village; however, ratings are appropriate where a farmer is comparing a new variety against a control in a binary fashion such as “better” or “worse”. It was also pointed out that the analysis of this type of data is not obvious and merits further research. A question posed was whether pair-wise comparisons (rating/ranking) can be made within a group of varieties. The answer was that the data would probably be more inconsistent than ranking/rating all varieties. Another important question was whether there were trade-offs between the use of ratings versus rankings. This is an aspect that has not been widely considered.

Another question was related to the appropriateness of non-parametric tests for rating/ranking. The answer was that such tests do not work for unbalanced designs; they are only testing procedures and do not allow an estimation of the effects of factors on the response. They are therefore less powerful than the methods described in the lecture. Furthermore, these procedures can be extended to genotype x environment analysis. It was pointed out that if ratings are assumed to be normally distributed, then a probability approach could be used. (Editor’s note: I assume this point refers to the use of parametric procedures). This would depend on whether the researcher is happy with the assumption that there is a continuous variable underlying the ratings, and that, in fact, the mean of a 3 rating and a 5 rating is 4.

The final discussion covered matrix ranking. This is a popular participatory technique used with farmer groups for assessing varieties. Group discussion and consensus are often very useful to begin with, particularly for an initial synthesis or screening inventory, but cannot provide an analysis of the heterogeneity present. In some cases consensus may be imposed by the structure of the exercise. Consensus is not necessarily of great interest as often there is more benefit in examining heterogeneity. The value of matrix ranking is limited because it is carried out with groups where consensus is sought, and the results cannot be analyzed statistically since there is no variation (few degrees of freedom). A question was asked on whether the same people are used for ranking the different characteristics. The answer was that usually this is not the case—different people rank different characteristics, compounding the difficulty of analyzing the results.

Analysis of the Demand for Crop Characteristics by Wealth and Gender: A Case Study from Oaxaca, Mexico

MAURICIO R. BELLON

Abstract

Small-scale farmers in developing countries are an important target for participatory breeding efforts. These farmers usually require multiple traits from a key crop. For successful participatory breeding, therefore, it is critical to identify and assess the multiple traits important to farmers and how these traits are supplied by the available germplasm. In addition, once "new" varieties are available it may be relevant to identify which are of most interest to farmers and hence have the highest potential for adoption. This paper describes a set of methods to: (1) identify crop traits important to farmers in a particular area; (2) assess their relative importance, particularly to different farmer groups; (3) assess the distribution of the traits among the varieties grown by farmers; and (4) identify "new" varieties that may be of interest to farmers. These methods are illustrated using data from a project on on-farm conservation of maize landraces in the Central Valleys of Oaxaca, Mexico. The methods were used during the diagnosis phase of the project. Although the examples presented refer to an on-farm conservation project, they are also valid in the context of participatory plant breeding.

Introduction

Many small-scale farmers in developing countries depend on one key crop, such as maize, rice, wheat, or potatoes, for their subsistence and for a source of income. Farmers require multiple traits from this key crop, since it plays various roles and fulfills different needs (Bellon 1996). These multiple traits have to be considered in breeding for successful new germplasm—farmers may not adopt a new variety if it only performs very well for one trait and poorly for many other important traits. Identifying these multiple traits and assessing their relative importance to farmers is not always a simple task. Unlike their counterparts in the developed world, where markets are relatively efficient and the value of different crop traits is reflected in prices, small-scale farmers in the developing world mostly operate under conditions of imperfect markets, where prices do not reflect the value of traits. Furthermore, while these traits may be obviously important to farmers, this information may not be articulated to outsiders, e.g., plant breeders, and therefore not easily recognized.

Participatory plant breeding (PPB) aims to reach farmers that have not been served by conventional breeding (Weltzien et al. 2000). It has been argued that a particular advantage of PPB derives from the strong links generated between scientists and end users. By making selection criteria more relevant to end user needs, participatory breeding can reach poor households that have not yet benefited from modern varieties (Kornegay et al. 1996; Sperling et al. 1993; van Oosterom et al. 1996). Therefore, two fundamental components of participatory breeding efforts are the identification and the assessment of traits that are important to small-scale farmers, particularly for farmer groups that may have been previously left out, have special needs, or face unique conditions. For PPB to be successful there should be methods of assessing so-called "subjective traits". In food crops, these traits typically include taste, aroma, appearance, texture, and other characteristics that determine the suitability of a particular variety for culinary use. Because such traits are a function of human perception, they are difficult to measure quantitatively. This poses a major problem for plant breeders—before breeders can select for a trait, it must be well identified and subject to measurement.[1] Identification and evaluation of subjective traits requires close collaboration between plant breeders, social scientists, and farmers. Social scientists traditionally have played a relatively minor role in plant breeding, but their contribution is fundamental when it comes to identifying subjective traits because of their knowledge of human perceptions and preferences.

The objective of this paper is to describe a set of methods to:

- identify crop traits that are important to farmers in a particular area;
- assess the relative importance of these traits to different groups of farmers;
- assess the distribution of the traits among the varieties grown by a group of farmers; and
- identify "new" varieties with traits that may be of interest to farmers.

These methods are illustrated with data from a project on on-farm conservation of maize landraces in the Central Valleys of Oaxaca, Mexico, and were used during the diagnosis phase of the project. The rest of the paper is divided into five parts. First, background information on the project is presented. Four sections follow this including a description, a presentation of some key results, and a short discussion of each method. Finally the conclusions are presented.

Background

The project "CG Maize Diversity Conservation: A Farmer-Scientist Collaborative Approach"[2] was a pilot study carried out between 1997 and 2002 with small-scale farmers in the

---

[1] Depending on the context, the plant breeder's concern may be to improve these "subjective" traits or simply to maintain them while other traits are being improved. In either case, however, the breeder will need to be able to identify and evaluate the subjective traits.

[2] CG is an abbreviation of CGIAR: the Consultative Group on International Agricultural Research.

Central Valleys of Oaxaca. Its aim was to determine whether it is possible to improve maize productivity (in terms of yield, stability, or other characteristics of interest to farmers) while maintaining or enhancing genetic diversity. An important concept of the project is that while productivity is always important, it is fundamental to establish productivity within a specific context (for example, yields of grain or fodder under high or low input levels, or on steep slopes or flat lands). In other words, productivity is contextual, not absolute, and participatory methodologies can help to establish this context.

The Central Valleys of Oaxaca were chosen for the project because of the importance of the Bolita race (Wellhausen et al. 1952), which has been described as one of the most interesting and productive races of maize in Mexico, although it has not been widely studied or collected (Ortega 1995). Modern varieties have had an almost negligible impact in this region. This should not be construed as farmer conservatism, as discussions with farmers in the region revealed that available modern varieties do not meet their agroecological and cultural requirements. They believed that improved varieties have a long cycle that is not compatible with rainfall patterns in the area, or that these varieties are not well suited to the special preparations and culinary tastes that are very important in the region. The region is ethnically diverse and experiences a wide variation in precipitation (535-1,126 mm/yr). Despite the economic importance of labor migration to the local economy, communities in the Central Valleys place a recognizable emphasis on culture, including culinary practices for maize. However, this diversity also faces threats, mainly from demographic and economic changes. While there are still strong incentives for farmers here to maintain their landraces, there is no guarantee that they will remain interested in maintaining maize diversity as their economic and social conditions change. Therefore, it is important to explore policy options or technical interventions that might support them.

The project was implemented in six communities of the Central Valleys. These communities were selected to represent contrasting agroecological and socioeconomic conditions present in the region. Because results differed across communities, this paper focuses on only one community, Santa Ana Zegache, for simplicity.

The project was divided into three components: (1) diagnosis, (2) interventions, and (3) impact assessment. The diagnosis was made during an earlier phase of the project. It included a collection of landrace samples representative of the regional maize diversity; an agronomic evaluation in scientist designed, farmer managed trials; a participatory exercise to identify a subset of landraces representative of the diversity present in the collection and those most likely to be valuable to farmers, i.e., a group of “elite” landraces; and a baseline survey. The baseline survey included a systematic evaluation of the characteristics farmers considered important in maize landraces and the distribution of these characteristics among the landraces grown.

The interventions component involved facilitating farmer access to a set of “elite” landraces through sales during demonstrations and field days. It also included farmer training in basic principles of maize reproduction, seed selection in the field and in the house (including hands-on exercises in the field), and seed and grain storage principles and techniques.

The impact assessment component included a baseline survey and monitoring a sample of farmers who participated in each of the interventions. Monitoring consisted of yearly systematic interviews with sample farmers about their participation, their perceptions, and the collections of maize landraces they purchased and grew. To link project interventions to the conservation of genetic diversity, a study of the genetic structure and diversity of these landraces was conducted.

Identifying Crop Traits Important to Farmers

Crop traits important to farmers in the Central Valleys were identified during the collection of landraces representative of the regional maize diversity. The landraces were collected from certain villages in the area. The villages were selected by local scientists to include sites representing a range of agroecological and social conditions and wide variation in local maize materials. Selection criteria included physical features such as rainfall and elevation and social factors such as ethnicity and the diversity of maize use by farm households. Time and financial resources limited the number of villages visited to 15.

The different landraces grown in each village were identified with the help of key informants, mainly local authorities involved with farming and people in charge of maize milling outlets. Farmers in each village willing to donate samples were also identified. During collection, these farmers were asked about the traits and uses of each maize type. Their responses were used to form the basis for identifying crop traits important to farmers including grain quality for cooking, fodder quality, and resistance to pests or abiotic stress.

Table 1 presents the responses of farmers who donated maize samples. An analysis of the frequency of farmers' statements shows that they considered many traits relevant. The most frequently cited positive characteristics were related to consumption, such as taste and suitability for special preparations (e.g., atole, a maize based drink), followed by high yield characteristics such as grain weight and short duration. The most frequently cited negative traits were low yield and poor resistance to storage pests. Farmers identified 11 different uses for their maize, including 9 special preparations (Table 2). The salience of consumption characteristics in farmers' statements and the high number of food uses discussed highlight the cultural importance of maize in the region.

Assessing the Relative Importance of Crop Traits to Different Farmer Groups

During the baseline survey, in which a random sample of 40 households in each of 6 villages were interviewed, male and female household members were asked to rate 25 traits in terms of importance (i.e., very important, somewhat important, or not important). This was undertaken to measure the extent to which farmers value these characteristics.

Table 1. Farmers' perceptions of the positive and negative characteristics associated with landraces collected in the Central Valleys of Oaxaca, Mexico.

	Positive characteristics			Negative characteristics		
Concern	Farmers' answers	Percentage responding	Cumulative response	Farmers' answers	Percentage responding	Cumulative response
Consumption	Good for atole	3.87				
	Good quality	1.66				
	Grain color	8.29				
	Good for pasture	2.76				
	Good taste	12.15				
	Good for tortillas	13.26				
	Good dough	1.66				
	Good for tostadas	0.55	44.20			
Yield	Thick grain	0.55		Low production	7.41	
	Produces cobs	0.55		Low yield	18.52	
	High weight	12.71		Small cobs	7.41	
	Good production	1.66		Few rows	3.70	37.04
	Good yield	0.55				
	Good yield by volume	18.23				
	A lot of grain	0.55	34.80			
Duration	Early	10.50	10.73			
Sale	Sells well	2.21	2.26			
Processing	Easy to shell	1.66	1.69			
Adaptation	Well adapted	1.66	1.69			
Abiotic stress	Withstands drought	1.10		Tall plants (lodging)	14.81	14.81
	Withstands cold	0.55	1.66			
Biotic stress	Withstands pests	0.55		Tall plants (lodging)	14.81	14.81
	Withstands weeds	0.55	1.10			
Storage	Stores well	2.21	2.21	Rot	22.22	
				Not resistant to weevils	7.41	
				Cob rots	11.11	
				Grain rots	3.70	44.44
Total		100.00			100.00	

Source: Bellon et al. (2002).

Table 2. Special culinary preparations and uses of maize identified by farmers during the collection of maize landraces, Central Valleys of Oaxaca, Mexico.

Preparation/use	Description
Tortillas	Flat bread made of maize
Atole	Maize based drink
Nicuatole	Type of gelatin made of maize
Buñuelos	Fried thin wafer
Nixtamal	Dough to make tortillas
Pinole	Powder made of maize
Pozole	Maize based drink
Tamales	Steamed maize mixture with a filling wrapped in a maize husk or banana leaf
Tejate	Sweet maize drink with cacao
Tlayudas	Special type of tortilla
Animal feed	Grain fed to poultry, leaves and stalk fed to cattle

Source: Bellon et al. (2002).

The list of characteristics included all those identified across the region. The reader should note that this list of 25 characteristics includes those that were not identified explicitly by farmers, but were included by researchers who perceived them to be important (which in fact they were). The list included yield stability ("produces something even in a bad season"), yield of tortillas by kilogram of dough, and suitability for all uses identified in the region (special dishes and preparations).

To examine the importance of these traits to different groups within the sample, all households were grouped into wealth classes through a wealth ranking exercise assisted by key informants in the project communities. The informants discussed local perceptions of wealth and, based on these views, identified three wealth classes (well-off, poor, and those in between) and characteristics of each. They were then asked to classify each participant into one of the three categories. The categories used to rank wealth were very similar among communities and are summarized in Table 3.

To test for gender differences in the demand for the aforementioned maize characteristics, the ratings given by male and female participants to each characteristic were compared using a Wilcoxon signed ranks test for two related samples. To test for differences in the demand for characteristics according to wealth, the ratings were grouped by wealth class. These ratings were compared for male and female groups according to the wealth class of the household they belonged to. The comparison was made using a Kruskal Wallis analysis of variance by ranks.

Table 4 compares the ratings of the important maize characteristics by men and women in farming households. The table shows the average rating, based on the following scale: 1 = very important, 2 = somewhat important, and 3 = not important. A Wilcoxon signed ranks test (a non-parametric statistical procedure) was used to test for statistically significant differences between male and female ratings for a characteristic.[3]

Table 3. Variables used to rank households by wealth, Central Valleys, Oaxaca, Mexico.

- Source of monetary income: off-farm labor, non-farm labor, ownership of business, remittances from the USA or other parts of Mexico.
- Access to land: quantity and quality of land, property rights over land, and access to irrigation.
- Family demographics: availability of family labor, consumption demands of family members.
- Ownership of animals, including access to animal manure.
- Access to bullocks or tractor for land preparation.
- Level of education.
- Access to transportation, particularly ownership of a pickup.
- Diversity of crops on the farm.
- Access to finance and savings.
- Importance of agriculture.
- Interest in agriculture.
- Ability to hire labor or purchase inputs.

---

[3] The table reports the mean or average rating, which makes it easier to identify differences and trends, but the test is based on the null hypothesis that the median (not the mean) of the population of differences is zero (Daniel 1978: 135-9). A non-parametric test, such as that used here, is more appropriate because the ratings are ordinal and their underlying distribution is unknown and is not likely to be normal. In this case this test is used because males and females were not selected independently of each other, but were members of the same household (they were related).

Table 4. Average ratings of importance of maize characteristics by males and females, Santa Ana Zegache, Oaxaca, Mexico.

Concern	Characteristic	Average rating			Top five characteristics	
		Males	Females	P value†	Males	Females
Consumption	Taste of tortillas	1.78	1.38	0.01		
	Good for atole	1.80	1.55	ns		
	Good for tlayudas	2.23	1.63	0.00		
	Ease of shelling	2.08	2.68	0.00		
	Good for storage	1.08	1.50	0.00	2	
	Good pasture	1.90	1.70	ns		
	Good feed	1.20	1.53	0.02	5	
	Nixtamal quality	2.05	1.33	0.00		5
	Good for tamales	2.25	2.23	ns		
	Good for tejate	2.73	2.38	0.01		
	Good for pozole	2.95	2.80	0.03		
	Good for nicoatole	2.90	2.70	0.02		
Yield	Yield by weight	1.25	1.05	0.03		2
	Yield by volume	1.28	2.03	0.00		
	Yield of tortillas	1.98	1.45	0.00		
	Yield stability	1.13	1.03	0.10	4	1
Duration	Duration	1.40	1.55	ns		
Sale	Ease of sale	1.85	1.53	0.03		
Abiotic stress	Withstands drought	1.03	1.08	ns	1	3
	Withstands wind	2.55	1.88	0.00		
	Withstands cold	2.75	2.30	0.00		
Biotic stress	Withstands weeds	2.45	2.35	ns		
	Withstands pests	2.40	1.60	0.00		
Management	Produced with little labor	1.40	1.85	0.01		
	Produced with little money	1.10	1.18	ns	3	4

† P value associated with a Wilcoxon signed ranks test for two related samples; ns = not significant.
Source: Bellon (2001).

A comparison of men's and women's ratings shows highly significant differences for most characteristics. Of the 25 characteristics, only 7 had no statistically different ratings. Of the 5 most highly rated characteristics, however, men's and women's ratings coincided in 3: tolerance to drought, yield stability, and low cash investment. Men also included storage properties and suitability for feed in the top 5 characteristics, while women included yield by weight and nixtamal[4] quality. These results also show that men and women value many characteristics: the average ratings for 14 and 17

---

[4] Nixtamal is the dough used to make tortillas. The milled maize is first soaked in water and lime.

characteristics for men and women, respectively, were between “very” and “somewhat important”.

These results show important gender differences in the demand for maize characteristics. Failure to recognize these differences would lead to biased interventions. In the Oaxaca project, if males alone had participated in the voting exercise to identify landraces for distribution, it is very likely that their choices would have been of more interest to them than for women. These results also have implications for breeding. Improvements in yield stability or drought tolerance would be beneficial for both men and women, but any improvements that come at the cost of nixtamal quality, for example, could negatively affect women more than men.

The large number of characteristics rated as “very important” or “somewhat important” also suggests that both men and women demand diversity, since it is unlikely that one maize type will be good at supplying all of their valued characteristics. Therefore there may not be a best or ideal maize type. Farmers in the study area require a range of maize types, and this motivates researchers to give farmers access to diversity in the Oaxaca project.

Similar analyses can be conducted using any farmer grouping or classification. Table 5 classifies men and women separately by wealth rank (i.e., well-off, intermediate, poor) and reports the average rating for each rank based on the following scale: 1 = very important, 2 = somewhat important, and 3 = not important. A Kruskal Wallis one-way analysis of variance by ranks (a non-parametric statistical procedure) was used to test for differences among the ratings, i.e., whether or not each rating for a characteristic was statistically equal among the three wealth groups.[5]

The analysis showed that the ratings of characteristics among wealth groups were not statistically different.[6] Not surprisingly, of the top five characteristics, all wealth ranks among men and among women agreed on the importance of yield stability, tolerance to drought, and low cash investment. Men across all wealth categories agreed on the importance of storage properties. Women across wealth categories agreed on yield by weight. For poor women, the taste of tortillas and nixtamal quality were also particularly important.

These results suggest that improvements in any of the traits may benefit all farmers equally. If differences between wealth groups had emerged for certain characteristics, however, the improvement of those characteristics would have benefited some groups more than others. It is also important to note that losses in some characteristics may be more negative for some groups than for others. For example, if resistance to lodging is rated significantly higher by the rich farmers, the introduction of a more resistant variety may benefit them more than the other groups. On the other hand, if the poor farmers rate resistance to storage pests significantly higher, and a new variety has

---

[5] The table reports the mean or average rating from which it is easier to identify differences and trends. However, the test is based on the null hypothesis that the three population distribution functions are identical against the alternative hypothesis that they do not all have the same median (Daniel 1978: 200-5).

[6] The exception to this is the characteristic “good for tamales” among men, where the poor rated it higher than the other wealth classes.

Table 5. Average ratings of importance of maize characteristics by wealth rank for males and females, Santa Ana Zegache, Oaxaca, Mexico.

		Males by wealth rank					Females by wealth rank				
Concern	Characteristic	Well -off	Interme- diate	Poor	Total	P value†	Well -off	Interme- diate	Poor	Total	P value†
Consumption	Taste of tortillas	1.79	1.83	1.83	1.81	ns	1.38	1.54	1.00	1.38	ns
	Good for atole	1.64	1.92	1.67	1.75	ns	1.38	1.69	1.33	1.50	ns
	Good for tlayudas	2.21	2.42	2.17	2.28	ns	1.62	1.54	1.67	1.59	ns
	Ease of shelling	2.21	2.00	2.00	2.09	ns	2.54	2.77	2.67	2.66	ns
	Storage properties	1.14	1.08	1.00	1.09	ns	1.31	1.62	1.50	1.47	ns
	Good pasture	1.93	2.00	1.50	1.88	ns	1.46	1.92	2.00	1.75	ns
	Good feed	1.29	1.17	1.00	1.19	ns	1.46	1.54	1.67	1.53	ns
	Nixtamal quality	2.07	2.08	2.17	2.09	ns	1.46	1.31	1.00	1.31	ns
	Good for tamales	2.50	2.25	1.83	2.28	0.06	2.46	2.08	2.17	2.25	ns
	Good for tejate	2.86	2.75	2.67	2.78	ns	2.54	2.23	2.33	2.38	ns
	Good for pozole	3.00	2.92	2.83	2.94	ns	2.85	2.85	2.67	2.81	ns
	Good for nicoatole	2.86	3.00	2.83	2.91	ns	2.69	2.69	2.50	2.66	ns
Yield	Yield by weight	1.36	1.08	1.33	1.25	ns	1.15	1.00	1.00	1.06	ns
	Yield by volume	1.29	1.50	1.17	1.34	ns	2.15	1.85	2.00	2.00	ns
	Yield of tortillas	1.93	2.00	2.00	1.97	ns	1.62	1.54	1.17	1.50	ns
	Yield stability	1.14	1.00	1.00	1.06	ns	1.08	1.00	1.00	1.03	ns
Duration	Duration	1.29	1.58	1.50	1.44	ns	1.46	1.54	1.50	1.50	ns
Sale	Ease of sale	1.71	2.00	1.83	1.84	ns	1.31	1.85	1.83	1.63	ns
Abiotic stress	Withstands drought	1.00	1.00	1.00	1.00	ns	1.00	1.15	1.17	1.09	ns
	Withstands wind	2.43	2.58	3.00	2.59	ns	2.08	1.69	2.00	1.91	ns
	Withstands cold	2.71	2.50	3.00	2.69	ns	2.31	2.38	2.17	2.31	ns
Biotic stress	Withstands weeds	2.14	2.67	2.50	2.41	ns	2.15	2.31	2.67	2.31	ns
	Withstands pests	2.36	2.33	2.67	2.41	ns	1.31	1.85	1.50	1.56	ns
Management	Produced with little labor	1.36	1.42	1.50	1.41	ns	1.92	1.77	1.67	1.81	ns
	Produced with little money	1.07	1.08	1.00	1.06	ns	1.15	1.23	1.17	1.19	ns

† P value associated with a Kruskal Wallis one-way analysis of variance by ranks for males and females separately; ns = not significant.
Source: Bellon (2001).

substantially lower resistance to these pests, the cost of adopting the new variety will be higher for the poor group than for the others.

By analyzing the ratings of these characteristics, as shown here, researchers have a method for predicting how the costs and benefits of introducing new germplasm are likely to be distributed among different groups of farmers and/or members of farming households.

Assessing the Distribution of Important Traits among Varieties Grown by Farmers

Varieties are not homogenous in supplying the characteristics that farmers want, otherwise one variety would be enough. Farmers are keenly aware of the differences in performance of their varieties and that these differences entail trade-offs. For example, a variety may be very good for resistance to lodging but poor for husk coverage, while another may have the opposite combination. To assess the extent to which each of the maize types supply the 25 characteristics, farmers were asked to rate each of their maize types into three categories (very good, regular, or poor). These ratings were compared among landraces, according to grain color (the main taxonomic variable used by farmers), for male and female groups separately. The comparison was made with a Kruskal Wallis analysis of variance by ranks.

Table 6 compares farmers' ratings of the performance of Blanco (white), Amarillo (yellow), Negro (black), and Belatove (red) maize types by gender group. Each maize type was rated for each characteristic, based on the following scale: 1 = very good, 2 = intermediate, 3 = poor. For the characteristics related to labor and cash investments, the rating scale was: 1 = little, 2 = intermediate, 3 = a lot. The table reports the average rating per maize type,[7] except for yield by weight, yield by volume, yield of tortillas, anthesis (days to male flowering), and days to harvest (an indicator of duration), for which the means of estimates provided by farmers in the appropriate units are used. A non-parametric Kruskal Wallis one-way analysis of variance by ranks for the ratings and a parametric one-way analysis of variance for the continuous variables were used to test for statistical differences across the different maize types for each characteristic.

Men's assessments of the four types showed statistically significant differences for most characteristics. The Blanco type is superior to the others for all characteristics, except for having the longest growing cycle or duration. On the other end of the spectrum, the Belatove type is inferior to all others, except for having the shortest duration. Amarillo and Negro are intermediate. The assessment shows a gradient of performance from Blanco to Amarillo to Negro to Belatove. These results suggest a trade-off between duration and good performance for other traits. All types, however, are considered particularly inferior for storage properties. These results are consistent with those obtained from the folk maize taxonomy exercise, in which farmers expressed that planting date—and therefore the uncertainty of the duration of the growing season—was very important. While Blanco had a high yield, multiple uses, and was easy to sell, it also had the longest growing cycle. Its longer duration was a disadvantage if rains were delayed and it had to be planted late because of increased risk of exposure to drought and frost. The other maize types had shorter growing cycles (Blanco>Amarillo>Negro>Belatove) and provided farmers with the flexibility to respond to the uncertain onset of rains, even though they were inferior for other characteristics.

Women's assessments of the four maize types showed statistically significant differences for a lower number of characteristics than did men's assessments. For example, unlike men,

---

[7] The table reports the mean or average rating, which makes it easier to identify differences and trends. However, the test is based on the null hypothesis that the three population distribution functions are identical, against the alternative hypothesis that they do not all have the same median (Daniel 1978: 200-5).

Table 6. Male and female farmers' ratings of the performance of different maize types according to several characteristics of importance, Santa Ana Zegache, Oaxaca, Mexico.

		Males Maize type						Females Maize type					
Concern	**Characteristic**	**Blanco**	**Amarillo**	**Negro**	**Belatove**	**Total**	**P value†**	**Blanco**	**Amarillo**	**Negro**	**Belatove**	**Total**	**P value†**
Consumption	Taste of tortillas	1.00	1.11	1.00	1.33	1.04	0.01	1.03	1.07	1.00	1.00	1.03	ns
	Good for atole	1.00	1.47	2.46	2.33	1.42	0.00	1.00	1.33	2.40	3.00	1.32	0.00
	Good for tlayudas	1.00	1.17	1.00	1.00	1.04	0.09	1.00	1.00	1.00	1.00	1.00	ns
	Nixtamal quality	1.00	1.22	1.29	1.67	1.13	0.00	1.00	1.07	1.00	1.00	1.02	ns
	Good for tamales	1.00	1.06	1.93	2.33	1.24	0.00	1.00	1.07	1.10	1.00	1.03	ns
	Good for tejate	1.00	2.00	2.36	2.33	1.55	0.00	1.03	1.80	2.20	2.00	1.39	0.00
	Good for pozole	1.00	1.83	2.43	2.33	1.52	0.00	1.03	1.20	1.80	1.00	1.18	0.00
	Good for nicoatole	1.00	2.11	1.50	3.00	1.44	0.00	1.00	1.87	2.50	3.00	1.46	0.00
	Ease of shelling	1.05	1.11	1.36	1.00	1.12	ns	1.45	1.07	1.00	1.00	1.29	0.01
	Storage properties	1.75	2.06	2.71	3.00	2.05	0.00	1.85	2.20	2.90	3.00	2.11	0.00
	Good pasture	1.00	1.00	1.93	2.33	1.23	0.00	1.08	1.07	1.90	3.00	1.23	0.00
	Good feed	1.00	1.00	1.07	1.00	1.01	ns	1.00	1.00	1.00	1.00	1.00	ns
Yield	Yield by weight‡	653.80	544.90	520.40	461.30	595.10	0.01	395.80	296.00	230.00	156.70	346.90	0.01
	Yield by volume§	4.00	3.99	3.99	4.00	3.99	ns	3.97	3.97	3.98	4.00	3.97	ns
	Yield of tortillas††	38.37	38.78	39.14	39.00	38.64	ns	36.05	36.80	38.00	40.00	36.58	ns
	Yield stability	1.08	1.56	1.86	2.00	1.37	0.00	1.63	1.33	1.20	1.00	1.48	0.04
Duration	Anthesis‡‡	79.90	74.60	62.90	60.00	74.60	0.00	74.00	65.90	53.50	45.00	68.90	0.00
	Harvest§§	121.90	116.20	97.40	95.00	114.90	0.00	127.50	118.30	97.10	96.00	120.50	0.00
Sale	Ease of sale	1.00	1.28	2.00	2.00	1.29	0.00	1.00	1.20	1.80	2.00	1.18	0.00
Abiotic stress	Withstands drought	1.35	1.89	2.64	2.33	1.76	0.00	1.54	1.47	1.60	2.00	1.54	ns
	Withstands wind	1.25	1.33	1.21	1.33	1.27	ns	1.48	1.60	1.20	2.00	1.47	ns
	Withstands cold	1.13	1.11	1.14	1.00	1.12	ns	1.25	1.47	1.40	1.00	1.32	ns
Biotic stress	Withstands weeds	1.63	2.06	2.00	1.67	1.80	0.01	1.80	1.93	1.60	1.00	1.79	ns
	Withstands pests	1.45	1.56	1.71	1.33	1.52	ns	1.58	2.07	2.11	3.00	1.78	0.00
Management	Produced with little labor	2.50	2.33	2.50	2.00	2.44	ns	2.30	2.33	2.40	2.00	2.32	ns
	Produced with few purchased inputs	2.58	2.56	2.57	2.00	2.55	ns	2.33	2.40	2.40	2.00	2.35	ns

† P value associated with a Kruskal Wallis analysis of variance test for the ratings, except for yield by weight, yield by volume, yield of tortillas, anthesis, and harvest where it is associated with a parametric analysis of variance; ns = not significant.

‡ Expected yield (kg/ha) calculated from the best, worst, and more frequent yields declared by farmers for each maize type using the triangular distribution method (Hardaker et al. 1997).

§ kg/local unit of volume (almud).

†† Number of tortillas/almud.

‡‡ Number of days to anthesis (male flowering).

§§ Number of days to harvest.

Source: Bellon (2001).

the women did not consider differences for consumption qualities, such as taste of tortillas, nixtamal quality, tlayudas, and tamales, but they did for ease of shelling. All of these characteristics relate to aspects of maize preparation that women are responsible for. Women provided much lower estimates for yield by weight and duration, but their ordering of these characteristics was similar to that given by the men. An important difference is that the women considered Amarillo, Negro, and Belatove to have higher stability than Blanco. In general they rated colored maize types much more highly than men did. In particular, women perceived colored maize types to perform better compared to Blanco than men did, so the trade-off between good performance and duration is not as strong among women as men. It appears that colored maize types are considered more important by females than males, therefore, women may be playing an important role in the conservation of these maize types.

The study suggests that the performance of any new variety introduced into the Central Valleys of Oaxaca could be rated according to the aforementioned characteristics by a panel of farmers. The farmers could predict how the variety might fit into the production system, which varieties it might displace, and how it would complement other varieties. For example, a shorter duration white maize type, equal in other respects to the currently grown white type, could displace the colored maize types since it would decrease the trade-off between desirability and duration. On the other hand, improving the storage quality of colored maize types may encourage their conservation.

Identifying "New" Varieties that May Have Interesting Traits to Farmers

While it is important to know how currently planted varieties perform in terms of important traits, it may be desirable to identify "new" varieties that may be of interest to farmers. In the Oaxaca project, these new varieties were simply landraces from the region that captured regional diversity (collected at the beginning of the project) but were mostly unknown in the project villages.

A set of 170 materials was evaluated by breeders and farmers in farmers' fields during the rainy season of 1997. The set comprised 152 maize populations (landraces) collected from the area, 16 landraces from the gene bank, and 1 improved variety derived from the Bolita landrace. A crucial problem was the impossibility of working with such a large number of entries to promote on-farm conservation—the entries do not contribute similarly to diversity, nor are they of equal interest to farmers. There was a need to identify a subset that would be interesting to farmers and contribute to maize diversity in the region (see Bellon et al. 2002 for details).

To identify landraces of more interest to farmers among the landraces collected, farmers from the region were invited to evaluate the 170 populations at physiological maturity and harvest at 3 trial locations during the agronomic evaluation of 1997. At harvest, 213 farmers (54% females) participated. All ears from the inner rows of the experiment were harvested and laid out in front of the stand, so farmers could judge grain yield and examine the ears. Farmers walked through the trial, observed the landraces, and recorded the plot numbers of their preferred landraces. Each plot recording was interpreted as a vote for that landrace. The purpose of this exercise was to obtain a rapid sort or classification of landraces according to farmers' expressions of interest. The exercise enabled us to systematically deal with many materials (170) and many farmers (approximately 70 per field day) in a relatively brief period (2-3 hours). Field days were open to all those who wished to participate. We assumed that the number of participants' votes reflected the value of a landrace. Originally we intended to ask participants to rank the materials for a set of traits; however, this was precluded by the large number of materials exhibited and the large number of participants per demonstration.

Figure 1 compares the votes made by male and female participants. Each point is a collected landrace. The x-axis represents the percentage of men who voted for a particular landrace, while the y-axis represents the same information for women. While there was a high correlation between male and female votes (0.70, $p<0.0001$), their voting patterns were not the same. For example, a higher number of females voted for particular landraces compared to males, suggesting that there was a higher level of agreement among females in their choices relative to males.

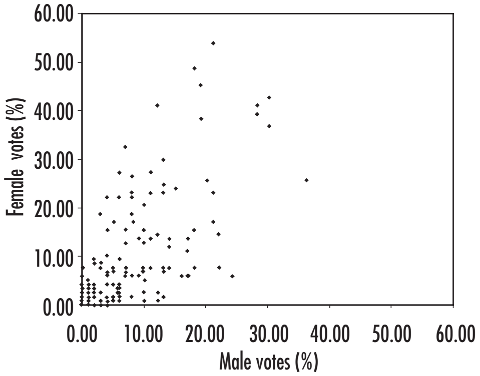


Figure 1. A comparison of the votes cast by men and women participating in field days in the Central Valleys, Oaxaca, Mexico.
Source: Bellon et al. (2002).

The most desirable landraces (those chosen frequently by both men and women) are located in the upper right section of the graph. The least desirable landraces are located in the lower left section. Results from the voting exercise show that there is great interest in diversity among participating farmers, as well as gender differences (Table 7). While male farmers voted for many landraces, women, on average, chose more landraces. Only about a third of the men and half of the women voted for the most popular landraces. Few landraces received no votes, suggesting that participants want a range of materials and that there is no "best" landrace. The fact that only a few participants voted for no landraces indicates a high degree of interest in all entries exhibited.

Breeders evaluated the landraces with the help of a selection index based on yield (Mg/ha), ear rot (%), erect plants (%), and moisture (%) calculated for each landrace in the set. The selection index was used to account for grain yield, grain quality, and standability. The selection index is a linear function of the

Table 7. Voting patterns of male and female participants in field days in the Central Valleys, Oaxaca, Mexico.

Landrace votes	Males	Females
Average number	10.800	13.700†
Maximum number	40.000	38.000
Minimum number	0.000	0.000
Landraces with zero votes (%)	12.500	15.800
Maximum votes awarded to one landrace (%)	36.400	53.900
Participants who did not vote for any landrace (%)	5.100	3.400
Correlation between votes received and selection index for landraces	0.369‡	0.362‡

† t-test for equality of means significant at the 0.01 level (2-tailed).
‡ Correlation significant at the 0.01 level (2-tailed).
Source: Bellon et al. (2002).

four variables mentioned above, adjusted to give a score of 100 to the best entry. The indices for the rest of the entries express the gap between their performance and that of the best entry. Farmers' votes were compared to this selection index (Table 7). There was a significant and positive correlation between the votes received by a landrace and its selection index. It is evident that farmers appear to consider agronomic performance, and that their voting patterns are far from random.

Conclusions

Although the examples presented in this paper refer to an on-farm conservation project, they are also valid in the context of PPB. In PPB there is a need to identify and assess the importance of crop traits, as well as to understand their distribution among current varieties. This should allow breeders and other scientists to identify gaps in the performance of these varieties, and to target and prioritize the traits that merit improvement. For new germplasm to be of interest to farmers, it has to provide a benefit that is lacking in the current germplasm or be superior for at least some traits. An important aspect of the complementarity between new and current germplasm is the extent to which the incorporation of a new variety into the set of varieties planted by a group of farmers reduces the trade-offs between varieties in the set. This may be analyzed by the methods presented here. For example, for a variety that yields well but has poor taste qualities, there is a trade-off between yield and taste. Introducing a variety with good yield and improved taste lessens the trade-off between the traits. Furthermore, by analyzing the importance of various traits to different social groups, one can assess how the improvement of certain characteristics could benefit or harm some groups more than others.

Allowing farmers to see and evaluate new varieties, and systematically capturing their assessment, as illustrated by the "voting" method, should allow breeders and other scientists to identify traits and varieties that are interesting and potentially valuable to farmers. The selection index of farmers may differ to that of breeders, both in terms of traits included and the weighting given to each trait. Also, the farmers' selection index is not written down anywhere and therefore may not be easily elicited. However, by allowing farmers to see and assess new materials and have a systematic method of rating their assessments, their selection index can be revealed. A discussion of how to transform the voting exercise into a farmer selection index is beyond the scope of this paper, but remains an area for further study. The method described here may serve in the initial phases of a participatory breeding project to identify suitable donor or parent materials. Alternatively it could be used as part of a participatory varietal selection exercise. While there are other approaches of assessing new materials by farmers, the method described here is easy to implement. It can deal with many entries and many farmers in a relatively short time and requires minimal investment in training farmers to rate varieties.

Acknowledgements

This work was carried out with a grant from the International Development Research Centre, Ottawa, Canada. It is part of the project "CG Maize Diversity Conservation: A Farmer-Scientist Collaborative Approach", implemented jointly by the International Maize and Wheat Improvement Center (CIMMYT) and Mexico's National Institute of Forestry, Agriculture, and Livestock Research (INIFAP). The author gratefully acknowledges the work of the project team (Flavio Aragón, Julien Berthaud, José Alfonso Aguirre Gómez, Jorge Mendoza, Irma Manuel Rosas, Melinda Smale, and Suketoshi Taba) without who this paper could not have been written.

References

Bellon, M.R. 1996. The dynamics of crop infraspecific diversity: A conceptual framework at the farmer level. *Economic Botany* 50:26-39.

Bellon, M.R. 2001. *Participatory research methods for technology evaluation: a manual for scientists working with farmers.* Mexico, D.F.: CIMMYT.

Bellon, M.R., J. Berthaud, M. Smale, J.A. Aguirre, S. Taba, F. Aragón, J. Díaz, and H. Castro. 2002. Participatory landrace selection for on farm conservation: An example from the Central Valleys of Oaxaca, Mexico. *Genetic Resources and Crop Evolution* (in press).

Daniel, W.W. 1978. *Applied nonparametric statistics.* Boston, U.S.A.: Houghton Mifflin.

Hardaker, J.B., R.B.M. Huirne, and J.R. Anderson. 1997. *Coping with risk in agriculture.* London: CAB International.

Kornegay, J., J.A. Beltran, and J. Ashby. 1996. Farmer selections within segregating populations of common bean in Colombia. In P. Eyzaguirre and M. Iwanaga (eds.), *Participatory plant breeding: Proceedings of a workshop on participatory plant breeding.* Rome, Italy: IPGRI. Pp. 151-159.

Ortega, R. 1995. Origen de la agricultura e importancia de los valles centrales de Oaxaca. In M.A. Vásquez Dávila (ed.), *La Tecnología Agrícola Tradicional.* Instituto Indigenísta Interamericano/Consejo Nacional de Ciencia y Tecnología, Oaxaca, México. Pp. 189-200.

Sperling, L., M.E. Loevinsohn, and B. Ntabomvura. 1993. Rethinking the farmer's role in plant breeding: local bean experts and on-station selection in Rwanda. *Experimental Agriculture* 29:509-519.

van Oosterom, E.J., M.L. Whitaker, and E. Weltzien R. 1996. Integrating genotype by environment interaction analysis, characterization of drought patterns, and farmer preferences to identify adaptive plant traits for pearl millet. In M. Cooper and G.L. Hammer (eds.), *Plant adaptation and crop improvement.* Wallingford, U.K.: CAB International. Pp. 383-402.

Weltzien, E., M.E. Smith, L.S. Meitzner, and L. Sperling. 2000. *Technical and institutional issues in participatory plant breeding from the perspective of formal plant breeding. An analysis of issues, results, and current experience.* CGIAR Systemwide Program on Participatory Research and Gender Analysis for Technology Development and Institutional Innovation Working Document 3.

Wellhausen, E., J. Roberts, L.M. Roberts, and E. Hernández. 1952. *Races of maize in Mexico: their origin, characteristics, and distribution.* Cambridge, U.S.A.: The Bussey Institution, Harvard University.

Discussion Summary

The discussion following the presentation focused on how wealth, gender, and agroecological differences were dealt with in the case study. It was surprising to some workshop participants that there were no differences between wealth categories in terms of the importance of traits—in other settings there are usually clear differences, particularly between the rich and other socioeconomic groups. In terms of gender, differences between the results of this case study and those from other areas were also noted. For example, in Yemen, males and females selected the same varieties because the females visited the trial prior to the official evaluation, and felt that they had to agree with the men's selection. In Oaxaca, there were obvious differences between many of the men's and women's selections. The key point here is that results may be cultural and location specific.

A question was raised on whether the researcher intervention raised productivity, but the author pointed out that the main aim of the project was to give farmers access to diversity and, in some cases, to give back lost varieties. Furthermore, self selection was encouraged in the project, and participants seemed to be more interested in diversity than in the population at large. Discussion then focussed on possible biases related to the types of farmers participating in the study. Even though everybody was invited to participate, only certain farmers chose to do so. One of the issues of participation is that, by definition, it entails a bias because participants are those who are willing to participate. In the Punjab, for example, wealthy farmers are more likely to participate and to buy. Nonetheless, a systematic monitoring of the participants was conducted to identify possible biases. Finally, it was noted that all materials sold were local landraces, though not necessarily from the communities where the participants lived. The materials were chosen to represent the regional landrace diversity.

Identifying Farmers’ Preferences for New Maize Varieties in Eastern Africa

HUGO DE GROOTE, MOSES SIAMBI, DENNIS FRIESEN, AND ALPHA DIALLO

Abstract

To bridge the gap between breeders and farmers and to ensure that new varieties satisfy farmers’ preferences and suit their socioeconomic situations, the International Maize and Wheat Improvement Center (CIMMYT) is developing and adapting participatory methods for identifying farmers’ maize variety preferences in East Africa. Several methods are being used: breeding on-station under stress conditions (simulating farmers’ conditions); participatory rural appraisals; farmers’ evaluation of new varieties on-station; and mother and baby trials. The latter is a new approach consisting of a central researcher-managed “mother” trial comprising all tested varieties, and satellite or “baby” trials, which are farmer managed and test a subset of varieties from the mother trial. The results show that farmers are eager to participate in selecting new varieties but their selection is very different from that of breeders. Moreover, farmers’ evaluations and testing on-farm both show very high variance. The methodology used in the study to identify farmers’ criteria and to facilitate farmer evaluation of the varieties was very convenient for data collection but not for data analysis. For future trials, sufficient resources need to be made available to assure enough high quality data for statistical analysis. These data will make it possible to adjust the breeders’ index in order to make it more responsive to farmers’ preferences.

Introduction

Improving communication between farmers and breeders

The purpose of plant breeding is to develop improved genetic material, particularly with higher yield. This is achieved through a more efficient use of resources by the plant, or through increased resistance to pests and other types of stress. In developing new materials and extending them to farmers, classical plant breeding faces two major obstacles. First, new varieties can be disappointing to farmers where undesirable traits go undetected during the breeding process. In Kenya, for example, several late maturing maize varieties have been released over the last 10 years, but none has surpassed the very popular H614D—a hybrid released in 1986 based on H614C, which was released in 1976. Secondly, breeders necessarily discard many crosses and varieties during the selection process because of traits considered undesirable;

however, these traits may actually be of interest to farmers. On the Kenyan coast, for example, a technician brought home a discarded cassava variety and this variety is now being rapidly adopted.

These examples illustrate the communication gap between breeders and farmers, which is also reported elsewhere (Kamara et al. 1996). However, efforts are being made to reduce this gap, in particular by a process called participatory plant breeding (PPB). The aim of PPB is to improve communication between farmers and breeders so that farmers' concerns and preferences are incorporated earlier in the research process; research is accelerated; and the adoption rate improves (Sperling et al. 1993; Eyzaguirre and Iwanaga 1996). Farmers are not only asked for their opinion (the consultative approach; see Biggs 1989 for definitions) and collaboration (collaborative approach), but are actively invited to help set the research agenda (collegiate approach). By inviting farmers to make decisions in the research process, it is assumed that they will not only adopt, but also, and more importantly, adapt the available technology to their own needs and environment (Ashby 1991). For this communication to succeed, new tools are needed, along with a common language that suits both breeders and farmers.

Breeders usually select their material by analyzing large amounts of data, which are systematically obtained from highly controlled situations to reduce variability. Farmers, however, select varieties based on small experiments and observations in the field and from anecdotal evidence, using intuitive multi-factor analysis (Sumberg and Okali 1997). Social scientists play an important role in bridging the gap between farmers and breeders, and this paper presents an interdisciplinary approach applied to maize breeding in East Africa.

Maize is a very important food crop in East Africa, but productivity has not kept up with population growth in the region. Soil fertility is declining and pest pressure is increasing, therefore maize breeding needs to focus on developing varieties tolerant to drought and nitrogen stress (Bänziger et al. 2000). A large number of varieties have been developed and, since 1999, farmers have been invited to collaborate in their evaluation.

Breeding maize for East Africa

Breeding stress tolerant varieties for East Africa started with physiological studies, determining the important traits, and then selecting a large number of promising varieties from the CIMMYT genebank in Mexico. Methods were developed to test these varieties under stress conditions on-station, simulating farmers' conditions as closely as possible. Using these methods, varieties were crossed and further selected and a large number have now been tested in Zimbabwe—CIMMYT's regional center for Southern Africa. CIMMYT tested the varieties using mother and baby trials (described later).

CIMMYT's maize breeding program in East Africa started in 1997. Diagnostic surveys and expert interviews revealed that food security is one of the major problems facing farmers in the region, and that maize production is their main strategy to combat this. The major constraints to maize production, as perceived by farmers, are drought, soil fertility, seed availability, and pests (De Groote et al. 2001b). Generally, the main pests are stem borers and weevils (the major storage pest), while around Lake Victoria the predominant pest is striga, a parasitic weed that attacks maize and sorghum. To deal with these constraints, CIMMYT and its partners have developed various projects. The most important of these are the African Maize Stress (AMS) Project, the Insect Resistant Maize for Africa (IRMA) Project, the Striga Project, the Seed Supply Systems Project, and the Quality Protein Maize (QPM) Project (Table 1). A team of five CIMMYT scientists (two breeders, two agronomists, and one economist) works on the different projects in collaboration with colleagues from the national programs, mostly within the East and Central Africa Maize and Wheat Network (ECAMAW). Social scientists form a working group within this network and focus on participatory research and analysis of the maize sector.

The participatory work involves testing new technologies, particularly varieties, but also cultural practices such as soil and water conservation and pest control. In the maize sector study, an analysis is made of the system to ensure that the new technologies to be introduced are appropriate. This includes studying maize marketing, seed production and distribution methods, fertilizer and pesticide markets, and credit availability.

In the Africa Maize Stress Project, new varieties are being developed for different agroecological zones. For the dry areas, 50 varieties were tested on 4 stations in Kenya in 1999. These materials are usually crosses or double crosses with the most popular improved open pollinated variety (OPV), Katumani Composite B (KCB), an early maturing variety to which all new materials are compared. Farmers were invited to all four stations to evaluate the material. Of the 50 varieties, 16 were retained for testing on-farm in 2000, using the mother and baby methodology. In 2001, four sets of varieties were tested in four different agroecological zones. At the end of the process, successful varieties will be proposed by the Kenya Agricultural Research Institute (KARI) for the National Performance Trials (NPTs) where they will be compared with varieties from other breeding programs and private seed companies. The best of these varieties will be selected for release.

Participatory Methods in Maize Breeding

New materials developed for drought stress and low nitrogen are tested and evaluated using a number of techniques.

Table 1. Constraints to maize production and related projects in East Africa.

Constraints†	Project
Drought, low nitrogen	AMS (Africa Maize Stress Project)
Pests such as stem borers and weevils	IRMA (Insect Resistant Maize for Africa Project)
Striga	Striga Project
Availability of quality seed at an affordable price	Seed Supply Systems Project
Nutrition	QPM (Quality Protein Maize Project)† Identified by farmers and scientists from national program

† Identified by farmers and scientists from national programs.

Classical breeding took place at four stations in the semi-arid region of Kenya under optimal and stress conditions, the latter to simulate farmers' conditions. In the last stage of the on-station research, 50 varieties were evaluated on 4 stations (Katumani, Kiboko, Kampi Ya Mawe, and Kitui) in 1999 in the dry areas of eastern Kenya. Methods and results of these breeding processes are discussed elsewhere (Bänziger et al. 2000).

To involve farmers more closely in variety development and selection, and therefore to increase the likelihood of adoption, one sublocation (smallest administrative unit in Kenya) close to each of the stations was selected for study. Participatory methods included participatory rural appraisals (PRAs), farmers' evaluation on-station, mother and baby trials, and community seed production (Table 2).

Initially, a PRA was conducted in each of the four sublocations, consisting of interviews with key informants, and group interviews of men and women on their maize variety preferences and selection criteria. In the second stage, the farmers or their representatives were invited to the station to evaluate the different test varieties, according to their own criteria. Breeders' and farmers' evaluations were then combined to select the varieties to be tested on-farm.

Table 2. Components of participatory methods used in maize breeding in eastern Kenya.

Component	Activities/data collected
Participatory rural appraisals	Farmers' variety preferences Maize selection criteria Perceived problems and pests of maize Interest in variety evaluation and testing
On-station:	Breeders' evaluation Farmers' evaluation
Mother trial	Breeders' evaluation (emphasis) Farmers' evaluation
Baby trials	Breeders' evaluation: yield data, timing of practices, and inputs, only Farmers' evaluation
Community seed production	Comparison of available varieties in mother and baby trials Experimentation with different modes of seed production: individual, community, and commercial

On-farm testing was conducted using the mother and baby trial approach developed in Zimbabwe. Sites for the mother or central trial are located in farmers' fields, schools, or on group farms that are sufficiently large enough to accommodate the trials. The mother trial is researcher or collaborator managed and the data collected are similar to those from on-station trials. This trial usually has a large number of entries grown in two-row plots. The purpose of the mother trial is to generate biophysical data for the breeders, as well as to allow farmers to evaluate all of the test varieties in one location. Subsets of varieties from the mother trial (usually 4-8 varieties) are given to farmers to grow on their farms under their management. These farmer-managed trials are known as baby trials. Their objective is to test the varieties under farmers' conditions and to obtain farmers' opinions on how the varieties perform under those conditions

The final stage of the maize development program is the production of hardy, high performing varieties that meet farmers' and consumers' preferences. However, as farmers often have difficulty in obtaining good quality seed of these new varieties at a reasonable price, CIMMYT has also initiated a project to produce and disseminate improved seed in Kenya. This project is currently in its first stage, which involves the evaluation of a large number of released or about-to-be-released varieties in many locations using mother-baby trials. In the subsequent stages of seed multiplication and distribution, farmers will be included through individual and community participatory methods. Since these stages have not yet begun, they are not discussed in this paper.

In the next sections, we discuss the participatory methods used—PRAs, on-station evaluation, and mother-baby trials—focussing on methodology, applications in the semi-arid zone of Kenya, and results.

Participatory Rural Appraisals

Methodology

To involve farmers in variety selection, PRAs were organized in one sublocation near each of the project sites. Firstly, a literature review was conducted at the district and divisional level for each site, and one sublocation was selected in each of the four sites: Kitanga (Katumani station), Mulaani (Kambi Ya Mawe), Ngumo (Kiboko), and Itoleka (Kitui). With assistance from extension and administration staff, farmers' meetings were organized in December 1999 and January 2000. Separate group interviews were conducted with men and women (except in Kitanga, where it was not convenient to do so), using the same open questionnaire. First the farmers were asked to record the varieties grown and the criteria used to select these varieties. Then they scored the varieties for each of the criteria on a scale of 1 (very bad) to five (very good). Next, farmers were asked to rank the constraints they faced in order of importance and score them on a scale of one (less important) to three (very important).

After the group discussions, farmers were asked if they were interested in evaluating the test varieties. In all four sites farmers were enthusiastic to participate. They expressed preference for evaluating the varieties twice: once around the tasseling stage and once at harvest. Dates were fixed for the farmers' visits on-station.

Results of the participatory rural appraisals

Only the major results of the PRAs are discussed here, i.e., farmers' maize variety selection criteria and farmers' interest in taking part in participatory breeding. The other elements are discussed in more detail by Bett et al. (2000). The different criteria used by farmers to select their maize varieties are presented in order of importance in Table 3. Farmers used a wide range of criteria, and the ranking differed substantially between sites and between groups. For ease of presentation, the table shows only those criteria mentioned by more than one group. Early maturity and yield are clearly the most important criteria as they are the only two mentioned by all groups in all sites.

To gain a general appreciation of the importance of a particular criterion in the area, we combined rankings from different groups and villages into a derived score, devised to represent the number of times a criterion ranks highly. For each group, the criterion receives a value inversely related to its rank, i.e., when a criterion is ranked first, it receives a derived score of 5, when it is ranked second, it receives a score of 4, and so on. The mean derived score (mds) is an indicator of the overall importance of the derived scores, and ranges from 0 (criterion was not mentioned) to 5 (criterion was ranked first by all groups). In Table 3, the criteria are ranked in order of decreasing mds.

According to this analysis, early maturity (mds = 3.4) is by far the most important criterion, followed by yield (mds = 1.9). A second important group (mds >1) comprises yield-related criteria such as cob size and grain size, other grain and cob characteristics, and drought tolerance. Criteria mentioned by at least three groups are pest and disease resistance, taste, and processing characteristics.

Considered less important are plant characteristics such as vigor, flint, and cob length, which were each mentioned twice. Characteristics mentioned only once include red cobs, tolerance to low soil fertility, tolerance to water logging, and cost and availability of seed. No important differences between men's and women's rankings were observed, except that, surprisingly, only men mentioned processing characteristics as a selection criterion.

Drought, soil fertility, and pests are the most important constraints to maize production mentioned by farmers. Weevils are by far the most important pest, followed by stem borers and termites (Table 4). Less important pests are army worms, porcupines, squirrels, and humans (theft in the field is a problem in many areas).

To evaluate the overall importance of each pest, we again calculate the mean derived score. Weevils are easily the most important pest. They were ranked first in all groups, resulting in a maximum score of 5. Stem borers and termites came a distinct second (mds = 2.8).

Table 3. Maize selection criteria used by farmers in four sublocations in eastern Kenya, ranked in order of importance.

	Rank							
	Kiboko		Kitui		Katumani	Kampi Ya Mawe		Derived
Criteria	Women	Men	Women	Men	W + M	Women	Men	score[†]
Early maturity	1	3	2	2	4	3	2	3.4
Yield	2	6	6	9	2	6	7	1.9
Cob size	6	5	4	1		2		1.7
Grain size	8	1	1					1.6
Other grain characteristics	5	2			3		9	1.3
Number of lines[‡]	3	11			1			1.3
Drought tolerance			3	3			6	1.0
Other cob characteristics			5	4			1	1.0
Pests and diseases	4	10		6		4	10	0.7
Taste			6	8	6		12	0.6
Plant characteristics						5	3	0.6
Processing		9		7			11	0.4

[†] Every time a criterion is ranked first it receives a score of 5, each second ranking scores 4, each third ranking scores 3, each fourth ranking scores 2, and each other ranking scores 1.

[‡] Only important for the variety Kinyanya, which means "eight rows" in Kamba, the local language.

Table 4. Ranking of importance of maize pests by farmers in four sublocations in eastern Kenya.

Common	Katumani		Kitui	Makueni	Kiboko	Mean derived
pests	Women	Men	Women	Women	Women	score†
Weevils	1	1	1	1	1	5.0
Stalkborer	3	3		2	2	2.8
Termites	2	2		3	3	2.8
Army worms	2		2			1.6
Porcupine			3	4		0.8
Squirrel	4	4	6	5		0.8
Man			5	10	5	0.6

† Every time a criterion is ranked first it receives a score of 5, each second ranking scores 4, each third ranking scores 3, each fourth ranking scores 2, and each other ranking scores 1.

The use of derived scores has some conceptual problems. The scores rely on a very subjective appreciation of the value of different ranks, and are therefore hard to generalize. Moreover, their use lacks a theoretical base from which to derive statistical tests. For this reason, it was decided that derived scores no longer be used. Instead farmers would be asked directly whether they perceived a variety to be very important (a score of 3), of medium importance (a score of 2), of low importance (a score of 1), or not important.

On-Station Evaluation

Methodology

The trials were conducted on four KARI research stations in the arid and semi-arid area, namely Katumani (Machakos district), Kambi Ya Mawe and Kiboko (Makueni district), and Kitui (Kitui district). Fifty new entries were tested on each station, usually double backcrosses of new materials with Katumani. The entries were established in small blocks, each in two rows of 3 m (0.75 m between rows, 0.5 m between hills).

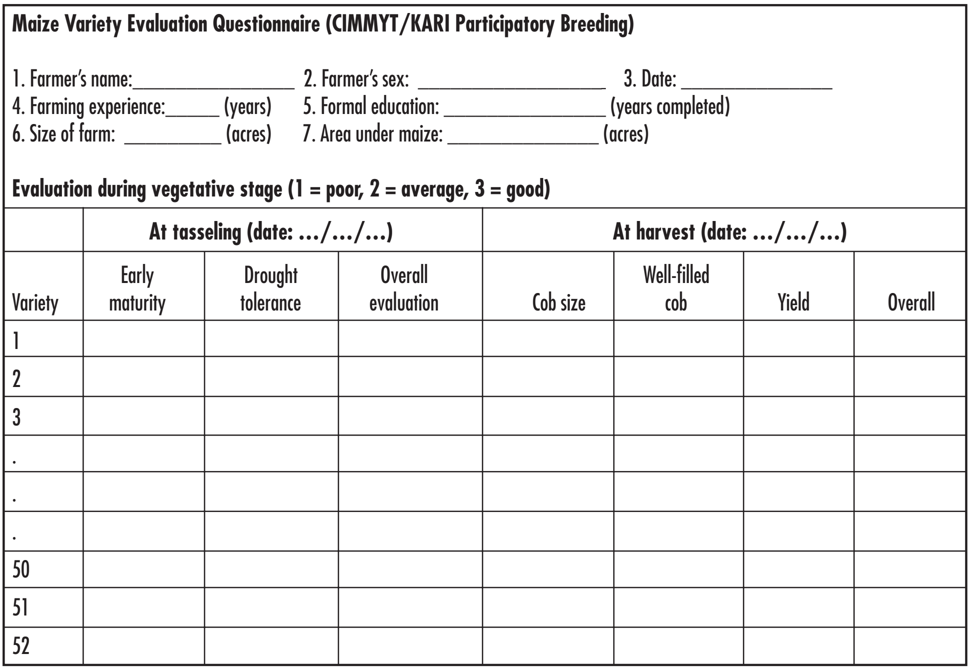
The statistical design used in the research is an alpha lattice, and the randomization is generated using Alpha software, produced by CIMMYT. A related software package, Fieldbook, generates forms in a spreadsheet for data entry (Barreti et al. 1997). The variables that can be integrated, depending on the breeders' strategy, are presented in Table 5. The software allows the calculation of a breeders' index, in which the breeder specifies which criteria (from the above variables) he or she finds important and weights each accordingly. The index has a 0-1 scale—the smaller the index, the better the variety is considered in terms of the traits included.

Farmers were invited to the station to evaluate the new varieties, assisted by researchers from KARI and CIMMYT, and Ministry of Agriculture extension staff. Initially farmers used three different colored ribbons to evaluate the varieties shortly before harvest. The green, blue, and red ribbons denoted varieties they appreciated, were moderately interested in, and did not like at all, respectively. The procedure went surprisingly well, and farmers were able to evaluate the 52 varieties in 10 to 30 minutes. Next farmers were given a maize variety evaluation form (Figure 1). For each variety, farmers were asked to make a cross against the qualities for which it was considered good. They were also asked for some personal information such as age and years of farming experience.

In later evaluations, changes were made to the procedure because farmers complained that they had to walk through the varieties twice. The ribbons were replaced by an extra column on the form in which farmers were asked to give a general evaluation (1 = poor, 2 = average, 3 = good) of each variety. Illiterate farmers were assisted by researchers and extension staff, or by students from a nearby secondary school. The evaluation procedure was conducted one or two weeks prior to harvesting and again at harvest. At the end of the evaluation, the participants were asked if they were interested in testing some of the varieties on their farms. They indicated that they had identified several worth testing, and that they would be able to test four or five varieties at a time.

Table 5. Variables used by breeders to evaluate drought tolerant maize varieties.

Variable (unit)	Description
Entry	A consecutive number from 1 to n (the total number of tested materials).
Pedigree	Parental background of the material.
Origin	
Index (0-1)	Breeders' index, derived from a weighted combination of the variables below.
Grain yield (t/ha)	
Ear aspect (1-5)	1 = good uniform cobs with the preferred texture of the area (i.e., flint for E. Africa); 5 = ugly cobs with undesirable texture for the area (i.e., dent).
Anthesis date (d)	Date when 50% of the plants shed pollen.
ASI (d)	Anthesis-silking interval.
Plant height (cm)	Measured from the ground to the first tassel branch on a representative plant.
Ear height (cm)	Measured from the ground to the insertion of the top ear on a representative plant (only for hybrids).
Root lodging (%)	% of plants with root lodging counted before harvest.
Stem lodging (%)	% of plants with stem lodging counted before harvest.
Ears/plant	Number of ears (having at least one grain) per plant at harvest.
Plant aspect (1-5)	1 = short plants with uniform and short ear placement; 5 = tall plants with high ear placement.
Ear rot score (%)	Scale of 1 (clean, no rot) to 5 (completely rotten).
GLS (1-5)	Score for gray leaf spot, a maize disease.
P. sorg (1-5)	Score for *Puccinnia sorghi* (rust, a maize disease).
E. turc (1-5)	Score for *Exserohilum turcicum* (leaf blight, a maize disease).
Leaf rolling score (1-5)	Measured twice before flowering when differences between genotypes are visible using following scale: 1 = unrolled; 2 = leaf rim starts to roll; 3 = leaf is shaped like a V; 4 = 60% dead leaf area; 5 = leaf is rolled like an onion.
Leaf senescence (1-10)	Three scores are taken after flowering when differences between genotypes are visible. Scoring is made on scale of 0 to 10, by dividing the percentage of dead leaf area by 10.
Grain texture (1-5)	Grain texture as preferred in the area.
Grain moisture (%)	Measured from a grain sample by an electronic moisture meter.



Maize Variety Evaluation Questionnaire (CIMMYT/KARI Participatory Breeding)

1. Farmer's name:________________ 2. Farmer's sex: ___________________ 3. Date: _______________
4. Farming experience:______ (years) 5. Formal education: ________________ (years completed)
6. Size of farm: __________ (acres) 7. Area under maize: _______________ (acres)

Evaluation during vegetative stage (1 = poor, 2 = average, 3 = good)

	At tasseling (date: .../.../...)			At harvest (date: .../.../...)			
Variety	Early maturity	Drought tolerance	Overall evaluation	Cob size	Well-filled cob	Yield	Overall
1							
2							
3							
.							
.							
.							
50							
51							
52							

Figure 1. Evaluation form for farmer evaluation of maize varieties on-farm.

Selection from on-station trials for the next cycle

The breeding program calculated the selection index for all varieties, resulting in a rank (Table 6). After a number of varieties had to be discarded because of undesirable traits, a final list was produced of varieties to be continued in the next cycle.

The farmers involved in variety evaluation were typical of the area. Most had completed at least primary education, with only small differences occurring between sites and between men and women. Farmers had, on average, 20 years of farming experience. Men had, on average, 19 acres of land and women had 14 acres. Average area under maize production ranged from 4.1 to 7.1 acres, with little difference between sites (for details see Bett et al. 2000). The farmers evaluated all varieties according to their own criteria. Their scores were averaged over the four sites, resulting in an average score, which allowed the varieties to be ranked and sorted (Table 6). No differences in rankings were found between farmers' preferences, or between farmer characteristics such as gender and farmer size.

Breeders' and farmers' selections in the study were very different. Breeders found 47 varieties to be better than the local check, whereas farmers only found 7. And of these 7, the breeders accepted only 3. Moreover, the breeders wanted to give 13 more varieties a chance in the next cycle, though they were rejected at this stage.

From Table 6 it is evident that there is very little relationship between the breeders' index and the farmers' evaluation. Statistically, the Pearson correlation coefficient (0.18) is not significantly different from zero (p = 0.287). Unfortunately, there was no discussion between breeders and farmers after the analysis to determine the cause of this lack of coherence. It is also unclear why the farmers' top three varieties were not retained for the next cycle.

Mother or Central Trial

Methodology

To increase farmer participation in the selection of drought tolerant varieties for eastern Kenya, the mother and baby trial approach was adopted for further variety evaluation (see De Groote et al. 2001a for a detailed description). The mother trials were planted on three KARI substations (Katumani, Kitui, and Kampi Ya Mawe) and on two farms, one

Table 6. Farmers' evaluation of new maize varieties at four research stations in arid and semi-arid eastern Kenya.

	Breeders' selection						Farmers' evaluation				
	Selection index		Yield (3 sites)		Overall		Site (score)				
Entry†	Value‡ (0-1)	Rank	t/ha	Rank	Score§	Rank	Kampi Ya Mawe	Katumani	Kiboko	Kitui	Final selection (breeder)
ECA-EE-43	0.51	27	2.7	35	1.64	1	1.55	1.75	1.57	1.69	
ECA-EE-40	0.40	12	3.2	12	1.62	2	1.44	1.79	1.4	1.85	
ECA-EE-51	0.81	51	2.3	46	1.51	3	1.12	1.75	1.6	1.58	
ECA-EE-45	0.44	17	2.8	29	1.48	4	1.51	1.71	1.17	1.54	Yes
ECA-EE-46	0.77	49	2.5	33	1.48	5	1.31	1.92	0.93	1.77	Yes
ECA-EE-29	0.45	18	3.2	8	1.45	6	1.78	1.42	1.7	0.88	Yes
ECA-EE-42	0.61	39	2.7	29	1.44	7	1.74	1.5	0.87	1.65	
ECA-EE-52	0.76	48	2.2	48	1.43	8	1.21	1.75	1.33	1.42	Yes
ECA-EE-38	0.38	9	3.0	14	1.43	9	1.55	1.46	0.77	1.92	Yes
ECA-EE-44	0.61	38	2.6	40	1.42	10	1.42	1.62	0.93	1.69	
ECA-EE-47	0.59	36	2.7	28	1.38	11	1.21	1.75	0.93	1.62	
ECA-EE-31	0.66	44	2.6	33	1.34	12	1.49	1.62	1.17	1.08	Yes
ECA-EE-30	0.73	47	2.9	22	1.31	13	1.34	1.46	1.50	0.92	
ECA-EE-23	0.67	46	2.7	34	1.30	14	1.67	1.42	0.93	1.19	
ECA-EE-4	0.56	31	3.0	20	1.30	15	1.19	1.26	1.43	1.31	
ECA-EE-37	0.65	42	2.6	35	1.30	16	1.21	1.54	0.67	1.77	
ECA-EE-18	0.43	14	3.0	25	1.28	17	0.84	1.83	1.3	1.15	Yes
ECA-EE-41	0.42	13	3.1	17	1.28	18	1.33	1.33	0.77	1.69	
ECA-EE-33	0.38	10	3.0	11	1.24	19	1.63	1.29	0.60	1.42	Yes
ECA-EE-28	0.59	35	2.7	31	1.23	20	1.58	1.58	0.80	0.96	
ECA-EE-35	0.58	34	2.6	28	1.23	21	1.25	1.5	0.63	1.54	
ECA-EE-21	0.48	23	2.9	19	1.22	22	0.84	1.5	1.40	1.15	Yes
ECA-EE-49	0.46	19	2.7	24	1.20	23	1.1	1.71	0.90	1.08	Yes
ECA-EE-3	0.49	24	2.7	28	1.18	24	0.97	1.17	1.53	1.04	
ECA-EE-20	0.78	50	2.5	32	1.16	25	1.07	1.33	0.73	1.50	

Note: Only the top 25 varieties according to farmers' preferences are shown.

† Code for the variety tested: ECA-EE = East and Central Africa, Extra Early Variety.

‡ The selection index is a weighted index of different desired traits—a lower value indicates good performance for these traits.

§ Scores are evaluations by farmers made on a scale of 0-2 (0 = poor, 1 = average, 2 = good).

belonging to Makindu Children's Home (near Kiboko) and the other to Emali Primary School. The mother trial consisted of 16 new varieties and 2 local checks, Katumani Composite B (KCB) and Dryland Composite 1 (DLC1). Each variety was grown in two rows of 5 m. Two seeds were sown per hill, which were later thinned to one plant per hill. The spacing between rows was 75 cm and between plants was 25 cm, giving a population of about 54,000 plants/ha. These were replicated three times, and grown under both optimal and non-fertilized conditions. Optimal conditions comprised recommended fertilizer rates and other cultural practices. Both farmers and breeders evaluated the varieties at the mother sites using the same methodology as for the on-station evaluation.

Farmers were invited to visit the mother trials at late silking and at maturity. At each evaluation an effort was made to encourage equal participation by male and female farmers. Of the 101 participants, 57 were women. At the beginning of the evaluation, farmers gathered to discuss what they thought were the important criteria for selecting a given variety at a particular development stage. These criteria were ranked and the top three were used for the evaluation. The criteria were translated into the local Kikamba dialect for ease of understanding.

The farmers were taken around the whole trial to get an initial feel for the project. Afterwards, they were divided into groups of five and taken around by the technical staff. Farmers who could not understand the labeling of the trials or English were assisted. Each criterion was scored from 1 to 5 (1 = very poor, 2 = poor, 3 = average, 4 = good, and 5 = very good). Farmers were also asked to give an overall score to each variety, i.e., their opinion of how good each variety was compared to all others. This was not simply an average of each variety's overall score, but a judgment of the variety in its entirety as a typical plant type. Farmers were also asked to select their best three varieties, and these were summarized at the end of the evaluation. Farmers were also requested to make comments on the whole exercise, including suggestions for improvement.

Results of breeders' evaluation

Table 7 shows a typical breeders' evaluation of a multilocational trial. The mean yield from the four sites indicates the extent to which the new varieties surpassed the two local checks. It can be seen that the best varieties yielded a ton more than the better local check (KCB). At the Makindu site, there were significant differences between the local checks and the new varieties, while at Emali, KCB was similar to the new varieties other than ECA-EE-45. At this site, the second local check (DLC1) was comparable in yield only to ECA-EE-45. At Kampi Ya Mawe, DLC1 produced very high yields, outperforming seven varieties including KCB. This was the best site for DLC1, where it produced the highest yield among the four sites. At Kitui, 12 varieties yielded higher than KCB, while the remaining varieties did not show significant yield differences. However, all varieties yielded significantly more than DLC1 at this site.

The selection index was used so that other important aspects were included in the selection. According to this index, the best five varieties were ECA-EE-21, -29, -40, -33 and 16 (Table 7). This ranking mostly reflects yield and ear characteristics.

Results of farmers' evaluation

The farmers who participated in the trial were typical of the area. On average, they had 18 years of farming experience, with the highest being 30 years. Farmers had, on average, 7 years of formal education, which is most likely higher than the regional average. Average farm size was 14 acres, but this was substantially smaller in Emali (7.5 acres) and Makindu (8 acres), the two sites closest to town. Of the sites characterized by smaller farm sizes (Emali, Ithookwe, and Katumani), about half of the farms were planted to maize. In areas with larger farm sizes, the proportion of farms under maize was lower. In general, few differences were observed between male and female farmers. While men's farms were, on average, larger (16.6 acres) than women's (15.1 acres), women's farms were larger in 3 of the 5 sites.

At silking, farmers evaluated the varieties for earliness, and made an overall assessment of each. At this stage, variety ECA-EE-13 was the best across all sites with a mean score of 3.36. It is

Table 7. Breeders' evaluation of 18 new maize varieties compared with two local checks (KCB and DLC1) in mother trials on four research stations in eastern Kenya.

		Grain yield under low N conditions						
Pedigree	Selection index (0-1)	Average	Emali (t/ha)	Kampi YaMawe (t/ha)	Kitui (t/ha)	Makindu (t/ha)	Ear aspect	Anthesis date
ECA-EE-21	0.36	5.10	4.10	6.20	6.20	3.90	3.60	55.00
ECA-EE-29	0.40	5.00	4.20	6.70	5.90	3.30	3.60	54.00
ECA-EE-40	0.40	5.00	4.20	5.80	6.30	3.70	3.50	54.00
ECA-EE-33	0.41	5.10	4.20	6.60	5.60	4.00	3.60	54.00
ECA-EE-16	0.41	4.70	4.10	6.10	5.20	3.40	3.70	55.00
ECA-EE-31	0.44	4.90	3.80	6.30	5.40	4.20	3.40	54.00
ECA-EE-46	0.46	4.90	4.10	6.30	5.70	3.50	3.70	53.00
ECA-EE-8	0.47	4.70	4.00	5.40	5.60	3.90	3.60	54.00
ECA-EE-6	0.48	5.00	4.60	6.00	5.60	3.60	3.50	54.00
ECA-EE-9	0.50	4.70	4.20	5.80	4.90	3.70	3.70	54.00
ECA-EE-13	0.51	5.00	3.70	7.00	5.50	3.90	3.30	53.00
Local check 1	0.53	4.10	4.20	5.20	4.40	2.60	3.00	54.00
ECA-EE-18	0.54	4.60	3.70	5.80	5.10	3.60	4.10	54.00
ECA-EE-38	0.56	4.60	4.20	6.00	4.50	3.40	3.80	54.00
ECA-EE-34	0.59	4.70	3.70	5.80	5.40	3.90	3.50	54.00
ECA-EE-36	0.62	5.00	4.10	6.20	5.90	4.00	3.70	55.00
Local check 2	0.89	3.90	3.30	6.10	4.00	2.30	2.70	52.00
ECA-EE-45	0.93	4.40	3.40	5.30	5.50	3.20	3.30	54.00
Mean			4.00	6.00	5.40	3.60	3.50	54.00
LSD			0.70	1.10	1.00	0.80		
CV			10.80	10.60	11.00	12.90		
Min			3.30	5.20	4.00	2.30	2.70	52.00
Max			4.60	7.00	6.30	4.20	4.10	55.30

important to note that two varieties, ECA-EE-13 and ECA-EE-6, were considered by farmers to be earlier than the local check (DLC1). In addition, another four varieties (-16, -33, -34, -46) were considered to be earlier than KCB. The remaining varieties matured later than the two local checks. In the overall assessment at silking, variety ECA-EE-13 was still considered to be the best. Variety EAC-EE-31, perceived by farmers to be later than the two local checks, scored more highly than KCB, while variety ECA-EE-6, ranked second for earliness, was considered comparable to KCB. Furthermore, variety ECA-EE-33, which was considered comparable to DLC1 for earliness, was ranked higher in the overall assessment. Overall, six varieties were considered to be better than or comparable to KCB.

For the evaluation at harvest, farmers developed the following criteria for all sites: cob size, cob fill (grain filling), and yield. Farmers were also asked to make an overall assessment of each variety independently. Based on cob size, seven varieties were considered to be better than or comparable to KCB. Variety ECA-EE-31 was perceived to have the best cob size, while DLC1 was considered least desirable for this criterion. Two varieties considered to be better for earliness (ECA-EE-6 and -18), were rejected, while four other entries (ECA-EE-8, -21, -25, -31, and -36) were ranked better than or comparable to KCB for cob size but not for earliness.

The next criterion was grain filling. Again variety ECA-EE-31 scored highly. Using this criterion, nine varieties were considered better than or comparable to KCB, while variety DLC1 was ranked the lowest. As observed earlier, the ranking of the varieties changed, with some remaining superior to KCB, while others that had not been considered better for cob size were ranked lower for grain filling, e.g., ECA-EE-25.

The top five varieties according to yield were the same as those ranked by yield components (cob size, grain filling). However, of the 11 varieties considered better than or comparable to KCB, only 7 were perceived to be higher yielding. Again, all varieties were considered to be higher yielding than DLC1.

In the overall assessment (Table 8), the top four varieties for cob size, grain filling, and yield were retained. About 10 varieties were considered to be better or comparable to KCB. This clearly indicates that farmers effectively use yield components for evaluating yield, and attach appropriate weightings to each component.

Comparison of breeders' and farmers' evaluations

It is important to know how closely breeders' and farmers' evaluations correspond and how they are linked in order to predict future adoption of new technologies. In the on-station evaluation, there was no correlation between farmers' and breeders' selections. In the mother trials, there was a small, though not significant, correlation of 20% (Pearson correlation coefficient = 0.2, p = 0.068).

To analyze the relationship between the farmers' and breeders' order of preference, we mapped each evaluated variety in a two-dimensional diagram, where the horizontal axis represents the farmers' rank and the vertical axis represents the breeders' rank (Table 9). The table shows, for example, how

Table 8. Overall assessment by farmers at harvest of 18 new maize varieties compared with two local checks (KCB and DLC1) in mother trials on four research stations in eastern Kenya.

Entry		Total mean score		Emali		Kitui		Kampi Ya Mawe		Katumani		Makindu	
number	Pedigree	Total	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank
8	ECA-EE-31	3.70	1	2.87	12	4.08	1	3.45	8	4.25	6	3.87	1
9	ECA-EE-33	3.66	2	3.24	3	3.69	3	4.03	1	4.21	8	3.11	7
7	ECA-EE-21	3.59	3	3.03	7	3.50	6	3.83	3	4.39	2	3.21	6
10	ECA-EE-34	3.48	4	2.95	10	3.61	4	3.53	6	3.93	11	3.37	3
16	ECA-EE-36	3.43	5	3.29	2	2.61	13	3.41	9	4.21	9	3.61	2
12	ECA-EE-46	3.40	6	3.16	5	3.39	7	3.33	12	4.39	3	2.71	14
5	ECA-EE-16	3.35	7	3.21	4	2.94	11	3.28	14	4.26	4	3.05	9
13	ECA-EE-25	3.32	8	2.95	11	3.89	2	3.55	5	3.81	12	2.42	15
4	ECA-EE-13	3.27	9	3.08	6	2.86	12	3.34	11	3.75	13	3.34	4
3	ECA-EE-9	3.27	10	2.82	14	3.17	10	3.22	16	4.25	7	2.89	12
17	KCB	3.22	11	3.03	9	3.33	9	3.47	7	4.26	5	2.00	17
2	ECA-EE-8	3.16	12	2.82	13	3.53	5	3.31	13	3.07	18	3.08	8
6	ECA-EE-18	3.13	13	2.71	16	2.19	16	3.28	15	4.57	1	2.89	11
1	ECA-EE-6	3.09	14	3.34	1	2.50	14	3.40	10	3.48	15	2.74	13
15	ECA-EE-38	3.05	15	2.71	17	2.42	15	3.67	4	3.43	16	3.00	10
11	ECA-EE-45	3.04	16	3.03	8	2.14	17	3.21	17	3.57	14	3.24	5
14	ECA-EE-29	2.99	17	2.76	15	1.97	18	3.95	2	3.96	10	2.29	16
18	DLC1	2.83	18	2.63	18	3.33	8	3.19	18	3.07	17	1.95	18

variety EE-EAC-31 was selected first by farmers, but only sixth by breeders. Varieties acceptable to both groups can be found in the top left corner of the table. There are clearly three varieties appreciated by both: EE-EAC-31, -33, and -21. Two more acceptable, but not outstanding, varieties are EE-EAC-16 and -46.

We can also express the correspondence of breeders' and farmers' preferences by calculating the overlap between the two. For example, farmers and breeders have no common variety in their first two choices. Comparing their first 3 choices, however, they have 1 variety in common (ECA-EE-21), or a 33% overlap. In their first 4 choices, there is a 50% overlap. Figure 2 shows the evolution of the overlap. It can be seen from the figure that among the first 10 choices, the overlap is still only 60%.

Satellite or Baby Trials

Methodology

A village close to each of the four mother sites was selected for each satellite or baby trial (Table 10). Farmers from these villages were asked to select 10 participants from among themselves to test a subset of varieties from the mother trials. Each of the 10 randomly selected farmers were given 250 g of each of 4 new varieties, ensuring that each variety was grown by at least 2 farmers. The 4 varieties were chosen randomly using an alpha lattice design with blocks of 4, and each of these blocks of 4 varieties was given to 1 or more farmers. Furthermore, farmers were given a local check (either KBC or DLC1) and were also asked to plant an additional plot of their own local maize. The latter variety differed between sites and farms. The checks were replicated twice, so most farmers had 8 plots, of which 4 were new varieties. Farmers were requested to manage all plots equally. Overall, the varieties tested were the same as those grown in the mother trial, and the same randomization process was used, i.e., an alpha lattice design.

Unfortunately the sample size of the trial was greatly reduced due to poor rains and resulting crop failure on all farms at Kitui, on 6 farms at Katumani, and on 3 farms at Kampi Ya Mawe and Kiboko. Moreover, only yield data from Kampi

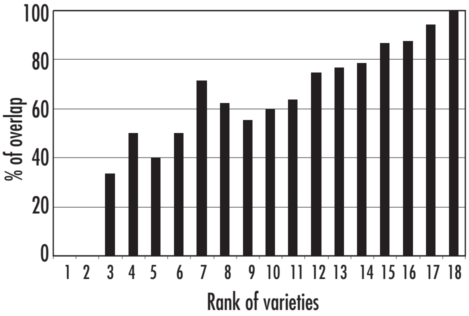


Figure 2. Comparing farmers' and breeders' variety evaluations.

Table 9. Ranking of maize varieties according to farmers and breeders.

Breeders' rank	Farmers' rank																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1			21															
2																	29	
3																		
4		33																
5							16											
6	31																	
7						46												
8												8						
9														6				
10										9								
11									13									
12											KCB							
13													18					
14															38			
15				34														
16					36													
17																		DLC1
18																45		

Note: Numbers in the table refer to varieties tested in the mother trial; KCB and DLC1 are two local checks. The variety that ranked third in breeders' assessment did not even appear among the first 18 varieties ranked by farmers.

Table 10. Sample size of the mother-baby trial established in eastern Kenya.

Location of mother trials	Location of baby trials	Initial no. of farmers	No. of farmers at sites where harvest was measured	No. of plots evaluated at tasseling	No. of plots evaluated at harvest
Katumani	Kitanga	10	4	24	64
Kampi Ya Mawe	Mulaani	10	7	56	88
Kiboko	Nguumo	10	10	52	0
Kitui	Kitui	10	0	0	0
Total		40	21	132	152

Ya Mawe and Kiboko were obtained—a transport problem stopped field workers visiting all other sites at harvest. Yields of 85 plots were measured, and the adjusted yield was calculated as:

Adjusted yield = cob weight*0.8*(100-moisture content)/((100-15)*area)*10000

where:

adjusted yield = estimated yield (kg/ha), adjusted for grain coefficient (grain weight/cob weight, assumed to be 80%) and moisture content (measured from three samples) and then calibrated to 15%

cob weight = weight of the harvested cobs (kg)

moisture content = mean moisture content of three samples
area = area harvested (m^2)

Farmers evaluated the varieties at tasseling and at harvest. At tasseling, farmers scored each variety for earliness; at harvest varieties were scored for cob size, cob fill (grain filling), and yield. Again, scores were given on a scale from 1 (very poor) to 5 (very good). At both tasseling and harvest, farmers were asked to give an overall assessment of each variety, using the same scale. This overall assessment, not to be confused with the mean score of the other criteria, can be considered to be similar to the selection index developed by the breeder during data analysis. The index takes into consideration more plant aspects at the given development stage than the criteria of importance. To analyze the overall evaluation and assess the weightings of individual criteria, regression analysis was performed. The overall evaluation was used as the dependent variable and the scores for the individual criteria were used as independent variables.

Results of breeders' evaluation of the baby trials

Because of the small sample size, no extensive analysis was performed other than a yield comparison (Table 11). Five varieties outperformed the best local check (KCB), although only one variety achieved this in more than one site. Since the sample sizes are very small (1-3 repetitions per variety), statistical analysis is not really possible.

Results of farmers' evaluation of the baby trials

Farmers' evaluation of earliness and overall value at tasseling are shown in Table 12. Results indicate that 13 varieties scored higher, on average, than the best local check, indicating an appreciation of breeding for extra early material. The relationship between the scores for overall evaluation and early maturity is highly significant, but the low correlation coefficient (0.271, $p = 0.002$) indicates that many other factors play a role here.

At harvest, 11 varieties outscored the best local check overall, and 7 scored higher than the local check in more than one location (Table 13). This indicates that factors other than yield are important to farmers.

The three varieties (ECA-EE-29, -45, -46) preferred by farmers and retained by breeders from the on-station evaluation also performed well in the farmers'

Table 11. Maize yield (adjusted for water content) from baby trials at two sites in eastern Kenya.

	Kiboko			Kampi Ya Mawe			Total		
Variety	Mean	Std dev	N	Mean	Std dev	N	Mean	Std dev	N
9	5,890	528	2			0	5,890	528	2
29	5,499		1				5,499		1
26	4,917		1				4,917		1
38	4,829	1,110	2			0	4,829	1,110	2
43	4,770		1				4,770		1
18	5,074	1,243	2	4,053		1	4,734	1,058	3
49	4,571	670	2	4,985		1	4,709	530	3
37	5,249	641	2	3,106		1	4,534	1,318	3
KCB	4,392	1,494	15				4,392	1,494	15
13	4,364		1				4,364		1
31	4,169	1,340	2				4,169	1,340	2
33	4,820	1,666	2	2,721		1	4,120	1,690	3
36	4,704	1,364	2	2,923		1	4,110	1,410	3
42	5,181	829	2	1,480		1	3,947	2,216	3
45	4,764	536	2	1,700		1	3,742	1,809	3
6	3,669		1				3,669		1
LOCAL	4,234	614	14	2,207	1,247	6	3,626	1,255	20
21	4,529	969	2	1,091		1	3,383	2,100	3
46	4,580	1,070	2	667		1	3,276	2,382	3
8	5,658		1	309		1	2,984	3,782	2
16	3,246		1	911		1	2,078	1,651	2
DLC1				1,365	313	6	1,365	313	6
34	1,118		1	417		1	768	496	2
Total	4,510	1,114	61	1,908	1,251	24	3,775	1,644	85

Table 12. Farmers' evaluation of baby trials at tasseling in three sites in eastern Kenya.

	Katumani		Kiboko		Kampi ya Mawe		Overall	
Variety	Overall evaluation at tasseling	N	Overall evalation at tasseling	N	Overall evaluation at tasseling	N	Overall evaluation at tasseling	N
46	5.00	1	4.00	1	5.00	2	4.67	4
29	4.00	1	5.00	2			4.50	3
42			5.00	1	4.00	2	4.50	3
45			5.00	1	4.00	0	4.50	1
49	5.00	1	3.00	1	5.00	2	4.33	4
9	4.00	2	4.00	1	3.50	2	3.83	5
36	3.00	1	3.50	2	5.00	1	3.83	4
26			3.50	2			3.50	2
37			3.00	1	4.00	2	3.50	3
16	3.00	1	4.00	3	3.00	2	3.33	6
21	2.00	1	5.00	1	3.00	2	3.33	4
13	2.5	2	4.00	2			3.25	4
18			4.5.0	2	2.00	2	3.25	4
LOCAL	2.75	4	2.86	14	3.92	12	3.18	30
33			.	0	3.00	1	3.00	1
38	2.50	2	.	0	3.00	2	2.75	4
KCB	2.50	4	2.93	14			2.72	18
8			3.00	2	2.00	2	2.50	4
31	3.00	2	2.00	1			2.50	3
DLC1					2.29	14	2.29	14
34	1.00	1	3.00	2	2.50	2	2.17	5
6	1.00	1	2.50	2			1.75	3
43			1.00	1			1.00	1
Mean	2.88	24	3.27	56	3.25	52	3.13	132

overall evaluation (Table 12), although the latter two yielded lower than KBC. In the mother trial, however, only ECA-EE-46 scored higher in the overall evaluation at harvest. Further analysis by variety is not really feasible, however, due to the very small sample size.

On-farm evaluation most likely best represents farmers' conditions and predicts future adoption of technologies, therefore it is important that it is analyzed and understood. The overall evaluation of individual varieties can be seen as the farmers' selection index. To analyze this index, the overall score at harvest was regressed on the score of the individual criteria, i.e., yield, well filled cob (grain filling), cob size, and vigor (Table 14). Yield had the highest coefficient (0.5), followed by vigor (0.2) and grain filling (0.2). Cob size was not significantly different from zero. The results show that the model predicts a large amount of the variation (R2 = 62%), but also that the individual criteria do not capture a number of elements. This highlights the importance of including

Table 14. Regression analysis of the overall evaluation of varieties.

Independent variables	Coefficient	Std error
Yield	0.526	0.098***
Vigor	0.191	0.088*
Well filled cob	0.189	0.083*
Cob size	0.034	0.099
Constant	0.191	0.310
R2	0.624	
Std error of estimate	0.659	
N	84.000	

Note: * significant at the 5% level; ** significant at the 1% level; *** significant at the 0.1% level.

Table 13. Farmers' evaluation of baby trials at harvesting in two sites in eastern Kenya.

	Kiboko		Kampi Ya Mawe		Total	
Variety	Overall evaluation at harvest	N	Overall evaluation at harvest	N	Overall evaluation at harvest	N
36	5.0	2	5.0	1	5.0	3
45	4.0	2	5.0	1	4.5	2
37	3.5	2	5.0	1	4.3	3
49	3.5	2	5.0	1	4.3	3
13	4.0	1			4.0	1
29	4.0	1			4.0	1
31	4.0	2			4.0	2
43	4.0	1			4.0	1
18	3.5	2	4.0	1	3.8	3
33	3.5	2	4.0	1	3.8	3
46	3.5	2	4.0	1	3.8	3
LOCAL	2.9	16	4.3	6	3.6	22
9	3.5	2	.	0	3.5	2
21	3.0	2	4.0	1	3.5	3
42	3.5	2	3.0	1	3.3	3
KCB	3.2	16	.	0	3.2	16
6	3.0	1			3.0	1
8	5.0	1	1.0	1	3.0	2
26	3.0	1			3.0	1
34	4.0	1	2.0	1	d3.0	2
DLC1				2.8	6	
2.8	6					
16	2.0	1	3.0	1	2.5	2
38	2.5	2	.	0	2.5	2
Mean	3.3	64	3.7	24	3.5	88

an overall evaluation score. The individual coefficients represent how much the overall evaluation increases with an increase in score of an individual criterion. For example, when a variety's score for yield increases by 1, its overall score increases by 0.5; when its score for vigor increases by 1, its overall score increases by 0.2, all other factors being equal. Thus, the coefficients can be seen as the weights in a selection index. The non-significance of the large cob criterion comes as a surprise after the group discussions, but it does make sense since larger cobs do not necessarily equate to more or better quality food for a household.

These results show how a farmers' selection index can be approximated and then compared to the breeders' index to make the breeders' index more responsive to farmers' needs.

Discussion and Conclusion

Two major points need to be discussed. Firstly, did we identify new and promising varieties through this process? And secondly, did the participatory methods contribute to the process? If so, how can they be improved?

In answer to the first question, the results are ambiguous. Although some promising materials were identified, the methodology resulted in a large body of often inconsistent data that are difficult to interpret. Few of the varieties consistently outscored the local checks. Two major factors make it difficult to identify promising material: the poor correlation between farmers' and breeders' evaluations, and the high variability in the farmers' evaluation and in performance in farmers' fields.

On the second point, the participatory methods clearly showed how classical breeding has problems responding to farmers' variety preferences. So far, the two approaches have not converged to form a method compatible to both. Scientists like to control for many factors so that they can accurately state that, under their very controlled circumstances, a limited number of traits have improved. A problem here is that these highly controlled circumstances are not often representative of farmers' conditions, and the limited number of traits might not represent farmers' preferences. This is highlighted by the very poor correlation between farmers' and breeders' evaluations in the study. The exercise, however, provides very useful insights into how to bring the two evaluations together by improving their respective methodologies.

In the study, the main classical breeding tool—the breeders' index—does not seem to represent farmers' preferences well. It could be improved through changes to the functional form, as well as to the variables included and/or the weights attached to different variables. The linear function is not always an appropriate form since some materials should be rejected when they do not pass certain levels. This, however, can be achieved through a multiplicative or multistage index, and there should be no problem in programming this into the existing software. Secondly, the selection process should be transparent so that the pathway of choice can be retraced, analyzed, and improved. Thirdly, an effort should be made to try alternative formulations, including different variables with different weights, to try to better match farmers' evaluations. Finally, and most importantly, breeders should have more frequent discussions with farmers to compare their respective preferences.

From the farmer's perspective, the PRA methodology and the evaluation used on-station and in the mother trials are well developed, although some modifications could be made. The baby trials, however, need serious work. The basic system of asking farmers to define their selection criteria, and to score new varieties on a scale of 1 to 5, was very convenient for data collection but cumbersome for analysis. The criteria collected during the PRAs need to be more concordant so that farmers' responses can be classified into a number of categories. This would simplify the analysis of farmers' evaluations of new varieties. The high variability associated with farmers' evaluation on-station or in mother trials needs to be managed by increasing the number of farmers (preferably to more than 50) participating. Sufficient resources need to be made available to make a rapid analysis possible. The people conducting the analysis will need to have appropriate software installed and have sufficient training, since proper statistical analysis differs from conventional methods. The analysis should then be included in the selection of varieties for the next cycle, and an attempt be made to include the farmers in that selection. An index to combine the evaluations at tasseling and at harvest still needs to be developed.

The baby trials need some serious thought. In the study the data are not very useful: the variance is high, the sample size is small, and a lot of data were lost due to bad weather and poor organization. The experience indicates that enough resources have to be made available to allow for regular site visits to assure the quality and quantity of data collected, and rapid data entry and analysis. The process could be improved by increasing the amount of data collected by the farmer. This could be achieved through well structured questionnaires, combined with proper training to enable farmers to fill them unassisted. A simplified yield measurement taken by farmers should also be tested. Farmers could include more evaluation criteria in the baby trials than is possible in the mother trials, and these data would be very useful for improving the selection index.

Finally, the experience revealed that farmers are eager to participate in selecting new varieties. The methodology still needs work, but it clearly demonstrates the potential for bringing together breeders' and farmers' selection indices. The collaboration between breeders, farmers, and social scientists shows promise for improving the selection procedure by taking into account farmers' preferences early in the process.

Acknowledgements

We would like to thank Wilson Muasia, Charles Bett, Josephine Malelu, and Daniel Mulwa from KARI for their help with the field work; the farmers from Kampi Ya Mawe, Katumani, Kiboko, and Kitui for their collaboration in evaluating new varieties; and Rick Coe, biometrician with the International Centre for Agroforestry (ICRAF), for his helpful insights. This research was made possible through financial support from the United Nations Development Fund to the African Maize Stress Project, and the Syngenta Foundation to the Insect Resistant Maize for Africa Project.

References

Ashby, J. 1991. Adopters and adapters: the participation of farmers in on-farm research. In R. Tripp (ed.), *Planned change in farming systems: progress in on-farm research.* New York: John Wiley and Sons.

Bänziger, M., G.O. Edmeades, D. Beck, and M. Bellon. 2000. *Breeding for drought and nitrogen stress tolerance in maize: from theory to practice.* Mexico, D.F.: CIMMYT. 68 pp.

Barreti, H.J., G.O. Edmeades, S.C. Chapman, and J. Crossa. 1997. The alpha lattice design in plant breeding and agronomy: Generation and analysis. In G.O. Edmeades, M. Banziger, H.R. Mickelson, and C.B. Peña-Valdivia (eds.), *Developing drought- and low N-tolerant maize.* Mexico, D.F.: CIMMYT.

Bett, C., H. de Groote, A. Diallo, W. Muasya, and K. Njoroge. 2000. Participatory plant breeding for drought resistant maize varieties in eastern Kenya. Presented at the 7th KARI Biennial Scientific Conference, 13-17 November 2000, Nairobi, Kenya.

Biggs, S.D. 1989. *Resource-poor farmer participation in research: A synthesis of experiences from nine national agricultural research systems.* Comparative Study Paper No. 3. The Hague: ISNAR-OFCOR.

De Groote, H., D. Friesen, M. Siambi, and A. Diallo. 2001a. Participatory plant breeding guidelines for the AMS on-farm trials for 2001 in East Africa. Mimeo.

De Groote, H., J.O. Okuro, C. Bett, L. Mose, M. Odendo, and E. Wekesa. 2001b. Assessing the demand for insect resistant maize varieties in Kenya, by combining Participatory Rural Appraisals and Geographic Information Systems. Presented at the International Symposium on Participatory Plant Breeding and Participatory Plant Genetic Resource Management: An Exchange of Experiences, 7-10 May 2001, Bouake, Côte d'Ivoire.

Eyzaguirre, P., and M. Iwanaga (eds.). 1996. *Participatory plant breeding.* Proceedings of a workshop on participatory plant breeding. Rome: IPGRI.

Kamara, A., T. Defoer, and H. de Groote. 1996. Selection of new varieties through participatory research, the case of corn in South Mali. *Tropicultura* (September 1996) 14 (3):100-105.

Sperling, L., M.E. Loevinsohn, and B. Ntabomvura. 1993. Rethinking the farmers' role in plant breeding: local bean experts and on-station selection in Rwanda. *Experimental Agriculture* 29:509-519.

Sumberg, J., and C. Okali. 1997. *Farmers' experiments, creating local knowledge.* Boulder/London: Lynne Rienner Publishers.

Discussion Summary

The discussion following the presentation centered on two themes: trial design and farmers' evaluation of the trials. The design of the baby trials included controls and, though these trials were evaluated in the same way as the mother trial, the control variety only performed better than other varieties in the mother trials. This could be due to variations in the management of the baby trials. Mother trials are intensively managed and this may influence variety performance. Baby trials provide contrasting management conditions, i.e., low management intensity. Mother trials can be used for testing agronomic practices, but may not be as useful for breeding purposes. Measuring yields in baby trials is difficult, prompting the question of whether farmers' yield scores could be used instead of yield measurements. This point was discussed in several presentations and merits further investigation. It was noted that yields from mother trials and baby trials may not correlate well, and this may be an important reason to have both, i.e., to provide different information.

In terms of farmers' evaluation of baby trials, the point was made that the farmer participants should ideally be selected randomly. As this is usually not possible, there is a risk of bias in the evaluation. While there is little that we can do about this, we need to be aware of the problem. A question was asked on how farmers in the case study ranked the varieties. The answer was that a group process was used to reach an agreement on the order of importance. It was then pointed out that group discussions are useful but may impose consensus when there is not necessarily agreement among farmers, and that there may be high variation in scores among farmers because they select for different purposes. Some farmers, for example, select for storage, while others select for selling properties, and these differences may influence the variability of scores. It was pointed out that farmers can be more precise than breeders when scoring varieties. It was also noted that care should be taken when using an index that combines farmers' opinions on different traits. Such an index can be misleading because it makes assumptions that one trait may compensate for another, when not all traits have the same importance to farmers. An alternative scoring method asks farmers to accept or reject varieties, as breeders do. The point was raised that it would be better if farmers evaluated trials 3-4 weeks before harvest as well as at harvest, but this may be location specific. For example, among maize farmers in Mexico, evaluations before harvest are not considered useful.

Participatory Plant Breeding: Setting Breeding Goals and Choosing Parents for On-Farm Conservation

BHUWON STHAPIT, KRISHNA JOSHI, SANJAY GYAWALI, ANIL SUBEDI, KEDAR SHRESTHA, PASUPATI CHAUDHARY, RAM RANA, DEEPAK RIJAL, MADHUSUDHAN UPADHAYA, AND DEVRA JARVIS

Abstract

Participatory plant breeding (PPB) is a strategy for strengthening on-farm conservation by encouraging farmers to continue searching, selecting, and managing local crop populations. This paper describes the preliminary results of PPB processes that address the difficulties of setting breeding goals and choosing parents in diverse research areas. Participatory methods were developed to gain an understanding of the extent and distribution of local crop diversity and the processes by which farmers maintain crop diversity on their farms. The study addresses farmers' needs, conservation issues, and diverse users' issues in the process of choosing parents and selecting for variable populations.

Introduction

On-farm conservation can generally be described as the process by which farmers maintain, manage, and improve the traditional crop cultivars that they have developed (Jarvis et al. 2000). Participatory plant breeding (PPB) is a strategy used to strengthen on-farm conservation by encouraging farmers to continue to search, select, and manage local seed supply systems. The PPB process aims to consolidate the farmer's role in setting breeding goals and selecting diverse genetic materials. In so doing it offers skills and opportunities for farmers to search for new diversity and to select and exchange variable populations that match their local preferences and needs (Sthapit and Jarvis 1999).

Participatory plant breeding has been advocated as a strategy for maintaining or enhancing the level of genetic diversity on-farm (Brush 2000) because it can be used to breed divergent cultivars for subtly different environments and for diverse end uses. Also PPB can add value to traditional landraces that would otherwise be lost from the system (Sthapit et al. 2001). On the other hand, the success of PPB products may stem from the addition of just a few major genes, corresponding, for example, to pest resistance or plant height (Brown and Young 1999). It is also commonly assumed that the replacement of farmers' unique varieties with modern cultivars may affect the intraspecies genetic diversity by reducing the genotypic diversity and/or the allelic diversity within the cultivated crop. Both participatory varietal selection (PVS) and PPB may have negative impacts on landrace diversity, because both methods are intended to change the structure of the local crop population to make it more productive and useful to farmers (Joshi et al. 2000b). It has been argued that successful PPB products may swamp a significant fraction of local diversity, leading to a short term gain in productivity, loss of local unimproved populations, and hence increased future vulnerability (Brown and Young 1999). The creation of new varietal diversity by PPB, which has already been successfully used in difficult environments (Sthapit et al. 1996), is a powerful method that can be employed in high potential production systems (Witcombe et al. 2001). However, the impacts of PPB on the extent and distribution of local crop diversity are still poorly monitored and documented.

The PPB program in Nepal was designed to investigate: (1) whether farmers' cultivars are being conserved, (2) whether the PPB process encourages farmers to maintain processes that allow crops to evolve and change over time, (3) whether PPB contributes to the enhancement and conservation of traditional varieties in situ and provides community benefits, and (4) whether landraces can be conserved by improving preferred traits, hence making them more desirable in relation to modern cultivars.

The purpose of this paper is to describe participatory methods that build local capacity to understand the extent and distribution of local crop diversity and the processes by which farmers maintain diversity on their farms. The study addresses farmers' needs, conservation issues, and diverse users' issues in the process of choosing parents and selecting variable populations.

The studies form part of the value-adding activities of the global project "Strengthening the Scientific Basis of In Situ Conservation of Agricultural Biodiversity in Nepal". The study was initiated in 1998 and has been jointly implemented by the Nepal Agricultural Research Council (NARC), Local Initiatives for Biodiversity Research and Development (LI-BIRD), and other local institutions.

Study Sites

Studies are underway in two contrasting villages in Nepal to test if PPB will meet both development and conservation goals. Begnas village is situated in the Pokhara valley in Kaski district (600-1400 meters above sea level), the central mountain ecosystem of Nepal. Pokhara valley, in the western hills of Nepal, is known for high quality rice diversity and harbors more than 69 rice cultivars. Kochorwa is situated in Bara district (54-100 masl) and lies within the fertile Indo-Gangetic plain (100-200 masl) near the southern border with Indian. This eco-site maintains about 35 rice cultivars. Production potentials are high, and farmers have adequate access to inputs and technologies.

The Participatory Plant Breeding Process

A series of brainstorming sessions were organized in Nepal to integrate the PPB approach into the national crop breeding strategy. A primary stakeholder meeting was held in April 1998 to develop a PPB process for on-farm conservation and to agree upon the roles of formal and informal institutions (Joshi et al. 2000a).

There are three essential components to PPB: (1) one parent is a locally adapted cultivar, (2) selection is decentralized in the target environment, and (3) farmers participate in the plant breeding process. Ceccarelli et al. (1996) argued that the most complete decentralization could be achieved when farmers participate in the breeding process in their own fields. Often plant breeders seek functional collaboration with farmers to obtain quick results. Community participation is central to on-farm conservation of agricultural biodiversity. The local capacity to enhance genetic diversity (i.e., to search, select, and exchange) is a key element of sustainable agriculture.

A broad PPB process was used in Nepal to develop local farmers' capacity to assess, select, and share local crop diversity, and to build sustainable partnerships between farming communities and researchers (Sthapit et al. 2000). The process can be divided into the following steps:

- Locate rice agroecosystems and identify interested communities.
- Organize "diversity fairs" for locating rice diversity, and collect germplasm and local knowledge.
- Understand and monitor local crop diversity through a community biodiversity register.
- Analyze options for adding benefits.
- Set breeding goals for PPB and parent selection with representative community participation.
- Agree on roles of farmers and researchers involved in the breeding process.
- Select for diversity.
- Strengthen farmers' seed systems for rapid diffusion.

In 1997 a "diversity fair" was organized at each site to raise community awareness of crop genetic resources, to locate genetic diversity and its custodians, and to promote the value of landrace diversity in the context of local food culture, market forces, and socioeconomic and agroecological considerations. The materials collected from the fairs were grown in farmers' fields as "diversity blocks". The blocks were used to assess field performance and to analyze preferred and undesirable traits through participation from male and female farmers, representing all socioeconomic strata. The diversity block could also be used to select for appropriate parent plants and as a seed source for crossing programs.

Appreciating local crop diversity

Understanding the value of local crop diversity is a key step prior to the initiation of participatory goal setting. After the diversity fair, local communities were motivated to keep a community biodiversity register (CBR). The CBR is a record, kept in a register book by community members, of all landraces in a community, including information on their custodians, passport data, and use value (Rijal et al. 2001). It aims to monitor the level of diversity held by farmers in a community over time, as well as the number of households and area covered by each cultivar within the community. Local communities are encouraged to use CBR information to understand factors influencing farmers' decision making on dynamic changes in local crop diversity and to develop their own on-farm conservation strategies (Sthapit et al. 1999).

We found that farming communities maintained a substantial level of rice diversity at the community level in both the Begnas middle hill eco-site (69 cultivars; 91% local, farmer-named cultivars) and the highly accessible Bara site (53 cultivars; 62% local landraces). More local than modern cultivars were maintained at the household level in both sites. The area of rice landraces in Bara was 17% of the total and in Kaski was 73%. We found that some landraces competed strongly with modern cultivars in certain niches and that these landraces could be promoted in similar areas without further improvement.

Setting breeding goals

The dilemma prior to initiating PPB was deciding which rice varieties to include: all varieties, only those with high market demand, those maintained by only a few farmers in small areas, or those grown by many farmers in larger areas? Since no one knew which group was more important from a conservation perspective, and it was not possible to include all of the traditional cultivars, the team decided to divide the varieties into categories according to certain criteria. There was no literature available to guide this research, so the team elected to divide the landraces into four groups, based on the average planted area and the average number of households growing each landrace (Figure 1). This design ensured that at least one variety from each group, representing a different use value, would be included in the crossing program. The process used is listed as follows:

- List existing local diversity.
- Categorize local crop diversity into four cells according to the planted area of each cultivar and the number of households growing each cultivar.
- Classify local crop diversity by use.
- Perform assessment analysis for preferred and unwanted traits using preference ranking and paired matrix ranking.
- Analyze potential benefits of local diversity and threats of genetic erosion.

Choosing parents

A key element of the PPB program is the choice of the local landrace parent. A consultative participatory process used to assess farmers' needs and the project goal highlighted the conflict of interest between choosing parents and setting breeding goals. The following broad

Figure 1. Conceptual framework for understanding the extent and distribution of local crop diversity (according to farmers' use values) at the community level.

Area planted

Number of farmers

Large planted area, many farmers (common)	Small planted area, many farmers
Large planted area, few farmers	Small planted area, few farmers (rare)

Source: Sthapit et al. (2000).

criteria were used in the parent selection process to avoid bias from influential groups in the community:

- Community interest and priorities
- Program goal and objectives
- Technical feasibility
- Availability of genetic variability
- Consumer interest

Focus group discussions (FGDs), attended by male and female farmers, were held at Kachorwa and Begnas sites with the objective of identifying the landrace parent for PPB. Both male and female farmers from all socioeconomic strata were consulted in the discussions. Participatory approaches were used to select at least one landrace from each of the four cells. The lists of farmer-named cultivars were analyzed using preference matrix ranking (Guerrero et al. 1993) for preferred and undesirable traits in order to identify traits needing improvement. The next step was to identify the best landraces from each of the four cells using preference matrix ranking. During the discussion, the preferred traits of each landrace parent were documented, while the traits needing improvement were thoroughly analyzed using paired matrix-ranking methods. The landrace parents preferred by farmers for inclusion in the PPB program were short-listed. A relatively large number was selected from the cell representing landraces grown by many farmers over a large area. The exotic parent was then identified by looking at the traits to be improved, the adaptability of the variety in the area, and other farmer preferred traits. Finally, cross combinations for Bara and Kaski sites were finalized by the team.

Distribution of diversity and selection of parents

One of the key steps to success in the PPB project is to understand the distribution of diversity according to farmers' use values in the selection of parents. Participatory rural appraisals (PRAs) were carried out to (1) understand key factors affecting farmers' decisions to maintain local cultivars, including information on market development, and (2) ascertain key limiting factors to production systems reliant on local crop cultivars. As described above, to understand the relative importance of specific landraces, farmers' varieties were grouped into four broad categories using baseline survey data and CBR (Figure 1). This approach could be carried out at the species or variety level.

This type of broad distribution analysis helps us to understand, for example, why some landraces are grown in a small area by many farmers, while others are grown in a small area by few farmers. It is very important for plant breeders to understand such distribution patterns and the underlying decisions made by farmers before starting to design a PPB program to promote on-farm conservation.

The above information can also be accurately collected using participatory methods. Figure 2 presents the traditional rice cultivars, with their perceived desirable and undesirable traits, considered for PPB by the Begnas farming community. Three landraces (Jetho budho, Ekle, and Mansara) from the commonly grown group were selected to address the needs of many households and to ensure that benefits would flow to a diverse range of people. Two landraces from the rare group (Birmaphool and Sano gurdi) were chosen. Aanga was chosen for its unique medicinal value and Gurdi was selected to represent good coarse varieties in each cell.

Results indicated that few culturally important cultivars (e.g., Anadi in Kaski and Sathi in Bara) are grown by many households in small patches for household use. These landraces could be conserved per se; however, the population size of Anadi has increased in recent years with the linking of the project to markets and with sales being made under its brand name in departmental stores.

Table 1 shows the traditional varieties selected in the study area in 1998 using the described participatory process. The exotic parent or local modern variety was selected by plant breeders according to farmers' preferences.

Table 1. Cross combinations selected for participatory plant breeding programs in Kaski and Bara sites, Nepal, 1998.

Begnas, Kaski (600-1400 masl)	Kochorwa, Bara (80-90 masl)
Pusabasmati/**Jetho budho**	**Mansara**/Rampur masuli
Ekle/Khumal-4	**Lajhi**/IR62161-22-1-2-1-1
Biramphul/Himali	**Lalka Basmati**/IR59606-119-3
Thulo gurdi/NR10286	**Dhudhisaro**/BG1442
Sano gurdi/NR10286	**Lajhi**/IR62161-22-1-2-1-1
Mansara/Khumal-4	**Dudhisaro**/BG1442
Aanga/NR10291	

Local seed system

Baniya et al. (2001) found that 96% of farming households in Begnas and Kochorwa villages are dependent upon informal seed sources. Seed flow occurs through farmers' social networks (Subedi et al. 2001). These communities manage their rich rice diversity through bartering, gifts, borrowing seed or seedlings, and purchase (Subedi et al. 2001).

Social networks of germplasm research and exchange

The study found that certain farmers maintain more diversity than other farmers in the community. These farmers are active in searching and selecting for new diversity and in maintaining and sharing it within and outside the

Figure 2. The extent and distribution of local rice diversity in Begnas, Kaski, Nepal, 1998.

Number of farmers	Area planted (large)	(small)
(many)	**Jetho budho** High market demand for quality traits such as softness, aroma, and taste, but low yielding, and prone to lodging and neck blast disease. **Ekle** Stable yield, good straw, good milling recovery, but high water and input requirements, late maturing, and prone to storage pests. **Mansara** Adapted to rainfed, poor land and low input conditions, early maturing, but poor taste, yield, and milling recovery.	**Gurdi** Good taste, good straw yield, good milling recovery, adapted to irrigated conditions, but low yielding.
(few)	**Aanga** Medicinal value, adapted to very poor soils, but poor yield, taste, and shattering. **Anadi** Valued for sticky rice, but poor milling recovery, high input requirement, and low yielding.	**Biramphool** High quality with good aroma, softness, and medicinal value, but extremely low yielding. **Sano gurdi** Adapted to *tari* rainfed conditions, tolerates shade, good taste, and good milling recovery, but low yielding.

community. These farmers are described as the "nodal" farmers of the community (Subedi et al. 2001). They are perceived to be more knowledgeable than other farmers in relation to seed and production environment matters and as being "diversity minded". Nodal farmers mainly belong to the resource-endowed farmer group, which is characterized by ownership of larger landholdings, more land parcels, and more livestock than less well-off farmer groups (Rana et al. 2000). Moreover, resource endowed farmers generally have a higher education level and participate more frequently in the local market (Gauchan et al. 2001). Some nodal farmers are women. Nodal farmers are spatially distributed within the community; together they can act as a conservation farming network and their farms are used as "field gene banks". They can be involved very effectively in community biodiversity registration and linked to development opportunities. Nodal farmers can enhance farmer to farmer dissemination of genetic materials, be effective as resource persons for farmer to farmer training, and provide information on local crop diversity. Their expertise and knowledge can be effectively utilized in the development of training and extension materials on local cultivars, and they can also be involved in public awareness on agrobiodiversity.

Who should test materials?

The 13 segregating materials selected for PPB were distributed to nodal farmers and a few other interested farmers for growing F_2 or F_3 bulk of their choice. Field performance of the different populations was assessed during farm walks by researchers and farmers. In Begnas village, F_3 segregants of Mansara (locally adapted to marginal, drought prone, rainfed, and low input conditions) and Khumal-4 (good quality modern cultivar with parentage of local variety Pokhereli masino) showed promising results in the upper areas (1000-1300 masl). The breeding goal was to incorporate the good eating quality and yield potential of Khumal-4 into Mansara without losing the latter's adaptive traits. Three distinct population types were selected jointly by farmers and researchers and these were further advanced in farmers' fields. The spread of M x K-4 bulks is being monitored, and the selection history is being documented. The cross between Pusa basmati and Jetho budho (local high quality rice with high market demand) also did well in *Khola-ko-chewn* environments (650-690 masl). In Bara district, six large segregating populations were evaluated in farmers' fields, of which farmers selected the populations of Lajhi and IR62161-22-1-2-1-1 for lodging resistance and post harvest traits.

Discussion and Conclusion

The participatory method used to divide local crop diversity into four categories is useful for understanding the extent and distribution of farmers' classification of local crop diversity. This is a simple method for teaching farmers about the concept of local common and rare traditional cultivars. It is also useful for analyzing options for introducing other benefits such as increased community awareness and motivation, seed networks, new germplasm, market links, and PPB.

Social networks highlight the effectiveness of decentralized selection and informal seed systems. Nodal farmers' expertise in selecting and maintaining genetic materials can be effectively used in PPB. Furthermore, capacity building of nodal farmers in PPB may enhance crop diversity on a large scale.

Participatory methods such as biodiversity fairs, diversity blocks, and community biodiversity registers raise local awareness, strengthen local capacity for understanding the value of local crop diversity, and strengthen the roles of farmers and the informal sector in the process of local crop development. Such community participation helps the benefits of on-farm conservation to reach poor and biodiversity-based livelihoods.

It unlikely that a single product of PPB will outcompete a large proportion of local crop diversity. The diverse agroecology existing in an area, combined with farmers' preferences for different cultivars with varying characteristics, should help to maintain biodiversity.

Acknowledgements

This work is part of the International Plant Genetic Resource Institute's global project "Strengthening the Scientific Basis of In Situ Conservation of Agricultural Biodiversity On-Farm: Nepal Country Component", which is funded by the Directorate-General for International Cooperation, the Netherlands.

References

Baniya, B.K., R.K. Tiwari, D. Pandey, J. Bajracharya, A. Mudwari, B.R. Sthapit, and D.I. Jarvis. 2001. Rice seed supply systems in Terai and high hills of Nepal. Presented at the National Workshop of In Situ Crop Conservation, 24-26 April 2001, Lumle, Nepal.

Brown, A.H., and A. Young. 2000. Participatory plant breeding—does it achieve or does it compromise genetic conservation? Presented at a participatory plant breeding meeting, Mexico.

Brush, S. (ed.). 2000. *Gene in field: on-farm conservation of crop diversity.* IPGRI/IDRC. N.p.: Lewis Publishers.

Ceccarelli, S., S. Grando, and R.H. Booth. 1996. International breeding programmes and resources and resource-poor farmers: Crop improvement in difficult environments. In P. Eyzaguirre and M. Iwanaga (eds.), *Proceedings of a workshop on participatory plant breeding.* Wageningen, The Netherlands: IDRC/FAO/CGN/IPGRI. Pp. 99-116.

Gauchan, D., and P. Chaudhary. 2001. A participatory marketing system research approach to analysing market based incentives and disincentives: a case of rice, Bara, Nepal. Presented at the National Workshop of In Situ Crop Conservation, 24-26 April 2001, Lumle, Nepal.

Guerrero, M.P., J.A. Ashby, and T. Gracia. 1993. Farmer evaluations of technology: preference ranking. Instruction Unit No. 2. Cali, Columbia: CIAT.

Jarvis, D., B. Sthapit, and L. Sears (eds.). 2000. *Conserving agricultural biodiversity in situ: a scientific basis for sustainable agriculture.* Proceedings of a workshop. Rome, Italy: IPGRI.

Joshi, K.D., B.R. Sthapit, and J.R. Witcombe .2000a. How narrowly adapted are the products of decentralised breeding? The spread of rice varieties from participatory plant breeding programme in Nepal. *Euphytica* 122(3):575-588.

Joshi, K.D., D.K. Rijal, R.B. Rana, S.P. Khatiwada, P. Chaudhary, K.P. Shrestha, A. Mudwori, A. Subedi, and B.R. Sthapit. 2000b. Adding benefits: PPB, seed networks and grassroots strengthening. In D. Jarvis, B. Sthapit, and L. Sears (eds.), *Conserving agricultural biodiversity in situ: a scientific basis for sustainable agriculture.* Proceedings of a workshop. Rome, Italy: IPGRI.

Rijal, A., M.P. Subedi, R.B. Upadhyaya, P. Rana, P.R. Chaudhary, R.K. Tiwari, B.R. Tiwari, B. Sthapit, and D. Gauchan. 2001. Community biodiversity register: developing community based database for genetic resources and local knowledge in Nepal. Presented at the National Workshop of In Situ Crop Conservation, 24-26 April 2001, Lumle, Nepal.

Sthapit, B., and D. Jarvis. 1999. PPB for on-farm conservation. ILEIA Newsletter 15(3/4):40-41.

Sthapit, B.R., K.D. Joshi, and J.R. Witcombe. 1996. Farmer participatory crops improvement, III. Participatory plant breeding, a case study for rice in Nepal. *Experimental Agriculture* 32:479-496.

Sthapit, B.R., K.D. Joshi, R.B. Rana, and M.P. Upadhaya. 2001. Enhancing biodiversity and production through PPB. In *An exchange of experiences from South and South East Asia. Proceedings of the International Symposium on Participatory Plant Breeding and Participatory Plant Genetic Resource Enhancement.* CGIAR-PRGA.

Sthapit, B.R., P. Sajise, and D. Jarvis. 2000. Strengthening scientific basis of in situ conservation on-farm: Learning experiences from Nepal and Vietnam. In *Links between cultures and biodiversity. Proceedings of the Cultures and Biodiversity Congress 2000.* Pp. 338-361.

Subedi, A., P. Chaudhary, B.K. Baniya, R.B. Rana, D.K. Rijal, R.K. Tiwari, and B.R. Sthapit. 2001. Who maintains crop genetic diversity and how? Implications for on-farm conservation and participatory plant breeding. Presented at the National Workshop of In situ Crop Conservation, 24-26 April 2001, Lumle, Nepal.

Witcombe, J.R., K.D. Joshi, R.B. Rana, and D.S. Virk. 2001. Increasing genetic diversity by participatory variety selection in high potential production systems in Nepal and India. *Euphytica* 122:575-588.

Discussion Summary

The discussion dealt mainly with a number of questions regarding the presentation. The first question was who decides on the improvement program? The answer was the community, which always looks at new diversity, but keeps old landraces to maintain the gene pool. Currently a PhD student is looking at the genetic variability in the study area with molecular markers, and this should help to identify good landraces to be incorporated into the participatory plant breeding program. A question was raised on the danger of losing vulnerable varieties. It was pointed out that common varieties are most vulnerable to new introductions. The number of different varieties grown by many farmers in small areas seems to be a good predictor of their potential to be maintained. There was a question on how long the farmers in the study have been using their varieties. While there is no exact information to answer this, there is data on how frequently they change materials. The author pointed out that, regardless of varietal value, there is a tendency towards conservation among farmers. In terms of the farmer network analysis, it was noted that there was less movement of materials between some households. The explanation given was that seed exchange among farmers is affected by caste and neighbor relationships. Also it was noted that there seems to be a higher variability in upland areas because of the many agroecological zones and uses of varieties in these areas—factors that promote genetic variability.

A Quantitative Method for Classifying Farmers Using Socioeconomic Variables

JOSE CROSSA, MAURICIO R. BELLON, AND JORGE FRANCO

Abstract

Small-scale farmers in developing countries are heterogeneous in the resources they control, in the constraints they face, and, hence, in the crop varieties they require. This poses the challenge to participatory plant breeding (PPB) of identifying farmers that experience comparable conditions and needs, and therefore require similar varieties. Practitioners of PPB need methods for classifying farmers into homogenous groups with similar variety demands. This paper presents a statistical method for classifying individuals—farming households in this example—into homogenous but distinct groups. The method allows the use of different types of variables, provides a systematic approach to decide the number of groups present in the data, and assigns a probability that an individual belongs to a group. The method assumes that data have been collected from a sample of farmers in the target socioeconomic or agroecological environments. In the example presented, the method is used to divide a random sample of small-scale maize farmers in Mexico into homogenous groups.

Introduction

Small-scale farmers in developing countries are not always homogenous, even within a community. Ownership of resources such as land, labor, and capital is not equal between households, nor is the sharing of knowledge and information. Consequently, goals and constraints differ between farming households. Variability—spatial and temporal—is another fact of life for every farmer and his/her household. Soils and topography vary and seasons change. All of these factors influence the type of crop varieties that farmers want and need. In failing to recognize differences between farming households, breeders may overestimate the potential impact of their varieties because they may end up working with a smaller and possibly unrepresentative subset of the farmers they hope to serve, or they may have a very static view of farmers' resources and/or constraints.

Recognizing and addressing the heterogeneity of conditions faced by small-scale farmers and their different needs are key to developing appropriate germplasm through participatory plant breeding (PPB). This creates the challenge, however, of identifying farmers that experience comparable conditions and needs and, hence, require similar varieties. In other words, the challenge is to identify the recommendation domains for new germplasm.

Practitioners of PPB need methods for classifying farmers into homogenous groups with similar variety demands. There are a number of different approaches to achieving this. A typology can be created based on some *a priori* knowledge of the conditions and needs of farmers. This requires that the researcher decides which variables are important, how they fit together, and what cut-off values in each variable should be used to divide each type. Another approach is to rely on the farmers' own views of their differences and to let them define the groups. A wealth ranking exercise is one example. Key informants in a community identify farmer groups based on wealth, as well as the characteristics defining each group. Then the informants classify farmers into each group (see Bellon, this proceedings). A third approach is to apply geometric or statistical clustering methods. Usually these methods rely on continuous or categorical data, requiring that measurements of farmers' characteristics (attributes) are available. A problem with this method, as with the typology, is that the researcher has to decide which variables should be included. A common mistake is to use all available variables, which makes the interpretation of each group difficult.

In this paper we present a powerful method for classifying "individuals" (e.g., farmers, households, etc.) into homogenous but distinct groups based on their attributes. The method allows the use of both continuous and categorical variables. This is important because certain attributes are measured as continuous variables (e.g., landholdings), while others are discrete (e.g., gender) or include multiple categories (e.g., membership in different farmer associations). It provides a standardized process for deciding how many groups exist, and, as well as assigning each "farmer" to a group, it gives the probability that each farmer belongs to that group. The method, however, does not solve the problem of deciding which variables to include. This can be addressed by using variables associated with the characteristics identified by a wealth ranking exercise or other participatory methodologies which identify farmer types in an area. This approach relies on variables that are relevant to farmers, but builds on the structure of the actual data to identify the number of groups and to classify each farmer according to these groups. This, in turn, should simplify the interpretation of the meaning of the groups.

The method presented here is illustrated using data from the on-farm conservation project described in Bellon (this proceedings), and involves the use of variables associated with characteristics identified in a wealth ranking exercise. It should be pointed out, however, that the link between the groups formed using the method and the farmers' variety requirements is not automatic. It depends on the variables chosen for the classification and their relationship with the demand for crop traits, and requires the association between the groups and the crop traits to be tested. This, however, is beyond the scope of this paper. For more information see Bellon (this proceedings).

The paper begins with a brief introduction to classification methods. This is followed by a description of mixture models and the method used to form the initial groups for the classification. The location model, the homogenous conditional mixture model, the independent mixture model, and the canonical variate analysis are then described. Finally, an example is presented of farmer classification using a sequential clustering strategy with categorical and continuous socioeconomic variables.

A Brief Introduction to Classification Methods

Classification methods are grouped in two main categories: cluster analysis and discriminant analysis. Discriminant analysis allocates new individuals to previously defined groups by finding a mathematical function, based on a linear combination of the original variables, that minimizes the chance of misclassification. Cluster analysis is the partition of a heterogeneous population into homogeneous subpopulations using hierarchical or nonhierarchical methods. In hierarchical methods, the individuals or groups are organized in a hierarchy or “tree” and are fused, one by one, to other individuals or groups with the most similar patterns for all attributes. These methods can be used to form a fixed number of groups by truncating the tree at a fixed level. The nonhierarchical clustering methods guess the number of groups and then use a certain method or algorithm to improve the previous classification by optimizing a particular objective function.

The agglomerative hierarchical clustering method starts with an original dissimilarity (distance or dispersion) matrix among all the individuals, and fuses the two individuals which have the smallest dissimilarity between them to form a group with two members. Next, the group-individual dissimilarity between this new group and the remaining individuals is calculated. This set of dissimilarities is added to the matrix of dissimilarities among the remaining individuals to form a new dissimilarity matrix that is one row and column smaller than the original. A new fusion procedure is carried out, and, when two or more groups are present, group-group dissimilarities must be computed. The procedure ends when all of the individuals are in one group. The method used for calculating the group-individual and group-group dissimilarity is called clustering strategy. A number of agglomerative clustering strategies have been proposed such as the single linkage (or nearest neighbor), the maximum linkage (or furthest neighbor), the unweighted pair group arithmetic averaging (UPGMA), and the Ward method (incremental sum of squares). The Ward method uses the within-groups sum of squares as the objective function. It fuses the two groups that increase the within-group sum of squares the least and increase the among-group sum of squares the most, over all of the possible functions.

Classification methods require a measure of association among individuals calculated from measurements of a number of attributes of each individual. The effective use of classification methods requires an understanding of the properties of the forms and type of data collected as well as of the measures of association. Data form consists of a two-way table of n individuals (farmers) and p attributes (or variables), and the type of attribute can be continuous or categorical. Categorical data may be binary or nominal. The two-way table of n individuals and p variables can have one type (only categorical or only continuous) or a mixture of types (continuous and categorical). Classification based on all available information on the individuals is much more reliable and trustworthy than that based on only some attributes.

Franco et al. (1998; 1999) proposed a two-stage sequential strategy for classifying and studying genetic resources. In the first stage, initial groups are formed using an agglomerative hierarchical clustering method such as Ward or UPGMA and includes all of the continuous and categorical variables. Then a statistical method such as the location model (LM) or the modified location model (MLM) for improving the initial groups is used. These statistical models allow the use of continuous and categorical variables. This two-stage sequential clustering strategy is usually called Ward-MLM or UPGMA-MLM.

The objective of the study was to use the two-stage sequential Ward-MLM strategy for classifying 240 farmers using 9 categorical and 22 continuous socioeconomic attributes.

Mixture Models

Agglomerative hierarchical clustering techniques use proximity (distance) matrices for finding groups of objects, and are basically exploratory (or geometrical) methods that do not use any probabilistic density models. Mixture models, on the other hand, cluster the data using a particular probability density function without the need to explicitly use any proximity measurement.

To illustrate the use of mixture models, consider a random sample of farmers, including samples taken from regions where two notoriously different farmer types based on income level can be found, i.e., low income (L) and high income (H) farmers. The attribute, number of hectares, was measured for each randomly selected farmer. Since we know that low income farmers have fewer hectares than high income farmers, the probability density function (pdf) for number of hectares should take this into consideration. If we assume that the attribute is normally distributed with specific mean (μ) and variance (σ^2), and that one farmer from the sample can come from either the low or high income subpopulation with probability or $(1-\alpha)$, respectively, then the pdf for any farmer can be written as:

$$\text{pdf of number of hectares} = (\alpha)[N(\mu_{L,}\ \alpha_L)] + (1-\alpha)[N(\mu_{H,}\ \alpha_H)] \quad (1)$$

In equation 1 there are five parameters to be estimated: a, the proportion of low income farmers in the population ($1-\alpha$ is the proportion of high income farmers in the population); and $\mu_{L,}$, σ_L, $\mu_{H,}$, and σ_H, which correspond to the means (μ) and the standard deviations (σ) of the number of hectares for low income (L) and high income (H) farmers, respectively. N means normal distribution. The function represented by equation 1 is commonly called "finite mixture density", and in this particular case the distribution of the number of hectares variable results from a weighted mixture of two underlying normal distributions, where the α and $(1\text{-}\alpha)$ are called mixing proportions $[\alpha + (1\text{-}\alpha) = 1]$. Note that the parameter α is the relative frequency of the underlying distribution $N(\mu_L, \sigma_L)$, and $(1\text{-}\alpha)$ is the relative frequency of the other underlying normal distribution $N(\mu_{LH}, \sigma_H)$. Assuming there are $i = 1, 2, \ldots, g$ farmer groups, equation 1 can be extended to:

$$\text{pdf of number of hectares} = \sum_{i=1}^{g} (\alpha_i)[N(\mu_i\, s_i)] \qquad (2)$$

The model in equation 2 can be extended to multivariate data in such a way that the univariate normal distribution is replaced by the multivariate normal (MVN) density with mean vectors μ_i and variance-covariance (dispersion) matrix Σ_i, such that:

$$\text{pdf of number of hectares} = \sum_{i=1}^{n} (\alpha_i)[MVN(\mu_i\, \Sigma_i)] \qquad (3)$$

$$\text{where} = \sum_{i=1}^{n} \alpha_i = 1$$

Parameter estimation of the mixture models—the maximum likelihood estimation

The parameters of the distribution under a mixture model are estimated by maximum likelihood (ML) procedures, in which case equation 3 gives the likelihood function of the unknown parameters as a function of the observed values. In ML we consider a $p \times 1$ random vector of observations $x' = x_1, x_2, \ldots, x_n$, and ask about the vector of parameters Θ (the true proportion, α; the mean, μ; and the dispersion matrix, Σ, under the normal probability density function). The maximum likelihood estimate of an unknown parameter is the linear combination of the observations that maximizes the likelihood of the parameter given the observations. The ML estimates of the unknown parameters Θ , $\hat{\boldsymbol{\Theta}} = (\hat{\alpha}, \hat{\boldsymbol{\mu}}, \hat{\Sigma})$, is the value of $\Theta = (\alpha\alpha, \mu\Sigma)$ corresponding to the maximum of $l(\Theta | x)$. It is usually easier to find the maximum of the logarithm of the maximum likelihood function $L(\Theta | x) = \ln[l(\Theta | x)]$ than to find it from the function itself, because of the mathematical properties of the logarithmic function. For many ML estimation problems, a simple solution for the ML estimator can be obtained by solving the equation $\partial[L(\Theta | x)] / \partial(\Theta) = 0$.

Forming the initial groups

The question of how to form the *a priori* subpopulations used in the mixture models has been examined by Franco et al. (1997a; 1997b) in the context of genetic resource conservation. The initial groups are the starting points of the iterative process by which a solution that corresponds to a global (or local) maximum of the likelihood function is found.

Franco et al. (1997a) compared the performance of several hierarchical and nonhierarchical classification strategies for forming the initial groups and then compared the application of the mixture of normal distributions to these initial groups. The authors found that the initial groups formed using the Ward clustering method with Gower's distance (so that all continuous and discrete attributes can be used in the classification) recovered a good percentage of the true groups. Furthermore, the authors applied the mixture models to those initial groups and found a great deal of reallocation of individuals among groups and, thus, the formation of more compact, homogeneous, separate, and well characterized groups. They called this a sequential clustering strategy, where the initial groups are formed using the Ward method, and the mixture normal distribution is applied to the groups to improve the classification.

From a statistical perspective the Ward method seems better than other hierarchical clustering strategies. This is because it has an objective function to minimize the within-group sum of variability and therefore to maximize the among-group variability; thus, it gives a natural connection to the analysis of variance. Furthermore, the Ward method is appropriate for multi-normal data distribution. One problem with this sequential clustering strategy, however, is that while all variables, continuous and discrete, are used to form the initial groups using the Ward method, only the continuous variables can be used in the mixture models.

Location Model

In practice there is a mixture of attribute types. Some attributes used for classifying individuals are continuous and others are categorical. A distance measurement such as Gower's distance can be used for mixed variable types, thus any hierarchical clustering algorithm such as the Ward method could be employed for clustering the individuals and forming the initial groups. While the mixture of normal distributions is appropriate for modeling only continuous variables, the binomial, trinomial, or multinomial distributions should be the natural probability density functions for modeling categorical variables. Therefore, for modeling mixed types of variables, a combination of these probability density functions should be the most appropriate modeling strategy.

A joint distribution of a set of continuous and categorical variables can be written as the product of the marginal distributions of some, and the conditional distribution of others, given the values of the selected variables. For example, for two variables A (continuous) and B (categorical), the joint probability is $P(A \cap B) = P(A|B)P(B)$. Olkin and Tate (1961) proposed a model where the joint distribution of continuous and categorical variables [P(A, B)] is the marginal distribution of the categorical variables [P(B)] multiplied by the conditional distribution of the continuous variables, given the categorical variables [$P(A|B)$]. This is known as the location model (LM) (Krzanowski 1988). The categorical variables are arranged in a contingency table where the table categories follow a multinomial distribution and the continuous variables are assumed to follow a multivariate normal (MVN) distribution. However, the parameters of these MVN distributions depend on their location in the contingency table of the categorical variable.

Homogenous Conditional Mixture Model

Recently Lawrence and Krzanowski (1996) proposed the homogeneous conditional Gaussian mixture (HCM) model which is based on the original location model of Olkin and Tate (1961) for clustering n observations into g underlying subpopulations using a mixture of continuous and categorical variables. The method combines all levels of the categorical variables into one multinomial variable with m multinomial levels (or cells). The algebraic details of this model, named the location model for simplicity, are given by Franco et al. (1998).

The HCM model (1) requires the estimation of a vector of means in each of the m x g cells (a total of m x g x p means), (2) has a likelihood function that compares each observation with the cell mean and not with the subpopulation mean, and (3) estimates the means of cells that may be empty and thus are not represented in the sample.

The Independent Mixture Model

Franco et al. (1998) proposed a model where the means, variances, and covariances depend not on the specific (is)th cell but rather on ith subpopulation. The main difference between the independent mixture (IM) model and the HCM model is that the vector of means and the dispersion matrix of the IM are assumed to be equal for all multinomial cells within a subpopulation, whereas for the HCM model, the vector of means and the dispersion matrix are assumed to be different in each multinomial cell within subpopulations.

As previously mentioned, each observation (y_{sj})is compared with the mean of the subpopulation (μ_i)and not with the cell mean (μ_{is}) as for the HCM model.

Canonical Variate Analysis

Canonical variate analysis is an ordination method for graphical display that allows groups on the data matrix and focuses on the separation among groups such that it can be used for discriminant analysis. Assume that p attributes are measured in each of the n individuals (matrix of n x p) [i.e., the p attributes measured on the j^{th} individual are represented by $x_j' = (x_{j1}, x_{j2}, \ldots, x_{jp})$] and consider that the n individuals are grouped into g clusters ($i = 1, 2, \ldots, g$). One objective is to examine whether there are differences between the g groups of n_i individuals ($j = 1, 2, \ldots, n_i$; where $n = \sum_{i=1} n_i$). Also, it is assumed that any direction in the p-dimensional sample space is specified by $a' = (a_1, a_2, \ldots, a_p)$, thus we will focus (for the j^{th} individual) on the linear combination $y_j = a_1x_{j1} + a_2x_{j2} + \ldots + a_px_{jp}$. The more separate the groups are in the space, the easier it will be to distinguish the various groups. One major aim is to find a low-dimensional representation of the data that will approximate the high dimensional configuration where the various groups are distinguished. Canonical variables attempt to explain complex relationships in terms of a smaller number of attributes, and if the correlation coefficient between the original and the canonical variables can be adequately interpreted, it will help to characterize the various groups in terms of the attributes associated with the canonical variables.

The Sequential Clustering Strategy for Classifying Farmers Using Categorical and Continuous Socioeconomic Variables

To illustrate the methodology, data from the on-farm conservation project described in Bellon (this proceedings) is used. The project included a random survey of a representative sample of farming households in six communities of the Central Valleys of Oaxaca, Mexico. A total of 240 households were surveyed. An exercise to rank the sample farmers according to wealth was carried out with assistance from key informants in each of the communities. The key informants identified characteristics pertaining to well off, intermediate, and poor farmer groups. These characteristics were related to variables in the survey, though this was not possible in all cases. For example, having interest (or motivation) was a characteristic of the well off, according to the key informants, however this is a difficult characteristic to measure. The variables used to create the classification of farmers in the study included age, education, family demographics, landholdings, animal holdings, sources of nonagricultural income, land quality, and ownership of agricultural implements such as plows, trucks, and tractors. Farmers were classified based on 9 categorical and 22 continuous socioeconomic variables (Table 1).

Estimation of the optimal number of initial groups

The two rules described by Franco et al. (1998; 1999) were used: the upper tail approach (Wishart 1986) and the likelihood profile, associated with the likelihood ratio test. Every hierarchical procedure performs n-1 fusions, and it is possible to arrange these values in increasing order. These sets of values have a mean and a standard deviation which are then used for selection of the best partition. The upper tail criterion selects as the best partition that which has a distance of fusion within the interval (1-α)100% of the distribution of the fusion values. Therefore, a partition with one group less requires a fusion outside the□interval.

The likelihood profile is used as a graphical display for observing the changes to the log-likelihood function in relation to the number of groups. The optimal number of clusters occurs when the log-likelihood function shows its highest increase. Using the Ward method on the 240 farmers, the upper tail approach determined the existence of 7-10 groups, and the changes in the likelihood profile showed that 5 groups is where the highest increase occurs (Figure 1).

The dashed line indicates the value of the log-likelihood for the five groups.

Relevant variables for discriminating among groups

A stepwise discriminant analysis was performed to examine the importance of the 22 continuous variables on the delineation of the 5 groups. Results indicated that the most relevant attributes were:

- C1: Male farmer's age
- C2: Years of education completed by male farmer
- C4: Years of education completed by female farmer
- C5: Family members less than 5 years old
- C7: Family members 16-60 years old

Table 1. Code, type, and description of the attributes used to classify 240 farmers.

Code	Type	Description
Q1	Binary	Male farmer knows how to read and write
Q2	Binary	Female farmer knows how to read and write
Q3	Binary	Has oxen
Q4	Binary	Has tractor
Q5	Binary	Has truck
Q6	Multistate	Importance of agricultural work outside farm
Q7	Multistate	Importance of nonagricultural work outside farm
Q8	Multistate	Importance of money from relatives living in Mexico
Q9	Multistate	Importance of money from relatives living outside Mexico
C1	Continuous	Male farmer's age
C2	Continuous	Years of education completed by male farmer
C3	Continuous	Female farmer's age
C4	**Continuous**	**Years of education completed by female farmer**
C5	**Continuous**	**Family members more than 5 years old**
C6	Continuous	Family members 5-16 years old
C7	**Continuous**	**Family members 16-60 years old**
C8	**Continuous**	**Family members more than 60 years old**
C9	Continuous	Number of hectares in *ejido*[†]
C10	**Continuous**	**Number of hectares in communal lands**
C11	**Continuous**	**Number of hectares in small holdings**
C12	Continuous	Proportion of irrigated maize
C13	Continuous	Number of cattle
C14	Continuous	Number of horses
C15	**Continuous**	**Number of goats**
C16	Continuous	Number of pigs
C17	Continuous	Proportion of land category 1 (good)
C18	Continuous	Proportion of land category 2 (medium)
C19	Continuous	Proportion of land category 3 (regular)
C20	Continuous	Proportion of land category 4 (poor)
C21	Continuous	Proportion of land category 5 (very poor)
C22	**Continuous**	**Proportion of area planted with maize**

Note: The variables in bold face are those that had the greatest influence in discriminating between farmers and therefore the most influence for forming the groups using the Ward method.

[†] An *ejido* consists of land distributed to rural communities after the Mexican Revolution in the early part of the 20th century. By law, *ejido* land was held and worked communally. Title did not reside with individual members of the *ejido* (known as *ejidatarios*) but with the *ejido* as a government entity. Constitutional reform in the late 20th century made it possible for individual *ejidatarios* to claim title to their land and dispose of it as they pleased.

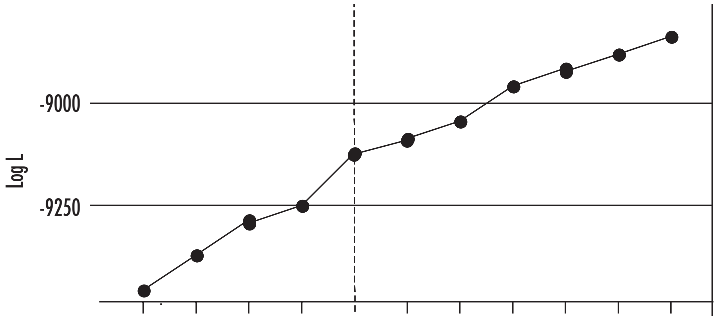


Figure 1. Profile of the log-likelihood function for the number of groups obtained using the Ward method.

- C8: Family members more than 60 years old
- C10: Number of hectares in communal lands
- C11: Number of hectares in small holdings

Years of education, age, and number of hectares seem to be the most important continuous variables for discriminating among the farmers in different groups.

A chi squared test to determine the relevant categorical variables for separating the five groups showed that three binary variables and two multistate variables were the most important discriminators:

- Q1: Male farmer knows how to read and write
- Q2: Female farmer knows how to read and write
- Q3: Has oxen
- Q6: Importance of agricultural work outside farm
- Q7: Importance of nonagricultural work outside farm

Years of education and outside support are the most important categorical variables influencing the groupings.

Ward-MLM

The initial groups formed by the Ward method changed their composition after the MLM method was applied (Table 2). For example, while the initial group of farmers belonging to group G1 comprised 98 farmers, 59 of them remained in G1 after MLM, 16 formed part of final G2, 1 formed part of G3, 18 formed part of final G4, and 4 formed part of final G5. Similarly, while the initial G2 had 43 farmers, 34 remained in the final G3, but 7 formed part of final G2, and 2 formed the final G5. None of the initial G2 farmers moved to final groups G3 and G4. A total of 32% of farmers were moved from one initial group to another final group. Only three observations were classified in a given group with less than 0.75 probability, i.e., 98.75% of the observations were classified in the 5 groups with at least 0.75 probability.

Characteristics of the five final groups

The 5 final groups, in terms of the 5 binary variables, the 4 multistate variables, and the 22 continuous variables, and after the two-stage sequence clustering strategy Ward-MLM, can be characterized as shown in Table 3.

Final group G1 is characterized by low values for C5 variables; high values for C9, C15, and C18 variables; and a high proportion of YES for the binary variables Q1, Q2, and Q3 (Table 3). This means that households in this group have few very small children. On average they own the highest amount of *ejido*[1] land, most of it of

Table 2. Number of farmers that moved from the initial groups formed by the Ward methods to the final groups obtained after MLM analysis.

Initial groups	Final groups G1	G2	G3	G4	G5	Total
G1	59	16	1	18	4	98
G2	7	34	0	0	2	43
G3	3	1	19	0	0	23
G4	2	0	0	27	4	33
G5	1	0	0	17	25	43
Total	72	51	20	62	35	

Note: Numbers on the diagonal are the farmers that remained in the same group after the modified location model (MLM) analysis.

---

[1] An *ejido* consists of land distributed to rural communities after the Mexican Revolution in the early part of the 20th century. By law, *ejido* land was held and worked communally. Title did not reside with individual members of the *ejido* (known as *ejidatarios*) but with the *ejido* as a government entity. Constitutional reform in the late 20th century made it possible for individual *ejidatarios* to claim title to their land and dispose of it as they pleased.

good to very good quality. Many households in this group have oxen and the average number of goats. Most of the male and female heads know how to read and write. These families seem to be in the middle of the demographic cycle and have good access to agricultural assets.

Final group G2 has high values for variables C6, C7, C10, C13, and C19; low values for variables C9 and Q6; and a high proportion of YES for variables Q1, Q2, and Q3. These families have access to family labor, given that most of their members are in the age groups of 5-16 years and 16-60 years (some have small children, but most are teenagers and above). They own, on average, the highest number of cattle and largest landholdings in communal areas, and may depend on cattle farming, though on a small scale. Their land is distributed among all land quality types. Most of the male and female heads know how to read and write. Most have oxen. Off farm labor and, to a lesser extent, non farm labor are important sources of income for these households. The availability of family labor and cattle seem to be important components of their livelihoods.

Final group G3 showed high values for variables C2, C4, C5, C12, and C16; low values for C1, C3, C8, C14, C15, C18, and C19; low values for Q7; high values for Q6, Q8, and Q9; and the highest proportion of YES for Q1, Q2, and the lowest for Q3. These families are the youngest and the best educated of the sample. Of all groups they rely the most on non farm labor. They own the highest proportion of irrigated land and high quality land; however, on average, they own the smallest landholdings. They also own, on average, the highest number of pigs, but no oxen. These farmers are

Table 3. Mean value of the 5 final groups for 22 continuous variables (C1-C22), 4 multistate variables (Q6-Q9), and the proportion of 5 binary variables (Q1-Q5) for each case.

Group	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13
G1	55.2	3.5	47.8	3.6	**0.3**	1.2	3.1	0.3	**2.7**	0.5	0.4	0.2	1.7
G2	41.4	5.0	34.9	5.1	0.7	**2.1**	**2.7**	0.1	**0.7**	**2.4**	0.4	0.1	**1.7**
G3	**38.1**	**8.9**	**32.6**	**9.0**	**1.1**	1.8	2.3	**0.0**	0.9	0.9	0.6	**0.3**	1.1
G4	**68.2**	**1.4**	**62.5**	**0.8**	0.3	**0.5**	**1.3**	**1.7**	**1.6**	0.4	0.8	0.1	1.0
G5	55.3	1.6	51.1	**0.2**	0.3	1.0	**4.6**	0.1	0.9	**0.1**	**2.4**	**0.1**	**0.4**

Group	C14	C15	C16	C17	C18	C19	C20	C21	C22	Q6	Q7	Q8	Q9
G1	1.4	**4.1**	1.1	0.4	**0.1**	0.1	0.1	0.2	0.8	2.8	2.4	2.7	2.5
G2	1.8	3.3	1.2	0.4	0.1	**0.2**	0.2	0.1	0.9	**1.9**	2.1	2.6	2.9
G3	**0.6**	**1.9**	**5.1**	0.6	**0.0**	**0.0**	0.2	0.1	0.7	**3.0**	**1.3**	**2.9**	**3.0**
G4	1.5	1.9	1.2	0.4	0.1	0.1	0.3	0.2	0.2	2.8	2.7	2.5	2.6
G5	1.2	2.7	1.5	**0.7**	0.1	0.1	**0.1**	**0.0**	**0.9**	2.5	**2.6**	2.5	2.5

	Q1		Q2		Q3		Q4		Q5	
Group	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
G1	6	**94**	4	**96**	33	**67**	96	4	90	10
G2	4	**96**	2	**98**	27	**73**	98	2	94	6
G3	0	**100**	0	**100**	**100**	0	95	5	50	50
G4	38	62	68	32	53	47	100	0	92	8
G5	37	63	89	11	17	**83**	100	0	88	12

Note: Values in bold represent the highest or the lowest values for each variable.

probably the better off families, with strong links to the non farm economy, practicing a more suburban type of agriculture.

Final group G4 had high values for variables C1, C3, C8, and C9; low values for C2, C4, C6, and C7; a high value for variable Q7; and high proportions of NO for Q4 and Q5. Families of G4 can be seen as the opposite of those in G3. They are the oldest and least educated and have the highest number of members in the oldest age group. They plant the smallest area to maize and have, on average, the second smallest landholdings and little irrigated land. These families are probably the worst off: they are old, have little available labor, and few assets. Remittances do not seem to be an important source of income either.

Final group G5 showed high values for variables C7, C11, and C22; low values for variables C4, C10, C12, C13, C20, and C21; a high value for Q7; a high proportion of YES for Q3; and high proportion of NO for Q4 and Q5. These households have the highest average number of members in the most productive age group (16-60 years), they own the largest area of private land, but one of the lowest in communal lands. They have the lowest cattle ownership, but most have oxen. Most of their land is planted to maize. While they own a small proportion of irrigated land, most of it is of very high quality. This group may be the most agriculturally oriented, particularly with respect to maize—in general they have good labor availability, a team of oxen, and private land that can be used as collateral. They may have the highest agricultural productivity potential given the high quality of their land.

The classification method has created a set of groups with similar resources and constraints. It reflects the importance of the household demographic structure, education, and access to agricultural assets. Rather than interpreting this classification exclusively in terms of poverty or wealth—although these were the basis for eliciting the variables used, and the patterns of wealth and poverty in it are obvious—the classification can be better interpreted in terms of a diversity of livelihood strategies that respond to the control of different assets. In any case, it is clear from the data that the resources controlled by the sample are relatively modest, even for the better off farmer group.

Canonical analysis and canonical plot

The canonical analysis involves only the 22 continuous variables. The first 2 canonical variables explained almost 90% of the variability existing in the entire data set. The pair-wise squared distances between the 5 final groups are shown in Table 4.

Clearly the final groups that are farthest apart are G3 and G4, followed by G3 and G5. The groups that are closest, with a large overlap between farmers, are groups G1 and G2, closely followed by G1 and G5, and G2 and G3. Group G4 seems to be fairly compact and well separated from the others.

The canonical variables are shown in Table 5. The first canonical variable is positively and highly correlated with

Table 4. Pair-wise squared distance between the five (G1-G5) final groups after the Ward-MLM† strategy.

	G1	G2	G3	G4	G5
G1	0.00	11.24	24.81	21.14	12.11
G2	-	0.00	14.23	45.14	25.93
G3	-	-	0.00	71.20	54.95
G4	-	-	-	0.00	28.10
G5	-	-	-	-	0.00

† MLM = modified location model.

continuous variables C1, C3, and C8, and negatively and highly correlated with C2 and C4. This indicates that the first canonical variable is associated with the age and education of the male and female household heads and the age group of household members in the oldest category. The second canonical variable is positively correlated with C4 and C8 and negatively correlated with C7. This shows that this canonical variable is associated with the education level of the female household head and the demographic composition of the household (particularly the members in the most productive and the oldest age groups). The third canonical variable is negatively correlated with variable C10, which is the area owned in communal lands. The demographic structure of the household and education of household heads are fundamental components of the classification.

Table 5. Canonical variables for each of the 22 continuous variables.

Continuous	Canonical variables		
variable	Can1	Can2	Can3
C1	**0.458224**	0.029365	0.221170
C2	**-0.389385**	0.294195	0.168153
C3	**0.437052**	0.000497	0.173032
C4	**-0.444267**	**0.391052**	0.209704
C5	-0.094468	0.093268	-0.085827
C6	-0.150245	-0.001007	-0.164493
C7	-0.090059	**-0.407822**	0.123471
C8	**0.486446**	**0.471022**	-0.219938
C9	0.048433	0.025511	0.326002
C10	-0.130322	0.099257	**-0.435796**
C11	0.051708	-0.234342	-0.017993
C12	-0.042150	0.119581	0.224169
C13	-0.034151	0.043551	-0.006289
C14	0.012748	0.004001	-0.194377
C15	-0.021852	-0.031214	0.031700
C16	-0.052913	0.065072	0.113935
C17	-0.021000	-0.129863	0.114719
C18	0.034930	-0.014545	0.032909
C19	-0.025951	-0.047472	-0.119569
C20	-0.002225	0.124978	-0.164192
C21	0.023587	0.073985	0.078142
C22	-0.059653	-0.145481	-0.176298

Note: Values represent correlations between canonical variables and the original variables. Values in bold represent the highest or the lowest values for each variable.

Figure 2 shows the plot of the first canonical variable against the second. This graphical representation is useful for visualizing the relationship between groups. It is clear that G4—older households with a low education level—forms a very compact group, well separated from the others, as well as G3—younger households with the highest education level—and G5. Groups G1 and G2 represent intermediate groups with several overlapping observations in terms of the two canonical variables.

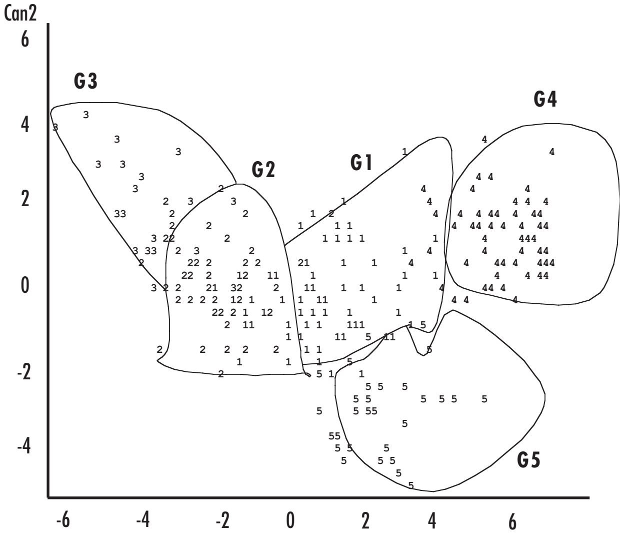


Figure 2. Plot of the first two canonical variables for 240 farmers from the canonical analysis using 22 continuous variables.
Final groups after the Ward-MLM clustering strategy are G1-G5.

Conclusions

This paper presents a method for classifying individuals—farming households in this example—into homogenous but distinct groups. The method allows the use of different types of variables, provides a systematic approach to decide the number of groups present in the data, and assigns a probability that an individual belongs to a group. The value of this method to practitioners of PPB is that it should allow them to group farmers into homogeneous groups, hopefully with similar variety requirements. It should be pointed out, however, that the link between the groups formed and the variety requirements is not automatic. It depends on the choice of variables used in the classification and their relationship with the demand for crop traits. It requires testing of the association between the groups and the crop traits and, hence, the varieties demanded.

References

Franco, J.E., J. Crossa, J. Villasenor, S. Taba, and S.A. Eberhart. 1997a. Classifying Mexican maize accessions using hierarchical and density search methods. *Crop Science* 37:972-980.

Franco, J.E., J. Crossa, J. Díaz, T. Taba, J. Villaseñor, and S.A. Eberhart. 1997b. A sequential clustering strategy for classifying gene bank accessions. *Crop Science* 37:1656-1662.

Franco, J.E., J. Crossa, J. Villaseñor, S. Taba, and S.A. Eberhart. 1998. Classifying genetic resources by categorical and continuous variables. *Crop Science* 38:1688-1696.

Franco, J., J. Crossa, J. Villaseñor, S. Taba, and S.A. Eberhart. 1999. A two-stage, three-way method for classifying genetic resources in multiple environments. *Crop Science* 39:259-267.

Krzanowski, W.J. 1988. *Principles of multivariate analysis. A user's perspective.* Oxford, U.K.: Oxford University Press.

Lawrence, C.J., and W.J. Krzanowski. 1996. Mixture separation for mixed-mode data. *Statistics and Computing* 6:85-92.

Olkin, I., and R.F. Tate. 1961. Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics* 32:448-465 (correction in 36:343-344).

Wishart, D. 1986. Hierarchical cluster analysis with messy data. In W. Gaul and M. Schader (eds.), *Classification as a tool of research.* New York: Elsevier Science Publishers BV.

Discussion Summary

The discussion centered on how to identify and characterize socioeconomic environments that are appropriate for participatory breeding. We do not have enough understanding of the way different socioeconomic variables determine or influence the demand for varietal traits by farmers. This understanding should be the basis for selecting variables to be included in a classification exercise, using the methodology described here, to identify the "social environments" that the breeding should target. It was suggested that one way of looking at the data is to set up hypotheses proposing that certain socioeconomic variables (e.g., animal ownership) would affect cultivar adoption. By generating relevant hypotheses from current varietal adoption patterns, multivariate data (variables) for effective classification can be identified. There is a lot of literature on the factors affecting adoption, but it is still not clear how to make these data relevant to participatory breeding. Another option is to use participatory diagnostics to determine the parameters to use.

There was also discussion on the advantages and disadvantages of looking at the impact of individual factors on the adoption or demand of traits versus using multivariate groupings of factors. It was pointed out that there is no inherent contradiction between both approaches. Clearly in the former approach, the impact of one factor at a time can be tested (keeping the others constant), while the latter does not allow this. However, in many cases factors are highly correlated (i.e., there is high multicollinearity in the data), hence, testing individual factors is difficult. No matter which approach is chosen, it is important to make clear hypotheses at the beginning.

It was emphasized that the most important objective of this presentation was to make participants aware of a powerful methodology to generate homogenous groupings. At the end of the day, any participatory breeding effort has to deal not only with biophysical heterogeneity, but also with socioeconomic and even cultural variability. For practical reasons it is important to segregate this variability into units that can be identified, characterized, and targeted. Any participatory plant breeding effort should target certain "recommendation domains" and therefore should have tools to accomplish this.

Appendix 1.
Current Participatory Breeding Projects Conducted by the Centers Represented at the Workshop

Farmers and Scientists—Building a Partnership for Improving Rainfed Rice in Eastern India - Phase 1

THELMA PARIS AND GARY ATLIN (IRRI)

The farmer participatory breeding project "Farmers and Scientists—Building a Partnership for Improving Rainfed Rice in Eastern India" was conducted by the International Rice Research Institute (IRRI) and six national agricultural institutions under the Indian Council of Agricultural Research (ICAR) in response to the problem of low adoption rates of improved released cultivars in rainfed rice environments. The main reasons often suggested for this poor rate of adoption are: (1) varieties selected on research stations may not outperform traditional varieties under farmer management; (2) improved varieties may not meet farmers' end use and cooking quality requirements; and (3) farmers may not have access to or information about seeds of new varieties. The project was based in eastern India, which hosts the world's largest concentration of rainfed rice.

The goal of this project was to enhance food security and to promote biodiversity. The main research objectives were to (1) test the hypothesis that farmer participation in rainfed rice breeding can help develop suitable varieties more efficiently, and (2) identify stages in a breeding program where farmer participation has the most impact.

The project also involved a social science component with a gender perspective and a plant breeding component. Household surveys and participatory ranking of useful traits using graphic illustrations of traits were used to understand how farmers' different socioeconomic and biophysical situations influence their preference for certain varietal characteristics. They were also used to understand rice varietal diversity in the region.

The plant breeding component applied participatory varietal selection (PVS) and participatory plant breeding (PPB) methods to promote partnerships between female and male farmers and breeders and social scientists, and to develop and evaluate rice varieties suited to rainfed environments. For PVS, 15-25 elite lines and a local check were included in the trials. Two to three farmers per site grew these varieties under their normal management practices. At two or three phenotypic stages of plant growth, farmers and breeders ranked the same set of varieties grown on-station and on-farm. Ranking was ordered from best to worst. Breeders recorded duration, plant height, and yield for each trial. In addition farmers' comments on the characteristics they liked or disliked and reasons for ranking were recorded in diaries. The objective was to get farmers to share their experiences and perceptions of the breeding lines tested on their fields. Five to ten farmers in the village evaluated (ranked) the same set of rice lines on the station and in farmers' fields at specific phenotypic stages. In some of the sites, female farmers were included as farmer cooperators and rankers.

Kendall coefficient of concordance was used to test the influence of farmer participation in the breeding process. Spearman rank correlation coefficient was used to compare farmers' and breeders' rankings and their rankings to the observed value. Analysis of genotype by environment interaction (GEI) for yield was also conducted. Sensory evaluation was carried out to test the cooking and eating quality of the PVS lines in two sites with both female and male farmers.

Research results indicate that farmer participation in varietal evaluation improves the selection of suitable varieties by ensuring that farmers' selection criteria for rice varieties are better understood by breeders. Furthermore, selecting varieties on farmers' fields minimizes the influence of GEI and ensures that lines are tested and selected in representative environments. Hydrological conditions and land type, as well as the usefulness of the variety to meet specific needs, are the major factors determining farmers' choices. Different varieties fulfill different livelihood functions, and farmers respond to the multiplicity of needs by growing a range of varieties. While men and women were in agreement that grain yield and duration of the variety were most important, women gave more importance to traits related to their specific roles such as competitiveness to weeds, quantity and quality of straw from rice, milling recovery, ease of dehusking and threshing, suitability for different food preparation, and storage quality. Meeting different farmers' needs may be better tackled by creating different varieties rather than trying to produce multipurpose varieties.

Scientists involved in the project were: T. Paris, S. Sarkarung, G. Atlin, K. McAllister, G. McLaren, R.K. Singh, B. Courtois, C. Piggin (ex-member), and S. Pandey (ex-member) (IRRI); Abha Singh, V.S. Sisodia, O.N. Singh, S. Singh, and H.N. Singh (NDUAT, a national agricultural research program in eastern India); R.K. Sahu, V.N. Sahu, S.K. Sharma, and M.L. Sharma (Indira Gandhi Agricultural University); R.K.P. Singh, R. Thakur, and N.K. Singh (Rajendra Agricultural University); D. Chaudhary and S. Ram (Central Rice Research Institute); and A.T. Roy and D.C. Pradhan (Orissa University of Agricultural and Technology). This project was funded by the International Development Research Centre, Canada, from 1997-2000, and is part of the SWI-PRGA, International Center for Tropical Agriculture, Colombia.

CG Maize Diversity Conservation: A Farmer-Scientist Collaborative Approach - Phase II

MAURICIO R. BELLON (CIMMYT)

The International Maize and Wheat Improvement Center (CIMMYT) and the Institutio Nacional de Investigaciones Forestales, Agrícolas y Pecuarias (INIFAP-Mexico) have conducted a pilot study on participatory maize improvement in the Central Valleys of Oaxaca, Mexico, since 1997. The goal of this project is to determine whether it is possible to improve maize productivity while maintaining or enhancing genetic diversity. Maize productivity was broadly defined in terms of yield, stability, and other characteristics of interest to farmers. The project conducts and compares different participatory interventions with small-scale farmers in six communities in this region.

Participatory methodologies have been widely used. First, to elicit the traits that farmers value in their maize landraces, during collection of 152 landraces representative of the regional diversity, farmers donating samples were asked to list the advantages and disadvantages of each. This led to the compilation of a list of 25 traits. Second, to measure both the extent to which the farmers demand these traits and the landraces supply the traits, a random sample of farmers, both male and female, was asked to rate each of the 25 traits in terms of importance (very important, somewhat important, not important). All farmers were also asked to rate each landrace in terms of performance with respect to each trait (very good, regular, poor). Since data were collected from male and female farmers in the same household, a Wilcoxon ranked test for two related samples was used to test for differences in ratings of importance by gender. The ratings of performance were compared with a Kruskal Wallis one-way analysis of variance by ranks for grain color, the main taxonomic characteristic used by these farmers. Furthermore, all farmers in the sample were ranked using a wealth ranking method. Then, using a Kruskal Wallis one-way analysis of variance by ranks, the ratings of importance of traits for the three wealth ranks (rich, medium, and poor) were compared (Bellon 2001).

A participatory approach combined with conventional agronomic evaluation was also used to select a subset of landraces for improvement and distribution. It was impossible to work with all 152 landraces collected, so 17 were chosen with good agronomic performance, from different agromorphological groups (a proxy for diversity), and were of interest to farmers. All landraces collected were evaluated in researcher managed trials in farmers' fields in all areas where collection took place. To gauge farmers' perceptions of each landrace, field days at harvest were organized, and farmers were invited to attend. During the field days, participants walked through the trial, observed the landraces, and recorded the numbers of plots that contained populations they liked. All ears from the inner two rows of the experiment were harvested and laid out in front of the stand, so farmers could judge grain yield and examine the ears. The purpose of this exercise was to obtain a rapid "sort" or classification of landraces according to farmers' expressions of interest. The exercise enabled us to systematically deal with many materials (170) and many farmers (approximately 70 per field day) in a relatively brief time period (2-3 hours). We viewed the participants' choices as votes and assumed that the higher the percentage of farmers voting for a landrace, the more potentially valuable it is to participants (Bellon et al. 2002).

Finally, simple experiments were conducted with farmers in their fields. Researchers provided the seed and a simple experimental design, and farmers provided the fields and the management. Each farmer agreed to plant three of the varieties from the field day plus one of his/her own varieties and to manage them in exactly the same way. Each variety was planted in four rows of approximately 10 m. One of the varieties was a common check. Researchers and farmers measured yield at harvest. Farmers kept a management diary during the duration of the experiment and they rated the performance of each landrace according to the 25 traits previously identified.

References

Bellon, M.R. 2001. *Participatory research methods for technology evaluation.* Mexico, D.F.: CIMMYT.

Bellon, M.R., J. Berthaud, M. Smale, J.A. Aguirre, S. Taba, F. Aragón, J. Díaz, and H. Castro. 2002. Participatory landrace selection for on farm conservation: an example from the Central Valleys of Oaxaca, Mexico. *Genetic Resources and Crop Evolution* (in press).

Participatory Varietal Selection in West and Central Africa

HOWARD GRIDLEY (WARDA)

Introduction

West and Central Africa (WCA) encompasses a vast, heterogeneous, and poverty-stricken area. Despite its natural endowments and agricultural potential, rice production still lags behind rice consumption in the region. WCA now imports about 3.2 million tons of rice a year, at a staggering cost of one billion US dollars—a cruel strain on the region's economies. The challenge is to develop sustainable rice production in a competitive world economy, and thereby enhance the region's self-sufficiency in rice, or at least significantly reduce its imports.

Rice production in WCA was based originally on the African indigenous rice *Oryza glaberrima*, but is now dominated by the Asian species *O. sativa*, although landraces of *O. glaberrima* are still grown in small traditional production systems in rainfed and deep-water ecosystems. During the last 40 years, rice breeders in WCA have concentrated on developing and releasing improved varieties of *O. sativa*, mainly using conventional breeding methods such as introduction, hybridization, and selection. Crop improvement scientists at the West Africa Rice Development Association (WARDA) systematically evaluate germplasm from both within and outside Africa, generate breeding materials, select superior lines, and test early and advanced breeding materials on-station and on-farm. WARDA's strategy for rice improvement is to combine specific agroecological adaptations of local rice varieties with the yield potential of introductions.

Oryza glaberrima represents a rich reservoir of useful genes for resistance to biotic and abiotic stresses. In 1991 WARDA initiated interspecific breeding to cross-introgess important traits between *O. glaberrima* and *O. sativa* resulting in the development of several highly promising, fertile, and stable interspecific lines that catalyzed the need for rapid dissemination to farmers.

A Partnership through Participatory Varietal Selection

To provide farmers with rapid access to new *O. sativa* and interspecific lines, WARDA eschewed the conventional top-down approach to technology transfer and initiated farmer participatory varietal selection (PVS). Participatory varietal selection aims primarily to accelerate the transfer of new lines to farmers' fields and determine the lines farmers wish to grow, the agronomic and quality traits farmers value, and the magnitude of gender differences. In the first year a rice garden is established in a village with up to 60 lines sown in an unfertilized and fertilized block. Farmers visit the garden at maximum tillering and maturity to select lines and their selection criteria are recorded. Farmers then receive seed of their selections for the next two seasons to sow on-farm and their selection criteria continue to be monitored.

The first PVS project was installed in 1996 at Boundiali in Cote d'Ivoire where farmers appreciated the concept of sharing responsibilities for rice research. Encouraged by the results, WARDA expanded PVS activities in Cote d'Ivoire in 1997, initiated PVS in Guinea, Ghana, and Togo in 1997, and in 1999 in Burkina Faso, Cameroon, Chad, Guinea-Bissau, Mali, Mauritania, Niger, Nigeria, Senegal, Sierra Leone, and The Gambia.

The selection criteria most cited by participating farmers are higher yield, short growth cycle, plant height, high tillering, weed competitiveness, and grain quality. Female farmers who are responsible for harvesting prefer tall rice to facilitate single-panicle harvesting. Participatory varietal selection has instigated the adoption (and release) of many lines, and farmers in many countries are especially enthusiastic about the new interspecifics, now termed NERICAs (New Rice for Africa). An example of a PVS success story is Guinea. Participatory varietal selection was initiated in 1997 and by 2000 NERICAs covered about 8,000 ha. The expected production is in the region of 15,000 t, of which one-third is supposed to be kept as seed; the production in 2000 generated a minimum gain of US$ 2.5 million over pre-NERICA production. The Guinean authorities project that 300,000 t of NERICA will be produced in 2002, with surplus available for export to neighboring countries, where the demand for seed is also increasing rapidly.

Other countries in WCA are also advancing in their PVS trials. In 2001, there will be approximately 24 NERICAs in advanced testing in farmers' fields in 13 countries. By 2001, 7 NERICAs will have been officially released in Cote d'Ivoire and Guinea, and several are in the pipeline for release in Togo, Benin, and Sierra Leone. Research shows that 10% adoption in only 3 countries—Guinea, Côte d'Ivoire, and Sierra Leone—can return an extra US$ 8 million per year. Adoption by 25% of farmers will return US$ 20 million.

Cassava Selection by Participatory Plant Breeding Methods in Southern Africa

N.M. Mahungu (IITA)

The Southern Africa Root Crops Research Network (SARRNET) is a regional organization operating within the 14 countries of the Southern Africa Development Community (SADC) implemented by the International Institute of Tropical Agriculture (IITA) in collaboration with the International Potato Center (CIP). The Network is involved in research and development activities on cassava and sweet potato.

Germplasm development, through plant breeding and introductions from IITA, Ibadan, or the region, is one of the major research activities carried out by IITA/SARRNET. Variety breeding for specific use characteristics (including fresh roots, flour, starch, and feed) is carried out in main research stations/centers. Clones are tested for adaptability in different agroecological zones of the region for root yield, disease/pest resistance, agronomic, and quality characteristics.

Since selected/improved varieties have to be adopted and used by farmers and other producers, these groups need to be satisfied with the technologies developed. Participatory research approaches are therefore advocated to ensure that the technologies generated meet the end users' expectations. This approach accelerates technology dissemination and adoption.

SARRNET advocates participatory plant breeding (PPB) because this approach involves all major stakeholders such as farmers, extension, processors, entrepreneurs, and researchers in several sequential stages of plant breeding, unlike other methods, which involve them in only the final selection processes (Mahungu and Kanju 1997). In cassava breeding, farmers are involved as early as in the selection of segregating materials of F_1 populations. While researchers record quantitative data, most farmers' assessments are visual and verbal comments/observations.

The following examples show how participatory methods have been exploited for cassava breeding in SARRNET and IITA programs:

In Bukoba region of Tanzania, farmers' participation in the selection of cassava varieties from on-farm trials resulted in the selection by farmers of varieties for specific interests, i.e., for high root yield, intercropping, leaves as a vegetable, or processing qualities. However, some of the selections were dropped after farmers were briefed on the disease susceptibility of the varieties (Kapinga et al. 1997).

In a participatory variety selection at Mansa Station in Zambia, farmers selected 15 clones of which 13 were among the 14 clones selected by the breeder. However, after analysis of dry matter, only 12 clones were finally selected that satisfied both farmers and researchers (Mahungu 1999).

Currently SARRNET, Malawi, is involving a timber company in the selection of cassava varieties with good flour characteristics for plywood filler/ binder. Of the four varieties tested, the company has selected two with characteristics similar to those of wheat flour for plywood binding.

References

Kapinga, R., P. Bart de Steenhuijsen, S. Kajiru, J. Chirimi, C. Rugutu, and N.M. Mahungu. 1997. Selection of cassava varieties by farmers in the Lake Zone of Tanzania. *African Journal of Root and Tuber Crops* 2 (1 & 2):248-253.

Mahungu, N.M., and E. Kanju. 1997. *Cassava breeding manual.* Reference document at the regional workshop on cassava breeding, Kibaha, Tanzania, August 25-30 1997 (draft).

Mahungu, N.M. 1999. Cassava germplasm enhancement in Southern Africa. In Akoroda and Teri (eds.), *Food Security and Crop Diversification in SADC Countries: The Role of Cassava and Sweet Potato.* N.p.:N.p.

Participatory Plant Breeding at the International Center for Research in the Dry Areas

MALIKA A. MARTINI (ICARDA)

Participatory plant breeding (PPB) started in Syria in 1996 and is now being conducted in five other countries in the region covered by the mandate of the International Center for Research in the Dry Areas (ICARDA). In these countries PPB has focused only on barley, except in Yemen where selections have also included lentils. The selection process in these countries has involved male farmers only, except, again, Yemen, where women participated in 2000. Barley entries included in the selection process are both fixed lines and segregating populations.

Entries were chosen to test farmers' and breeders' preferences for different attributes and/or characteristics. The entries were planted in nine farmers' fields and two ICARDA research stations. Unreplicated plots of 8 rows of 7.5 m (12 m^2) in 4 strips of 52 plots were established on-farm and on-station. There was one exception (Sauran), where each entry was arranged in 8 strips of 26 plots. Farmers were given a field book to record daily precipitation and plot evaluations. Different scoring methods were used in collecting the data. A numeric scale from 0 to 5 (0 = worst, 5 = best) was used. Qualitative scoring was also used (bad, medium, good, very good, and excellent).

Many types of selections were performed:

- individual selection by each participating (host) farmer alone in his own field;
- selection by each participating farmer in Breda and TelHadya (research stations);
- selection by the senior barley breeder of DASR, Ministry of Agriculture, in each of the nine locations and in Breda and TelHadya; and
- group selection by neighboring farmers at five of the nine locations.

Selection was conducted exclusively between entries.

The data available from the research to date are restricted to the year of selection, farmers' scores of the different entries, and farmers' reasons (up to five) given to support their choices. The top 15 varieties have been coded and entered onto a computer database. A profile questionnaire was designed, and approximately 140 farmers were interviewed over 3 years (1996-97, 1997-98, and 1998-99). Information on seed scoring was also collected (positive, negative, and neutral).

We are interested to know if other people have used the same PPB methodology, and, if so, how they analyzed the data on farmers' preferences and farmers' scoring of varieties. At present we lack the expertise to analyze this information. As the first step, we are in the process of collecting information on the socioeconomics of PPB in Syria and Yemen.

Participatory Variety and Clone Evaluation within Farmers' Field Schools in San Miguel, Peru

Oscar Ortiz (CIP)

Background

The International Potato Center (CIP) and CARE-Peru initiated a collaborative project on the integrated management of late blight (IDM-LB) through farmers' field schools (FFS) in 1997. This project aimed at developing IDM technologies and adjusting the FFS methodology to work with potato farmers in the Andes. As part of the technology development with farmer participation, the work focused on facilitating farmer access to resistant genotypes and evaluating control practices (optimizing fungicide use).

Participatory Variety and Clone Evaluation

During 1997 and 1998, farmers evaluated 12 varieties with different degrees of resistance to LB. In 1999, CIP's breeding program provided 54 promising clones with resistance to LB, which were divided into groups and evaluated by 13 farmer groups. In 2000, farmers continued evaluating 25 selected clones from the previous season. Farmers also evaluated clones originating from true potato seed. Experiments had 2 or 3 repetitions and each involved between 10 and 15 genotypes.

Farmers evaluated genotypes at harvest, focusing on yield and tuber characteristics (tuber shape, color, and proportion of different sizes). After harvest, farmers also evaluated culinary quality.

Taking into consideration that most people in the San Miguel area have difficulties in reading and writing, a visual evaluation methodology was used. Each farmer expressed his or her opinion using small cards with drawings of human faces on them. If the farmer liked the genotypes, he/she used the card with a happy face drawn on it; if he/she considered the clone as regular (not good, but not bad either), he/she used a card with a serious face on it. If the farmer disliked the clone, he/she used the card with a sad face on it. The evaluation was made on an individual basis and each evaluator put the card in a paper bag located near the genotype that was being evaluated. At the end of the evaluation, all participants should have evaluated each genotype.

After the evaluation, the facilitator counted the cards (according to type of face) in each genotype. Results were written on paper or a blackboard and presented to the group. Each genotype had a total number of happy, regular, and sad faces, which allowed the facilitator to see which genotypes were preferred by farmers. The facilitator asked farmers why they liked or disliked each of the genotypes, taking notes of the farmer criteria.

With the purpose of quantitative analysis, a value was given to each type of card. In this way, each genotype was represented by an index, which was ordered to see which were the preferred genotypes. The ranked genotypes, according to the index, were useful to compare opinions among farmer groups that evaluated the same genotypes. Non-parametric tests were used for comparison purposes. The method was also useful to see if female preferences were similar to male preferences.

Participatory Breeding with Sorghum in Mali: Statistical and Analytical Aspects

E. Weltzien Rattunde (ICRISAT)

The research on participatory sorghum (*Sorghum bicolor*) improvement in Mali focuses on two key objectives: (1) the modification of the priorities and objectives of sorghum improvement research for Mali to better meet farmers' needs and preferences, and (2) farmers' assessment of specific new varieties of sorghum in a wide range of production zones. The work thus involves variety evaluations conducted by farmers and in depth discussions on criteria for choosing or rejecting new varieties as a basis for understanding which characters are priority traits. Included in these discussions are the advantages and disadvantages of the local varieties being grown and changes in the production system that have occurred.

We thus have a large body of three types of data:

1. Quantitative data from measurements taken in farmers' plots. We use alpha designs to increase the number of participating and contributing farmers in each village. We use REML-based procedures for the analysis of these quantitative data, to deal with the relatively larger number of missing plots, and to be able to include each farmer's choice of control variety in the analysis.

2. We have increasingly multi-facetted data from farmers' rankings of the tested varieties for a large and variable number of criteria, over many locations, by different members of the participating farm household, and for up to three years of consecutive testing. We use the ranks primarily to establish whether or not a new variety is superior to the local control. To date the numbers of responses per agricultural zone seem too low to attempt statistical analysis.

3. The third type of data is listings of variety traits and characteristics that farmers use when they describe the advantages and disadvantages of a variety in a trial. For each trait we determine the frequency with which it has been used by the farmers evaluating the varieties. Together with this frequency, we also record whether the test variety was judged superior, similar, or inferior to the control variety for the particular trait. With this type of data we can make frequency comparisons for specific traits, for groups of traits, by specific types of farmers, in specific production regions, to allow for a quantitative analysis of key issues. However, we have focused more effort on a qualitative analysis of these results so far. The aim here is to gain a better understanding of what

farmers’ concepts are of specific traits or groups of traits, e.g., grain quality or requirements for adaptations. This also requires a detailed understanding of the production systems from the farmers’ perspective. Thus, this type of qualitative data is used to gain detailed understandings of key issues, to formulate hypotheses for testing, and for detailed analysis.

Participatory Improvement and Dissemination of Maize Varieties with Resistance to Stem Borer in Southeastern Nigeria

Sam O. Ajala (IITA)

Maize cultivation in southeastern Nigeria is plagued by two major constraints: stem borers and acid soils. Acid soils, which occur in pockets throughout the region, can be easily managed through soil improvements including chicken manure. Stem borers, however, are usually controlled using insecticides or by planting local varieties—both approaches are often considered inadequate by the local farmers. Although southeastern Nigeria has a bimodal rainfall pattern, stem borer infestation is so intense in the second season that maize is generally not planted.

Cassava is the major food crop grown in southeastern Nigeria because it survives in the poor (acidic) soils; however, a large proportion of the people interviewed for this study emphasized the importance of maize. Maize is primarily a cash crop; green ears are sold for consumption fresh, roasted, boiled, or made into breakfast cereal. The farmers interviewed wished to increase their maize cultivation but were limited by the lack of improved varieties with resistance to stem borer, especially in the second planting season.

Trials were conducted in association with petroleum industry-based nongovernmental organizations (NGOs) in the region. A number of maize varieties with resistance to stem borer and local checks were planted in different locations and seasons. Farmers were invited to assess the varieties at maturity. Preference ranking and pair-wise comparison were used to assess and select varieties. At the end of the exercise, three varieties were selected (through participatory methods) and established on-farm in the next season. Two of the three varieties have been adopted and are now commonly grown in the region.

Farmers want to retain these varieties because they are higher yielding and perform well in the second season. However, the concerns expressed by different farmer groups on the fresh maize and milling qualities of the varieties highlighted the need to breed varieties for different niches. Initial discussion on how to select the new maize varieties revealed farmers’ desire to become partners in the improvement process. Recurrent selection in maize involves the generation and selection of desirable progenies for recombination to form a new but improved cycle of the same population. Because maize is an open-pollinated crop, however, the appearance of the final product after recombination will be different to that of the selected progenies. It is therefore imperative to come up with innovative ideas for involving farmers in the actual participatory improvement process. Issues at stake include the type of progenies to generate, the number of progenies to evaluate with individual farmers or groups of farmers, the choice of checks, and the design/analysis of data generated to aid effective selection of progenies for recombination. All of these issues are being addressed by a group of maize scientists working under the auspices of the regional network, the West and Central Africa Maize Network (WECAMAN).

Appendix 2. Workshop Participants

Sam O. Ajala
IITA
c/o L.W. Lambourn & Co.
Carolyn House
26 Dingwall Road
Croydon CR9 3EE
England

Gary Atlin
IRRI
M.C. P.O. Box 3127
1271 Makati City
Philippines

Mauricio R. Bellon
CIMMYT
Apdo. Postal 6-641
06600 Mexico, D.F.
Mexico

Merideth Bonierbale
Crop Improvement and Genetic Resources Department
CIP
Apartado 1558
Lima 12
Peru

Richard Coe
ICRAF
P.O. Box 30677
Nairobi
Kenya

José Crossa
CIMMYT
Apdo. Postal 6-641
06600 Mexico, D.F.
Mexico

Lan Fang
Institut für Agrarpolitik und Marktforschung
Justus-Liebig-Universität
Senckenbergstrasse 3
D-35390 Giessen
Germany

Steve Franzel
(could not attend, but his paper was presented by R. Coe)
ICRAF
P.O. Box 30677
Nairobi
Kenya

Howard Gridley
WARDA
01 BP 2551
Bouake 01
Cote d'Ivoire

Hugo de Groote
CIMMYT
ICRAF House
P.O. Box 25171
Nairobi
Kenya

Nina Lilja
CIAT-PRGA Program
Apartado Aéreo 6713
Cali
Colombia

N.M. Mahungu
IITA/SARRNET Coordinator
Chitedze Research Station
P.O. Box 30258
Lilongwe 3
Malawi

Malika A-Martini
ICARDA
P.O. Box 5466
Aleppo
Syria

Ernst-August Nuppenau
Institut für Agrarpolitik und Marktforschung
Justus-Liebig- Universität
Senckenbergstrasse 3
D-35390 Giessen
Germany

Thelma Paris
IRRI
M.C. P.O. Box 3127
1271 Makati City
Philippines

Rafael Omolehin
Institut für Agrarpolitik und Marktforschung
Justus-Liebig-Universität
Senckenbergstrasse 3
D-35390 Giessen
Germany

Oscar Ortíz
CIP
Apartado 1558
Lima 12
Peru

Sieglinde Snapp
Department of Horticulture
Michigan State University
A440-A Plant & Soil Sciences Building
East Lansing MI 48824-1325
USA

Bhuwan Sthapit
IPGRI Regional Office for Asia, the Pacific and Oceania
P.O. Box 236, UPM Post Office
43400 Serdang, Selangor Darul Ehsan
Malaysia

Eva Weltzien Rattunde
ICRISAT
Eltgeshohl 3
53424 Remagen
Germany

John Witcombe
DFID Plant Sciences Research Programme
Centre for Arid Zone Studies
University of Wales
Bangor LL57 2UW
United Kingdom

Monika Zurek
Institut für Agrarpolitik und Marktforschung
Justus-Liebig-Universität Giessen
Senckenbergstr. 3
35390 Giessen
Germany

ISBN: 970-648-096-X

CIMMYT
International Maize and Wheat
Improvement Center

PRGA
Participatory Research and Gender Analysis

IRRI
International Rice Research Institute