630 US ISSN 0197-9310 November 1985

Design and Analysis of an International Experimental Network: Legume Inoculation Trials in the NifTAL Project, the INLIT Experience

R. J. Davis
F. B. Cady
C. L. Wood
C. P. Y. Chan

000180
S

Research Series 042

HITAHR · College of Tropical Agriculture and Human Resources · University of Hawaii

Library of Congress Cataloging in Publication Data
Main entry under title:

Design and analysis of an international experimental network.

(Research series ;)
1. Legume inoculation--Research--Tropics.
2. Legumes--Tropics--Field experiments. I. Davis, R. J. (Robert J.) II. Title: NifTal project. III. Title: Inlit experience. IV. Series: Research series (Hawaii Institute of Tropical Agriculture and Human Resources) ;
SB177.L45D47 1985 633.3'048947 85-5555

ACKNOWLEDGMENTS

This research was supported by the U.S. Agency for International Development through contract numbers AID/ta-C-1207 and DAN 0613-C-00-2064-00. The success of the trials was due to the research efforts of our numerous cooperators.

THE AUTHORS

R. J. Davis is Head, International Outreach Program, NifTAL Project, University of Hawaii.

F. B. Cady is Visiting Biometrician, NifTAL Project.

C. L. Wood is Associate Professor, University of Kentucky, Lexington, Kentucky.

C. P. Y. Chan is Computer Specialist/Data Analyst, Benchmark Soils Project, University of Hawaii.

CONTENTS

	Page
Abstract	3
I. The INLIT Experience	3
The NifTAL Project	3
History of INLIT	4
Coordination of INLIT	4
INLIT Experiments	5
Field Testing for Legume Inoculation Response	5
Objectives of INLIT	7
II. Design of the INLIT "A" Experiment	7
General Considerations	7
INLIT Treatment Selection and Experimental Design	9
Discussion of Design Alternatives	11
III. Intrasite Analysis of an INLIT Experiment	13
Inoculation Questions of Interest	13
Comparisons Among Treatment Means	15
Glycine max (Soybean) Case Study	15
IV. Nonorthogonal Comparisons Among Treatment Means	21
V. Multilocational Analysis of INLIT Data	26
Preliminary Combined Data Analysis	27
Environment Means Data Analysis	31
VI. *Cicer arietinum* (Chickpea) Multilocational Case Study	32
VII. Summary and Conclusions	42
VIII. References Cited	43

DESIGN AND ANALYSIS OF AN INTERNATIONAL EXPERIMENTAL NETWORK: LEGUME INOCULATION TRIALS IN THE NifTAL PROJECT, THE INLIT EXPERIENCE

R. J. Davis
F. B. Cady
C. L. Wood
C. P. Y. Chan

ABSTRACT

Described is a system of worldwide trials for testing response to legume inoculation, the International Network of Legume Inoculation Trials (INLIT). The basic INLIT experiment is described in detail: formulation of the research design, the coordination and servicing of trials, and the analysis of data on a single and multiple location basis. Included is a treatment of the data analysis of field experiments containing nonorthogonal comparisons. Secondary benefits of the trials have been identifying research workers and policy makers in developing countries, and organizing these persons into an organization for expediting research and information flow. Acceptance of the concept by many legume inoculation workers has indicated the validity of the concept, and feedback from cooperators has indicated that the INLIT program has been successful.

Keywords: biological nitrogen fixation, networking, legume inoculation trials, international experimental network, agricultural development, nonorthogonal comparisons, multilocational analysis, environmental means data analysis.

I. THE INLIT EXPERIENCE

The NifTAL Project

The NifTAL (Nitrogen Fixation by Tropical Agricultural Legumes) Project is a United States Agency for International Development (USAID) centrally funded project contracted to the University of Hawaii. NifTAL is funded to provide support to legume inoculation workers in the developing countries of the tropics. Conducting research on specific problems of immediate importance to legume workers and implementing the transfer of known technology to tropical countries are two major means by which NifTAL provides such support. A major NifTAL effort is the International Network of Legume Inoculation Trials (INLIT). INLIT is a major vehicle by which NifTAL achieves both its research and technology transfer functions. It is a loose association of all legume inocu-

lation workers that facilitates information flow and enhances the aggregate research effort. NifTAL promoted the organization of INLIT, and it continues to provide coordination and service support.

History of INLIT

In the late 1970's USAID decided to increase its activities in biological nitrogen fixation (BNF). In considering where to put emphasis, a consensus developed within USAID that the most pressing need in the tropics was for definitive field trials in legume inoculation.

In January 1979, a workshop was held at NifTAL (Harris, 1979) to consider this idea. In attendance were legume inoculation workers from developed and developing countries, representatives from various international agricultural research centers and international donor agencies, and others interested in international agriculture. The latter included agronomists, biometricians, management specialists, social scientists, and soil scientists whose expertise or interest impinged upon the subject. This group further endorsed the idea of worldwide inoculation trials.

The need for a coordinated worldwide program in legume inoculation research in developing countries stemmed from the following factors: (a) there was insufficient demand for legume inoculants in the developing world to support substantial commercial inoculant production such as exists in North America and Australia; (b) there was no knowledge about the farmers' perceptions of the benefits of inoculation; (c) inoculant production facilities had been installed in some countries in the tropics but product demand had not developed; (d) few studies with tropical legumes had conformed to the criteria considered necessary for a definitive inoculation trial; (e) there was at that time a rising interest in BNF because of the rising cost of fertilizer nitrogen; and (f) it was recognized that these individual trials would be more meaningful if coordinated in an international network.

Therefore, workshop participants recommended the establishment of an international network of all legume inoculation workers. Network members would be urged to conduct definitive field trials using a uniform experimental design so that data could be analyzed across locations, on either a countrywide, regional, or global basis. Each individual cooperator would collect data for his location, but would gain further benefit from the composite analyses. The workshop also suggested that for such a network to succeed, some organization would have to actively promote, organize, coordinate, and provide service support to the activity. It was recommended that NifTAL be given this mandate. This was approved by USAID, and NifTAL's organizational structure was realigned to handle the increased responsibility. The International Network of Legume Inoculation Trials (INLIT) was initiated in calendar year 1980.

Coordination of INLIT

To coordinate INLIT, NifTAL designated certain staff members as Network Coordinators, with regional responsibilities for Latin America, Africa, and Asia. Later the African region (which included the mid-East) was divided into two parts—mid-East North Africa and sub-Sahara Africa—and one person was designated Head of International Outreach to manage all INLIT activities.

The Network Coordinator is the cooperator's link with NifTAL and with others in INLIT. Network Coordinators actively promoted the network by extensive travel throughout their region. NifTAL attempted

to identify all legume workers and explain the concept and operation of the INLIT program to them. Later Coordinators' activities were directed more towards visiting experiments and consulting with cooperators. Regional cooperators try to visit as many cooperators, and as many trials, as possible. In between visits cooperators' questions are answered by mail. No attempt was made to pressure researchers into joining INLIT. Rather the advantages of membership were explained and allowed to stand on their merits. Research workers would, through NifTAL, be tied into a worldwide communications network. If all workers conducting trials would employ one standardized experimental design, trials could be analyzed on a countrywide, regional, or worldwide basis. Each researcher would get at least as much information out of his individual experiments as he could by designing his own trials, but by producing data that could be analyzed on a worldwide basis information would be obtained that would also be useful to policy makers and donor agencies. In addition, the standard trial package could be used by legume workers not conversant with inoculation techniques and procedures.

We believe that the success of the INLIT has been in no small measure due to the personal touch provided by this organizational approach. The Network Coordinator causes the cooperator to perceive both NifTAL and INLIT as being composed of real people who are interested in his activities. This provides motivation that would not be provided cooperating with a faceless bureaucracy.

INLIT Experiments

INLIT consists of three components, the "A," "B," and "C" Experiments, designed to determine just what advantages can accrue, with the present state of the art, from inoculating leguminous crops in the tropics. The "A" Experiment is designed to ascertain field response to inoculation using the best available strains of *Rhizobium*. It is therefore a test of the present state of the art. It requires no microbiological expertise, so it can be conducted by any INLIT cooperator. The "B" and "C" Experiments are concerned with a search for better strains in cases where neither "A" Experiment inoculation nor control treatments give maximum yields. The "B" Experiment, a pot experiment to test local isolates, requires some microbiological expertise. The "C" Experiment, formatted like the "A" Experiment, evaluates in the field the best strains from the "B" Experiment against the "A" Experiment inoculum. This bulletin describes the design and analysis of the "A" Experiment.

Many cooperators in the INLIT will conduct only the "A" Experiment, either because they lack microbiological expertise or because results from the "A" Experiment do not indicate the need for further research. Often the best available strains from various major culture collections will be equal or superior to local isolates. Such strains represent the end product of many isolations over the years and have been tested against a wide range of plant varietal and agroclimatic conditions. Even if the local conditions contain some stress factor, the bacteria will often rapidly adapt to the stress because of their short generation time. However, as indicated above, one should not stop with a negative response in the "A" Experiment.

Field Testing for Legume Inoculation Response

There is no substitute for field testing to evaluate the response to inoculation. Although this is considered a truism by many *Rhizobium* workers, NifTAL Net-

work Coordinators continually find workers making the assumption that *Rhizobium* strains indigenous to a given area will be superior to introduced strains. Acceptance of this assumption obviates the need for tedious testing, but to do so is invalid. Strains in major culture collections, having been tested over a wide range of plant germplasm and agroclimatic conditions, will often be superior to new isolates. This is not surprising since they represent the end product of an exhaustive testing program in which many isolates were discarded. Therefore, any new isolate should be tested extensively under field conditions before using it in an inoculant. The situation is often exacerbated by defining the "area" in a geographic or political rather than an agroclimatic sense. Therefore, there is no substitute for adequate field testing to determine the response to inoculation.

Field testing must be extensive. A complex experiment with many objectives may, if pursued to a successful conclusion, give much information about a specific crop–location response. But the data will not be of much value in making decisions on an areawide or countrywide basis. This requires extensive trials, although the individual experiments may be simple. In fact, as is elaborated below, the most extensive trials system possible requires the simplest possible experiment. The above is true whether the purpose of the trials is to determine the need to inoculate using established *Rhizobium* strains or to evaluate potentially superior strains.

The necessity of determining the need for inoculation can be obviated by the simple expedient of inoculating all leguminous crops. Here, the rationale is that inoculation will never decrease yields. Therefore, inoculation can be considered insurance since its cost will be very small compared with the value of the crop or the cost of nitrogen fertilizer. Such a policy is a generally recommended practice in the developed countries where quality inoculant is readily available through commercial channels, the farmers understand the concept of insurance, there is an effective extension system, and the farmers are literate, with a level of sophistication enabling them to make a value judgment as to whether they wish to inoculate.

In the developing countries the above conditions generally do not exist. There is no infrastructure for inoculant production, distribution, and use. Where inoculant production facilities are available, there is no system for distributing the inoculant to the user in a viable condition. In addition, there is no user demand for the product. In such cases large-scale inoculant production facilities, unless designed on a modular basis, can become white elephants.

To develop demand for inoculant entails an educational program to create an awareness on the part of the potential users, both of the need to inoculate and of the precautions necessary so that the *Rhizobium* are still viable when they contact the seedling root. The educational aspects of the problem, even where an extension infrastructure exists, are enormous. There are generally many more farmers per unit of land area than in the developed world. These tend to be less educated and sophisticated so they are less able to absorb educational materials that are not specifically tailored for them. And, of course, education about inoculation must compete with myriad other concerns such as pest control, fertilization other than nitrogen, disease problems, water management, etc.

Each developing country must make its own decision with respect to inoculant recommendations. When deciding what crops and acreage to inoculate, decision

makers must avoid the mistake of equating inoculant production capacity with acreage inoculated. Inoculant production capacity does not equate to crops inoculated. Because of these considerations it is prudent for a developing country to first determine the likelihood of a response to inoculation. Scarce developmental resources can then be employed in those situations where there is a likelihood of a cost-effective response to inoculation. The determination should indicate not only whether an increase in yield is possible under the existing conditions, but also whether there is a potential for getting a yield increase if other inputs become available. If there is not, then scarce developmental resources should be channeled into more productive pursuits. If an economical yield increase from inoculation is theoretically possible, even if not presently attainable, resources should be put into research to determine and overcome the impediment.

Determining response from inoculation, whether testing best available strains or determining suitability of promising *Rhizobium* isolates for use in inoculants, requires a properly designed field experiment. A *Rhizobium* strain pot test with soils from areas of potential use will reduce the number of strains that need to be tested in the field. Many *Rhizobium* isolates that show promise in sand or water culture in a green- or shade-house under sterile conditions will not live up to their initial promise when tested in nonsterile soil. Single row plot trials are also useful in reducing the number of strains. However, they are only marginally superior to a soil pot test in evaluating field performance. In fact, the choice between a pot test and a single row plot test is often dictated by whether or not the soil to be tested is in a field adjacent to the laboratory or at a considerable distance.

Objectives of INLIT

The primary objective of INLIT is to determine under definitive conditions whether or not a yield increase can be obtained through legume inoculation using the present state of the art. It was recognized that the factor generally lacking in trials that had been reported in the literature was an assessment of the quality of the inoculant when it was used.

Secondary objectives are:

(a) To identify all research workers in the tropics who are currently or are contemplating conducting research on legume inoculation.

(b) To convince all workers conducting field trials on the effect of legume inoculation to utilize a common experimental design, thereby enhancing the overall value of the data obtained without decreasing its value to the individual research worker.

(c) To improve the quality of research work and the expertise of research workers in legume inoculation by providing various support services to legume workers.

II. DESIGN OF THE INLIT "A" EXPERIMENT

General Considerations

The principles of experimental design are well developed and familiar to most researchers. Randomization, replication, and blocking in comparative experimentation are emphasized in experimental design books such as Cochran and Cox (1957).

Replication of treatments in a completely randomized design allows the calculation of experimental error while randomization is necessary for its validity. In addition, use of randomization also avoids bias in treatment means from undue influence of confounding factors. Experimental error divided by the number of replications is the

variance of a treatment mean. The standard error is then proportional to the square root of replication number. Grouping of field plots into homogeneous blocks decreases the magnitude of experimental error. However, unless treatments are replicated within a block, a direct measure of experimental error is not available. Only by an assumption (on the residual mean square from the analysis of a randomized block design) is an estimate of the experimental error available. Additional restrictions on the allocation of treatments to the plots or on block sizes result in latin square and incomplete block designs. These more complex experimental designs usually are not warranted in field inoculation experimentation. Their increased operational requirements and additional difficulties in the data analysis with unforeseen complications such as missing plots encourage the use of simpler experimental designs.

The general objectives in field inoculation trials are to determine the magnitude of response to inoculation of leguminous crops and, equally important, to calculate a realistic measure of variability (the standard error) associated with the inoculation response. The crux of trials to determine inoculation response, or to compare strains, is to have a properly designed experiment that addresses the primary hypothesis. The choice of the experimental design for allocating inoculation treatments to the field plots is usually a choice between a completely randomized design and a randomized complete block design. The magnitude of the standard error can be controlled to some degree by the experimenter through use of blocking or increasing the number of replications.

How the data are analyzed will depend not only on the experimental design, but also on the selected treatments. If the field layout employed is straightforward, the experimenter can do a better job of planning the treatment structure and subsequent data analyses. It is generally recognized that field data require statistical analysis. However, the utility of designing treatments so that the primary questions can be answered directly from the data analysis often does not receive sufficient attention during the planning stage. Weaknesses in treatment design recognized during data analysis and interpretation equate to increased cost and time of experimentation.

Two shortcomings in analysis of experiments with inoculation treatments are sufficiently common to warrant special mention. One is the utilization of computing procedures for arraying treatment means in assessing inoculation response. Pairwise mean separation procedures should not be used for analyzing an experiment with a treatment structure designed for specific comparisons among treatment means (Chew, 1980). These procedures should be restricted to arraying means such as in a strain elimination test.

A second problem involves use of default options in data analysis programs. Modern statistical computing packages not only relieve the experimenter of burdensome calculations, but also allow additional analyses that otherwise might not be undertaken. On the other hand, these efficient programs also require experimenters to understand the computing procedures sufficiently to elicit output that directly addresses the formulated hypothesis. The output from default options of most programs is of general interest, but usually does not yield estimates and variability measures to answer the specific objectives of the experimenter. For example, a small probability value associated with a large F value for the ratio of the treatment mean square to the experimental error mean square simply reveals that statistically at

least one of the treatment means differs significantly from at least one of the other treatment means. The treatment structure of the INLIT "A" Experiment trials is designed to give specific comparisons among the treatment means. Computer output should give the estimates and the associated standard errors of these same specific treatment comparisons.

An ideal experimental site will never be found. Although certain obvious problems should be avoided, generally the site should approximate the soil and climatic conditions expected on farmers' fields. If one is testing soybean inoculation in a country where most fields contain no soybean *Rhizobium* it would be meaningless to use experimental land that had had inoculated soybean trials for several seasons. But a less than ideal condition can, and should, be tolerated if it is representative of the land on which farmers will be growing the crop.

INLIT Treatment Selection and Experimental Design

The "A" Experiment is the key to answering the question of whether or not, with the present state of the art, a yield increase can be expected from inoculating tropical legumes with *Rhizobium*. The primary question asked in the "A" Experiment is whether, with the best available *Rhizobium* strains, an increase in yield can be expected from inoculation with *Rhizobium*, and if so, whether the yield obtained is the maximum yield attainable. Secondary questions are whether, when an increase is not obtained from inoculation, the yield obtained without inoculation represents the maximum yield possible, and if not, why the maximum yield is not obtained. The maximum yield, with respect to N, is obtained by incorporating a fertilizer, or "bag" nitrogen treatment. This is an arbitrary high level of nitrogen applied to all nitrogen plots regardless of the level of other fertility factors. It is applied in split dosage so that nitrogen will not be limiting any time during the course of the experiment. Where there has been a chance for excessive leaching, such as high rainfall immediately after an application, the cooperator is instructed to reapply the nitrogen. The purpose of the nitrogen treatment is not to see how much nitrogen the crop needs, but to ascertain that N is not limiting so as to have a maximum value of yield with respect to N. The cause of less than maximum yield can be inferred from nodule information, although further experimentation will sometimes be necessary.

These questions are all considered at two fertility levels. The lower level, designated F, for farmer, represents current farmer practice or, where applicable, current recommendations to farmers. The higher level, designated M, for maximum, represents a level of fertility, other than nitrogen, to ensure maximum yield. Where yield curves have been determined for the particular crop and agroclimatic conditions, the values from these experiments should be used. When such values are not available, the usual case, the M level treatment consists of arbitrary high levels of P and K (1000 kg/ha P, 100 kg/ha K for volcanic ash soils, 100 kg/ha each for other soils), minor elements if suspected of being lacking, and lime, if needed, to adjust the soil pH to an optimum value for growth of the crop. This gives six treatments as shown:

Treatment	Fertilizer Nitrogen	*Rhizobium* Inoculation
F1	–	–
F2	+	–
F3	–	+
M1	–	–
M2	+	–
M3	–	+

These treatments are incorporated into a randomized complete block design with four replications. It was felt that the likelihood of the loss of any given plot in most experiments was such that fewer than four replications or use of an incomplete block design was not wise. Cooperators are urged to conduct the "A" Experiment for at least three years at each location, but on different plots. Because of variation between years, any field experiment conducted for only a single season can yield unwarranted conclusions. Cooperators are also urged to conduct follow-up experiments on the same plots to study the course of introduced *Rhizobium* strains over time. The "A" Experiment should be continued even if the first year's data indicate the desirability of going on to the "B" and "C" Experiments.

Plot size is 2.4 m × 7.5 m, so one "A" Experiment requires an area of 15 m × 28.8 m (0.04 ha). Early harvest is taken from 1 m of length, and 5 m is retained for final yield data. The remaining 1.5 m gives 0.5 m borders at the ends and between the early harvest and final harvest area. Side borders are 0.6 m on each side. Where an institution has standardized on a different plot size, the plot size can be modified without interfering with the integrity of the design. However, treatment information given in the instruction booklet will then have to be recalculated, and the cooperator must indicate the actual plot size employed on the data sheets.

The "A" Experiment was purposely kept simple, with treatments at a minimum, so that it could be managed by cooperators with little or no previous experience in handling either *Rhizobium* or field experiments. The wisdom of this decision has been borne out since many cooperators have had experience in either field or laboratory research, but not with both. Also, size of the experiment was kept small so that it would not be overly burdensome for institutions with limited resources, since success of the experiment necessitated the maximum number of experiments. If meaningful composite analyses are to be conducted for a country or geographical region, the trials must be extensive. One location intensively studied would only give much data for that one location. To make inferences about response over a wide area requires an extensive series of experiments.

The randomized complete block design allows for expansion by individual cooperators. If the cooperator desires additional treatments of special interest to himself, these can be incorporated. Care, of course, must be taken not to eliminate any of the six basic treatments and the plots must be rerandomized. Good management practices, such as proper tillage and weed control operations, timely irrigation in non-rainfed experiments, disease and pest control, use of adapted crop varieties, adherence to proper planting dates, making good field notes, timely harvest, and proper recording of data, are essential.

NifTAL supplies each cooperator with a combination field methods–data record booklet. This is designed so that anyone can follow the directions stepwise and conduct a successful experiment. The data sheets are in triplicate so that if data are recorded directly, the cooperator will have one copy for himself, one to send to NifTAL, and one for any other purpose such as the Coordinator of an in-country network or the parent institution in the case of a field station.

NifTAL supplies a three-strain inoculant composed of the best available strains. These strains are selected for nitrogen fixation efficiency, diverse origin, and serological distinctness. If the cooperator has already developed a high quality inoculant

using his own strains he can substitute this product. Integrity of the experiment only requires use of a high quality inoculant, that is, one with sufficient organisms per gram to insure a sufficient number of viable *Rhizobium* is available to each germinating seedling. However, generally there will be an advantage to using the NifTAL-supplied inoculant since NifTAL also supplies typing antisera for its strains. Most cooperators have elected to use the NifTAL-supplied inoculant in the initial experiments. If there is no response to inoculation over the control, one wants to know why. Such a lack of response is often due to native soil organisms being more competitive in occupying nodule sites. Typing nodules can give a good indication whether or not this is the situation. It is also useful to type nodules in follow-up experiments to monitor establishment and survival of strains in inoculant.

The most important observation taken in the "A" Experiment is the yield of final product—grain yield, forage yield, wood yield, etc. However, recording of other observations that could help explain the results obtained is encouraged. Most useful are nodule data and plant dry weight at early harvest. The nodule data give an indication of the abundance and efficiency of soil organisms, and the dry weight gives an indication of the inoculant effect if a crop failure precludes taking final yield data. Space for recording data on the most likely characteristics is provided on the data sheets given to the cooperators.

All measured characteristics are analyzed by NifTAL. A copy of the analysis is supplied to the cooperator and the data are stored for later composite analyses. The entire procedure is shown diagrammatically in the flow chart in Figure 1.

Discussion of Design Alternatives

The number of replications could have been decreased or increased. A simplification would have been to have had only three replications, as statistically valid data could theoretically be obtained with only three replications. However, personal experience and consultation with other workers who had conducted field research in the tropics indicated the need for four replications. The possibility in many tropical areas of losing some plots, if not an entire replication, due to factors beyond the control of the research worker, indicated the need to start with at least four replications. This might be construed as justification for even more than four replications. However, adding too many replications would increase the experimental size to such an extent that some potential cooperators would be eliminated. Therefore, four replications represented a good compromise.

One treatment often employed in legume inoculation trials is starter nitrogen. Some workers like to supply a small amount of nitrogen to help the seedlings until their nodules start fixing nitrogen. In fact, some workers use "starter" doses of nitrogen sufficient to supply the need of the entire crop. Use of starter nitrogen is still a debatable subject among legume workers, and consideration of all of its ramifications is outside the scope of this paper. We do not believe use of starter nitrogen is needed or advisable. Whether or not the above opinion is correct, adding such a treatment here would not contribute anything towards answering the primary question about the need for inoculation. Therefore, it would be an irrelevant treatment, yet it would either increase the size of the experiment or decrease precision of the data collected.

The use of an incomplete block design would allow addition of treatments without increasing the size of the experiment or

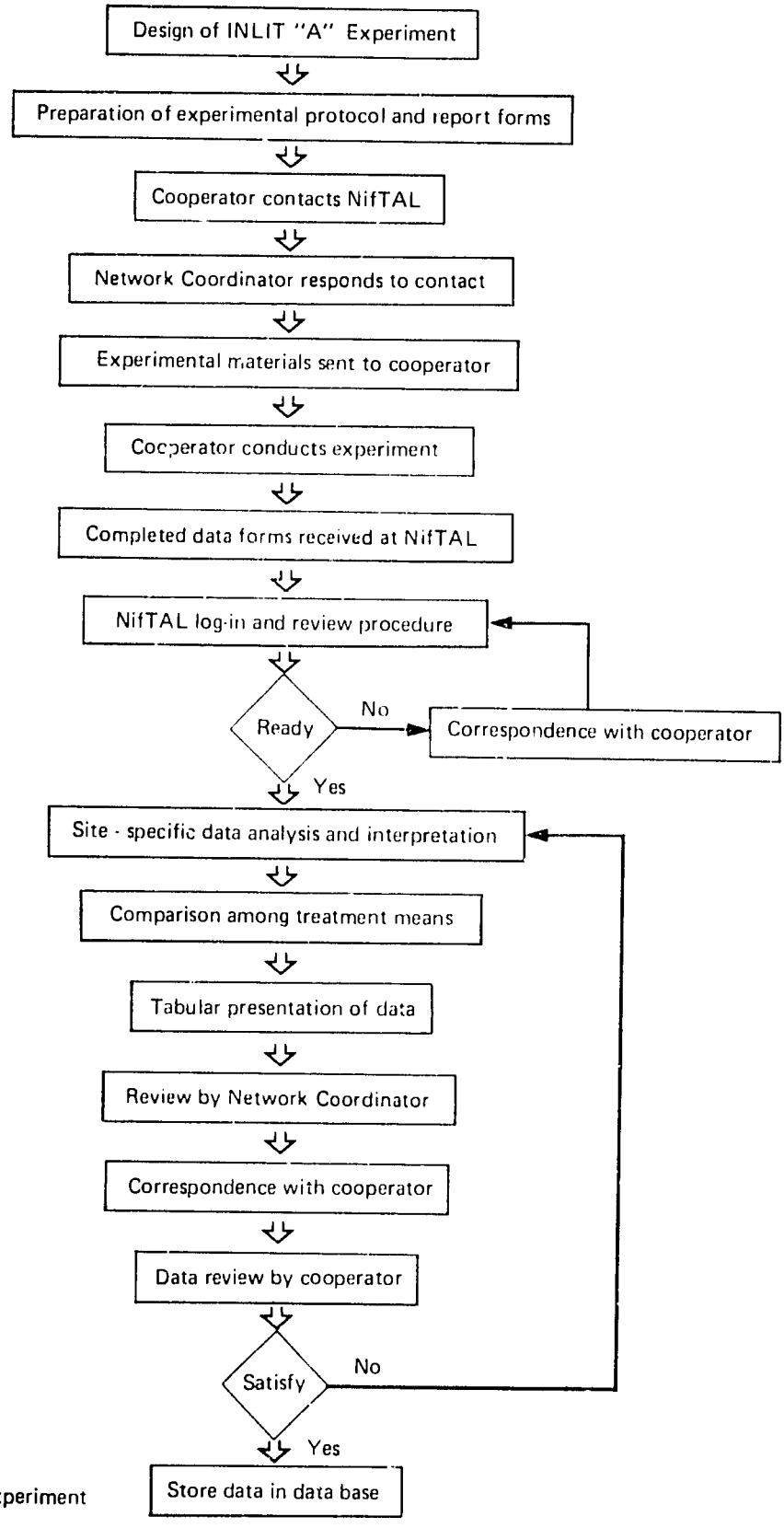

Figure 1.
Flow chart for INLIT "A" experiment

Figure 1. Flow chart for an INLIT "A" Experiment.

decreasing the number of replications, thus increasing the amount of information obtained. However, there are no other treatments that would contribute directly to answering the primary question. Additional treatments could be assigned to provide ease of data analysis and interpretation through elimination of nonorthogonality. But we believe the purpose of field experimentation is to get maximum usable data in the field, not ease of data manipulation. The moot point, with respect to keeping a small experiment that could be handled by the maximum number of cooperators, would not be whether to add treatments with attendant partial replication, but rather whether to reduce the number of replications.

The experimental treatments appear to lend themselves to a split plot design. A valid experiment could have been designed with split plots, the main plots being either fertility or nitrogen treatments. Generally, a split plot design is utilized when field operations either require or are facilitated by use of such a design. Since such is not the situation with the "A" Experiment, there would have been no operational advantage to utilizing split plots. Putting nitrogen treatments in the main plots would cause loss of precision in comparing treatments in the main plots. Since precision is gained in the subplots comparisons, a case could be made for putting nitrogen treatments there, although there would have been no operational advantage. However, this would have eliminated the flexibility of the present design, which allows those cooperators with sufficient reason and resources to incorporate treatments of special value to themselves.

III. INTRASITE ANALYSIS OF AN INLIT EXPERIMENT

The "A" Experiment has six treatments, which include both *Rhizobium* and fertilizer (bag N) sources of nitrogen along with a control treatment at each of two levels of fertility. These treatments can be displayed as:

	Nitrogen Source		
Fertility	**Control**	**Bag N**	***Rhizobium***
Farm	F1	F2	F3
Maximum	M1	M2	M3

Inoculation Questions of Interest

In the analysis of the data we have chosen five major comparisons among the treatment means that are pertinent to the main question asked in this experiment (Table I): Does legume inoculation enhance plant growth and yield? The two most important aspects of the main question are:

(a) Is the average of *Rhizobium* inoculation treatments (F3 and M3) different from the average of the control treatments (F1 and M1)?

(b) Is the average of the *Rhizobium* inoculation treatments (F3 and M3) different from the average of the bag nitrogen treatments (F2 and M2), i.e., is there a nitrogen source effect?

Interpretation of these questions assumes that the crop's response to the inoculation treatment is not affected by the fertility status. To check the validity of this assumption two additional questions need to be considered.

(c) Is the magnitude of the inoculation response relative to the control the same for both levels of fertility?

(d) Is the magnitude of the inoculation response relative to the bag nitrogen the same for both levels of fertility?

If the magnitude of the inoculation response relative to the control (or bag N)

Table 1. Treatment comparisons

Description	Treatment Comparison
1) INOCULATION BY FERTILITY INTERACTION	$\frac{(M3-M1)}{2} - \frac{(F3-F1)}{2}$
If interaction is not important, go to 2.	
If interaction is important, go to 2A and 2B.	
2) INOCULATION VS. CONTROL	$\frac{(M3+F3)}{2} - \frac{(M1+F1)}{2}$
A) EFFECT OF INOCULATION AT FARM FERTILITY	F3 - F1
B) EFFECT OF INOCULATION AT MAXIMUM FERTILITY	M3 - M1
3) N SOURCE BY FERTILITY INTERACTION	$\frac{(M3-M2)}{2} - \frac{(F3-F2)}{2}$
If interaction is not important, go to 4.	
If interaction is important, go to 4A and 4B.	
4) N SOURCE: INOCULATION N VS. BAG NITROGEN	$\frac{(M3+F3)}{2} - \frac{(M2+F2)}{2}$
A) EFFECT OF N SOURCE AT FARM FERTILITY	F3 - F2
B) EFFECT OF N SOURCE AT MAXIMUM FERTILITY	M3 - M2
5) MAXIMUM VS. FARM FERTILITY WITH INOCULATION	M3 - F3

depends on the fertility, we say that the inoculation response interacts with the fertility. If either interaction is present, then the comparison of the inoculation responses relative to the control (or bag N), averaged over fertility levels (main effects), can be misleading. In this case, we should examine the inoculation effects separately at each fertility level, farm and maximum.

The last main query of the "A" Experiment concerns only *Rhizobium* at differing levels of fertility.

(e) Is the yield of the *Rhizobium* inoculation treatment the same at each fertility level?

Comparisons Among Treatment Means

These queries can be formulated as linear contrasts among the treatment means. For example, the first question can be stated (using the previous treatment notation for convenience) as

"Is (F3 + M3) 2 - (F1 + M1) 2 = 0 ?"

Similarly, each of the other questions can be formulated to ask whether a linear contrast among the treatment means is equal to zero. These contrasts, hereafter called treatment comparisons, are given in Table 1.

The treatments have been selected for certain designed comparisons among the treatments. If the interaction comparisons (comparisons 1 and 3) are judged to be not important, then treatment differences can be attributed to treatment comparisons 2 and 4 (often referred to as the inoculation and N source main effects). The comparison of maximum and farm fertility levels with inoculation (comparison 5) completes the set of planned comparisons among the treatment means. If the interactions are judged to be important, then comparisons 2A, 2B, 4A, and 4B (often referred to as the inoculation and N source simple effects) should be examined. Since the set of comparisons used in the interpretation depends on the evaluation of the interaction comparisons as the first step, all nine comparisons are calculated during the data analysis. In practice, however, only comparisons 1, 2, 3, 4, and 5 are utilized if both interaction comparisons are judged to be not important. If the interactions are important, the simple effect comparisons, 2A, 2B, 4A, and 4B, replace comparisons 2, 4, and 5.

Statistically, we will run the t test to evaluate whether a treatment comparison, estimated from the observed treatment means, is significantly different from zero. In order to construct this test, we first determine the standard error of the estimated treatment comparison. For example, from the coefficients of treatment means that appear in the treatment comparison *Rhizobium* vs. control, namely, (F3 + M3) 2 and −(F1+M1)/2, the standard error of the estimated treatment comparison is the square root of

$$[(1/2)^2 + (1/2)^2 + (-1/2)^2 + (-1/2)^2]\,\sigma^2/r$$

where σ^2 denotes the experimental error of the experiment and r is the number of replications. An estimate of σ^2 can be found as the experimental error from the analysis of variance table. Similarly, the standard error can be calculated for each treatment comparison. Then, for each experiment, the major questions regarding *Rhizobium* inoculation would be answered.

Glycine max (Soybean) Case Study

Soybean grain yields from experiment AS 620, a randomized complete block design experiment with four replications, will be used as a case study of an intrasite data analysis.

Output 1 gives the analysis of variance and other summary statistics. The model

EXPERIMENT NO.: AS620
NAME OF COOPERATOR: J.R. LOCKMAN
INSTITUTION: TSAAP, PAKISTAN
LOCATION: LAHORE, PAKISTAN
CROP: SOYBEAN, GLYCINE MAX
VARIETY: WILLIAMS
SOWING DATE: MARCH 4, 1981

ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: SEED_YLD SEED YIELD (KG/HA)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PR > F	R-SQUARE	C.V.
MODEL	8	7368172.83333333	921021.60416667	6.05	0.0014	0.763515	15.8864
ERROR	15	2282160.12500001	152144.00833333		STD DEV		SEED_YLD MEAN
CORRECTED TOTAL	23	9650332.95833334			390.05641686		2455.29166667

SOURCE	DF	ANOVA SS	F VALUE	PR > F
BLOCK	3	97807.12499999	0.21	0.8849
TREATMTS	5	7270365.70833334	9.56	0.0003

Output 1. Analysis of variance table and other summary statistics for an INLIT "A" Experiment (intrasite analysis).

sum of squares includes (a) the block (BLOCK) sum of squares with 3 degrees of freedom, and (b) the treatment (TREATMTS) sum of squares with 5 degrees of freedom. The error mean square with 15 degrees of freedom is the estimate of the experimental error and it is used for evaluating the importance of the treatments (F value with 5 and 15 degrees of freedom). The value under PR > F is the probability of an observed value larger than the calculated value on chance alone. The F value of 9.56 indicates there are treatment differences that are statistically significant at the 1 percent level; that is, there is less than one chance per hundred that these effects were due to chance alone rather than to the effect of the treatments. The actual probability in this case was 0.0003, or three chances out of ten thousand of the effect being due to chance alone.

Included in Output 2 with each of the nine treatment comparisons are the estimate of the comparison, the standard error of the estimate, the 95 percent confidence interval, the *t* statistic, the PR > *t*, the probability of an observed value larger than the absolute value of the calculated *t* on chance alone. This probability is called the significance level and is associated with a two-tailed alternative hypothesis. If the experiment were repeated a large number of times and a 95 percent confidence interval calculated for each experiment, then we would expect 95 percent of these calculated confidence intervals to include the value of the true difference of the comparisons. The estimate, the standard error of the estimate, and the lower and upper limits of the 95 percent confidence interval are in the same units as the unit of measurement, i.e., kg/ha for grain yield. The *t* statistic for each comparison is the ratio of the estimate to the standard error of the estimate.

Judging the importance of the two interactions (comparisons 1 and 3) can be based on the magnitude and significance level of the *t* statistic or the width of the confidence interval. Quick procedures for evaluating the importance of an interaction include determining whether the confidence interval includes zero and whether the *t* value has a significance level greater than 0.05. If either of the above is true, then the interaction is judged not to be important. Remember, however, that the width of the interval depends not only on the estimate and the standard error but also on the level of confidence.

If there is an inoculation by fertility level interaction, then the inoculation effect is not the same at both fertility levels and one should look to see at what fertility level the effect is more important. Likewise, with the N source by fertility interaction, one should ascertain that the interaction is important, and then go on to investigate the N source difference at both farm and maximum levels of fertility.

The two main effects, inoculation and N source, and the maximum vs. farm fertility with inoculation comparison, are self-explanatory and are the pertinent questions with respect to the main question: whether or not, with the present state of the art, legume inoculation affects yield. These comparisons will indicate whether or not inoculation increases yields, whether higher yields can be obtained with inoculation than with fertilizer nitrogen or vice versa, and, if inoculation increases yields, whether it does so at both high and low fertility levels.

In Output 3, the treatment yields are printed out by replication along with the means and standard errors (within parentheses). Once it has been determined that certain comparisons are important, one can go to the table of means and see the actual magnitude of treatments, N source,

EXPERIMENT NO.: AS620

SEED YIELD (KG/HA)

COMPARISONS AMONG TREATMENT MEANS

	COMPARISON	ESTIMATE	STANDARD ERROR OF ESTIMATE
1:	INOCULATION BY FERTILITY INTERACTION	-125	195.03
2:	INOCULATION VS. CONTROL	1167	195.03
2A:	EFFECT OF INOCULATION AT FARM FERTILITY	1292	275.81
2B:	EFFECT OF INOCULATION AT MAXIMUM FERTILITY	1042	275.81
3:	N SOURCE BY FERTILITY INTERACTION	379	195.03
4:	N SOURCE: INOCULATION N VS. BAG N	922	195.03
4A:	EFFECT OF N SOURCE AT FARM FERTILITY	543	275.81
4B:	EFFECT OF N SOURCE AT MAXIMUM FERTILITY	1301	275.81
5:	MAXIMUM VS. FARM FERTILITY WITH INOCULATION	303	275.81

	COMPARISON	95 % CONFIDENCE INTERVAL LOWER LIMIT	95 % CONFIDENCE INTERVAL UPPER LIMIT	T FOR HO: PARAMETER=0	PR > |T|
1:	INOCULATION BY FERTILITY INTERACTION	-541	290	-0.64	0.5304
2:	INOCULATION VS. CONTROL	751	1583	5.98	0.0000
2A:	EFFECT OF INOCULATION AT FARM FERTILITY	704	1880	4.69	0.0003
2B:	EFFECT OF INOCULATION AT MAXIMUM FERTILITY	454	1630	3.78	0.0018
3:	N SOURCE BY FERTILITY INTERACTION	-36	795	1.94	0.0708
4:	N SOURCE: INOCULATION N VS. BAG N	506	1338	4.73	0.0003
4A:	EFFECT OF N SOURCE AT FARM FERTILITY	-45	1131	1.97	0.0679
4B:	EFFECT OF N SOURCE AT MAXIMUM FERTILITY	713	1889	4.72	0.0003
5:	MAXIMUM VS. FARM FERTILITY WITH INOCULATION	-285	891	1.10	0.2889

Output 2. Estimates, standard errors, confidence intervals, and significance tests for planned comparisons among treatment means for an INLIT "A" Experiment (intrasite analysis).

or fertility means.

The individual treatment comparisons focus on where the yield increases (or decreases) occur. Referring again to the outputs, the treatment comparisons (Output 2) and means (Output 3) enable one to evaluate effects. Treatment comparison 1 indicates that the inoculation by fertility level interaction is not important; that is, the magnitude of the increase due to inoculation is essentially the same at each level of fertility. That the interaction is not important is indicated by the fact that zero is bracketed by the two 95 percent confidence limits (−541, +290). The *t* statistic is another way of looking at the same thing. The *t* of −0.64 indicates a probability of 0.5304. In other words, there is about a 50 percent chance of getting this value without regard to treatment effects, the same probability as with chance alone. Since this interaction is not important, the simple effects 2A and 2B can be ignored. Sometimes, however, whether or not a comparison is important will not be as obvious.

Treatment comparison 2, inoculation vs. control, is another easily determined case. The output shows that the 95 percent confidence interval is nowhere close to zero and that the *t* value has a probability of 0.0000. That is, there is less than one chance in ten thousand of this value being obtained by chance alone. Therefore, there is an effect due to inoculation. The estimated magnitude of this effect is an increase of 1167 kg/ha with a standard error of 195 kg/ha. From the 95 percent confidence interval, we would expect the increase to be between 751 and 1583 kg/ha.

Whether fertilizer nitrogen and *Rhizobium* inoculation have the same effect at both fertility levels (treatment comparison 3) is not quite as obvious as treatment comparisons 1 and 2 above. The 95 percent confidence interval contains zero, but zero is close to the lower limit. The *t* statistic has a probability of 0.07, which is relatively close to the 0.05 level generally used as a cutoff point in field research. In this case, if one chose the 90 percent confidence limit for assaying importance, the zero would be outside the confidence limits (the actual significance level of the *t* test is 0.0708) and the comparison would be deemed important. In this situation, the researcher would go on to treatment comparisons 4A and 4B (N source simple effects).

Here, with treatment effect 4A, going through the same reasoning process used above, one sees that most of the effect is at the maximum fertility level. The 95 percent confidence interval encompasses zero for the farm fertility level, but just barely. The probability of the *t* value being due to chance alone is 0.07 (0.0679). If one were looking for any trends that might be there, one might accept a 7 percent level of significance as important in this situation, and estimate the effect to be 543 kg/ha.

With treatment effect 4B, the N source simple effect at the maximum fertility level accounts for most of the effect due to N source. The 95 percent confidence interval falls nowhere near zero and the probability of the *t* value being due to chance is 0.0003 (three times out of ten thousand). The actual effect here was 1301 kg/ha; stated in other words, one could expect to get, at 95 percent confidence, an increase compared with bag nitrogen somewhere between 713 and 1889 kg/ha by practicing *Rhizobium* inoculation at the maximum fertility level.

At this point one should avoid the temptation to attempt conversion of the above yield increase into equivalent pounds of nitrogen fixed by the crop. The increased yield of the inoculated plots over the bag nitrogen plots could be the result of insufficient N being applied to attain maximum yield, but this is unlikely here since instruc-

EXPERIMENT NO.: AS620

SEED YIELD (KG/HA)

	NITROGEN		
	CONTROL	BAG N	RHIZOBIUM
FERTILITY			
	F1	F2	F3
FARM	1052	2756	2959
	1811	2211	3017
	2211	2581	2841
	1757	2281	3183
TREATMENT MEANS:	1708 (195)	2457 (195)	3000 (195)

FARM FERTILITY MEAN: 2388 (113)

	M1	M2	M3
MAXIMUM	2957	2453	3038
	2250	1943	3136
	1951	2087	3328
	1888	1525	3711
TREATMENT MEANS:	2262 (195)	2002 (195)	3303 (195)

MAXIMUM FERTILITY MEAN: 2522 (113)

CONTROL MEAN	BAG N MEAN	RHIZOBIUM MEAN
1985 (138)	2230 (138)	3152 (138)

OVERALL MEAN: 2455 (80)

*** IN PARENTHESES ARE THE STANDARD ERRORS ***

Output 3. Plot yields, treatment means, fertility means, nitrogen means, overall mean, and standard errors for an INLIT "A" Experiment (intrasite analysis).

tions call for applying enough for maximum yield. It could be the result of the nitrogen leaching or otherwise being made unavailable to the crop. Or it could indicate that the crop was unable to take up adequate amounts (quantities) of mineral nitrogen under the prevailing conditions.

If, when evaluating comparison 3, the interaction had not been deemed important, one would have considered only comparison 4. Comparison 4 is obviously important. The 95 percent confidence interval is far from zero and the probability of the t value being due to chance alone is again 0.0003, or a probability of three in ten thousand. Therefore, evaluating this main effect, the estimate ± 1 standard error gives a yield increase of 922 ± 195 kg/ha, or an increase somewhere between 727 and 1117 kg/ha due to *Rhizobium* inoculation over fertilizer nitrogen. (This interval, ±1 standard error, is roughly a 67 percent confidence interval.)

Comparison 5, the maximum vs. farm fertility effect with *Rhizobium*, is obviously not important. The 95 percent confidence interval includes zero (although zero is well off center) and the probability of the t value being due to chance alone is 0.3 (three times out of ten). Therefore, one can conclude that under the conditions of this experiment inoculation was equally effective at both fertility levels.

IV. NONORTHOGONAL COMPARISONS AMONG TREATMENT MEANS

The major questions posed in an INLFT experiment lead to a nonstandard analysis of the data. In particular, the treatment comparisons outlined in Table 1 form a nonorthogonal set of treatment comparisons. Table 2 displays the coefficients of the six treatments for two comparisons given in Table 1, inoculation vs. control (comparison 2) and inoculation N vs. bag N (comparison 4).

Each comparison sums to zero (the definition of a contrast), but when the sum of cross-products is not zero, the two comparisons are said to be nonorthogonal. Two immediate ramifications of nonorthogonal treatment comparisons arise:

(a) When each of the five major treatment comparisons from Table 1 is evaluated separately, as in Output 2, the five associated sums of squares will not add to the treatment sum of squares with 5 degrees of freedom given in Output 1. If the treatment comparisons are evaluated sequentially, the interpretation of the treatment comparison can change.

(b) The experiment-wise error rate, i.e., the probability of declaring at least one treatment comparison different from zero when, in fact, there are no differences among the six treatment means (Chew, 1980), is impossible to calculate analytically and consequently is unknown. Only by computer simulation studies can the experiment-wise error rate be even approximated.

If the standard deviation, σ, was known and each test of five *orthogonal* comparisons was run at $\alpha = 0.05$, a false rejection of one contrast would not affect possible false rejections of the other contrasts. Thus the experiment-wise error rate would be $[1 - (1.0 - 0.05)^5] = 0.23$. This follows since the orthogonality of comparisons and the assumption of normal experimental error, which is assumed here, jointly imply that the estimated treatment comparisons are statistically independent. However, the five treatment comparisons of interest in this investigation are not orthogonal and, hence,

Table 2. Comparison coefficients for comparisons 2 and 4

	Comparison Coefficients						
	F1	F2	F3	M1	M2	M3	Sum
Comparison 2	-0.5	0.0	0.5	-0.5	0.0	0.5	0.0
Comparison 4	0.0	-0.5	0.5	0.0	-0.5	0.5	0.0
Product	0.0	0.0	0.25	0.0	0.0	0.25	0.5

are not independent. The correlation between any pair of comparisons is equal to the sum of the cross-products of the comparison coefficients divided by the square root of the product of the two sums of squares of the comparison coefficients. For example, the correlation between comparisons 2 and 4 is equal to 0.25 + 0.25 divided by the square root of (1)(1). All the correlations may be displayed as shown in Table 3, where 0.707 is the reciprocal of the square root of 2. From consideration of the table rows (or columns) we note that each treatment comparison is correlated with one or two other treatment comparisons. The interaction comparisons 1 and 3 are correlated with each other and with comparison 5. The main effect comparisons 2 and 4 are correlated with each other, while comparison 5 is correlated with the two interaction comparisons. Note that all the correlations in the table are positive.

This implies that if one falsely rejects a contrast being equal to zero and then considers another contrast with which it is positively correlated, the probability of also rejecting the latter contrast is greater than the quoted significance level.

Here we consider the probability of falsely rejecting comparison 2 when we have already falsely rejected comparison 4 and, hence, calculate the experiment-wise error rate considering only these two contrasts. This calculation is based on the multivariate Student t distribution.

The multivariate version of the Student t distribution must be used because the five contrasts are nonorthogonal and, hence, are correlated; and in reality, we must estimate the experimental error. Using computer simulation the procedure is to estimate the proportion of joint rejections of the comparisons for a given possible α = 0.05 critical point for the individual t tests. This was done using the GGMNR subroutine of IMSL (International Mathematical and Statistical Library). Twenty thousand sets of five independent normal variables, each with mean zero and variance = 1/4, were generated. For each set of six variables, the five contrasts were calculated and the corresponding t tests were computed. The estimate of σ^2 was randomly generated as the square root of a chi-squared variable with 15 degrees of freedom. The five t values then have the same multivariate t distribution as would the five t values arising in the analysis of an experiment.

In particular, for comparisons 2 and 4, if comparison 4 is falsely rejected, then the probability of falsely rejecting comparison 2 is 0.211. In the orthogonal case, this value equals 0.098. One might think that the latter value should be 0.05 for α =0.05. However, since the experimental error must be estimated, the t tests are not orthogonal. The case of known experimental error was also considered in the same simulation. Here the corresponding probabilities were 0.18 and 0.06. The latter value is slightly off from the expected value of α = .05, but within acceptable simulation error. Similar results were found for the other pairs of comparisons.

The second major ramification of the nonorthogonality of the five major treatment comparisons in Table 1 is that the sum of squares for the individual contrasts do not add to the treatment sum of squares with 5 degrees of freedom. In particular, the sums of squares for a contrast comparison may be computed as:

$$SS\ (\text{comparison}) = t^2\ (\text{Exptl. Error M.S.})$$

where t is the ratio of the estimated comparison to the standard error of the comparison and M.S. is mean square.

In this case, one cannot interpret the sums of squares for the comparisons as a partitioning of the treatment sum of

Table 3. Correlations among treatment comparisons

Treatment Comparison No.	Treatment Comparison 1	2	3	4	5
1	1	0	0.5	0	0.707
2	0	1	0	0.5	0
3	0.5	0	1	0	0.707
4	0	0.5	0	1	0
5	0.707	0	0.707	0	1

squares, which represents the total amount of variability in the treatment means.

The problems encountered in interpretation are especially evident if the analysis is done using a computer regression program; e.g., the regression procedure of the Statistical Analysis System (SAS, 1982). One method for computing the sums of squares due to our five comparisons would be to enter the comparisons directly as independent variables in a regression program (Allen and Cady, 1982). For example, the treatment coefficients for comparison 1, as shown in Table 1, are 0.5, 0.0, -0.5, -0.5, 0.0, and 0.5 for the F1, F2, F3, M1, M2, and M3 treatments, respectively. These coefficients then become the values of the independent variable X1 in a regression program. Four additional independent variables, X2, X3, X4, and X5, can also be constructed from Table 1 and displayed as in Table 4.

The set of values in this display represents one replication of the experiment and must be repeated three additional times for the analysis of the 24 observations. The regression analysis would then be run with these five X-variables as the independent variables and the 24 measured yield responses as the dependent variables. If the five treatment comparisons were orthogonal, the order of the five X-variables would not make any difference. The analysis could be done one X-variable at a time or all five X-variables together in a multiple regression analysis. The estimated comparisons and the associated sums of squares would not depend on the sequencing of the X-variables. In regression terminology, the sequential and partial regression coefficients (the estimated comparisons) would be the same. The sequential and partial sums of squares would also be the same (Allen and Cady, 1982).

With a nonorthogonal set of comparisons, however, the ordering (sequencing) of the five X-variables is important. For example, with the ordering X1, X2, X3, X4, X5, the sequential sums of squares (Type I sums of squares in SAS) for X1 and X2 are the sums of squares associated with the comparisons as we would expect from Table 1, namely, the inoculation by fertility interaction and the inoculation vs. control

Table 4. Regression variable representation of the five major comparisons shown in Table 1

Treatment	Independent Variable X1	X2	X3	X4	X5
F1	0.5	-0.5	0.0	0.0	0.0
F2	0.0	0.0	0.5	-0.5	0.0
F3	-0.5	0.5	-0.5	0.5	-1.0
M1	-0.5	-0.5	0.0	0.0	0.0
M2	0.0	0.0	-0.5	-0.5	0.0
M3	0.5	0.5	0.5	0.5	1.0

comparisons. However, the sequential sum of squares for X4, after X1 and X2, would be the sum of squares for another comparison that compares the average of control and *Rhizobium* with bag N. Similarly, X3, after X1 and X2, is the interaction of this comparison with fertility level. Recall that comparison 1 was set up to compare bag N and *Rhizobium*. However, any regression program will automatically orthogonalize, or adjust, each comparison to the previous comparisons.

The comparisons that result from an ordering of X1, X2, X3, X4, X5 are the familiar analysis-of-variance orthogonal comparisons for a 2 × 3 factorial treatment design, i.e., (1) the interaction between *Rhizobium* vs. control and farm vs. maximum fertility; (2) the main effect of *Rhizobium* vs. control; (3) the interaction between the average of *Rhizobium* and control vs. bag N and farm vs. maximum fertility; (4) the main effect of the average of *Rhizobium* and control vs. bag N; and (5) the main effect of fertility level. Even though these are a straightforward set of comparisons and are orthogonal, they do not strike at the main questions of the INLIT investigation and we forego them in favor of the nonorthogonal set of comparisons.

Finally, we should note that the sum of squares for comparison 3 and comparison 4 can be computed in a regression analysis from the ordering X3, X4, X1, X2, X5. The sequential sum of squares for X2, after X3 and X4, would be the sum of squares for the comparison between the average of *Rhizobium* and bag N vs. control; and X1, after X3 and X4, would be the interaction of this comparison with fertility level.

V. MULTILOCATIONAL ANALYSIS OF INLIT DATA

In the INLIT, the same experiment is carried out at several locations and is also repeated over several seasons. After an intrasite data analysis emphasizing the five basic treatment comparisons at each environment (an experiment during a given season at a given location) we now want to combine the data from all the environments into one multilocational analysis. Note that the terms environment and experiment, and also site when referred to as intra- or intersite, are used interchangeably.

Each experiment reflects conditions of its particular environment. For example, previous inoculation and low pH are soil conditions reflecting the particular environment, excess rainfall is a particular weather condition, and good weed control is a particular experimenter condition. Consequently, there is motivation to investigate the consistency of our treatment comparisons over the environments to understand that treatment comparisons can vary from environment to environment. For example, do all environments in the network show a similar response to inoculation?

A basic motivation for a multilocational analysis is the conceptual need to make inferences to a larger region than that represented by a series of individual localities. Each environment is envisioned as a sample from a population of environments. Then the environment is the basic experimental unit and is considered a replication of all the experimental conditions affecting the treatments at an individual environment. The results of the multilocational data analysis need to be communicated to all the cooperators to have an exchange of information through the network linkages.

INLIT multilocational data analysis is developed in the remainder of this section, starting with the preliminary combined data analysis, followed by the environment means data analysis. The entire procedure is shown diagrammatically in Figure 2.

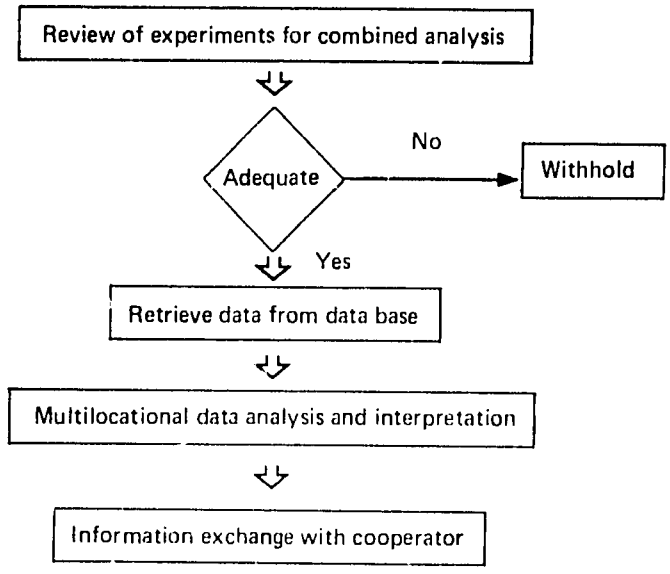

Figure 2. Flow chart for multilocational analysis of INLIT data.

Preliminary Combined Data Analysis

General information on the analysis of a series of experiments is given by Cochran and Cox (1957). Though certain aspects of data analysis have changed in recent years, e.g., the emergence of statistical computer packages, the principles and general approach of combining experiments given in Chapter 14 of Cochran and Cox are still useful for experimental networks.

Major objectives of the preliminary combined data analysis include:

(a) Evaluation of the consistency of individual environment treatment comparison estimates across the environments.

(b) Estimation of treatment comparisons averaged over the environments.

The two objectives have to be approached sequentially. If the consistency evaluation shows that a particular treatment comparison is not consistent over the environments, i.e., an interaction exists between the treatment comparison and environments, then the interpretation of the treatment comparisons averaged over the environments is not clear.

Each environment has a data analysis structure for assignment of yield variability to certain sources. The intrasite data analysis developed previously in Section III is summarized in Table 5 as the basis for the preliminary combined data analysis.

When intrasite data analyses are combined into a multilocational data analysis, two new components have to be considered. The average yield for each environment varies; consequently, part of the yield variability, and at times a major factor in accounting for yield variability, is the variation among the environment average yields.

Table 5. Summarization of intrasite data analysis

Source of Variation	Degrees of Freedom
Blocks (Reps)	3
Treatments	
Inoculation by fertility interaction (COMP 1)	1
Inoculation vs. control (COMP 2)	1
N source by fertility interaction (COMP 3)	1
Inoculation N vs. bag N (COMP 4)	1
Maximum vs. farm fertility with inoculation (COMP 5)	1
Experimental Error	15

Secondly, and the major reason for the preliminary combined data analysis, is the recognition that the yield response for each treatment comparison might not remain the same over all the environments. The analysis will need to include five intersite interaction terms between each treatment comparison and environments.

In addition to the two new components, the three original sources of variation of each intrasite analysis need to be included in the combined analysis. These are blocks, treatment comparisons, and experimental error. As expected, the treatment comparisons will represent the average effect of the comparisons over all the environments since each treatment comparison is exactly the same comparison at each environment. However, each block at each environment is not necessarily the same identity. Consequently, the blocks sum of squares from each environment are added or pooled over the environments; in data analysis terminology, the blocks are nested within the environments. The degrees of freedom are calculated as the sum of the blocks degrees of freedom from the intrasite analyses. If each experiment has four blocks, then the degrees of freedom in the combined analysis are the product of 3 and the number of environments. Similarly, the experimental errors for the intrasite analyses are pooled together in the combined analysis. Then, the analysis of variance table for the preliminary combined data analysis with p environments and four blocks for each environment is shown in Table 6. Hereafter, the interactions between a treatment comparison and environments will be called intersite interactions. For example:

COMP 1 is the *intrasite* interaction between inoculation and fertility.

The COMP 1 by environments interaction will be known as the COMP 1 *intersite* interaction.

Two criticisms of this preliminary combined data analysis are explored in detail by Cochran and Cox (1957), namely, heterogeneity of the experimental error variances and heterogeneity of the treatments by environments interaction with 5(p-1) degrees of freedom.

With INLIT, or with international experimental networks in general, the experiments cover a wide range of environmental conditions and are carried out by researchers who have different backgrounds and research experiences. Despite a common experimental protocol, the assumption of homogeneity of all the intrasite experimental errors is not realistic and the use of the pooled experimental error as an estimate of σ^2 can be questioned. Statistical theory of significance tests indicates that probability levels should not be considered as exact probabilities. Unfortunately, the extent of deviation from the stated significance levels is usually unknown. Fortunately, the pooled experimental error will be used only as an indicator of treatment comparison intersite interactions. Adherence to exact probability levels will not be necessary. Furthermore, the final multilocational analysis will be modeled so that dependence on the pooled experimental error will be lessened.

The second criticism, heterogeneity of the treatments by environments interaction, also needs to be handled in the INLIT analysis. First, the INLIT treatments are immediately set up as five treatment comparisons, and the treatments by environments interaction also would be analyzed as five potentially heterogeneous interaction terms. In fact, heterogeneity of the interaction terms can work to our advantage. Consider the following scenario. The COMP 2 (inoculation vs. control) intersite inter-

Table 6. Analysis of variance table for the preliminary combined data analysis

Source of Variation	Degrees of Freedom
Environments	(p-1)
Blocks within environments	3p
Treatments	
Inoculation by fertility interaction (COMP 1)	1
Inoculation vs. control (COMP 2)	1
N source by fertility interaction (COMP 3)	1
Inoculation N vs. bag N (COMP 4)	1
Maximum vs. farm fertility with inoculation (COMP 5)	1
Treatments by environments	
COMP 1 by environments interaction	(p-1)
COMP 2 by environments interaction	(p-1)
COMP 3 by environments interaction	(p-1)
COMP 4 by environments interaction	(p-1)
COMP 5 by environments interaction	(p-1)
Experimental error within environments (pooled experimental error)	15p

action is important, but other intersite interactions are not. Remembering that replication of an environment is not possible, the nonimportant intersite interactions can serve as an estimate of experimental variability. The influence of the pooled experimental error will be lessened and the assumptions of significance testing will be more closely approximated (see Table 14.9 and discussion of Cochran and Cox, 1957).

Before proceeding, the nonorthogonality of the INLIT set of treatment comparisons also has to be handled in calculating the sums of squares of the preliminary combined data analysis. The pooled experimental error sum of squares is usually calculated by statistical computer programs as the difference between the total sum of squares and the model sum of squares. It is then important that the five treatment comparison sums of squares add to the treatment sum of squares with 5 degrees of freedom; and that the five treatment comparison intersite interaction sums of squares add to the treatments by environments sum of squares with 5(p-1) degrees of freedom. Computer programs handle this chore by calculating sequential sums of squares (Allen and Cady, 1982). Ordering of the treatment comparisons is important in calculating sequential sums of squares. As we know from Section IV, however, the interpretation of COMP 3 and COMP 4 after COMP 1 and COMP 2 is not the same N source comparison as when COMP 3 and COMP 4 are estimated alone or in a sequence before COMP 1 and COMP 2.

The primary reason for the preliminary combined data analysis is to evaluate the heterogeneity of the five treatment comparison intersite interactions. This evaluation forms the basis for determining the model and the composition of the error term in the environment means data analysis.

Environment Means Data Analysis

If no intersite interaction is deemed important during the preliminary combined data analysis, then the treatment comparisons, averaged over the environments, can be evaluated following the development presented in Section III. However, several choices of an error term for significance testing and confidence interval estimation are now available from the preliminary combined data analysis. Three alternatives are considered here:

(a) The pooled experimental error mean square.

(b) The intersite interaction mean square that corresponds to the treatment comparison.

(c) The average mean square of some or all intersite interactions.

Each alternative has limitations. The pooled experimental error is usually in violation of the homogeneity assumption. Both (b) and (c) can be estimating more than σ^2, the assumed plot-to-plot experimental variability. This inclusion of an additional component of intersite variability is perhaps more realistic and is consistent with previous statements that an environment should be considered the experimental unit for making inferences. No pooling of potentially heterogeneous quantities is required in (b), but the (p-1) degrees of freedom of (b) might not be sufficient. Alternative (c) is used in the INLIT multilocational final analysis.

In general, one or more of the intersite interactions will be judged important and an analysis in addition to the preliminary combined data analysis will be needed. The new analysis will be called the environment means data analysis and will utilize a data base of the six treatment means from each environment, i.e., each environment is one replication with six treatment means as observations.

Features of the environment means data analysis include treatment means for observations and a model formulated from the preliminary combined data analysis, but reduced in number of parameters. One or more of the intersite interactions will be utilized as a measure of experimental variability for significance testing and confidence interval estimation. The most promising candidates are the interactions between environments and treatment comparisons 1 and 3, as these two intersite interactions would not be expected to be large.

Interactions between environments and treatment comparisons 2, 4, and 5 are more likely, especially the COMP 2 comparison of inoculation vs. control. Some environments will give a response to *Rhizobium* whereas others will not. The magnitude of the COMP 5 intersite interaction will depend on the relative differences between farm and maximum fertility across the environments. An interaction between environments and comparison 4 is more difficult to understand.

The environment means data analysis should be sufficiently flexible to include additional variables to explain or at least account for important intersite interaction sums of squares. For example, if environments give differential inoculation responses, then one or more factors, not controlled at each environment but varying among the environments, need to be identified.

INLIT was designed to include measurements at an early stage of growth in addition to final crop yields. Nodule number and nodule fresh weight were among the early determinations. These data can be used to classify the previous inoculation status of each environment. Both variables are continuous variables, but due to field experimental problems, the nodule data have been used to form a dichotomous variable, either a recent history of the particular crop or no previous history. Then 1 degree of freedom of the COMP 2 intersite interaction degrees of freedom is associated with the interaction variable between COMP 2 and previous crop history.

The reduced model for the environment means data analysis will then include (p-1) variables for environments and variables for both treatment comparisons and intersite interactions that have been shown to be important in the preliminary combined data analysis. The (p-1) environmental variables are indicator variables (Allen and Cady, 1982), constructed so that the set of variables represents an orthogonal set. If one environmental variable has been formulated as a dichotomous variable to account for an intersite interaction, the remaining (p-2) environmental variables are constructed as indicator variables orthogonal to the dichotomous variable.

With more than one dichotomous variable or with one or more continuous environmental variables, calculation of a standard error for a treatment comparison is more complex due to lack of orthogonality. However, the general strategy of the preliminary combined data analysis for model formulation input to the environment means data analysis remains the same. From this final analysis, estimates and standard errors of the important treatment comparisons can be calculated.

The next section gives a case study for chickpea (*Cicer arietinum*) demonstrating an INLIT multilocational data analysis.

VI. *CICER ARIETINUM* (CHICKPEA) MULTILOCATIONAL CASE STUDY

Grain yields from 20 chickpea experiments have been received from cooperators. Six experiments were not used in the multi-

Table 7. Cooperators, locations, and sowing dates for chickpea experiments in multilocational analysis

Experiment	Cooperator	Location	Sowing date
AS480	B. R. Gupta	Kanpur (UP), India	Oct. 1980
AS487	Shankar Lal	Fyzabad (UP), India	Nov. 1980
AS496	A. S. Tewari	Gwalior (MP), India	Nov. 1980
AS498	N. K.Gupta	Jhansi (UP), India	Nov. 1980
AS507	H. N.Singh	Shahjahanpur (UP),India	Nov. 1980
AS510	Dy. Dir. Agric.	Bareilly (UP), India	Nov. 1980
AS51.	R. P. Pareek	Pantnagar (UP), India	Nov. 1980
AS572	H. G. Singh	Kanpur (UP), India	Nov. 1980
AS603	K. A. Malik	Faiselabad, Pakistan	Nov. 1980
AS755	A. K. Habibullah	Mymensingh, Bangladesh	Nov. 1981
AS762	O. P. S. Shilhodie	Shahjahanpur (UP),India	Oct. 1981
AS769	S. N. Gupta	Jhansi (UP), India	Oct. 1982
AS962	K. A. Malik	Faiselabad, Pakistan	Nov. 1982
AS968	Md. Shahjahan	Joydepur, Bangladesh	Nov. 1982

Table 8. Means, standard deviations, treatment comparison estimates, and standard errors for *Cicer arietinum* grain yields [a]

Environment	Mean	Std Dev[b]	Treatment Comparison[c] 1	2	3	4	5	Std Err[d] of COMP 1-4	Std Err of COMP 5
AS 480	2896	146	-17	1063	-351	668	24	73	103
AS 487	2269	247	-40	15	69	202	-25	124	175
AS 496	1130	113	-116	211	97	-73	-53	56	80
AS 498	1764	573	177	344	250	271	354	287	405
AS 507	1060	383	252	89	215	-118	483	192	271
AS 510	2493	145	-115	260	-10	198	-42	73	103
AS 513	2128	437	-302	-94	31	52	-271	219	309
AS 572	1869	180	-182	611	-248	284	-278	90	127
AS 603	2097	442	221	-87	300	183	608	221	313
AS 755	1904	264	27	360	2	202	25	132	186
AS 762	2536	238	-42	-187	10	17	42	119	167
AS 769	3576	336	-52	448	-135	-240	-208	168	238
AS 962	2291	247	-40	1210	-262	625	242	124	175
AS 968	1656	219	-10	198	104	146	208	110	155

[a] All values are in units of kg/ha.

[b] Standard deviation

[c] The five treatment comparisons are: (1) inoculation by fertility interaction, (2) *Rhizobium* vs. control, (3) N source by fertility interaction, (4) *Rhizobium* vs. bag N, and (5) farm vs. maximum fertility with *Rhizobium*.

d Standard error

locational analysis due to extremely low yields, high coefficients of variation, or other experimental field problems. Table 7 lists the cooperators, locations, and sowing dates for the remaining 14 experiments. Means, standard deviations, and treatment comparison estimates and standard errors are presented in Table 8. All calculations are done with the statistical computer package Statistical Analysis System (SAS, 1982).

From Table 8, the magnitudes of the intrasite interaction comparisons 1 and 3, specifically inoculation by fertility and N source by fertility, do not appear to be important relative to their standard errors. If so, interpretation can focus on the main effects, COMP 2 (inoculation vs. control) and COMP 4 (N source), and the effect of COMP 5 (farm vs. maximum fertility with inoculation). The yield responses for these comparisons are not as consistent over the environments, especially for the inoculation effect of COMP 2.

The combined analysis of variance table (Cochran and Cox, 1957) of Output 4 is a decomposition of the model sum of squares into four terms including experiments, blocks within experiments, treatments, and treatments by experiments interaction. The F statistic with 65 and 210 degrees of freedom evaluates the treatments by experiments interaction, i.e., the consistency of intrasite treatment comparison estimates across the environments (labeled as experiments in the output). The F value is the ratio of a mean square to the error mean square (pooled experimental error) while PR > F is the probability of an observed value exceeding the calculated value of F by chance alone. Associated with a small probability level, the F value of 2.55 is unlikely to occur by chance alone.

If the composite evaluation of the treatments by experiments interaction had not been important, one approach would have been to proceed directly to the five comparisons at the bottom of the output. These estimates are the intrasite estimates averaged over the 14 experiments. For example, the average difference between the *Rhizobium* and control treatments (COMP 2) is 317 kg/ha (with a standard error of 41.63). However, interpretation of the COMP 2 estimate is not clear without additional examination of the treatments by experiments interaction.

The preliminary combined data analysis, outlined in Section V, contains a breakdown of the treatments by experiments interaction sum of squares into five potentially heterogeneous intersite interaction terms. Each of the last five sums of squares in Outputs 4A and 4B is an interaction between experiments and one of the five major treatment comparisons. With the nonorthogonal nature of the INLIT set of treatment comparisons, the five intersite comparisons need to be calculated sequentially. The "TYPE I SS" are sequential sums of squares (Allen and Cady, 1982) with the five treatment comparisons appearing in a COMP 1, COMP 2, COMP 3, COMP 4, COMP 5 ordering in Output 4A, while Output 4B reflects a COMP 3, COMP 4, COMP 1, COMP 2, COMP 5 ordering. The sums of squares for each of five treatment comparisons and intersite interactions for both outputs will add to the sums of squares for treatments and treatments by experiments of the classic combined analysis of variance table of Output 4. Since COMP 2 and COMP 4 are the major questions for INLIT, only the two orderings of Outputs 4A and 4B need to be considered.

The values of the sequential sums of squares depend on the order of the input variables. While the COMP 2 intersite interaction sum of squares is large regard-

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: SEED_YLD

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE
MODEL	125	176489834.80790792	1411918.67846326	14.55
ERROR	210	20379847.99593814	97046.89521875	PR > F
CORRECTED TOTAL	335	196869682.80384606		0.0001

R-SQUARE	C.V.	ROOT MSE	SEED_YLD MEAN
0.896481	14.6985	311.52350669	2119.42228836

SOURCE	DF	TYPE I SS	F VALUE	PR > F
EXPTS	13	135443513.81492710	107.36	0.0001
BLOCKS(EXPTS)	42	18639256.68725741	4.57	0.0001
TRTS		6315656.85707148	13.02	0.0001
TRTS*EXPTS	6	16091407.44865193	2.55	0.0001

PARAMETER	ESTIMATE	T FOR H0: PARAMETER=0	PR > |T|	STD ERROR OF ESTIMATE
COMP1	-17.09476183	-0.41	0.6818	41.62907964
COMP2	317.15903135	7.62	0.0001	41.62907964
COMP3	5.16672785	0.12	0.9013	41.62907964
COMP4	172.69715803	4.15	0.0001	41.62907964
COMP5	79.15324651	1.34	0.1802	58.87240902

Output 4. Preliminary combined data analysis including the combined analysis of variance table and estimates of five treatment comparisons averaged over the experiments.

SOURCE	DF	TYPE I SS	F VALUE	PR > F
EXPTS	13	135443513.81492710	107.36	0.0001
BLOCKS(EXPTS)	42	18639256.68725741	4.57	0.0001
COMP1	1	16364.92939234	0.17	0.6818
COMP2	1	5633031.66550373	58.04	0.0001
COMP3	1	14043.06617504	0.14	0.7040
COMP4	1	14881.65097791	0.15	0.6958
COMP5	1	637335.54502247	6.57	0.0111
COMP1*EXPTS	13	1196494.25114468	0.95	0.5043
COMP2*EXPTS	13	8877649.73352130	7.04	0.0001
COMP3*EXPTS	13	1919228.70452781	1.52	0.1111
COMP4*EXPTS	13	2231196.73418649	1.77	0.0496
COMP5*EXPTS	13	1866838.02527165	1.48	0.1264

Output 4A. The treatment comparisons and intersite interaction sequential sums of squares for the ordering of the comparisons COMP 1, COMP 2, COMP 3, COMP 4, COMP 5.

SOURCE	DF	TYPE I SS	F VALUE	PR > F
EXPTS	13	135443513.81492710	107.36	0.0001
BLOCKS(EXPTS)	42	18639256.68725741	4.57	0.0001
COMP3	1	1494.92429215	0.02	0.9013
COMP4	1	1670161.26992941	17.21	0.0001
COMP1	1	28913.07127523	0.30	0.5858
COMP2	1	3977752.04655224	40.99	0.0001
COMP5	1	637335.54502247	6.57	0.0111
COMP3*EXPTS	13	1984986.66188654	1.57	0.0943
COMP4*EXPTS	13	3313431.39176653	2.63	0.0021
COMP1*EXPTS	13	1130736.29378595	0.90	0.5585
COMP2*EXPTS	13	7795415.07594126	6.18	0.0001
COMP5*EXPTS	13	1866838.02527165	1.48	0.1264

Output 4B. The treatment comparisons and intersite interaction sequential sums of squares for the ordering of the comparisons COMP 3, COMP 4, COMP 1, COMP 2, COMP 5.

less of ordering, the importance of the COMP 4 intersite interaction does depend on the ordering. Fitted first, COMP 4 is important (Output 4B), but fitted after COMP 2 (Output 4A), "COMP 4" is actually the comparison between bag N and the average of control and *Rhizobium* (see Section IV). Then, the "COMP 4" intersite interaction sum of squares is reduced. The COMP 4 intersite interaction is clearly not important. However, the intersite interaction with COMP 3 is associated with a probability level of approximately 0.1. Closer examination of experiments 480, 572, and 962 reveals a larger *Rhizobium* response for farm fertility than for maximum fertility, whereas the bag N response is either the same or larger for maximum fertility. Even though intersite interactions with both COMP 3 and COMP 5 are potentially interesting, they were not deemed sufficiently important, either by significance level or yield magnitude, to be retained in the model.

The next step was to determine a rationale for the intersite interactions with COMP 2 and COMP 4. One possibility investigated was whether or not the location had a recent history of chickpea cultivation as indicated by sufficient *Rhizobium* to produce adequate inoculation. This determination was made from previous crop information supplied by cooperators where available and otherwise indirectly from early harvest measurements. Control plot nodule numbers and nodule fresh weights equal to those on the *Rhizobium* plots were accepted as evidence of a recent prior chickpea crop. Lesser numbers or lower weights, in terms of statistical or biological significance, were accepted as indicating no recent history of chickpea cultivation. Note that these determinations were made a priori to final grain yield measurement and data analysis.

When the environments were so separated, seven experiments were classified as having a recent history of chickpea cultivation and seven as having no such history (the equal numbers were fortuitous). Then, one of the 13 environment degrees of freedom was utilized as a dichotomous environment variable with +1 assigned to those environments without a recent history of chickpea cultivation (480, 496, 510, 572, 755, 762, and 962) and −1 assigned for such a recent history of inoculation. The remaining 12 environment indicator variables were arbitrarily constructed to be orthogonal with the dichotomous environment variable. Consequently, the sequential (Type I) sums of squares are equal to the partial (Type II) sums of squares for the 13 environment variables (E1 to E13) in Outputs 5A and 5B, and the sum of the 13 environment sums of squares will add to the experiments (EXPTS) sum of squares with 13 degrees of freedom in Outputs 4, 4A, and 4B.

Two reduced models are considered in the environment means data analysis. Model 5A includes variables for environments, COMP 2, COMP 2 by environments, COMP 4, COMP 4 by environments, and COMP 5. Model 5B deletes the variables associated with COMP 4. The parameter estimates, standard errors, and other information are shown in Outputs 5A (Model 5A) and 5B (Model 5B). These outputs are from the SAS regression program (PROC REG).

We wish Model 5A parameter estimates could be interpreted directly; that is, the inoculation effect without a recent history of chickpea cultivation is the sum of the coefficients (308 + 181 = 489) for COMP 2 and the interaction between COMP 2 and the dichotomous environment variable, E1. With environments that have had a recent history of chickpea cultivation, the

DEP VARIABLE: YMEAN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	42	38093925	906998	27.169	0.0001
ERROR	41	1368719	33383.402		
C TOTAL	83	39462645			

ROOT MSE	182.711	R-SQUARE	0.9653
DEP MEAN	2119.422	ADJ R-SQ	0.9298
C.V.	8.620805		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|	TYPE I SS	TYPE II SS
INTERCEP	1	2119.422	19.935432	106.314	0.0001	377323870	377323870
E1	1	40.586310	19.935432	2.036	0.0483	138369	138369
E2	1	122.714	11.509727	10.662	0.0001	3794802	3794802
E3	1	-181.526	13.618493	-13.329	0.0001	5931276	5931276
E4	1	68.558858	16.679179	4.110	0.0002	564038	564038
E5	1	-93.643053	21.532728	-4.349	0.0001	631370	631370
E6	1	-169.869	30.451875	-5.578	0.0001	1038795	1038795
E7	1	122.106	52.744195	2.315	0.0257	178920	178920
E8	1	31.768078	11.509727	2.760	0.0086	254321	254321
E9	1	-56.635802	13.618493	-4.159	0.0002	577371	577371
E10	1	-260.880	16.679179	-15.641	0.0001	8166982	8166982
E11	1	-78.703704	21.532728	-3.655	0.0007	445988	445988
E12	1	-173.032	30.451875	-5.682	0.0001	1077848	1077848
E13	1	960.069	52.744195	18.202	0.0001	11060800	11060800
COMP2	1	307.747	56.385917	5.458	0.0001	1408258	994438
COMP2E1	1	181.425	56.385917	3.218	0.0025	489308	345606
COMP2E2	1	80.593509	32.554424	2.476	0.0175	364965	204602
COMP2E3	1	-15.698989	38.518914	-0.408	0.6857	47887.984	5545.311
COMP2E4	1	-52.249873	47.175843	-1.108	0.2745	45328.591	40950.738
COMP2E5	1	49.695201	60.903751	0.816	0.4192	16953.513	22226.517
COMP2E6	1	-40.499303	86.130910	-0.470	0.6407	15134.710	7380.871
COMP2E7	1	-728.905	149.183	-4.886	0.0001	976052	796954
COMP2E8	1	-40.266755	32.554424	-1.237	0.2232	15597.331	51074.464
COMP2E9	1	22.237654	38.518914	0.577	0.5669	45289.388	11126.549
COMP2E10	1	13.101852	47.175843	0.278	0.7826	609.592	2574.878
COMP2E11	1	-96.990741	60.903751	-1.593	0.1189	58741.681	84664.834
COMP2E12	1	-233.565	86.130910	-2.712	0.0097	112295	245486
COMP2E13	1	295.139	149.183	1.978	0.0546	31250.000	130660
COMP4	1	18.823523	56.385917	0.334	0.7402	3720.413	3720.413
COMP4E1	1	11.052359	56.385917	0.196	0.8456	1282.624	1282.624
COMP4E2	1	25.249444	32.554424	0.776	0.4424	20082.335	20082.335
COMP4E3	1	-48.508621	38.518914	-1.259	0.2150	52944.442	52944.442
COMP4E4	1	9.285680	47.175843	0.197	0.8449	1293.358	1293.358
COMP4E5	1	-24.216133	60.903751	-0.398	0.6930	5277.790	5277.790
COMP4E6	1	-19.449423	86.130910	-0.226	0.8225	1702.260	1702.260
COMP4E7	1	60.653214	149.183	0.406	0.6865	5514.580	5514.580
COMP4E8	1	41.991843	32.554424	1.290	0.2043	55544.419	55544.419
COMP4E9	1	33.233025	38.518914	0.863	0.3933	24849.763	24849.763
COMP4E10	1	-37.245370	47.175843	-0.790	0.4344	20808.264	20808.264
COMP4E11	1	54.050926	60.903751	0.887	0.3800	26293.523	26293.523
COMP4E12	1	193.519	86.130910	2.247	0.0301	168522	168522
COMP4E13	1	-340.278	149.183	-2.281	0.0278	173683	173683
COMP5	1	39.576623	34.529181	1.146	0.2584	43856.655	43856.655

Output 5A. Environment means data analysis. The model includes variables for environments, COMP 2, COMP 4, COMP 5, and intersite interactions with COMP 2 and COMP 4.

DEP VARIABLE: YMEAN

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	28	37532405	1340443	38.194	0.0001
ERROR	55	1930239	35095.256		
C TOTAL	83	39462645			

ROOT MSE	187.337	R-SQUARE	0.9511
DEP MEAN	2119.422	ADJ R-SQ	0.9262
C.V.	8.839073		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|	TYPE I SS	TYPE II SS
INTERCEP	1	2119.422	20.440173	103.689	0.0001	377323870	377323870
E1	1	40.586310	20.440173	1.986	0.0521	138369	138369
E2	1	122.714	11.801139	10.398	0.0001	3794802	3794802
E3	1	-181.526	13.963296	-13.000	0.0001	5931276	5931276
E4	1	68.558858	17.101476	4.009	0.0002	564038	564038
E5	1	-93.643053	22.077910	-4.241	0.0001	631370	631370
E6	1	-169.869	31.222880	-5.441	0.0001	1038795	1038795
E7	1	122.106	54.079614	2.258	0.0279	178920	178920
E8	1	31.768078	11.801139	2.692	0.0094	254321	254321
E9	1	-56.635802	13.963296	-4.056	0.0002	577371	577371
E10	1	-260.880	17.101476	-15.255	0.0001	8166982	8166982
E11	1	-78.703704	22.077910	-3.565	0.0008	445988	445988
E12	1	-173.032	31.222880	-5.542	0.0001	1077848	1077848
E13	1	960.069	54.079614	17.753	0.0001	11060800	11060800
COMP2	1	317.159	50.067994	6.335	0.0001	1408258	1408258
COMP2E1	1	186.951	50.067994	3.734	0.0004	489308	489308
COMP2E2	1	93.218231	28.906770	3.225	0.0021	364965	364965
COMP2E3	1	-39.953300	34.202951	-1.168	0.2478	47887.984	47887.984
COMP2E4	1	-47.607033	41.889889	-1.136	0.2607	45328.591	45328.591
COMP2E5	1	37.587135	54.079614	0.695	0.4900	16953.513	16953.513
COMP2E6	1	-50.224014	76.480124	-0.657	0.5141	15134.710	15134.710
COMP2E7	1	-698.589	132.467	-5.274	0.0001	976052	976052
COMP2E8	1	-19.270833	28.906770	-0.667	0.5078	15597.331	15597.331
COMP2E9	1	38.854167	34.202951	1.136	0.2609	45289.388	45289.388
COMP2E10	1	-5.520833	41.889889	-0.132	0.8956	609.592	609.592
COMP2E11	1	-69.965278	54.079614	-1.294	0.2012	58741.681	58741.681
COMP2E12	1	-136.806	76.480124	-1.789	0.0792	112295	112295
COMP2E13	1	125.000	132.467	0.944	0.3495	31250.000	31250.000
COMP5	1	39.576623	35.403418	1.118	0.2685	43856.655	43856.655

Output 5B. Environment means data analysis. The model includes variables for environments, COMP 2 and COMP 5, and intersite interactions with COMP 2.

inoculation effect is the difference between the two coefficients (308 – 181 = 127). (The unweighted sum and difference of the coefficients result from the +1 and –1 assigned values of E1.) However, the parameter estimates of Output 5A are partial regression coefficients (Allen and Cady, 1982). Both the partial coefficients and the partial sums of squares (Type II SS) reflect the COMP 2 comparison after fitting COMP 4. Consequently, the two coefficients do not reflect the desired COMP 2 comparison, but a comparison between the control treatment and the average of the two N sources, bag N and *Rhizobium* N (see Section IV). The partial COMP 4 coefficients of Output 5A are more dramatic. The values 18.8 ± 11.1 are the sum and difference of the two partial coefficients for the partial COMP 4, which is the comparison between bag N and the average of the control and *Rhizobium* treatments rather than the desired comparison between N sources.

To cope with the interpretational problem of the partial regression coefficients, two alternative approaches are available, namely, calculate sequential regression coefficients (Allen and Cady, 1982) directly or fit reduced versions of Model 4A by deleting either the COMP 2 or COMP 4 variables. The SAS regression procedure has an option that calculates sequential coefficients. The estimated sequentials are not shown in Output 5A but, for example, if the COMP 4 and COMP 4 by E1 interaction variables are ordered before the COMP 2 and COMP 2 by experiments interaction variables, then the sequentials will give the desired COMP 4 comparisons, namely, a 274 (172 + 102) kg/ha *Rhizobium* N response relative to bag N for environments without a prior history of chickpea and a 71 kg/ha *Rhizobium* response for environments with such a prior history. The variance of each sequential coefficient is:

$$(33383.4)(1/14)[0^2 + (-0.5)^2 + 0.5^2 + 0^2 + (-0.5)^2 + 0.5^2]$$

where 33383.4 is the error mean square, 14 is the number of experiments, and the remaining values are the comparison coefficients. The standard error of either the sum or the difference of the two sequential coefficients is the square root of the sum of the two variances, or 69.06.

Model 5B includes the same parameters as Model 5A except that the COMP 4 and COMP 4 intersite interactions variables have been deleted. The partial coefficients given in Output 5B in the parameter estimate column are also the sequential coefficients since each variable in Model 5B is orthogonal to each other variable. The *Rhizobium* response relative to the control is 504 (317 + 187) kg/ha for environments without a prior history of a chickpea crop and 130 kg/ha for environments with such a prior history. The standard error of either coefficient is 50.068 from Output 5B. Then the standard error of either the sum or the difference of the two coefficients is the square root of the sum of the two variances, or 70.81. Using the error term from Output 5A, the standard error of either the sum or the difference is 69.06.

Is there a difference between environments with or without a prior history of chickpea cultivation? The estimate of this difference is 374 (504 – 103) kg/ha with a standard error of 100.1 kg/ha.

As discussed in Section V, the error term of the environment means data analysis is expected to be larger than an error utilizing solely the pooled intrasite experimental errors. The standard deviation is 183 from Output 5A and 187 from Output 5B instead of the 311 from Output 4. If no sources of variability were part of the error sum of squares other than the pooled intrasite

experimental errors, then the standard deviation in Outputs 5A and 5B would have been 311/2, the standard deviation divided by the square root of the number of replications at each environment. (Each observation in the data analyses of Outputs 5A and 5B is a treatment mean.) Finally, note that COMP 5 is not important relative to its associated standard error in the environment means data analysis.

An anomaly with COMP 4 needs explanation. The effect due to *Rhizobium* vs. bag N is 274 kg/ha without prior crop history vs. 71 kg/ha with such a history. The most logical explanation for bag N being less effective than *Rhizobium* is the tendency for excessive leaching in tropical soils. There is, however, no apparent reason why there should be a differential effect with the efficiency of bag N being less with environments without a prior history of chickpea cultivation. It may turn out not to be as important with additional environments in the analysis. However, it is still an interesting phenomenon. If additional data continue to support the differential effect, a logical explanation will have to be sought.

For this chickpea data set, the sequential coefficients and their standard errors could have been hand calculated from Table 8 and the output from the preliminary combined data analysis. One of the following conditions would make hand calculations prohibitive: the data set for each environment is not complete (missing plots), partial coefficients are desired, or the environment variables cannot be constructed as an orthogonal set. In general, SAS or another statistical computer package that will give sequential coefficients, in addition to estimates and standard errors of linear combination of partial coefficients, would need to be used.

The case study of this section and the soybean example of Section III can serve as guidelines to intrasite and multilocational data analyses for other inoculation trials. Appropriately modified to incorporate specific questions of interest, the data analysis methodology of Sections III and V can be applied to other experimental networks where hypotheses of interest can be formulated as comparisons among treatment means. The goal of any analysis is to elicit output that directly addresses the formulated conjectures. Analysis is simplified with orthogonality among the treatment comparisons. Although desirable, a more important criterion in data analysis is whether the comparisons are meaningful and were preplanned. The methodology developed here for intrasite and multilocational data analyses is sufficiently general to handle both orthogonal and nonorthogonal comparisons among treatment means.

VII. SUMMARY AND CONCLUSIONS

This bulletin has described the International Network of Legume Inoculation Trials (INLIT), a system of worldwide trials for testing response to legume inoculation. Considered were the history of the trials, the need for the trials, the formulation of the actual components of the trials, and the general objectives.

The primary objective of INLIT is to determine whether under definitive conditions a yield increase can be obtained through legume inoculation using the present state of the art. The basic INLIT experiment, the "A" Experiment, is described in detail. This includes the formulation of the research design, the coordination and servicing of the trials, and the analysis of data on a single and multiple location basis. The primary objective is being achieved. Preliminary results indicate that a response often can be obtained from inoculation with proper *Rhizobium*.

This even has been noted with some traditional crops where the presence in the soil of *Rhizobium* of the correct cross-inoculation group would be expected.

The secondary objectives of identifying all research workers in the tropics interested in legume inoculation, convincing them of the advantages of conducting uniform field trials, and improving the quality of research work and expertise of research workers by providing support services, have all been achieved.

A comprehensive recapitulation of the results of the trials has not been given, although preliminary data have been used to illustrate and validate the methodology used. Data from the trials are still being received at NifTAL. Data from individual experiments are analyzed as soon as they are received, and a report sent to the cooperator. Results of multilocational analysis will be reported as data are accumulated. In the meantime INLIT progress, as well as other cooperator-specific information, is communicated periodically to all cooperators.

The trials have brought all legume inoculation workers throughout the tropics into one network, with a uniform research design used by all those conducting field research in legume inoculation.

The acceptance of the INLIT concept by many legume inoculation workers has shown that a need was perceived for such coordinated trials. Feedback from cooperators has indicated that the INLIT successfully met the demand, both with the concept of a valid, uniform research design, and by providing sufficient support services for the trials.

VIII. REFERENCES CITED

ALLEN, D. M., and F. B. CADY. 1982. Analyzing experimental data by regression. Lifetime Learning Publications, Belmont, California.

CHEW, V. 1980. Testing differences among means: correct interpretation and some alternatives. HortScience 15:467-470.

COCHRAN, W. G., and G. M. COX. 1957. Experimental designs. John Wiley and Sons, Inc., New York.

HARRIS, S. W. 1979. Planning an International Network of Legume Inoculation Trials. Proceedings NifTAL Workshop. Jan. 15-19, 1979. NifTAL Project and U.S. Agency for International Development.

SAS INSTITUTE, INC. 1982. SAS user's guide: Statistics, 1982 edition. SAS Institute, Inc., Cary, North Carolina.