Expl Agric. (2000), volume 36, pp. 151–163
Printed in Great Britain
Copyright © 2000 Cambridge University Press

DESIGN PERSPECTIVES IN LONG-TERM EXPERIMENTATION

By JANET RILEY†

IACR-Rothamsted, Harpenden, AL5 2JQ, UK

(*Accepted 5 January 1999*)

SUMMARY

The aims and design features of historic long-term experiments are discussed and modern perceptions are presented of long-term studies to address sustainability and research impact in farming communities. A research strategy is proposed combining on-station and on-farm research whose results influence dynamically the directions of the component studies. Guidelines are given for choosing well-controlled designs, appropriate data and sampling procedures and for maintaining quality control in data collection and indicator formation over anticipated lengthy periods of time.

INTRODUCTION

During the last decade increasing interest in global change has been demonstrated. Changes to the natural environment stem from climate change resulting from natural disasters, influences of pollution and agrochemical usage upon the production of gases. Biological change to renewable natural resources, such as land, forestry and wetlands, is influenced by land management and marine productivity. Population changes result from the effects of urbanization, the impact of fertility controls and fluctuations in health caused by population pressures and the declining availability of natural resources. Of immediate importance is the conservation of biodiversity, assessment of long-term sustainability and the management of sustainable development.

Long-term experiments have been run in many parts of the world since the latter part of the last century and early this century (Steiner, 1995). Their contribution to the debate on global change is invaluable as they can provide detailed information on changing conditions over substantial periods of time, as well as providing large databases to support precise studies of sustainability and impact assessment of research.

However, new demands are now changing the priorities of political and developmental institutions. Political change has resulted in the movement of aid funds to new areas of the world and inadequate availability of funds for the continuation of existing projects. Yet developmental change has resulted in greater emphasis being placed on participatory studies to assess the uptake of

† Email: Janet.Riley@bbsrc.ac.uk

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

interventions by farmers and the long-term social and economic impact of research upon communities. The need to maintain the resource base appears to oppose productivity improvement and the assessment of multidisciplinary integrated systems over lengthy periods of time is crucial to the resolution of this conflict.

The design of long-term experiments to incorporate multidisciplinary systems and research impact upon communities is necessarily complex. System components may include soils, atmospheric gases, crops, pests, disease, trees, animals, fish, people and organizations far beyond the system, such as schools, markets and governments, both local and national, and upon whom changes to the system might impinge. The estimation of component behaviour in any system is necessarily complex because of the many sources of variation and their interactions. Unlike well-controlled long-term trials on research stations, such studies of farmer-managed systems are at greater risk of generating data which cannot provide precise and consistent estimates of change.

The conflict between short-term research perspectives and long-term sustainability assessment raises numerous questions. How can long-term sustainability be estimated and impact assessments be made if long-term projects cannot be funded? How can precise data be guaranteed from short-term on-farm studies when variability is large and its sources are at times unknown? Can historical data be used to predict future sustainability patterns? Can short-term data be used in place of longer series? Yet, the placing of experiments on farmers' fields provides an opportunity to verify research station results and to examine dynamically a broader range of agronomic components whilst determining changes to productivity and the resource base at the same time.

This paper proposes a research strategy for long-term dynamic multidisciplinary studies and discusses some relevant aspects of design for assessment of sustainability and impact.

RESEARCH STRATEGY

Existing long-term trials typically involve lengthy monitoring of formal plot experiments on research stations. Those in the UK, USA, Philippines and India reviewed by Barnett *et al.* (1995b) began in the years between 1843 and 1972. Many such trials in other parts of the world have existed for up to 30 years. A feature of many of these experiments is that they were not planned to study sustainability, nor are their original objectives still relevant. All of the trials relate to improvement of productivity. Many are supported by weather data for which their relationships with production variables have been estimated, sometimes incorporating economic summaries relevant to the time period.

Emphasis by donors is now given to the examination of interventions upon all components of farming systems, their productivity and sustainability and the ultimate social and economic impact of change upon communities. Research upon whole-farm systems and their long-term sustainability under population pressure

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

and climate change is essential. Such multidisciplinary studies appear complex and do not appear to lend themselves to the use of standard, documented designs. A broader range of designs and design strategies may be needed together with powerful analysis techniques provided by statistical software. Before this is done however, a research strategy is needed to guarantee clear objectives and relevant data.

The determination of farmer priorities, and of constraints to development, is the essential initial step. These can be determined through participatory rural appraisal methodology in farmers' fields, by talking to farmer groups and by obtaining donor suggestions (Pretty *et al*., 1995). Detailed observational surveys or trials can then follow with different degrees of input from the farmer. Awareness of changing farmer priorities should be maintained through discussion and interaction. Specific approaches to design relating to the objectives must be chosen (Riley and Alexander, 1997). These can be categorized broadly as follows:

- Detailed studies on one or a few isolated system components with controlled modifications in time. These may be planned by the researcher and done on farmers' fields either by the researcher or by the farmer.
- Whole-system studies which require monitoring of all system components and their inter-relationships in time. These may involve the monitoring of the existing system managed by the farmer or the monitoring of changes introduced to the system by the researcher.
- Detailed, controlled on-station studies planned by researchers with inputs to the design and interpretation by farmers to reflect their changing constraints and perceptions.

Existing long-term studies may be adapted to accommodate these or may be reduced in size to provide a number of observation plots for detailed modelling of sustainability patterns.

Identification of changing priorities at any stage will influence the objectives of both on-farm and on-station studies and information gained from each of these types of study should influence the direction of the other type. Results obtained from specific biological processes studied in some detail on-station can support observations made on farm. The important factors are that dynamic changes can be incorporated into any of the studies at any appropriate point in time and that changing effects must be estimated in an unbiased way. To ensure this, design and data collection strategies must be chosen carefully.

FEATURES OF EXPERIMENTAL DESIGN

Design features vary according to the aims of an experiment and the environment in which it is done. Typically researcher-designed, researcher-managed on-farm experiments will have design structures not dissimilar to those done on research stations. Variation caused by existing environmental features and the influence of farm management patterns will cause the major differences in design and analysis

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

requirements. Studies done by farmers and studies of whole systems have designs which differ most markedly from those of on-station studies. The structure of a plot experiment on fertilizers replicated across farms in one region will differ from that of a socio-economic survey of farmers' current management practices in a number of regions. However, certain important design requirements are common to all studies. These relate to:

- the specification of clear objectives
- the identification of sources of variability and its control through design
- the choice of sampling strategy robust to influences in time
- the maintenance of good trial management to ensure that objectives are adhered to and that mishaps and undesirable changes to treatments do not occur
- the choice of indicators and the scale and frequency of measurement for estimation of sustainability and impact upon the environment
- the establishment of appropriate storage structures to handle large data sets

For each of these design requirements some major features are highlighted below with emphasis upon long-term experiments.

The specification of clear objectives

All experiments, however simple, are based upon one or more objectives which can be quantified or qualified through the experimental procedure. Initial decisions are necessary to ensure the usefulness of the experiment in time and thus its longevity. When experiments are complex, multidisciplinary and lengthy in time, their objectives need to be specified very clearly so that they may be achieved consistently and with the desired precision. Simplicity of design is preferred as the design will then be more robust to changes in time. Thus few treatments and non-complex objectives will have a greater chance of long-term maintenance than a complex factorial which may not be robust to the influence of future scientists or farmers or environmental changes. The desired environment needs to be very clearly identified as results may not be directly applied to another environment. Thus the objective to determine for three years the difference in yields for organic and inorganic fertilizers on moderately wealthy farms in the Kakamega area of Kenya is not the same as that to determine for three years the acceptability of inorganic fertilizer by all farmers – wealthy, moderately wealthy or poor – in the whole of western Kenya.

For some long-term experiments objectives may change in time with changing farmer priorities or development constraints. In multidisciplinary studies, some components of the system may require changes; others may not. Controlled modifications which incorporate such changes without loss of the basic design structure include :

- minimum changes to treatments or husbandry to ensure longevity
- incorporation of innovations in practice whilst maintaining original objectives

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

- changes of objectives to address current questions
- changes to the experiment to achieve more short-term information, sacrificing longevity
- basic maintenance of treatments with later relevant modifications made in response to identified needs (Leigh *et al.*, 1994)

These range from the simple to the drastic, but with care and provided they are appropriate they can be employed to maintain the relevance of long-term studies.

Designs to control variability

Historic long-term trials typically have involved lengthy monitoring of formal plot experiments on research stations where fertilizers are applied to particular species or where different crop rotations are compared. Designs for these trials vary. Simple but robust randomized block design structures are commonly used and many of the early experiments have unreplicated sets of treatments (Steiner, 1995). Other experiments are more complex in structure. The long-term rotational and manurial trial at Serere Research Station in Uganda involved cotton (*Gossypium* spp.), finger millet (*Eleusine corocana*) and groundnuts (*Arachis hypogaea*) in five cycles of five year rotations with three different resting periods, five types of resting cover and farmyard manure at three levels (McWalter and Wimble, 1976).

These experiments have provided a wealth of information about efficiency of design and which design can form the basis of on-station or researcher-managed on-farm studies.

A common feature of historic long-term trials is their non-anticipation of the possibility of changing patterns of environmental variation or of future damage to experimental plots. Design of any long-term trial should aim to maintain control of both spatial and temporal variability and to minimize the amount of labour required in a study which potentially generates vast quantities of data. Even relatively short-term experiments such as forestry trials need to be designed to ensure adequate estimation of treatment effects (Williams and Matheson, 1994) if environmental influences change in time. Control of variability can be achieved by a large degree of stratification, which if not ultimately required does not impair the design or bias the treatment estimates. Fig. 1 shows a row-column design from Williams and Matheson (1994) with four contiguous replicates of 20 treatments in four rows of five columns. The design is said to be 'latinized' as no treatment appears more than once in the long, sixteen-plot columns. If a whole column is damaged, its effect can be removed in the analysis; if it is not, this extra stratification does not impair the analysis. Optimal or near-optimal cyclic designs which allow this can now be designed simply using the computer software CycDesigN.

Designs for multidisciplinary whole-system studies need to be simple, to remain robust in time. Studies like this are not yet widely documented and the influence of different environmental factors must also be determined. Their stratification and

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

	Column					
Row	1	2	3	4	5	
1	11	8	19	18	12	
2	13	5	20	2	10	Replicate 1
3	1	7	16	14	6	
4	17	3	9	4	15	
1	15	4	3	11	20	
2	5	10	17	9	18	Replicate 2
3	14	16	12	8	13	
4	19	1	2	6	7	
1	18	9	10	1	19	
2	3	2	14	20	8	Replicate 3
3	16	15	7	5	17	
4	6	13	11	12	4	
1	7	17	8	10	2	
2	9	14	1	3	11	Replicate 4
3	12	20	15	19	16	
4	4	6	18	13	5	

Fig. 1. A row-column design for 20 treatments; rows within replicates are incomplete blocks; columns within replicates are latinized within long columns.

continuous documentation is crucial. Shepherd and Roger (1991) examined the importance of different sources of variability in on-farm agroforestry trials and showed that when genetic, environmental and management factors were considered separately, the patterns in variability caused by the trees, plots, experimental blocks, farms and regional locations differed relative to each other. Changes in time need to be determined. If wealth groups are shown to be the most important source of initial variation in a group of farmers it is essential to know whether this feature remains the most important as time passes, or whether some other stronger environmental factor emerges.

Many long-term trials have been done without true treatment replication. For long-term on-farm trials, enough replication is needed within the levels of the identified major sources of variability so that the study is robust to farmer drop-

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

out or failure to monitor the system at any time. Thus if wealth category is identified as the major source of variability, the choice of ten rich farmers, eleven of medium wealth and two poor farmers is unlikely to be representative of a typical community and will not be robust if one poor farmer drops out of the study. As the replication required is usually large, limited resources often dictate that only few treatments or interventions will be manageable.

Sampling strategies

To ensure unbiased data, a sampling strategy needs to be determined both in space and in time. For plot experiments documented sampling procedures may be used (Dyke, 1988) but consistency of the sampling strategy is essential. The historic long-term experiments had original plots of various sizes (Table 1) but over the years they were split to accommodate new treatments. The resultant plot sizes for some of the trials are very small with a consequent loss of precision of estimation of treatment effects (Steiner, 1995).

Researcher-managed on-farm plot trials are likely to be subject to less control of variation. Using 100 on-farm fertilizer trials on vegetables in Jamaica, Fielding and Riley (1997) showed that much larger plot sizes would have enabled more precise yield measurements to be obtained in the first year of the study. Table 2 shows for five of the crops, the relative plot sizes suggested by the study in relation to those which generated the data and those plot sizes used on-station.

Fielding and Riley (1997) also examined the problems of multiplying up yields from small harvest areas on farms. Yield estimates for two varieties of broccoli (*Brassica oleracea* var. *italica*) in St Kitts using small plots were compared with yields from neighbouring commercial plots. The percentage errors in Table 3

Table 1. Design features of historic long-term experiments.

Location	Crops	Design	Original plot size
Auburn, USA, 1896	cotton–legumes	variable replication	41.5 × 6.5 m^2
IRRI, Philippines, 1963	rice	split plots + nitrogen	various
Oregon, USA, 1931	wheat	two replicates of nine organic treatments	40.2 × 11.6 m^2
Sanborn Fields, USA, 1888	maize–wheat–clover rotations	unreplicated + fertilizers	30.5 × 9.5 m^2
India, 1970	maize–wheat–rice–legume rotations	RCB†, four replicate blocks + factorial fertilizers	160 to 300 m^2
Park Grass, UK, 1856	uncultivated grass	20 plots, split plot + factorial fertilizers	67 to 202 m^2
Woburn, UK, 1877–1926	wheat + barley	PK + forms of fertilizers + FYM	1000 m^2
Broadbalk, UK, 1843	wheat	factorial fertilizers single replicate	1214 m^2

† Randomized complete block design.

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

Table 2. Relative sizes of on-station and on-farm plots to achieve similar precision in Jamaican vegetable trials.

Crop	Average area used in the study (m^2)	Suggested harvest areas for on-farm trials (m^2)	Average areas used for on-station trials (m^2) in Jamaica
Cabbage	37.1	75	16.7
Groundnut	67.2	300	14.9
Sweet pepper	39.7	100	50.2
Tomato	48.4	150	5.7
Yam	31.8	100	15 plants

Table 3. Broccoli (*Brassica oleracea* var. *italica*) yields in kg ha^{-1} extrapolated from small farm and large commercial plots.

Variety	kg ha^{-1} extrapolated from small area	kg ha^{-1} extrapolated from large area	% error in small area estimates
A	6320	3480	+82
B	6530	4800	+36

show that use of yields estimated from small areas can give quite misleading results. An understanding of the likely changes in plot variability in time would help to determine appropriate initial plot sizes.

For studies such as surveys of farmer practice or multidisciplinary trials involving whole systems, plots are unlikely to be relevant. Data are observational and samples should be taken which are representative and random. However, in many situations the choice of farms may be limited, farmers may drop out of the study in time and data may be biased in different ways at different time points. Barnett and Riley (1995) indicated that size-biased sampling is a common phenomenon in environmental studies. Here, the distribution of data is typically biased upwards if larger units, such as prominent plants, more enthusiastic farmers or larger sheep in a flock, are chosen.

Another problem with sampling is that it can take much time and involve expensive analytical procedures. Two methods of sampling which save time and expense are composite sampling and ranked set sampling. Composite sampling involves the combination of several sample units and their analysis as a group in order to give an estimate of the mean. Patil *et al.* (1992) give a summary of environmental sampling methods including interesting examples of the spread of viruses by mosquitoes and the sampling of milk supplies in different locations.

The second method, ranked set sampling, aims to sample over as wide a space as possible thus removing randomness. Repeated sets of data are chosen and certain ranked values selected from each set. The use of these ranked sample values has been shown to provide more efficient estimates of means than equivalent sample sizes from simple random samples. Examples of its statistical and financial efficiencies are presented by Patil *et al.* (1992).

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

Good trial management

In discussing good trial management of the Rothamsted long-term classical experiments Leigh *et al.* (1994) note that both 'care and commitment' are essential to ensure that the trials are perceived to have integrity. Care in design, sampling, management and data recording is needed together with commitment to ensuring good results, proper interpretation and publication. For trials with plots, accurate location of plot boundaries (Dyke, 1988) and repeated use of sampling schemes in subsequent years (such as return to the same farmers) will ensure comparability of data over time. Maintenance of observations such as changes in local communities will supplement core data such as yields or farmer opinions. The maintenance of harvesting and weighing machines and their calibration if methods are changed is essential to provide comparable data for estimates of yield change or sustainability of a management practice.

Recording of data is a practice which, if faulty, can be very misleading. A study by the author involved a group of thirteen scientists measuring the height of a 53.5-cm potted oilpalm in a nursery. Fig. 2 shows the achieved measurements. Six scientists underestimated the height by more than two centimetres; two underestimated by up to six centimetres. Training in measurement (and in scoring) is essential to ensure that both accuracy and consistency over time is achieved. Fielding (1992) gives an interesting example of differences in ability to score damage in leaves and how this was improved with training.

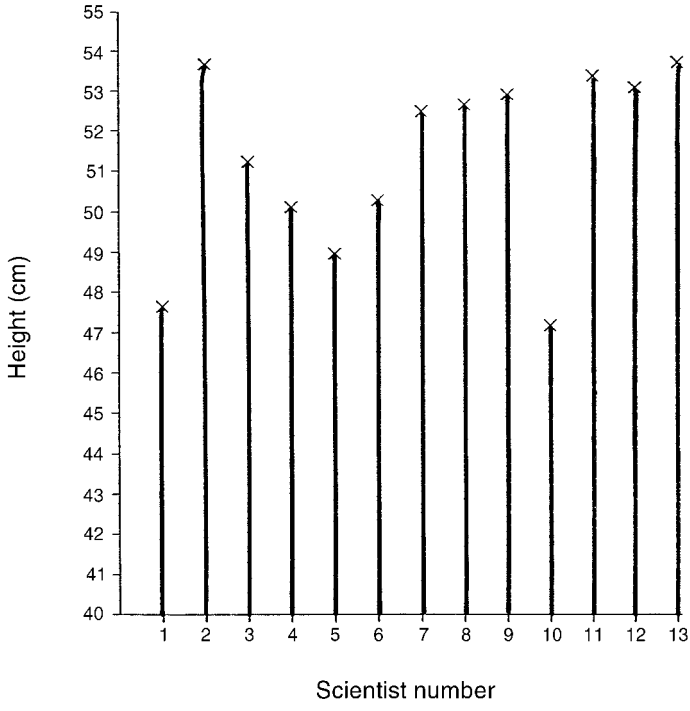


Fig. 2. Height in cm of a nursery oilpalm measured by thirteen scientists.

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

The choice of data and sampling procedure must be the same at each measurement date. Thus the same units – branches, farm areas, quadrats – should be monitored each time, otherwise inconsistency of material will prevent reliable estimates of sustainability from being made. If the same sample points cannot be used in subsequent years, this should be noted so that any possible adjustments can be made.

The choice of indicators

If productivity, sustainability, impact or development achievement of a project are to be identified, some observation or measure must be examined. Such indicators may be quantitative, such as the number of farmers using a technology throughout the time of the programme or the productivity of edible beans at project termination. They may also be qualitative such as the popularity of introduced cropping systems. Economic and political indicators abound in all areas of assessment. Social development indicators include measures of greater confidence, reduced conflicts in communities and reduced out-migration. Quality-of-life indicators may include improved quality of food, consumer food preferences, technology acceptance, or increased labour availability. Indicator type clearly depends as much on project objectives as on project type but also must be meaningful to the farmer; it is often valuable to ask the farmers to suggest indicators of value to themselves.

Indicators of environmental sustainability involve changes in waste levels, industrial accidents, water quality, gas emissions, reduced soil erosion, changes in resource stocks, ecosystem fluctuations and climate control. Underlying these is the need to define clearly the meaning of sustainability and the anticipated scope of system impact. Then the choice and use of appropriate methods must be made to estimate sustainability or impact upon the environment including local communities. Each of these demands clear definition of indicators for continuous assessment (Riley and Alexander, 1998). Kmietowicz (1995) shows that for many developing countries, agricultural production indices may be unreliable as it is impossible to distinguish year-to-year change from the average variability in the index.

System components generate different data which are sometimes weighted and combined to form ratios, indices or indicators. Such complex combinations are difficult to interpret and often have no biological meaning. If their statistical distribution is unclear, it can be difficult to determine an appropriate analysis. Cox *et al*. (1992) recommend using simple weighting schemes within components and disciplines followed by robust analysis techniques and sensitivity analysis. Barnett *et al*. (1995a) used an economic index, total factor productivity, and successfully produced statistical summaries for long-term agricultural data.

Scale and frequency of measurement

Relationships estimated for data which are grouped, for example according to area or region, will often not be the same as those relationships estimated for the

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

same data before grouping. This is known as the ecological fallacy and Steel and Holt (1996) give a review of applications where this has been found. A simple example can be seen by the fitting of Taylor's power law to berry borer infestation counts in central American coffee. For spatially distinct samples this relates the sample variances to their means according to

$$\text{variance} = \text{a} \,.\, \text{mean}^{\text{b}}$$

On the logarithmic scale, the relationship can be estimated with linear regression. When the scale of sampling is changed from the branch to the tree, to a marked area of specified size or to a larger area of twice the specified size the estimates of the regression coefficients change as shown in Table 4.

For many species the estimated Taylor's power law coefficients are known to lie within specified ranges. This is clearly dependent upon scale and the use of an accepted set of coefficients at a different scale may invalidate predictions of abundancy. Similar studies need to be done for relationships for long-term studies estimated in time where data are aggregated over discrete intervals chosen subjectively. The effect of scale has been examined mainly for regression relationships for quantitative data, but explorations are also needed for its effects upon analyses of qualitative data from socio-economic studies.

Long-term trials propose the question of how often data should be measured. An initial response is to measure annually, but depending upon facilities and the chosen variable it may be necessary to measure more or less frequently. To estimate trend the objectives must be referred to. A short-term trend indicating a pattern or response which may be used as a basis for a more detailed longer study will be valuable. A long-term trend or estimate of sustainability must be estimated with care. Monteith (1990) gives an indication of how to estimate the length of time necessary to identify trends in data, but stresses that the method depends upon the assumption that the trend is almost constant in the period of estimation and that this is unlikely to be so in practice. Riley and Alexander (1998) propose a scheme for determining core indicators and subsidiary indicators for monitoring throughout a project and beyond it in both time and space.

Table 4. Estimates of Taylor's power law coefficients for samples of different scales.

Sampling unit	$\log_{10}$a	b
Branch	0.343	1.39
Tree	0.728	1.64
Area	−0.073	1.86
Double area	−2.290	2.35

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

Data storage structures

Long-term experiments generate large amounts of data. On-station plot experiments generate agricultural and weather data at the very least. Multidisciplinary studies generate much more. Underpinning any long-term experiment is the need for good management and accurate data collection and storage. The construction of models and their combination in compounded models leads to the need for large-scale databases. Unless carefully coordinated the storage of linked databases from different disciplines can result in a database where multidisciplinary use is impossible. Inaccurate models used to underpin geographic information systems can lead to inappropriate predictions. Key database construction issues are discussed in Gaunt *et al.* (1997).

The electronic Rothamsted archive (ERA) database (Potts *et al.*, 1996) provides a format to ensure good database construction for any experimental regime. This provides secure storage for disparate data relating to the Rothamsted long-term experiments as well as contemporary data from other sources. Constructed using the Oracle database running under Unix, it uses cooperating software systems to perform specific tasks like data entry, description and extraction.

CONCLUSIONS

Historically, long-term experiments have been on-station plot-based studies and were not originally planned with the aim of determining sustainability patterns, although some are now valuably used in this respect. Pressure for multidisciplinary studies to determine impact and sustainability of both production and development is leading to interest in long-term farmer-participatory studies. Research strategies need to be flexible, incorporating detailed scientific studies and participatory community-based studies whose results can supplement each other to respond to changing priorities of farmers and of development agencies. Such multidisciplinary studies appear complex, difficult to design and maintain over long periods of time. Standard designs are likely to be relevant for only some categories of study. However simple designs can be determined which will remain robust in time if objectives are appropriately specified and if quality control is maintained through appropriate specification of data choice, training of both scientists, field operators and farmers and careful design of databases. A sound foundation will then be available for the use of modern methods of statistical analysis.

Acknowledgements. IACR-Rothamsted is grant-aided by the UK Biotechnology and Biological Sciences Research Council. Detail of the software CycDesigN can be accessed on the World Wide Web at: http://www.ffp.csiro.au/tigr/software/cycdesign/ and detail of the Electronic Rothamsted Archive (ERA) can be accessed at http://www.res.bbsrc.ac.uk/era/

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press

REFERENCES

Barnett, V. & Riley, J. (1995). Statistics for environmental change. *Experimental Agriculture* 31:117–130.

Barnett, V., Johnston, A. E., Landau, S., Payne, R. W., Welham, S. J. & Rayner, A. I. (1995a). Sustainability – the Rothamsted experience. In *Agricultural Sustainability. Economic, Environmental and Statistical Considerations*, 171–206 (Eds V. Barnett, R. Payne & R. Steiner). Chichester: Wiley.

Barnett, V., Payne, R. W. & Steiner, R. (Eds) (1995b). *Agricultural Sustainability. Economic, Environmental and Statistical Considerations*. Chichester: Wiley.

Cox, D., Fitzpatrick, R., Fletcher, A. E., Gore, S. M., Spiegelhalter, D. J. & Jones, D. R. (1992). Quality-of-life assessment: can we keep it simple? *Journal of the Royal Statistical Society, A*, 155:353–394.

Dyke, G. V. (1988). *Comparative Experiments with Field Crops*. New York: Oxford University Press.

Fielding, W. J. (1992). Damage assessment by eye: some Caribbean observations. *Field Crops Research* 30:183–186.

Fielding, W. J & Riley, J. (1997). How big should on-farm trials be and how many plots should be assessed? *PLA Notes* 29:19–22.

Gaunt, J. L., Riley, J., Stein, A. & Penning de Vries, F. W. T. (1997). Requirements for effective modelling strategies. *Agricultural Systems* 54:153–168.

Kmietowicz, Z. W. (1995). Accuracy of indices of industrial production in developing countries. *The Statistician* 44:295–308.

Leigh, R. A., Prew, R. D. & Johnston, A. E. (1994). The management of long-term agricultural field experiments: procedures and policies evolved from the Rothamsted classical experiments. In *Long-Term Experiments in Agricultural and Ecological Sciences*, 253–268 (Eds R. A. Leigh & A. E. Johnston). Wallingford, Oxon: CAB INTERNATIONAL.

McWalter, A. R. & Wimble, R. H. (1976). A long-term rotational and manurial trial in Uganda. *Experimental Agriculture* 12:305–317.

Monteith, J. L. (1990). Can sustainability be quantified? *Indian Journal of Dryland Agricultural Research and Development* 5:1–15.

Patil, G. P., Gore, S. D. & Sinha, A. K. (1992). Environmental sampling and statistical modelling with examples. In *Proceedings of the XVIth International Biometric Society Conference, Hamilton, New Zealand, 7–11 December 1992*. Washington: International Biometric Society.

Potts, J. M., Verrier, P. J. & Payne R. W. (1996). The Electronic Rothamsted Archive and its use in agricultural research. In *Proceedings of ICCTA'96/VIAS/NNAA Congress on ICT Applications in Agriculture, Wageningen, June 16–19, 1996*, 171–176. Wageningen, The Netherlands: ICCTA.

Pretty, J., Guijt, I., Thompson, J. & Scoones, I. (1995). *Participatory Learning and Action. A Trainer's Guide*. London: International Institute for Environment and Development.

Riley, J. & Alexander, C. J. (1997). Statistical literature for participatory on-farm research. *Experimental Agriculture* 33:73–82.

Riley, J. & Alexander, C. J. (1998). Guidelines for an assessment method for the optimum uptake of research. *Journal of Sustainable Agriculture* 12:99–117.

Shepherd, K. D. & Roger, J. H. (1991). *Approaches to On-Farm Testing and Evaluation of Agroforestry Technology. ICRAF Working Paper No. 67*. Nairobi: International Council for Research in Agroforestry.

Steel, D. G. & Holt, D. (1996). Analysing and adjusting aggregation effects: the ecological fallacy revisited. *International Statistical Review* 64:39–60.

Steiner, R. A. (1995). Long-term experiments and their choice for the research study. In *Agricultural Sustainability. Economic, Environmental and Statistical Considerations*, 15–22 (Eds V. Barnett, R. Payne & R. Steiner). Chichester: Wiley.

Williams, E. R. & Matheson, A. C. (1994). *Experimental Design and Analysis for Use in Tree Improvement*. Melbourne: CSIRO.

https://doi.org/10.1017/S0014479700002039 Published online by Cambridge University Press