The Analysis of Results from Paired Paddock Comparisons

David R. Kemp
University of Sydney, Australia

P. M. Dowling
NSW Agriculture, Australia

D. L. Michalk
NSW Agriculture, Australia

G. Millar
NSW Agriculture, Australia

The XIX International Grassland Congress took place in São Pedro, São Paulo, Brazil from February 11 through February 21, 2001.

Proceedings published by Fundacao de Estudos Agrarios Luiz de Queiroz

ID#32-10

THE ANALYSIS OF RESULTS FROM PAIRED PADDOCK COMPARISONS

D.R. Kemp[1], P.M. Dowling[2], D.L. Michalk[2], and G.D. Millar[2]

CRC for Weed Management Systems, Pasture Development Group

[1]University of Sydney, Faculty of Rural Management, david.kemp@orange.usyd.edu.au

[2]NSW Agriculture, Orange Agricultural Institute; Orange NSW 2800 Australia.

Abstract

Paired-paddock comparisons are a common way of evaluating new grassland practices at a scale more relevant to farmers. They are also being used to replace or augment designed experiments and can be the only source of information available on a practice. However, it is often uncertain if the differences between paddocks are significant. Importantly, a current trend among funding organisations is to support paddock comparisons. The need for valid procedures to compare unreplicated treatments is increasingly urgent. It is suggested that a range of tools be used to infer statistical significance from using typical error values from related studies or subsampling, through to multivariate techniques to follow trends. Local 'rules of thumb' could be developed and data evaluated with calibrated models. A final judgement on treatment effects would need to be based upon the use of several criteria to achieve a 'balance of probabilities'. Consideration of these problems suggests that paired-paddocks should only be used to evaluate contrasting treatments where large effects are expected and not small variations within a practice.

Keywords: paired-paddocks, evaluation, multivariate, inference

Introduction

Grassland research tends to follow a common sequence. Initially ideas are explored in small plots, the more successful are then scaled up to paddocks and then ideally tested in a grazing system. Paddock-scale grazing experiments are usually done once the range of treatments is narrowed down and when the technology involved is sufficiently different to common practice that it needs to be fully evaluated. The expense of running paddock-scale grazing experiments often limits their use.

Paired-paddock comparisons have been widely used in grassland studies for many years. They are often used to: demonstrate at a local farm scale the better treatments coming from research programs *e.g*. as a simple contrast with some standard common practice; bypass designed experiments and test ideas coming from smaller scale studies; extend the range of treatments and information in parallel studies to a core experiment; and evaluate on-farm innovations developed by farmers. In each case there is a problem in evaluating the significance of the results obtained *i.e*. are the differences real? This paper aims to outline some of the issues involved and consider some methods that could help to form an opinion as to the scientific merits of any paddock differences obtained.

Criteria for evaluation

Grasslands are ecosystems that are typically assessed using a range of criteria *e.g*. saleable animal products, water balance, nutrient leakage and the retention of desirable species.

To enable a reasonable comparison it is assumed that paired paddocks have been established using best practice *i.e*. a random allocation of treatments and there are as many factors in common as possible so that the paddock comparison better reflects the treatments.

Before applying treatments, some initial measurements of the system should always be taken to see if the paired paddocks are similar, else the results will be suspect.

Saleable animal products enable an economic evaluation to be done on treatments. However, a simple evaluation of animal products often contains the assumption that the stocking rate and grazing pressure were optimal for that system. It is doubtful if that is commonly the case. A better analysis could be done if other data were available on the likely response curve to stocking rates for that system so that the position of the paired-paddocks on the response curve could be estimated. The data obtained could also be tested against established, calibrated models (Donnelly *et al*., 1997) provided sufficient measurements are taken to run the model effectively.

To decide if the stocking rates were optimal, additional local criteria could be used. These criteria would include the levels and quality of herbage mass available throughout the year – to test if the grassland was being over- or under-grazed. Tools such as the 'Pasture Management Envelope' (Spain *et al*., 19; Kemp *et al*. 1991) can be developed from local and research information to decide upon appropriate boundary values within which grasslands are being effectively managed and as a way of testing if stocking rates were appropriate.

The use of animal data requires that each paddock of the comparison is a closed system. If the animals are off the paddocks for any significant periods then animal performance information can only be effectively used for the periods they are in the paddocks.

Total grassland production provides another measure of the performance of each paddock. Ideally, this needs to be the net primary growth of the grassland determined with the use of exclusion cages. Standing herbage mass is of less value as that is the balance between herbage growth and animal consumption which then complicates any evaluation.

Grassland composition often determines the animal, sustainability and biodiversity performance of the system and provides a valuable way of assessing the impact of treatments. Problems can arise though if only one component is monitored without due consideration of how the whole grassland ecosystem is responding to treatments. A significant change in one component may not always be reflected in significant changes in other components and this results in the possibility of a type II statistical error.

Other criteria such as components of the water balance, nutrient status and pathways will depend upon the aims of a paired-paddock comparison. In this and the other cases considered above, they need to be measured with sufficient accuracy to reflect the status and trends in each paddock under study. Often parameters are not measured with the intensity and rigour required to get a reasonable estimate of the mean.

Statistical inferences

Commonly, paired-paddock comparisons involve two treatments and no replication. The main information available from which inferences can be drawn is the difference between the paddocks. The level of error that could be used to test the significance of any difference usually needs to be inferred from other sources.

Subsample errors can be estimated for criteria such as herbage mass or variation in a soil nutrient *etc*., from the samples taken within each paddock to give an indication of the general variation likely for that parameter. Other estimates could come from related experiments done in a similar environment and at a similar scale, particularly when there is only one estimate of a parameter from each paddock *e.g*. water yield or bulked animal product. This approach has some affinity with modern decision analysis based on Bayesian theory – an approach that does not appear to have been widely used in grassland studies. An additional approach that has not been widely used is bootstrapping where sampling information can be used to estimate errors.

Many paired-paddock comparisons are done by advisory staff and, or farmers who are not familiar with statistical procedures. In this case, it can be better for research staff to provide some general guidance on the differences that are typically required to infer significance. Often the coefficients of variation and measures such as the least significant difference (as a percentage of the mean) are large in large paddock comparisons, particularly where the intensity of subsamples is low; *e.g*. coefficients of variation of >50% could be expected in many cases based upon experiments where treatment and replication numbers are low. A local review of related experiments would be required to decide a 'rule of thumb' that can provide guidance. A reasonable first guess in many paired-paddock comparisons of different grazing systems would be that any differences would need to be >30% of the 'control' to approach significance. A corollary of this is that paired-paddock comparisons are really only appropriate for comparing major contrasts and not for sorting out more subtle variations on a management practice.

Where there are several factors operating in concert *e.g*. species composition data, there are statistical techniques available (*e.g*. non-metric multidimensional scaling) that can be used to monitor trends and give some indication of treatment impacts. Ordination techniques have been used by ecologists for many years to analyse data that are similar in structure to that from paired-paddock comparisons *i.e*. no replication and none or few sample sites. Certainly, the power of the analysis is increased with more sites. Experience with such procedures (Kemp *et al*., unpublished data) has generally shown that only treatments that have been significant in experiments have shown separation in a multivariate analysis. These techniques are more useful if data is collected over time so that trends become more apparent. They also have the advantage over conventional univariate statistics that trends in the whole community are portrayed reducing the risk of type II errors mentioned earlier. The use of multivariate statistics does require that researchers be involved in analysing paired-paddock data sets, but they do provide additional support to the rules of thumb suggested above.

Discussion

Paired-paddock comparisons can be a useful supporting tool in providing original or additional information about innovations in grassland management. They are not a replacement for properly constructed experiments, though unfortunately there has been a recent tendency in Australia to do that. The ability to decide if the differences between paddocks are likely to be significant will depend upon the quality of the data obtained *i.e.* intensity and frequency of measurements, the criteria used to compare paddocks and then the use of a few different approaches to estimate significance. The evaluation of paired-paddock comparisons could draw on the analogy with law – that of deciding upon the 'balance of probabilities'. The more criteria that can be employed in making a judgement, the more confidence the users would have in the results. Investigators using paired-paddocks should aim to improve the quality of the data obtained than is often the case, to form a better judgement as to treatment effects.

Acknowledgements

The difficulties of evaluating paired-paddock comparisons have been the subject of numerous discussions with colleagues within the CRC for Weed Management Systems and the Sustainable Grazing Systems program coordinated by Meat and Livestock Australia.

References

Donnelly, J.R. (1997). Decision support: Delivering the benefits of grazing systems research. Proc. XVIIIth Int. Grassld. Congr., Canada, Session 26.

Kemp, D.R. (1991). Defining the boundaries and manipulating the system. Proc. 6th Ann. Conf. Grassld. Soc. of New South Wales, 24-30.

Spain, J., Pereira J.M., and Gualdron R. (1985). A flexible grazing management system proposed for the advanced evaluation of associations of tropical grasses and legumes. Proc. XVth Int. Grassld. Congr., Kyoto, Japan, 1153-5.