Chapter (non-refereed)

Lindley, D. K.; Howard, D. M.. 1988 Some statistical problems in analysing cotton strip assay data. In: Harrison, A. F.; Latter, P. M.; Walton, D. W. H., (eds.) *Cotton strip assay: an index of decomposition in soils.* Grange-over-Sands, NERC/ITE, 25-27. (ITE Symposium, 24).

Some statistical problems in analysing cotton strip assay data

D K LINDLEY and D M HOWARD
Institute of Terrestrial Ecology, Merlewood Research Station, Grange-over-Sands

1 *Summary*
The statistical methods suitable for data derived from cotton strip assay are those which are also appropriate for soil profile data, and the problems encountered are similar. Precision can be increased by using stratified sampling for survey work and blocking for designed experimental studies. The need for adequate replication of the experimental plots, rather than increasing the number of replicate strips, is emphasized. Repeated measurement techniques will be required to analyse depth and time effects, and multiple regression may help to explain differences between sites.

2 *Introduction*
Some of the statistical problems in analysing data from the cotton strip assay are inherent, whilst some are enlarged by the observer or experimenter. The distinction between these 2 types is made because the observer obtains the information by a survey approach, whereas the experimenter collects the information from a designed experiment.

The problems can be grouped under the following headings:

i. within site variability;
ii. problems with depth and time;
iii. the use of appropriate experimental designs;
iv. use of the global approach in bringing diverse data together.

3 *Within-site variability in surveys*
The survey approach to cotton strip assay studies in soil suggests a situation where a number of strips are placed at random on a site and the scientist is only concerned with an average measure of loss of tensile strength. A mean and standard error are initially calculated for a specific site. Inspection of the first 78 sets of data collected together for the Workshop shows a high degree of variability on many sites. Using the coefficient of variation (CV), SD/$\bar{x}$, as the measure of variability and looking at the top and bottom of strips only, Figure 1 shows the distributions of the 2 sets of CVs. The top substrips show more variability than those from the bottom substrips, and some extremely variable sites have been encountered. The CV has no use in testing or estimating, but does provide a basis for appreciating the precision possible in an experiment, thereby aiding decisions on the size of the sample and allowing for comparisons of variability.

Because of the considerable variation encountered in the past on individual sites, it may be possible, in any future work, to group the sampling points into strata, in such a way that variation within a stratum can be expected to be less than variation between the strata. If successful, this method will increase the precision by which the site means can be measured.

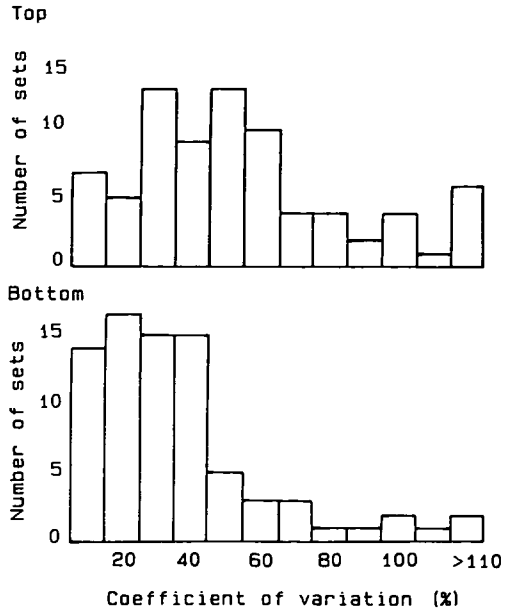

Figure 1. Histograms of coefficients of variation of top and bottom substrips of 78 decomposed cotton strips

4 *Experimental design*
At the start of every experiment, the question to be answered must be clearly defined. In other words, the experiment must always have a preliminary idea or a hypothesis to test. For example, the question may be: 'do the conditions of felling and not felling affect the decomposition rate of cotton strips inserted in the soil?' Some thought must also be given to the population of interest, so that the conclusions drawn are not used outside this population.

Ineson *et al.* (1988) noted that there are a large number of sets of data available about which there is knowledge of variability. This knowledge should, if possible, be used to help in the design of experiments. Is there any point in carrying out an experiment with inadequate replication in which, because of the inherent variability of the data, it is impossible to detect differences between treatments? This approach could be justified, if the experiment was planned as a pilot trial.

4.1 Example
The critical point about the importance of adequate replication can be illustrated by the following example.

It is assumed that there are 2 treatments, felling (F) and not felling (NF), and that these 2 treatments (t) have both been allocated to 3 blocks (r). A block, by definition, is a physical unit containing one complete replication of treatments. Plots within blocks must be homogeneous and differences between blocks made to account for as much as possible of the systematic variation between plots.

Five cotton strips (s) can be randomly placed in the 6 experimental plots so that the 30 cotton strips in the field form the experiment. On completion of the experiment, the following data were produced (Table 1). The grand mean is a tensile strength of 16.40. For treatment F, the mean is 19.29, and for NF it is 13.51. Can the differences between the 2 treatments be detected? Table 2 presents a skeleton analysis of variance (Steel & Torrie 1980).

Table 1. Example of a cotton strip assay data set prior to analysis

Treatment		Blocks I	II	III	
Felling replicate	1	23.67	19.99	21.74	
	2	25.28	17.17	15.25	
	3	25.63	9.34	16.96	
	4	10.92	19.99	22.87	
	5	13.22	23.18	24.19	
	Σ	98.72	89.67	101.01	
	x̄	19.74	17.93	20.2	x̄F = 19.29
Not felling replicate	1	18.67	11.89	17.90	
	2	8.00	8.08	11.98	
	3	7.13	14.45	14.36	
	4	13.90	18.32	21.04	
	5	15.60	8.25	13.13	
	Σ	63.30	60.99	78.41	
	x̄	12.66	12.20	15.68	x̄NF = 13.51

Table 2. Skeleton analysis of variance

Source of variation		Degrees of freedom
Blocks	r–1	2
Treatments	t–1	1
Experimental error	(r–1) (t–1)	2
Sampling error	rt (s–1)	24
Total		29

The F ratio or variance ratio used to detect whether or not there is a significant difference is the ratio of the treatment MS/experimental error MS, in this case 250.56/4.11 = 60.92. At the 5% level, the level in the F tables is 18.51. Therefore, the result is significant. It is important to note that analysis of variance is a statistical technique for testing differences between 2 or more treatment *means*. If the analysis is reworked, based on the means, the skeleton analysis shown in Table 3 is obtained. The degree of significance is the same.

Table 3. Skeleton analysis of variance based on means

Source of variation		Degrees of freedom
Blocks	r–1	2
Treatments	t–1	1
Experimental error	(r–1) (t–1)	2
Total	rt–1	5

The important statistical point to observe is that, by taking 5 samples within each plot, the experimental (treatment) replication has not been increased.

It would have been more satisfactory to deploy the 30 cotton strips over 5 blocks, with 3 strips allocated to each plot. In this way, the replication of the treatments has been increased from 3 to 5. The degrees of freedom for the experimental error would then rise from 2 to 4. However, it is recommended that the degrees of freedom for experimental error should be 10–20.

5 *Problems with depth and time*

When analysing the results of surveys and designed experiments using cotton strip assay, the structure of the sources of variation is often oversimplified. Simplification is usually one of 2 types: (i) analysing factorial experiments as if completely randomized, and (ii) ignoring the correlations of errors induced by sampling the same cotton strip at different depths. The first type of mistake leads to distorted probabilities in making inferences about either treatment or depth factors. The second type constitutes a serious mistake only if the correlations are not homogeneous between depths, and affects only the tests for differences between depths.

The method for analysing data which lack independence has been presented in a number of sources (Winer 1971; Gill & Hafs 1971). An appropriate analysis of variance is that for a split-plot design, as comparisons between treatments (between cotton strips) are free to vary more than comparisons between depths (within cotton strips).

For an effect lacking independence (depth or interactions), the degrees of freedom for both numerator and denominator of the global F tests (eg 'all levels are equal' *vs* 'at least 2 differ') should be multiplied by a correction factor, epsilon (Greenhouse & Geisser 1959). However, Boik (1981) has shown that the adjustment is limited to global F tests and should not be used for specific contrasts. He recommended testing each contrast against its own variance. Barcikowski and Robey (1984) recommended also using partitioned errors in *a posteriori* tests.

During the course of the Workshop, we will have seen several examples of how the cotton strip assay performs over time. Marked seasonal trends have been detected on several sites (Brown & Howson 1988; Lawson 1988). In these situations, a number of cotton strips would have been sampled at specified times throughout the season, and a site mean produced for each time. If there are several sites and the question is whether there are differences between the trends at different sites, it may be possible to subdivide the main effect of time into polynomial components and to test whether they have a significant interaction with sites.

6 *Use of the global approach in combining diverse data*

In decomposition studies of cotton strips in the tundra, Heal *et al.* (1974) successfully used the techniques of multiple regression and principal component analysis to 'explain' tensile strength losses on 24 sites. However, we have no means of knowing, from their paper, whether their predictive equation would be successful if it were applied to an independent set of data. In the intervening interval since 1974, their equations have been used on another data set (Smith & Walton 1988), and possibly also by other workers.

A major problem in combining diverse data sets is the difficulty in obtaining a complete data set. It is one of the benefits of the Workshop that an attempt can be made to construct a useful data set. Ineson *et al.* (1988) demonstrate the application of multiple regression techniques to the data from the Workshop.

If sufficient complete sets can be assembled, it would be valuable to construct predictive multiple regression models for the major geographical regions, eg tundra, tropics, temperate zone, etc. Provided that these models can be successfully assembled, then we should be able to test for differences between the regions. A useful approach to adopt would be a detailed study of the residuals, ie the difference between the observed and predicted value for each site. Extreme values, either positive or negative, could be examined in detail, and further clues might emerge to explain any large difference obtained at that site.

7 *Conclusions*

In this broad review, we draw attention to some statistical problems raised in the collection and analysis of cotton strip assay data. It has been noted that a high degree of variability can be expected when cotton strips are laid down on individual sites. Notice should be taken of this variability when organizing surveys and designing experiments, so that they are adequately replicated. Multiple regression techniques are suggested as being one of the better ways of comparing data from diverse sources, in spite of all the difficulties likely to be encountered in obtaining a complete data set.

8 *References*

Barcikowski, R.S. & Robey, R.R. 1984. Decisions in single group repeated measures analysis: statistical tests and three computer packages. *Am. Statistn,* **38,** 148-150.

Brown A.H.F. & Howson G. 1988. Changes in tensile strength loss of cotton strips with season and soil depth under 4 tree species. In: *Cotton strip assay: an index of decomposition in soils,* edited by A.F. Harrison, P.M. Latter & D.W.H. Walton, 86-89. (ITE symposium no. 24.) Grange-over-Sands: Institute of Terrestrial Ecology.

Boik, R.J. 1981. *A priori* tests in repeated measures designs: effects of nonsphericity. *Psychometrika,* **46,** 241-255.

Gill, J.L. & Hafs, H.D. 1971. Analysis of repeated measurements of animals. *J. Anim. Sci.,* **33,** 331-336.

Greenhouse, S.W. & Geisser, S. 1959. On methods in the analysis of profile data. *Psychometrika,* **24,** 95-112.

Heal, O.W., Howson, G., French, D.D. & Jeffers, J.N.R. 1974. Decomposition of cotton strips in tundra. In: *Soil organisms and decomposition in tundra,* edited by A.J. Holding, O.W. Heal, S.F. MacLean & P.W. Flanagan, 341-362. Stockholm: Tundra Biome Steering Committee.

Ineson, P., Bacon, P.J. & Lindley, D.K. 1988. Decomposition of cotton strips in soil: analysis of the world data set. In: *Cotton strip assay: an index of decomposition in soils,* edited by A.F. Harrison, P.M. Latter & D.W.H. Walton, 155-165. (ITE symposium no. 24.) Grange-over-Sands: Institute of Terrestrial Ecology.

Lawson G.J. 1988. Using the cotton strip assay to assess organic matter decomposition patterns in the mires of South Georgia. In: *Cotton strip assay: an index of decomposition in soils,* edited by A.F. Harrison, P.M. Latter & D.W.H. Walton, 134-138. (ITE symposium no. 24.) Grange-over-Sands: Institute of Terrestrial Ecology.

Smith, M.J. & Walton, D.W.H. 1988. Patterns of cellulose decomposition in four subantarctic soils. *Polar Biol.* In press.

Steel, R.G.D. & Torrie, J.H. 1980. *Principles and procedure of statistics. A biometrical approach.* New York: McGraw-Hill.

Winer, B.J. 1971. *Statistical principles in experimental design.* New York: McGraw-Hill.