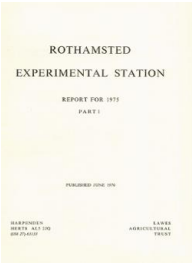
ROTHAMSTED RESEARCH

Report for 1975 - Part 1

Full Table of Content



Statistics Department

J. A. Nelder

J. A. Nelder (1976) *Statistics Department ;* Report For 1975 - Part 1, pp 321 - 333 **- DOI: https://doi.org/10.23637/ERADOC-1-132**

STATISTICS DEPARTMENT

J. A. NELDER

Staff

Head of Department J. A. Nelder, D.SC.

Senior Principal Scientific Officers

D. A. Boyd, PH.D.
J. C. Gower, M.A.

Principal Scientific Officers

B. M. Church, B.SC.
J. H. A. Dunwoody, B.A.
F. B. Leech, D.SC., M.R.C.V.S.
W. J. Lessells, B.SC.
G. J. S. Ross, B.A.
H. R. Simpson, M.A.
R. H. Wimble, M.SC.

Senior Scientific Officers

N. G. Alvey, M.I.S.
B. Gilliver, B.SC.
R. A. Kempton, B.A., B.PHIL.
H. H. Spechter, PH.D.
A. J. Vernon, B.SC.

Higher Scientific Officers

C. F. Banfield, B.SC.
Mary G. Hills, B.A.
R. W. Payne, B.A.
A. D. Todd, M.SC.

Scientific Officers

Mrs. Jill F. B. Altman, M.I.S.
R. J. Baker, M.SC.
C. J. Dyer, B.SC.
P. W. Lane, B.A.
Penelope K. Leech
Janet Riley, M.SC.
K. Ryder, M.I.S.
P. E. Sparrow, M.SC.
R. P. White, B.SC.
J. Wood, M.SC.

Assistant Scientific Officers

Mrs. Janet I. Elsmere
Linda C. Jackson
Mrs. Pamela Tett

Secretary

Ruth T. Hunt

Clerical Staff

Mrs. Doris A. Court
Mrs. Elsie Davies
Mrs. Janet A. Day
Mrs. Jennifer D. Dewhirst
Mrs. Janet Harrison
Susan T. I. Hyde
Linda M. King
Mrs. Elizabeth H. Mayton
Mrs. Patricia K. Neill
Mrs. Sheena Perry
Mrs. Jean H. Povey
Genevieve M. A. Rogers
Mrs. Brenda J. Watler
Janice R. Woods

Introduction

Much of our work involves collaboration with workers in many disciplines, and at all levels from the small statistical query to joint projects extending over several years. This collaborative aspect is supplemented by our work on general computer programs, now increasingly used throughout the world, and by research into statistical theory, most of which arises directly from the problems we encounter. This report makes no attempt to list all our activities, confining itself to exemplifying the main lines of our work; if day-to-day activities are under-represented this should not be taken to mean that they are unimportant, only that their diversity defies adequate summarisation.

Practical applications

Collaboration with other departments. We are experimenting with the allocation of individuals to other departments to act as specialist statisticians, and to be the main contact between those departments and ourselves. Three such contacts have been set up and first reactions are favourable on both sides; some examples of work arising from the scheme follow.

https://doi.org/10.23637/ERADOC-1-132

The population dynamics of the potato cyst-nematode are being studied using data from long-term rotation experiments at Woburn and results from pot experiments. A model has been set up to describe the changes in egg density from year to year when potatoes are grown continuously, and has been successfully fitted to data from pot experiments. The model is being extended to study the effect of crop rotation, and also the competition factor, when two species of potato cyst-nematode are present.

Work has begun on the problem of estimating mortality due to fungal parasitism in the cereal cyst-nematode. Multivariate techniques have been used to describe the affinity between populations of potato cyst-nematode based on larval and cyst measurements, and a paper is in preparation on the use of canonical variate analysis in separating the two known species of this nematode. (Kempton with Jones, Kerry, Evans, Franco and Stone of Nematology Department)

Some strains of mildew are becoming resistant to fungicide. An investigation of this involved measuring the response of different strains of mildew to fungicide dose using probit analysis techniques. Initially probit lines were fitted to the numbers of spores reaching a certain stage of development, but this involved use of a microscope which was tedious and time-consuming. Consequently it was decided to simply count the numbers of infections which can be assessed by eye and to use the technique of Wadley's problem to estimate the total number of spores. This increased the variability per test, but increased the final accuracy because it allowed more data to be collected and more tests to be performed.

Tolerance was not a result of any detectable increase in immunity in the population but seemed to be more an upward shift of the probit line with the slope remaining roughly the same. Immunity detected in the field experiment was not consistent within plots with the same strain of the fungus and same seed dressing treatment although other parameters were consistent. This suggests that the apparent immunity might be an artefact, possibly because the leaves in the test were not completely covered with the fungicide. There were no detectable changes between samples taken at the beginning of June and those taken at the beginning of July which suggests that the initial population in a crop tends to determine the population for the rest of the season. (Payne, with Bainbridge and Smith of Plant Pathology Department)

From a theory supplied by O. Talibudeen (Chemistry Department) for the diffusion of potassium in soil over long periods of time, a three-compartment diffusion model has been formulated, and fitted to data using the Maximum Likelihood Program. The results can be displayed by the graph plotter. A paper is being prepared. (Lane, with Beasley, Computer Department)

Multivariate analysis. The techniques of principal coordinate analysis, canonical variate analysis and centroid cluster analysis were used on three qualitative field observations and nine quantitative laboratory measurements made on each of 88 samples of peat. Suitable criteria were required for the classification of peat samples and the multivariate techniques were found useful for grouping the samples on their similarities. Linear discriminant functions were found for distinguishing the different groups of peat. Procrustes rotation was used to establish whether the field observations and laboratory measurements characterised the samples in a similar way, but this was found not to be so. (Banfield, with Bascomb, Soil Survey)

Canonical variates analysis was also applied to six variates measured from a gas chromatograph of the anal sacs from 51 foxes. It was established that the heights of peaks in the chromatograph could be used as variates for discriminating between different groups of foxes. (Banfield, with Mr. H. J. H. McFie of the Meat Research Institute)

Our advice was sought on the use of multivariate and clustering techniques in a survey

https://doi.org/10.23637/ERADOC-1-132

of the economic and agricultural resources of certain regions of India. Classifications of the regions based on different criteria needed to be compared using recently developed techniques for comparing dendrograms. A study of the way variates for the regions change with time was also investigated by Procrustes rotation of the configurations of regions produced by principal coordinate analysis at different times.

An experiment on the effect of nitrate on nitrogen uptake produced 27 variates measured on eight sets of 60 plants. Five or six different levels of nitrate had been applied to groups of six plants and then the plants were dissected and analysed for nitrogen content at different times of the year. Principal component analysis was used to see if the plants clustered according to the level of nitrate applied, and linear discriminant functions were sought using canonical variate analysis to separate the groups of plants. However, plants were not found to form distinct clusters according to their nitrate application. (Banfield, with Mr. L. Taylor and Dr. J. Dancer of Portsmouth Polytechnic)

Fertiliser requirements of crops. Recent increases in the price of phosphate relative to crop value caused us to look at current recommendations in *ADAS Bulletin* 209. Examination of recent experiments suggests that recommended rates ought now to be about half those of *Bulletin* 209. A set of recommendations had been produced to allow for future changes in the relative price of phosphate. (Wood)

Among the series of experiments summarised during 1975 were those involving nitrogen on cereals, phosphate on potatoes and magnesium on herbage; in particular, Sparrow is preparing papers on the nitrogen response of spring barley (with Mr. P. Needham, ADAS), the effects of long-term residues of phosphorus fertilisers (with Mr. R. D. Russell, ADAS) and the effect of lime on magnesium uptake by herbage (with Mr. E. Jones, ADAS).

Forecasting virus yellows in sugar beet. Past data on the percentage of virus yellows and its possible relation to weather in the preceding winter and spring was re-examined, using the new facilities of the GLIM program. The superiority of the complementary log-log transformation over the traditional angular transformation for data on disease incidence was confirmed, but attempts to improve on the original predictors, number of frost days and mean April temperature, by using various forms of accumulated temperature, were not successful. Though any sensible predictors based on pre-1974 data would have forecast 1974 as a bad year for virus yellows, they all underestimated considerably the actual incidence of the disease. (White, with Dr. M. A. Watson)

Livestock experiments. While the amount of new data on livestock experiments, mostly from ADAS, shows no sign of diminishing, it is encouraging to be able to report that the backlog has this year been entirely cleared. Experiments dealt with included those on dairy cattle, beef cattle, calves, pigs and sheep. The following paragraphs exemplify this aspect of our work.

Work has continued in the development of a flexible set of program modules, using the more advanced facilities of Genstat, for the analysis of the bulky data arising from milk experiments. It has been possible to standardise most of the tasks and a standard run now provides for analysis of variance and covariance of milk yield and milk composition data for time periods specific to the experiment, the merging of the yield and composition data for the calculation and analyses of yields of milk components and SCM yield, tabulation of food data, graphs of milk yields, both actual and cumulative, liveweights, and weekly food intakes. The graphs show both the individual treatment means and the means of each factorial treatment factor. The type of analysis is determined by pre-set parameters, and restriction to a subset of the data is possible. (Spechter)

https://doi.org/10.23637/ERADOC-1-132

An analysis was made of data from eight commercial farms assessing the value of oestrus control applied to suckler herds, and investigating the effect of level of feeding on fertility. Whereas fertility rates showed a clear response to feeding there was no evidence that it was associated with either weight or condition of cow. (Altman and Lessells)

We were also involved in the analysis of results from a large-scale dairy-cattle experiment sponsored by the ARC and done by Mr. C. L. Johnson at Leeds University. This was carried out over four years, involved 96 cows, and compared different patterns of feeding throughout the lactation. The results will produce valuable evidence of how the effect of subsequent feeding pattern varies with initial yield. (Lessells and Spechter)

The experiments at four EHF's to relate the level of feed concentrate with milk production continued, and showed interesting differences in the response of cows at different centres; reasons for these differences will require further investigation. (Lessells and Spechter) Data on 80 miscellaneous silages allowed predictive equations to be developed for estimating digestibility from laboratory determinations, and their accuracy to be determined. (Altman)

Surveys. We continue to be consulted about, and to deal with data from, surveys on a range of topics. Our major commitment remains the Survey of Fertiliser Practice.

Fertiliser practice. The final figures for the extended survey of 1974 were completed in January and published in Part 2 of the *Rothamsted Report for 1974* (pp. 195–199) as well as in the usual form of a report to the ADAS Soil Analysis and Fertiliser Committee (Paper 1). Further analyses of the 1974 data classifying them by farm type region (Paper 2) and compound fertiliser type have been done. A draft report on fertiliser use on grassland was circulated in October and a final version has been prepared (Paper 3). This summarises work we have done on grass used for grazing and conservation on holdings of different farm type and size. We have tried to link an estimate of the stocking rate for each farm with the amount of N used on each grass type; the results are sufficiently encouraging for us to attempt improvements in the detailed calculations for 1976. Much work remains to be done on the grassland information where preliminary estimates for 1975 show that increases in the use of N continue despite recent price rises. The 1975 survey, which was based on 1300 farms and was organised and funded as in 1974, also shows a decrease in the use of phosphate, particularly on winter cereals. This change was to be expected as the price of phosphate has trebled and some firms already have compounds containing less P on the market.

Considerable time has been spent on advance planning for 1976, when preliminary figures are required by 1 October (two months earlier) and metric units will be used for the first time. (Church and Hills)

Milk quality. The field work for the National Milk Quality Survey was completed and we now have all the data on magnetic tape. Routine monthly and updating analyses were completed and circulated, and we have done some 20 analyses investigating relationships between different tests, inter-laboratory differences, associations between features of milking technique and the keeping quality of the milk produced, and have produced tables showing numbers of 'failures' in samples produced by individual farms. We have also provided a summary sheet for despatch to each farmer, giving the results from his farm with national averages for comparison. This work is continuing and is expected to lead to the publication of a monograph. (F. B. Leech and P. K. Leech)

Other surveys. The data for the survey of dystokia in Friesian heifers (*Rothamsted*

https://doi.org/10.23637/ERADOC-1-132

Report for 1973, Part 1, 222) are now complete. Assistance was given with a survey of animal diseases in Nigeria. (F. B. Leech and P. K. Leech)

Routine analysis. Much of our effort goes into routine data processing, chiefly analysis of experimental results; this service is used by almost all the other departments at Rothamsted, ADAS, other outside organisations and, through our ODM unit, by overseas workers.

The period between harvest and Christmas has always been a busy one but the seasonality of this work has been accentuated in recent years as ADAS Regions and EHFs have set earlier deadlines for their annual reporting of experimental results. Autumn 1975 has been especially difficult as much poultry work from Gleadthorpe had to be processed at the time of peak demand for crop results. An additional burden resulted from an ADAS request for help in analysing the results of the tomato experiments done at EHSs in 1975. In all we analysed just over 1000 replicated experiments for ADAS.

The volume of work rose by 15% compared with 1974, and the total amount of data reached 1·73 million items.

With the increase in work, and a 15% decline in assistant staff, those concerned did well to hold the average turn-round time to within half a day of the 1974 figure. (Dunwoody, Dyer, Riley, M. T. Simpson and Todd)

Statistical programming

Genstat. Two releases were made, 3.05 in January and 3.06 in June. Now that the system is becoming available on several machine ranges, it is vital for its subsequent upkeep that the time that convertors have to spend in mounting new releases should be cut to a minimum. The code has now been extensively annotated to contain the variants required for different machines, and an improved version of the conversion program generates just those alternatives required for each range. The CDC 7600 version of release 3.05 was completed by Dr. K. Y. Kwok of the Manchester Regional Computer Centre in a few weeks, compared with the many months required for the original conversion. A version for the Univac 1108 was prepared by Mr. R. Cormier of the Statistical Research Service of Agriculture in Ottawa in about three months. A version of release 3.03 for the ICL 1906A was completed by Dr. P. Griffiths of the Oxford University Computing Laboratory, and released to a selected number of university installations; this version has produced many problems, but again most of the necessary variants have been incorporated in the latest release. We are greatly indebted to our convertors for their work in making Genstat available on other machine ranges.

The latest release (3.07) became available at Rothamsted in December, and tapes have been sent to the convertors. The IBM 360/370 version, which we produce ourselves, has been delayed by the transfer of the Edinburgh 370 computing service to the machine at Newcastle upon Tyne.

Work has begun on the transfer of release 3.07 to the ICL 4–75 at Edinburgh using the FORTE compiler developed by Mr. G. E. Millard. (Simpson) Experimental use of an earlier version on the ICL 4–75, a paged machine, has shown that interactive use, even of such a large program as Genstat, is possible and reasonably efficient, given a suitable operating system and software. Genstat is now being used more than 100 times a day, both at Rothamsted and on the CSIRO machine in Canberra.

New facilities. A new option allows structures to be concatenated when printing, to give increased flexibility in presenting the output. A corresponding directive JOIN allows

https://doi.org/10.23637/ERADOC-1-132 *pp 6*

the results to be stored rather than printed, giving the user the ability to build up compound headings. (P. K. Leech) A new function for sorting variate values has been added, and GRAPH has been amended to give smoother 'lines'. (Alvey) The code for the CLASSIFY directive has been shortened and its efficiency improved. (Banfield) A new version of the garbage-collector, which recovers unused space internally was completed. (Wedderburn and Lane) Substantial improvements were made to the section dealing with the analysis of designed experiments; a new method of preliminary analysis of the design improves the numerical accuracy and greatly reduces the amount of computation required for standard orthogonal designs. The EXTRACT directive now allows the user to extract from univariate analyses of variance all the information required for a multivariate analysis of variance and a suitable macro is being written. (Payne and Wilkinson) A new directive RELATE was added to the clustering directives. Results from principal coordinate analysis are related to each data variate, which may be qualitative or quantitative. A pseudo F-statistic indicates whether there is a noticeable association between them. (Ross)

Genstat has links in both directions with RGSP (the Rothamsted General Survey Program). Considerable alterations were made to the subroutines of the link to simplify the machine-dependent sections. (P. K. Leech)

Numerous other small improvements were incorporated and various errors corrected; however very few serious errors have been uncovered by the many users of the system.

Documentation. The ninth User Guide, the Genstat Macro Library, was issued during the year, and contains information on using the library together with the first six macros (Paper 5). Five of these were listed last year, and the sixth calculates orthogonal polynomials. The macro library is issued with the program and is thus easily available to all users. (Todd)

There have been several requests for a primer for beginners, at a simpler level than that of the User Guides. A first draft has been produced and this is being widely distributed to encourage response from a range of potential users of Genstat. (Alvey)

Distribution. Twenty licences have been issued, 13 to organisations in the UK, four in the rest of Europe, two in North America and one in Australia. The licensees are mostly universities and research institutes, but now include two commercial firms and one computing bureau. The procedures developed for the licensing of Genstat are being extended to cover our other public programs. (Alvey)

Generalised Linear Interactive Modelling (GLIM). The second release of this program (*Rothamsted Report for 1974*, Part 1, p. 294) by the Numerical Algorithms Group (Oxford), took place in September. It contains substantial enhancements, including the definition and use of macros, and simple plotting of scatter diagrams. The package is now available in 44 centres in ten countries, and user reaction has been encouraging. The new release is available on eight machine ranges, including now a version for the PDP 11, and individual conversions have been made for Burroughs machines and the Univac 1108. The power of the new release was well shown by an exercise in which the analysis of a complex dilution assay, the subject of a recently published paper, was programmed and completed from the terminal in a single session of 40 min.

We are indebted to Mrs. Margaret Wood of Imperial College for devising a link program, which allows tables produced by the Rothamsted General Survey Program to be presented to GLIM for further analysis. Such links between systems provide users with greatly increased computing power. Work is now in progress on further enhancements, including branching and looping facilities, based on the use of macros. (Baker, Nelder

https://doi.org/10.23637/ERADOC-1-132

and Wedderburn, in collaboration with members of the Royal Statistical Society Working Party on Statistical Computing)

Maximum Likelihood Program (MLP). The interactive version was provided with a second channel, controlled by directives ECHO, INPUT and OUTPUT. The program was reorganised to allow as many as possible of the facilities of the full version to be available interactively, including small-scale plots and contour diagrams. The full version of the program was given a larger data store, and the maximum number of parameters in a model increased to eight. The syntax for models and derived variates was extended to allow labels and symbolic names, declared previously by a NAMES directive. Two arithmetic and three new logical functions were provided. The module FIT ASSAY was added, to provide a direct method of estimating herbicide concentrations or similar dilution assay problems involving continuous variables. New facilities have been added to Wadley's problem, mainly for use with the data on resistant strains of mildew collected by the Plant Pathology Department. These allow for natural immunity to be fitted, for graphical output, the analysis of heterogeneity for replicated doses and the use of heterogeneity factors in the calculation of S.E.s and fiducial limits of LD50s etc. A new curve, the generalised hyperbola, was added. A new option allows the approximate support limits to be computed for each function of the parameters, assuming the parameters are stable. (Ross, Payne, Kempton and White)

Genkey. Work has been concentrated on making this program more easily transportable between machines. Machine dependencies have been isolated in a few subroutines, and the output reorganised for more flexibility. Some of the selection criteria were modified following work on the theory by Gower and Payne. The University of Manchester Computing Centre is preparing a version of the program for the CDC 7600. (Payne)

Other programs and algorithms. Minor modifications were made to the cluster analysis program (CLASP) to facilitate input, and to save paper in the output. (Ross and White) Two programs were written both in Fortran and GENSTAT to perform growth-invariant canonical variate analysis. The first employed a principal components technique to estimate the growth effects and the second required the input of a matrix of growth effects after they had been estimated using a factor analysis program. Both versions of the two programs have been used extensively on problems in paleontology. (Banfield, with Professor R. A. Reyment of Uppsala) An algorithm in Fortran to form ultrametric distances for a single linkage dendrogram from a given matrix of distances has been accepted for publication. This algorithm had been used extensively in previous work by Gower and Banfield when studying the empirical distributions of goodness-of-fit criteria for hierarchical classifications. (Banfield) A program for plotting data requires an algorithm to determine for each axis a scale which uses as much of the space as possible while giving round numbers for the scale marks. A new algorithm, appreciably shorter than existing published ones, has been devised (Paper 16). (Nelder)

General-purpose subroutines were written to fit exponential and logistic curves by the method of stable ordinates. These subroutines incorporate their own optimisation sequences. (White)

When large programs, such as Genstat, are distributed, information must be transmitted on how they are to be segmented to fit into available space on the machine. This task, previously done by hand, has now been automated by a program, CHART, which derives the form of the overlay tree from computer output and displays it in a form suitable for publication. (Simpson)

https://doi.org/10.23637/ERADOC-1-132

Theory

Models for fertiliser response. Research and advisory workers have been turning their attention to the gap between average crop yields and their potential yield when as many limiting factors as possible have been eliminated. One suggested line of attack envisages very large complex experiments to evaluate the separate contribution of each husbandry factor. As the size and complexity of the experiments would depend very much on what interactions can be expected between the factors, we have been looking at the results of factorial experiments already done to see what could be learnt about the form of interactions, how important they are and how often they occur. Because many recent multi-level fertiliser experiments have been completed, we have much information about the factors in crop nutrition; at ADAS request, other work is being done on set sizes and populations of maincrop potatoes in relation to variety and other husbandry factors.

In fertiliser experiments most interactions can be interpreted in terms of one or two quite simple models. In the first type when a soil deficiency of, say, nitrogen, is made good by fertiliser dressings, response usually increases more or less linearly with increasing dressings until soil supplies of a second factor, say phosphorus, are exhausted and response to them ceases. On adding P fertiliser the same initial response to N appears but continues further before flattening off. Thus with successive increments of P fertiliser, the N response curve shows a series of turning points until a 'plateau' is reached where no additional amounts of N or P affect yield. This type of model, known as the 'Law of the Minimum' or the 'Law of Limiting Factors', was put forward by Liebig more than a century ago. It predicts that the dressing of N required for full yield will increase with increasing P. In the second type, the initial rate of response to the first factor increases in proportion to the amount of fertiliser added to make good a deficiency of the second factor. With this type of model the optimum dressing of N may *decrease* as the dressing of P increases. Indeed, with barley and potatoes adding more K fertiliser can so increase the efficiency of N use as to double yield while halving N requirement. Maximising crop production might seem at first sight to call for larger inputs of each factor but this need not be so if the optimal combination of factors leads to each one being used more efficiently. Of course, crop responses do not always take the form of intersecting straight lines; curvilinear responses to nitrogen could result from a limiting factor such as leaf disease that increased more than proportionately with increasing fertiliser N. In some experiments disease control has both increased the efficiency of N use and restored the 'two-straight-lines' form of response.

The importance of interactions between fertiliser nutrients depends on the ability of the soil to supply the crop's needs. Most barley crops can obtain the whole of their P and K requirements from the soil and so can responsive crops like potatoes on sites rich in nutrients; in such conditions yields are solely on the 'plateau' portion of the response curve and interactions are usually too small to be measurable. By contrast when the soil is deficient in two or more nutrients, fertilisers may give very large increases in yield and the optimal amounts are much affected by the form of the response. (Boyd)

Multivariate analysis. When n characters are measured on a set of p individuals, the resulting measurements can be represented as a set of p points in n-dimensional space. An important class of procedures in multivariate analysis are those which seek a configuration of points in a few dimensions (perhaps 2 or 3) which is, in some sense, closest to the original configuration in n dimensions. If the reduced configuration is close to the original, it simplifies visual inspection for detecting pattern, and shows important aspects of the data. Two kinds of procedure for reducing dimensionality are metric and non-metric scaling; metric scaling, also known as principal coordinate analysis, defines a

https://doi.org/10.23637/ERADOC-1-132

distance between every pair of points in the original configuration, and between corresponding points in the original and reduced configuration, and these distances can be used to define closeness. Non-metric scaling uses the ordering of points (larger or smaller) without the notion of distance and gives rise to a criterion, known as stress, for measuring the closeness of the original and reduced configurations. Non-metric scaling makes fewer assumptions than does metric; however, it gives rise to much more computation in deriving the reduced configuration. Metric scaling is known to lead to a unique result and uses a standard algorithm, whereas non-metric scaling may not produce a unique result and uses a more complex iterative algorithm. It is therefore important to establish whether the advantage of non-metric scaling outweighs the disadvantages. To this end an empirical comparison was made of configurations of individuals produced by metric and non-metric multidimensional scaling. The techniques of principal coordinate analysis and J. B. Kruskal's multidimensional scaling were used on several sets of distances and similarities and the results compared by Procrustes rotation in Genstat. Cases were found where the metric results were more consistent with a non-metric result than the non-metric were with themselves. A paper is being prepared to show these results. (Banfield)

Canonical variate analysis is a technique for producing those linear combinations of measurements that best discriminate between groups (e.g. species in a mixture of individuals). Problems arise when the individuals are of varying sizes and part of this variation reflects differing ages; this variation cannot help in the discrimination sought, and indeed will often hinder it. The removal of this unwanted variation gives rise to growth-invariant canonical variate analysis. Applications of this have been studied in the identification and classification of fossils, though the method applies of course equally to the study of living organisms. The multidimensional space that accommodates the variation between organisms can be partitioned so that variation due to the organisms' age or growth is separate. Canonical variates are then obtained to discriminate between populations of fossilised organisms in the space orthogonal to that of growth. Two techniques were tried for the estimation of growth effects, principal components and factor analysis. Discriminant functions between different samples of both *Globoconusa daubjergensis* and *Subbotina pseudobulloides* fossils were then computed after the growth effects had been removed. It was possible to quantify changes in the fossils of the organisms due to environmental factors alone. A paper will be published. (Banfield, with Professor R. A. Reyment, Uppsala)

The relative abundance of species. Much of the raw material for our understanding of the fluctuating numbers of different species of insects consists of the catches made in traps, such as those of the Rothamsted Insect Survey. Work on characterising species frequency distributions has concentrated on specifying more precisely the nature of the sampling variation between catches. Most previous theory has assumed that the population abundances are fixed, but, if inferences are to be made about the habitat, their variation from year to year must be described. A theoretical model set up to include this variation, suggests that some of the statistics used for measuring diversity are less suitable than hitherto thought. These theoretical results are strongly supported by empirical evidence from the Rothamsted Insect Survey (Paper 18). The distribution of species counts in a sample of trapped insects is usually derived on the assumption of a fixed length of time for trapping. An alternative sequential method, assuming a fixed number of individuals in the sample, has led to the derivation of some interesting new results and a paper has been submitted for publication. (Kempton and Wedderburn)

Non-linear inference. Classical linear models with Normal errors have a geometrical

https://doi.org/10.23637/ERADOC-1-132

representation in which the n data values become a point in n-dimensional space and the set of points for which the model gives an exact fit, the solution locus, forms a flat surface (hyperplane). The discrepancy between the data and the model is measured by the shortest distance of the data point from the solution locus. For non-linear models the solution locus is curved, and for non-Normal errors the shortest distance does not measure the discrepancy exactly. The theory of stable parameters was developed to minimise the effects of these added complexities, and it has now been unified for models whose log likelihood is approximated by a weighted sum of squares. The log likelihood is related to the distance in sample space from the data to points on the solution locus. Stable parameters give coordinate systems on the solution locus that are approximately orthogonal and regularly spaced. Stability is unaffected by moderate variation in the weighting, which affects the scale of each data coordinate. A similarity principle, formulated for continuous models, expresses the idea that a system that is stable for one model is also stable for other models of the same dimensionality, provided the solution loci are approximately parallel in the neighbourhood of the data. The similarity principle justifies the use of stable ordinates in curve fitting and the use of similar moment functions in distribution fitting.

Since stable parameters are not necessarily the quantities of practical interest, it is necessary to develop methods for recovering information about other parameter systems and functions of parameters. While there is no difficulty in obtaining maximum likelihood estimates directly, and satisfactory estimates of variances and covariances are obtained from first derivatives of the function, it is desirable to estimate support limits when the functions are non-linear. If the parameters are strictly stable, in the sense that the log likelihood is exactly quadratic, then support limits, which are maxima and minima of the function on a critical contour, are obtained by optimisation with respect to angular coordinates in one fewer dimension than that of the model. If the parameters are not very stable, constrained optimisation methods must be used.

Stable parameters can be defined to form an hierarchic sequence of models, analogous to orthogonal polynomials. Such systems have been investigated for various curves. (Ross)

Commonwealth and Overseas

The Ministry of Overseas Development (ODM) continues to support a unit within the Department to give a service to research workers in agriculture overseas, chiefly within the Commonwealth. Activities involved visits to some of the countries concerned, the processing and interpretation of experimental and survey data using the Rothamsted computer, collaboration in writing up results, and the training of visiting workers.

During the year, at the invitation of the Swaziland Government, Wimble visited Swaziland to advise the Research Division of the Faculty of Agriculture of the University of Botswana, Lesotho and Swaziland on the analysis of data from field experiments and on statistical aspects of the design of future programmes of experimentation. A report of his visit was submitted to the Swaziland Government. Walker spent two weeks at CIMMYT (International Centre for Maize and Wheat Improvement, Mexico) in March; subsequently he visited Nicaragua and had discussions with the ODM forestry team at Matagalpa and others. In September Walker began a two-year secondment to CIMMYT where he is in charge of a statistical service unit providing biometrical and computational services to staff of CIMMYT in their world-wide programme of plant-breeding and related experimentation.

Data were received during the year from Botswana, Fiji, Malawi, Mexico, Nicaragua, Nigeria, Philippines, Swaziland, Tanzania, and Zambia. Major tropical crops, like cotton,

https://doi.org/10.23637/ERADOC-1-132

sorghum, groundnut and maize, predominated; other data related to sunflower, millet, legumes, wheat, tobacco, soya bean, and pineapple, and to grassland and natural forests. Some data from cattle studies were received from Kenya.

Genstat has been extensively used for the analysis both of long-term trials and of sets of trials repeated over sites and years. Examples are long-term fertiliser trials on maize and sunflower from Zambia and the Swaziland district fertiliser trials. (Ryder)

Walker completed three collaborative jobs, with Dr. J. B. Hall (Nigeria) on the clustering of natural forest quadrats, with Dr. R. G. Lowe, also of Nigeria, on the prediction of growth of natural forest, and with Mr. P. H. Giles (Nicaragua) on the storage of maize (with Lane). He began work on the classification of sites and varieties from experiments conducted in several countries, using data from CIMMYT, and this work is being extended to larger data sets. (Gilliver)

Consultation at the planning stage is important, and contact has been established with agricultural officers in Nepal who are starting to collect information on the growing of crops in two regions of the country. This may involve designing district trials to be carried out on sloping land with small terraces, starting with the next planting season.

From the collaborative work two jobs of more than routine statistical interest have been selected to illustrate the work of the unit.

The relationships between species composition of herbage and certain chemical constituents of the soil in which the herbage grew were analysed using data from samples drawn from six soil parent materials. Principal component analysis of the herbage composition scores identified two main contrasts among the species. These were related to the parent materials by analysis of variance and to the soil chemical constituents by regression. The main species contrast was found to be associated with soil copper. (Wimble, with Professor Beavington, Nigeria)

'Rainfall confidence limits' were estimated for one station in Swaziland for periods of three weeks throughout the year. These limits are amounts of rain within which future rainfall for the period may be expected to lie with 50% probability. The usual method for doing this is to apply a transformation of the log $(x + c)$ family to recorded rainfall data in the past, then to estimate upper and lower quartiles on the transformed scale, assuming normality, and to transform these back. The method fails in areas with substantial dry seasons, because the rainfall distribution is then J-shaped, and no transformation to normality is possible. Estimation of the limits from the data without transformation gives better results. (Wimble, with Mr. J. S. Watson, Swaziland)

It is important that the large-scale programs that run on our large computer should be supplemented by smaller but flexible programs able to run on local machines. In response to a request by the biometrics unit of the Ministry of Agriculture Research Department of Malawi, work is nearly complete on a program for the analysis of a partially confounded design. Overseas workers often consult the Department on methods of recording data from field trials. A recording sheet is being designed which could be used either directly or with minor modifications in many situations and which would ease the transference of results from paper to either cards or paper tape. A trial description sheet has also been proposed which would ensure that the description of the trial is complete and contains all the necessary information to allow us to process the data. With information on trials being received from a large number of countries on numerous annual crops, the forms need to be flexible enough to cope with the many types of recordings that may arise. (Gilliver)

Two overseas workers spent substantial periods in the Department, one from the Bureau of Soils in the Philippines, the other from Malawi. In addition, 17 overseas research workers visited the Department for discussion of their work.

https://doi.org/10.23637/ERADOC-1-132 *pp 12*

Staff and visiting workers

We were greatly saddened by the sudden unexpected death of R. W. M. Wedderburn in June. The Department has lost a highly gifted research worker.

Mary T. Simpson and G. N. Wilkinson left. B. Gilliver was appointed to fill the vacancy arising from Walker's secondment to CIMMYT. Other appointments were R. J. Baker, Janet Riley and K. Ryder.

Verona Roberts left in January after 32 years of devoted work for the Survey of Fertiliser Practice.

Gower spent the year as Research Leader in the Division of Mathematics and Statistics, at Adelaide, South Australia.

Nelder spent a month in Australia, visiting the CSIRO Divisions of Mathematics and Statistics, and of Computer Research in Perth, Brisbane, Sydney, Adelaide and Canberra. While there he attended the First Conference of the Division of Mathematics and Statistics and he and Gower gave papers (Papers 4, 10 and 11). He also visited Scandinavia, giving seminars at the Mathematical Institute, Aarhus University, Denmark and the Department of Mathematical Statistics, Lund University, Sweden. Church spent two months in Australia assessing sample survey work in CSIRO and the organisation of survey data processing, with particular reference to the setting up in Canberra of the Rothamsted General Survey Program. Whilst on a private visit to Australia, Leech gave six lectures on disease eradication as a statistical problem. Banfield, Ross and Wilkinson all attended the 40th Session of the International Statistical Institute in Warsaw where Ross gave a paper and Banfield attended in his capacity as U.K. Regional Editor for the ISI Statistical Theory and Method Abstracts. Immediately before this meeting Ross attended a symposium on Statistical Methods in Agriculture at the College of Agriculture, Poznan, Poland. Earlier in the year he gave a talk on statistical programs developed at Rothamsted at a meeting of the Société Adolphe Quetelet, Gembloux, Belgium. Banfield spent a month in the University of Uppsala at the invitation of Professor Reyment of the Paleontological Institute.

Six Genstat courses were given during the year by Alvey and Simpson. Two were at Rothamsted, the others at Manchester, University College London, Edinburgh Regional Computing Centre and the University of Bristol Computer Centre.

Five temporary workers spent varying periods in the Department, four of them from overseas.

Publications

General Papers

1 Church, B. M. (1975) Fertiliser use on farm crops in England and Wales, 1974. London: *Ministry of Agriculture, Fisheries and Food* (*SS/SAF/14*), 14 pp.

2 Church, B. M. (1975) Fertiliser use in farm type regions of England and Wales, 1974. London: *Ministry of Agriculture, Fisheries and Food* (*SS/SAF/17*), 30 pp.

3 Hills, M. G. (1975) Fertiliser use on grassland, 1974. London: *Ministry of Agriculture, Fisheries and Food* (*SS/SAF/16*), 22 pp.

4 Nelder, J. A. (1975) An introduction to Genstat. *The Mathematical Scientist*. Supplement to Vol. 1, No. 1.

5 Todd, A. D. (1975) *The GENSTAT Macro Library, GENSTAT User's Guide No. 9*. Inter-Universities/Research Councils Series, Report No. 25. Program Library Unit, Edinburgh Regional Computing Centre, 25 pp.

https://doi.org/10.23637/ERADOC-1-132

PAPER IN ROTHAMSTED REPORT, PART 2

6 CHURCH, B. M. (1976) Use of fertilisers in England and Wales, 1975. *Rothamsted Experimental Station. Report for 1975*, Part 2, 91–95.

RESEARCH PAPERS

7 (CARLISLE, R., ANEKE, C.) & WALKER, P. (1975) Ectopic beats in Nigerian patients on and off Digoxin therapy. *East African Medical Journal* **52,** 19–25.

8 GOWER, J. C. (1975) Generalised Procrustes analysis. *Psychometrika* **40,** 33–51.

9 GOWER, J. C. (1975) The determinant of an orthogonal matrix. *Applied Statistics* **24,** 150–153.

10 GOWER, J. C. (1976) Procrustes rotational fitting problems. *The Mathematical Scientist*. Supplement to Vol. 1, No. 1.

11 GOWER, J. C. (1976) Multivariate analysis in Genstat. *The Mathematical Scientist*. Supplement to Vol. 1, No. 1, 59–62.

12 GOWER, J. C. & PAYNE, R. W. (1975) A comparison of different criteria for selecting binary tests in diagnostic keys. *Biometrika* **62,** 665–675.

13 (INNES, N. L.) & WIMBLE, R. H. (1976) Breeding for reduced seedcoat fuzz in Upland cotton. (*Gossypium hirsutum*). *Annals of Applied Biology* **83,** 1–9.

14 (INNES, N. L.), WIMBLE, R. H. & (GRIDLEY, H. E.) (1975) Estimates of genetic parameters for lint quality in upland cotton. *Theoretical and Applied Genetics* **46,** 249–256.

15 JENKINSON, D. S., POWLSON, D. S. & WEDDERBURN, R. W. M. (1976) The effects of biocidal treatments on metabolism in soil. III. The relationship between soil bivolume, measured by optical microscopy, and the flush of decomposition caused by fumigation. *Soil Biology and Biochemistry* **8,** 189–202.

16 NELDER, J. A. (1976) A simple algorithm for scaling graphs. *Applied Statistics* **25,** No. 1, 94–96.

17 ROSS, G. J. S. & GASSER, J. K. R. (1975) The distribution in the soil of aqueous ammonia injected under grass. *Journal of the Science of Food and Agriculture* **26,** 719–729.

18 TAYLOR, L. R., KEMPTON, R. A. & WOIWOD, I. P. (1976) Diversity statistics and the log-series model. *Journal of Animal Ecology* **45,** 255–272.

19 WALKER, P. (1973) Problems of experimental design in tropical regions. *Proceedings of the 10th Colloquium of the International Potash Institute*, pp. 359–373.

20 WATSON, M. A., HEATHCOTE, G. D., LAUCKNER, F. B. & SOWRAY, P. A. (1975) The use of weather data and counts of aphids in the field to predict the incidence of yellowing viruses of sugar-beet crops in England in relation to the use of insecticides. *Annals of Applied Biology* **81,** 181–198.

21 WEDDERBURN, R. W. M. (1976) On the existence and uniqueness of the maximum likelihood estimates for certain generalised linear models. *Biometrika* **63,** 27–32.

https://doi.org/10.23637/ERADOC-1-132