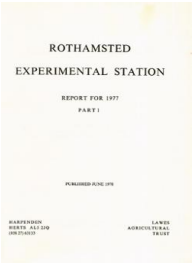
ROTHAMSTED RESEARCH

Report for 1977 - Part 1

Full Table of Content



Statistics Department

J. A. Nelder

J. A. Nelder (1978) *Statistics Department ;* Report For 1977 - Part 1, pp 297 - 313 **- DOI: https://doi.org/10.23637/ERADOC-1-134**

STATISTICS DEPARTMENT

J. A. NELDER

Staff

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

Senior Principal Scientific Officers
J. C. Gower, M.A.
R. H. Wimble, M.SC.

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.

Senior Scientific Officers
N. G. Alvey, M.I.S.
C. F. Banfield, B.SC.
Miss Mary G. Hills, B.A.
R. W. Payne, B.A.
H. H. Spechter, PH.D.
A. J. Vernon, B.SC.

Higher Scientific Officers
Mrs. Jill F. B. Altman, M.I.S.
C. J. Dyer, B.SC.
P. W. Lane, B.A.
J. G. Pearlman, M.SC.
K. Ryder, M.I.S., B-Tech.
P. E. Sparrow, M.SC.
A. D. Todd, M.SC.
P. J. Zemroch, M.SC.

Scientific Officers
R. J. Baker, M.SC.
Mrs. Linda C. Bassill, M.I.S.
Miss Penelope K. Leech
J. N. Perry, M.SC.
Miss Janet Riley, M.SC.
Miss Patricia J. Wherry, B.SC.
R. P. White, B.SC.
J. Wood, M.SC.

Assistant Scientific Officers
Mrs. Janet I. Elsmere
S. A. Watson

Temporary Worker
I. A. Rubanov

Secretary
Mrs. Barbara I. Merrett

Clerical Staff
Mrs. Anne Arnold
Miss Michelle Bateman
Mrs. Jillian M. Bridgman
Mrs. Doris A. Court
Mrs. Elsie Davies
Mrs. Janet A. Day
Mrs. Susan Greenland
Miss Susan T. I. Hyde
Mrs. Genevieve M. A. Jeffrey
Miss Linda M. King
Mrs. Geraldine F. Kirby
Miss Moira W. Livings
Mrs. Elizabeth H. Mayton
Mrs. Janice R. Messer
Mrs. Patricia K. Neill
Mrs. Brenda J. Watler

Introduction

Our main activity continues to be the application of statistical methods to many kinds of biological data. These may arise from field experiments, from surveys of farms, their fields and animals, from experiments in laboratories and controlled-environment chambers, or from collections of organisms made both at home and abroad. A common feature of the analytic procedures is the exposure of pattern; in contrast to the commonly held view of statistics as being concerned with the establishment of significant differences, the discovery of pattern is much more concerned with the establishment of significant sameness, that is the discovery of invariant features.

We have continued to develop our computer programs as aids to statistical analysis; their distribution has been considerably extended and their documentation improved. The development of relevant statistical theory, though not a major activity, is nonetheless an important part of our work. The problems tackled arise directly from the applications, and solutions are required to improve the statistical techniques involved in those applications.

https://doi.org/10.23637/ERADOC-1-134

Practical applications

A major part of our programme involves the application of statistical ideas and techniques to research problems as they arise at Rothamsted, at other Agricultural Research Council Institutes (ARC), or at the many Agricultural Development and Advisory Service (ADAS) centres with whose experiments we are involved. The topics that follow will serve to illustrate the variety of applications, but do not attempt to catalogue all our activities in these areas.

Interference between pheromone traps. The planning of experiments involving pheromones as insect attractants must take account of the problems of interference between traps. Experiments have been done on the spacing of pheromone traps for pea moth using a change-over design, two Latin squares with an extra row, which enabled residual effects of the previous day's treatment to be estimated in addition to site, day and treatment effects. The treatments used were an isolated trap and two lines of three traps, one aligned along and the other across the wind. In addition to comparing the isolated trap with the central traps of the triplets, the design allowed comparisons to be made of the three traps within each triplet. A first-stage mathematical model can explain differences in trap catches caused by interference, and work is in progress to extend this to allow for variation due to diffusion and wind fluctuations. (Perry, with Wall, Entomology Department)

In addition, work is in progress with 1976 data on catches of pea moths from pheromone traps set up for monitoring and the correlations between sites measured over time are being evaluated. Techniques employed include graphs and regressions of correlations against distance between sites, contour maps and principal coordinate analyses. Sites are also compared after grouping by geographical region. The purpose of the analysis is to gauge the minimum distance necessary between sites for monitoring. (Perry, with Macaulay, Entomology Department)

Competition between wheat bulb fly larvae. Competition between wheat bulb fly larvae when invading shoots has been investigated using pot experiments. A model was built allowing for density-dependent mortality and pot effects, based essentially on the classical occupancy problem. The mean and variance of the number of uninvaded shoots were calculated theoretically. It was possible to formulate the model as an instance of a generalised linear model, and hence to fit it to data by use of the extended facilities in the new release of the program GLIM. (Perry and Baker, with M. G. Jones, Entomology Department)

South American nematode populations. The data comprised measurements on 1180 Peruvian nematode larvae sampled from 89 populations. The analyses included an unbalanced hierarchical analysis of the variability of several characters of 300 larvae sampled from ten cysts which had been collected from two sub-populations of the nematodes. It was established that for most of the characters the variability between the sub-populations was greater than amongst the cysts which in turn was greater than that amongst the larvae. A canonical variate analysis was also used on all 89 populations to see if the characters observed could separate the populations in accordance with the geographical regions from which the populations were sampled.

Canonical variate analyses were also made of subsets of the data. Twenty-seven populations from each of six valleys in Peru were combined and it was shown that the characters measured separated the populations according to their different environments. Another analysis separated those populations from Ecuador, Northern and Central Peru from those of Southern Peru and Northern Bolivia. (Banfield, with Franco and Evans, Nematology Department)

https://doi.org/10.23637/ERADOC-1-134

Carbon flow through leaves. A compartmental model is being constructed, using data from a series of tracer experiments on sunflower leaves. In these, incorporation of ^{14}C into some of the intermediates of photosynthesis and photorespiration was measured over a period of time for differing water-stress conditions. These conditions were held constant for each experiment, and consequently it was possible to model the ^{14}C incorporated as a linear dynamical system for any postulated biochemical pathway. However, none of the pathways put forward so far can completely explain the section of the data involving photorespiration. A possible explanation has been found in mathematical terms, but as yet there is no biochemical evidence to support it. (Pearlman, with Lawlor, Botany Department)

Multivariate methods. Considerable use has been made of our programs for multivariate analysis, and the special requirements of a diversity of users are helpful in indicating directions that future developments of these programs should take.

Principal component analysis was applied to peak heights on gas chromatograms of the organic acid fraction derived from 113 strains of bacteria. The strains included reference strains, unknown strains and some further replicates of the reference strains. The compounds detected by the chromatograph were acetic, propionic, formic, pyruvic, lactic, oxaloacetic and succinic acids. Many of the unknown strains were identified by comparison of their principal-component scores with the reference strains and their replicates. (Banfield, with Mr. J. E. Hine of the Central Public Health Laboratory, Colindale)

From data collected on locusts at various times since fledging, a method was required for deciding whether a locust was sexually mature or immature from observations of fractions separated by gas chromatography. A canonical variate analysis showed that a subset of just nine of the 15 observed fractions discriminated well between the two groups. These nine fractions also had the best correlations with the colour of the locusts, which is a simple indicator of maturity. (Banfield, with Mr. R. Greenwood of Portsmouth Polytechnic)

Canonical variate analysis was also applied to the identification of three subspecies of the African Lechwe antelope, and it was possible to show from measurements of horn size and shape that the subspecies *smithemani leche* and *kafuensis* are distinct. An additional skull thought to be the only authentic specimen of the believed extinct subspecies *robertsi* was shown to be different from skulls of the other three subspecies and most like *smithemani*. Additionally it was shown that *smithemani* and *leche* resembled each other more closely than they did *kafuensis*. (Banfield, with Mr. W. F. H. Ansell of Zambian Game Department)

A recently developed technique for the analysis of asymmetric matrices found use in a geological context, and we were able to apply it, using a macro written in Genstat, to analyse a matrix of the occurrence of unlike nearest-neighbouring depositional environments in the Mississippi delta of Louisiana. This allowed the skew-symmetric part of the asymmetry to be displayed graphically for further geological interpretation (Banfield, with Prof. R. A. Reyment of Uppsala University, Sweden)

Optimum seed rate for potatoes. An investigation into optimum seed rate for various seed sizes and varieties of potatoes has been completed. Information from many ADAS regions and EHFs was used and three reports were prepared, one for each of the varieties Pentland Crown, King Edward and Record. An evaluation of the results of experiments on Maris Piper and Désirée was also made, but separate reports were not written as the information on these two varieties was less extensive. An economic analysis of all these results was undertaken resulting in the production of tables of recommendations for each

https://doi.org/10.23637/ERADOC-1-134

variety for seed sizes from 10 to 28 sets kg^{-1} and for cost/value ratios of the price of seed to the value of ware from one to three. These tables will form the basis of an *ADAS Short Term Leaflet*, and provide the first recommendations to be based on extensive sets of results; previously the various experimental centres had tended to evaluate their own results separately. (Wood, with Mr. F. Shotton and Mr. J. Gunn of ADAS)

National grassland manuring trial, GM20. Analysis has continued of this extensive series of experiments described in the *Rothamsted Report for 1976*, Part 1, 328. Results show that there was little variation in annual dry-matter yield from year to year at each site; the between-site variation was large and analyses, based on incomplete data, indicate that little of this variation will be explained either by total winter or total summer rainfall.

Dry-matter yields within each cut can be related either to the amount of nitrogen available, part of which are residues of previous nitrogen applications, or to the amount of nitrogen applied. Either relationship gives similar predicted maximum yields though the evidence indicates that available N is on the whole the better explanatory variable. Yields are highest in the early part of the growing season. Work has begun in relating these dry-matter yields to environmental factors. The effect of rainfall through the growing season has been expressed by using a model proposed by Mr. J. Sinclair of the Grassland Research Institute.

The *Technical Report*, which will present all relevant site information, is near completion, and a report on the progress of the analysis was presented to the joint ADAS/ARC Committee on Grassland Manuring in March. (Sparrow, with Mr. J. Morrison, Grassland Research Institute)

Surveys. The survey group has dealt with a variety of tasks, with the Surveys of Fertiliser Practice and Milk Quality continuing to be a major commitment.

Fertiliser practice. The 1977 survey followed closely the pattern established in 1976. Provisional estimates of fertiliser use, fertiliser stocks and other basic tabulations, based on 765 farms visited by Farm Research Ltd., were available by mid-September; final figures (including a further 550 farms visited by members of ADAS) were circulated at the beginning of December. Compared with 1976, increases of the order of 15% were recorded in N use on winter wheat and on both temporary and permanent grass. On grassland, this continues the long-term trend checked by the drought in 1976, but for winter wheat, the increase came after 5 years in which N applied to cereals remained fairly constant. There was also an increased use on sugar beet of complete fertilisers containing magnesium and sodium; further work on this is required.

The fertiliser information recorded in 1976 has been analysed by farming type regions and, together with similar work for previous years, compared with recent recommendations and reported to the ADAS Soil Analysis and Fertiliser Committee (Paper 2).

One of the main objectives of the Survey has always been to monitor actual fertiliser use against current advice. Such monitoring is now extremely difficult when the 'potential' yield (if established) may be much in advance of rapidly moving options and 'break-even' positions. The introduction of high yielding wheats such as Huntsman may well be the reason behind the change in use of N on wheat. It may consequently be necessary to collect information on variety in future. (Church and Hills)

The development of these surveys since 1942 was reviewed, information on recent trends in the use of fertilisers summarised and current practice compared with published recommendations (Paper 12).

https://doi.org/10.23637/ERADOC-1-134

Soil sample data from surveys since 1969 provide useful information on the proportions of soils in different nutrient categories. These data are now being examined to assess the feasibility of monitoring general trends in nutrient status by resampling the same fields. A draft report has been prepared for discussion. The practical problems are those of maintaining consistent field sampling and analytical procedures over several years. (Church)

Milk quality. A comprehensive report on the National Milk Quality Survey was circulated to the main interested parties. When critical comment has been received, a shorter version will be drafted for publication.

This survey embraced three linked studies on a sample of 446 farms. Laboratory tests on monthly milk samples from farm bulk vats provided data on the keeping quality of milk awaiting collection at the farms. Tests on rinses of milking equipment, taken once in winter and once in summer, provided estimates of bacterial contamination in the milking plant and bulk vat that would contribute to the total bacterial contamination of milk samples taken from the vat. The third study was a record of features of dairy hygiene, taken on the same days as the two sets of equipment rinses.

About 80 analyses were needed to summarise the three sets of results and elucidate the complex inter-relationships between the three studies and between different items in the same study.

The report on the milk sample tests shows the proportions of rejections that might occur if particular criteria of milk quality were applied. By associating data from rinses with data from relevant milk samples, it appeared that on a very significant proportion of farms, the milking equipment was responsible for a large proportion of the bacteria found in samples of milk taken from bulk vats. When the rinse data were associated with the surveys of dairy hygiene, it appeared that on average newer techniques of cleaning equipment were more effective than older techniques, though results on different farms ranged from poor to excellent, whatever the cleaning technique. Records of methods of teat and udder hygiene did not show clear and consistent relationships with levels of bacterial contamination found in the bulk vat milk samples. (F. B. Leech and P. K. Leech, with Mr. J. J. Panes and Mrs. D. R. Parry of ADAS)

Other surveys. Final plans for the National Survey of Grass Weeds, 1977, for which the Weed Research Organisation have primary responsibility, were completed; the basic tabulations and a preliminary report were produced before the year end. (Church and P. K. Leech, with Mr. J. G. Elliott of Weed Research Organisation)

We have continued to collaborate with F. Yates and J. Beasley (Computer Department) on the development and use of The Rothamsted General Survey Program (RGSP), and are represented on the Secretariat, newly formed to administer it. We assisted in two external projects which used the program; one was the Seychelles census, in which the effectiveness of RGSP for the rapid analysis of a small census (60 000 records) was demonstrated, with validation and basic table production being completed within a month. (Church, with J. Beasley, Computer Department and Mr. T. Jones, ODM) The other project was a survey of Nigerian farms (see under 'Commonwealth and Overseas'). Following some practical experience here with the program, the originators for both projects have been able to specify further supplementary analyses from overseas.

Tabulations of milk yield data were produced for the survey of Dystokia in Friesian Heifers and also final tables needed for completion of a report on the Bloodstock Survey. (P. K. Leech)

Routine analysis. We continue to provide a service for the processing of data from

https://doi.org/10.23637/ERADOC-1-134

experiments and surveys, and this year at least 118 workers from other Departments and from Broom's Barn made use of it. Data comprising 1·45 million items were involved, a 5% drop on last year, caused mainly by an agreed reduction in the amount of raw data from ADAS poultry experiments. The average turn-round time was again reduced by about 1 day, from 9 to 8 working days. (Dunwoody, Dyer, Todd and Wherry)

As an aid to the inspection of data from field experiments we followed up a suggestion by G. V. Dyke (Field Experiments) that plot residuals should be indexed by specified row and column coordinates. This means that, where a field plan of an experiment is provided, the residuals can be printed on their relative field positions. Allowance can be made for gaps or non-contiguous sets of plots. (Dunwoody and Todd)

The 1976 edition of *The Yields of the Field Experiments*, which gives summaries of the results of the Rothamsted and Woburn trials, was published for the first time simultaneously with the Annual Report. The improved schedule resulted from a tightening of the timetable for intermediate scrutiny of the data and better use of direct computer output. The standard of presentation was improved, and further improvements are anticipated following the acquisition of a daisy-wheel printer with paper-tape attachments during the year. (Dunwoody and Todd)

Summaries of experiments. During the year we prepared summaries of several series of experiments, mostly from ADAS. These included:

(i) The SS57 series on NPK manuring of spring barley from sites with non-chalk soils. (Zemroch)
(ii) A series on nitrogen manuring of oilseed rape, involving 25 sites over 3 years.
(iii) A long-term trial of wild oats in continuous winter wheat from Drayton EHF. (Dyer)
(iv) A survey of nitrogen responses in crops. Responses of crops grown under widely varying conditions were compared, mainly by graphical methods. (Lane, with Miss B. Benzian)

Livestock experiments. Work on the design and analysis of ADAS animal experiments and investigations has continued. This has included assistance in planning a series of experiments to be done by Messrs. Eastwoods in conjunction with ADAS, for exploring the possible value of poultry waste as a source of protein in the feeding of beef cattle. Assistance has also been given in setting up a coordinated breed evaluation trial involving Holsteins and Friesians. (Lessells and Altman)

A report reviewing the problems of dairy cattle experimentation was circulated to relevant ADAS experimental farms as a basis for discussion. This was concerned chiefly with two problems; the first was how to achieve the best compromise between the need to reject extreme cows of various types and the need to obtain the smallest possible range of calving date, while the second concerned the need to block on calving date whilst producing an even balance of yield potential over treatments. (Lessells and Altman)

The first year's results from a 2-year investigation, in seven herds of Friesian cows, into the effect of progesterone treatment on the calving-to-conception interval, showed that the treatment shortened the interval by $14 \pm 2{\cdot}3$ days without affecting pregnancy rate. (Altman, with Mrs. Bridget Drew, ADAS)

We are responsible for the design and analysis of a large breeding experiment with pigs, which is being conducted at Great House, Trawscoed and Terrington EHFs, under the direction of Mr. F. K. Deeble (ADAS). It is now in its 3rd year and may continue for a further 6 years. The aim is to assess the relative merits of first-cross and 'criss-cross' Landrace and Large White pigs in terms of litter size, birth weight and weaning weight.

https://doi.org/10.23637/ERADOC-1-134

The first-cross females averaged larger first litters than the 'criss-cross', but this difference did not reappear in the second litters. Differences between farms were large compared with differences between breeding lines. (Spechter)

Eight climate rooms at Gleadthorpe EHF were used for a series of experiments on broiler chickens with varying ventilation rates, room temperatures and densities of chickens per m^2 of floor space. We proposed a suitable scheme for sampling the large populations of free-roaming birds in climate rooms without causing excessive disturbance. These samples are used to estimate population growth rate, and current results indicate that the method gives unbiased estimates.

The first experiment in these rooms showed that live weight per unit of food consumption increased with increasing restrictions on ventilation rate, the limiting factor apparently being the tolerance of farm workers—rather than chickens—to ammonia in the rooms. (Spechter, with ADAS poultry specialists)

Modifications to routine analyses of animal experiments have been introduced. These aim to summarise results more concisely than before, and to clarify the effects of treatments when successive observations on several variates are involved over a period of time. Methods appropriate to the analysis of growth curves in animal experiments are being explored. (F. B. Leech and Spechter)

Theory

Unfolding. Multidimensional unfolding is a method for analysing rectangular matrices, to give a geometrical representation of two sets of points, one set associated with column labels and the other with row labels. In this representation the distance between two points, one from each set, is chosen to approximate the corresponding entry in the original matrix. Thus the method resembles the problem of drawing a map of town configuration in two countries given the inter-country but not the intra-country town-distances.

What constitutes a good approximation can be defined in various ways, the two-most important of which give different forms of analysis termed metric and non-metric unfolding.

The investigation of properties of metric unfolding has continued. Deficiencies in the standard treatment have been revealed and the theory is now better understood, though no reliable computational algorithm is yet available. This is surprising in view of the intrinsic simplicity of the model.

A new use for unfolding, which has promising properties, is for the representation of the elements of symmetric matrices. In this representation the row and column labels are plotted separately so that each entity occurs twice. No attempt is made to force self-distance to be zero. Such diagrams have three advantages over traditional distance diagrams:

(1) They allow some high-dimensional configurations to be represented in few dimensions (e.g. a regular simplex only requires one dimension).
(2) They provide information on the inherent dimensionality of samples (e.g. an inherently two-dimensional sample is represented by pairs of coincident points in two dimensions, but higher dimensional symmetry requires more complex geometry).
(3) Sample structure is revealed, e.g. sub-samples may be detected by their requiring different axes of symmetry.

Preliminary applications of this technique (using non-metric unfolding algorithms) are most promising. (Gower)

https://doi.org/10.23637/ERADOC-1-134

Diagnostic keys. When an identification or diagnostic key is used, incorrect identifications may arise either because tests are incorrectly used or observed, or because the specimen examined is aberrant or incomplete. A method of checking for the occurrence of such errors has been developed which involves constructing a check key for each end point of the main key. Use of the appropriate check key confirms or corrects the initial identification.

The extensive literature on identification keys and diagnostic tables is being summarised and a review paper is in preparation. (Payne, with Dr. D. A. Preece, University of Kent)

Residuals in generalised linear models. In linear models with Normal errors, residuals defined by $y-\hat{\mu}$, where y is observed and $\hat{\mu}$ the fitted value predicted by a model, are useful in investigating the goodness-of-fit of that model. With generalised linear models involving non-Normal distributions for the ys, some generalisation of residuals is required which would allow techniques developed for Normal residuals to be applied to this wider class of models. Definitions of generalised residuals of the form $f(y)-f(\hat{\mu})$, suitably scaled to give constant variance, where the function $f(.)$ is chosen to make the distribution of $f(y)$ approximately Normal, are well-known. An alternative definition has been developed using the formula

$$\text{sign}\,(y-\hat{\mu})\times\sqrt{d}$$

where d is the contribution to the deviance (= twice the log likelihood ratio) of the observation in question. These residuals have the property that their sum of squares equals the deviance, which is itself a generalisation of the residual sum of squares for Normal errors. It has been found that with Poisson, binomial or gamma distributions, these deviance residuals are remarkably close to those derived from a normalising transformation for a wide range of values of y given μ. (Ross and Nelder)

Updating techniques in multivariate analysis. Techniques for updating regression equations following the addition or removal of a variate or unit from the data matrix are now well-known. They are particularly useful when working with an interactive program. A start has been made in developing similar techniques for use in the procedures of multivariate analysis, in which a basic requirement is the updating of the latent roots and vectors of a sums-of-squares-and-products matrix when a unit or variate is added or removed. The core of the solution of this problem is a method that gives the latent roots of a diagonal matrix plus a symmetric matrix of rank two. Such a method has been devised and the resulting algorithm is being incorporated into programs for several multivariate analyses. (Banfield and Gower)

Information on parameters. Models are fitted to data in order to estimate certain quantities, either parameters in the model or quantities derived therefrom. Each observation provides some information on these quantities, but some observations provide more than others. The allocation of the observations is a problem of experimental design, and we wish to know which observations provide the most information about each parameter. In a non-linear multiparameter model we need some provisional estimate of the parameters before we can suggest effective designs.

For a model of the form $E(y_i) = f(x_i, \boldsymbol{\theta})$, with approximately normal independent errors having variance σ^2/w_i, the dispersion matrix of the maximum-likelihood estimate $\hat{\boldsymbol{\theta}}$ is approximately

https://doi.org/10.23637/ERADOC-1-134

$$(V_{jk}) = \sigma^2 \left(\sum_i w_i \frac{\partial f_i}{\partial \theta_j} \frac{\partial f_i}{\partial \theta_k} \right)^{-1} = \left(\frac{w_i}{\sigma^2} \sum_i \pi_{ij} \ \pi_{ik} \right)$$

where $\pi_{ij} = \sum_k V_{jk} \frac{\partial f_i}{\partial \theta_k}$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$.

The vectors π_j show how each point provides information on each parameter, for models in which $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The vectors $\partial f / \partial \theta_j$ are also of interest, and show how changes in each parameter affect each fitted value.

For example, in fitting an exponential curve

$$\mathrm{E}(y) = \alpha + \beta \rho^x$$

it is found that π_α increases with x, so that, as expected, points near the asymptote provide most information about the asymptote, while π_ρ shows that ρ is estimated as a contrast between outer points and inner points close to where $\mathrm{E}(y)$ is intermediate in value. (Ross)

Graphical representation of parameter solutions. For each point x_i there are values of parameters $\boldsymbol{\theta}$ for which the fitted value exactly equals the observed value of y. These parameter values form a locus in parameter space, and in two-parameter models (or models in which only two parameters are studied while the remainder are fixed) the locus is a curve which may be plotted, directly if the equation $\mu = y$ has an analytic solution, otherwise indirectly by constructing contours of the residual $y - \mu$ and interpolating to find the zero contour.

If all the parameter loci are drawn on the same graph it is easy to see whether there is a well-defined solution indicated by convergence of loci, or whether there are outliers (a few loci inconsistent with the remainder) or multiple solutions (several knots of intersection) or no solution (divergence). The shape of the likelihood function may be inferred from these plots. (Ross)

The 1/6th replicate of a 6^3 design. This design, introduced by Boyd, is intended for investigating response surfaces using enough levels of each factor to allow the data to indicate the shape of the surface while restricting the number of plots to an acceptable figure (36) for general use in the field. As with any fractional-replicate design, problems of interpretation may arise from the aliasing of effects with each other, caused by the use of only a fraction of the full set of treatment combinations. With factors at six levels the aliasing structure is complicated by non-orthogonality, and the selection of treatment combinations to be used is not straightforward. Study of the alias structure of this design has enabled improvements in treatment selection to be made, and designs have been found in which main effects and simple components of interactions are nearly orthogonal. A paper is being prepared. (Wimble, Zemroch and Sparrow)

Statistical programming

The distribution of our programs has been considerably extended during the year. The techniques we have evolved to help the conversion to other machine ranges are reducing the amount of work required both to mount new versions on existing machine ranges and to convert programs for new ranges. (Simpson)

Genstat. The main event of the year was the release of version 4·01. No major enhancements were added for this release, but various errors discovered in release 3.09 were corrected and some other internal tidying-up took place. Facilities for mathematical

https://doi.org/10.23637/ERADOC-1-134

optimisation were included on a trial basis, and details will be released to selected users. The ICL 4–70 version went into service at Rothamsted in October, and that for the IBM 370 was distributed in November by Mr. N. M. McLaren of the Cambridge University Computing Laboratory. A version for the ICL 4–75 is being produced on the ERCC System 4 machine; progress is slower than hoped, because of difficulties in working at a distance through a terminal. This version is important because it uses the compiler (FORTE of the ERCC) which will also be available on the new ICL 2900 range. Preliminary work on version 3.07 showed that transfer from the 4–75 to the Edinburgh 2970 machine was workable, though defects in the latter's software prevented the full system from being put together.

Conversions for other machines are in progress at eight other sites, as follows:

TABLE 1

Genstat: conversions in progress for various machines at eight sites

Machine	Site	Converter
CDC 7600	Manchester Regional Computer Centre	Mr. J. Lloyd-Jones
CDC Cyber 76	CSIRO, Canberra	Mr. R. I. Baxter
CDC 6000	Waite Institute, S. Australia	Dr. P. Baghurst
ICL 1900	Oxford University Computing Laboratory	Dr. P. Griffiths
Univac 1108 (standard compiler)	Shell Research, Sittingbourne	Mr. G. Paterson
Univac 1108 (new IBM-compatible compiler)	University of Copenhagen	Dr. J. Wasniewski
DEC-10	University of York	Mr. A. J. Bullen
Burroughs B6700	University of Otago, New Zealand	Dr. B. G. Cox

We are encouraged by the enthusiasm of the converters, and much indebted to them for their time spent in making Genstat more widely available. (Alvey, Banfield, Lane, P. K. Leech, Nelder, Payne, Ross, Simpson, Todd and White)

Monitoring of use and performance. Improvement in the efficiency of use of programs depends upon information about the way they are actually used, and on how they perform in different machine environments. The monitoring of Genstat usage at Rothamsted (*Rothamsted Report for 1976*, Part 1, 332) covered 800 jobs run during 1 week. More than half ran successfully, in that no diagnostic faults were reported; of these about 60% concerned the analysis of data from designed experiments, and a further 30% used the regression facilities. About 5% of jobs failed because of syntax errors in the program; of those that failed while executing, nearly half did so because of faults in the data, while one-fifth ran out of space. (Simpson, with Beasley, Computer Department)

A preliminary investigation into the behaviour of Genstat on a virtual-memory machine was made on the ICL 2970 at the Edinburgh Regional Computing Centre. On such machines, a program, instead of occupying a fixed part of the store during execution, exists on a set of 'pages', which are being continually brought into and taken out of the working area as execution proceeds. Very little is known about the behaviour of large programs on paged machines. Sub-systems of Genstat were run with different sizes of data area; small areas (1, 4 and 16 K double words) showed little difference in the running time of a standard set of examples, but with larger areas (64 and 256 K) the program did run more slowly. The main result found concerned the use of short

https://doi.org/10.23637/ERADOC-1-134

integers, for which the 2900 series, unaccountably, provides no direct accessing instructions. Replacement of these by long integers decreased running times by about 20%, thus illustrating the very large fraction of time which computers spend retrieving and storing information compared to that spent processing it. (Simpson, with Mr. H. I. Greenwell, University of Edinburgh)

Macro library. Genstat allows the user to define and to use macros, i.e. sets of Genstat directives which can be named and then inserted in the program by referring to that name. Such a facility allows procedures not built into the system to be implemented in terms of lower-level operations which are available. Sets of macros may be stored together on a disc file to make a macro library and centres can, and do, construct their own macro libraries to cover their special needs. A standard macro library is now distributed with Genstat and is thus available to all users. It now contains 11 macros, of which five have been added during the year. CLASSF obtains a starting classification (for the CLASSIFY directive) using furthest-class nuclei; MANOVA carries out a multivariate analysis of variance; FIELLER estimates the LD50 with associated limits when a probit or logit model is fitted using the regression package; CENSOR analyses censored data and ALIAS gives details of aliased model terms in an Analysis of Variance. The macro GLM, for fitting generalised linear models, is no longer required, being superseded by the extensions to the regression section of Genstat, which now contains these models built-in. A new macro GLMODEL extends the class (as in GLIM) by allowing the user to define his own link function and variance/mean relation. (Todd, with Lane, Banfield and Payne)

Documentation. The new manual described in the *Rothamsted Report for 1976*, Part 1, 333, was issued during the year to coincide with the release of version 4.01. The original has been extensively revised to meet comments and criticisms by readers having varied knowledge and experience of the system; master copies were produced within the Department, using a Diablo printer with paper-tape attachment, and reproduced by a printing firm using the offset litho process. (Alvey, Tett and Watson)

Distribution. There are now 40 licensees in ten countries, for seven different machine ranges. UK licences have risen from 13 to 23, the others being to centres in Australia (four), the USA (three), New Zealand (two), Netherlands (three), with one each in Bulgaria, Czechoslovakia, France, Germany and Sweden. (Alvey, Tett and Watson)

Generalised linear interactive modelling (GLIM). Release 2 of this program is now in use in more than 110 centres in 21 countries, and it is encouraging to find that several Universities have successfully developed teaching courses centred on the class of models which the program fits, with the program being used interactively by the students throughout the course.

Release 3, the main extensions for which were given in the *Rothamsted Report for 1976*, Part 1, 333, was completed during the year and extensively tested on a variety of real data sets. The facility whereby the user may define his own link functions and variance/mean relations has found extensive use and greatly extends the class of models that the program can fit. The user's manual was also completed; it includes five chapters giving the statistical background to the models, an introductory guide to all the directives in the language, and a reference section giving the full syntax and semantics of each directive, together with a set of worked examples. The master version of the program and the manual have been delivered to the Numerical Algorithms Group (Oxford) who will

https://doi.org/10.23637/ERADOC-1-134

prepare and publish the manual, as well as distributing the program. (Baker and Nelder, with Mr. M. R. B. Clarke of Queen Mary College, London)

Maximum Likelihood Program (MLP). Version 3.04 was successfully implemented on the following machine ranges: IBM 370 at Cambridge, Cyber 7600 at CSIRO, ICL 1906 at Reading University, and Burroughs 6600 at Wellington, New Zealand. Conversion is in progress for the CDC 6000 series at Imperial College. Assistance was given to the various implementors. A reduced version written for an ICL 1901A by Mr. F. B. Lauckner was adapted for use on a small 1901 machine at Mauritius. The reduced version fits curves only.

Version 3.05 was developed throughout the year and its release will be timed to coincide with the manual, a draft of which is in preparation. This version incorporates several minor syntax changes to improve the consistency and ease of use of the language. Restrictions on operations on partitioned data variates have been removed and dynamic redefinition of the data structure is allowed. One-dimensional transects through the likelihood may be plotted. Fourier curves and Gaussian curves have been added, and stable ordinate analysis now minimises the sums-of-squares of correlations to obtain an optimal set of ordinates. Assay analysis now includes the complementary-log-log transformation, and dilution series allow negative binomial errors. The parts concerned with frequency distributions and genetic models have been extensively revised. Gamma and negative binomial errors are allowed in general models, with estimation of a linear scale parameter in the case of gamma errors. The method of finding confidence limits of functions of parameters has been improved (using Lagrangian multipliers). All non-normal weighted residuals are now based on the likelihood ratio. First derivatives of fitted values with respect to parameters are now multiplied by the dispersion matrix giving a more useful set of vectors for interpreting the importance of each data point in providing information on the parameters. (Ross, White and Payne, with Mr. R. A. Kempton, Plant Breeding Institute, Cambridge)

Genkey. Version 3.01 was completed and released to the University of Reading, where it has been converted to the ICL 1900 computer, and to CSIRO, for conversion to the CDC 7600. This version contains several new facilities.

(1) Test results can now more easily be specified as partially variable, i.e. more than one result is possible but others do not occur.
(2) An algorithm has been incorporated for constructing irreducible test sets—these are sets of tests such that if any test is removed from the set some pair of taxa will no longer be separable. Such sets can be constructed either to distinguish a particular taxon from every other taxon, or else to distinguish every taxon from every other taxon. The latter type can be used as the basis of either a key or a diagnostic table containing a minimum number of different tests.
(3) Keys or diagnostic tables may be constructed for a subset of the taxa.
(4) Tests may be ranked and keys constructed using only tests with up to a specified rank. Used in conjunction with the already existing facilities for extending keys, this would, for example, enable a key to be constructed to identify the taxa as far as possible using only the easiest tests then extended in stages, allowing more difficult or costly tests to be used at each stage, until complete identification is achieved. A difficult test would thus be used only if essential.
(5) An algorithm has been developed to select sets of tests for probabilistic identification. The set selected is such that identification would be achieved with $x\%$ confidence, $y\%$ of the time, where x and y are specified by the user. It is also possible to specify

https://doi.org/10.23637/ERADOC-1-134

that certain tests must be in the set. For example, in medical screening tests like age and sex which are always known can be specified.

(6) Keys can now be constructed, such that up to n (specified) errors can be made when using the key without an incorrect identification resulting. (Paper 18) (Payne)

Commonwealth and Overseas

The senior post in the unit sponsored by The Ministry of Overseas Development (ODM) remained vacant almost throughout the year, and, although now filled, the occupant will not take up his position until well into 1978. In addition, P. Walker, under secondment to CIMMYT (International Maize and Wheat Improvement Centre) has now accepted a permanent post with the Centre. These staffing difficulties have inevitably curtailed somewhat the service we have been able to provide during the year. The pattern remains one of assistance with the design and analysis of experiments and surveys, using our computing facilities at Rothamsted, supplemented by visits to countries requesting assistance, and training periods for overseas staff.

Visits abroad. K. Ryder visited The West African Rice Development Association in Monrovia, Liberia, for 3 weeks in January. His main tasks were to review their network of variety, insecticide and herbicide trials and to advise on the statistical analysis of the results. He visited research stations in Liberia and Sierra Leone and spent a good deal of time in working through a local set of computer programs used to analyse the data and to produce summary tables of the trials. A report on the visit was sent to the Association.

In March, K. Ryder spent 2 weeks at the main agricultural station at Gaborone, Botswana advising three new research teams on their data-collection problems and on how the local computer service could be improved or augmented to cope with the increase in data. The work of the three teams ranged from basic experimental work on general agronomic practices, through evaluation of a new general purpose farm implement on farmers' fields, to a survey on the day-to-day running of a series of farms and the interaction between the farmers and the extension officers. A report has been submitted to the Botswana Government, including recommendations on data collection and on possible future development of the computing service.

In July, R. H. Wimble, at the request of the Government of Malawi, visited the Biometrics Unit of the Malawi Ministry of Agriculture at Makota Research Station. A report has been prepared for the Ministry concerning staffing, organisation, statistical consulting and computing.

Analysis. Data were submitted from the following countries during the year: Botswana, Bolivia, Gambia, Ghana, Libya, Malawi, Malaysia, Nepal, Nigeria, Tanzania, Thailand and Zambia. The crops involved included wheat, barley, sorghum, maize, millet, soya bean, cowpea, pigeon pea, groundnut, sunflower, oil palm, sugarcane, tea, cotton and potatoes. Nearly all of the work from Tanzania came from a series of experiments to study the effect of fertiliser and irrigation upon several varieties of sugarcane on differnt soil types. A series of 40 agronomy trials on groundnuts was received from Malawi, together with data from the second year of a fertiliser × spacing trial on tobacco. (Riley)

Two substantial pieces of work resulted from data received from Botswana and Nigeria. Two plots of nearly 1000 sorghum plants were grown in Botswana to study the effects of the time and intensity of attacks by three insects (*Heliothis arrugera*, *Chilo portellus* and *Lonquinquis sacchari*). The yield and 100 grain weights were recorded for each plant; these were then classified by six factors, namely number of heads, levels of attack for each of the three insects, time to reach growth-stage seven and time of death of flagleaf. The new extended regression facilities in Genstat were used to find which of the main effects

https://doi.org/10.23637/ERADOC-1-134

and two-factor interactions were significantly affecting yield and grain weight. (Ryder and Riley)

The data from an agricultural survey on 100 farms in Nigeria carried out by researchers from International Institute of Tropical Agriculture were analysed using RGSP. The farms were from three villages near Ibadan and were visited at least once a week for 56 weeks. On each visit, details of labour used, seeds and fertiliser bought and produce sold were recorded. The labour data were recorded for individual fields and included the type of work done, the crop involved, the age and sex of the workers, the wages paid and meals provided. The data occupied over 20 000 punch cards. Tables were produced for each village and for three farm sizes, traditional, medium and progressive, from which it is hoped to look at the efficiency of the farms and the patterns of agricultural practice in Nigeria. (Ryder and Church)

Visitors. We had visits from workers in Botswana, El Salvador, Gambia, Nepal and Tanzania, either before beginning projects abroad with ODM, or to discuss experiments and reports already completed. We gave training to an agricultural statistician from Nigeria, Mr. O. E. Asiribo, mainly in the use of our programs for the statistical analysis of data.

Other work. Some modifications were made to the new data-recording forms first introduced last year. These now include space for a sketch of the site plan, something which is very often omitted but which may be of great importance for analysis. Examples of the modified forms were sent to overseas workers and the comments received were favourable. (Riley)

A survey of the available papers on intercropping is being made, and a catalogue of the literature on both intercropping and competition has been produced and is being updated as information becomes available. New methods of analysis may be necessary for intercropping experiments, but work on these must await further work in the field and the availability of more data. (Riley)

Staff and visiting workers

Patricia J. Wherry joined the Department as a Scientific Officer in March. Pamela Tett left in September, to be replaced by S. A. Watson, in October.

J. A. Nelder visited the Division of Mathematics and Statistics of CSIRO in Perth, Adelaide, Brisbane, Canberra and Sydney during November and December, and afterwards attended the 41st Session of the International Statistical Institute held in New Delhi. He was also awarded the Guy Silver Medal (1977) of the Royal Statistical Society.

Genstat courses were given by N. G. Alvey at The Plant Breeding Institute, Cambridge, and at the University Computing Company London; he also lectured on Genstat and Use of Statistical Packages to the Reading and Avon groups of the Royal Statistical Society. Further courses were given at Leeds and Oxford by H. R. Simpson and P. W. Lane, who also contributed to a symposium at Brunel University. Lectures on Genstat were also given by C. F. Banfield at Bristol University, and a Seminar on analysis of variance using Genstat was attended by R. W. Payne at the Civil Service College, Sunningdale.

R. J. Baker attended a Symposium on Dose-response analysis, held in Bremen, where he presented a paper (Paper 7).

C. F. Banfield gave three talks during the multivariate analysis course organised by the Institute of Statisticians at the University of Kent, Canterbury, during March. The titles were 'Canonical variate analysis and procrustes rotation', 'Metric scaling' (principal

https://doi.org/10.23637/ERADOC-1-134

coordinate analysis), and 'Applications of multivariate analysis'. He also attended the IMA Conference on Numerical Software at the University of Sussex in September, where he presented a paper (Paper 9).

Jill F. B. Altman attended the ADAS Livestock Officers' Conference in November, where she presented a paper.

J. C. Gower presented invited papers at the 10th European Meeting of Statisticians, Leuven, Belgium, in August, and at the conference on the analysis of spatial data at Bath University in September.

R. W. Payne also attended the 10th European Meeting of Statisticians, where he contributed a paper on robust diagnostic keys.

G. J. S. Ross visited the Division of Mathematics and Statistics of CSIRO in Perth, Adelaide, Canberra, Brisbane and Sydney in February and March, and presented a paper at the 10th Symposium on the Interface between Statistics and Computing at Gaithersburg, Maryland, in April. In September, he was an invited speaker at the First IRIA Colloquium on Data Analysis and Information at Versailles, France, and also at the IMA Conference on Numerical Software at the University of Sussex (Paper 22).

R. H. Wimble was invited by ADAS Soil Scientists, Eastern Region, to give a 1-day seminar on the estimation of responses to fertiliser by field crops. He also lectured on design and analysis of field experiments to the training course for ADAS Soil Scientists, Grades A and B, at Bristol in January.

J. Wood was part-author of a paper which was read at the Annual General Meeting and Spring Conference of the World's Poultry Science Association, UK Branch, in April.

P. J. Zemroch presented a shortened version of his paper written in collaboration with Dr. R. A. Bailey and Dr. D. A. Preece, to the British Combinatorial Conference, Royal Holloway College, in July.

I. A. Rubanov from the Plekhanov Institute of National Economy, Moscow, joined the Department as a temporary worker in September.

Publications

General Papers

1 Alvey, N. G., Banfield, C. F., (Baxter, R. I.), Gower, J. C., (Krzanowski, W. J.), Lane, P. W., Leech, P. K., Nelder, J. A., Payne, R. W., (Phelps, K. M.), (Rogers, C. E.), Ross, G. J. S., Simpson, H. R., Todd, A. D., (Wedderburn, R. W. M.) & (Wilkinson, G. N.) (1977) *Genstat Manual*, Statistics Department, Rothamsted Experimental Station, 469.

2 Hills, M. G. (1977) *Fertiliser use in farm type regions of England and Wales* 1976. London: Ministry of Agriculture, Fisheries and Food (SS/SAF/24), 33 pp.

3 Baker, R. J. & Nelder, J. A. (1978) *The GLIM System Manual*, Release 3 Oxford: Numerical Algorithms Group.

4 Nelder, J. A. (1977) My kind of statistics. *Bias* **4**, No. 2, 77–81.

Paper in Rothamsted Report, Part 2

5 Church, B. M. (1978) Use of fertilisers in England and Wales, 1977. *Rothamsted Experimental Station. Report for 1977*, Part 2, 73–78.

https://doi.org/10.23637/ERADOC-1-134

Research Papers

6 Baker, R. J. (1977) Algorithm AS112. Exact probabilities derived from two-way tables. *Applied Statistics* **26,** 199–206.

7 Baker, R. J. (1978) The GLIM System. *Proceedings of the Bremen Symposium on Dose-response Analysis* 59–65.

8 Banfield, C. F. (1978) Multivariate analysis in Genstat. *Journal of Statistical Computation and Simulation* **6,** 211–222.

9 Banfield, C. F. (1978) Singular value decomposition in multivariate analysis. *The Proceedings of the Institute of Mathematics and its Applications on Numerical Software—Needs and Availability*. London: Academic Press, pp. 137–149.

10 Banfield, C. F. & Bascomb, C. L. (1977) Characterisaticiis of peaty materials from organic soils (Histosols) in England and Wales. *Geoderma* **19,** 131–147.

11 Banfield, C. F. & Bassill, L. C. (1977) Algorithm AS113. A transfer algorithm for non-hierarchical classification. *Applied Statistics* **26,** 206–210.

12 Church, B. M. & (Lewis, D. A.) (1977) Fertiliser use on farm crops in England and Wales: Information from the Survey of Fertiliser Practice—1942–1976. *Outlook on Agriculture* **9,** 186–193.

13 Gower, J. C. (1977) The analysis of three-way grids. In: *The Measurement of intrapersonal space by grid technique*. Vol. 2 *Dimensions of intrapersonal space*. Ed. P. Slater, London, New York: Wiley, pp. 163–173.

14 Gower, J. C. & (Laslett, G.) (1978) Explicit singular value decompositions, spectral decompositions and inverses of some skew-symmetric matrices. *Utilitas Mathematica* **13,** 33–48.

15 Leech, F. B., (Smith, I. M. & Hill, R.) (1977) The effect of dietary-supplements of copper, zinc and iron sulphates alone or with a chelator on survival in experimental fowl typhoid in the chick. *Avian Pathology* **6,** 425–434.

16 Nelder, J. A. (1977) A reformulation of linear models. *Journal of the Royal Statistical Society* (Series A) **140,** 48–76.

17 Nelder, J. A. (1977) Multi-dimensional contingency table with one factor as a response. *The Statistician* **6,** 41–42.

18 Payne, R. W. & (Preece, D. A.) (1977) Incorporating checks against observer error into diagnostic keys. *New Phytologist* **79,** 201–207.

19 Payne, R. W., Smith, J. E. & Bainbridge, A. (1977) Ethirimol sensitivity in populations of Erysiphe Graminis from plots of spring barley. *Annals of Applied Biology* **87,** 345–354.

20 Payne, R. W. & (Wilkinson, G. N.) (1977) A general algorithm for analysis of variance. *Applied Statistics* **26,** 251–260.

21 Ross, G. J. S. (1977) Models and designs for vaccine trials. *Pesticide Science* **8,** 401–404.

22 Ross, G. J. S. (1978) Curve fitting using the Rothamsted maximum likelihood programme. In: *Proceedings of the IMA Conference on 'Software—Needs and Availability'*, London & New York: Academic Press, pp. 203–223.

23 Sparrow, P. E., (Johnson, P. S. & Boyd, D. A.) (1977) Manurial experiments with spring barley on chalk soils in Southern England. *Experimental Husbandry* **32,** 8–18.

24 Sparrow, P. E. & (Jones, E.) (1977) Magnesium uptake by herbage. *Journal of the Science of Food and Agriculture* **28,** 905–910.

https://doi.org/10.23637/ERADOC-1-134

25 SPARROW, P. E. & (RUSSELL, R. D.) (1977) Tests of long-term residues of phosphorus fertiliser. *Journal of the Science of Food and Agriculture* **28,** 687–693.

26 SPARROW, P. E., (PHILLIPS, F. T., ETHERIDGE, P., KAVADIA, V. S. & SETHI, G. R.) (1978) Translocation of ^{14}C-dieldrin in leaves of young cotton plants from spray droplets. *Annals of Applied Biology* **89,** 51–59.

27 WHERRY, P. J., (CLIFFE, R. E., SMITH, J. W. G. & TILLETT, H. E.) (1977) Excess mortality associated with influenza in England and Wales. *International Journal of Epidemiology* **6,** 115–218.

28 WHERRY, P. J. & (SMITH, J. W. G.) (1978) Poliomyelitis surveillance in England and Wales 1969–1974. *Journal of Hygiene* **80,** 155–167.

29 ZEMROCH, P. J. & (MARDIA, K. V.) (1977) Table of maximum likelihood estimates for the Bingham Distribution. *Journal of Statistical Computation and Simulation* **6,** 29–34.

30 ZEMROCH, P. J., (SCHORAH, F. J., SHEPPARD, S. & SMITHELLS, R. M.) (1978) Leucocyte ascorbic acid and pregnancy. *British Journal of Nutrition* **39,** 139–149.

https://doi.org/10.23637/ERADOC-1-134