

Thank you for using eradoc, a platform to publish electronic copies of the Rothamsted Documents. Your requested document has been scanned from original documents. If you find this document is not readable, or you suspect there are some problems, please let us know and we will correct that.



ROTHAMSTED
RESEARCH

Report for 1974 - Part 1

[Full Table of Content](#)



Statistics Department

J. A. Nelder

J. A. Nelder (1975) *Statistics Department* ; Report For 1974 - Part 1, pp 287 - 299 - **DOI:**
<https://doi.org/10.23637/ERADOC-1-131>

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.
P. Walker, B.A.
G. N. Wilkinson, M.Sc.
R. H. Wimble, M.Sc.

Senior Scientific Officers

N. G. Alvey, M.I.S.
H. H. Spechter, PH.D.
A. J. Vernon, B.Sc.
R. W. M. Wedderburn, B.A.

Higher Scientific Officers

C. F. Banfield, B.Sc.
Mary G. Hills, B.A.
R. A. Kempton, B.A., B.PHIL.

Scientific Officers

Mrs. Jill Altman, H.N.C.
C. J. Dyer, B.Sc.
P. Lane, B.A.
Penelope K. Leech
R. W. Payne, B.A.

Mary T. Simpson, B.Sc.
P. E. Sparrow, M.Sc.
A. D. Todd, M.Sc.
R. P. White, B.Sc.
J. Wood, M.Sc.

Assistant Scientific Officers

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

Secretary

Ruth T. Hunt

Clerical Staff

Jill A. Cain
Mrs. Anne M. Carnegie
Mrs. Doris A. Court
Mrs. Elsie Davies
Mrs. Janet A. Day
Mrs. Jennifer D. Dewhurst
Mrs. Susan M. Hatfield
Susan T. I. Hyde
Linda M. King
Mrs. Patricia K. Neill
Mrs. Sheena Perry
Mrs. Jean H. Povey
Mrs. Verona A. Roberts
Carolyn Smith
Mrs. Betty E. Spratt
Mrs. Brenda J. Watler
Janice R. Woods

Introduction

The department has again been concerned with a wide range of topics, some of immediate practical agricultural value, some related to longer-term developments, and some to the development of statistics itself as a subject. The interchange of ideas at various levels is essential to progress in the development of quantitative models for the interpretation of data.

Practical applications

Crop responses to fertiliser. Examination of results of more than 100 experiments from three series of multi-level nitrogen experiments with spring barley showed that most though not all crop responses to nitrogen were well represented by linear segments (split lines) (in collaboration with Messrs P. Needham and P. A. Johnson, ADAS); this finding can be related to early theories of Liebig and Blackman on the role of limiting factors. However, the chief object of these and other series of multi-level fertiliser experiments is to formulate more accurate practical recommendations for fertiliser use. The suggestion in the 1973 report that grouping results of different sites need not result in the loss of the split-line form of response has been strengthened by results of experiments with sugar beet, potatoes and barley (with Messrs D. J. Eagle and R. D. Russell, ADAS

ROTHAMSTED REPORT FOR 1974, PART 1

and Dr. A. P. Draycott, Broom's Barn). When experiments with these crops were grouped by soil series, previous rotation and region, the amount of fertiliser needed to reach the turning point from the rising segment to the 'plateau' segment differed little from field to field or year to year, and so the group averages retained the split line form. This means that unless there is a several-fold increase in the relative cost of N, recommended N dressings should remain almost unchanged. By contrast, except for crops on deficient soils or for some high-value crop like potatoes, the average crop response to P fertiliser is already insufficient to cover the cost of the dressing, though some may still be needed to maintain soil reserves.

Decreased fertiliser recommendations might be one way of diminishing agriculture's unwanted contribution to the N and P content of our water resources; at an ADAS/ARC Conference on Water Quality the rates of response/kg N and P of the principal crops were used to estimate the likely losses to the farmer and the nation from sub-optimal dressings (Paper 14).

Although official fertiliser recommendations have not long been published, ways of making them more accurate and also simpler are already emerging from the many experimental results that have appeared recently. With ADAS colleagues we have begun work on the relation of N requirement to amounts of summer and winter rainfall and to the N Index system. (Boyd, Dyer, Sparrow and Victor)

The diversity of species. Work continued on indices of species diversity, and used data from the Rothamsted Insect Survey. Yearly catches from 100 traps were used to compare the fit of four different models for describing the distribution of species frequencies in a sample. A generalised form of Fisher's logarithmic series, based on the beta distribution of the second kind, is proposed for fitting data on species-abundance with exceptionally long tails, and gives a better general fit to light-trap data than the log-normal model (Paper 21). (Kempton)

Multivariate analysis. The substantial facilities now available for multivariate analysis were again in demand. The technique of Procrustes rotation for the comparison of a set of configurations of points was applied to 46 configurations obtained from 23 taxonomists, 11 experienced and 12 inexperienced, each using both intuitive and numerical scores. The configurations represented a non-metric scaling applied to the characters of 12 Caminalcules (a group of imaginary animals created by Joseph H. Camin). By rotating each configuration to best fit, a distance measure was obtained between each pair of taxonomists. A principal coordinate analysis then revealed consistencies and inconsistencies in the taxonomists' assessment of the same organisms. The same technique was applied to data on forest trees from the Institute of Terrestrial Ecology to compare the growth at different sites.

A non-hierarchical classification using a sums-of-squares criterion was used on termite measurements to classify them into varying numbers of classes from two to ten. Both metric and non-metric scaling techniques were applied to dissimilarity matrices relating to material from 20 archaeological sites, and to matrices derived from skeletal remains from mesolithic sites in Western Europe. In all cases the results were found to be very similar, showing the much simpler algorithm of metric scaling to advantage. Comparisons of these kinds are being collated.

C. L. Bascomb (Soil Survey) supplied data for soil sample classification. A series of nine tests was applied to the 1469 samples giving a binary key for each. A sample was then allocated to the class that was defined by a similar key to that of the sample. This procedure of testing variate values and allocating samples to classes was easily performed using the CALCULATE directive in Genstat. (Banfield)

STATISTICS DEPARTMENT

Cluster analysis was applied to strains of bacteria, soils, and higher plants. The connections between data, clusters and principal co-ordinates were related to models of non-random sampling and aggregation. (Ross, Lane and White)

Combining information in assays. When an assay is used repeatedly on a routine basis, information about the parameters, particularly the slope, accumulates, but is often not used. Use of such prior knowledge is specially important where assay units are expensive, as with assays on foot-and-mouth disease vaccines. Vaccines may lose their protective ability during storage and regular testing is necessary. For assay groups of cows are vaccinated at a series of dilutions, and after inoculation with the disease are examined for presence of symptoms. All cows taking part in the assay must be killed. The combination of information on the slope from different assays increases the amount of information given per animal; when the slope can be taken as known the total number of responses is sufficient for the median protective dose and simple charts can be made for all outcomes of the assay. (Ross, Lane and F. B. Leech)

Livestock experiments. We have helped in revising the ADAS Advisory Bulletin on the composition and nutritive value of feeding stuffs. This involved checks on the internal consistency of data for individual feeds, and the calculation of revised estimates of digestibility coefficients and metabolisabilities. (Altman and Lessells)

Work on the design and analysis of ADAS animal experiments and investigations has continued. A projected ten-year co-ordinated experiment comparing different systems of cross-breeding pigs has involved detailed preparation of forms and coding instructions to ensure consistent and easily handled data. (Lessells)

After collection data must be checked for possible errors, stored, then later retrieved, merged and analysed. Where any operation can be standardised, a standard program module can be written and pre-tested, and its action on a particular set of data controlled by variable parameters stored with the data themselves. Several such program modules have been written in Genstat for the analysis of milk data, which may include grass and yard phases varying in length between animals. The aims of speeding up analyses and reducing programming errors are being achieved, and all outstanding winter-calving dairy experiments have now been analysed. (Spechter and Lessells)

The coordinated experiments comparing the performance of Limousin and Simmental cross-bred cattle with native breeds provided extensive data for analysis. Four ADAS Experimental Husbandry Farms were involved together with the Norfolk Agricultural Station. All animals were individually fed *ad-libitum* on the same complete diet, and serial slaughter was adopted to ensure that the degree of finish covered a wide range within each breed. One side of the carcass of each of 132 animals was jointed and fully dissected by the Meat and Livestock Commission; the results initially were expressed in 190 measurements, reduced finally to 15.

When compared at the same degree of finish, as indicated by the fat percentage of the carcass, the food cost per pound of carcass weight was lower for Herefords and Limousins than for Friesians and Simmentals. The carcass composition of the Limousins was superior to that of all other breeds in having more meat, less bone, and a higher proportion of high-priced cuts. In addition to these breed comparisons the data provide valuable information concerning the optimum degree of finish at which cattle should be slaughtered. (Lessells and Spechter)

Surveys. The Rothamsted General Survey Program has again been widely used for surveys on many topics.

Fertiliser practice. Survey information about how farmers use fertilisers is particularly

ROTHAMSTED REPORT FOR 1974, PART 1

valued now that estimates of total consumption are no longer available from subsidy statistics, and of immediate interest because of very large increases in fertiliser prices.

In 1974, the new series of collaborative surveys begun in 1969 with the ADAS and the Fertiliser Manufacturers Association (FMA) was therefore extended. Of the increased sample of 1400 farms, the FMA contribution to field work (800 farms) was done on their behalf by a market research company. This, and the need to use temporary coding staff, required complete re-design of survey forms and instructions (with more explicit questions and precoding) and was a valuable experience. Preliminary figures based on 1100 sample farms were available early in December: they showed a continuing increase in use of 'straight' N on grassland, and some reduction in use of P, particularly on grassland. Estimates of fertiliser stocks held on farms were also obtained. The results of the 1973 survey have been written up (Papers 2 and 3), and analyses are proceeding on the soil analysis data for 1971-72, which was received during the year. (Church and Hills)

Other surveys. We studied the feasibility of a multi-purpose cereal survey for the MAFF, and a survey of the land disposal of sewage sludge for the Department of the Environment. Consulting continued on ADAS surveys of pesticides and cereal foliar diseases (both from MAFF Plant Pathology Laboratory) and on wild oat surveys (ADAS, Northern Region). The main analysis of the 1973 National Potato Harvester Damage Survey, covering 665 farms was completed; it was estimated that 23% of the harvested crop had sustained flesh damage and that 13% of the otherwise undamaged potatoes had internal bruising. (Church and Hills)

A National Survey of the Keeping Quality of Bulk Milk started in February. An updated analysis is prepared each month and copies circulated to laboratories in England and Wales and to members of an informal consultative group. The survey of dystokia in Friesian heifers entered its second year; the results confirm those already reported (*Rothamsted Report for 1973*, 222) and also highlight herd differences. (F. B. Leech and P. K. Leech)

Routine analysis. The amount of data handled this year rose by 5% following a 20% rise in 1973. There was less work from overseas projects but more from ADAS. Turnround time was again reduced, this year by half a day to 9½ working days. (Dunwoody, Dyer, Simpson, Sowray and Todd)

Treatment plans produced by the design program can now be read directly by Genstat, thus avoiding re-transcription with the associated risk of error. (Todd)

Theory

Statistical inference. Work has continued on the basic principles of statistical inference, and on Fisher's theory of fiducial probability. Fiducial distributions cannot be transformed like classical probability distributions, because of a requirement that a statistic and its corresponding parameter must retain a certain monotonic ordering relation under transformation. This restricts transformability and also implies a restricted additivity of fiducial probability measures. Consequently fiducial distributions, even one-dimensional ones, need not be unique and this non-uniqueness represents an intrinsic and empirically verifiable property of finite data. It has been formulated as an *inconsistency principle* which resolves many difficulties of fiducial theory. However, at the same time it has adverse implications for Bayesian inference, the fundamental question being whether subjective prior information can be properly represented by a probability distribution, and whether coherency, however desirable mathematically, is achievable in actual procedures of inference.

STATISTICS DEPARTMENT

Papers have been prepared on the fiducial combination of information on means and variances, and on more general principles of statistical inference. (Wilkinson)

The theory of stable parameters (*Rothamsted Report for 1973*, Part 1, 223) was further developed, and sheds light on the conflict of objectives between the statistician and the biologist. The biologist asks for more complex models as a better approximation to reality while the statistician simplifies models to obtain more reliable parameter estimates. The statistician can ask for more data, better design, ancillary information or less random variation, according to what is practicable. Modelling is seen as a three-phase cycle of definition, estimation and re-design. Practical implications of the theory have been worked out for several examples.

Curve fitting in terms of stable ordinates works because each parameter mainly represents a different segment of the curve. Constraints on the possible ranges of stable ordinates reveal that curves such as the generalised hyperbola, the double hyperbola and the generalised logistic can tolerate only a small range of error, which limits their usefulness in biology. (Ross)

Multivariate analysis. Work continued on the empirical distributions of goodness-of-fit criteria for hierarchical classifications. Three groups of criteria were found from those studies. Kendall's Tau, Lerman's number of inversions and the cophenetic correlation formed one group, criteria associated with ordination formed another group and the residual sum of squares after a Procrustes rotation was the remaining criterion. A particularly simple relation was found for the expected value of the last criterion in terms of the numbers of units and variates in the data matrix. This work formed the basis of an invited paper at the 8th International Biometrics Conference, and will be published in the conference proceedings (Paper 18). (Banfield and Gower)

The Procrustes rotation method for the comparison of related configurations of points is proving a valuable practical technique in multivariate analysis, but, as with hierarchical classification techniques, assessment of any particular set of results is difficult without knowledge of those that would be obtained in the absence of any real structure. The general problem is very difficult, but some results have been obtained for the simplest Procrustes rotation problem where samples from distinct multinormal populations with means μ_1 and common dispersion matrix Σ give sample estimates m_1 and S ; here explicit expressions have been derived for the moments of the null distribution of the rotation statistic, which measures the goodness-of-fit under rotation of the k true canonical means with the k sample canonical means. (Gower, with Professor K. V. Mardia, University of Leeds)

Many distributional forms arising in multivariate analysis can be obtained only by simulation studies; such studies often require the generation of samples from multivariate normal distributions, conditional on the values of certain statistics. The problem of generating such restricted samples can be reduced to that of generating random orthogonal transformations, and this problem has been solved using a technique applicable also to groups other than the group of orthogonal matrices. A paper has been submitted for publication, together with an algorithm (see below). (Wedderburn)

A simple device to obtain statistics conditional on the data is to perform a series of randomisations of the data. The resulting distributions of statistics of interest provide an empirical test for the existence of structure in a data matrix, and statistics such as the largest eigenvalue and the mean nearest neighbour are readily compared. (Ross, Lane and White)

Robust regression. It may happen in fitting a model to data that the data are homogeneous except for a few outliers or 'bad' points, caused perhaps by gross errors in

ROTHAMSTED REPORT FOR 1974, PART 1

recording. It may also happen that the model fits the data well, except at a few points having extreme levels of one or more treatment factors. Use of ordinary regression methods for fitting the model followed by examination of residuals is not a very sharp diagnostic tool for investigating such situations, and better ones can be devised (robust estimation procedures) in which the discrepant points are iteratively discounted to produce a fit in which they are given little weight. A new type of procedure has been devised in which the fitting process can be reduced to the iterative least-squares procedure used in fitting generalised linear models; a by-product of this procedure is a new way of fitting linear models by minimizing the sum of absolute deviations (L1-norm). The properties of this procedure are being compared with others proposed elsewhere, using an adaptation of the program GLIM (see below). (Lane and Nelder)

Analysis of variance. The algorithm in Genstat used to produce the analysis of variance of data from designed experiments is restricted to generally balanced designs. By contrast the algorithm used for regression makes no assumptions whatever about possible symmetries, and so fails to take any short cuts that the recognition of such symmetries makes possible. Some bridge-building between these extremes is required, and an iterative method has been developed for analysing designs that are not generally balanced; numerical studies indicate that very good convergence rates can be obtained when the canonical efficiency factors for any one term of the model are not too different. (Wilkinson)

Transformations for exponential families. Exponential families of distributions, which include the Normal, Poisson, Binomial, and gamma distributions, can be characterised by the variance function $V(\mu)$ expressing the variance V as a function of the mean μ . The use on non-Normal data of techniques developed for Normal data may require simultaneous transformation of the data and the corresponding theoretical means. Four transformations are of interest, namely those that normalise the probability distribution, normalise the likelihood, stabilise the variance, and stabilise the information. It has been found that these can all be expressed as integrals of powers of the variance function; this result subsumes several well-known but hitherto isolated cases. (Wedderburn)

Diagnostic keys. In a non-probabilistic key the outcome of a test on a species is always the same, whereas in a probabilistic key the outcome is more or less uncertain and probabilities must be assigned to the possible results. Several existing criteria for selecting binary tests in non-probabilistic keys have been investigated and found to have disadvantages; this work led to the discovery of a new criterion without these disadvantages. A paper has been prepared (Gower and Payne). A new criterion has also been devised for probabilistic keys and will be described elsewhere. (Payne)

Statistical programming

Genstat Mark 3. Though a few new facilities have been added, this has been mainly a year of consolidation. Two releases were made, 3.03 in January and 3.04 in July. The efficiency of the version for the IBM 360/370 range has been improved by use of the H compiler and in other ways. (Simpson) Further improvements in efficiency have come from recasting the code, and this has often been combined with reduction in size and greater generality for the user. In particular the revision of the general CALCULATE directive has abolished many constraints on the types of operands allowed, and extended the operations on tables. (Alvey) Similar improvements have been made in the handling of the many tables produced internally during the construction of an analysis of variance. (Wilkinson)

STATISTICS DEPARTMENT

Among the multivariate procedures the ROTATE directive was extended and the maximal predictive criterion was added to the CLASSIFY directive. Again space was saved by reducing existing code. (Banfield)

A LOCAL directive has been implemented which restricts the scope of identifiers. This allows macros to be developed for general use without risk of an accidental clash of identifiers in the user's program and the macro. (Simpson)

A new directive CONTOUR gives contour plots of functions defined on a rectangular grid, and improved versions of the procedures for hierarchical cluster analysis and the calculation of minimum spanning trees were introduced. (Ross) For the fitting of non-linear models a new directive OPTIMISE was developed and tested with a generalisation of CALCULATE which defines models succinctly. The syntax has proved practicable and the directives will be included in a future release. (Alvey and Ross)

Documentation. There are now eight User's Guides available for Genstat, the three new titles being *Structure Storage and Retrieval*, *Input and Output* and *Analysis of Designed Experiments* (Papers 5, 6 and 7). All aspects of the system are now covered by these introductory guides, and in addition a prospectus has been prepared, together with an information leaflet on availability and licensing conditions. During the year an implementor's manual has been written, describing the internal structure of the system. This serves to document the program both for the programmer concerned with maintaining or extending the system, and for the implementor concerned with transferring it to a new machine range.

A macro library. The Genstat language is sufficiently general to allow new procedures to be written in it using the existing facilities of the system, and to allow these procedures to be defined as macros and stored in a library for general use. Any user may create such a library for himself, but macros likely to be of general use will be included in a special library and the text distributed with the program. A user's guide for this library is in preparation (Todd), and the nucleus of macros deals with canonical correlation, generalised Procrustes analysis, the evaluation of missing values in multivariate data (Banfield and Gower), iterative least squares (Wedderburn), and hierarchical analysis of variance. (Lane)

Transfer to other machines. The transfer of version 3.03 to the CDC 7600 was successfully completed during the year, thanks to the excellent work of Dr. K. Y. Kwok under the direction of Mr. H. C. Stone at the University of Manchester Regional Computer Centre. The inevitable unforeseen incompatibilities occurred, but much has been learnt from the exercise, and many of the changes in code have been marked in the program in such a way that the transfer of future versions will be done automatically by a simple editing program. Version 3.05, due for release at Rothamsted in January 1975, has been fully annotated in this respect, and also makes use for the first time of the two levels of core storage which characterise the CDC 7600.

A tape of version 3.05 has been supplied to the Statistical Research Service, Agriculture Section, Ottawa for possible transfer to the Univac 1108. The Fortran on this machine is highly compatible with that of the program, and the prospects of successful transfer are good.

Distribution. The licensing procedure is now in operation, and organisations pay a standard handling charge to the distributor, together with an annual licence fee, varying with the type of organisation. The Program Library Unit of the Edinburgh Regional

ROTHAMSTED REPORT FOR 1974, PART 1

Computing Centre acts as distributor for the IBM 360/370 version, and the University of Manchester Regional Computer Centre for the CDC 7600 version. The program is now also available at the University of London Computer Centre, the Universities of Bristol, Cambridge and Newcastle, University College, London, and the MRC Institute of Occupational Medicine. Licences have been issued to CSIRO (Australia) and to Reckitt & Colman Ltd. in the UK. A number of other enquiries have been received.

Generalised linear interactive modelling (GLIM). The first release of this program, described in the *Rothamsted Report for 1973*, Part 1, 227 was made by the Numerical Algorithms Group (Oxford) during the autumn. Versions are available for the following ranges: IBM 360/370, ICL System 4, ICL 1900, ICL 1906A, CDC 6000 series, and CDC 7600. There has been a gratifying interest in the facilities that GLIM offers, both inside and outside the Agricultural Research Service, and the program is now available on the 4-70 for both interactive and batch use. Work is in progress on a second release, which will include a plotting facility, simple macros, and other enhancements. (Nelder and Wedderburn, in collaboration with members of the Royal Statistical Society Working Party on Statistical Computing)

Maximum likelihood program (MLP). An interactive version was released, containing as much as possible of the program, including contour plots. A new HELP directive lists the facilities and the rules. Mnemonics were introduced to simplify the specification of options. Directives QUICKPLOT and PRINT provide useful output, the former being able to provide a visual analysis of variance in a one-way classification.

Shift operators and single-value functions of variates were provided, and likelihood plots were arranged to show 95% contours around the solution in planes defined by all pairs of axes. True 95% limits can be found by searching for the points of contact of tangent primes normal to each axis, but this process requires much more computation than the original solution. Correlations between successive observations can be specified to allow a modification of the likelihood, leading to more realistic parameter estimates when observations are on the same subject. Stable ordinates for curves can be evaluated. (Ross)

The sections dealing with probit analysis and models specified by transition matrices have been revised and simplified to allow them to be included in the interactive version. (Payne) A new facility allows the bivariate normal distribution to be fitted to two-way contingency tables, and the procedure is extendable to other distributions if required. (Kempton)

This program has been released to selected users elsewhere for trials, and documentation is now being developed for a general release.

Cluster analysis program (CLASP). New devices in HIERARCHY save substantial time and the NEIGHBOURS directive has been extended. IDENTIFY and ADDA-POINT were combined to improve efficiency. The hierarchy threshold can be defined automatically to produce a suitable number of stages. The FILLIN directive estimates missing values in the similarity matrix. New plotting facilities allow group numbers or principal co-ordinate scores to be plotted on geographical axes. Minimum spanning trees can now be produced on the incremental plotter. RANDOMISE reassigns variate values at random for empirical tests of structure. (Ross, Hawkins and Lane)

Genkey. The new test selection criterion for keys (see above) has been included in this program, and the input of data made easier and more efficient. The new representations of keys described in Paper 4 are all available in the latest version, which is now mounted

STATISTICS DEPARTMENT

also at Edinburgh. The Program Library Unit at the Edinburgh Regional Computing Centre has agreed to publish the manual for us. (Payne)

Other algorithms. An orthogonal matrix can represent either a pure rotation or a rotation combined with a reflection according as its determinant is 1 or -1 . Gower has devised an algorithm to distinguish these two cases. Algorithms for the generation of random orthogonal transformations (Wedderburn), and the scaling of graphs (Nelder) have been submitted for publication.

Additional routines have been developed and tested in collaboration with Dr. F. Yates for use in the Rothamsted General Survey Program. Those for data validation and updating and amendment of tabulation are proving especially useful. Other routines will file any set of tabular data in a form accessible to RGSP table-manipulation and print instructions. (Church)

A large program, such as Genstat, must be overlaid in order to fit into the available space, and possible overlay structures depend upon the inter-relationships of the constituent sub-programs. A program has been developed that scans the output from the Fortran composer on the 4-70, constructs a table of links between sub-programs, and displays it in a readable form to assist the construction of suitable overlay structures. (Simpson)

Commonwealth and overseas

The work we do for agriculturists overseas is supported by the Ministry of Overseas Development (ODM), and its variety is as great as ever. We received data this year from Cameroon Republic, Fiji, Iran, Lesotho, Malaysia, Malawi, Mexico, Nicaragua, Nigeria, Philippines, Sabah, Sarawak, Sudan, Swaziland, Tanzania and Zambia; major tropical crops like cotton, maize, cocoa, sugarcane and groundnuts predominated, together with alfalfa, tobacco, sunflowers, okra, cocoyam, teak, obeche, and gladiolus. Some data on the growth of chickens came from Nigeria, and other projects concerned soil measurements, the growth of timber in natural forests, and the description of natural vegetation.

Again much of our work has concerned the interpretation of data from many years and many sites, and may involve substantial numbers of figures. The successful handling of such data depends heavily on the computing tools available, particularly the ability to merge and select sets of data easily. Typical of jobs undertaken during the year were the analysis of five years of cotton trials from Gezira Research Station, Sudan, a set of 34 screening trials for cotton progenies in Tanzania, and a long-term rotation trial from Lesotho. (Macpherson)

Another long-term experiment is that from Serere in Uganda, where data are available from several five-year rotations over 29 seasons. The first five cycles have been summarised and a paper prepared. Sets of plot residuals of a given crop arising in different cycles were examined for correlation. All crops examined showed positive correlations of about 0.4 between successive cycles, and for cotton, correlations of this order persisted in crops grown 15 years apart. (Wimble, with Mr. A. R. Walter, ex-Uganda)

Data from Zambia giving the percentage cover in quadrats occupied by grass, herb and sedge species were analysed using CLASP. This provided the ecologist who collected the data with suggestions for groupings of species to use in the construction of a vegetation map of part of the Lochinvar National Park. Biologically meaningful groupings were derived, both for rainy and dry season data, and plots of the occurrence of important species referred to principal co-ordinates and to spatial co-ordinates were provided. (Wimble, with Miss Wendy Rees, Zambia)

Data on the growth of economic trees in natural forest in Nigeria for the period 1959-73 have been examined to determine whether the prediction of basal area increment

ROTHAMSTED REPORT FOR 1974, PART 1

could be improved; with Khaya species it has been found that an assessment of stem form status improves the accuracy of the basic predictor, initial basal area. (Walker, with Dr. R. G. Lowe, Nigeria)

Cluster analysis techniques were applied to a set of 46 quadrats of natural forest in Southern Nigeria using stock densities (measures of species dominance) of 153 tree species; the first two principal co-ordinates were related to soil type and species diversity (itself a measure of age). Further analysis is continuing with a reduced species list in an attempt to identify any other important causes of grouping. (Walker, with Mr. J. B. Hall, Nigeria)

We were consulted on the planning and preliminary analysis of a pilot survey of live-stock losses in Southern Ethiopia (Church), and on biometric problems connected with sugar-cane in Mauritius. (Ross)

During the year 20 overseas research workers called here for substantial discussions on their work and training was given to two overseas students.

Staff and visiting workers

Diana M. Hawkins, Jacqueline S. Edwards, C. E. Rogers, P. Anne Sowray and Agnes Victor left. Jill F. B. Altman, P. Lane, H. H. Spechter, J. Wood and R. P. White were appointed. Walker returned from a three-year secondment to the Federal Department of Agricultural Research, Nigeria.

Wilkinson was awarded the Guy Medal in Bronze of the Royal Statistical Society.

Nelder, Banfield, Ross and Wilkinson attended the 8th Biometric Conference held in Constanza, Romania. Nelder organised a session on statistical computing and computer simulation of biological processes. Banfield presented a paper jointly prepared with Gower. Nelder attended a symposium on computational statistics, COMPSTAT, held in Vienna and gave a paper (Paper 23). Gower attended the 8th International Conference on Numerical Taxonomy held in Lisbon and gave a paper (Paper 16). Wilkinson organised a weekend seminar on statistical inference where papers were read by himself, Professor G. A. Barnard, Dr. O. Barndorff-Nielsen and Professor A. Birnbaum. Boyd and Church attended an ADAS/ARC Conference on Agriculture and Water Quality and gave papers (Papers 14 and 12). Church attended a symposium on the study of agricultural systems at Reading University. F. B. Leech attended the First International Symposium in Equine Reproduction and gave a joint paper (Paper 22). At a meeting of the British Society for Animal Production he gave a paper with Mr. J. F. Fullbrook and Mr. D. L. Stewart on an investigation into the incidence of dystokia in Friesian heifers, an abstract of which was published in the Society's Proceedings. Leech also gave a paper at a conference on applications of the computer in veterinary medicine organised by the Association of Veterinary Teachers and Research Workers. Alvey attended a conference at Sheffield organised by the Institute of Statisticians on the use of computers in statistics.

Nelder spent three weeks in the Caribbean on a Ministry of Overseas Development (ODM) assignment acting as Assessor for the Biometrics Research Scheme project R2688 (1972-76). He visited centres in Jamaica, Barbados, St. Lucia, Guyana and Trinidad where he gave seminars, both general and statistical, to research workers. He also took and helped to get working the program package GLIM for use in the project.

Gower spent a month at the invitation of the Université de Paris-Sud in the Laboratoire de Biometrie at the Centre National de Recherches Zootechniques, Jouy-en-Josas. He was consulted on the use and implementation of Genstat and also on the use of multivariate techniques. He gave seminars on both subjects. He also attended the Summer

STATISTICS DEPARTMENT

School of the Swedish Statistical Society and gave lectures on maximal predictive classification and on multiplicative analysis of variance.

Sponsored by ODM, Walker visited the Soils Bureau, Manila, Philippines for four weeks where he advised on the design and analysis of experiments.

A course on Genstat was given by Alvey and Banfield at the Edinburgh Regional Computing Centre.

Four temporary workers spent varying periods in the department, three of them from overseas.

Publications

BOOK

- 1 NELDER, J. A. (1974) *Computers in biology*. London & Winchester: Wykeham Publications (London) Ltd., 149 pp.

GENERAL PAPERS

- 2 CHURCH, B. M. (1974) Use of fertilisers in England and Wales, 1973. London: *Ministry of Agriculture, Fisheries and Food (SS/SAF/9)*, 10 pp.
- 3 HILLS, M. G. (1974) Fertiliser types 1973: increases in the use of 'straight' N. London: *Ministry of Agriculture, Fisheries and Food (SS/SAF/11)*, 17 pp.
- 4 PAYNE, R. W. (1974) *GENKEY*. Inter-University Research Councils Series, Report No. 24. Edinburgh: Program Library Unit, Edinburgh Regional Computing Centre, 24 pp.
- 5 ROGERS, C. E. (1974) *Structure storage and retrieval*. *GENSTAT User's Guide* No. 7. Inter-University Research Councils Series, Report No. 21. Edinburgh: Program Library Unit, Edinburgh Regional Computing Centre, 19 pp.
- 6 SIMPSON, H. R. (1974) *Input and Output*. *GENSTAT User's Guide* No. 8. Inter-University Research Councils Series, Report No. 23. Edinburgh: Program Library Unit, Edinburgh Regional Computing Centre, 20 pp.
- 7 WILKINSON, G. N. & ROGERS, C. E. (1974) *Analysis of Designed Experiments*. *GENSTAT User's Guide* No. 2. Inter-University Research Councils Series, Report No. 16. Edinburgh: Program Library Unit, Edinburgh Regional Computing Centre, 16 pp.

PAPERS IN ROTHAMSTED REPORT, PART 2

- 8 CHURCH, B. M. (1975) Use of fertilisers in England and Wales, 1974. *Rothamsted Experimental Station. Report for 1974, Part 2*, 195-199.
- 9 JOHNSTON, A. E. & WEDDERBURN, R. W. M. (1975) The Woburn Market Garden experiment, 1942-69. I. A history of the experiment, details of the treatments, and the yields of the crops. *Rothamsted Experimental Station. Report for 1974, Part 2*, 79-101.

RESEARCH PAPERS

- 10 BANFIELD, C. F. & (HARRIES, J. M.) (1975) A technique for comparing judges' performance in sensory tests. *Journal of Food Technology* **10**, 1-10.
- 11 (BARNETT, J. A., BASCOMBE, S.) & GOWER, J. C. (1975) A maximal predictive classification of Klebsiellae and of the yeasts. *Journal of General Microbiology* **86**, 93-102.

ROTHAMSTED REPORT FOR 1974, PART 1

- 12 CHURCH, B. M. (1975) Use of N and P fertilizers on farm crops in England and Wales. (Given at ADAS/ARC Conference on Agriculture and Water Quality, December 1974.) *Ministry of Agriculture, Fisheries and Food, Technical Bulletin No. 32.*
- 13 DURRANT, M. J., DRAYCOTT, A. P. & BOYD, D. A. (1974) The response of sugar beet to potassium and sodium fertilizers. *Journal of Agricultural Science* **83**, 427–434.
- 14 (EAGLE, D. J., RUSSELL, R. D.), BOYD, D. A. & DRAYCOTT, A. P. (1975) Using response curves to estimate the effect on crop yield and profitability of possible changes in fertilizer recommendations. (Given at ADAS/ARC Conference on Agriculture and Water Quality, December 1974.) *Ministry of Agriculture, Fisheries and Food, Technical Bulletin No. 32.*
- 15 GOWER, J. C. (1974) Maximal predictive classification. *Biometrics* **30**, 643–654.
- 16 GOWER, J. C. (1975) Goodness-of-fit criteria for classification models and other patterned structures. *Proceedings of the 8th International Conference on Numerical Taxonomy, Lisbon, August 1974.*
- 17 GOWER, J. C. (1974) The mediancentre. *Applied Statistics* **23**, 466–470.
- 18 GOWER, J. C. & BANFIELD, C. F. (1975) Goodness-of-fit criteria for hierarchical classifications and their empirical distributions. *Proceedings of the 8th International Biometric Conference, Constanza, August 1974.*
- 19 HOLFORD, I. C. R., WEDDERBURN, R. W. M. & MATTINGLY, G. E. G. (1974) A Langmuir two-surface equation as a model for phosphate adsorption by soils. *Journal of Soil Science* **25**, 242–255.
- 20 KEMPTON, R. A. (1974) A model of selective migration. *Heredity* **33**, 79–86.
- 21 KEMPTON, R. A. (1975) A generalized form of Fisher's logarithmic series. *Biometrika* **62**, 29–38.
- 22 (LAING, J. A.) & LEECH, F. B. (1974) The frequency of infertility in thoroughbred mares. *Proceedings of the First International Symposium in Equine Reproduction, Newmarket, July 1974.*
- 23 NELDER, J. A. (1974) Genstat—A statistical system. *COMPSTAT 1974: Proceedings in Computational Statistics*. Ed. G. Bruckmann, F. Ferschi and L. Schmetter. Vienna: Physica Verlag.
- 24 NELDER, J. A. (1974) A user's guide to the evaluation of statistical packages and systems. *International Statistical Review* **42**, 291–298.
- 25 NELDER, J. A. (1974) Statistical packages: introductory remarks. *Bulletin of the Institute of Mathematics and its Applications* **10**, 165–166.
- 26 NELDER, J. A. (1974) Log linear models for contingency tables: a generalisation of classical least squares. *Applied Statistics* **23**, 323–329.
- 27 PAYNE, R. W. (WALTON, E. & BARNETT, J. A.) (1974) A new way of representing diagnostic keys. *Journal of General Microbiology* **83**, 413–414.
- 28 ROGERS, C. E. & WILKINSON, G. N. (1974) Numerical problems in recursive analysis of variance algorithms. *Bulletin of the Institute of Mathematics and its Applications* **10**, 141–143.
- 29 ROSS, G. J. S. (1975) Fitting models to ecological data. *Proceedings of the 8th International Biometric Conference, Constanza, August 1974.*
- 30 SADOVSKI, A. N. (1974) L1-norm fit of a straight line. *Applied Statistics* **23**, 244–248.
- 31 WEDDERBURN, R. W. M. (1974) Quasi-likelihood functions, generalised linear models, and the Gauss-Newton method. *Biometrika* **61**, 439–447.

STATISTICS DEPARTMENT

- 32 WEDDERBURN, R. W. M. (1974) Alternative derivation of the sum of squares in a non-full rank general linear hypothesis. *Journal of the Royal Statistical Society, Series B* **36**, 108–109.
- 33 WEDDERBURN, R. W. M. (1974) Generalized linear models specified in terms of constraints. *Journal of the Royal Statistical Society, Series B* **36**, 449–454.