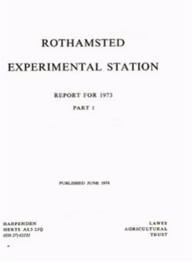
ROTHAMSTED
RESEARCH

Report for 1973 - Part1

Full Table of Content



ROTHAMSTED

EXPERIMENTAL STATION

REPORT FOR 1973

PART 1

PUBLISHED JUNE 1974

HARPENDEN
HERTS AL5 2JQ
(058 27) 63133

LAWES
AGRICULTURAL
TRUST

Statistics Department

J. A. Nelder

J. A. Nelder (1974) *Statistics Department ;* Report For 1973 - Part1, pp 219 - 229 **- DOI: https://doi.org/10.23637/ERADOC-1-130**

STATISTICS DEPARTMENT

J. A. NELDER

Our activities are one step removed from the agricultural process because our concern is not with crops or the pests of crops, themselves, but with information about them. We are concerned with how this information is collected, how its reliability can be assessed, how to expose patterns in data, how to develop theories for future prediction, and how theories can be matched against data. Numbers constitute our raw material, and the computer our prime working tool.

Practical applications

The topics described below have been chosen to illustrate the wide range of problems on which the department is consulted.

Crop responses to fertilisers. In the past, few experiments provided detailed information on the shape of the fertiliser response curve and so, when their results were used for making practical recommendations for fertiliser use, some arbitrary smooth curve had to be used to estimate the optimum. Recently several series of new types of experiment capable of giving detailed information on the form of response to one or more fertiliser nutrients have been done by ADAS soil scientists. The form of response differs between sites, but most sites show an approximately constant rate of increase up to a turning point (usually the optimum) beyond which yield changes little or, more often, slowly decreases. Grouping the sites by kind of soil, soil nutrient content and crop rotation accounts for much of the variation in the amount of fertiliser needed to attain optimum yield. Within these groups, seasonal differences in optima have proved fairly small, and so in making practical recommendations we can meaningfully represent response by two line segments joined by a sharply-inflected curve. This finding is of special importance in view of recent large changes in crop and fertiliser prices because it implies that farmers would be unwise to over-react to price changes. For example, from results of several years' experiments with barley grown in a cereal rotation on chalk and limestone soils in southern England, a change of ±50% in the relative prices of barley and nitrogen changes the optimal nitrogen dressing by only about ±5%. Even with up to five-fold increases or decreases in the current crop : fertiliser price ratio the recommended dressing would remain within the range 75–125 kg/ha N. (Boyd and Sparrow)

The development of mathematical models. An important theoretical component of fertiliser action is the adsorption of phosphate in soils. In collaboration with Holford (Chemistry Department) a model was developed assuming a mixture of two kinds of site on which the adsorption takes place, and this gave both an excellent fit to experimental data and readily interpretable parameters in the model. (Wedderburn)

In a joint project with Botany Department on photosynthesis, problems were recognised as analogous to those in electrical network theory and by applying relevant theorems from this theory the problem of developing a model was simplified. The model specified produced a non-linear equation, and a program was written to solve this. (Wedderburn)

The use of many different indices to describe the spatial aggregation of a single species, and species aggregation or diversity in a single sample, has led to slow progress in the analysis of these forms of ecological pattern. Indices should have sample stability and

https://doi.org/10.23637/ERADOC-1-130

allow comparison between sites; many of the statistics used at present, e.g. those based directly or indirectly on the sample variance, are highly unstable when applied to the very skew distribution met with in studies of animal populations. Studies of yearly light-trap samples from the Rothamsted Insect Survey have shown that the cumulative distribution function for species plotted on a logarithmic scale of abundance is almost linear except in the extreme tails. As the slope of the line appears to be a characteristic of the site and is unaffected by the size of the sample, it should provide a suitable index of species diversity. When the log-series and log-normal models of species abundance were fitted to the light trap data, the log-series better fitted samples from stable environments, while those from changing environments were more highly skewed and gave a closer fit to the log-normal model. When the log-series gave a good fit, the slope of the cumulative abundance curve could be expressed in terms of one of the parameters of the model; year-to-year variation in this parameter was small compared with variation between sites. It is hoped to link variation in the slope parameter with environmental characteristics of the sites (17). (Kempton, with Taylor, Entomology Department)

ADAS data on wheat bulb fly egg densities from several thousand fields sampled over the past 20 years in eastern England were analysed. For both light and heavy soil types over 70% of the variation in egg density between years could be accounted for by variation in three factors: availability of egg-laying sites, autumn rainfall (affecting date of sowing of winter wheat) and January soil temperature (affecting date of hatching of the eggs) (16). (Kempton, with Bardner, Fletcher and Jones, Entomology Department and Mr. F. E. Maskell, ADAS, Cambridge)

Probit analysis is a standard technique for analysing insecticide assays. When different strains of insect are cross-bred a compound response may occur. From the form of response to one such cross the genetic cross-over probabilities for three loci, one for resistance and the others for marker characters, were estimated, an allowance being made for possibly different probabilities that the different genotypes would reach maturity. The distance on the chromosome of the resistance locus from that of the markers could then be calculated, and the analysis suggested that the two markers were in fact alleles of the same factor. (Payne, with Farnham, Insecticides and Fungicides Department)

Work continued on models of mildew development on wheat and barley through the four stages, spore, germinated spore, appressoria, and hyphae, using data from plants grown in controlled-environment cabinets at different temperatures. Better agreement with the data was obtained by assuming that the probability of germination decreases with time, eventually becoming zero. Such time-dependence in the probability of change from one state to the next is likely to be a general feature of the system, and is being further explored. (Payne)

Multivariate methods. Multivariate techniques again found a range of applications. A paper has been prepared using data collected at the Meat Research Institute on the visual assessment of sides of beef by judges of various kinds. Rotational fitting gives measures of consistency between pairs of judges, and principal coordinate analysis then allows the judge's abilities to be displayed graphically. The same combination of techniques has now been applied to tasting tests on meat. (Banfield, with Dr. J. M. Harries, Meat Research Institute)

Multidimensional scaling, a technique for producing configurations in a few dimensions that are as close as possible to an original configuration in many dimensions, was applied to data supplied by the Experimental Cartography Unit of the Royal College of Art on mineral deposits from 138 sites. Work is continuing on how these graphical representations can be used to produce maps displaying the multivariate nature of the original data.

https://doi.org/10.23637/ERADOC-1-130

The same technique was applied to several matrices derived from different ways of assessing the resemblances between the 20 amino acids occurring in polypeptide chains. Two of these matrices were based on chemical properties only and the remaining matrices were calculated from an evolutionary model by Mr. R. Jorre and Professor R. N. Curnow of the University of Reading. In this model, the codons which specify the amino acids are equally likely to mutate to any other codon, and this may or may not change the amino acid. From these transition probabilities the average number of steps it took for one amino acid to mutate into another could be predicted and these averages formed the basis for nine more resemblance matrices. Multidimensional scaling gave a representation of the amino acids as described by each of the 11 resemblance matrices, and the agreement between each pair of representations was found by rotational fitting. Finally a principal coordinates analysis showed in just two dimensions how consistent were the different ways of measuring resemblance. The resemblances obtained from the evolutionary/mutational model agreed well amongst themselves but differed substantially from resemblances based on chemical properties alone. (Banfield)

Considerable use was made of cluster analysis, and classifications were obtained for soil profiles, species of *Peperomia*, strains of yellow rust, and *Salmonicida* and soil bacteria. (Hawkins)

Other examples of the application of multivariate techniques included the use of canonical variate analysis with data on tree pollen from Manitoba, principal component analysis on termite measurements, and both on data from Mr. M. Hills of the British Museum (Natural History) on skulls, stone tools and dog families. (Banfield)

Groups of experiments. Recent developments in experimental methods have been associated with increasing interest in the planning and design of series of experiments and in the analysis and interpretation of their results (12.11). The notion of doing experiments in groups goes back well into the last century, but because Fisher's ideas were developed in a context where each experiment tended to be considered in isolation, statistically designed experiments were first conceived as a means of providing valid results within ascertained limits of accuracy from single experiments. Important as these new methods were, they led, as Yates pointed out, to undue preoccupation with significance tests; they also affected experimental design by overemphasising the need for formal replication and so limiting the number of factors and levels of each factor that could be tested.

The shortcomings of this approach become obvious when we try to interpret the results from series of experiments, now an important part of the department's work. During the year, results of investigations done on Experimental Husbandry Farms (EHFs) and by ADAS agronomists and soil scientists were summarised; in particular, summaries of the results of several series of fertiliser experiments with maincrop potatoes and a national series of multilevel N tests with spring barley should provide a sound basis for advisory purposes. (Boyd, Dyer, Sparrow and Victor)

Livestock experiments. We continue to be involved in ADAS projects, including this year the planning of a long-term coordinated pig-breeding experiment. Data are being studied on the performance of Charolais cross-bred cattle compared with native breeds over three years on 40 farms, and on calvings from Limousin and Simmental bulls imported in 1970–71.

Advice has been given on the design of a coordinated series of experiments on EHFs comparing the performance of Limousin and Simmental cross-bred cattle with native breeds. In an earlier investigation with Charolais crosses, the degree of finish at slaughter differed between breeds, depending on the rate at which they mature. To secure some overlap in the degree of finish, and thus provide more objective comparisons, there will be a range of slaughter dates for each breed. (Lessells)

Extensive data on poultry from experiments done at Gleadthorpe EHF have been handled during the year. Heat loss and maintenance requirements were estimated from 1971–72 results with controlled environments and for the 1972–73 experiments we did period-by-period analysis of food consumption, mortality, egg production and weight. (Hills)

Surveys. We are now well equipped with the necessary computing tools for the rapid processing of survey data, and clients with small surveys can be instructed how to organise their own analyses using our standard programs. Consultation at the planning stage remains vital; among other things it enables us to write and test programs for analysis in advance of receiving the data. This procedure has been followed with a new survey on potato harvester damage sponsored by the Potato Marketing Board, for which the complete coded data are expected shortly.

Rothamsted insect survey. Records prior to 1968 were originally punched on paper tape but never analysed due to difficulties over interpretation and editing. A program was written to interpret these tapes and re-write them on magnetic tape so that they can now be merged with later records. Punching of all records is now keeping up with the inflow of data and the magnetic tapes are used daily by the Entomology Department to extract information, both for their own use and to satisfy the requests of outside researchers. Problems with tape corruption are now minimal. (Kempton)

Fertiliser practice. The fifth annual survey of fertiliser practice on a sample of 614 farms covering the whole of England and Wales was completed, in collaboration with ADAS soil scientists and representatives of the Fertiliser Manufacturers' Association.

Having assembled data about fertiliser use on individual crops from almost 3000 farms over the last five years, we can now provide estimates of recent average use, and of the proportions of major crops receiving different amounts of fertiliser nutrients for any defined farm type or geographical area.

The main trend over the last five years has been increased use of N on grassland. The survey results show a fairly steady increase in total N use during this period whereas official statistics based on subsidy claims show wide fluctuations, probably because of changes in stocks held on farms.

Survey information on the nutrient status of soils in England and Wales was summarised and a review paper prepared on the history of the surveys. (Church and Hills)

Other surveys. A national survey of the keeping quality of milk supplied for bulk tanker collection, planned and organised in conjunction with microbiology and dairy husbandry specialists of ADAS will start on 1 January 1974. The main purpose is to anticipate the results of applying tests of keeping quality specified under EEC regulations. The survey should also detect associations between test results and conditions on the sample farms, and so indicate causes of unsatisfactory milk supplies.

Analysis of two surveys by ADAS of the relationship between the body condition of ewes at tupping time and their subsequent fecundity showed that the better the condition at tupping time the larger the lamb crop; but breeds for which twinning is the norm did not produce more triplets.

A survey of dystokia in Friesian heifers mated to bulls of different breeds (including Friesians) has been analysed for ADAS. This showed that Friesian heifers calving for the first time outside of the age limits of about two to two-and-a-half years, more often had difficult parturitions and more often produced non-viable calves than when parturition occurred within this rather narrow age interval. The age at parturition for optimum results depended to some extent on breed of bull. Survey results for a second year are being collected and this will improve the reliability of the estimates of calf mortality rates. Papers have been prepared on the two last-mentioned surveys. (F. B. Leech and P. K. Leech)

We continued to be associated with both the pesticide surveys and surveys of foliar diseases in cereals organised by the MAFF Plant Pathology Laboratory, the survey of wild oats and blackgrass organised by the ARC Weed Research Organisation, and Potato Marketing Board surveys of seed–tuber diseases. (Church and Hills)

Routine analysis. After a fall last year, the amount of data handled this year rose by 20%, the increase coming mainly from overseas projects. Turn-round time has been reduced on average by at least two days to ten working days, and the sampling of jobs to monitor the process of analysis continues. (Dunwoody, Dyer, Sowray and Todd)

A new technique for handling the Station's field experiments has been developed whereby programs for analysis are partly developed and tested in advance, and then stored as macros within the Genstat system. When the results come in the pre-written programs are recovered and executed; this greatly reduces the turn-round time for the initial analyses. (Todd)

New forms for recording experimental data, with associated information, were introduced on a trial basis during the year. The aim is to minimise recording errors, and to simplify the description of data and instructions for analysis. (Dunwoody)

Theory

Statistical inference. Argument about the fundamental principles of statistical inference continues. A new resolution of some of the basic logical difficulties has been found, the key points of which are (1) that inferential probability distributions are essentially univariate, and relate to univariate functions of the unknown parameters and (2) when there is more than one unknown parameter, the totality of univariate inference distributions cannot in general be represented by a single multivariate distribution for all the parameters. This approach shows the alleged inconsistencies of the Fisherian theory of fiducial inference to be a strength rather than a weakness, and suggests that with some reformulation the theory could provide the kernel of a unified theory of inference.

Fisher's principle of total information, that all relevant information must be used in a scientific inference, and also that no spurious information (e.g. arbitrary prior distributions) should be introduced, has a corollary that may be termed the principle of reflexivity of inference. This states that a complete statement of inference must be reflexively related to the empirical information from which it is derived, so that it must be possible to argue in reverse from the inference to the relevant information on which it is based. The principle implies that an inferential probability distribution is not a complete statement of an inference, but needs to be supplemented by its sampling distribution, this being needed both for the combination of further empirical information and in any decision process arising from the inference. (Wilkinson, with Professor A. T. James, Adelaide University)

Non-linear models. The theory of stable parameters unifies many ideas in linear and non-linear inference. It is important to distinguish between three types of parameter: defining parameters that describe the algebraic formulation of a model, natural parameters that estimate quantities of interest to the investigator, and stable parameters that correspond to well-defined but contrasting features of the data. Stable parameters help to reconcile the defining parameters of the statistician with the natural parameters of the statistician's client. They are easily estimated numerically and their probability distribution can usually be assumed to be multivariate normal. The natural parameters or the defining parameters are obtained from them by transformation. Experimental designs for non-linear models should be such as to make the natural parameters stable, and if this is not possible then the natural parameters may never be adequately estimated.

The detection of stable parameters is aided by studying the derivatives of fitted values with respect to each parameter, weighted if necessary. The largest values, of either sign, show which data points contribute most to the estimation of each parameter. Similar patterns for different parameters point to poor conditioning and the need for transformations or improvements in design.

The relationship between the likelihood function and the 'solution locus' was studied, and a graphical representation was devised. Beale's non-linearity criterion was shown to be unaffected by parameter transformation, which limits its usefulness as an indicator. Transformation to stable parameters is a process of linearising the coordinate system within the solution locus, and the likelihood function, as a generalised distance measure from data point to solution locus, is approximately normal when the solution locus is nearly linear and the parameter contours are linear within it. Multiple solutions exist if there is more than one normal from the data point to the solution locus. (Ross)

Multivariate analysis. One of the perennial problems of practical multivariate analysis is that of assessing whether patterns, perhaps in the form of minimum spanning trees, extracted from a set of data, are likely to represent a repeatable characteristic of that data or are just artefacts of no significance. One approach to this problem is to simulate structureless data and create sampling distributions for statistics that describe certain kinds of pattern; their behaviour in the null case can then be assessed and used as a guide line, mainly to counteract over-optimism that real effects have been found. This process was applied to the so-called agreement statistics (Kendall's tau etc.) when used for comparing two representations of distance matrices, by minimum spanning trees and ultrametric distances. Some of the statistics were found to be closely related, others hardly related at all; clearly they do not all describe the same aspects of the data. (Banfield and Gower)

Progress has been made in analysing sets of multivariate samples, thus generalising some classical multivariate methods. A method of analysing three-way grids provides a first step in unifying the multiplicative analysis of variance with individual scaling models. The Procrustes method, whereby two configurations of points are stretched and rotated to a position of best relative fit, has been generalised to an arbitrary number of configurations; a Genstat program to do the necessary calculations has already been applied to data from two sources. Rotational fits are peculiar to multivariate samples and have no counterpart in univariate samples where only differences between means (translations) and differences in scale (dilations) are meaningful. However, in multivariate analysis both these aspects can be combined with rotations to give a useful orthogonal analysis of variance. (Gower)

Diagnostic keys. Although an optimum diagnostic key may be defined in more than one way, none of the definitions leads to practicable methods of construction. In practice heuristic methods are used which construct the key in stages, and at each stage the best test for division is selected according to some criterion based on the way each test divides the remaining species. Several selection criteria are in use, three of which have been compared with a fourth method, which in certain cases distinguishes which tests are clearly better. All three of the other criteria have errors of two kinds: they sometimes classify better tests as poorer and sometimes classify poorer tests as better. In the region where the most useful tests are found all three criteria are comparable. The Shannon-Information criterion tends to accept more poor tests than desirable. This finding supports recent criticisms of the assumption that Shannon Information necessarily measures the kind of information relevant to every problem. (Gower and Payne)

The relative abundance of species. The log series model for the relative abundance of different species can be derived by Poisson sampling from a mixed population described by a gamma distribution. This model has been generalised by using a beta distribution of the second kind in place of the gamma distribution. The resulting model has the same number of parameters as the log-normal model but preliminary results suggest that it provides a better general fit to data. An equivalent generalised form of the negative binomial was also found useful for describing the spatial distributions of parasites. (Kempton)

Analysis of variance of designed experiments. The algorithm used in Genstat to detect the various kinds of symmetry in data from designed experiments depends on the preliminary analysis of artificial variates, one for each term in the model. A new method has been discovered in which the analysis of a single artificial variate generates all the efficiency factors and orthogonality relations for generally balanced designs. The new method is more efficient and has better numerical properties than the old, though it does not generate such complete information when sets of effects are wholly or partly aliased with one another. (Rogers and Wilkinson)

Quasi-likelihoods. In generalised linear models (see *Rothamsted Report for 1971*, Part 1, 231) the error term is assumed to have a distribution belonging to the exponential family. In the maximum likelihood equations for estimating parameters in the model the relation between the variance and the mean plays a crucial role, and this relation can be used to define a quasi-likelihood. Quasi-likelihoods enable models to be fitted with weaker assumptions than are required for maximum-likelihood estimation. With this approach the exponential family of distributions appears as the least informative given the variance–mean relation. A paper has been prepared. (Wedderburn)

Statistical programming

Genstat Mark 3. The first release of this version was made in July, and a second release in November. The facilities described in last year's Annual Report are all working; in addition the user can now restrict some operations in the system to subsets of his data, and change the subsets dynamically. The scope of this facility is being steadily increased, it being currently available in analysis of variance, regression and tabular output. Several functions have been added under the CALCULATE directive, including some for quick calculation of simple regression coefficients; pseudo-factor levels can now be easily generated for use in the analysis of lattice designs. A new directive HIERARCHY allows hierarchical clustering using several different algorithms. (Nelder, Alvey and Ross)

The section on the analysis of designed experiments has been improved by the addition of a directive for extracting components of the output and assigning them to standard data structures. Submodels for treatment effects are now specified in the factorial treatment model formula, and a new operator has been introduced for specifying the pseudo-factorial structures required for analysing partially balanced lattice designs. (Rogers and Wilkinson)

A syntax-checking version of the compiler is now available on the Rothamsted 4-70; this allows complex Genstat programs to be checked before a full run is attempted. (Simpson)

Work on the third release is now almost complete, and this contains some substantial extensions. It will now be possible to interleave freely the compiling and executing phases during a job. This means that compilation of a current block of instructions may be broken off at any point in order to compile and execute an inner block (perhaps to read in and so obtain the size of the dimension of some structure). Similarly execution of a block may be interrupted to compile and execute an inner block, giving in effect a run-time macro. In consequence substitutions, which previously had to be done at compile-time, e.g. of formats, can now be made fully dynamic. This facility has already been found useful in writing general procedures for multivariate analysis and for data storage and retrieval. (Nelder and Simpson)

A new algorithm for classifying units into groups on the basis of multivariate data about them has been added to the system. (Banfield and Gower)

The graphical routines have been re-written to give an improved layout and the additional facility of printing up to 2 × 2 graphs per page. (Alvey)

Efforts are being made to hold the total size of the system at about its present level, so that if new facilities are to be added, space must be saved elsewhere by tightening the code. Considerable savings have been made in the graphical routines, the regression part and in the interpreter for principal component analysis. (Alvey, Banfield and Wedderburn) A restructuring of the general calculation routines is also expected both to reduce the size of the program and to increase the generality of the operations. (Alvey and Nelder)

General programs written in the Genstat language include some for diallel analysis, which can provide the Jinks–Hayman form of analysis, also Jones's analysis for the half-diallel. Information can be combined over blocks and sites. (Alvey)

Documentation. The reference manual was published during the year, and contains a full description of all the facilities of the system, beginning with a description of the language (12.8). Appendices list the diagnostics, summarise the syntax and give information on running jobs on different machines. There is a full index. Though the reference manual can be read sequentially as a single document, it is not intended to be the means by which a potential user learns how to use the system. For this purpose we are publishing a set of introductory *User's Guides*, of which four have so far appeared (2, 3, 6, 9), covering the Genstat language, matrix operations, table operations and regression. Three others are nearly complete. (Alvey, Gower, Nelder, Ross and Wedderburn) Other documentation provided for users of the system includes an annotated set of worked examples with the output that is produced when they are run, and a notice board giving the current state of known restrictions and faults. Both these items are supplied as files stored by the computer, and users may ask for copies as required.

Distribution. The version of Genstat for the IBM 360/370 series is distributed for us by the Program Library Unit of the Edinburgh Regional Computing Centre. The distribution is now being formalised by issuing a licence-to-use, renewable annually, for a fee to be negotiated. The System 4 version has been supplied direct to the Bristol University Computer Centre.

Link to RGSP. Two directives have been added to Genstat which allow tables to be passed in both directions between Genstat and Yates' Rothamsted General Survey

Program. The user can now easily make use of the special facilities of the two systems in one job. (F. B. Leech and P. K. Leech)

Church has collaborated in the testinf of new facilities in RGSP for the graphical display and amendments of tables, and for the formation of tables needed for calculating multi-stage sampling errors.

Transfer to other machines. The study begun last year with Imperial College on the possible transfer of Genstat to the CDC 6000 range of computers had to be abandoned because of unsatisfactory performance of the operating system and staff changes. During the year two new projects were begun, for transfer to the ICL 1900 and CDC 7600 series. The former is being undertaken by Mrs. L. Hayes and Mr. P. Griffiths of the Oxford University Computing Laboratory and the latter by Mr. H. C. Stone and Mr. K. Y. Kwok of the University of Manchester Regional Computer Centre. The start of both projects was delayed by difficulties in transferring magnetic tapes from one machine to another, but these have now been overcome.

Other programs

Generalised Linear Interactive Modelling (GLIM). This program was designed by the Working Party on Statistical Computing of the Royal Statistical Society to fit the generalised linear models of Nelder and Wedderburn referred to above. Several members of the department have contributed to it, and the program has reached the pre-release state with a draft *User's Manual*. It is now being tested at selected sites, and will eventually be distributed by the Numerical Algorithms Group Project, which is responsible for issuing standard libraries to British universities and others. The program uses the Genstat notation for specifying linear models and a simple interpretive language designed for interactive use at a terminal. The program has been mounted experimentally on the Rothamsted 4-70 in an interactive form for testing and development (12.7) (Nelder, Rogers and Wedderburn)

Maximum Likelihood Program (MLP). The general model fitting routines were extended and improved. The existence of these facilities proved exceedingly useful and new models could be fitted immediately without recourse to Fortran programming. First derivatives of fitted values were incorporated using numerical differencing. This made it possible to provide standard errors of fitted values. The 95% support limit facility was improved, and Beale's non-linearity criterion was incorporated. A FUNCTION directive enables functions of parameters to be computed, together with their standard errors and correlations. The program is now available at the Edinburgh Regional Computing Centre. MLP has been used to produce a package for fitting transition-matrix models. These models may be appropriate when individuals can be completely described by assignment to one of a finite group of states, and the whole system described at any time by the number of individuals in each state whose elements give the probability that an individual currently in one state will move to another state in the next time interval. The model consists of the transition matrix, and a vector giving the initial probabilities of individuals being in each state. The elements of the matrix and vector may be constants or functions of the parameters to be fitted. They may also be time dependent. (Kempton, Payne and Ross)

Using the experience gained in developing this program, the problem of including optimization procedures in Genstat were studied. Preliminary studies showed that the existing interpretive language, though capable of describing the necessary steps, was too slow in execution in its present form, so that special directives will be required; the form of these is being investigated. (Ross)

Cluster analysis program (CLASP). A new graphical presentation of the minimum spanning tree has been implemented, and this makes the output much more readily comprehensible. Contour density plots are now provided and give a convenient visual summary of the separation of major clusters. Derived-variate facilities are also available. (Hawkins and Ross)

Genkey. This program constructs diagnostic keys from a matrix of responses of a set of species to a set of tests, together with some ancillary information. Data input has been made more flexible, and the output more informative and more under the user's control. A paper describing the system has been prepared (12.21). (Payne)

Other routines. An algorithm to compute the median centre of a set of points has been accepted for publication (Gower) and three subroutines concerned with computing the latent roots and vectors of a symmetric matrix and with the singular-value decomposition of a rectangular matrix have been included in the Computer Department's System Library. (Banfield)

Commonwealth and Overseas

The Overseas Development Administration continued to support the work we do in advising on the design and analysis of agricultural experiments overseas and in providing a data-processing service. Data came from Bolivia, Fiji, Gambia, Ghana, Kenya, Malawi, Sabah, Sarawak, Tanzania, Uganda and Zambia, on crops as diverse as bananas, cocoa, coffee, cotton, groundnuts, kenaf, maize, millet, oil palms, sorghum, sugarcane, soya beans and tobacco. We had visits from 24 overseas research workers during the year, and one ODA student spent a training period here.

Much of the work arising from the experiments concerns the interpretation of results from many sites obtained over several years, and for this the storage and retrieval facilities in Genstat have been extensively used. For example, for fertiliser experiments in Zambia on continuous maize lasting five years we have provided annual analyses, investigated trends over the five years, and examined associated soil data and their relationships with yield. (Wimble and Macpherson, with Mr. A. Prior, Zambia)

The diversity of jobs undertaken for overseas clients is well shown by the following two examples:

Four years' data from 22 experiments on the rate of development of bacterial wilt in potatoes in Kenya were examined by fitting two models for the regression of proportion infected on time to first wilting. One model was the probit transformation with log (time) and the other, Van der Plank's 'Simple Interest' model. The probit model gave better linearity due to its better representation of the rather slow early development of the disease. Regression analysis of the slopes of the fitted lines on environmental variables showed that there was an optimum soil temperature for the rapid spread of the disease, the rate falling off with both higher and lower soil temperatures. (Wimble, with Miss E. Hind, now in Zambia, and Mr. A. H. Ramos, Kenya)

Several analytical procedures, including Hayman and Jinks analyses, were tried out on cotton data from 9 × 9 diallel cross experiments at two sites in Uganda. There was clear evidence of interaction between genotype and site in most variates, but interpretation of the crosses in genetical terms proved very difficult since the Jinks analysis indicated that a simple additive genetical model of parental and dominance effects was not consistent with the data and the dropping of particular parents apparently giving rise to difficulty did not lead to any marked improvements in fit. (Wimble and Macpherson, with Mr. H. Gridley, ex-Namulonge, Uganda)

We have also been consulted about the grouping of tree species in Zambia, and a long-term soil fertility experiment in Uganda for which there are data for 25 years. (Macpherson and Wimble)

We did extensive analysis of the data from two surveys of streptothricosis of cattle collected by the University of Ibadan, and another on tuberculin reactions in Uganda cattle. (F. B. Leech and P. K. Leech)

Staff and visiting workers

Margaret A. Currie and F. B. Lauckner left. Pamela Jones, Diana M. Hawkins and Jacqueline S. Edwards were appointed.

Gower, Nelder and Wilkinson attended the 39th Session of the International Statistical Institute in Vienna. Nelder organised a meeting on Aspects of Computing in Statistics at which Wilkinson was an invited discussant. Gower gave a paper on classification problems (12.13) and was an invited discussant for the multivariate meeting.

Gower was joint organiser of a conference on Multivariate Analysis and its Applications sponsored by the General Applications Section and the Multivariate Study Group of the Royal Statistical Society. Nelder gave a paper and Gower reviewed various aspects of multidimensional scaling. Gower and Professor K. V. Mardia of the Department of Statistics, University of Leeds reported on the conference in a joint paper (12.5).

Nelder attended the Third ARC Data Logging Symposium held at the National Institute of Agricutural Engineering and gave a paper (12.20). Gower was joint organiser of a conference on Computational Problems in Statistics held at the University of Essex and arranged by the Institute of Mathematics and Its Applications. Rogers and Wilkinson gave a paper.

Nelder and Gower helped to organise a joint meeting of the Royal Statistical Society and British Computer Society on the Assessment and Verification of Statistical and Other Scientific Software. Nelder gave a paper on the evaluation of packages and systems.

A course on Genstat lasting three days was given for the Edinburgh Regional Computing Centre by Alvey, Banfield, Simpson and Wedderburn.

Wimble spent three weeks in Zambia advising the Research Branch of the Department of Agriculture on statistical problems, experimental design and data analysis. Leech spent a week at Ibadan University attending an International Symposium on Dermatophilus Infection and gave a paper (12.18).

Five temporary workers spent varying periods in the department, two of them from overseas.