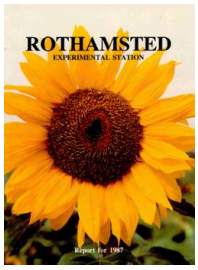
ROTHAMSTED RESEARCH

Rothamsted Experimental Station Report for 1987

Full Table of Content



Biomathematics Division

J. C. Gower

J. C. Gower (1988) *Biomathematics Division ;* Rothamsted Experimental Station Report For 1987, pp 29 - 42 **- DOI: https://doi.org/10.23637/ERADOC-1-28**

BIOMATHEMATICS DIVISION

STAFF

Head of Division J.C. Gower

Computing and Electronics Department

Head of Department
P.J. Verrier, B.Sc.

Senior Scientific Officers
K. Bicknell, B.Sc.
W.J. Moore, B.Sc.

Higher Scientific Officer;
Christine F. Thomas, B.Sc.

Scientific Officer
E. Wilson

Assistant Scientific Officer
Caroline Mitchell
A. Ball

Executive Officer
Denise V. Legg

Higher Professional and Technical Officer
R.N. LeFevre, B.Sc.

Technical
D.J. Peck, B.Sc.

Technical II
P.D. Compton
S.B.N. Day

Honorary Scientist
F. Yates, Sc.D., C.B.E., F.R.S.

Personal Secretary
Janet Why

Administrative Assistant
Gwyneth Frier

Statistics Department

Head of Department
J.C. Gower, B.A.

Senior Principal Scientific Officers(UG6)
Rosemary A. Bailey, D.Phil.
B.J. George, B.Sc.

Principal Scientific Officers (UG7)
R.W. Payne, B.A.
J.N. Perry, M.Sc.
Janet Riley, M.Sc.
G.J.S. Ross, B.A.
H.R. Simpson, M.A.

Senior Scientific Officers
P.G.N. Digby, B.Sc.
C.D. Kershaw, Ph.D.
P.W. Lane, M.A.
K. Ryder, B.Sc.
H.H. Spechter, Ph.D.
A.D. Todd, M.Sc.

Higher Scientific Officers
Anne E. Ainsley, Ph.D
C.J. Dyer, B.Sc.
Penelope K. Leech, B.Sc.
R.P. White, B.Sc.

Scientific Officers
Suzanne J. Clark
Janet I. Elsmere
K.W. Coleman, B.Sc.
S.A. Harding, B.Sc.
S. Langton, M.Sc.
Rosemary Poultney, B.Sc.
Gina L. Smith, B.Sc.
Suzanne Welham, B.A.

Visiting Workers
P.R. Krajewski, M.Sc.
T. Miwa, D.Eng.
H. Monod, D.Eng.
G. Bove, Laurea

Visiting Scientist
B.M. Church, B.Sc.

Sandwich Students
Ethna Prosser
Seema Sehdev

Departmental Secretary
Valerie Payne

Personal Secretary
Diane P. Dawkins

Specialist Typist
Christine Jaggard
Angela C. Piears

Administrative Officers
Anne E. Arnold
Brenda J. Watler

Administrative Assistants
Wendy L. Barthorpe
Karen J. Hakesley
L. Vera Wiltsher

INTRODUCTION

The Division is the only one of its kind within the AFRS. In March the Rothamsted Division together with the statistical and computing posts at Long Ashton merged to form a new IACR Division of Biomathematics. In recognition of both the additional emphasis on electronics development and the wider role required to encompass the computing and electronic needs

https://doi.org/10.23637/ERADOC-1-28

of the several sites of the new Institute these activities have been grouped under a new Computing and Electronics Department. This report is concerned with activities at the Rothamsted site but it must be recorded that liaison between the two sites has begun well and that in the future we can hope for a symbiotic effect that will benefit both Long Ashton and Rothamsted. Already this year about 20 Long Ashton experiments were analysed at Rothamsted and returned in good time for the publication of the *Preliminary results of the field experiments*. At Rothamsted we have noted with interest the Farm Management System developed by Lucey and Moody at Long Ashton. This promises to be an extremely useful way of automating some of the management of field experiments, enabling an up-to-the-minute state of planning and progress to be available to all concerned, much reducing the paper work and eliminating some of the possibilities of error. Of course the system is a prototype and has been in use at Long Ashton for only one season but it is already clear that the idea is worth exploring and perhaps generalizing for much wider use.

In September a third VAX 11/750 was installed. This has made only a marginal improvement in computer response-time. Rapid response is the single most important factor from the point of view of the computing scientist. Poor response not only directly slows down a computing run but it also interrupts the train of thought, so encouraging errors and further wasting time. Much scientific work is exploratory and the associated computing needs to be done in interactive mode, but effective interactive computing demands rapid response and good graphics. Innovative work postponed to await better computing probably explains the very rapid take-up of the additional resource of the third VAX and it is clear that our computing needs remain unsatisfied.

Rothamsted took delivery of its first Sun workstations in June. These are being coupled to an Ethernet, itself to be linked to AGRENET. The Sun workstations are now coming on stream and are available to a few scientists and statisticians with specially heavy computing demands. The relatively cheap dispersed workstation concept allows individual research groups to finance their own computing hardware provided that the infrastructure for the connectivity and associated file server and other services can keep pace with the needs. People have quickly adapted to using a new operating system (UNIX) and first reports are very favourable. UNIX based workstations are rapidly becoming the preferred computing engines in universities so the Sun workstations provide an entry to the wealth of applications software mounted internationally on UNIX systems.

The commercial release of Genstat 5 is a major event of the year which is highlighted in the General Report as well as below.

As in most years, the Statistics Department has done collaborative work with all the Rothamsted departments; it has also been consulted by at least six of the eight AFRC Institutes and by more than half the different sites. Professor A.A. Greenfield, an experienced statistical consultant in the industrial and medical fields, spent three months in the Statistics Department. It was useful to have an outside view of our consulting activities. Based on a review of recent Rothamsted publications, Professor Greenfield gave a talk to the Station, mainly criticizing obscurities of statistical presentation, lack of information and analyses that might have been done better. Effort put into improving such matters is certainly worthwhile but perhaps the more serious problem is persuading scientists to seek statistical and mathematical advice in novel areas of research where we have much to offer. Devolved computing brings with it problems of incorrect analyses, the need to answer queries and help with interpretation—all of which make additional demands on our stretched resources. The problem is not peculiar to Rothamsted; a similar tendency is noted below regarding our work with ADAS and indeed it is an international phenomenon.

The Overseas Development Administration (ODA) has decided to centralize its technical services at a new site in Chatham and the question arose as to whether the ODA-funded

https://doi.org/10.23637/ERADOC-1-28

Biometrics Unit, which has been at Rothamsted for over 20 years, should move there. In the event it was decided that the statistical and computing expertise available at Rothamsted together with library support, more than made up for the slight financial saving of moving, so the Unit will remain. In the following report there are examples which illustrate the benefit of such cross-fertilization.

COMPUTING AND ELECTRONICS DEPARTMENT

The Department provides a computing and electronics design, development and advisory service and support for equipment. The Electronics Group (i) designs, develops and produces equipment in support of computerization and instrumentation projects and (ii) provides a repair and installation service for computing equipment and laboratory instruments. (LeFevre) The Systems Group (i) develops data capture, small systems integration and system utilities software and (ii) provides a service for general purpose scientific computing and word processing. (Moore) The Applications Group (i) develops general applications and graphics software and (ii) provides advice and services for systems analysis and user training. (Bicknell) Projects frequently require expertise from two or more groups, so their boundaries are loosely defined.

Services

The Systems Group now supports three DEC VAX 11/750 supermini computer systems and a network of eight Sun workstations all with access to AGRENET, the AFRC wide-area computer network, which itself provides a gateway to JANET, the universities' academic network, plus PSS, the British Telecom X25 UK network, and IPSS the International X25 network. These networks give users access to any computer system worldwide that is also connected to one or other of these networks; this allows access to specialized software packages, databases and/or collaboration with research workers elsewhere; it is also a means of load-shedding to other systems, particularly within the AFRC.

This group also supports 24 stand-alone word processor systems, eight data preparation key-to-disc stations, nearly 100 stand-alone microcomputer systems, 110 multi-access terminals, associated serial printers, plotters, line printers and media conversion systems. All these devices are dispersed throughout the campus and the associated serial network, broad band networks and related hardware are installed and supported. This service is a major task of the Department.

Development

With staff again at full complement the Department has been very busy completing delayed projects and starting new ones. Some examples of developments follow.

The image analysis system obtained last year has been augmented with a scanning photo-microdensitometer allowing high resolution data to be extracted from gel images. Software is being developed to couple the systems to general purpose computers for further analysis and graphical investigations. (Bicknell, Ball)

Database packages are once again proving very useful, enabling large or small, and simple or complex data to be held and manipulated with relative ease. The Informix ESQL package has been used on a Torch XXX UNIX system to marshall the Broom's Barn sugar beet data. (Wilson) The Datatrieve package mounted on the DEC VAX systems provided databases for use of Library journals and elemental content of soils. (Thomas) The dBASE III+ package mounted on IBM PC microcomputers has been used for a meteorological database, further major developments of the financial cost forecast package and a new special grants forecast/reporting package. (Verrier)

https://doi.org/10.23637/ERADOC-1-28

The availability of very high quality, low cost laser printers has provoked an investigation into desk top publishing systems which provide typeset-quality camera-ready output. The Framemaker package has been acquired for the Sun systems to provide a suitable environment for the preparation of journal articles, pamphlets and manuals. (Verrier) New non-scientific wordprocessing applications now use Displaywrite 4.

Genstat development has continued this year with version 5.1 containing high quality graphics output utilizing the GHOST and Numerical Algorithms Group (NAG) J06 graphics libraries. Genstat 5.02 has been implemented on the Sun systems under UNIX utilizing the NAG J06 and Sun-GKS graphics packages. Postscript and HPGL drivers are in preparation in order to give the widest possible coverage of equipment for Genstat. (Bicknell) Further minor developments of the Genstat editor were undertaken. (Verrier)

Other graphics activites have included developments of R$TRANPLOT and R$SURF, both being local graphics suites required by users. (Bicknell, Thomas) The need for high quality graphics output for mapping has resulted in the purchase of a Versatec electrostatic colour plotter which when connected to the Sun or VAX systems will allow high resolution images to be produced suitable for publication. (Bicknell)

Network development has been required to cope with the ever increasing demand for high speed data highways between machines and to reduce the burden of cabling inherent in a serial-data star-network. Ethernet provides a 10 Mbit per second broad-band data-highway enabling the direct connection to a common cable of different manufacturers' computing engines spread throughout the campus. In addition, terminals and output devices are connected to the Ethernet indirectly via specialized devices or already connected machines. The Rothamsted Ethernet already connects IBM PC microcomputers, Torch XXX microcomputers, Sperry PC word processors and Sun workstations and associated peripheral devices; the VAX systems should be added next year. The Ethernet, together with a common carrier protocol, allows rapid transfer of data between computer systems, easy access for users and most importantly, the ability to provide a file server. Commercial network software has been under beta-test evaluation, and it is expected that the academic community 'pink book' protocol for terminal traffic will be available early next year. (Moore, LeFevre, Verrier)

The Sun systems provide much needed computer power to small research groups who can now undertake investigations previously prohibitively expensive or impossible. To aid Sun users, the Department has mounted training courses and has provided utilities and converted various software packages, in particular, Genstat, MLP and kriging routines. (Bicknell, Thomas, Ball, Moore, Verrier, Wilson)

The Electronics Group has been much in demand for both development and repair work. Developments include an electroporation unit designed for genetic engineering research (LeFevre, Peck), a smoke detector monitor controller (LeFevre), a solid state rain detector for glasshouse control (Peck, Compton), and network line and line driver/receiver test equipment (Peck).

Staff

R.J. Higgins, Nicole Ford and Christine Alvis left and A.M. Ball, S.B.N. Day and Janet Why joined the Department this year. K. Bicknell is a member of the GHOST Technical Committee and a member of the Eurographics Board and was invited to speak at the NAG Users' Association on the use of NAG graphics within Genstat and gave a presentation at the Genstat User Conference in Pavia.

https://doi.org/10.23637/ERADOC-1-28

STATISTICS DEPARTMENT

Practical applications

The core of our work is collaboration with scientists in other departments, mainly organized through the liaison statistician system. The following are some examples.

The modelling of soil movement in the Market Garden Experiment at Woburn, was completed, using a two-dimensional diffusion model. A paper is being prepared summarizing the modelling procedures, and showing that nearly all the heavy metals present in the sewage sludge applied to this experiment between 1940 and 1961 were still recoverable from the soil in 1984. (Lane with McGrath, Soils)

A series of Genstat programs was written to check, store and analyse the data on flow rate and nitrogen content in the drains under the Brimstone experiment: for the 85-86 and 86-87 seasons. The amount of nitrogen was calculated by regression from voltage measurements. Flow rate and nitrogen content were combined and interpolated, and then integrated to estimate the total amount of nitrate lost from the eight plots under study. There was little difference between the direct-drilled and the ploughed plots in the first season, but a slightly greater loss for ploughed plots in the second. (Lane with Goss, Soils)

The next example shows the difficulties that can arise in sampling and design. An experiment in 1986 seemed to show that a certain insecticide unexpectedly caused population growth and increased reproduction in *M. persicae* at the end of the season. To investigate the effects of the individual chemical constituents of this insecticide a further experiment was incorporated into one that had already begun; a hurriedly produced design confounded some treatment effects with blocks. That an untreated area, not part of the randomization, was being sampled came to light midway through the experiment, vitiating the data already collected. A further complication was that clip cages put onto the plots, each containing three adult aphids of a particular resistance class, were sprayed (or not) in the laboratory beforehand with the appropriate treatment. Only highly resistant aphids survived and reproduced so that intended comparisons between the number of offspring for different resistance classes could not be made. A new sampling scheme was suggested so that the best use could be made of the remaining month of the season. (Clark with Entomology and Nematology)

Work was done on design and analysis of experiments to detect strength of insect attraction to odours, which brought to light some methodological problems in the conduct of such experiments. (Kershaw with members of the Insecticides Department)

Laboratory data were analysed to establish differences in shoot formation and callus production for diploid and tetraploid potatoes. When measurable (some regenerated tissues failed to produce shoots) differences were large. A field trial was then designed to compare 15 derivatives formed by electrical and chemical fusion methods. The chromosome number for each line was determined and the data were analysed to establish the effect on shoot and tuber formation of both the fusion method and chromosome numbers. This has been publicized in both the farming and general press, since it was one of two pioneering experiments done in this country involving the release of 'genetically engineered' organisms on a field scale. (Spechter with Fish, Biochemistry)

Neighbour effects (see Theory, below) can have many causes—notably fertility trends and competition. Several problems have been handled this year, which demonstrate the value to other projects of having special expertise available locally.

In the agroforestry project (see Overseas, below) treatments essentially have a two-factor factorial structure, one factor being tree species and the other intercrop species. Because different tree species can have very different heights, methods of design and analysis of experiments that allow for the effect of the tree species on each neighbouring plot, but not

https://doi.org/10.23637/ERADOC-1-28

for the entire neighbour treatment, need investigation. Uniformity data with various superimposed treatment models and with several randomizations of either neighbour-balanced designs or of unconstrained designs, has been used to investigate empirically the relative merits of classical and neighbour analyses for various tree-neighbour effects. (Langton) Methods of constructing and randomizing designs balanced for tree neighbours have been investigated; for small numbers complete enumeration is possible; for larger numbers the group-generation method previously used by Bailey for quasi-complete Latin squares has proved very useful. Concentration has so far been on one-dimensional designs, partly to find out the possibilities, partly because two-dimensional designs have a much greater cost-differential between plans with, and without, border plots. In October, an urgent request was received from ADAS for a design for a wheat variety × seed-rate trial in which wheat variety was thought to have a neighbour effect. Because of our previous work, a suitable design was found almost immediately. (Bailey with Mr H. Monod)

Experiments were designed to examine the effect that healthy, diseased and missing potato plants had on the yield and tuber size of their neighbours for two different diseases. Each individual plant was used both for yield and as a neighbour of up to six other plants. The design allowed the effects to be estimated of all 45 possible patterns of the first two neighbours on either side of a plant within a furrow, as well as the effect of first neighbours across furrows. A further experiment has been sponsored. (Ainsley with Hide, Plant Pathology)

By counting the number of pixels on a representation of a root screened on the Magiscan, it was hoped that root length could be estimated. This is not the case because each point on the root can generate two adjacent pixels. A simple algorithm was written to adjust for this effect. (White with Verley, Soils)

A scheme to predict the stages of apical development of winter wheat is under development. Observations from the current year are matched with models based on data from previous years. The components of variance due to errors in current observations and to errors in fitting the model were combined to estimate the variance of prediction. (Ross with Travis, Crop Production)

Routine analysis. Problems with the DEC Rainbow machines have continued despite the use of higher grade discs. It is now clear that the basic problem is with the drive mechanism. Discs moved between machines or even between drives on the same machine produce read errors. This negates the intended use of one machine for archiving, and causes delays and frustration. The data processors have had to find alternative ways of moving data between machines. The volume of Rothamsted data analysed has increased slightly this year but turn-round has remained steady at about five days. Less data were entered into Epson machines and transferred onto cassettes; clients seem to be unhappy with the Epsons and are reverting to paper records. More robust machines and new software need investigation (see ADAS, below). If we are to improve methods of collecting data in field experiments at Rothamsted we must invest both time and money to provide flexible and attractive systems.

Work for ADAS

In addition to the statistical work arising from all the MAFF commissioned projects at Rothamsted, the Department also provides a separate service to ADAS R & D. This consists of analysis and advice on the design of experiments, mainly on arable crops but also on, and related to, livestock. This work includes a major commitment to give statistical advice at all stages from initiation to the final report.

https://doi.org/10.23637/ERADOC-1-28

Crop experiments. About 700 analyses were done for Soil Science this year, a decrease of about 20%, partly due to more clients performing their own analyses; the danger of incorrect analyses or interpretations is a matter of some concern. Although the decrease has meant less data-preparation the demand for advice has increased and this trend is likely to continue.

A major task has been the combined analysis from 62 field experiments on milling wheat varieties drawn from experiments done in the period 1977–85. The original experiments had various objectives and covered different rates and timing of nitrogen application. The data were first organized in a consistent way, followed by two basic analyses. In the first analysis exponential curves of constant shape, but with differing location, were fitted to the grain yield data. Only one set of data showed a curvature markedly different from the others; three other sets were so irregular that no reliable curve could be fitted. In the second analysis double exponential curves with constant exponents were fitted to the grain nitrogen data. These gave S-shaped curves that accommodate the well-known decrease in grain nitrogen at low levels of nitrogen application. Only one set of data gave an unsatisfactory fit, one of those which also failed on the yield fitting. Other sets, with large intervals in the nitrogen application rate, had minima in regions with no data and therefore without positive support, but this was immaterial because only the higher levels of grain nitrogen were of interest. From these basic analyses the effect of changes in level and rate of premium payment on the value of the crop were evaluated to determine potential recommendations to farmers. (George with Dr R. Sylvester-Bradley, ADAS)

A smaller study combining results of several experiments was concerned with nitrogen uptake monitored on four soil types in Essex for two successive and contrasting growing seasons. Three fertilizer treatments were applied: 0, 130 and 200 kg ha^{-1} N for winter wheat, and five agronomic properties were determined (grain yield, harvest index, 1000 grain wt, percentage grain protein and percentage fertilizer recovery). Results indicated, as expected, that climate is the most decisive growth factor, with soil type important mainly in special conditions such as drought, when the clay loam and silty loam soils with high field-moisture capacities perform best. (Dyer with Dr W.S. Wilson and Miss M.K. Marsden, Unversity of Essex)

Data entry. Further developments of the FIELD program are described below. The proportion of data sent on paper has decreased steadily and now all data are submitted on discs produced by FIELD; in addition, ADAS clients who formerly used SuperCalc have switched to FIELD. Users of dBASE II can transfer data to FIELD if all units are uniquely labelled but flock records are often hierarchic and progeny are identified only by the maternal label. FIELD cannot accept such data but modifications are in hand to cover records of this type. (George)

Nutritional chemistry. Two sets of data were compared, one from East of Scotland College of Agriculture (ESCA) and the other from Institut National de la Recherche Agronomique (INRA), on the energy value of compound feedstuff for pigs. A joint equation for predicting metabolizable energy in terms of convenient chemical analyses was found, which is being proposed as a working standard for the labelling of feedstuffs for EEC regulations. (George with Stranks, ADAS)

This is one example of work in nutrition chemistry which, after the difficulties reported in 1986, we can once again handle, having filled the vacant post concerned. It has been possible to take on new projects for the ADAS Feed Evaluation Unit and to analyse two of last year's projects. (Coleman with Mr D.I. Givens, ADAS)

Livestock experiments. Livestock work has been handled by Spechter (poultry, pigs and some dairy work); Gina Smith (sheep and beef cattle), Coleman (beef cattle) and George

https://doi.org/10.23637/ERADOC-1-28

(dairy hygiene, animal welfare, rabbit control and deer farming). One example must suffice. Pressures from the animal welfare lobby, pending legislation and new Ministry guidelines, all strongly support the search for alternatives to the battery cage system. The aviary and perchery developed at Gleadthorpe Experimental Husbandry Farm provide feasible alternative systems despite many practical problems. Further studies were done in 1987 to evaluate the effects of stocking density and other factors interacting with these systems. A third system—the Elson tiered terrace system, developed chiefly by A. Elson, (ADAS), is being tested. This system aims to retain the advantages of battery cages while allowing the birds greater freedom of movement. The Elson system is part of an international project run in collaboration with the Institut für Kleintierzucht, Celle, Federal Republic of Germany. Statistical advice was given on the design and evaluation of the prototype and help with the design of a behavioural study to evaluate the use of the prototype. (Spechter with Mr A. Elson, Dr A.J. Ballantyne, Dr J. A. Hill, ADAS)

Surveys

Fertilizer practice. Part 2 of this report contains a summary of the findings of this year's Survey of Fertilizer Practice. (Elsmere) Surveys of fertilizer practice done in England and Wales, and in Scotland during 1986 were reported. These annual surveys are done in collaboration with the Fertiliser Manufacturers' Association (FMA) and ADAS Soil Scientists. In 1987 the number of farms sampled was 1165 in England and Wales and 256 in Scotland (1170 and 254 respectively in 1986). Field work was done from June to August by Farm Research Ltd on behalf of the FMA. As usual, a one-page report was sent to all cooperating farmers, and a full report to those who requested one. In addition to the usual provision of an assistant for the coding of the survey forms the whole of the data entry was taken over by the data processing section last year. (Kershaw and Leech)

Sugar beet surveys. The 1986 data tape was received from the British Sugar Corporation and translated for use on the VAX. As well as the usual analysis, grid references and sample yields were extracted and fed into the program SURFACE II to produce contour maps of yield distribution. Further data on virus yellows and powdery mildew were extracted. (Leech with Dewar and Asher, Broom's Barn)

The relationships between sugar beet yields, growers' practices and other factors recorded in survey data have been examined and series of experiments analysed and interpreted. A paper has been written on results of trials assessing a technique for comparing seed drills. (Mr B.M. Church for the British Sugar Corporation)

Representative soil sampling. The representative soil sampling scheme continued again this year with just over half the sample being resampled from fields visited in 1981, the other half being entirely new. This scheme of resampling fields at five-year intervals will be continued on an annual basis. (Elsmere and Kershaw)

Straw incorporation. In 1984 the National Farmers' Union (NFU) in collaboration with ADAS and the AFRC, set up a register of farmers who had already incorporated cereal straw or would shortly do so. Questionnaires were sent to these farmers in 1984, 1985 and 1986. Results from the 1986 data were presented at an annual NFU conference for members of the register. (Kershaw)

Theory

The development of statistical theory is a small but very important part of the work of the Department. It underpins both our collaborative work, which often generates novel statistical

https://doi.org/10.23637/ERADOC-1-28

problems, and our development of general-purpose statistical software, which itself requires the generalization of methodology. This component of our work ensures a lively statistical ambience which sustains the quality of all our work. It plays a vital role in our ability to recruit and hold good-quality staff even with a national shortage of statisticians and with the uncertainties of agricultural research funding. Rooted as it is in practical requirements, our theoretical work is immediately useful and rarely of a very abstract nature. Progress has been made in the following areas during 1987.

Analysis and design of experiments. A New Initiative project has been started to assess the effectiveness of models that assume correlation between adjacent field-plots. Proposals to base the analysis of experiments on neighbour models, as they are termed, go back to 1934 and recently the idea has re-emerged. Some specific problems where neighbour effects are of interest were discussed above. Initial theoretical work done at Rothamsted in 1985 gave little encouragement to the idea that neighbour analyses, of which there are several variants, are better than those based on classical models but work done elsewhere, principally with variety trials, has suggested otherwise. It has become necessary to examine the issues thoroughly. Work started with an investigation of the use of time-series and variance-distance (semivariogram) methods for specifying variance patterns in a set of long-term Rothamsted experiments. The estimation and use of semivariograms play a central role in neighbour methods. Theoretical and simulation work on the expected moments of various statistics for estimating semivariograms from the results of field experiments has shown that using residuals to estimate semivariograms can lead to systematic bias in the results. Non-standard use of the time-series part of Genstat 5 is being investigated for fitting a particular neighbour model (exponential variance) that has not previously been fitted to individual trial results. (Ainsley and Bailey)

Previous work of Bailey had constructed sets of quasi-complete Latin squares for all odd numbers of treatments less than 20 except 15. In such squares, every treatment is next to every other treatment exactly twice in the east-west direction and exactly twice in the north-south direction also. These sets are valid in the sense that we may use proper randomization as a substitute for exact knowledge of the fertility patterns in the field. Attempts were made to extend this work to even numbers of treatments in the range 4–20. A construction was found which gives large numbers of quasi-complete Latin squares. However, detailed and extensive results on the randomization of these squares suggest that, among the squares so constructed, there is no subset which forms a valid randomization set for any method of analysis. (Bailey with Prof. C.E. Praeger, University of Western Australia)

Non-linear modelling. Non-linear models for fitting relationships among two, or more, variables and for fitting a variety of probability distributions are continually arising in the work of the Station and indeed throughout the AFRC. Both the numerical processes of fitting such models and the problems of statistical interpretation generate major difficulties. Few centres, internationally, have more expertise and experience than Rothamsted and it is gratifying that a book *Non-linear estimation* embodying this knowledge has been accepted for publication. (Ross) Detailed progress is as follows:

Parallel model analysis. Parallel model analysis is used to compare the fits of similar non-linear models to several sets of data such as measurements of plants over time on different plots of a field experiment, or responses to doses of different insecticides. When there is a single classification, such as types of insecticide, the analysis is conceptually straightforward (although the computations may be lengthy); some parameters take common values while others are separately estimated for each data set. Parallel model analysis has been generalized

https://doi.org/10.23637/ERADOC-1-28

to take into account the effects of nested and crossed factors. If the factors are nested the extension is straightforward, but if the factors are crossed the analysis resembles a non-orthogonal analysis of variance. The common model is usually of special interest. (Ross)

Reparameterization of mixture distributions. Long-tailed distributions (LTD) are characterized by a preponderance of low valued observations with occasional large observations. They occur in many contexts, from the width of cracks in soils to the survival times of organisms. The large observations may be the most important; in statistical terms we must estimate the tail probabilities of the distributions. Sometimes the data might be mixed, a proportion obeying a short-tailed distribution (STD), the remainder an LTD. Mixtures are difficult to fit, because the smaller observations in the LTD are indistinguishable from those in the STD. Complex mixture models were fitted by reparameterization, using expected moments and percentiles. The least precise estimate was of the proportion of each component in the mixture; estimated tail probabilities were very sensitive to choice of model. With large samples, tail values might be fitted better by using a truncated distribution. (Ross)

Multivariate analysis. Multivariate analysis is concerned with the simultaneous analysis of data observed on several variables. Classical multivariate analysis, largely developed in the 1930s, generalized univariate methods, replacing variance by a matrix of variances and covariances or correlations and by developing a panoply of significance tests based on unrealistic assumptions. Over the past twenty years more practically useful methods of multivariate data-analysis have been developed in which the concept of distance (or dissimiarlity) plays a central role.

The possibilities latent in the exploitation of the concept of distance have still not been fully recognized. In 1981 Digby and Gower showed how, by appealing to distance concepts, a classical method known as canonical variate analysis, which contrasts variation between and within groups, could be generalized in certain respects that required fewer assumptions. This was the first step towards replacing a multivariate analysis of variance by an analysis of distance. Progress has been made in generalizing the idea to handle more complex sampling structures. (Digby)

Multivariate data analysis appeals strongly to graphical methods of display. One such method, the biplot, represents a multivariate sample by a set of points and the variables by concurrent straight lines. Biplots, which display the information in a multivariate sample, are a useful generalization of the idea of scatter plots among two variables. A disadvantage is that a biplot is implicitly restricted to one form of distance; the absence of a similar method using other ways of measuring distance has long been felt. A method of non-linear biplots has now been developed which is based on a simple idea that promises to be useful in a much wider context. (Gower and Harding)

Mathematical ecology. Ecological investigations have long been a part of the work of Rothamsted. They are now of increasing importance, particularly in the study of the impact of farming practice on the environment. The collection of reliable data and its interpretation has generated interests in population dynamics, community studies and fundamental problems of how to measure ecological phenomena. The Statistics Department is currently active in all these areas. Animals, unlike plants, move. This dynamic element leads to complications. A new index of aggregation was investigated, suitable for data concerning counts of (invertebrate) animals in sample units, where the spatial relationship of the units may or may not be defined. The index is a function of (i) the number of movements between spatial (or sample) units required for a given sampled population to achieve complete aggregation (all individuals crowded into a single sample unit) and (ii) the number required to achieve

https://doi.org/10.23637/ERADOC-1-28

randomness. For simplicity randomness is defined as the condition when sample variance equals sample mean, although formally this is only necessary, not sufficient. Investigations of the index and some algorithms for its computation are proceeding. (Perry)

Statistical computing

Our longstanding involvement in statistical computing (see Gower, *Rothamsted Report for 1985*, 221–235) is rooted in the need for a flexible, reliable and convenient means for doing all the very varied kinds of statistical analysis arising in our daily work. What was originally produced for our own needs also fills the needs of others, and our major product, Genstat, is marketed internationally. Genstat is the only major general-purpose statistical system not of United States origin. The commercial side of this undertaking is handled by the Numerical Algorithms Group (NAG), Oxford who deal with contracts, marketing and, since 1985, with the organization of the conversion of the Rothamsted computer code to run on a large range of computers, although we still handle conversions for VAX, Sun and IBM/PC computers. The chief events of the year are the commercial release of Genstat 5, a major revision of the system (see the General Report, p. 15) with documentation including a 777 page Reference Manual and the negotiation by NAG of a contract with Honeywell, whereby Genstat 5 becomes the statistical component of the applications software for Honeywell computers. More details of this and other computing work are given below.

Genstat 4.04B. Eleven more copies of the VAX version were sent out; the total has now reached 123. (Simpson)

Genstat 5. The main work this year has been in validating the code and in completing the documentation. We are grateful for the help of our colleagues both at Rothamsted and at other AFRC sites, many of whom are now using Genstat 5 routinely. This field-testing in many different types of work has allowed us to correct many bugs that arise only in unusual circumstances, making us confident of the version now released commercially by NAG. The comments of the users of these test versions have been overwhelmingly favourable, so we trust that Genstat 5 will lead to a further increase in the number of Genstat installations around the world. (Ainsley, Bicknell, Digby, Harding, Lane, Leech, Payne, Simpson, Todd, Verrier and White, with Dr G. Tunnicliffe Wilson, Lancaster University) The new version was put through many stringent checks on programming standards, particularly by Mr R. Iles on the Apollo at NAG and by Mr A. Mann on the Prime at Scottish Agricultural Statistics Service (SASS). These showed up many transgressions of the Fortran standard that were acceptable to VAX computers. (Simpson)

Courses. The new version of Genstat has led to a demand for courses of various kinds. Introductory courses have been held at Agriculture Canada (Charlottetown), South African Department of Agricultural and Water Management (Pretoria), Institute of Food Research (Norwich and Reading), Institute for Grassland and Animal Production (Aberystwyth, Hurley and North Wyke), Institute of Horticultural Research (Littlehampton), Institute of Arable Crops Research (Rothamsted and Long Ashton). Courses to assist users to convert from Genstat 4 to Genstat 5, have been held for the Netherlands agricultural statisticians (Wageningen), and for the Welsh Plant Breeding Station. A course has also been designed to teach the non-statistical use of Genstat for graphics and calculation; this was first given at AFRCCC in November. (Bicknell, Digby, Harding, Lane, Payne, Simpson, White with Arnold LARS, Kirby IFRR, Patefield IGAP, Fenlon and Townley IHR and Franklin IFRN)

https://doi.org/10.23637/ERADOC-1-28 *pp 12*

Documentation. The Genstat 5 Reference manual and a Reference Summary, giving concise details of the syntax have been published. (Ainsley, Bicknell, Digby, Gower, Harding, Lane, Leech, Payne, Simpson, Todd, Verrier, White with Dr G. Tunnicliffe Wilson, Lancaster University and Dr L.J. Paterson, Heriot-Watt University) *Genstat 5: an introduction*, has been written in the same form as the introductory book for Genstat 4 and publication by Oxford University Press is expected in January 1988 (Lane with Mr N. Alvey and Dr N. Galwey, Cambridge University). A new book, *Genstat 5: a second course*, is nearing completion (Lane and Digby with Dr N. Galwey, Cambridge University) The Conversion Manual to support those who convert Genstat 5 to run on different ranges of computer was revised. (Simpson with Mr A. Stalewski, NAG)

The *Genstat Newsletter*, published twice a year, is a useful vehicle for communication between Genstat users. It contains news, information on current developments and technical articles, often of high quality, on new uses of Genstat. It has two editors: Mr K.I. Trinder of NAG, and P.W. Lane who took over from R.W. Payne in June. The 1987 issues include articles by Ainsley, Digby, Harding, Lane, Payne and Simpson.

Extending Genstat. One of the strengths of Genstat is its powerful high-level computing language which allows new methods to be programmed as *procedures*. The Procedure Library is overseen by an editorial committee with Payne as chairman and Digby as secretary and is to be upgraded every six months; an initial version was prepared for the first release of Genstat 5. The Library is self-documenting, and four procedures have been written to allow users to obtain information about the Library and about procedures from within Genstat itself. (Payne)

The procedure facility is but one of three ways of extending Genstat. Another method allows users to write their own Fortran code and embed this in Genstat either using the OWN facility or designing their own commands. The third method uses the PASS facility which allows an external program to be run from within Genstat. (Lane)

Genstat, shorn of its statistical facilities and coupled with the OWN and PASS mechanisms, produces a general environment suitable as an interface for general-purpose subprogram libraries. This potentially powerful use of Genstat is being investigated. (Lane with Mr R. Iles & Dr J.A. Nelder of NAG)

MLP. This program is also marketed by NAG; 39 licences were current in November 1987. A new Manual has been written and published. A version for the Sun workstation was produced. Facilities to aid interpretation and to take advantage of technological advances have been developed. (Ross)

At the core of MLP are programs to find the optimum of a numerical valued function. Methods of combining Newton with conjugate gradient methods of optimization were studied. It was found that by updating the inverse Hessian and using gradient information only, substantial savings could be made, but that it was advisable to recompute the Hessian completely from time to time, to avoid accumulated errors. (Ross)

Optimization of non linear functions might be improved by searching along suitable curves rather than straight lines. The positions after successive iterations may indicate a curved valley, which may be generated as a function of a single parameter. Polynomial functions are the easiest to use but do not always give good results as they are not invariant to changes in scale. Power functions have been investigated as possible alternatives, but these may be used only when the angle beween two successive vectors is small. This strategy is at worst harmless and at best quite useful. (Ross)

The work on expert systems in non-linear modelling continues but progress depends on software developments. (Ross with Prof. C. Berzuini, University of Pavia)

https://doi.org/10.23637/ERADOC-1-28

CLASP. An interactive version of CLASP was prepared and directives added, to assist in file handling and data modification. Methods used in CLASP for hierarchical clustering were compared with recently proposed procedures. These were slightly faster for large data sets, but required extra work to obtain dendrograms. It is therefore doubtful if CLASP needs revision, at least in this respect. (Ross)

Genkey. A start was made on preparing a Fortran 77 version, but major restructuring will be required to exploit the character-handling facilities now available; this will however simplify future conversions to other machine ranges. New sites during the year included Kew, University of Pavia and INRA, Montpellier. (Payne)

Rothamsted General Survey Program (RGSP). As stated in the Report for 1986, it was decided to upgrade the survey program. Progress has been good and the program has been converted to Fortran 77 and runs on the VAX, Sun workstation and on the IBM/PC (or compatible machines). The standard data reading facilities have been enhanced so a much wider range of surveys can be analysed without having to supply a Fortran execution program. An interactive program allows users to set up RGSP instruction files without awareness of the detailed syntax. Documentation is in preparation for a new release of the program in 1988. (Kershaw)

FIELD. The FIELD data-entry program has been extended by a comprehensive HELP facility and an option for tabulations of various summary statistics classified by up to four factors. Data entry from disc files has been modified to allow many variables in parallel, either in unit or in random order. The raw data files can be produced by transfer from Epson or Hunter machines using programs for initial data-entry written by ADAS staff. Versions of the FIELD program can now be run on a variety of microcomputers under CP/M and MS-DOS operating systems. (George)

Overseas

The Overseas Development Administration (ODA) continued to fund the Biometric Unit. J. Riley and S. Langton visited Kenya to attend a Workshop on Intercropping and visit other institutes in the country. J. Riley visited Zambia and Thailand to assess ODA-funded projects; she visited India to give statistical advice on an ODA tree crops project. S. Langton visited Brazil to initiate collaboration in agroforestry research.

The amount of annual crop and animal work has remained steady, coming from the following countries: Bolivia, Botswana, Dominican Republic, Falkland Islands, Indonesia, Kenya, Nepal, Sierra Leone, Solomon Islands, Sudan, Tanzania, Thailand, Yemen Arab Republic, Zanzibar and Zimbabwe. (Riley, Ryder, Prosser, Sehdev and Poultney)

Requests for statistical advice and analyses for perennial crop data have continued to arrive in the Unit; these requests came from Bolivia, Dominica, Ecuador, Indonesia, Kiribati, Nepal, Papua New Guinea, Solomon Islands, Tanzania and Thailand. (Riley, Ryder, Prosser, Sehdev and Poultney)

The number of personal visits to the Unit has increased remarkably. The number of overseas staff visiting the Unit for short periods of training has also increased, with overseas workers coming from Sri Lanka, Botswana and Nepal. (Ryder and Riley)

The survey of statistical software for microcomputers continues and testing has been greatly helped by the Unit's new Apricot Xen microcomputer. (Ryder)

The intercropping research project has continued with increasing collaboration with overseas countries. Datasets have been collected from a total of 11 countries on both perennial and

https://doi.org/10.23637/ERADOC-1-28

annual crops; examination of features of these data and the suitability of analysis by available techniques is in progress. (Langton and Riley)

Staff and visitors

K.W. Coleman, Suzanne Welham and Rosemary Poultney were appointed during the year; Ethna Prosser and Seema Sehdev worked for six months as sandwich students. Professor A.A. Greenfield spent three months with us as an adviser on consulting.

In August J.C. Gower was an invited speaker at a conference, organized in Seoul by the Korea University and the Ohio State University, on Recent Developments in Statistics and their Applications; he also visited the Korean Rural Development Administration. He then went to Tokyo for the International Statistical Institute and was an invited discussant at the First International Association of Statistical Computing World Conference on Computational Statistics and Data Analysis at Shizuoka, Japan. He was also an invited speaker at the 5th Symposium on Data Analysis and Informatics in Versailles and the First Conference of the International Federation of Classification Societies in Aachen; the latter was also attended by R.W. Payne. Bailey spent three months visiting the Universities of Sydney and Western Australia, CSIRO Canberra and The Ministry of Agriculture and Fisheries in Christchurch, New Zealand. P.G.N. Digby attended the ZUMA conference in Heidelberg as the Numerical Algorithms Group representative; he also visited the University of Cape Town and other sites in South Africa. P.W. Lane gave a presentation on Genstat at the Dutch Agricultural Statisticians Conference in Wageningen. He was seconded to NAG, Chicago for three months to act as Genstat Consultant and attended the Americal Statistical Association conference in San Francisco. He also visited the Université Paul Sabatier in Toulouse to give a seminar and to demonstrate Genstat 5. G.J.S. Ross gave a series of talks on non-linear modelling at the Institut National de Recherche en Informatique et en Automatique/Centre International de Mathématiques Pures et Appliquées (INRIA/CIMPA) summer school on Data Analysis Methods and Software at Nice, France. J.C. Gower gave a series of lectures at the first European Course on Advanced Statistics (Multidimensional Data Analysis) held in Capri. R.W. Payne spent three weeks with INRA in Montpellier to initiate a collaborative project on the identification of varieties of grape vine. The Genstat Conference in Pavia was attended by A.E. Ainsley, P.G.N. Digby, S.A. Harding, P.W. Lane, R.W. Payne, K. Ryder, G.J.S. Ross, H.R. Simpson, and R.P. White, who each gave talks and/or demonstrated Genstat 5.

J.C. Gower gave an invited paper on Statistics and Agriculture at the Royal Statistical Society Charter Centenary Conference also attended by R.A. Bailey, P.G.N. Digby, B.J. George, P.W. Lane, P.K. Leech, R.W. Payne, J.N. Perry and H.H. Spechter. He was also an invited speaker at an International Symposium 'Prospects in Systematics' organized by the Systematics Association in London to mark the Golden Jubilee of the Association and Huxley's influential book *The new systematics* (1940). R.A. Bailey read an invited paper at and A.E. Ainsley, J.C. Gower, P.K. Leech, J.N. Perry, J. Riley and K. Ryder attended the meeting in Edinburgh arranged by the Biometric Society in honour of Professor David Finney. Harding attended the British Computer Society conference on the future of graphics software. Members of the Department have as usual been active in presenting papers at meetings of learned societies and regional conferences.

Pavel Krajewski from the Institute of Plant Genetics, returned to Poznan in September and Dr Teshuishi Miwa from the National Institute of Agro-Environmental Sciences, returned to Ibaraki, Japan in August, Hervé Monod from INRA Versailles arrived in April and Guiseppe Bove from the University of Rome spent two periods in the Department in July and from October to December.

https://doi.org/10.23637/ERADOC-1-28