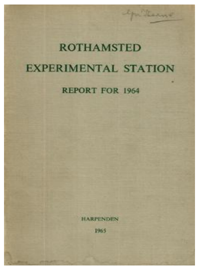
ROTHAMSTED RESEARCH

Report for 1964

Full Table of Content



Statistics Department

F. Yates

F. Yates (1965) *Statistics Department ;* Report For 1964, pp 203 - 211 **- DOI: https://doi.org/10.23637/ERADOC-1-57**

STATISTICS DEPARTMENT

F. YATES

R. T. Clarke, A. Frater, Molly Dyke (née Fretton) and Winifred A. Collingbourne (née Johnson) left, and Judith M. Stone and J. Rosemary Rossiter were appointed to the staff. A. G. Davies transferred from the staff of the department to International Computers and Tabulators. B. M. Church and A. J. Vernon were away on secondment during the whole year.

F. Yates spent three weeks at the Bell Telephone Laboratories, New Jersey, and was a member of a committee called by the World Health Organisation to advise on the epidemiological aspects of the proposed World Health Research Centre. F. B. Leech spent two months in Argentina as Consultant to the Government advising on farm animal disease surveys. The visit was arranged by the Department of Technical Co-operation under the British Technical Assistance Scheme.

Professor W. G. Cochran, a member of the Department before the war and now Professor of Statistics at Harvard University, is spending part of his sabbatical year in the department. A. J. B. Anderson, an Agricultural Research Council scholar, is spending a year in the department studying computation methods. Six other workers, two from overseas, spent various periods in the department during the year.

General

The department was very fully occupied in 1964 organising and programming the new Orion computer. The Orion is a much larger and more complex system than our previous computers and gives correspondingly greater opportunities for the rapid and effective analysis of the large amounts of numerical data that are collected in much agricultural and biological research. Processing data on a computer, however large and complex, does not, however, automatically provide meaningful results. As we said in a recent paper (12.20): "Computers are good servants but bad masters. There has been plenty of statistical nonsense produced on desk calculators, but this will be nothing compared with the flood that will emerge from computers if they are not wisely used and firmly controlled."

Apart from restraining the starry-eyed, our main tasks on the Orion are to run an efficient service, and to provide a good set of versatile general programmes which will enable statisticians and other research workers at Rothamsted and at other institutes to get statistical analyses done quickly and effectively, including many previously impossible or exceedingly tedious; also, to make available a large computer system for statisticians at other institutes to use directly in developing their own methods. In so far as we succeed in these tasks, we shall make a useful contribution to improving the effectiveness and speed of much agricultural research.

https://doi.org/10.23637/ERADOC-1-57

Among the statistical analyses completed by members of the Department in 1964, the following are of most direct agricultural interest: summaries of experiments on nitrogenous manuring of cereals (12.11, 12.12), and of 5 rotation experiments on the residual effects of potash (12.3); a report on the merits of Charolais bulls as sires of beef cattle in dairy herds (12.16); the results of surveys of potato-growing practices (12.15) and of the incidence of mastitis (12.10); and an investigation of the plasma calcium and phosphorus concentration in dairy cows (12.14).

The Computers

The Orion computer was delivered in December 1963 and passed its site acceptance test on 12 March 1964. We had considerable assistance during this period from Mr. S. Iles (H.M. Treasury Technical Support Unit). The terms of the acceptance test had been settled by Ferranti and differed from those normally used by the Technical Support Unit for Government computers. The problems of supervising acceptance tests usually arise from argument on the interpretation of incidents and their influence on the progress of the tests. Here Mr. Iles' experience was invaluable, particularly during the factory tests. The final site test went smoothly. Table 1 summarises the performance of the machine. These

TABLE 1

Orion performance, 1 *April*–31 *December*

Number of working days	187
	Hours
Scheduled work periods	1,216
Time lost due to faults during scheduled periods	235
Available time during scheduled periods	981
Additional late running	145
Time used (including 21 hours idle time)	1,126

The scheduled work period is from 11.0 a.m. to 5.30 p.m. So far as possible any time lost during the day because of machine faults is made up by late running.

figures overestimate the performance because the time taken to restart a programme after machine failure was not accurately assessed during the first few months, and also because no account was taken of faulty peripherals which could, for example, prevent time-sharing on jobs. However, some allowance should perhaps be made for the fact that the maintenance team has changed very frequently (we have had six second engineers).

The performance has unfortunately not shown steady improvement. For about four months after the acceptance test the machine performed well enough to cope with the work load, but from the middle of August to early November we were plagued with far too many incidents. This period accounts for 63% of the whole time lost over the nine months Eventually I.C.T. arranged for specialist engineers to work on site, and their patient examination restored the performance to something like its original level. It remains to be seen whether this improvement can be maintained; on most occasions the machine was in a serviceable state when switched off, which suggests that with longer running there will be improved performance. As the load on the Orion increases we must certainly look for improvement.

https://doi.org/10.23637/ERADOC-1-57

The actual time used (1,126 hours) was 93% of that nominally available during scheduled periods. No subdivision between time used for productive work and time used for programme development was made, but this is now being done. The proportion of time required for programme development can be expected to decrease as more general programmes become available; but there will be a continuing demand for development of programmes written in autocode for special jobs. The demands on the machine for productive work will undoubtedly increase rapidly as programmes become available—no survey analyses, for example, have yet been made on the Orion—so it seems certain that more hours will soon have to be worked, and arrangements are being made for evening work without engineers. This should suffice for the next few months; later it will probably be necessary to arrange two-shift working.

The Atlas and the Orion were among the first computers to adopt time-sharing, and the designers did much pioneering work in this development. Time-sharing is a technique by which two or more programmes can be run simultaneously, thus making more economic use of the central computer and peripheral equipment. Doubts have been expressed on the real value of this technique, and we have made efforts to assess its value on the Orion. It is difficult to do this, but our experience and that of other Orion users indicates that in a typical day's run the work would have taken 40–50% longer to complete had the jobs been run singly. A further gain from time-sharing is that we can run a small base-load programme with low priority which tests many functions of the machine and thus gives assurance that performance is satisfactory. Against this must be set the additional storage required; one-quarter of the drum store and one-sixteenth of the core store is occupied by the time-sharing and associated programmes.

The Orion proved slower than expected on many types of job, and we suspected that the times given for the basic operations in the specification were under-estimated. The actual times are difficult to measure, as for many operations they are variable, but Mr. G. Felton, I.C.T., has agreed that, for typical data-handling work, the stated times should be increased by about 40%. I.C.T. are now investigating whether these times can be reduced.

As we expected, there is an imbalance between the output punches and the printing equipment. At present all results are punched on one of two 7-track tape punches, and the resultant tapes are printed on a group of four Flexowriters. The increasing amount of productive work, particularly the analysis of experiments done in batches, produces long rolls of paper tape that take hours to print and require constant attention. To overcome this imbalance, the installation of a line-printer is now needed. This mechanism takes results directly from the machine and prints a line at a time at speeds of up to 1,000 lines a minute.

The 401 and 402 continued to give satisfactory service, but the small volume of survey work has lessened demand for the 402, particularly during the latter part of the year. The 401 is still in use during the greater part of each day. In all, the machines were used for production for 2,580 hours, compared with 3,528 hours in 1963. Both machines will be

https://doi.org/10.23637/ERADOC-1-57

disposed of in March or April 1965. It is unlikely that any further use will be found for the 401, which has no repertoire of general programmes, but the 402 will be given to Watford Technical College.

Before A. G. Davies joined I.C.T. he converted all our old 5-track tape equipment from the 401 code to the Orion code, but we have a conversion programme which enables the Orion to convert tapes in Orion code to 401 code. Having the 401 and 402 greatly eased the transition period and has given us time to develop really satisfactory programmes for Orion which take full advantage of its greater potentialities. Our computer operators did an excellent job in quickly learning to handle the Orion, and in keeping all three computers running simultaneously.

During the early months we were able to offer testing facilities on the Orion to programmers from Beecham Ltd., pending delivery of their own Orion. This resulted in a useful association, and Beechams gave us reciprocal aid when we were experiencing difficulties with our Orion later in the year.

Programme and Systems Development

The development of programmes and systems of operation for the Orion has, as expected, occupied much of our time. Although progress on our general statistical programmes has not been as fast as we had hoped, D. H. Rees has evolved a good operating system and the flow of work through the machine is excellent—moderate-sized jobs from outside users who have prepared their own tapes are usually cleared within the day.

Extended Mercury Autocode. Extended Mercury Autocode (EMA) was provided by I.C.T. in March and has been extensively used since. As is usual with new compilers, various errors were detected. Apart from correcting these (done by I.C.T.), we have been able to influence further development on this compiler. The original compiler was not fast, but only required 3,568 words of core store, and consequently by time-sharing two compiling jobs could be run on the machine at the same time. A faster compiler has now been written by I.C.T., using 5,488 words of core store. This shortens the time required in compiling by about 40%, though time-sharing on compiling jobs is no longer possible with the present core store. Other jobs not requiring much core store can still be time-shared with the compiler, and the new compiler is expected to speed compiling work. We thank Mr. Peter Hunt, I.C.T., for allowing Mr. David Foster, the I.C.T. EMA compiler writer, to undertake these developments. Mr. Foster has done much of his development work on the Rothamsted Orion, and the personal contacts that have resulted have been most helpful.

The existence of an effective autocode has proved immensely valuable. Without it our programming for the Orion would be in a critical state. It is of value not only for writing *ad hoc* programmes for special jobs but also for writing general programmes designed for repeated use, such as those for the analysis of standard types of experiment. No exact comparisons have been made, but the compiler seems to construct an efficient machine language programme.

Although the EMA language has been criticised, and suffers from the

https://doi.org/10.23637/ERADOC-1-57

historical fact that in its original form it was designed specifically for the Mercury computer, the extended language being made compatible with the original language (probably a mistake), we find it a simple and relatively elegant language to use, comparable with Fortran and greatly superior in this respect to Algol. The assembly and monitoring facilities of EMA are excellent, and locating and correcting errors in a programme rarely presents any difficulty, except for the barbarous procedure required to correct programme tapes. To overcome this we have written a correction programme that incorporates corrections in a programme tape and produces an amended tape, but this is not a wholly satisfactory procedure. It would be better to store a programme under development on magnetic tape and amend the magnetic-tape version as required. This development is now being discussed with I.C.T.

The existence of a good autocode has enabled statisticians at other research institutes to make direct use of the machine. The National Vegetable Research Station and East Malling Research Station, in particular, do much of their own programming and tape preparation and co-operate with us in the writing of general programmes; various other institutes also write programmes for special jobs. Exact records of the distribution of work have not been kept, but Table 2 shows the distribution of runs by institutes for November and December. Most of the runs assigned to other institutes used programmes written by these institutes.

TABLE 2

Distribution of runs by institutes, November–December

Institute	Runs No.	Runs %
Rothamsted Experimental Station *	997	61
National Vegetable Research Station	328	20
East Malling Research Station	109	7
Others †	194	12
Total	1,628	

* Including all experiment analyses using Rothamsted programmes. These are batched up irrespective of source.

† This includes:

Animal Breeding Research Organisation	John Innes Institute
Glasshouse Crops Research Institute	Milk Marketing Board
Grassland Research Institute	National Institute of Agricultural Botany
I.C.T. (for EMA Development)	National Institute of Agricultural Engineering

From 1 April to 31 December 7,653 runs were made. Of these, 4,333 (57%) were EMA runs, i.e., runs requiring the EMA compiler. These may be tests of programmes being developed, or "load and go" production runs, i.e., production runs for which the programme is compiled at the time it is used. The remainder consist of runs using previously compiled programmes, and programmes written in machine code (mostly test runs of our survey and experiments programme).

Other languages. The other two languages promised for Orion by I.C.T. were Symbolic, which is essentially an extension of Basic (the ordinary

https://doi.org/10.23637/ERADOC-1-57

machine code) and allows quantities to be referred to by symbols instead of by store addresses, and has certain other conveniences; also Nebula, which is a commercial programming language. Symbolic has just become available and will doubtless be used in preference to Basic for future work; Nebula is not likely to be much used by us, and requires a larger machine than ours for compiling. A Fortran compiler has been written for the Orion by the Harwell group. This also requires a larger machine for compiling, but the possibility of compiling already developed Fortran programmes at Harwell may be of some use.

Survey and experiments programme. Progress on our Survey and Experiments Programme (SEP) was slower than we had hoped, but the separate sections are now all tested, and a usable version, at least for survey analysis, should be finished soon. The slow progress is partly explained by the lack of a computer for testing in the early stages, partly by difficulties in organising an emendation procedure suitable for so large a programme and partly by underestimation of the time required for linking all the parts of a complex autocode into a coherent whole. (Gower, Simpson, Martin, and Ross)

Programmes for the analysis of experiments. Three general programmes have been written in EMA. Simplex analyses randomised blocks and Latin squares containing one or two factors; one of the factors may be on split plots. (Healy) The 2^n programme analyses 2^n designs of all types, and deals effectively with partial confounding, fractional replication, and factors with four or more levels. (Yates) The 3^3 single replicate includes 3^3 designs with extra treatments; it also has the novel feature of providing summaries of sets of experiments and permits partial summaries to be amalgamated into larger summaries. (Yates)

All these programmes use the same input programme; this is similar to the one we developed for the 401, and provides for the preliminary computations on the data that are almost always required. (Healy) A similar programme for the input of data recorded plot by plot instead of variate by variate was also written. (Healy) The programmes make provision for missing observations, which are simply indicated by *, and for full covariance analyses. We also endeavoured to standardise and simplify the conventions governing the specification of the design and type of analysis required, both to simplify work within the department and in the expectation that many workers outside the department will in future write their own specifications. For the first time we can give really informative and adequate headings to tables, etc., and indicate clearly the nature of any errors detected by the computer. This is a great help to those using the results and also saves much annotation of results by departmental staff.

Although constructing these programmes took some time, they should serve as models for constructing further programmes for other types of design. Experience shows that parts of a programme written in autocode can be readily modified for use in other programmes. Thus, the variate control and missing plot sections of the 2^n programme were used to

https://doi.org/10.23637/ERADOC-1-57

construct the parallel sections in the 3^3 programme. This procedure is not possible when writing in machine code. Instead it is customary to write formal routines which it is hoped will serve similar purposes later. As the introduction of excessive generality into routines complicates their use, and as future needs are usually not fully apprehended, this procedure is by no means ideal.

Latent roots and vectors. This programme was written to solve the matrix equation $(\mathbf{A} - \lambda\mathbf{B})\mathbf{x} = \mathbf{O}$. It was incorporated in various classification programmes, and a programme for calculating multiple discriminants.

Scanning facilities for EMA. Gower implemented his earlier suggestion (Comp. J. (1962), **4**, 280–286) of making the scanning facilities for multiway tables, which form a basic part of the 401 General Survey Programme and of SEP, available in EMA. These facilities have proved of great value in constructing experiment analysis programmes and the EMA Survey Programme.

EMA Survey Programme. This was written to provide a survey analysis programme for the Orion of the same general type as the 401 General Survey Programme, pending completion of our proper Survey and Experiments Programme. (Yates, Leech and Anderson) It was written in Autocode and the user has to write his own autocode instructions for the preliminary computations on the observed variates (derived variate computations). When the Survey and Experiments Programme is completed the EMA programme may well be scrapped; if not, it will require extension to make it more convenient for general use. Comparing the speed of this programme with SEP will be interesting, as it differs from SEP in two fundamental respects: the derived variate instructions are compiled once for all instead of being interpreted each time the computation is performed, and space is saved as far as possible by packing so that more tables can be held in the core store during formation.

Various other standard statistical programmes, such as the solution of normal equations and sums of squares and products, were written and others are being prepared.

Experiments

Somewhat to our surprise the number of variate analyses required for experiments analysed in 1964 was greater by 26% than in 1963 (Table 3). In part this may be because with the more rapid analysis possible on the Orion there was little backlog this December, but the improved facilities that are being provided by the Orion are undoubtedly attracting additional work.

We were consulted on many problems of design and interpretation. Lessells summarised experiments on the effect of nitrogen on winter wheat and on spring cereals (12.11, 12.12). Clarke summarised 10 years' results on 5 rotation experiments on the residual effects of potassium fertilisers

https://doi.org/10.23637/ERADOC-1-57

(12.3). Boyd and Frater reported on the results of an investigation on the relative merits of different methods of soil analysis for determining responses to potash manuring (12.2). Preece investigated some problems in experimental design, in particular the interpretation of additional treatments in Youden square designs. Boyd prepared a report on experiments

TABLE 3

Numbers of replicated experiments analysed in the department

	Number of experiments By hand	On computer	Total	Number of variates on computer	Variates per experiment
1934	115	—	115	—	—
1951	437	—	437	—	—
1955	384	419	803	834	2·0
1957	98	1,253	1,351	5,041	4·0
1959	67	2,649	2,716	11,102	4·2
1961	89	2,862	2,951	15,184	5·3
1963	72	2,770	2,842	14,357	5·2
1964	88	3,383	3,471	18,054	5·3

comparing Charolais bulls with bulls of British breeds as sires of beef cattle in dairy herds (12.16). Patterson completed an investigation on the use of split plots for subsidiary treatments in cyclic rotation experiments (12.13).

In co-operation with Mr. J. A. Nelder (National Vegetable Research Station) and Dr. S. C. Pearce (East Malling Research Station) we continued the investigation into the problems of recording large amounts of data obtained in some types of experiment, particularly in horticulture. No one system seems likely to meet all needs, because some workers strongly prefer written records; but for cauliflower experiments Mr. Nelder has developed a Port-a-punch system, using the Orion for preliminary summary and error detection, with which he is very satisfied. For those Experimental Horticulture Stations that prefer written records, the Ministry of Agriculture's Data Processing Division at Guildford is offering a punching service. The National Vegetable Research Station, East Malling Research Station and the Glasshouse Crops Research Institute will be responsible for the analysis of the data on the Orion.

In conjunction with J. M. Hirst and G. A. Hide, D. H. Rees devised a system for directly recording on to punched paper tape observations on fungus infection on potato tubers. This replaced earlier methods of recording on a form or on a battery of mechanical counters. The output tapes are processed on the Orion, which checks for errors in punching and prepares a summary.

Surveys

Fortunately, in view of our other commitments, relatively few analyses of the survey type were required in 1964, but we were extensively consulted on the design of very varied agricultural surveys, and requests are expected to analyse many of these. The Ministry of Agriculture (Economics Branch) asked us to undertake a series of analyses of data derived from June 4th Returns, etc., similar to the work we did for them in 1962, and the first of these is started.

The results of the 1964 Survey of Maincrop Potatoes were analysed and

https://doi.org/10.23637/ERADOC-1-57

a report was completed in co-operation with the Potato Marketing Board and the National Institute of Agricultural Engineering (12.15) (Boyd, Church and Hills). No new surveys of fertiliser practice were done in 1964, but the delayed results of two earlier surveys were analysed, and the results of previous surveys provided answers to numerous queries. Yates and Boyd gave a general account of the surveys from their inception in 1942 (12.19).

Vessey analysed and reported on a survey of the incidence of mastitis in the Reading area (12.10). In connection with plans to eradicate brucellosis from cattle in this country, we co-operated with the Ministry of Agriculture in planning a further survey on a small random sample of the herds in Britain, designed to determine the frequency of reaction to the serum agglutination test. The field and laboratory work of this survey was completed in December, and arrangements have been made to analyse the results as soon as the cards are received.

There was considerable delay (for which we were not responsible) in preparing the cards for the calf survey, and work on the analysis was further delayed in the hope that the SEP programme would be available. It is now planned to use the EMA survey programme, and most of the preliminary work has been done.

Commonwealth and Overseas

The results of various experiments from overseas were analysed, and we were consulted on problems of experimental design. We could not give this assistance to overseas workers without computer facilities. When examining the results sent to us for analysis, we can often suggest improvements in the design of future experiments.

A. J. Vernon and Marjory G. Morris's investigation of variability in West African cocoa farms was published (12.17).

Theoretical Work

There was little time for theoretical investigations, but Yates reviewed the development of the concept of fiducial probability by R. A. Fisher, and particularly his justification of Behrens' test, which has given rise to much controversy (12.18). F. Yates and M. J. R. Healy gave the opening paper of a symposium on the practical needs that should be met in the teaching of statistics (12.20). M. J. R. Healy described the use of discriminant functions in medicine (12.8).

Other Work

Among the many miscellaneous investigations in which members of the Department collaborated the following led to published papers:

(*a*) Analysis of plasma calcium and phosphorus concentration in the Compton dairy herds (12.14).

(*b*) Study of variation in the long-tailed field mouse in North-West Scotland (12.5).

(*c*) The relation between height, puberty, family size, social class and educational ability in schoolchildren (12.6, 12.7).

(*d*) Capture–recapture observations on noctule bats (12.4).

https://doi.org/10.23637/ERADOC-1-57