SIMULATION IN AN INTERACTIVE

COMPUTER ENVIRONMENT

A Thesis

submitted to the

UNIVERSITY OF NEWCASTLE UPON TYNE

for the degree of

DOCTOR OF PHILOSOPHY

D.R. Appleton

November, 1972.

ACKNOWLEDGEMENTS

The first part of the research embodied in this thesis was conducted while I was in receipt of a postgraduate studentship for which I am grateful to the Science Research Council.

In following a course of research which tries to look outward from Computing I have come into contact with many people in different departments of the University whose help I am very pleased to acknowledge.

Firstly I should like to thank my Supervisor, Prof. E.S. Page, for the opportunity to embark upon the research described here, for his direction and help during its execution, and for the provision of quite excellent computing facilities.

I am also indebted to my Head of Department for the last two years, Prof. D.J. Newell, for his encouragement and for making possible the writing of this report.

Dr. K. Mitchell, Chairman of the Advisory Committee on the Development Plan, collected some of the data used in the room usage study of chapter 3, and I am grateful to him and to Mr. J. Lynn, Consultant Architect/Planner for using the system which I developed.

For many discussions about genetics, for his encouragement, and for using the genetics teaching programs of chapter 4 with his students my thanks go to Dr. D.F. Roberts of the Department of Human Genetics.

I was introduced to the cell cycle by Dr. A.R. Morley and Dr. N.A. Wright of the Department of Pathology. To them my thanks for many stimulating discussions and for allowing me to use the data which they had so painstakingly collected. Dr. Morley kindly provided me with figure 5.1 and the original version of figure 5.4 from which the cytokinetic model developed.

Finally I must express my gratitude to Dr. G. Shearing who was responsible for the graphics software, and to Miss Barbara James for her work on the illustrations.

SUMMARY

This thesis is a report of an investigation into the possible advantages to be gained by running computer simulations interactively. Models relevant to administrative, teaching and research work were constructed for the study, and the merits of the interactive use of deterministic models, stochastic equations and Monte-Carlo simulations were examined.

In order to be able to draw worthwhile conclusions from the investigation it was necessary to study substantial systems which had proved to require computer simulation for their elucidation.

The first model described is a deterministic portrayal of the use of university lecture rooms which was written as an aid towards evaluating the need for additional rooms in an expanding university. The part played by interaction is to make easier the incorporation of human experience into the planning mechanism.

Population genetics provided the next system, and it is shown how a model using a set of stochastic equations in conjunction with fast graphical output may be of value in teaching. A Monte-Carlo approach is demonstrated to be unsuitable for interactive use.

The final simulation also employs stochastic equations, designed to represent a system from the field of cell cycle kinetics whose action is not fully understood. Using the model interactively allows the researcher to form an appreciation of the consequences of altering its parameters and to fit experimental data with more perception than is possible using purely algorithmic methods.

Introductions are given to the two biological systems dealt with, so that the results of using the models can be discussed in relation to the genetics and cytokinetics involved as well as purely in the context of the interactive use of simulation.

CONTENTS

Page

1. INTRODUCTION

1.1. INTERACTIVE USE OF SIMULATION

The research, the results of which are embodied in this thesis, was carried out in order to determine whether there were any types of simulation in which the user of an interactive terminal system had an advantage over his batch-mode counterpart.

The most likely situation in which this might occur is where the system being modelled is, or has previously been considered to be, an aggregate of algorithmic and non-algorithmic processes; where for instance some human decision is made on the basis of experience or intuition. For an adequate model of this system it is necessary to incorporate a similar stage of human decision making. The situation is like that encountered in business games, so a problem which arose in a planning context (the administration of a university) was used as a basis for the investigation of this sort of system.

The model of this system can also be used to throw some light on the change in attitude towards the use of computers which can be brought about in workers outside the computing field when they are exposed to interactive computing. Many are basically averse to computers and use them only because they see that it has become absolutely essential to do so. They are unlikely to have a natural facility in the use of computers, but with an interactive program prompting them in clear language for data and informing them of its operations they may begin to see the advantages computers can bring to their own disciplines.

Another type of system which is likely to be amenable to interactive simulation is one in which the model itself may be programmed, but flexibility in its use is essential. The second simulation considered is therefore designed to investigate how the flexibility available from a terminal system can be used by a teacher to instruct university students in some of the basic notions of population genetics. The combination of a terminal system and the capability of fast graphical output via a storage display tube makes this particular teaching application likely to succeed.

Other questions of interest which can be discussed in the context of an extension of the genetic model are whether Monte-Carlo simulations are suitable for interactive use in view of the large amount of program running time which this technique usually needs; and whether special purpose languages have anything to offer. First two simulation languages (GPSS and CSMP) are investigated followed by a language which is particularly well suited to interaction and which also has good simulation capabilities (APL).

Sometimes the advantages of interactive computing are rather intangible. The increased

facility for program debugging and development, the vast improvement in turnround time for short jobs with little input or output and the greater overall convenience bring a higher degree of user satisfaction. Such an improvement can lead to programs being written which otherwise might not have been begun. The final system investigated could have been modelled non-interactively, but familiarity with the system would have been more difficult to obtain, and the use of batch methods could easily have led to an unsatisfactory method of analysis. The problem which arises is to find the values of the parameters of a multiparametric model which best fit experimental observations. For reasons which are discussed in section 5.3.1. the use of standard curve fitting techniques and automatic optimisation procedures are inappropriate, while interactive methods may be able to provide a suitable alternative. Heuristic techniques are sometimes used merely because of the programming effort which is thereby avoided, but the situation here is different; the human factor is important to the success of the model.

1.2. FIELDS OF APPLICATION

This report is concerned with investigations conducted in three different directions in an attempt to determine in which circumstances it might be advantageous and in which disadvantageous to attack a problem by means of an interactive computer simulation. The directions chosen are not the model-dependent ones in which the system is represented by a deterministic model, a stochastic equation or a Monte-Carlo simulation, although all these are considered, but are application-dependent. The models considered are taken from administration, teaching and research.

The methods used in these three fields overlap to a considerable extent but in each the emphasis is different, for the questions raised in each application are not of the same nature. Models for use in an administrative context are likely to be large and possibly clumsy, frequently giving rise to data-base problems. Interaction plays its part in facilitating the incorporation into the model of the experience of those involved in the running of the system, aiding communication between model-builder and administrator, and generally leading to more acceptable results than an entirely algorithmic approach is capable of providing.

Teaching models, on the other hand, are at their best when they are simple representations of the system whose behaviour is to be understood. They must run quickly so that organisational difficulties are not allowed to obscure the main object of the lesson, while at the same time they must be acceptable substitutes for experiments with the system itself and not give an impression of unreality. Interaction provides for

students' interest in and involvement with the system.

Davies [1] points out in relation to models and simulation in quantitiative biology that "for the most part the construction of a model depends not on the use of a computer but on the modeller's understanding of the biological and physical characteristics of the problem and on his ability to express it in mathematically tractable form without resorting to undue simplification".

For this reason the background to the two biological models presented is given in some detail in chapters 4 and 5. It is especially important that anyone attempting to produce a model for research purposes should be familiar with the real-life system. Only in this way is he able to simulate the system accurately without obscuring its workings by modelling the associated noise. Research-oriented models frequently have two concomitant objectives; the estimation of the parameters of the model which fit experimental data, and the elucidation of the manner in which the system operates, by testing assumptions about it. An interactive model is particularly helpful as a method of investigating the second of these.

Taylor [2] discusses the differences between models intended for teaching purposes and those designed as research tools with respect to the degree of complexity which must be incorporated. Although he is dealing with models run on analogue and hybrid computers his observations have wider validity. He confirms the view that the research-oriented model must be a faithful representation of the system while the equivalent model used for teaching purposes may (or indeed should) sacrifice this precision for simplicity. The teaching model studied, while treating a system of the same order of complexity as the research-oriented one, is therefore dealt with by a less detailed model for which much of the analysis can be done analytically to enable very fast computer response.

1.3. TYPES OF SIMULATION

There are several different ways of simulating a system, more than one of which may be considered appropriate to a given case. For the room allocation problem chosen to exemplify the use of simulation in an administrative context a deterministic model was adopted. The reasons for following this course in preference to the other possibilities are given in section 3.2.3. Deterministic models are characterised by their relatively fast rate of execution in comparison to Monte-Carlo models and are especially useful when the variation in the system is of secondary importance.

When variation is an integral part of a system the use of a stochastic equation

approach is attractive. Such is the case in the teaching application considered. A vital part of the lesson was to convey the idea of the probability distribution of gene frequencies and to calculate the distributions for several values of the parameters of the genetic system in question. A Monte-Carlo approach also held some initial attractions and its possibilities were investigated.

In the research-oriented model, which describes the kinetics of cells in a biological system, the variation between cells was an important consideration rendering a deterministic model inappropriate, while the interactive nature of the model excluded the too-slow Monte-Carlo technique. It was, however, possible to construct a mathematical analogue of the system and use the computer to solve the resulting integral equations. A step by step technique of solution essentially brought about a fairly direct model of the system. Further details of the model chosen are in section 5.2.

1.4. SIMULATION FOR PARAMETER ESTIMATION AND INSIGHT

The type of simulation best known because of its use in industrial situations is carried out to see in what manner one particular parameter will vary if a working system is modified. Usually the problem may be concisely posed, although the system may be complex, and a straightforward answer can be expected. Such an experiment may involve the calculation of a parameter of the system which was previously undetermined (such as the number of calls lost to a telephone system because of congestion), but the nature of all the interactions of the system will be known before the task is undertaken.

At other times not only will quantitive information be required from the model, but it will also be expected to provide qualitative descriptions of a system which may, in the light of previous knowledge, be functioning in one of several ways. This is very often the case in biological applications of simulation. It is likely that the system to be modelled is imperfectly understood and the experimental data that is available must be used to try to determine which of a range of possible hypothetical systems could be operating. It is in this situation that interaction is most likely to play a prominent part, enabling the user of the model to observe the effects of altering parameters at first hand instead of necessarily confining himself to a predetermined strategy or accepting an intolerable response from many successive batch runs.

The construction of a model which is primarily intended to give insight into the workings of a system rather than estimate its parameters as accurately as possible calls for a great deal of judgement on the part of the model-builder. Without overwhelming the model with superfluous detail he must at the same time represent the system

sufficiently faithfully to enable it to be used in a predictive capacity.

1.5. SIMULATION AND OCCAM'S RAZOR

William of Occam's famous assertion "entia non sunt multiplicanda praeter necessitatem" (freely translated as "things should not be made more complicated than necessary") has a direct and very important bearing on the construction and use of simulations. Firstly, in the construction of a model, the lesson to be learned is to guard against modelling second order effects which are not under investigation. These merely act as noise and obscure the more important effects.

The main contribution which Occam's razor makes, however, is in connection with the use of such models. During their construction it is not always possible to determine which parameters will be crucial or what range of values will be of most interest. Consequently it can be desirable to incorporate a considerable degree of generality into a model, and a multiparametric model results. With an interactive simulation it can become apparent during the process of familiarising oneself with the model that certain parameters, at least within a theoretically acceptable range, are of minor importance to the resulting yield surface. If any independent evidence for the values of such parameters is available they should be fixed at these values when the model is used in a data fitting context.

While it is essential that the user should be aware of how many parameters are employed by his model, this number can sometimes be underestimated. Bode's rule, which states that the relative distances of the planets from the sun are found by adding 4 to the series 0, 3, 6, 12, 24, ... is rather less successful than at first sight it seems and as it was for long accepted. In fact the relationship may be expressed as $R \propto 4 + 3x\ 2^n$ for $n = -\infty, 0, 1, 2, \ldots\ldots$ and it can be seen that the model has 3 numerical parameters, and is also forced to use a slightly artificial sequence of values for n. Add to these points the fact that there is no real physical justification for the form of the relationship and it becomes apparent that the ability of the model to fit data for the first seven planets is hardly surprising; nor is its failure with the remaining planets unexpected.

The predictive ability of a simulation is one method of distinguishing a satisfactory simulation from one which has nothing worthwhile to offer. In biological systems with many parameters, models must also have many parameters and frequently data can be explained in more than one way. A multimodal response surface characterises such a situation. If it is impossible to choose between the alternatives on theoretical grounds or on the basis of greater simplicity then the model should be capable of suggesting further experimentation to differentiate between them.

1.6. DATA COLLECTION

Before any investigation can be carried out into the desirability of employing interactive capabilities in the running of computer simulations it is necessary to choose a fairly complex system about which worthwhile questions may be asked and about which there is a considerable quantity of relevant data available. There is little point in modelling a simple system, nor in merely watching the workings of a more complicated one. The three systems which are used as bases for building models in this thesis all have relevance in other areas. A necessary and non-trivial prerequisite for discussion of interactive use of the models is an understanding of the systems and the form in which data is available about them.

The model presented in chapter 3 which relates to lecture room usage in Universities arose from problems encountered in planning future building programmes. The model was designed and constructed in order to study limitations imposed on student numbers by existing accommodation, to allocate rooms efficiently to classes and to determine the most appropriate size of essential new lecture rooms. Data relating to the lecture rooms, the available courses, student numbers and the course timetable had to be collected and organised. Some of this was available from University administrative sources, while one aspect, the probability of a student on a given course attending a given lecture was more difficult to obtain. Further details of the data are given in section 3.3.1.

The two biological models illustrate a common problem connected with the use of computer simulation. Frequently there is difficulty in communication between the person who wishes to investigate a system via a model and the person who is to formulate, program and test the model. Two courses are open; the most profitable in the long-term may well be for the biologist, physicist, engineer or whoever understands the system to make himself acquainted with all the necessary computing techniques and to program the simulation himself. This is very often the case in the physical sciences, but biologists and social scientists often seem reluctant to invest time in acquiring the necessary skills. It therefore falls to the computer scientist to learn about the system he is to simulate. He must learn about it in depth, and not accept any part of it as a "black box". If the programming is not going to be understood by the data collector it is vital that the system and the data be understood by the simulator. It is also necessary for anyone making a judgement of the value of interaction with the models to have an appreciation of the systems involved; introductions to the genetic and cytokinetic systems will be found in sections 4.1 and 5.1 and glossaries of those words which may be unfamiliar to the computer scientist have been compiled and appear as appendices.

The stochastic equation used in the genetic simulation (the Fokker-Planck equation) has been dealt with in several research articles, and the data used with this model was also available in the literature. In a teaching situation it is convenient that this is so, while it is helpful in arousing interest if the model and the data have not previously been allied.

The data for the model of the cell cycle was gathered, some previous to the construction of the model, and some as a result of predictions made by the model, in the Department of Pathology, University of Newcastle upon Tyne, by Dr. A.R. Morley and Dr. N.A. Wright. The experimental protocol is briefly described in section 5.4.1.1.

2. COMPUTING FACILITIES

2.1. HARDWARE

Since the execution time and storage requirements of a program are important considerations if satisfactory use is to be made of it in an interactive situation, it is relevant to describe the physical characteristics of the equipment on which the programs discussed here were run.

The main computer available was an IBM 360/67 with 1 megabyte of main memory. A 2301 magnetic drum of 4 million bytes capacity, with a transfer rate of 1.2 million bytes per second, provided the paging device for the virtual memory of the timesharing system. On-line file storage of some 400 million bytes was on thirteen 2314 disks.

Execution times for arithmetic operations between the working registers of the 360/67 are 0.65 μsec for addition or subtraction, 4.45 μsec for multiplication and 8.45 μsec for division.

Batch jobs were run from a 2540 card reader (1000 cards per minute) and output on a 1403 line printer (600 lines per minute). An alternative service was provided for short batch jobs, a 2501 reader (1000 cpm) and a 1443 printer (200 lpm) being operated on an open-shop basis.

Thirty-one 2741 typewriter terminals and eight 2260 CRT terminals were supported for interactive work. Three 2701 parallel data adaptors served remote centres, and a further 2701 was linked to a PDP 11 of 12K 16 bit words which controlled a Benson-Lehner graph plotter and a Computek 400 storage display tube. The Computek was supported by software only as an output device.

2.2. SOFTWARE

For most of the day the 360 operated under the Michigan Terminal System (MTS). MTS is a timesharing system written in reentrant code which supports conversational and batch use. Each user is provided with his own virtual storage and CPU, and has access to a considerable number of public files containing compilers, program libraries and file editors. These can be used in conjunction with his own files through the medium of a simple but comprehensive command language. High level programming languages available included ALGOL 60, FORTRAN and GPSS; subsequently ALGOL W and APL were implemented. Subroutines which could be called by FORTRAN and ALGOL W programs were available for outputting to the graph plotter and the Computek. The latter device can be used in two modes. The simulation program may be designed either to

produce a file suitable for subsequent plotting or viewing on the Computek, or to send the output directly to the screen so that it is possible to examine the picture without halting execution of the program.. The second method is the more convenient if only one picture is produced by a given set of data, but is less useful if more than one graph is drawn. However, there are also advantages in storing the program output which compensate for having to terminate the modelling program, because such files are so constructed that graphs may be superimposed. This greatly aids the comparison of graphs derived from models with different parameters.

3. AN INTERACTIVE COMPUTER MODEL OF UNIVERSITY ROOM UTILISATION

3.1. INTRODUCTION

Although management interest in simulation is not new, Universities have been slow to incorporate into their administration techniques proven in business and industry. Until recently only timetabling problems had been tackled in any detail on a computer, but some reports have now been given of models dealing with aspects of university resource allocation. These include an ambitious model of Fairfax University by Scarborough and Daniel [6] which examines the general structure of the administration in an interactive model without investigating any one problem in great depth. Another model, of the University of California at Berkeley, has been described by Longworth Smith [3]; this is a mathematical model which treats student demand for courses as a stochastic variable and obtains information on the required distribution of classroom sizes. Unfortunately lack of data prevents any extensive validation of the model. The problems of timetable construction and student scheduling in the context of the American university course structure have also been dealt with by Holz [1] using the GASP program. Hopkins [2] has discussed the uses and limitations of a matrix oriented cost simulation model. The difficulties inherent in such a model (changing transformation matrices with time, and the need to collect a vast amount of data) are not present in the current situation. In section 3.2.2. it is shown how the interactive nature of the model allows much of the data collection to be circumvented. McNamara [5] presents an extensive bibliography of work done, mainly since 1969, which is relevant to his mathematical programming approach to the problem, and an interesting analytical development is the graph-theoretical approach of McDiarmid [4].

3.2. APPROACH TO THE PROBLEM

3.2.1. STATEMENT AND LIMITATIONS

In session 1969-70 the University of Newcastle upon Tyne catered for some 6000 students ofwhom 4670 were in the 39 departments considered, the remainder being in the Faculties of Medicine and Education whose development may be dealt with separately. It was required to know whether lecture rooms available during that session would be sufficient to house an increased number of students; which departments could (by this criterion) accept more students; and what would be the most appropriate sizes of any new rooms required if the student intake were to be increased by a specified amount.

The possibility of altering the existing class-teacher timetable was not considered

except in very special cases where a single class would otherwise necessitate the building of a new room. It was felt that such a revision was a major undertaking and a quite separate problem from that posed by raising student numbers.

Other factors which would be important in a real expansion, such as availability of staff, increased refectory facilities and extra student accommodation were not considered; these aspects could conveniently form parallel investigations. Nor were laboratories, drawing offices nor very small teaching rooms included in the study, as these rooms are frequently used by departments for specialised activities and may not be suitable for other purposes. They could however be treated in a similar manner to that in which lecture rooms were examined if some modifications were made in the method of formulating a room timetable.

3.2.2. ADVANTAGES OF AN INTERACTIVE MODEL

Simulation of a system has two principal benefits as a technique of investigation; firstly it can produce results not easily obtainable by analytical or other methods, and secondly the actual construction of a model brings new insight into the more complex interactions of the system. This latter advantage is emphasised when an interactive model is studied, when not only does the model builder benefit, but, more important, the users of the model are helped to understand the system more fully. The present model is typical of many in being written by someone with no immediate interest in the system being modelled and intended to be run by someone with no experience of computer simulations, and the use of interactive techniques facilitates the user's task.

Furthermore, the computer can be used more efficiently to deal with complex problems if it can be seen that human intervention is possible, for if recommendations decided on the basis of a computer model are to be acceptable it is as well that not only algorithmic methods but also the experience and expertise of those responsible for implementation should have been incorporated into the decision making process. In the present instance interaction is vital since only the user knows or can find out the possibilities for reconstruction of rooms, provision of new rooms, and some of the other constraints associated with a proposed increase in student numbers, such as the necessity to divide a class in two because of unwieldiness. To build a non-interactive model it would be necessary to ensure that data was available for every contingency which could arise in the course of its running. This would be an extremely wasteful procedure, for in the current situation it turns out that a great many facts need to be used typically only once or twice each.

3.2.3. THE TYPE OF MODEL USED

An important practical decision is whether to treat the system as deterministic or probabilistic. The sizes of many classes are not known until after the start of teaching and until then only estimates can be made. It would be possible to incorporate a random element into the model to take this uncertainty into account, but such a procedure entails considerable extra complexity without giving obvious compensating advantages. A practical reason for rejecting a Monte-Carlo model is that the need to replicate simulation runs is likely to occupy enough computer time to make interaction with the computer infeasible, and so in the present situation it was decided to employ a deterministic model. Similarly a stochastic model was rejected because the statistic of interest is the maximum number of students who can be accommodated. Thus it would be necessary to look for estimates not of the mean of a distribution but of, say, the 95th percentile. To get an accurate answer distributions would have to be known with a precision which is unlikely to be obtainable.

As has been indicated the aims of the simulation were more closely associated with providing a model which could answer several related problems than with producing a system designed to answer a single clearly posed question. One point of view was to ask how many more students could be accommodated in existing lecture rooms, while another was to find out how many new rooms would be needed to house a given increase in student numbers. The second approach had to recommend it the fact that the future size of the university had already been proposed, but on the other hand the increase need not be uniform in all departments, so the situation remained open. As far as this report is concerned it will be assumed that in any particular application of the model the student numbers have been determined in some fashion. An extension to the generalised problem may be thought of in terms of iterative use of the model.

It was originally intended to simulate the system fairly closely by updating supposed student numbers year by year. This was found to be unnecessary in what was principally a long-term planning exercise, so that the situation could be considered to be in equilibrium, and the temporal aspect of the model became submerged. Again, iterative use of the model could restore this side of the question to prominence.

3.3. THE ORGANISATION OF THE MODEL

3.3.1. DATA STRUCTURE

Lists were obtained from each university department of the lectures given by its staff and to its students by other departments. (This task was performed by Dr. K. Mitchell.)

For each lecture the time and the room were given and the number of students attending was categorised by their registration codes. (On registration at university each student is given a mnemonic code to describe his course for the year; there are 442 such codes, most of which are used in any year). For the computer model the fundamental unit is the lecture room, and so the data was arranged in the form of timetables of 35 weekly periods for the 115 rooms of capacity 20 or more which were considered.

A file of 14 pages (1 page = 4096 characters) was constructed to represent the time-tables, each period containing the address in another file of a list of the registration codes of students in the room at that time, with the percentages of total students registered under these codes who were actually present. The file containing this data occupied 9 pages, and a file containing total numbers of students for each of the relevant codes, along with explanations of the codes, used 11 pages. A further 4 page file held the capacities of the rooms and descriptions of their locations.

3.3.2. PROGRAM STRUCTURE

It is not necessary for the user of the suite of programs to understand the organisation of programs or data or any details of the MTS system. To initiate a session he issues a single command and thereafter responds to prompting and questionning. As a result of the initial command the CRT display is as shown in figure 3.1. (page 20). The user's reply causes the relevant program to be loaded. The use of such a controlling program enables dynamic loading which reduces the size of the object module needed in the virtual memory at any time, an important consideration in a time-sharing environment. This advantage would be nullified if a great deal more loading were thereby entailed, but in fact each program, especially the main program, tends to be employed for a substantial time.

The programming language used throughout was ALGOL 60 and the compiled object modules occupy 39 pages; the total storage required for the system, exluding source programs and library modules but including temporary space for updated files and line-printer output, is 142 pages. About half of this total storage may be required in virtual memory at one time. This is close to the limit of the Newcastle MTS system when the load upon the machine is heavy, but at off-peak periods it is quite acceptable.

The suite includes programs which will investigate the room usage at a superficial level as a screening technique, run the main program, perform housekeeping on the data, or transform the data from symbolic form into readily comprehensible room-orientated timetables. The first option (figure 3.1.) is the most important, but the other six programs are essential for its usefulness.

```
WHICH PROGRAM IS REQUIRED?
    1.   TO PERFORM REALLOCATIONS
    2.   TO PRINT OUT TIMETABLES
    3.   TO UPDATE STUDENT NUMBERS
    4.   TO ADD NEW ROOMS
    5.   TO CHANGE EXISTING ROOMS
    6.   TO UPDATE THE TIMETABLES
    7.   TO GATHER STATISTICS
    8.   TO TAKE NO ACTION
ENTER THE APPROPRIATE NUMBER
```

igure 3.1. The options open to the user of the room usage model when the system is initiated; as shown on the 2260 screen.

3.4. USING THE MODEL

3.4.1. AN EXAMPLE TERMINAL SESSION

The first program called up by the user in a typical session might be number 3. This would enable him to alter by some proposed ratio the student numbers in any specified departments or faculties. He would then run program 7 to gain an idea of the congestion in the rooms under investigation, through their percentage morning and afternoon utilisations and the frequency of overcrowding. If it seemed feasible he would alleviate this congestion by means of the main program (number 1); but it is possible that only the provision of a new room could cure it completely. Option 4 simulates the building of a room of the desired size and the user may then recall program 1 and conclude reallocations. The timetables can now be sent for printing using program 2, and while this is being done it is possible to store them (program 6) for future use. The user exits from the suite by calling option 8. The time spent in such a session is dependent on the number of rooms looked at and the number of reallocations to be made by the user. For an increase in class sizes of 50% in 20 moderately well used rooms reallocations might be performed and timetables produced in about 1½ hours.

Many other permutations of the programs are, of course, possible, and the particular one chosen will depend to a degree on the temperament and skill of the user as well as on the amount of congestion and the number of rooms involved.

3.4.2. THE ALLOCATION PROGRAM

One of the obvious advantages of an interactive system from the point of view of an inexperienced user is the way in which the system can guide him through his task by offering him options instead of forcing him to remember and make provision for entering the various items of data which are required when a particular run is chosen. All of the programs therefore prompt for data. A description is given of the main program alone as the techniques used in the others of the suite are basically similar. The program listing is in appendix 3.

When the allocation program is called up the following questions are put to the user.

1. "How many rooms are to be considered?"
   While the program will allow the user to deal with all the rooms at once it is envisaged that they will be dealt with in groups of about 20 selected because of their location on the university site.

2. "Enter their numbers."
   The user may enter a list of numbers, or shorten this to "from _____ to _____"; or any

combination of these two forms. It is important in preparing a system for interactive use that as many as possible of the obvious unambiguous answers to a question should be acceptable, although a clear manual emphasising the expected answers should have been studied by prospective users.

3. "At what time do you wish to start?"
Usually this will be answered by "9, mon", but it is also possible to start elsewhere in the week. The program takes each day in turn starting at 9 a.m. and finishing at 4 p.m. The option is available in case a previous run had to be curtailed, so that the system need not look at periods which are known to have been eliminated of congestion.

4. "Do you wish to make all room changes by hand?"
Normally the answer will be "no", when the program will make the simpler reallocations and will only return control to the user when it is difficult to place classes or if this cannot be done. The algorithm used by the program will be explained later in this section. The user will answer "yes" so that he can overrule the algorithm if a run is being carried out for the purpose of resiting a few particular lectures because consultation with departments has shown a certain room to be unsuitable.

When all periods have been dealt with the user must answer "yes" or "no" to three further questions.

5. "We have completed the week
Do you wish to look at an earlier time?"
The user may have found a period complicated to resolve satisfactorily, and he has the option of leaving the problem unresolved and continuing with the next period. It may be advisable to do this, because if such situations arise often it is likely that an extra room or rooms will have to be used. Having finished the week he may return to the problem period by answering "yes" to this question and "10, wed", for example, to the subsequent question "Which time?"

Notice the personification of the program in this question; a small point apparently, but it is nevertheless important that the user treat the program as a consultant with the timetabling facts at his disposal, and the attitude of shared responsibility for decisions is not another symptom of the condition in which the computer is blamed for human error, but is healthy use of the system.

6. "Do you wish to make any further changes?"
This is an option offered to the user to enable him to refine the earlier placing of classes

by the program, and gives him latitude to employ his specialised knowledge of any factors which might invalidate the decisions suggested by the computer.

7. "Do you wish to save the new timetables?"
   If the user has been successful in allocating classes to rooms he will answer "yes" and the timetables constructed will be saved in permanent disk-files.

After question 4 has been answered the program reads the data necessary for the rooms to be considered and begins to calculate the number of students in these rooms at the periods specified. When it detects that there are more students occupying a room than its official capacity it prints a message like:

"MONDAY AT 10
ROOM NO. 57 CAPACITY 40 44 STUDENTS"

No action is taken on this room until all further rooms have been dealt with for the same period; then, if the user is not dealing with all reallocations himself, the program takes the following course:

A. If the largest number of overcrowded students can be put in the largest empty room this is done. This step is repeated until it is impossible or no room is overcrowded.
B. If step A cannot be executed, then if the offending class can be exchanged with another so that both can be accommodated this is done and the program returns to step A. When neither step can be carried out the user is given control.

When control is passed to the user, either because he requested it or because the simple algorithm above has been unable to stop overcrowding he has the capability of issuing the commands shown in figure 3.2. (page 24), where the second form indicates the minimum necessary. The user has the choice of making his commands short or readily understood. (For clarity particular rooms and times are shown.) The action of the program when given one of these instructions is as follows:

I. continue No (further) attempt is made to place classes for this period, and the program continues with the next period.

II. stop The rest of the week is ignored and the program only gives the option of saving the timetables before terminating.

III – V. focus, does, when After III has been used to confine attention to a particular room, IV will be answered by the number of times that the same number of students use the room later in the week. V lists the times this situation occurs.

VI. how many This question may be used for any room at any period. With III – V it helps the user to allocate one room consistently for a class which is displaced

I	continue	c
II	stop	s
III	focus on room 12	f 12
IV	does the situation occur again?	d
V	when?	w
VI	how many are in room 12 at 10, friday?	h 12, 10, fri
VII	exchange classes in rooms 12 and 13	ex 12, 13
VIII	put class in room 12 at 10, friday in room 13	p 12, 10, fri 13
IX	list status	l
X	repeat	r
XI	move to 3, friday	m 3, fri

Figure 3.2. The command language in which the user interrogates the program and performs room allocations.

frequently, if this is possible and desirable. Like the possibility of overriding the program's placing of classes it is a refinement which is sometimes worthwhile if a department's preferences on the matter are known.

VII – VIII. exchange, put These are the most frequently used instructions. While VIII is capable of doing the work of VII, a straightforward exchange of classes in the period under investigation is so common that it pays to have a simple instruction for the procedure.

IX. list status This instruction gives the capacity and number of students for all rooms being considered, at the time in question. It is clearly a most valuable instruction but suffers from two slight drawbacks; it is more time-consuming than the others, which is of no importance if it is correspondingly more convenient, but also if more than 10 rooms are under investigation the answer to the query takes up more than the capacity of the 2260 screen, requiring the user either to scroll backwards and forwards over the screen or to copy down the output. Either way is less than convenient so that it turns out that this apparently all-purpose instruction is less used than one might suppose and is generally brought into play only as a last resort.

X. repeat The program will repeat the period just dealt with, taking any reallocations into account. This is helpful for checking that the changes made have had the desired effect, for clarifying a position which has become complex, or for returning control to the program to rehouse smaller classes whose problems have been shelved while a larger class has been dealt with.

XI. move All periods are acceptable in this instruction, so that the user may jump backwards or forwards. A greater degree of flexibility is thereby achieved, as it is inconvenient to work with any period other than that at which the program has given up control, due to the lack of instructions like "exchange" and "list". Alternatively, the instruction may be used when the critical periods have already been isolated and these alone may be considered, or only morning periods may be dealt with as in some faculties very few cases of overcrowding in lecture rooms occur in the afternoons when students are largely engaged in practical classes.

Part of a run of the allocation program, in which some of these commands are used, is shown in figure 3.3. (page 26), and further discussion of their use is presented in the following section.

```
WEDNESDAY AT  10
    ROOM NO.  41  CAPACITY  50     65 STUDENTS
    ROOM NO.  47  CAPACITY  30     39 STUDENTS

THE FOLLOWING ROOMS ARE EMPTY AND WILL TAKE
THE LARGEST OF THESE CLASSES
 NONE
THE FOLLOWING ROOMS ARE LARGE ENOUGH
ROOM NO.  44  CAPACITY  70     55 STUDENTS
ROOM NO.  46  CAPACITY  80     76 STUDENTS

WHAT ACTION IS REQUIRED?
LIST

STATUS OF ROOMS ON WEDNESDAY AT  10
ROOM NO.  41  CAPACITY  50     65 STUDENTS
ROOM NO.  42  CAPACITY  60     55 STUDENTS
ROOM NO.  43  CAPACITY  60     47 STUDENTS
ROOM NO.  44  CAPACITY  70     55 STUDENTS
ROOM NO.  45  CAPACITY  40      0 STUDENTS
ROOM NO.  46  CAPACITY  80     76 STUDENTS
ROOM NO.  47  CAPACITY  30     39 STUDENTS

WHAT ACTION IS REQUIRED?
EX 41,44

ROOM NO.  41  CAPACITY  50     55 STUDENTS
ROOM NO.  44  CAPACITY  70     65 STUDENTS

WHAT ACTION IS REQUIRED?
EX 41,43

ROOM NO.  41  CAPACITY  50     47 STUDENTS
ROOM NO.  43  CAPACITY  60     55 STUDENTS

WHAT ACTION IS REQUIRED?
PUT 47 AT 10,WED IN 45

ROOM NO.  45  CAPACITY  40     39 STUDENTS
ROOM NO.  47  CAPACITY  30      0 STUDENTS

WHAT ACTION IS REQUIRED?
CONT
```

Figure 3.3. An example of the use of the command language of figure 3.2.

3.4.3. USING THE ALLOCATION PROGRAM

In order to deal with the variety of situations which he may encounter in his attempt to make full use of the resources at his disposal and at the same time to determine when these resources are insufficient, the user must have available a wide-ranging yet uncomplicated set of commands. Just as he must be able to screen the overall situation (program 7) before attempting to improve it, so he needs commands to investigate a particular period at different levels. In the normal situation where the program returns control to the user after failing to find room for a class it tries to ease his problems by listing all the rooms large enough to take the class. None of these is empty or the algorithm would have been equal to the task, nor do any of them hold fewer students than the overcrowded room is capable of accommodating, otherwise the classes would have been interchanged. It is often the case however that the students in one of these rooms may be exchanged with a small number in a moderately sized room and these may then be exchanged with the overcrowded class. For instance if rooms 1, 2 and 3 have capacities of 50, 60 and 70 and are expected to hold classes of 65, 45 and 55, room 1 will be listed as overcrowded and room 3 as large enough to house the students. The user should be aware of the capacity of room 2 and may also know from his initial investigation that it is seldom used to capacity. He then inquires of the program how many students are currently occupying room 2 and on receiving the answer exchanges the classes in rooms 2 and 3. He may then either resolve the situation by exchanging the classes now in rooms 1 and 2 or he may repeat the period and allow the program to do so for him.

The foregoing case is the simplest with which the user will be required to deal. Figure 3.3. shows a slightly more difficult case with two rooms overcrowded. Here it is just possible to deal with the smaller room by fitting the students into an empty room. Because 40 is the second smallest classroom capacity and 39 the second smallest size of class it is impossible to provide a greater margin of error for the reallocation, but in similar cases the user may wish to do this, especially if he determines using the more detailed questions that the situation occurs frequently. If he suspects that class sizes are slightly underestimated he must exercise his judgement whether or not to allow such exact reallocations. (He must not assume that the whole class will seldom be present at a lecture so that an apparent crush is permitted; naturally if the room is uncomfortably crowded students may stay away.)

The user, then, has three types of command; those for interrogation, those for action, and those for the control of the program. He may use either of the forms shown in figure 3.2. according to temperament or he may use intermediate forms as in the "list" and "put" commands of figure 3.3. There is no such alternative for output, which is always immediately intelligible. A 2260 CRT enables this course to be adopted; using the same output on a 2741 terminal would

be tedious, and abbreviations are often obscure and irritating, especially for those unused to computer systems.

3.5. EXPERIENCE GAINED BY USING THE MODEL, AND ITS FUTURE USE

After the system had been tested on current student numbers and rooms the effect of increasing the intake of one faculty (Arts) by 100% was investigated. It was found that the simple reallocation algorithm used by the program was quite adequate, the user requiring to make relatively few allocations himself. The somewhat surprising result emerged that for such a large increase in students, given one change in the class-teacher timetable, only one new lecture theatre would be required, while several of the smaller rooms could be used for other purposes. It is unlikely that this finding would have been obtained without a computer model.

The timetables produced for the rooms considered (figure 3.4., page 29) proved a satisfactory method of communication with those responsible for administration; they are suitable for use by the planners in assessing the amount of travelling involved in the changes and in ensuring that no unacceptable congestion is likely in lifts or on staircases. They are also in a form suitable for scrutiny by departmental representatives who could detect any aspect of the room allocations which displeased them.

The use of the model demonstrated that time-sharing systems can lead to interaction in three basically similar but practically distinct ways. Firstly a user may interact with a program, as in the allocation program, by issuing commands and responding to questions put by the program. Secondly he may interact with a suite of programs, deciding which program is appropriate at a particular point in time; when to screen the data, to run the main program, perform housekeeping, or run a program which will aid communication with other interested parties, such as the timetable printing program. Thirdly the user may decide, possibly more slowly than at a terminal, what his next course of action is to be; for example whether the combination of two rooms to make a larger one is feasible. It is important that he should not feel bound to reply to the computer as quickly as it gives results and poses questions.

In this context, however, he should maximise his chances of interacting quickly with the computer by arriving at the terminal well prepared. He should know what he wishes to accomplish during the session and should have copies of the previous timetables for the rooms to be dealt with as well as listings of the files of room capacities and registration codes. Although the necessary information can be obtained at the terminal by using the MTS file editor it is much easier in almost all instances to look up a hard copy listing of a file to determine,

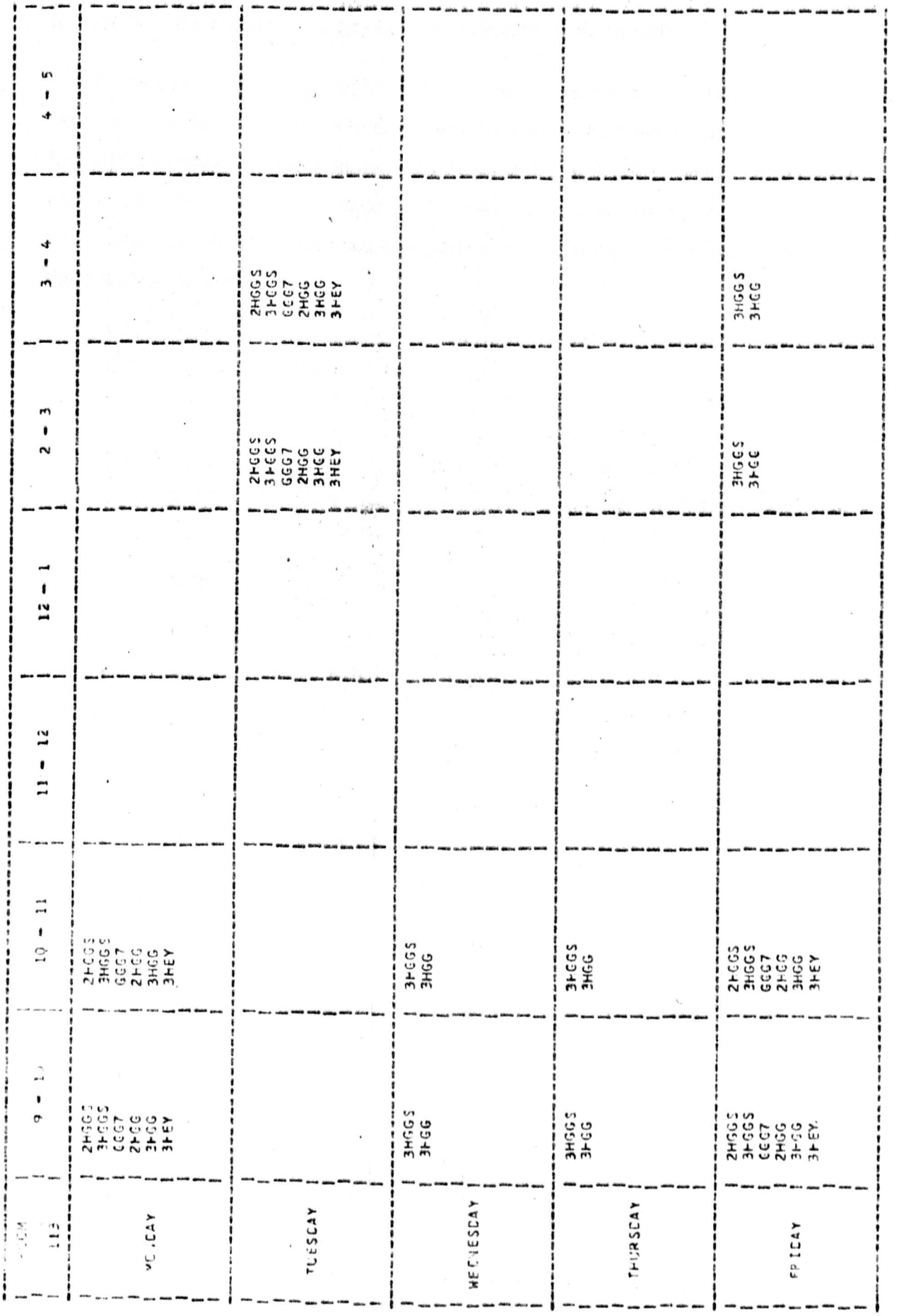


Figure 3.4. Line printer output of a typical room timetable, showing the registration codes of students attending lectures.

for instance, the meaning of a forgotten mnemonic registration code.

As well as continuing to employ the model in long-term planning situations it is hoped to use it to allocate lecture rooms to classes at the start of each university session. Registration codes are stored on the computer as a matter of course and such files could be used to provide an accurate breakdown of student numbers before lectures begin. Lectures which would be oversubscribed could be pinpointed and alternative venues allocated before the situation occurred.

4. AN INTERACTIVE COMPUTER MODEL FOR TEACHING POPULATION GENETICS

4.1. INTRODUCTION

4.1.1. MENDELIAN INHERITANCE AND THE HARDY-WEINBERG LAW

In 1865 Gregor Mendel, a Moravian monk, conducted an experiment in crossing peas of two different true-breeding lines. He found that crossing yellow-seeded peas with a green-seeded variety always gave yellow-seeded offspring; but these offspring when crossed with one another could themselves give rise to three different types of pea. As well as yellow-seeded peas like themselves they could produce true breeding yellow or green-seeded offspring. Mendel quantified his findings and propounded a hypothesis which explained his observations. His proposals for the mechanisms of inheritance were astonishingly accurate and the basis of the theory of inheritance of qualitative characteristics is due to him.

Mendel postulated, and it is now accepted to be true in many instances, that in each adult pea plant there are 2 entities (genes) determining each recognisable qualitative hereditary factor. If they are identical the plant is true breeding for that characteristic; such a plant is said to be a homozygote in the gene concerned. When the 2 genes are different the organism is called a heterozygote; sometimes a heterozygote has characteristics intermediate between the two homozygotes (semidominance), but in Mendel's peas as far as seed colour was concerned the heterozygotes behaved as did the yellow-seeded homozygotes. The gene giving rise to the yellow-seeded characteristic is said to be dominant over the green seed producing one which is recessive.

The genetic construction of an organism is termed its genotype, while the appearance to which this leads is its phenotype. The yellow-seeded variety of pea produced by crossing the two strains is genotypically different from the parental yellow-seeded variety, but is is phenotypically the same. Fig. 4.1. (page 32) represents the situation which Mendel observed.

In reproduction an offspring receives one gene from each parent, each parental gene having an equal chance of being transmitted. The parental zygotes split to form gametes and two of these combine to form the offspring zygote. The genetic information is carried on the chromosomes in the cell nucleus. One site on a chromosome is known as a locus, and the alternative forms of gene at a locus are called alleles. If more than one allele is present in a population the system is said to be a polymorphism.

Most of population genetics deals with the study of polymorphisms; the first important result is the Hardy-Weinberg law, put forward independently in 1908 by G.H. Hardy and

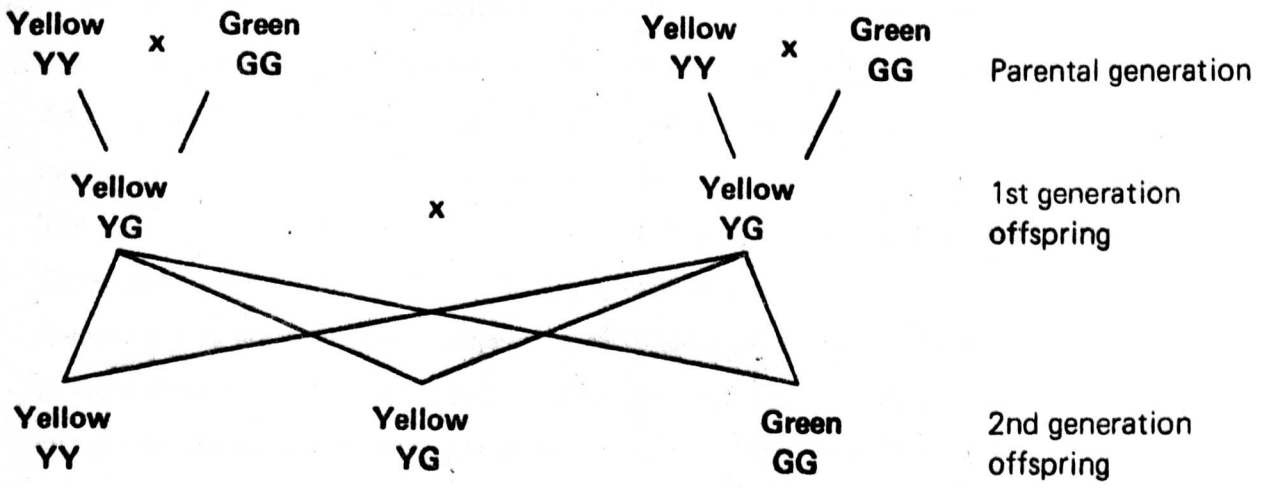


Figure 4.1. The inheritance of seed colour in peas observed by Mendel.

W. Weinberg. It deals with the situation at an autosomal locus – i.e. with genes not carried on the sex chromosome. Characteristics transmitted on this chromosome are sex-linked; in the autosomal case it is immaterial from which parent a gene was received. If at a diallelic locus the frequency (i.e. the proportion in the population) of genes A and a are p and q (= 1–p) respectively, the expected genotype frequencies in a (large) population are: AA, p^2 ; Aa, 2pq; aa, q^2. If this population undergoes panmixia (i.e. if mating is at random) the mating frequencies for the various genotypes are those given in table 4.2.(a) (page 34). The entries in this table may be partitioned to give the frequency of offspring genotypes. These are shown in table 4.2.(b). The Hardy-Weinberg law is the assertion that the gene frequencies in the succeeding generation are unaltered; and the polymorphism is said to be in Hardy-Weinberg equilibrium. In fact, a large population will attain equilibrium after only one generation of random mating irrespective of the initial gene frequencies.

4.1.2. DRIFT, SELECTION, MIGRATION AND MUTATION

While the determination of whether or not a polymorphism is in Hardy-Weinberg equilibrium is a necessary step forward in the investigation of a locus, attention is most often focussed on the investigation of reasons for the absence of this situation. It may not obtain for example because of non-random mating patterns, the smallness of the population so that variation due to the sampling of gametes in the reproductive process assumes importance, or because some form of directional force is acting on the gene frequencies.

In section 4.2.2. the case will be discussed of a diallelic locus at which gene frequencies fluctuate because of random sampling of gametes; such a process being termed random genetic drift, or the Sewall Wright effect in honour of Wright's paper on the subject in 1931 [18] although Fisher had discussed the phenomenon in 1922 [6]. Under drift one of the genes must eventually be lost to the population due to the workings of chance, so that all members of the population will be homozygotic. In section 4.2.3. those forces which can keep a polymorphism in equilibrium in the presence of sampling variation are examined. One such force is selection, wherein the parental genotypes are unequally favoured in their ability to be represented in the next generation whether by reason of longevity or superior fertility. The propensity of an organism of a particular genotype to have offspring is called the fitness of the genotype. This quantity is measured in relation to the other genotypes. Another pressure which can counter the effects of drift is migration; the influx to a population of new members whose mean genetic make-up differs from the current population mean can also give rise to an equilibrium situation. Mutation, or the development in an individual of a non-inherited gene, is very rare. Its result in terms of gene frequency is the same as an immigrant would produce, so that whenever a model is constructed which takes migration into account it may also be

(a)

Mating frequencies for the different genotypes of a diallelic locus under panmixia; frequencies of A and a are p and q respectively

		♀ AA	♀ Aa	♀ aa
♂	AA	p^4	$2p^3q$	p^2q^2
♂	Aa	$2p^3q$	$4p^2q^2$	$2pq^3$
♂	aa	p^2q^2	$2pq^3$	q^4

(b)

The frequency of offspring genotype for a diallelic locus under panmixia, from table 4.2 (a)

AA offspring

		♀ AA	♀ Aa	♀ aa
♂	AA	p^4	p^3q	–
♂	Aa	p^3q	p^2q^2	–
♂	aa	–	–	–

Total: $p^2(p^2+2pq+q^2)=p^2$

Aa offspring

		♀ AA	♀ Aa	♀ aa
♂	AA	–	p^3q	p^2q^2
♂	Aa	p^3q	$2p^2q^2$	pq^3
♂	aa	p^2q^2	pq^3	–

Total: $2pq(p^2+2pq+q^2)=2pq$

aa offspring

		♀ AA	♀ Aa	♀ aa
♂	AA	–	–	–
♂	Aa	–	p^2q^2	pq^3
♂	aa	–	pq^3	q^4

Total: $q^2(p^2+2pq+q^2)=q^2$

Figure 4.2.

assumed to cover mutation.

Because selection is effective through the medium of disease its action is the most important of the directional forces. The models which will be discussed have been designed so that an attempt can be made to estimate the genotype fitnesses if the gene frequency distribution is known. (The phrase "gene frequency distribution" is slightly unfortunate. It is used for "gene frequency frequency distribution" or "gene frequency probability distribution".)

4.1.3. MULTIPLE ALLELES AND THE ABO SYSTEM

Another complication in the study of the inheritance of qualitative factors which renders the models mentioned so far inappropriate, is that there are frequently more than two alternative forms of gene at a given locus. One important system in which this is so is the ABO blood group system. Knowledge about this system has application in blood transfusions, establishment of paternity, and the investigation of several diseases, the typing of the blood being carried out by the agglutination of red blood cells. There are 3 alleles denoted A, B and O at the locus. A and B are codominant and each is dominant over O, so that the 6 genotypes possible give rise to 4 phenotypes:– O (genotype OO), A (AA or OA), B (BB or OB) and AB (AB). For a locus with 4 alleles there are 10 genotypes and for the general system with k alleles there are ½k (k+1) genotypes.

Several studies, notably by Matsunaga [13] and by Cohen and Sayre [3], have shown that selection is acting at the ABO locus. Livingstone [11] has pointed out that it will be extremely difficult to associate any of the major diseases of history such as plague or smallpox with a particular blood group, and yet it is largely these diseases which have given rise to the different gene frequency distributions which exist in various parts of the world today. Nevertheless it may be fruitful to obtain estimates of the degree of selection which has occurred and the directions in which it was acting. Section 4.4.3. contains a discussion of a method of estimating ABO gene frequencies from observed phenotype frequencies and of how these may be used in conjunction with a stochastic model incorporating selection and migration to investigate those factors.

4.1.4. PREVIOUS MODELS OF GENETIC SYSTEMS

From the simple algebraic model of Hardy and Weinberg the simulation of genetic systems has grown in sophistication at a great pace. Haldane, Fisher and Wright in many papers, in the 1930's and 1940's especially, produced a plethora of mathematical models, and recently Kimura [8] has published many new formulations, mainly basing them on the Fokker-Planck

diffusion equation which will be dealt with in section 4.2.1. The basic parameter of any genetic system is the frequency of the relevant genes and because inheritance is carried out in a probabilistic manner it is pertinent to consider the gene frequency distribution. In the case of a diallelic locus under drift, for example, it is appropriate to find the probability distribution for the gene frequency x after t generations if the frequency in the original parent generation was p. Stochastic equation models are therefore very powerful tools for the investigation of genetic systems, but Monte-Carlo simulations have also been used with some success. Among the first to use such methods in genetics were Fraser [7], in Australia, and Martin and Cockerham [12] in America. These workers saw in Monte-Carlo methods an opportunity of examining the workings of systems through varying their parameters. The methods form a useful adjunct to the mathematical analyses, modelling greater detail in a more realistic manner, but they are tedious to use and their findings are more difficult to interpret, so that they are not well adapted to use in teaching, nor in an interactive environment. These points will be discussed more fully in section 4.6.

4.1.5. TEACHING POPULATION GENETICS

The dispersive effect of random genetic drift and its prevention by directional forces such as selection and migration may prove difficult phenomena to communicate to students of genetics. It is often impossible within a given syllabus to use suitable experimental material to illustrate population behaviour; the mathematics involved in the theoretical treatments in textbooks and research articles is sufficiently abstruse to deter almost all students, and the results contained in such sources are necessarily incomplete. A compromise solution was found by Crosby [4] who used Monte-Carlo simulation techniques, but even with a computer results in a comprehensible form do not become available sufficiently quickly, since it takes many successive simulations to build up a probability distribution. Arguably the time spent in this pursuit is not wasted as it can give the students an appreciation of what a probability distribution actually represents, but for complicated systems with many parameters the ability to investigate the effects and interactions of these parameters is lost.

An alternative approach is to teach students about probability distributions by means of a quincunx or picking coloured balls from a bag, and then to model the genetic system by the Fokker-Planck diffusion equation. In this way it is possible to investigate the behaviour of a diallelic locus under drift, to demonstrate the equilibrium which is produced by some values of genotype fitnesses, and to show the effects of altering population size or permitting immigration. Students should also be able to understand the influences exerted by the various factors at a locus where three alleles are present.

If the probability distribution found as the solution of the relevant diffusion equation is drawn directly on a graphical display unit by the computer program, intelligible results are available very quickly after the program has been given the necessary data on fitnesses, migration, population size and so on. This approach should prove much more satisfactory than presenting the student with an account like that given in section 4.1.2. or trying to explain the mathematics involved in a model.

4.2. THE MODELS

4.2.1. THE FOKKER-PLANCK EQUATION

In a finite population the changes in gene frequency which take place do so as a result of discrete events. It is nevertheless convenient to treat the gene frequency as a continuous stochastic variable and model the changes in it by a diffusion equation. If the population comprises N individuals the frequency of any gene may take any value of the form j/2N for j = 0, 1, 2, ..., 2N, and so if N is large the continuous approximation may be expected to be good.

The appropriate differential equation is the Fokker-Planck equation. Its most general form is

$$\frac{\partial \phi}{\partial t} = \tfrac{1}{2} \sum_{j=1}^{k-1} \frac{\partial^2}{\partial x_j^2} \{ V(\delta x_j)\, \phi \} + \tfrac{1}{2} \sum_{\substack{i=1 \\ i \neq j}}^{k-1} \sum_{j=1}^{k-1} \frac{\partial^2}{\partial x_i\, \partial x_j} \{ W(\delta x_i, \delta x_j)\, \phi \} - \sum_{j=1}^{k-1} \frac{\partial}{\partial x_j} \{ M(\delta x_j)\, \phi \} \quad (*)$$

where the notation used is as follows:

$\phi(x_1, x_2, \ldots\ldots, x_k; t)$ is the joint probability density function, t generations after some initial state, for the gene frequencies x_i of the k alleles A_i at an autosomal locus; $x_1, x_2, \ldots\ldots, x_{k-1}$ are taken as the independent variables, with $x_k = 1 - \sum_{i=1}^{k-1} x_i$.

δx_j is the change in x_j per generation;

$M(\delta x_i)$ is the expected value of δx_i;

$V(\delta x_i)$ is the variance of δx_i;

$W(\delta x_i, \delta x_j)$ is the covariance of δx_i and δx_j.

The equation only deals with the case in which all k alleles are present, i.e. $x_i \neq 0$, $1 \leq i \leq k$.

The first two terms on the right hand side of equation (*) give the rate of change with respect to time of the gene frequency distribution due to random fluctuations, while the third term represents the rate of change due to systematic pressure.

For the simplest case of a diallelic locus the equation reduces to

$$\frac{\partial \phi}{\partial t} = \tfrac{1}{2} \frac{\partial^2}{\partial x^2} \{ V(\delta x)\, \phi \} - \frac{\partial}{\partial x} \{ M(\delta x)\, \phi \}, \text{ for gene frequency } x.$$

4.2.2. A DIALLELIC LOCUS UNDER RANDOM GENETIC DRIFT

If there are no systematic pressures on the gene frequency x the expected change per generation is zero; and if the only source of variation is random sampling of gametes (in a population of effective size N) the variance is $\frac{1}{2N}\, x\,(1-x)$. The Fokker-Planck equation simplifies to

$$\frac{\partial \phi}{\partial t} = \frac{1}{4N}\,\frac{\partial^2}{\partial x^2}\{x\,(1-x)\,\phi\}$$

It is required to find the gene frequency probability distribution after t generations, $\phi(x, t; p)$, given that the initial frequency (at generation 0) was p.

i.e. $\phi\,(x, 0; p) = \delta\,(x - p)$ (Kronecker delta.)

Kimura [8, 9] has given the solution as

$$\phi\,(x, t; p) = \sum_{i=1}^{\infty} p(1-p)\, i\,(i+1)\,(2i+1)\, F(1-i, i+2, 2, p)\, F(1-i, i+2, 2, x) e^{-i(i-1)\,t/4N}$$

for $0 < x < 1$, where $F(.,.,.,.)$ is the hypergeometric function. Because of the exponential term the series converges reasonably quickly and a computer solution may be obtained. The probability that a gene becomes lost from the population may be calculated by numerical quadrature, or from Legendre polynomials using another of Kimura's results [8]. The computer program which evaluates the distribution for given values of N, p and t is in Appendix 3. The distribution is given on the Computek in graphical form as illustrated in fig. 4.3. (page 44).

Use has been made of the phrase "effective size" in talking about a population. This is a figure which depends on the age and sex structure of the population, on the mating pattern, and on the history of these factors. Essentially it is the average number of individuals in a population who contribute genes to the next generation. Almost without exception in models it is taken to be a constant with respect to time. Any temporary decrease in the actual population size has the effect of lowering the effective size thereafter [10, page 323].

4.2.3. A DIALLELIC LOCUS UNDER SELECTION AND MIGRATION

Before investigating the effects of selection it is necessary to decide how selection is to be represented in the diffusion equation model. The method which has been chosen is to ascribe weights (fitnesses) to the individual genotypes so that the ratios of the number of offspring of the genotypes AA, Aa, aa which themselves reach reproductive age are $w_1 : w_2 : w_3$. The "mean fitness" $\bar{w}$ is then given by $\bar{w} = w_1 x^2 + 2w_2 xy + w_3 y^2$ where x is the frequency of A and y ($= 1 - x$) is the frequency of a. The actual values chosen for the w_i are immaterial as long as the ratios are unaltered, but it is sometimes convenient to take the fitness of a genotype to be 1 minus its selective coefficient. This quantity is defined in appendix 1.

Li [10, p. 338] gives the probability density function for the gene frequency x at a diallelic locus in a population of effective size N, which is in equilibrium, as $\phi(x) = \frac{C\bar{w}^{2N}}{x(1-x)}$ where C is a constant such that $\int_0^1 \phi(x)\,dx = 1$. It is important to remember that both alleles must be present before the differential equation is a valid approximation, so that the solution also is only valid in the range $^1/2N \le x \le 1 - {}^1/2N$.

The corresponding expression if migration also occurs is $\phi(x) = \frac{C'\,\bar{w}^{2N}}{x^{1-4Nm\bar{x}}(1-x)^{1-4Nm(1-\bar{x})}}$.

Migration of population is adequately modelled for present purposes by assuming that a population in which the gene frequency is x exchanges members at random at a rate m per generation with a population where the frequency is $\bar{x}$. For given values of N, m, $\bar{x}$, w_1, w_2 and w_3 the distribution may be plotted on the Computek almost immediately.

The term involving $V(\delta x)$ in the differential equation is as in the previous section. $M(\delta x)$ is calculated assuming additivity of the effects of selection and migration. The most interesting case of selection is when the fitness of the heterozygote is superior to that of either of the homozygotes (overdominance). In this instance a stable polymorphism is possible. It is convenient to write $w_1 = 1 - s_1$, $w_2 = 1$, $w_3 = 1 - s_3$. (There is no loss of generality here as s_1 and s_3 may be negative if heterozygote advantage does not apply.) If x and y are the frequencies of A and a in one generation then the frequencies in the next are:

AA: $(1 - s_1)\,x^2$ Aa: $2xy$ aa: $(1 - s_3)\,y^2$.

The change in frequency between generations is therefore

$$\frac{1}{\bar{w}}\{(1-s_1)\,x^2 + xy\} - x = \frac{xy}{\bar{w}}(s_3y - s_1x).$$

It is worth noticing that Kimura, and others, in this situation take $M(\delta x) = xy\,(s_3y - s_1x)$, [8], an approximation which is valid only for small s_1, s_3. Since the case of lethal recessives ($s_3 = 1$) is also of interest in the present context it is undesirable to make this approximation. In any case it is pointless to do so when a computer is to be used, and the fact that $s_3y - s_1x$ is the derivative $\frac{d\bar{w}}{dx}$ means too that many theoretical treatments of the diffusion equation become not simpler but more difficult if the approximation is used.

The expected change in x due to migration is $m(\bar{x} - x)$.

4.2.4. A TRIALLELIC LOCUS UNDER SELECTION AND MIGRATION

Wright [19] has given the general expression for the equilibrium gene frequency distribution for a k-allelic locus under selection and migration. A derivation of this rather neglected but very important formula, from the diffusion equation, is given in the next section. It is

$$\phi(\underline{x}) = \frac{C\bar{w}^{2N}}{\prod_{i=1}^{k} x_i^{1-4Nm\bar{x}_i}}$$

extending the notation for the diallelic locus in a natural manner.

For the triallelic case (k = 3), if w_{ij} is the fitness of genotype A_iA_j (and $w_{ij} = w_{ji}$ for an autosomal locus) the nine parameters are N, m, $\bar{x}_1$, $\bar{x}_2$, and 5 of w_{11}, w_{12}, w_{13}, w_{22}, w_{23} and w_{33}. The system is therefore relatively complicated and gives rise to many results which, without the proposed interactive model have been difficult or impossible to communicate to students.

The probability distribution in this case is a bivariate distribution in the triangular region $0 < x_1 < 1$, $0 < x_2 < 1$, $0 < x_1 + x_2 < 1$. It was decided to represent this by a contour diagram, orientating the triangle in the manner adopted by Brues [2] in her investigation of the ABO locus. This method of depicting the distribution was regarded as better than that in which an equilateral triangle is used because students could be expected to be more familiar with rectangular axes. These alternatives are illustrated in figs. 4.5 and 4.6 (pages 46 and 47).

4.2.5. A DERIVATION OF THE EQUILIBRIUM GENE FREQUENCY DISTRIBUTION FOR A K-ALLELIC LOCUS

The general form of the Fokker-Planck equation was given in section 4.2.1. The steady state solution is required. Kimura [8] has introduced the idea of "probability flux"; let the probability flux along the x_i axis be $P(x_i)$. Then

$$P(x_i) = \tfrac{1}{2} \frac{\partial}{\partial x_i} \{ V(\delta x_i)\, \phi \} + \tfrac{1}{2} \sum_{\substack{j=1 \\ j \neq i}}^{k-1} \frac{\partial}{\partial x_j} \{ W(\delta x_i, \delta x_j)\, \phi \} - M(\delta x_i)\, \phi$$

For a steady state solution $\frac{\partial \phi}{\partial t} = 0$ and the equation to be solved is therefore

$\sum_{i=1}^{k-1} \frac{\partial}{\partial x_i} P(x_i) = 0$. The solution which is of interest is the "zero flux" equilibrium solution which, if it exists, is given by $P(x_i) = 0$ for $1 \leq i \leq k-1$.

If the expected change in x per generation due to selection is ${}^{s}M(\delta x_i)$, an argument corresponding to that in section 4.2.3. gives ${}^{s}M(\delta x_i) = \frac{1}{\bar{w}} \sum_{j=1}^{k} w_{ij} x_i x_j - x_i$.

If there is also migration, and the population exchanges members at random with a population in which the gene frequency is $\bar{x}$ at the rate of m per generation, the expected change in x_i due to migration may be written.

${}^{m}M(\delta x_i) = m\,(\bar{x}_i - x_i)$.

If the variance of δx_i arises only because of random sampling of gametes, then for a

population of effective size N,

$$V(\delta x_i) = \frac{1}{2N} x_i (1 - x_i) \text{ and } W(\delta x_i, \delta x_j) = \frac{-1}{2N} x_i x_j$$

The equation $P(x_i) = 0$ therefore becomes

$$\frac{\partial}{\partial x_i} \{x_i(1-x_i)\phi\} - x_i \sum_{\substack{j=1 \\ j\neq i}}^{k-1} \frac{\partial}{\partial x_j} \{x_j \phi\} = \{\frac{4Nx_i}{\bar{w}} \sum_{j=1}^{k} w_{ij}x_j - 4Nx_i + 4Nm (\bar{x}_i - x_i)\}\phi$$

which reduces to

$$x_i(1-x_i) \frac{\partial \phi}{\partial x_i} - x_i \sum_{\substack{j=1 \\ j\neq i}}^{k-1} x_j \frac{\partial \phi}{\partial x_j} = \{ \frac{4Nx_i}{\bar{w}} \sum_{j=1}^{k} w_{ij} x_j - 4Nx_i + kx_i + 4Nm (\bar{x}_i - x_i) - 1\}\phi \quad ..(1)$$

In order to solve this set of equations it is convenient to consider the matrix H defined by

$$H_{ij} = x_i x_j, \; i \neq j \; ; \; H_{jj} = x_j (1 - x_j)$$

H is proportional to the variance-covariance matrix of δx_i and δx_j. It is a symmetric non-singular square matrix of order $k - 1$, having determinant $\prod_{i=1}^{k} x_i$. Its inverse, K, may be shown to be given by

$$K_{ij} = \frac{1}{x_k} , \; i \neq j ; \; K_{jj} = \frac{1}{x_k} + \frac{1}{x_j} .$$

Since equation (1) is equivalent to $H \frac{\partial \phi}{\partial \underline{x}} = \phi \underline{f}$

where $\frac{\partial \phi}{\partial \underline{x}} = (\frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2}, \ldots, \frac{\partial \phi}{\partial x_{k-1}})'$ and $\underline{f}$ is a $(k - 1)$ vector with

$$f_i = \frac{4Nx_i}{\bar{w}} \sum_{r=1}^{k} w_{ir}x_r + kx_i - 4Nx_i + 4Nm (\bar{x}_i - x_i) - 1, \text{ it follows that } \frac{\partial \phi}{\partial \underline{x}} = \phi K \underline{f} \quad ...(2).$$

Now $(Kf)_j = \sum_{i=1}^{k-1} K_{ij} f_i = \frac{1}{x_k} \sum_{i=1}^{k-1} f_i + \frac{f_j}{x_j}$

$$= \frac{1}{x_k} \frac{4N}{\bar{w}} \{ \bar{w} - x_k \sum_{r=1}^{k} w_{kr} x_r \} + \frac{1}{x_k} \{(k-4N-4Nm)(1-x_k) + 4Nm (1-x_k) + 1-k\}$$

$$+ \frac{4N}{\bar{w}} \sum_{r=1}^{k} w_{jr} x_r + (k-4N-4Nm) + \frac{4Nm\bar{x}_j - 1}{x_j}$$

$$= \frac{4N}{\bar{w}} \sum_{r=1}^{k} (w_{jr} - w_{kr}) x_r + \frac{1-4Nm\bar{x}_k}{x_k} - \frac{1-4Nm\bar{x}_j}{x_j}$$

Furthermore $\sum_{r=1}^{k} (w_{jr} - w_{kr}) x_r = \tfrac{1}{2} \frac{\partial \bar{w}}{\partial x_j}$, and, writing $a_j = 1-4Nm\bar{x}_j$ for convenience

leads to $(Kf)_j = 2N \frac{\partial}{\partial x_j} (\log \bar{w}) + \frac{a_k}{x_k} - \frac{a_j}{x_j}$.

Therefore equation j from the system (2) is equivalent to

$$\frac{\partial \phi}{\partial x_j} = \{2N \frac{\partial}{\partial x_j} (\log \bar{w}) + \frac{a_k}{x_k} - \frac{a_j}{x_j}\}\phi \qquad ...(3).$$

If $G_j(\underline{x})$ is an arbitrary function of $x_1, x_2, \ldots, x_{k-1}$ but does not involve x_j, then integration gives

$$\log \phi = \log \bar{w}^{2N} - \log x_k^{\alpha_k} - \log x_j^{\alpha_j} + \log G_j(\underline{x})$$

$$\text{i.e.} \quad \phi = G_j(\underline{x}) \frac{\bar{w}^{2N}}{x_k^{\alpha_k} x_j^{\alpha_j}}$$

Since this must hold for $j = 1, 2, \ldots, k-1$ the only possible solution is $\phi = \frac{C\bar{w}^{2N}}{\prod_{i=1}^{k} x_i^{\alpha_i}}$ where C is chosen to make $\int_R \phi(\underline{x})\, d\underline{x} = 1$ for the region $R : 0 < x_i < 1, (1 \leq i \leq k-1)$, and $0 < \sum_{i=1}^{k-1} x_i < 1$. This is the result used in section 4.2.4.

Equation (3) leads directly to the conclusion that, if migration is not a major factor, modal equilibrium gene frequencies for a large population are given by $\phi \frac{\partial}{\partial x_j} (\log \bar{w}) = 0$, or more simply $\frac{\partial \bar{w}}{\partial x_j} = 0$, $j = 1, 2, \ldots, k-1$; which is the condition for the average fitness to be a maximum. This result is well-known, but the diffusion equation approach is a powerful method for producing extensions of well-known results such as the conditions for stability under selection and migration, which can also be deduced from equation (3).

4.3. OUTPUT FROM THE MODELS

The two models of the diallelic locus were programmed in ALGOL W with FORTRAN graph plotting routines linked in. The program listings are in appendix 3. Neither of the programs presents any computational difficulties, except that using small values for t/N in the first, say less than 0.1, made the program rather slower than desirable. Similarly for values of the gene frequency initially close to 0 or 1 computation is increased because of slow convergence of the series to be evaluated. For these cases the number of function evaluations was reduced. Most of the coding is concerned with the format of the output.

Six runs of the model of the diallelic locus under drift are shown in fig. 4.3. (page 44). The graphs are essentially the same as those given by Kimura, but it must be emphasised that in a teaching situation the parameters for successive runs may be decided upon by the teacher in order to illustrate any feature which is causing difficulty, or they may be chosen by the students. It is important that each graph should be scrupulously annotated to prevent misunderstandings due to differences of scale or confusion between models. As well as the graphical output the mean gene frequency and the standard deviation were printed on the 2741 terminal for the students' convenience. Points which may be brought out clearly using

the model include the dependence of the distribution upon the number of generations elapsed and the effective population size only through their quotient; the increasing probability of gene loss (or alternatively fixation) with time and with the departure of the initial frequency from 0.5; and the inevitability of final fixation.

Fig. 4.4. (page 45) shows the results of 6 runs of the model of the diallelic locus in equilibrium under directional pressures, which illustrate various effects of selection and migration. Again teachers and students should work out between them just which values of the parameters must be considered next so as to make a specific point. In the course of a one-hour session it would be quite feasible to draw some 20 such graphs, from which a great deal of information can be derived. The effects of lethal recessives can be investigated, and gametic selection can be discussed as well as superior heterozygote fitness, but it must be remembered that the model caters for those populations in which both genes are present. It may be thought instructive to show that the formula given by Li [10, page 339], which seems to indicate that only through the product of the population size and the heterozygote advantage is selection acting, is an approximation which should not be accepted as a general result.

The model of the triallelic locus under selection and migration was programmed in FORTRAN. Fig. 4.5. (page 46) shows the Computek output for 4 runs. The topics which can be covered using this model include the effect of population size and migration rate, as with two genes. Conditions on the genotype fitnesses for the stability of an equilibrium situation, which now become more involved, can be investigated. It is in the triallelic case that the use of such a model in an interactive context has the greatest advantage over other media in giving students insight into the system. In the simpler case of a diallelic locus the use of the technique described is convenient and useful, but it cannot be denied that the facts can be adequately taught without such a method. To understand the various ramifications of an alteration in parameters in the triallelic case, however, a quick method of producing intelligible results is imperative. Several alternatives were possible for the format of the output. Numerical output like that in fig. 4.8. (page 51) could have been given, but although such a table is useful on occasion it does not give an easily understood overall impression of the situation. Alternatively some attempt could have been made to draw a three dimensional representation of the probability distribution, but this presents so many difficulties that it is unlikely that a satisfactory, informative, unambiguous diagram would have resulted. A contour diagram of the distribution is a convenient compromise between these two extremes. An equilateral triangle could have been used as axes for the 3 gene frequencies, and an example is given in fig. 4.6. (page 47), but since the model was to be used in association with an

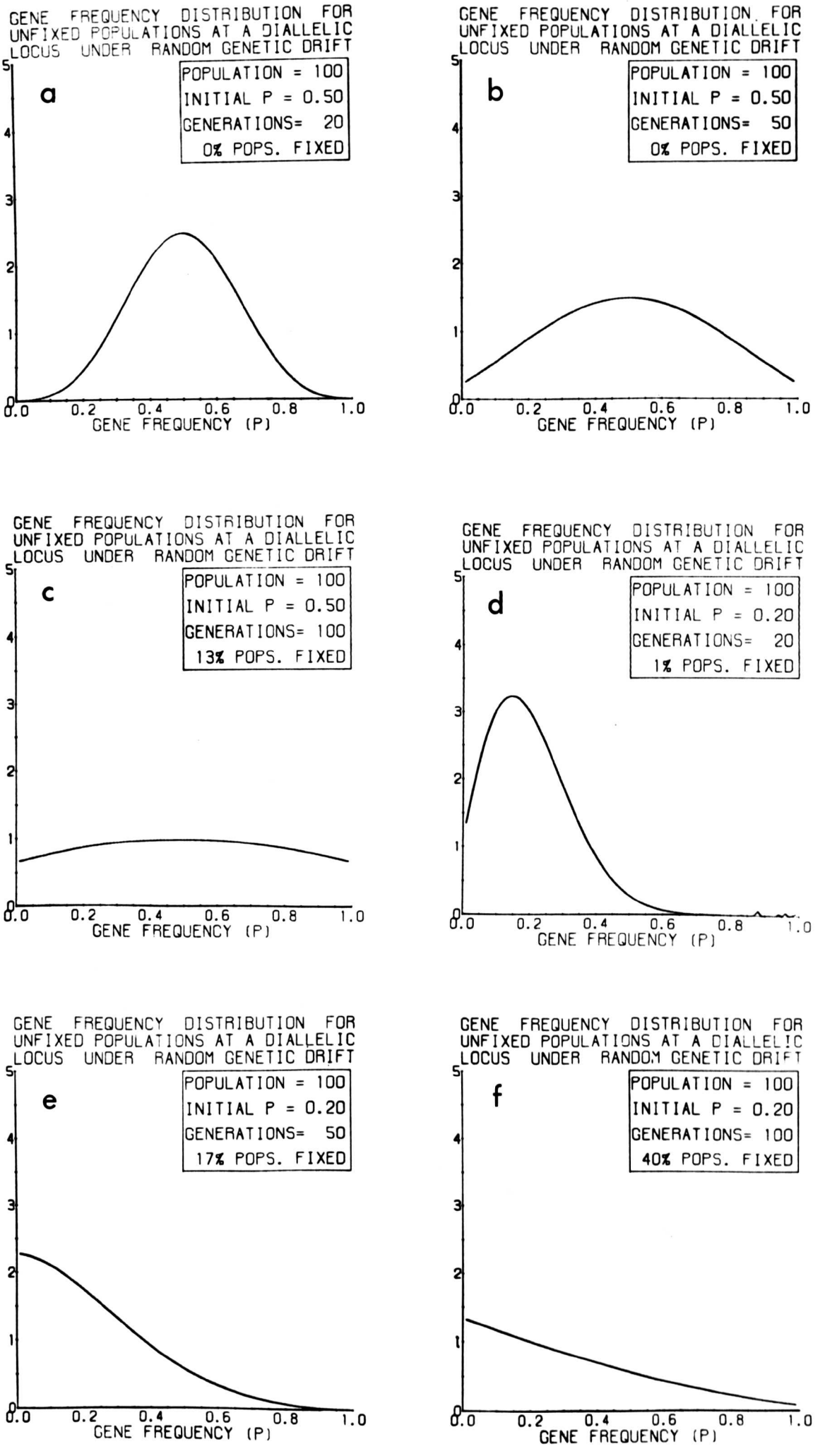


Figure 4.3. Sample output from the model of a diallelic locus under drift.

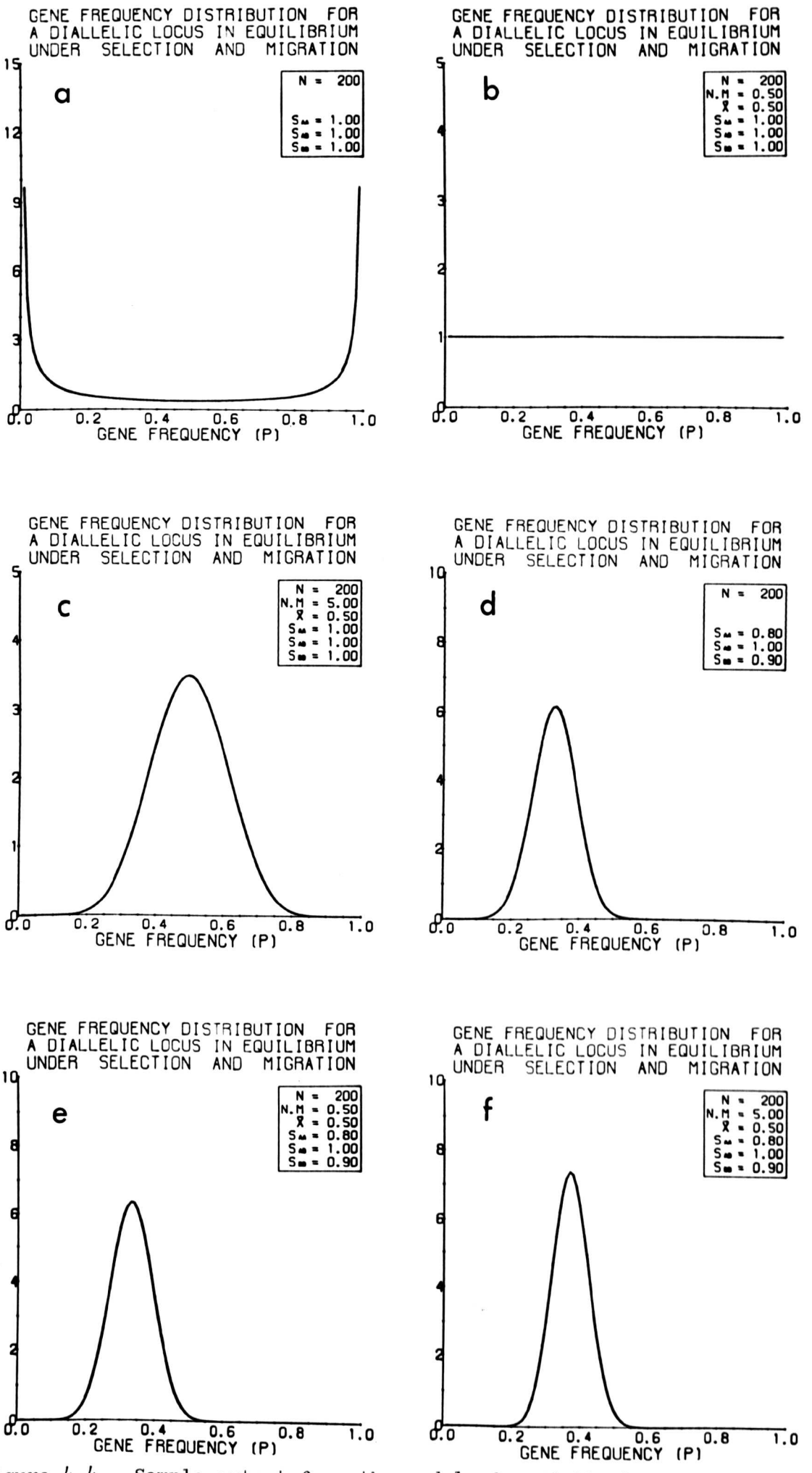


Figure 4.4. Sample output from the model of a diallelic locus under pressure.

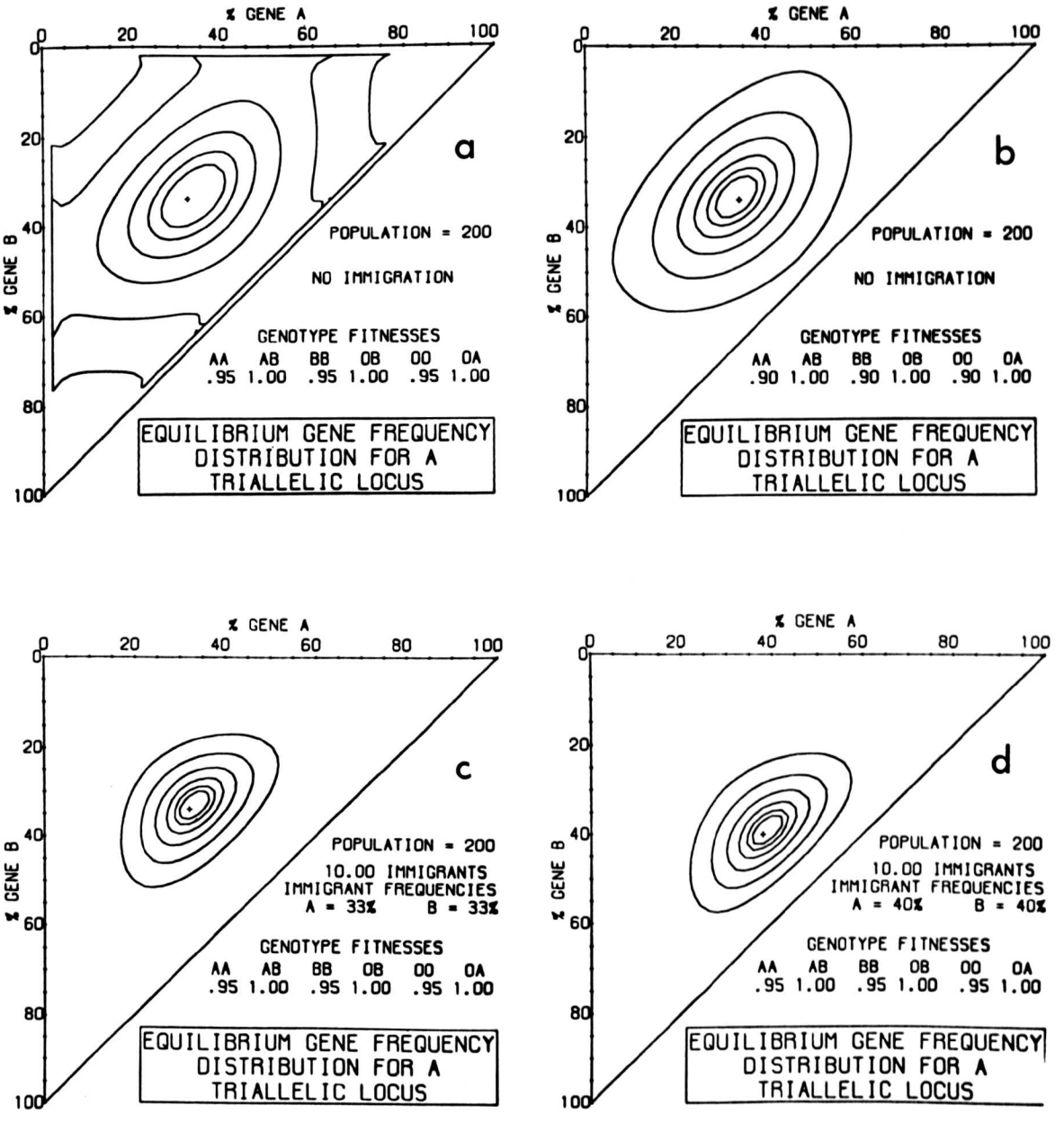


Figure 4.5. Sample output from the model of a triallelic locus under pressure.

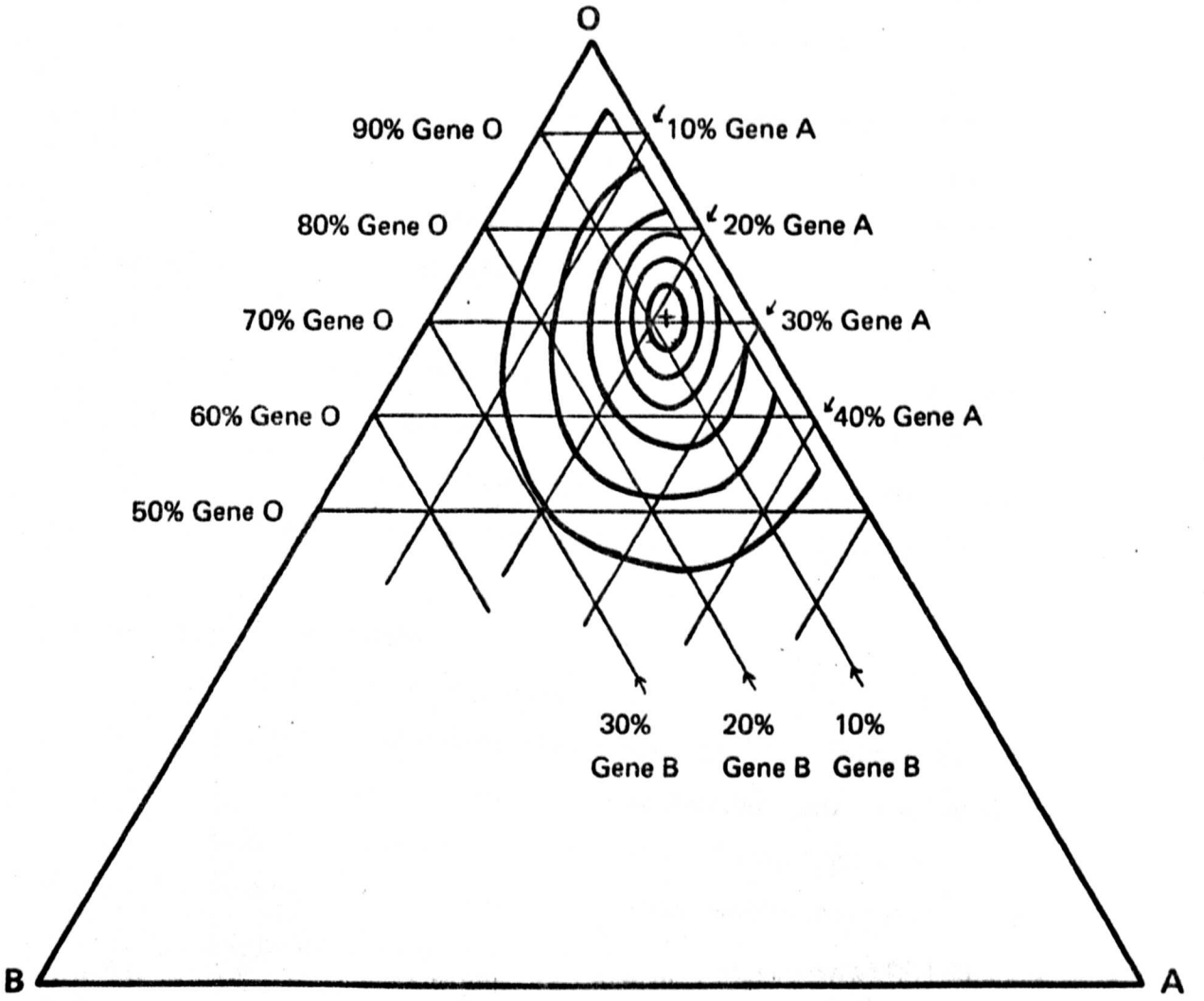


Figure 4.6. An alternative representation of the gene frequency distribution at a triallelic locus. The population size is 100 with 5 immigrants per generation from a population with 20% gene A and 10% gene B. Fitnesses of genotypes AA, AB, BB, OB, OO and OA are 0.90, 1.10, 0.90, 1.02, 1.00 and 1.05.

investigation into selection at the ABO locus, and one of the principle papers in the literature is that of Brues [2], it was thought to be appropriate and less confusing for the students to use a rectangular system of axes as was done in that article.

Because the expression for the frequency distribution involves the term $\bar{w}^{2N}$ small changes in the mean fitness, due to small departures from the equilibrium position, can cause large changes in the value of the probability density function. The contours drawn were not therefore equally spaced heights but in fact were chosen so that the representation of the distribution was more acceptable to the eye. The ratio of the maximum value of $\phi(\underline{x})$ to the value at any point was calculated and contours at equal intervals of the log-log of this quantity were drawn. Thus the outer contour of the 6 drawn in fig. 4.5. represents a very low probability indeed. It would have been pleasant to have drawn percentile contours but the calculation involved would have slowed down the response of the program unacceptably.

4.4. VALIDITY OF THE MODELS.

4.4.1. THE DIALLELIC LOCUS UNDER DRIFT

Mathematical models of systems often arouse suspicions in students (and others) that the simplifications necessary to make the model tractable render it so unlike the actuality as to be worthless. It is therefore important that the distributions produced by the models should be demonstrated to bear a reasonable resemblance to those found by observation. Nassar [15] has investigated the agreement between the Fokker-Planck equation solution and a Monte-Carlo model of a diallelic locus under drift and found it to be good even for quite small populations. Due to the time-dependent nature of the system little more can be done to convince students of the adequacy of the model than draw their attention to this finding.

4.4.2. THE DIALLELIC LOCUS UNDER PRESSURE

If certain polymorphisms in human populations are assumed to be in equilibrium they may be used in conjunction with a model to validate it. It is necessary to choose a locus at which selection and migration are known to be effective or the action of the parameters of the model which deal with these aspects will not be examined. The other main requirement is that a large body of data should be available so that the model is adequately tested. Such is the case at the ABO locus. Mourant, Kopec and Domaniewska-Sobczak [14] have collected an extensive quantity of data from published sources, and it is well known that certain diseases are associated with the ABO blood group giving rise to selective pressure. Although the ABO system is inherited via a triallelic locus the B gene is missing in pure-blooded American Indians: it may therefore also be used with the model of the diallelic locus in equilibrium

under selection and migration. Because such parameters as migration rate and population size are not known sufficiently accurately for the populations for which data is available on blood groups, the models cannot legitimately be used to estimate selective intensities for the different genotypes. It is nevertheless important to demonstrate that the results of the model are compatible with observations.

The gene frequency distribution for 106 samples from American Indians taken from the book by Mourant et al was calculated from the phenotype frequencies (see section 4.4.3.2); the distribution is given in fig. 4.7.(a). Numerical and graphical representations of the distribution produced by the model for the parameters specified are in figs. 4.8.(a) and 4.9.(a), (pages 51 – 52). The numerical results were calculated using a two dimensional analogue of the trapezoidal rule. The agreement between the distributions is close enough to indicate that the model is not unreasonable. No attempt was made to obtain a better fit by altering population size or introducing migration, simplicity of approach being considered more important than precision. The differences in requirements in terms of accuracy and simplicity of simulations used in teaching compared with those used in research are dealt with in section 6.1.

4.4.3. THE TRIALLELIC LOCUS

4.4.3.1. ABO BLOOD GROUP DATA AND BERNSTEIN'S METHOD

The data on the ABO blood group system is available in the form of phenotype frequencies. If in a sample of size n there are n_1, n_2, n_3, n_4 members with phenotype O, A, B, AB respectively the likelihood function for the frequencies p, q, r of genes A, B, O is

$$\frac{n!}{n_1!\, n_2!\, n_3!\, n_4!} \quad (r^2)^{n_1} (p^2 + 2pr)^{n_2} (q^2 + 2qr)^{n_3} (2pq)^{n_4} .$$

Letting $\frac{1}{n}(n_1, n_2, n_3, n_4) = (f_1, f_2, f_3, f_4)$ the maximum likelihood equations for p, q and r (subject to $p + q + r = 1$) can be written as

$$\frac{f_2 + f_4}{p} + \frac{f_2}{p + 2r} = \frac{f_3 + f_4}{q} + \frac{f_3}{q + 2r}$$

Simple inefficient estimates are

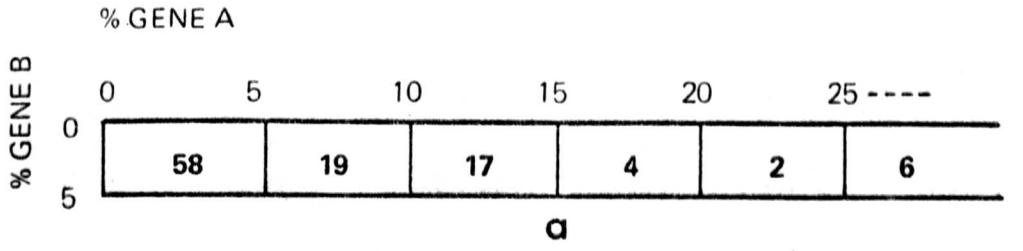
$$r = f_1^{\frac{1}{2}} ; \qquad p = 1 - (f_1 + f_3)^{\frac{1}{2}} ; \qquad q = 1 - (f_1 + f_2)^{\frac{1}{2}} ;$$

and Bernstein [1] has shown that better estimates are given by

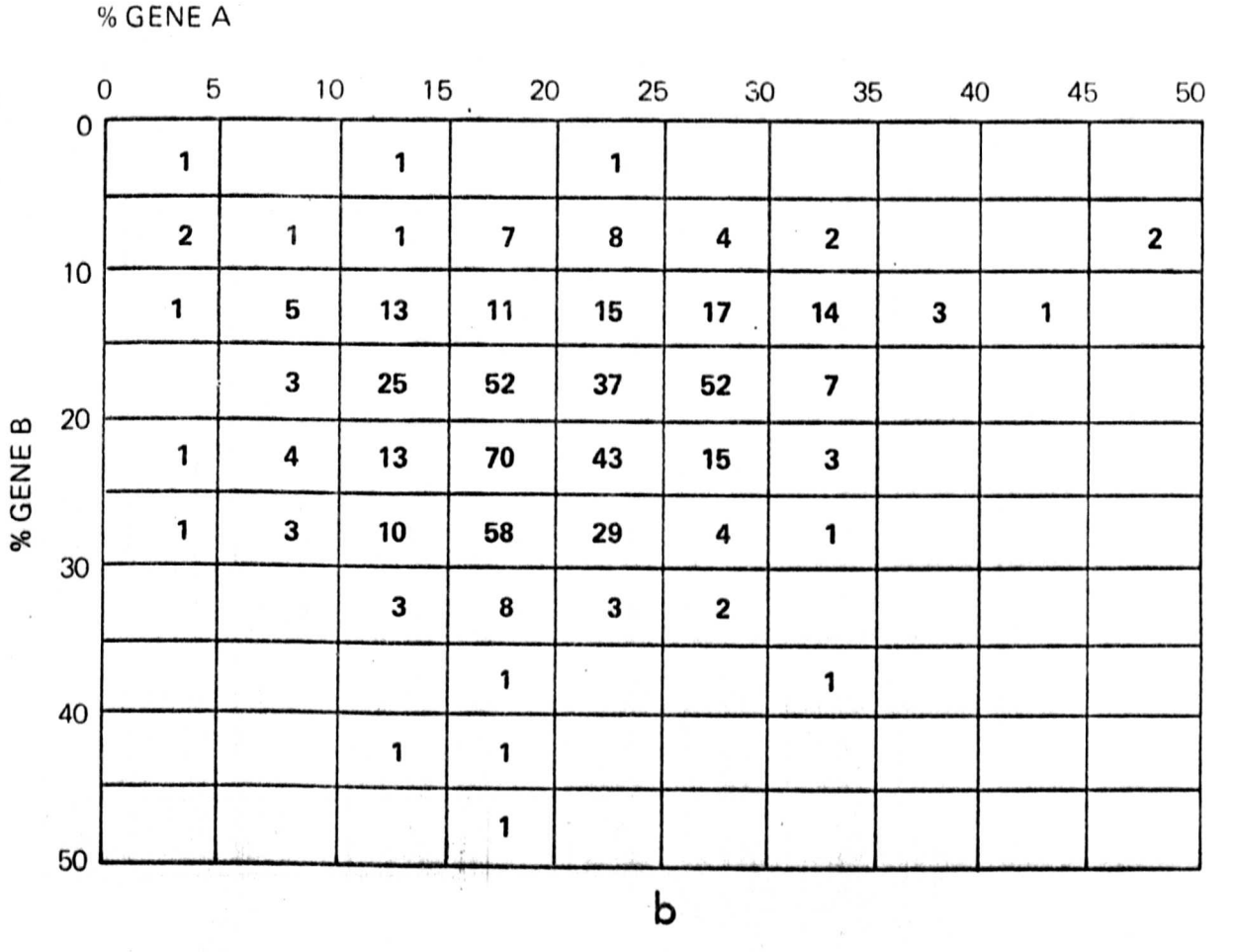
$$p = \{1 - (f_1 + f_3)^{\frac{1}{2}}\}\{1 + D/2\}$$

$$q = \{1 - (f_1 + f_2)^{\frac{1}{2}}\}\{1 + D/2\}$$

$$r = \{f_1^{\frac{1}{2}} + D/2\}\{1 + D/2\} \qquad \text{where } D = (f_1 + f_2)^{\frac{1}{2}} + (f_1 + f_3)^{\frac{1}{2}} - f_1^{\frac{1}{2}} - 1.$$

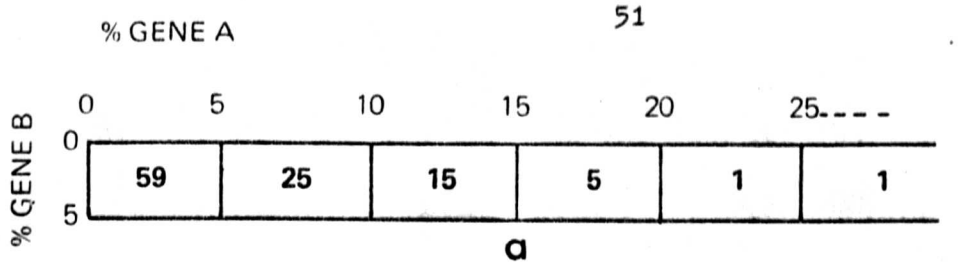


a

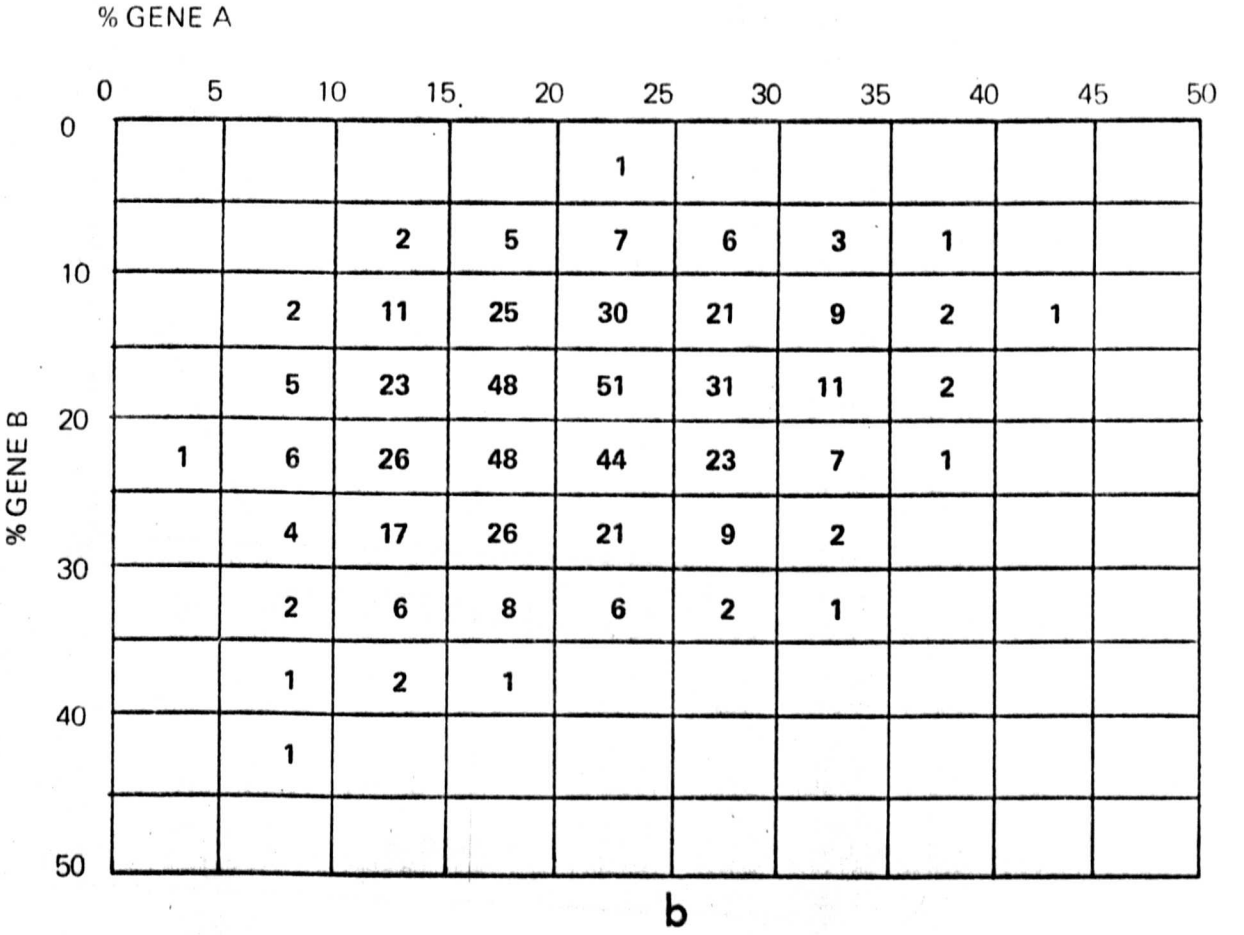


b

Figure 4.7. Gene frequency distributions calculated from observed phenotype frequencies using Bernstein's method for (a) 106 American Indian samples and (b) 562 Asian samples.

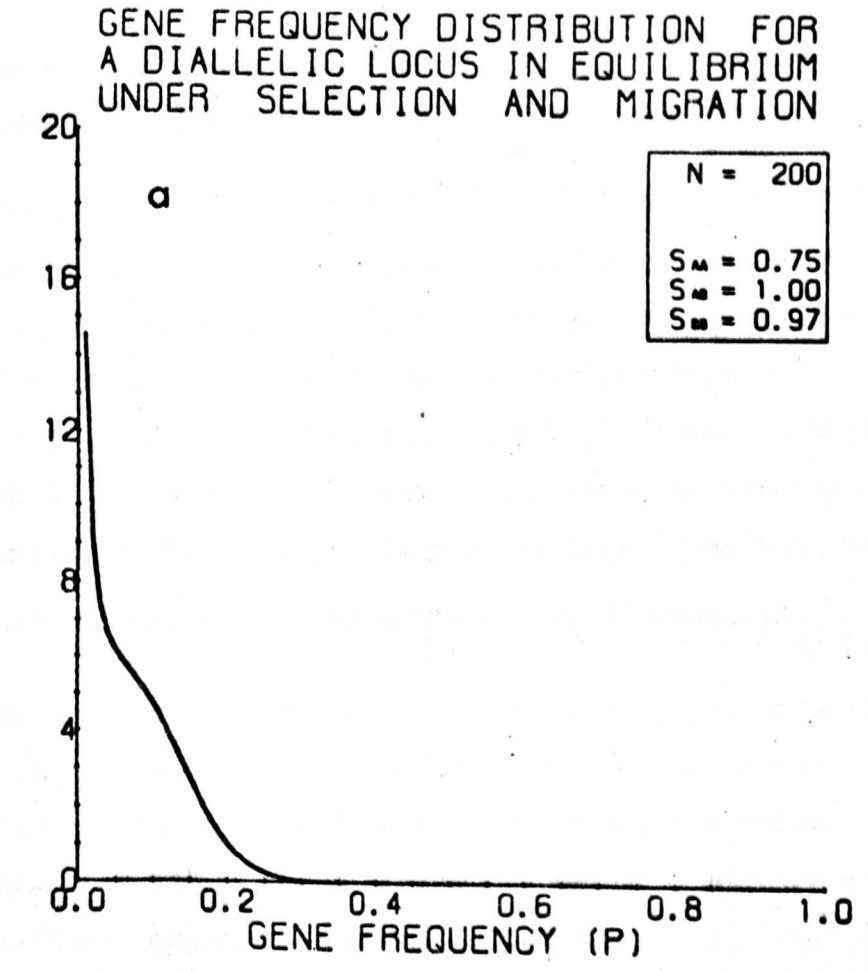


a



b

Figure 4.8. Gene frequency distributions from the models.
(a) 106 populations of size 200 with fitnesses of AA, AO and OO equal to 0.75, 1.00 and 0.97; no migration
(b) 562 populations of size 200, 5 immigrants per generation from a population with 20% gene A and 20% gene B; genotype fitnesses for AA, AB, BB, OB, OO and OA are 0.91, 1.06, 0.90, 1.05, 1.00 and 1.05.



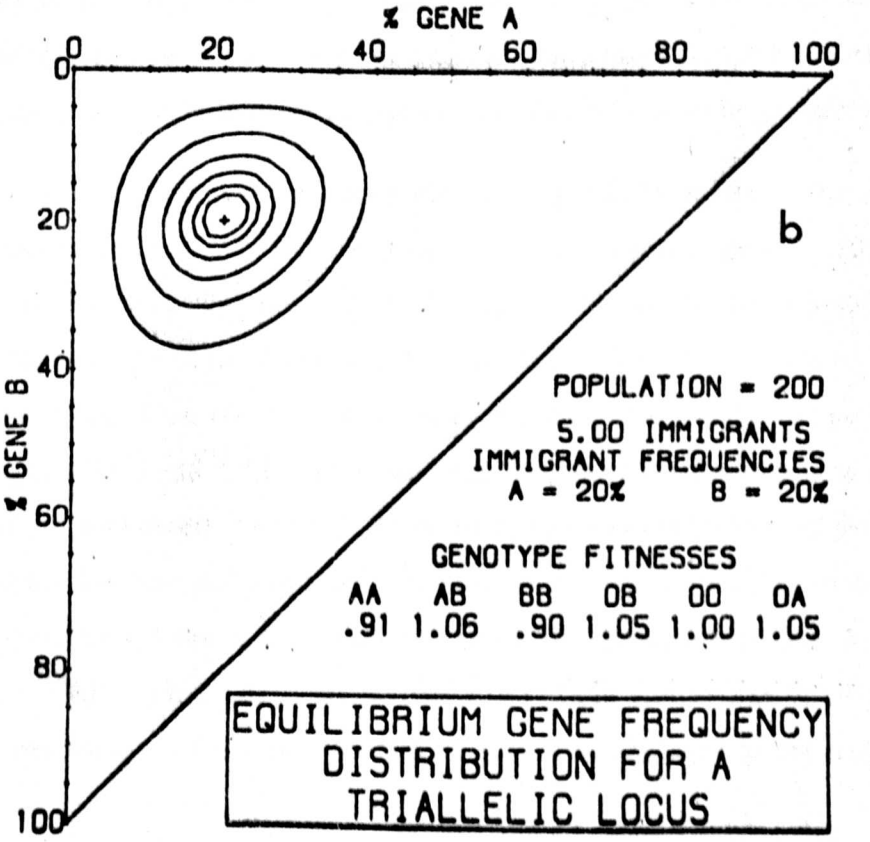


Figure 4.9. Graphical output from the models corresponding to the distributions of figure 4.8.

Bernstein also showed that a comparison of the observed phenotype numbers against those expected with the gene frequencies calculated was equivalent to testing

$\frac{pq}{2n(1-p)(1-q)}$ as a χ^2 deviate with 1 degree of freedom. (It is interesting to note that Bernstein only arrived at this result after making an unjustified approximation to compensate for a previous algebraic error.) In fact in the 2000 samples analysed (representing about 2.7 million individuals tested) D was too often positive; it is thought that this might be due to misclassifications of group AB as B. Roberts [16] gives a method which compensates for this. It is also conceivable that the discrepancy is due to selection causing a departure from Hardy-Weinberg equilibrium, but a proportion of the B phenotypes was reclassified as AB in a partially successful attempt to Normalise the distribution of $\sqrt{\frac{pq}{2n(1-p)(1-q)}}$.

Samples for which full details of the results were not given or for which information on the locality or type of population was lacking were rejected. So also were samples involving persons not indigenous to the place where the sample was taken, since a recent mixture of populations is likely to have perturbed the equilibrium situation. A χ^2 value significant at the 1% level also caused rejection of a sample as did an original sample size of less than 88. This latter value is somewhat arbitrary, but is in fact the result of tolerance limits considerations. About 90% of the samples considered are in the region $p < \frac{1}{3}$, $q < \frac{1}{3}$; if it is stipulated that values of p, q and r must be correct to ± 0.05 with a probability of 75% in this region the sample size restriction given can be calculated. It is important that the tolerance conditions are not made too rigid otherwise some small samples from anthropologically interesting populations would have to be discounted.

It is sufficient to deal here with those populations from Asia. 562 samples were retained and the gene frequency distribution for them is given in fig. 4.7.(b). It should be remembered when interpreting the results that the means of p and q were calculated independently of the sizes of the samples; they are not necessarily the values which would be obtained from the data produced by summing over the phenotype frequencies collected. Figs. 4.8.(b) and 4.9.(b) show the gene frequency distribution numerically and graphically for the parameters specified. Because there is no evidence to show that the effective population size and migration rate used in the model are close to actual values in Asian populations it would be improper to accept the fitnesses used as estimates of the fitnesses of the ABO genotypes in Asia. The shape of the distribution however does bear a sufficient resemblance to the observed data to give some credence to the model.

4.4.3.2. THE ESTIMATION OF FITNESSES

It has been emphasised several times that the model of the triallelic locus in equilibrium should not be used with ABO phenotype frequency data to estimate the genotype fitnesses at the ABO locus. It is however interesting to consider parenthetically how such a procedure could be carried out if the only unknown parameters were the fitnesses, and the observed gene frequency distribution did not also have to be estimated from a model. The investigation of such a topic is likely to shed some light on the possibility of using similar models for parameter estimation in an interactive environment.

Because only the ratios of the fitnesses are relevant there are 5 parameters to be estimated and the bivariate frequency distribution may conveniently be characterised by 5 quantities, namely the means of the two variables, their variances and their covariance. Without loss of generality the fitness of genotype OO may be constrained to be unity.

If $\underline{w} = (w_{AA}, w_{AB}, w_{BB}, w_{OB}, w_{OA})$, the required fitnesses,

$\underline{u}$ = (Exp(A), Exp(B), Var(A), Var(B), Cov(A, B)), calculated from the observed data, and $\underline{u}^{(0)}$ is an approximation to $\underline{u}$ arising from submitting fitnesses $\underline{w}^{(0)}$ to the model, then it is possible to find a better approximation to $\underline{w}$ from

$$\underline{w}^{(0)} + \alpha\, M^{-1}(\underline{u} - \underline{u}^{(0)})$$

where M is the matrix of the resultant increases in the u_i for a rise of α in any of the w_i, as shown for the Asian data in fig. 4.10. (page 55). The calculation of the entries in M assumes that there are no interactions between the effects of raising two fitnesses. This was confirmed to be the case for the African data by altering the fitnesses in the manner of a full 2^5 factorial experiment. In some situations it is possible to economise on runs by starting from a vector $\underline{w}$ which is symmetrical in two genes.

From a starting position of

$w_{AA} = 0.92$ $w_{AB} = 1.05$ $w_{BB} = 0.92$ $w_{OB} = 1.05$ $w_{OA} = 1.05$ (and $w_{OO} = 1.00$ throughout) new estimates

$w_{AA} = 0.907$ $w_{AB} = 1.061$ $w_{BB} = 0.902$ $w_{OB} = 1.050$ $w_{OA} = 1.049$

were obtained. These estimates gave values of (20.4, 20.1, 40.1, 38.6, −8.4) for $\underline{w}$ in comparison with the initial (20.7, 20.7, 45.1, 45.1, −13.9), both attempting to fit data giving (20.3, 20.1, 39.0, 37.0, −7.0). The new values of the parameters of the distribution are therefore considerably improved, and it may be supposed that the fitnesses are also estimated correspondingly more closely. A suitable method for calculating the elements of matrix M is a quarter replicate of a 2^5 factorial experiment. The results can be handled rapidly and simply by an APL program to give revised estimates for the fitnesses, but the method is

		Increase of 0.01 in				
		AA	AB	BB	OB	OA
Increase in	100 Exp (A)	**1.5**	**1.0**	**–0.5**	**–2.3**	**2.9**
	100 Exp (B)	**–0.5**	**1.0**	**1.5**	**2.9**	**–2.3**
	10^4 Var (A)	**2.3**	**–2.5**	**0.5**	**0.5**	**–5.5**
	10^4 Var (B)	**0.5**	**–2.5**	**2.3**	**–5.5**	**2.5**
	10^4 Cov (A,B)	**–0.8**	**3.8**	**–0.8**	**–1.3**	**–1.3**

Figure 4.10. Matrix of increases in the moments of the gene frequency distribution of genes A and B for a rise of 0.01 in the fitnesses of each of the genotypes.

essentially algorithmic and an automatic procedure is quite adequate. In a situation where there is interaction between the parameter main effects on one or more of the response variables heuristic methods may prove useful, but less in finding the optimal values of the parameters than in helping the user to understand how these interactions are caused.

4.5. USE OF THE MODELS FOR TEACHING

The models were used to teach a class of ten third year honours genetics students at the University of Newcastle upon Tyne. The students had previously heard two lectures on gene frequencies in populations and had done a simple Monte-Carlo simulation by drawing coloured balls from a bag. Considerable interest was shown in the lesson conducted from the Computek; students were keen to provide parameters for each successive graph and to suggest the form of the resulting distribution. Indeed, the lesson tended to proceed rather too quickly for the efficient retention of information, but this defect will easily be rectified in future years. In fact it is not serious, because suitable reinforcement of the main principles was given in a summary using hard copy output at the end of the lesson. The interactive nature of the lesson, however, was far more valuable than mere use of previously drawn graphs could be, for the students were able to discover facts for themselves. The ability of migration alone to counter the effect of drift surprised the class, and their discovery that this same effect was of second order importance in the presence of a stable polymorphism should leave a lasting impression of the dramatic effect of selection.

It is important that no organisational difficulties are allowed to obscure the main objects of a lesson, although students will tolerate a limited amount of these if an interesting new teaching technique is being tried out. A class of more than ten students would present such difficulties, as the area from which the Computek screen may comfortably be viewed is rather small. Closed circuit television might be employed to convey the lesson to a large audience, but since participation in the lesson by the student is so necessary it is probably as well to restrict the use of the model to small classes in any case.

The availability of a storage display screen for output of the graphs facilitated the use of the models to a considerable extent, but even if such equipment were not available some progress could be made towards using interactive computer models as teaching aids. If a standard line printer is available for output, and may be taken out of normal service for the duration of the lesson, graphs may be output by this means. Evans [5] has produced a program for drawing contour representations of bivariate probability distributions and fig. 4.11. (page 57) shows how it may be used in the present context to bring out the salient features of the distribution previously given in fig. 4.9. (b). It is unfortunately true that many

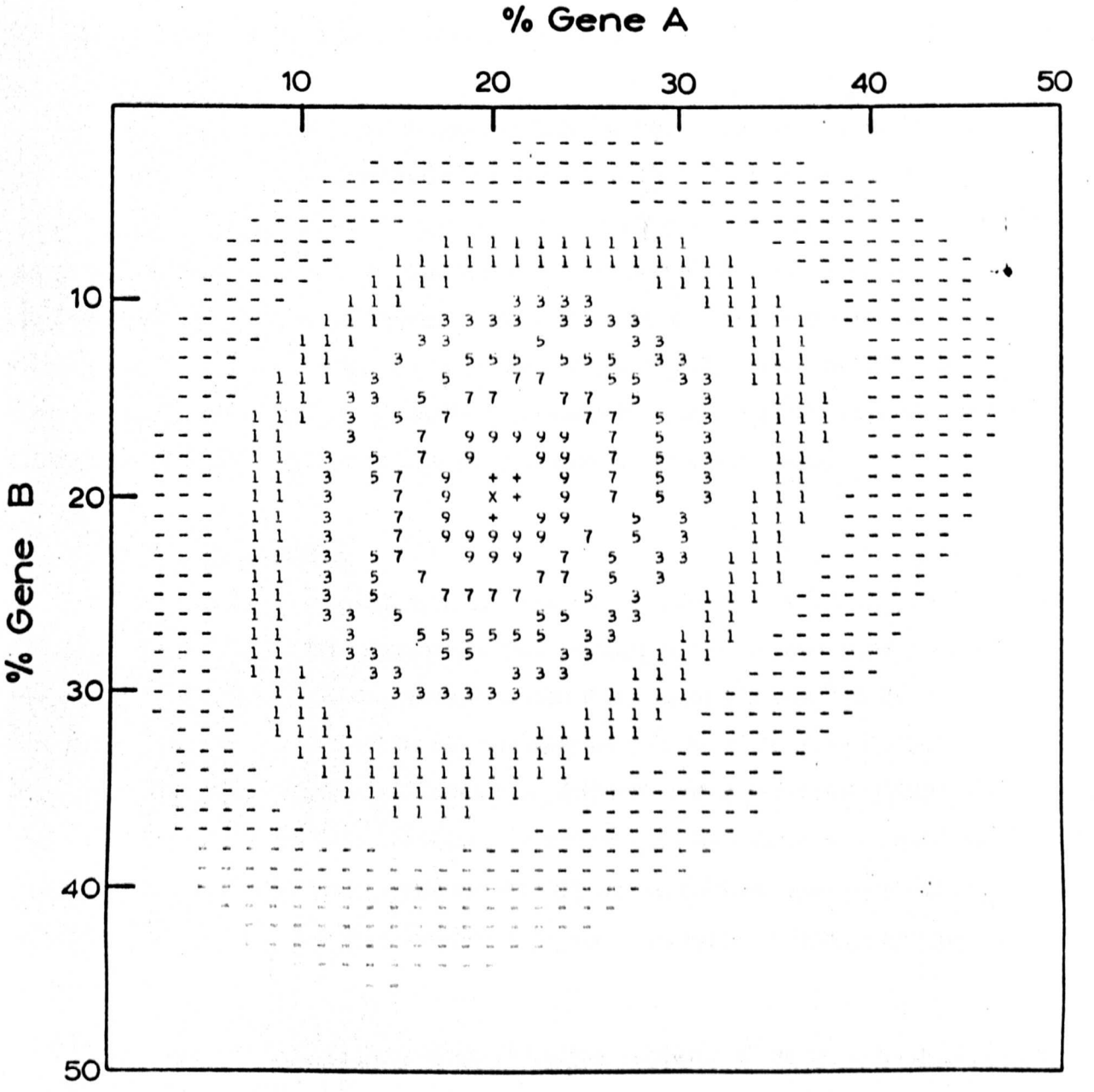


Figure 4.11. A representation of figure 4.9 (b) which may be output on a line printer.

installations wishing to use the models interactively would have to resort to some compromise like this, although the absence of facilities for interactive graphics can seldom be justified on economic grounds as they can be provided for as little as £10,000, ready to use.

4.6. A MONTE-CARLO APPROACH TO THE ABO SYSTEM

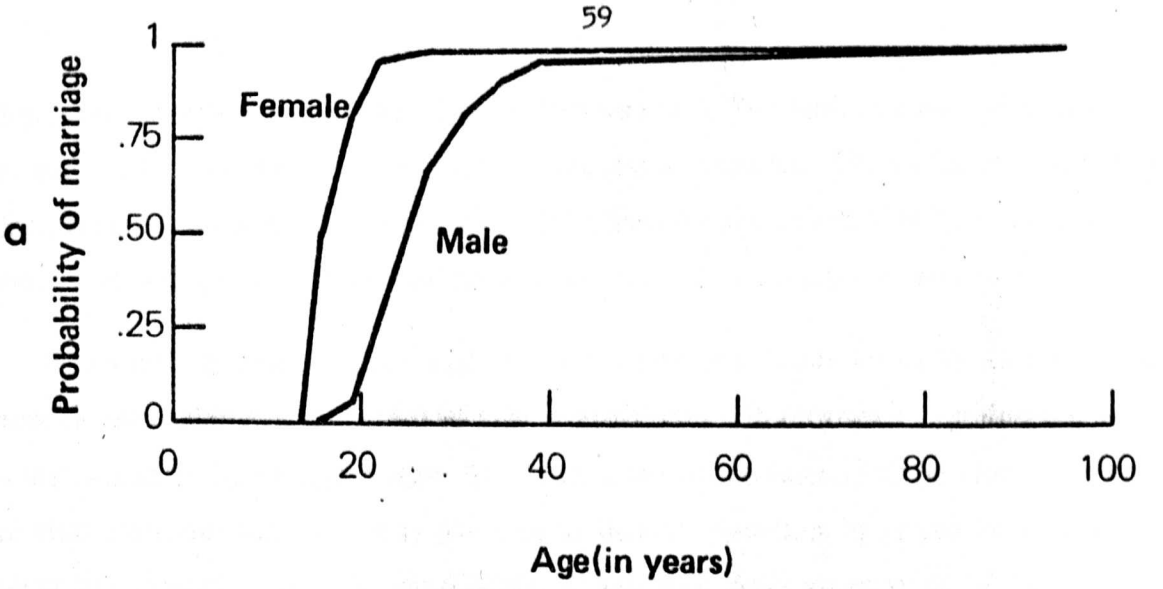
4.6.1. INTRODUCTION

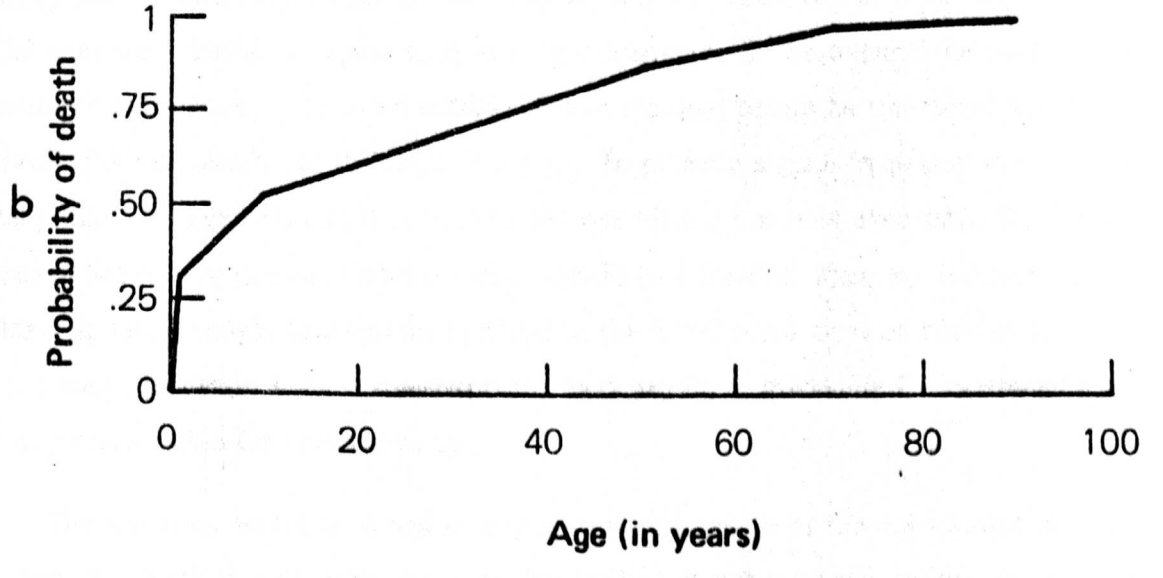
It was stated in section 4.4.3.2. that the diffusion equation model of gene flow was not well disposed to the task of estimating the fitnesses of the ABO genotypes because information was lacking on vital factors such as population size and migration rates. These considerations also apply to the model about to be discussed. As a model for investigating the use of interaction it is valid enough, but in the absence of good estimates of these important parameters the genetical results must be treated with scepticism. This investigation was in fact carried out prior to the treatment of the system by stochastic equation, the latter method being adopted after the lack of success with the more direct simulation.

4.6.2. DEMOGRAPHIC ASPECTS

The diffusion equation approach to the simulation of gene flow in a population is certainly adequate for many applications as a teaching device. It does however make several simplifying assumptions about the system it is intended to represent and is therefore unsuitable as a vehicle for investigating some facets of the system, such as the effect of a very small effective population size, different mating patterns or alternative modes of action of the selective pressures. To answer questions about such topics, and to test the efficacy of using a model with non-overlapping generations, a more detailed simulation is called for. Since variation plays such a large part in the system a Monte-Carlo approach is indicated.

In addition to the data on blood group phenotype frequency which has already been used, certain demographic factors must be taken into account. The birth rate, death rate, duration of the fertile (at risk) period and how these parameters have changed over the years must be known. Since the object of the exercise is to determine fitnesses which bring about the present equilibrium situation the simulation performed must cover several generations, and, the human generation time being some three decades, the demographic particulars must be ascertained for the past six hundred years or so. In fact the relevant distributions have been assumed to remain constant over that period as a first approximation to the truth. The distributions for age at death and beginning and end of fertility which were used for the Asian populations also dealt with by the stochastic model are given in figure 4.12 (page 59).





Age at end of fertility

c

Females	45 years
Males	60 years

Figure 4.12. Vital statistics which may be used in a Monte-Carlo simulation of (ABO) gene flow in a population.

They were constructed from data given by Vielrose [17]. The birth rate may be adjusted by the program to keep the population size approximately constant. Since it is generally death which terminates the at risk period the distributions for age at end of fertility need not be determined with great accuracy and these ages may safely be taken to be constants.

It is simple to transform these graphs into functions suitable for GPSS, IBM's General Purpose Simulation System, and it is also straightforward to program a flow diagram such as that shown in figure 4.13. (page 61) into a few of the common GPSS blocks. By adapting the vital statistics functions it is possible to simulate selection by increased fertility, longer life, maternal-foetal incompatibility, or any other mode which might be relevant.

4.6.3. RESULTS FROM THE MODEL

The simulation of 600 years of gene flow history in a population of 200 individuals involves some 5000 transits through a model like that of figure 4.13. This requires a considerable amount of processing by the GPSS interpreter, and since GPSS does not accept data from external sources the program must be edited before each successive run. The user must therefore expect to spend about 5 minutes at the terminal for each run. This would be quite acceptable if the number of runs required before he interacted with the simulation was small, but in fact it is not so. To produce a gene frequency distribution many runs are required, and it is seldom obvious after a few runs even when the fitnesses chosen are widely in error that the series should be curtailed. When the fitnesses are closer to those sought to maintain equilibrium the distribution must be built up more accurately, requiring more simulation runs. This aspect of the Monte-Carlo method makes it quite unsuitable for interactive use.

The situation would be eased if only a single parameter of the distribution were to be estimated, which is very often the case, but in the present instance where a distribution is required (and a bivariate one) the Monte-Carlo technique is definitely unsuitable and failed to produce any worthwhile results.

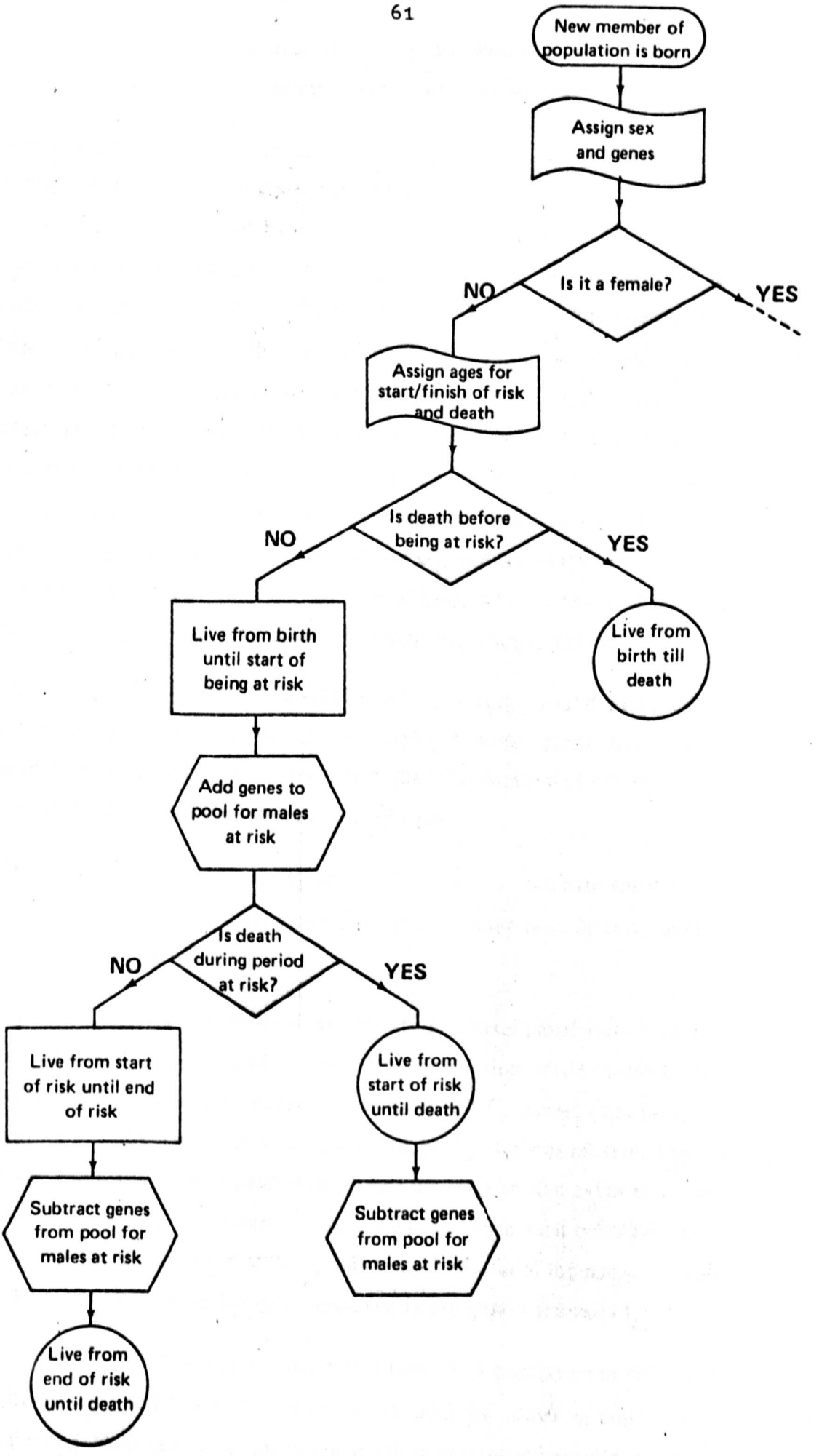


Figure 4.13. Logic diagram for a Monte-Carlo simulation of gene flow in a population.

5. AN INTERACTIVE COMPUTER MODEL OF A CONDITIONAL RENEWAL SYSTEM

5.1. INTRODUCTION

5.1.1. THE CELL CYCLE OF MAMMALIAN CELLS

Howard and Pelc [6] showed in 1953 that the reproductive cycle of root meristem cells could be observed to consist of distinct phases. The same phenomenon is apparent in many organisms including mammals. A cell which is taking part in the reproductive cycle divides to form two daughter cells, which may in turn divide to form two further daughters each. The act of division is called mitosis, and the period between successive mitoses is known as interphase. Before mitosis can occur a cell must increase in size and synthesise various proteins and DNA. That period during interphase in which DNA is synthesised is termed the synthetic phase or S-phase. The period between mitosis and the succeeding S-phase is the pre-synthetic phase, denoted by G_1, and the remainder of interphase, between the end of the S-phase and the start of mitosis is the post-synthetic phase G_2. During G_1 and G_2 the cell synthesises the other constituents necessary for mitosis.

Mitosis, or the M-phase, may itself be seen to comprise four distinct phases. These are known as prophase, metaphase, anaphase and telophase, but because the period of mitosis is short in comparison with the total duration of the cycle few models distinguish between the different stages of mitosis.

The symbols t_{G_1}, t_s, t_{G_2} and t_m are used to represent the time spent by a cell in the pre-synthetic, synthetic, post-synthetic and mitotic phases respectively, and their sum, the cell cycle time, is denoted by T_c.

Not all cells which are produced as a result of mitosis proliferate. Some alter their structure and function in a process known as differentiation, while others apparently lie dormant in G_1. Such cells are frequently said to be in a G_0 phase; experimentally they are indistinguishable from cells which pass through G_1 at the usual rate. The cells in a population which are actually progressing round the proliferative cycle at any given time is sometimes called the proliferative compartment, although such cells need not be spatially separate from the rest of the population. The ratio of the number of cells in the proliferative compartment to the cell population is the growth fraction, I_p.

One other time is of interest, namely that taken by a population of cells to double their number. This will depend on the growth fraction, the number of cells which differentiate and on the cell death rate. In a population which is growing exponentially (no cell loss) and has a growth fraction I_p, the doubling time T_D is related to T_c by $(1+I_p)^{T_D/T_c} = 2$, or

$$T_D = \frac{\log 2}{\log(1+I_p)} \; T_c.$$

5.1.2. PROLIFERATIVE INDICES AND FLM CURVES

During the S-phase cells incorporate thymidine, and it is possible to label these cells by providing them with a supply of tritiated thymidine. Autoradiographic techniques may then be used to detect the label and hence to determine the proportion of cells which were in the S-phase at the time of labelling. This parameter is known as the labelling index, I_L. Cells in mitosis may be observed because their morphology is different from cells in interphase, and the mitotic index I_m is defined to be the proportion of the cell population which is undergoing mitosis at any given time. The mitotic index is a more direct measurement of the degree of proliferation in a system than the labelling index, but because mitosis occurs relatively rapidly a great number of cells must be counted to obtain few mitoses. Typically the S-phase lasts ten times longer than the M-phase and so the labelling index is more convenient. Since cells in the S-phase proceed directly through G_2 to mitosis, I_L is equally valid as a measure of proliferative activity, and I_L and I_m together are known as the proliferative indices.

In a steady state system it can be shown that $\frac{I_m}{I_L} = \frac{t_m}{t_s}$

Another proportion which may be calculated from the same observations as I_L and I_m is the fraction of mitoses which are labelled. The variation of this index with time provides a powerful method of estimating the cell-cycle parameters (the individual phase times); it was first described by Howard and Pelc [6], and developed by Quastler and Sherman [13]. If a label is given for a very short period (a "pulse label") and the fraction labelled mitoses (FLM) calculated serially thereafter, a curve is obtained in which the FLM is zero while the labelled cohort passes through G_2, unity while it passes through mitosis, zero again for G_1, S and G_2 of the daughter cells and then unity for the next mitosis. Of course the real life situation is not as simple as this, because not all cells take the same time to pass through any given phase. Instead the phase times may be characterised by probability distributions, and several authors, most recently Gilbert [5], have derived methods of determining the cell-cycle parameters and their variances from experimental data. Figure 5.1. (page 64) shows a section of the testosterone stimulated seminal vesicle of a castrate mouse; mitoses, labelled cells and labelled mitoses are all present.

One other type of experiment should also be discussed at this stage; the determination of the cumulative labelling index. Tritiated thymidine labelling is carried out at intervals

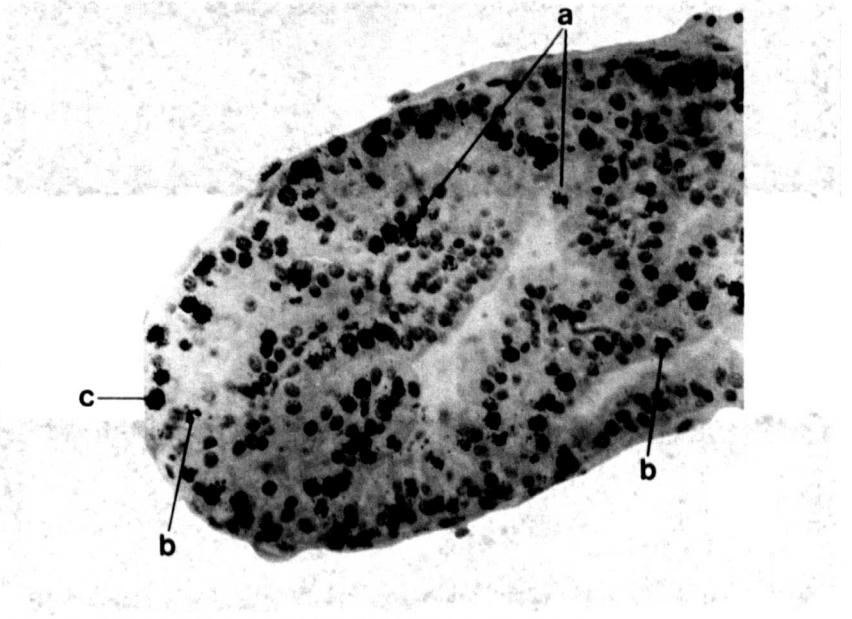


Figure 5.1. A section of the prostate of the mouse showing (a) unlabelled mitoses, (b) labelled mitoses and (c) labelled cells in interphase.

shorter than t_s so that all cells in the proliferative compartment are labelled as they pass through S. The proportion of labelled cells which is ultimately found in the population is not itself the growth fraction because on mitosis a labelled cell becomes two labelled cells, but the growth fraction may be calculated using this technique. The system is considered to be well determined when the growth fraction and the phase durations are known.

5.1.3. CONDITIONAL RENEWAL SYSTEMS

The particular example of the cell cycle which will be modelled in section 5.2. involves a conditional renewal system. This is an organ or tissue which normally has a very low mitotic index but which will respond to a stimulus, such as trauma or provision of hitherto lacking trophic hormone, with a wave of mitoses. Experimental data will be used which relates to two areas of the prostatic complex in the castrate male mouse when testosterone is administered. These areas are the seminal vesicle and the coagulating gland (see figure 5.2., page 66 , adapted from [12]). Morley, Wright and Appleton [10] have shown that the mitotic wave produced by the stimulus in these organs dies out even if the stimulus is continued. A similar situation obtains in the oestrogen stimulated uterus of the castrate female rat [8].

Partially hepatectomised rats exhibit a conditional renewal system wherein the liver grows by mitotic activity until it attains its original size, as demonstrated by many investigators, (see for example Johnson [7]).

A successful model of these systems will be able to estimate the cell-cycle parameters and the size of the proliferative compartment as well as giving some idea of the mechanism behind the rise and fall in the proliferative indices.

5.1.4. APPLICATIONS OF THE CELL-CYCLE

It would perhaps be profitable at this stage to indicate the reasons for studying the cell cycle. Cytokinetics is not merely an end in itself, comprising interesting autoradiographic techniques and stochastic models, but has applications in the field of cancer chemotherapy. It is important to know the kinetics of both the neoplasm to be treated and the related organs and tissues in order to plan a strategy whereby the tumour may be eliminated without irrevocable damage being caused to surrounding areas. Cancer chemotherapeutic drugs may, for example, block cells from entry into DNA synthesis. If the cell cycle time of the neoplastic cells is short compared to that of other cells which are affected, administration of the drug for the period of the tumour cell cycle will destroy the tumour, while at the same time preserving those other cells which have not reached the block. The proportion of cells

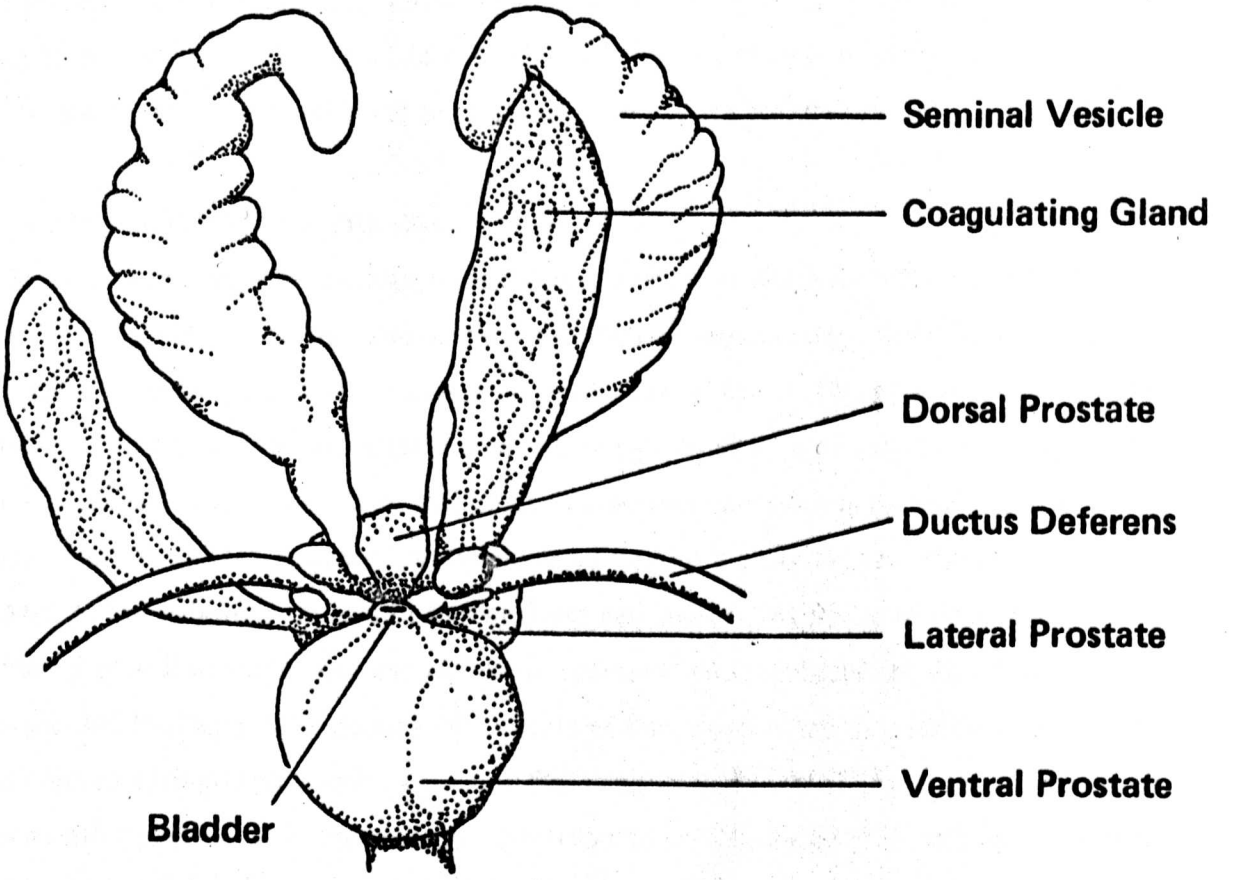


Figure 5.2. The prostatic complex of the (male) mouse.

able to repopulate the system can then be estimated.

It is also important to be able to study the stimulatory effect of hormones, as many diseases both acquired and inherited arise from hormone deficiency or excess. For this reason it is important that a model of a conditional renewal system's cytokinetics should provide insight into the system and reliable estimates of its parameters.

5.1.5. PREVIOUS MODELS OF THE CELL CYCLE

Since Howard and Pelc described their original model of the cell cycle, reproduced in figure 5.3. (page 68), many extensions to it have been constructed. Some, like Cairnie, Lamerton and Steel's model of the intestinal epithelium of the rat [3], deal with a particular physiological system in detail. Others deal with a particular type of experiment, usually the FLM experiment. A wide variety of techniques have been employed in formulating these models. Barrett [1] uses Laplace transforms in setting up his model, but favours a Monte-Carlo approach to the solution. He is followed by Steel and Hanes [14] who also deal with the continuous labelling curve, and use automatic optimisation techniques for data fitting. Pedersen and Hartmann [11] discuss two models of the mouse ovary in relation to FLM and continuous labelling experiments, assuming Gaussian distribution of phase durations, using a stochastic equation. This approach is also favoured by Takahashi [15], who uses a Gamma distribution for phase durations, and Gilbert [5], also using a Gamma distribution, is able to obtain a simple model of the FLM curve by means of the Laplace transform.

These methods are fairly clearly divided into models which are intended to clarify the workings of a system and those which are designed to estimate cell-cycle parameters. The results of most of the recent FLM models are indistinguishable from the point of view of how adequately they fit experimental data, and it must be doubtful whether any further benefit can be derived from models based on an experimental method rather than on a specific system. Too often the assumptions required by the former type of model are not satisfied in the particular instance in which the experimenter would like to use it.

None of the computer models so far published is suitable for interactive use; the Monte-Carlo approach is too slow, optimisation techniques are performed automatically, graphical output is usually absent, leading to slow assimilation of results. The principal reason, of course, is that as models for parameter estimation they do not need or want the particular aspects of interactive use.

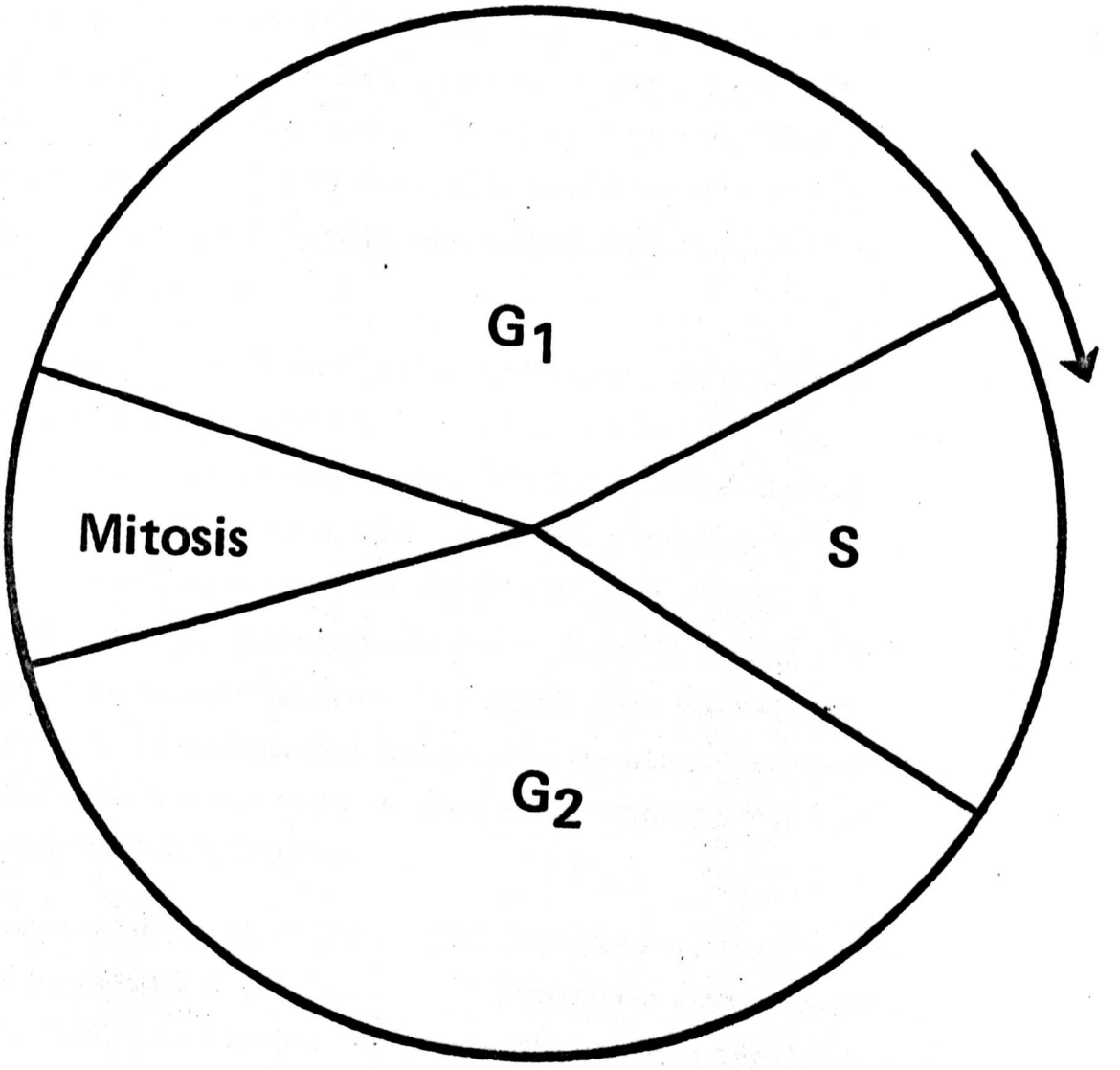


Figure 5.3. Howard and Pelc's model of the cell cycle.

5.2. A MODEL OF A STIMULATED CONDITIONAL RENEWAL SYSTEM

5.2.1. MODEL DESCRIPTION

If a conditional renewal system is deprived of a necessary hormone, mitotic activity is reduced to very low levels and the cells may be considered to be in a G_0 or long G_1 phase. After hormonal stimulation there is a delay followed by a burst of DNA synthesis and mitosis which then dies out. This is so even if the hormone is ensured to be in plentiful supply by administering it at regular intervals. The following model is proposed to describe the system (see figure 5.4., page 70).

Cells start in compartment G_0 and after stimulation some or all of them pass through G_1, S, G_2 and M phases. After mitosis a daughter cell may either leave the cycle and enter the differentiated compartment or undertake a further cell cycle. The numbers of cells taking these alternative paths is controlled by the "decycling probability". The idea of decycling probability is akin to Bresciani's distribution ratio [2] and corresponds to his $\vec{\eta}$. It is more basic and for many purposes more useful than the growth fraction, particularly when the distribution of cells over the phases of the cycle is unknown or changing, although both quantities give the same kind of information about the system. Some workers have stated that cells divide only once after stimulation, so the decycling probability would be identically 1; this assertion will be examined.

Simulations of the cell cycle have gained acceptance in recent years since Barrett's Monte-Carlo approach [1] and Takahashi's investigation of a stochastic equation technique for analysis of FLM curves [15] were reported. The model described in figure 5.4. was treated analytically because of the greater speed attainable and because more insight can be obtained from a mathematically defined model than from a Monte-Carlo model, which often tends to mask the important effects by modelling too closely the noise (random or other irrelevant variation) of the system under investigation. Graphical presentation of output was chosen because of its inherent advantage in ease of understanding and communication.

5.2.2. MATHEMATICAL TREATMENT OF THE MODEL

I. The system modelled.

i) All cells are initially in a G_0 phase.

ii) A proportion of the cells leave G_0 and pass through G_1, S, G_2 and M phases.

iii) The rate at which these cells leave G_0 is a function of time only.

iv) All cells take the same time to pass through a given phase.

v) Progeny may leave the cycle (differentiate) or proceed through a further cycle.

vi) The decycling probability at the end of each cycle may be constant, depend on the total

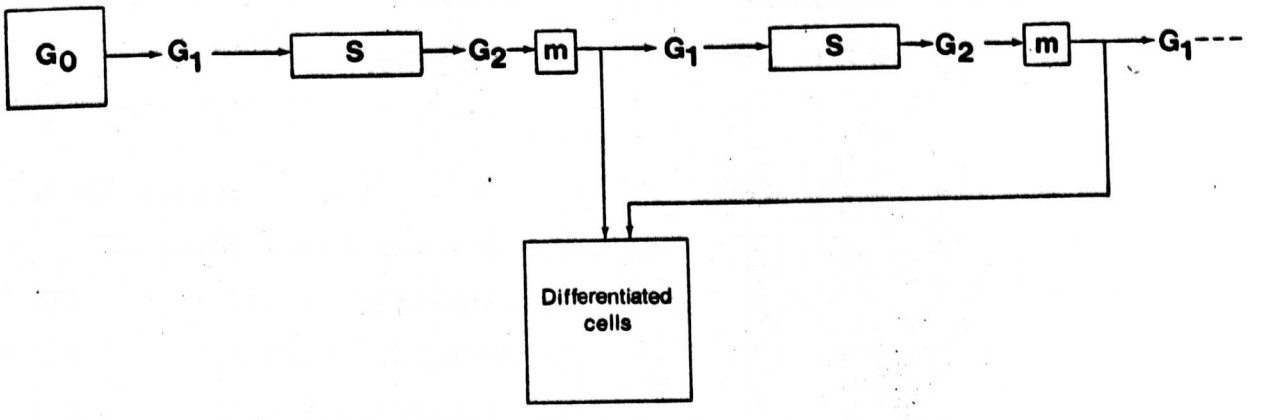


Figure 5.4. **Model of stimulated cell kinetic response.**

Cells start in G_0 and move through the cycle towards the differentiated compartment. After each mitosis a cell "decides" whether to differentiate or to repeat the cycle.

number of cells in the cycle or on the total number of cells which have left the cycle, or be any other function of time.

vii) Labelling of cells in an S-phase is assumed to be instantaneous. If this is not so, then the estimate of the duration of S (t_s) must be reduced by the duration of action of the labelling agent, and the estimate of the duration of G_1 (t_{G_1}) increased by the same amount.

II. Notation used.

The functions which must be calculated are:

$N(t)$: total number of cells in the system.

$I_L(t)$: proportion labelled of total.

$I_m(t)$: proportion in mitosis of total.

$I_p(t)$: growth fraction.

$FLM(t)$: proportion labelled of mitoses.

$I_{CL}(t)$: continuous labelling index.

$\vec{\eta}_r(t)$: decycling probability after the r^{th} cycle.

The following quantities are used in the analysis:

$D(t)$: number of cells which have left the system.

$S(t)$: number of cells in S-phases.

$M(t)$: number of cells in M-phases.

$G_0(t)$: number of cells in G_0.

$f(t)$: birth rate into the cycle for cells leaving G_0.

k : number of cycles possible (for convenience of calculation only; k may be taken to be infinite in theory).

N_0 : number of cells initially in G_0.

ρ : proportion of cells which leave G_0.

$L_{jr}(t)$: number of mitoses in r^{th} cycle labelled in j^{th} cycle.

$L(t)$: number of labelled mitoses.

τ : time of labelling (FLM curve).

$\alpha_r^{(m)}, \omega_r^{(m)}$: time from leaving G_0 to start/finish of r^{th} mitosis.

$\alpha_r^{(s)}, \omega_r^{(s)}$: time from leaving G_0 to start/finish of r^{th} S-phase.

(Thus, $\omega_r^{(m)} - \alpha_r^{(m)} = t_m$ in the r^{th} cycle;

$\omega_r^{(s)} - \alpha_r^{(s)} = t_s$ in the r^{th} cycle;

$\omega_r^{(m)} - \omega_{r-1}^{(m)} = T_c$ for the r^{th} cycle.)

For conciseness $P(\theta)$ is defined by

$$P(\theta) = \rho\, N_0 2^{r-1} \prod_{j=0}^{r-1} (1-\vec{\eta}_j(\theta+\omega_j^{(m)})) \; ; \; \omega_0^{(m)}=0, \; \vec{\eta}_0 = 0;$$

so that, for example, $P(t - a_r^{(s)})\, f(t - a_r^{(s)})\, dt$ is the number of cells which start the r^{th} S-phase between times t and t + dt.

$$r_{1jr} = \tau + a_r^{(m)} - \omega_j^{(s)} \; ; \qquad r_{2jr} = \tau + \omega_r^{(m)} - \omega_j^{(s)};$$

$$r_{3jr} = \tau + a_r^{(m)} - a_j^{(s)} \; ; \qquad r_{4jr} = \tau + \omega_r^{(m)} - a_j^{(s)}.$$

III. Solution

The following relationships hold.

$$G_0(t) = N_0(1 - \rho \int_0^t f(\theta)\, d\theta) \tag{1}$$

$$D(t) = 2\sum_{r=1}^{k} \int_{\omega_r^{(m)}}^{t} \vec{\eta}_r(\theta)\, P(\theta - \omega_r^{(m)})\, f(\theta - \omega_r^{(m)})\, d\theta \tag{2}$$

$$N(t) = N_0 + \sum_{r=1}^{k} \int_{\omega_r^{(m)}}^{t} P(\theta - \omega_r^{(m)})\, f(\theta - \omega_r^{(m)})\, d\theta \tag{3}$$

$$S(t) = \sum_{r=1}^{k} \int_{t-a_r^{(s)}}^{t-\omega_r^{(s)}} P(\theta)\, f(\theta)\, d\theta \tag{4}$$

$$M(t) = \sum_{r=1}^{k} \int_{t-a_r^{(m)}}^{t-\omega_r^{(m)}} P(\theta)\, f(\theta)\, d\theta \tag{5}$$

Notice that since $\vec{\eta}_r(t)$ and hence P(t) may depend on D(t) or N(t) equations (2) and (3) may be integral equations.

$$L_{jr}(t) = \begin{cases} 0, & t < r_{1jr} \\ \int_0^{t-a_r^{(m)}} P(\theta)\, \{f(\theta) - f(\tau - \omega_j^{(s)})\}\, d\theta, & r_{1jr} < t < r_{2jr} \\ \int_{t-a_r^{(s)}}^{t-\omega_r^{(s)}} P(\theta)\, f(\theta)\, d\theta. & r_{2jr} < t < r_{3jr} \\ \int_0^{t-\omega_r^{(m)}} P(\theta)\, \{f(\tau - a_j^{(s)}) - f(\theta)\}\, d\theta, & r_{3jr} < t < r_{4jr} \\ 0, & r_{4jr} < t \end{cases} \tag{6}$$

$$I_L(t) = \frac{S(t)}{N(t)} \qquad I_m(t) = \frac{M(t)}{N(t)}$$

$$I_p(t) = 1 - \frac{G_0(t - \alpha_1^{(s)} + \alpha_2^{(s)} - \omega_1^{(m)}) + D(t)}{N(t)}$$

$$I_{CL}(t) = 1 - \frac{G_0(t - \alpha_1^{(s)})}{N(t)}$$

$$L(t) = \sum_{r=1}^{k} \sum_{j=1}^{r} L_{jr}(t) \qquad FLM(t) = \frac{L(t)}{M(t)}$$

The form of the relationship for $I_p(t)$ is explained in section 5.4.1.

5.2.3. COMPUTER SOLUTION OF THE MODEL

As indicated previously, when the decycling probability is not a function of time alone but depends on the value of another variable in the system, the mathematical formulation of the system involves an analytically intractable integral equation and step by step simulation is required. This must be carried out quickly as it will be used several times to investigate the behaviour or find the parameters of a particular system. Use is therefore made of the facts that the decycling probability changes slowly and that small changes in the decycling probability cause only small changes in the numbers of cells in any part of the system. This enables the value of $\vec{\eta}_r(t)$ to be approximated by assuming linearity over t−2, t−1 and t, and hence the functions D(t), N(t), S(t), M(t) and L(t) may be calculated. The estimates of the $\vec{\eta}_r(t)$ are then improved; no further iteration is required. $G_0(t)$ may be calculated analytically; other integrals are evaluated using the trapezoidal rule, and economy is gained for D(t), N(t) and S(t) by using the values previously found for t−1. It is then simple to calculate the required indices.

The function f(t) has been chosen to be an Erlangian distribution with parameter 2. The advantages of this distribution are that it is not a priori an unreasonable approximation to the (unknown) distribution of times spent in G_0 after stimulation, and that it can be handled analytically; the program will accept other values of the parameter, or indeed other distributions should a better one be found. Although the distribution may be integrated analytically, this is not important in the method of solution chosen, but it makes the solution for the case of constant decycling probabilities obtainable in an explicit form (see section 5.2.4). An Erlangian distribution was preferred to the more general Gamma distribution not because an attempt can be made to justify it on the grounds of the "method of stages" as Takahashi does [15], but because it is much faster to evaluate than the Gamma distribution.

The number of cycles considered depends on the length of the simulation run. Efficiency is improved by not postulating a 4th cycle if no cells will have reached it in the time under investigation. The program listing is in appendix 8.

5.2.4. A SIMPLER MODEL

If the decycling probability at the end of any given cycle remains constant, equations (2) and (3) (section 5.2.2.) need no longer be treated as integral equations, and expressions for the various indices may be found in closed forms.

P_r is defined analogously to $P(\theta)$ by

$$P_r = \rho \; N_0 2^{r-1} \prod_{j=0}^{r-1} (1 - \vec{\eta}_j)$$

Writing $F(t) = \int_0^t f(\theta)\, d\theta$ (so that for the Erlangian distribution with parameter n and mean μ,

$$F(\theta) = 1 - \exp(-n\theta/\mu) \sum_{r=1}^{n-1} (n\theta/\mu)^r / r! \quad)$$

enables equations (1) to (6) to be simplified to

$$G_0(t) = N_0 (1 - \rho F(t)) \tag{1a}$$

$$D(t) = 2 \sum_{r=1}^{k} \vec{\eta}_r P_r F(t - \omega_r^{(m)}) \tag{2a}$$

$$N(t) = N_0 + \sum_{r=1}^{k} P_r F(t - \omega_r^{(m)}) \tag{3a}$$

$$S(t) = \sum_{r=1}^{k} P_r \{F(t - \alpha_r^{(s)}) - F(t - \omega_r^{(s)})\} \tag{4a}$$

$$M(t) = \sum_{r=1}^{k} P_r \{F(t - \alpha_r^{(m)}) - F(t - \omega_r^{(m)})\} \tag{5a}$$

$$L_{jr}(t) = \begin{cases} 0, & t < \tau_{1jr} \\ P_r \{F(t - \alpha_r^{(m)}) - F(\tau - \omega_j^{(s)})\}, & \tau_{1jr} < t < \tau_{2jr} \\ P_r \{F(t - \alpha_r^{(m)}) - F(t - \omega_r^{(m)})\}, & \tau_{2jr} < t < \tau_{3jr} \\ P \{F(\tau - \alpha_j^{(s)}) - F(t - \omega_r^{(m)})\}, & \tau_{3jr} < t < \tau_{4jr} \\ 0, & \tau_{4jr} < t \end{cases} \tag{6a}$$

Other equations remain unaltered from those in section 5.2.2.

5.3. USE OF THE MODEL

5.3.1. THEORETICAL CONSIDERATIONS

Unlike most FLM curve-fitting techniques, designed to estimate the durations of the phases of the cell cycle, the present model is to be regarded primarily as a means of gaining familiarity with and insight into the system it represents. It is true that it can be used to estimate the cell cycle parameters, and in fact a mitotic index curve provides a better estimate of the duration of mitosis than can be obtained from an FLM curve. But a mathematical model should do more than estimate parameters; it should also teach its users about the system it represents. Furthermore, a good model will be amenable to verification by suggesting other experiments. Decycling probabilities found from the model can be examined by means of grain-count experiments, and the total number of cells in the system at any time can be compared with DNA measurements.

In an examination of the effect on the final population size of changing the decycling probability or of postulating different quantities upon which it might depend, the user of the model will require to carry out several simulation runs. Frequently he will choose the parameters for one run only after seeing the results of the previous run. If such a procedure is to be carried out conveniently it is important that the computing facilities available should support man-machine interaction at a fairly sophisticated level. The model under discussion was programmed in ALGOL W with FORTRAN graph-plotting routines, and was designed to be run interactively from an IBM 2741 typewriter terminal with graphical output being produced on the Computek 400 storage display screen. Hard copy output could be obtained whenever a particularly informative set of parameters was used or when a fit was found for experimental data.

The time elapsing from the start of entering data in response to prompting from the program to the completion of the display of labelled and mitotic index curves was about 3 minutes, the exact time depending on the pressure from other users on the time-sharing system. The time for such a run was always roughly similar to that taken up by discussion of its output.

The use of a multiparametric model may legitimately be criticised as an unprofitable technique if parameters are altered arbitrarily until the best possible fit has been obtained. It was therefore thought advisable to restrict certain of the parameters of the model. If good estimates of t_s or T_c are already available it is not permissible to depart from these by more than a small amount; the time parameters for each cycle may conveniently be constrained, in the absence of any evidence to the contrary, to be the same, except for the duration of the first G_1 phase which is experimentally indistinguishable from G_0. The decycling probability

for each cycle will normally be the same and should only be taken to depend on the number of differentiated cells if an adequate fit cannot be achieved while it remains constant. In this fashion the freedom of the model is limited and its ability to fit the several sets of data presented to it becomes the more impressive. Where data has been fitted nine parameters have been varied. In using the model for other purposes many more possibilities may be investigated.

5.3.2. FITTING THE CURVES TO DATA

All indices which can be expressed as the proportion of "successes" in a number of "trials" may be treated identically in the matter of fitting curves to observed values. Thus the following argument applies equally to labelling index, mitotic index, continuous labelling index and FLM curves.

Suppose the experimental index at time t is λ_t, derived from ν_t observations, and the value predicted by the model is μ_t. For the FLM curve several methods are in use. Takahashi, Hogg and Mendelsohn [16] minimise the weighted sum of squares

$$\sum_t \nu_t (\lambda_t - \mu_t)^2$$

which does not take into account that for the same number of observations the sampling variance is higher near $\lambda_t = 0.5$ than it is near 0 or 1. McDonald [9] maximises the log-likelihood

$$\sum_t \nu_t \{\lambda_t \log \mu_t + (1 - \lambda_t) \log (1 - \mu_t)\}$$

which suffers because μ_t may be zero, and he has suggested to Steel and Hanes [14] that they minimise

$$\sum_t (\arcsin \sqrt{\lambda_t} - \arcsin \sqrt{\mu_t})^2 .$$

This last technique takes the differences in sampling variance into account, but it is preferable to extend it slightly so that the possibility of counting different numbers of cells is catered for, and consider the expression

$$\sum_t \nu_t \ (\arcsin \sqrt{\lambda_t} - \arcsin \sqrt{\mu_t})^2 . \qquad \ldots\ldots (1)$$

The authors cited use automatic optimisation procedures to search for the required parameters. It was decided not to do this for several reasons. Firstly, since data may be available for labelling, mitotic and continuous labelling indices, the situation is rather different from that in which the parameters of a model are adjusted to fit a single set of data, and it would be invidious to ascribe weights to the various sums of squared residuals to obtain one statistic suitable in all instances. Expression (1) may not actually be minimised, therefore, for a given set of data. (In the case of a single FLM curve too a non-optional set of estimates may be

useful if a precise estimate of a particular parameter is desirable.)

This leads on to the second reason for rejecting automatic techniques, namely that one of the benefits of simulation is the opportunity it gives to study the effects and interactions of the various parameters involved. This advantage is lost if the method is too computer-orientated; properly the computer is a tool in simulation studies and should be used in an interactive situation whenever possible, with the model builder in full control.

Thirdly, it is felt that automatic techniques involving sophisticated numerical optimisation procedures may give a spurious illusion of accuracy. It is probable that any apparent improvement in estimation of parameters which an automatic method provides will be essentially model-dependent, reflecting, for example, the particular distributions chosen for phase times, rather than offering any further description of the system being modelled.

Finally it must be recognised that where a variable is capable of sudden alteration the technique of minimising the residual sum of squares may be inappropriate. There is little justification for assuming that the indices follow a suitable distribution in the presence of considerable variation in the rate of response of the animals used. The transformations discussed do not take this variation into account, dealing only with the sampling variance. It is unavoidable, however, that each data point is produced by a different animal, and a correct analysis would involve the calculation of some norm by which the "distance" of a point from the simulated curve could be measured. It may well be that fitting the curve by eye, however prone to subjective error, will yield better results than the use of least squares methods including transformation of data.

5.4. RESULTS AND DISCUSSION

5.4.1. THE CASTRATE MOUSE SEMINAL VESICLE.

Figures 5.5. to 5.10. (pages 78 to 83) show the output of the model. For figures 5.5. (a) and (b), the labelling and mitotic indices, the experimental data for the seminal vesicle is presented. This data has been fitted in the manner described, and figure 5.6. shows the 2741 terminal print-out which produced the results.

Table 5.7. gives analyses of variance for the best fit to the data which has been obtained. Since the durations of G_1, S and G_2 influence the mitotic index curve only through their sum, and since the number of cells in the cycle may be adequately determined from the continuous labelling curve alone, the 9 parameters of the model give rise to only 5 degrees of freedom for the I_m curve fitting. In the same way there are 5 degrees of freedom for the labelling index case. Three parameters are directly involved in constructing both curves; namely the mean time in G_0,

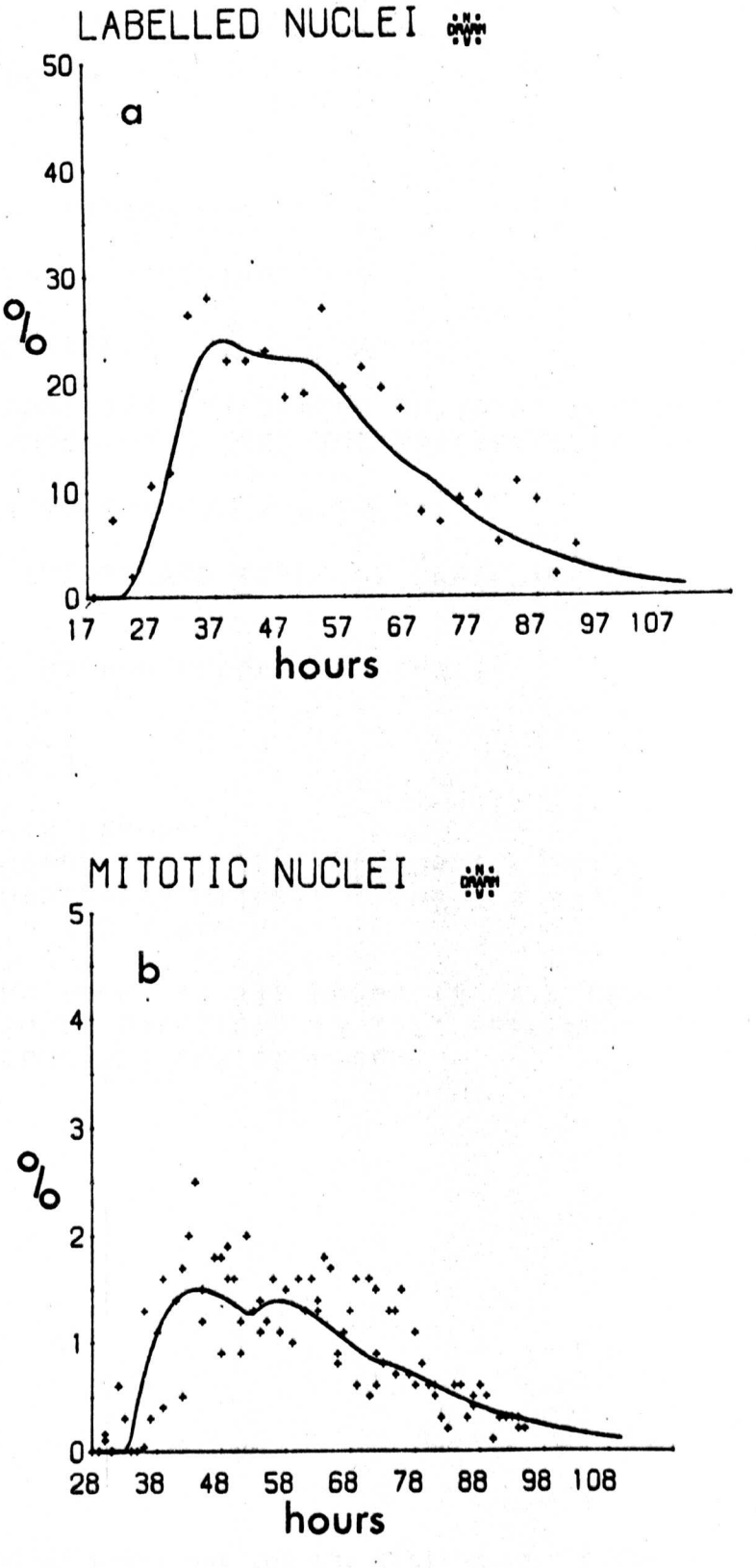


Figure 5.5 (a) Labelling and (b) mitotic index curves produced by the model using the parameters given in figure 5.6. The experimental points for the seminal vesicle are also shown.

```
 DURATION OF EXPERIMENT?
† 110
 MEAN TIME IN G0?
† 28
 PERCENTAGE OF CELLS LEAVING G0?
† 100
 PARAMETER FOR ERLANGIAN DISTRIBUTION?
† 2
 NUMBER OF CYCLES POSSIBLE?
† 3
 DOES DECYCLING PROBABILITY (X) DEPEND ON TOTAL NUMBER OF CELLS
 (T) OR NUMBER (D) WHICH HAVE LEFT THE PROLIFERATIVE CYCLE?
† "D"
 PARAMETERS FOR RELATIONSHIP OF X WITH D?
† 0.75    0.00075
 NUMBER OF FLM EXPERIMENTS AND TIMES OF LABELLING?
† 2
† 24    48
 DURATIONS OF G1, S, G2 AND M FOR EACH CYCLE?
†     22    10    1.3    0.7
†      7    10    1.3    0.7
†      7    10    1.3    0.7
 EXPERIMENTAL DATA?
† $CONTINUE WITH SEMVES RETURN
 RESIDUAL SUM OF SQUARES FOR LABELLING CURVE = 153.8
 RESIDUAL SUM OF SQUARES FOR MITOSIS   CURVE = 317.7
 FINAL VALUES OF X, D AND T ARE
 0.93   236   AND 241.
 CUMULATIVE LABELLING INDEX AT 110 HOURS IS 99.4 PERCENT
 MAXIMUM VALUE OF GROWTH FRACTION" IS 71.7 PERCENT
 ENTER "STOP" IF GRAPHS NOT NOW REQUIRED
†."CONT"
```

Figure 5.6. Computer terminal print out for the fitting of the seminal vesicle data. The user enters the lines marked †. Experimental data was contained in a file "semves". The decycling probability ($\vec{\eta}$) depended on the number (d) of differentiated cells: $\vec{\eta} = 0.75 + 0.00075d$.

Labelling Index

Source of Variation	Sum of Squares	D.F.	Mean Square	Variance Ratio
Regression	971.3	5	194.26	47.12
Residual	107.2	26	4.12	
Total	1078.5	31		

Mitotic Index

Source of Variation	Sum of Squares	D.F.	Mean Square	Variance Ratio
Regression	428.2	5	85.64	26.30
Residual	302.8	93	3.26	
Total	731.0	98		

Figure 5.7. Analyses of variance using the weighted arcsine transformation for the best set of parameters discovered for the fitting of the seminal vesicle labelling and mitotic index data.

The theoretical residual variance due to sampling of cells is 0.25, an indication that there is considerable inter-animal variation.

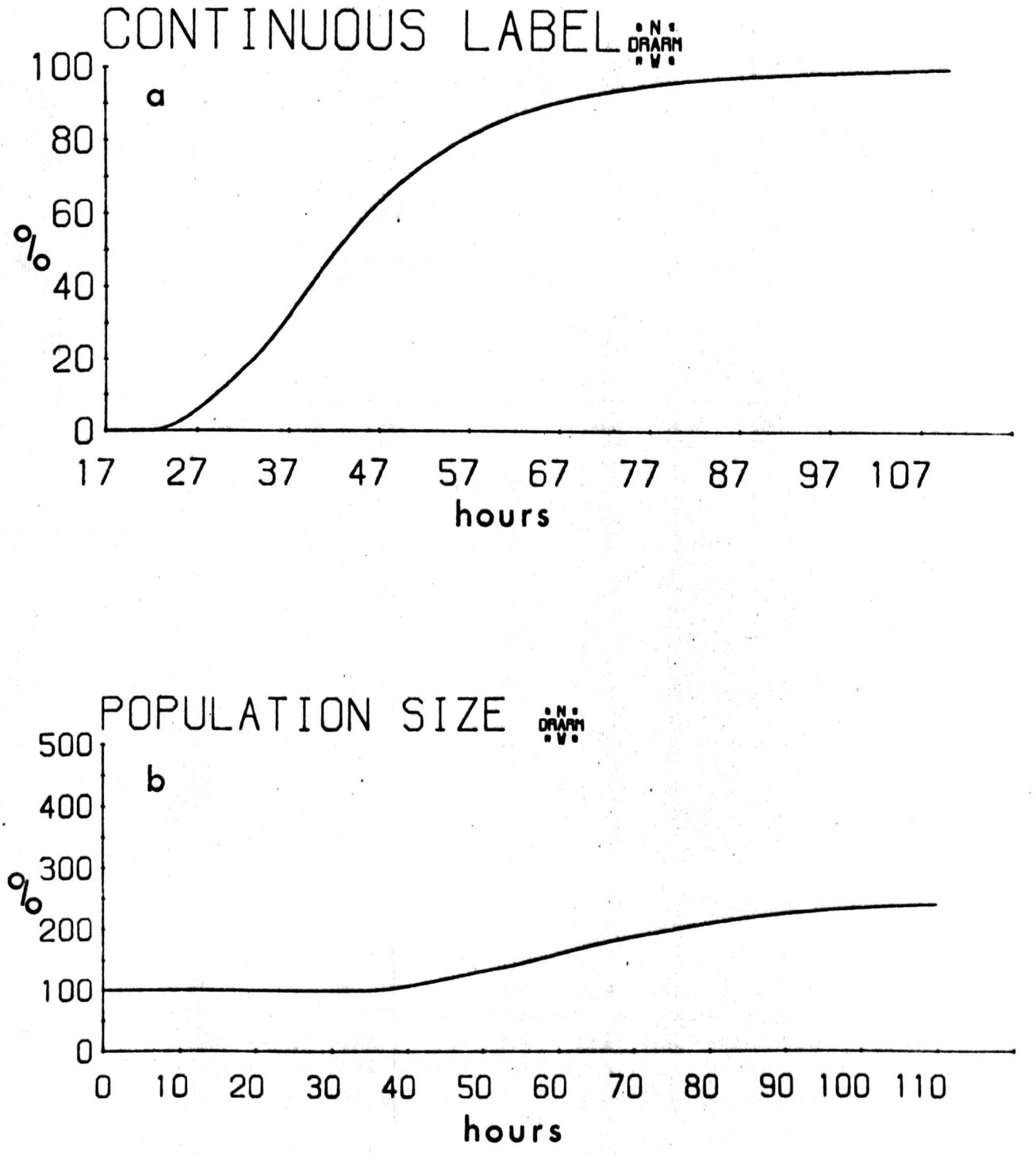


Figure 5.8. The curves predicted by the model (using the parameters given in figure 5.6) for experiments to investigate (a) the continuous labelling index and (b) the number of cells in the population as functions of time after stimulation of the system.

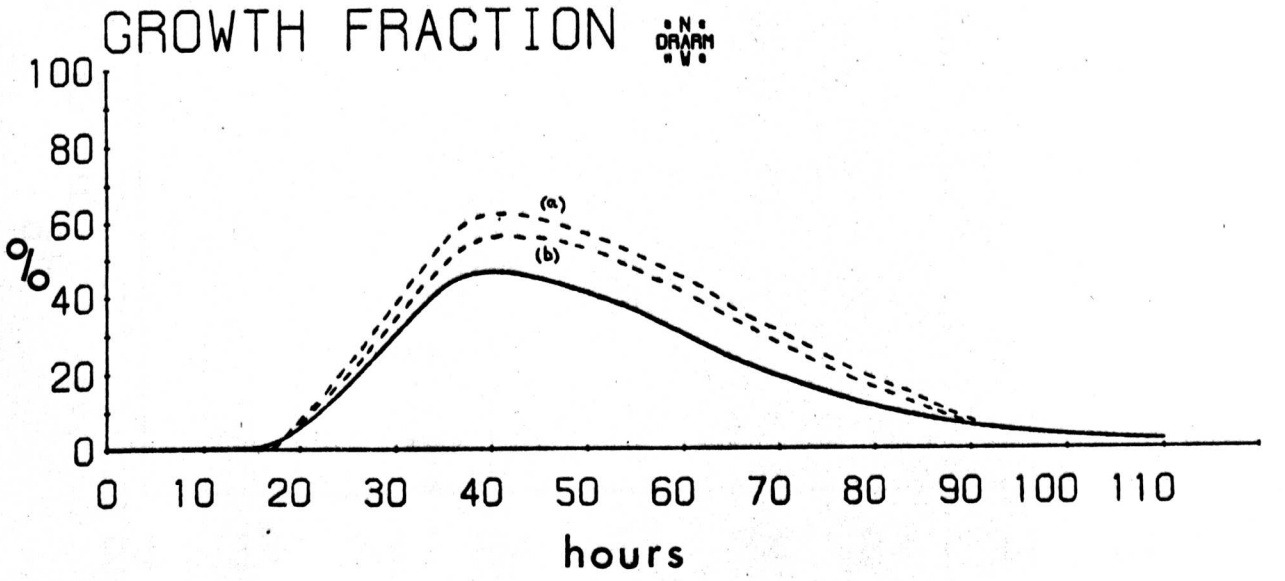


Figure 5.9. The continuous line represents the growth fraction calculated by the model using the parameters of figure 5.6. The dotted lines correspond to the growth fraction calculated from the observed seminal vesicle labelling index curve assuming (a) exponential growth and (b) steady state conditions.

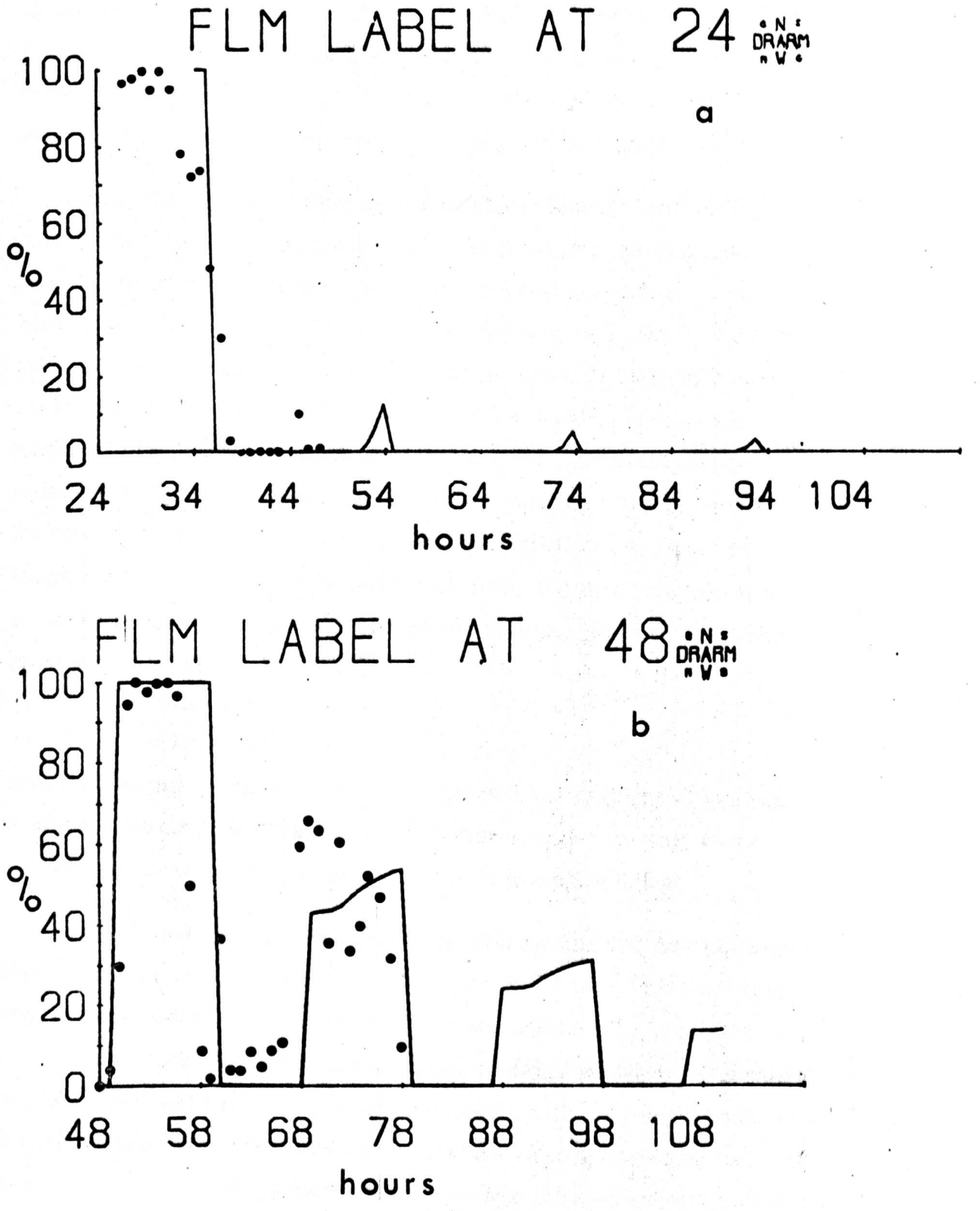


Figure 5.10. Fraction labelled mitoses curves for a pulse label at (a) 24 and (b) 48 hours after stimulation. The experimental data points for the seminal vesicle have been superimposed on the output produced by the model using the parameters of figure 5.6.

and those parameters expressing the relationship between the decycling probability and the number of differentiated cells. Experience with the model has made it clear that the I_L and I_m data are satisfied simultaneously by values of these parameters. While very small changes in a parameter may raise one residual variance and lower the other, changes which would be considered important cause the residuals to increase or decrease together.

Figures 5.8. (a) and (b) show the curves which should be produced by experiments which investigate the variation with time of the continuous labelling index and the total cell numbers. Figure 5.9. is drawn so that the conventional growth fraction can be examined. Because it is not possible to distinguish between cells in G_0 and G_1 the growth fraction calculated by the model may be defined in several ways. All cells which have left G_0 and are still in the cycle could be considered to be proliferating, or cells which have not yet reached the first S-phase could be excluded. In fact an intermediate position has been chosen, whereby only those cells which will not reach the first S-phase in a time equal to the second G_1 phase are excluded. This is equivalent to an assumption of G_1's of equal length and G_0 a constant plus an Erlangian variate. The growth fraction curve given by the model is rather lower (as is to be expected) than that produced by the relationship given by Cleaver [4],

$$I_p = \frac{I_L \text{ (observed)}}{I_L \text{ (theoretical)}} \quad ,$$

when the theoretical index is calculated either on the assumptions of exponential growth or steady state conditions. The agreement is closest during the early stages of proliferation and becomes worse as more and more cells fail to proliferate further.

Figures 5.10. (a) and (b) are FLM curves, with experimental data superimposed on them. Figure 5.10. (a) shows especially clearly the dangers of using FLM techniques to estimate cell cycle parameters if the growth fraction is not constant. The large influx of unlabelled cells reaching their first mitosis around 54 hours swamps the few labelled cells entering a second mitosis, and leads to the erroneous impression that only one division is undergone. It has been found that, when labelling is more than about 60 hours after the initial stimulation of the system, modelled FLM curves bear little resemblance to real life in second and subsequent peaks, because for this aspect of the model variation in cell cycle parameters is no longer overshadowed so effectively by the variation in the length of the G_0 phase. Two such experiments are therefore convenient; one near the time that cells begin to arrive in S, and one at the peak of DNA synthetic activity.

5.4.1.1. EXPERIMENTAL METHOD

The experimental protocol followed is fully documented in the report by Morley et al [10] referred to in section 5.1.3. Briefly it is as follows. Male mice from an inbred strain were castrated at 3 months and experiments begun 2 weeks later. Each animal was injected with 250μg of testosterone propionate every 24 hours till death. At intervals varying from 1 to 3 hours over 4 days animals were given tritiated thymidine and killed after an hour. Autoradiographs of the seminal vesicle were prepared, and 2000 cells were counted to obtain estimates of the labelling and mitotic indices. If necessary additional mitoses were examined so that the FLM could be calculated from at least 100 mitoses.

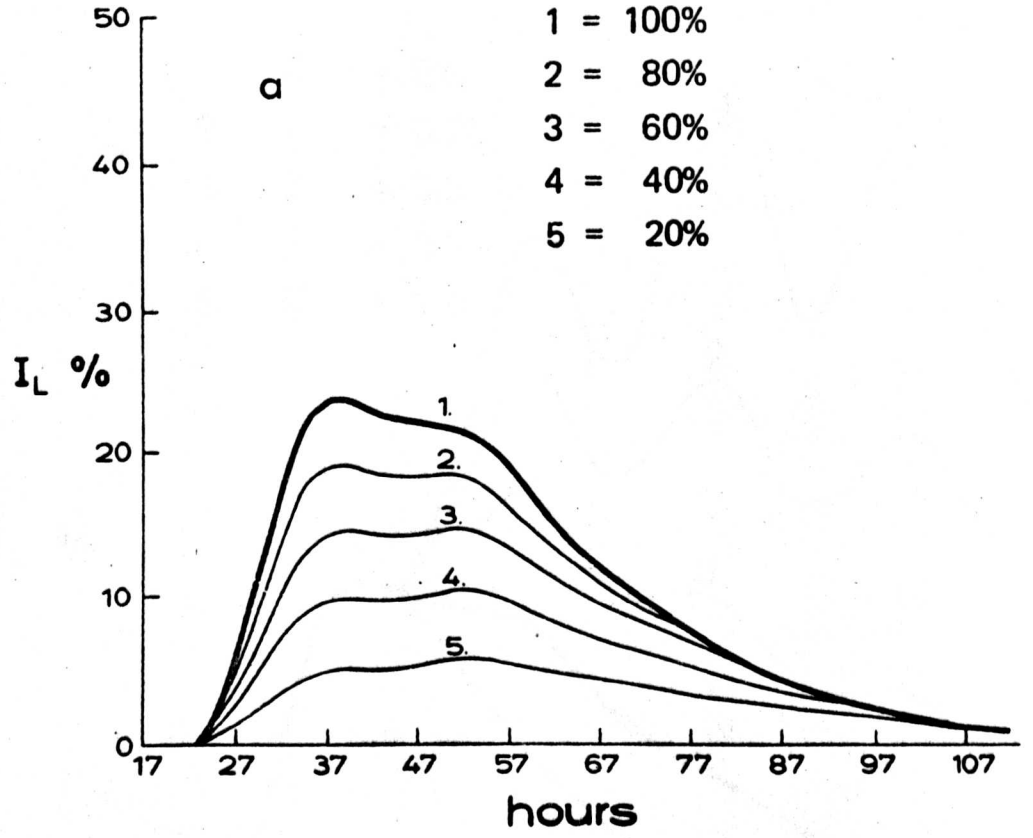
5.4.2. RESULTS OBTAINED WITHOUT DATA

One of the most profitable uses of the model is to observe the effects on the output curves of systematically altering parameters of the system. Figure 5.11. (page 86) shows the labelling index and continuous labelling index curves for the same parameters as were used to fit the seminal vesicle data, except that the percentage of cells entering the cycle varies from 100% to 20% in steps of 20%. The same effect can be observed in the mitotic index.

Figure 5.12. (page 87) again uses the parameters of figure 5.6. except that the decycling probability is kept constant, taking the values 0(0.25)1. It is clear that the decycling probability has a great influence on the shape of the resulting curve, and it is likely that it can be estimated fairly closely.

Classical ideas of growth control such as the Weiss-Kavanu model propose a negative feedback mechanism from the functional or differentiated compartment acting upon the proliferative compartment to control the rate of cell production. This feedback may depend on the number of cells in the differentiated compartment or on some related variable. In the present experimental situation an induced peak in the proliferative indices occurs, followed by a steady decrease despite the continuation of the stimulus. It is possible that any form of feedback producing this response operates via the decycling probability. In figure 5.13. (page 88) are shown I_L curves when $\vec{\eta}$ is made proportional to the number of cells in the differentiated compartment, with $\vec{\eta}$ changing at different rates. Such differences are reflected in the shapes of the curves. The factors determining the initial value of $\vec{\eta}$ are of great interest but are, as yet, obscure.

Figure 5.14. (page 89) illustrates two points; firstly that labelling and mitotic index curves can be produced which have more than one distinct peak. This is achieved by using a shorter mean time in G_0 and a longer first G_1 than was required for the seminal vesicle;



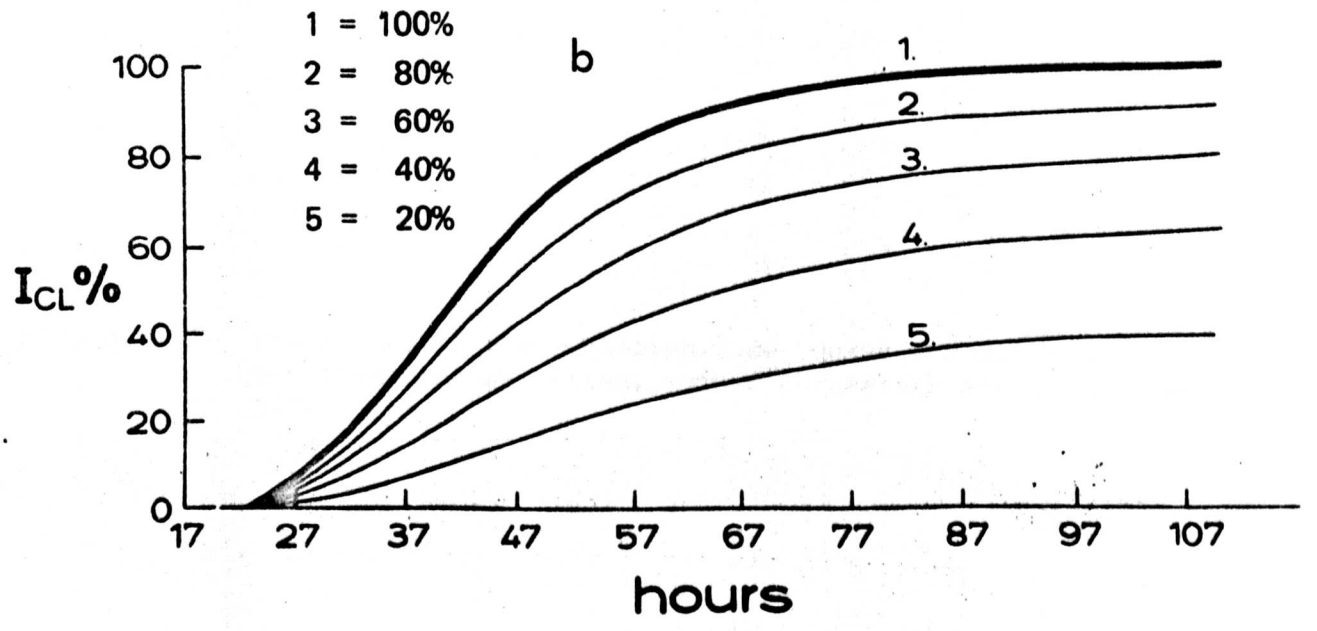


Figure 5.11. The change in (a) labelling index and (b) continuous labelling index curves, for different percentages of cells leaving G_o. Other parameters are as in figure 5.6.

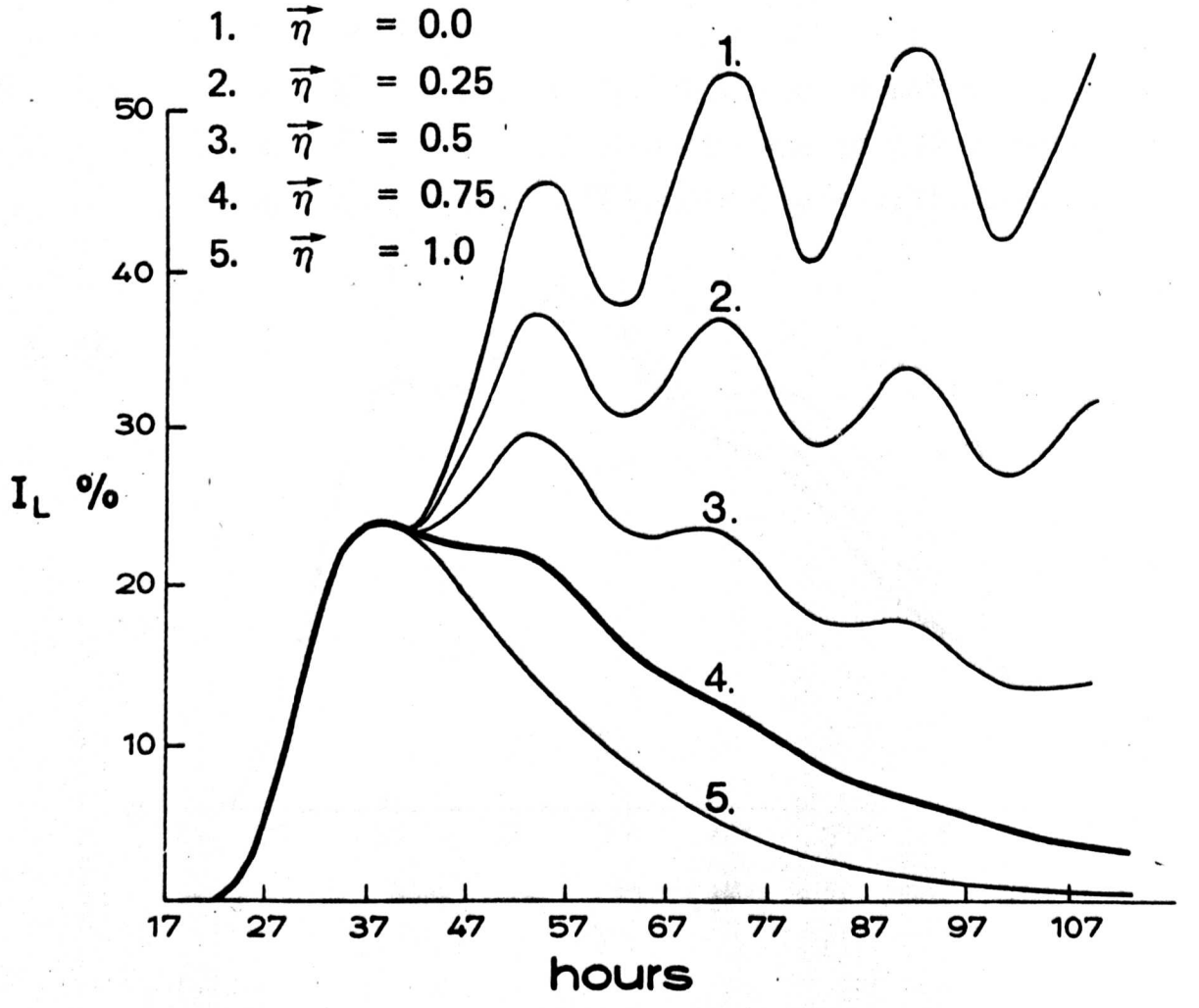


Figure 5.12. The change in the labelling index curve for different decycling probabilities. Other parameters are as in figure 5.6.

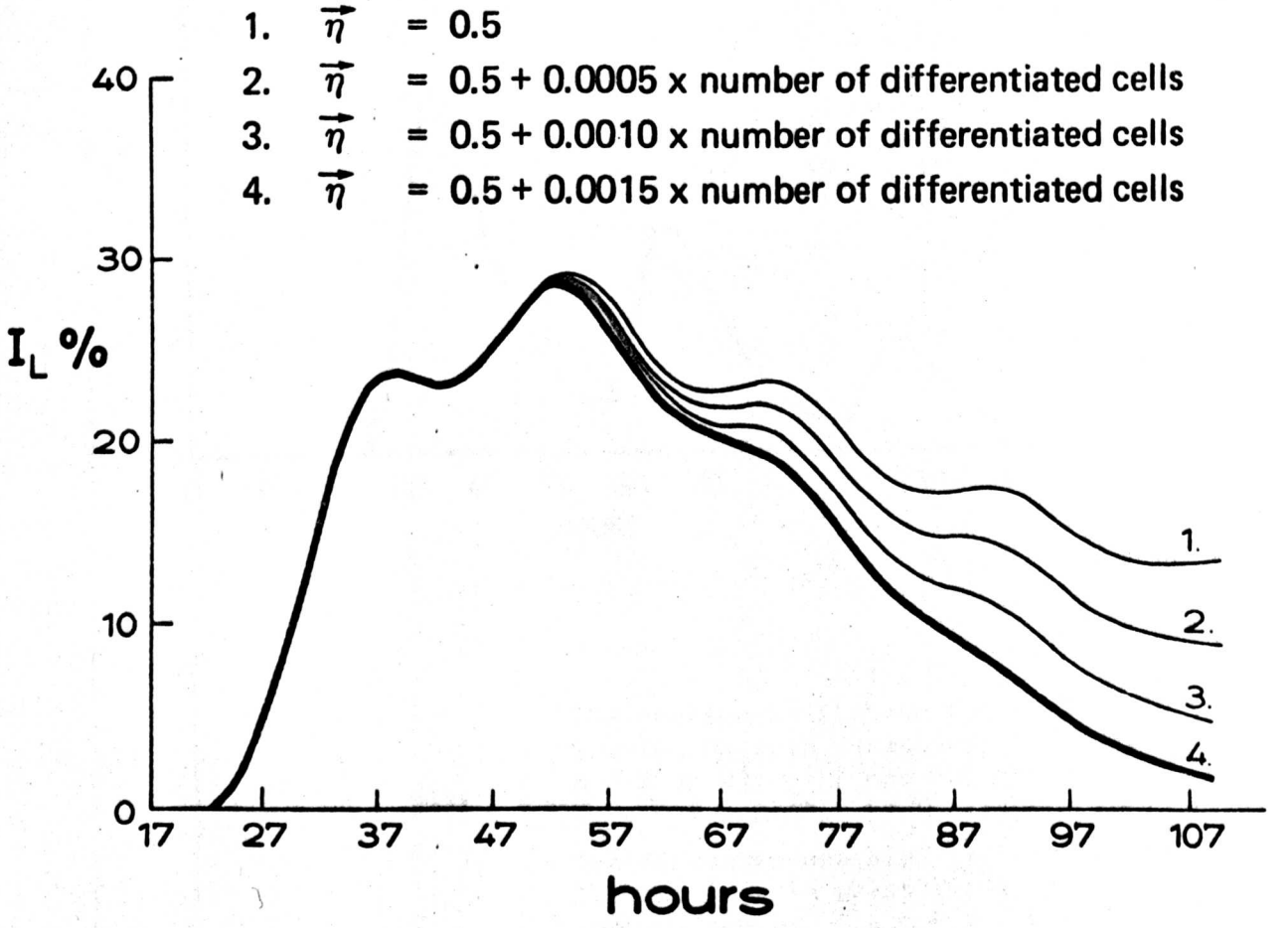


Figure 5.13. The change in the labelling index curve for different relationships between the decycling probability and the number of differentiated cells. Other parameters are as in figure 5.6.

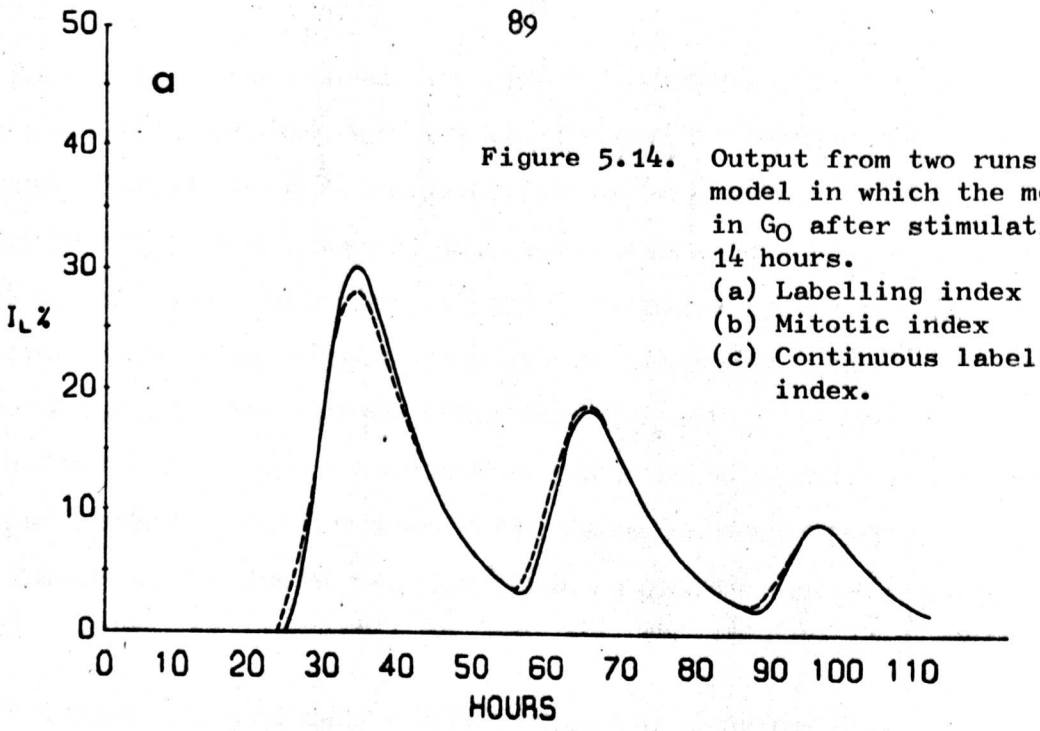
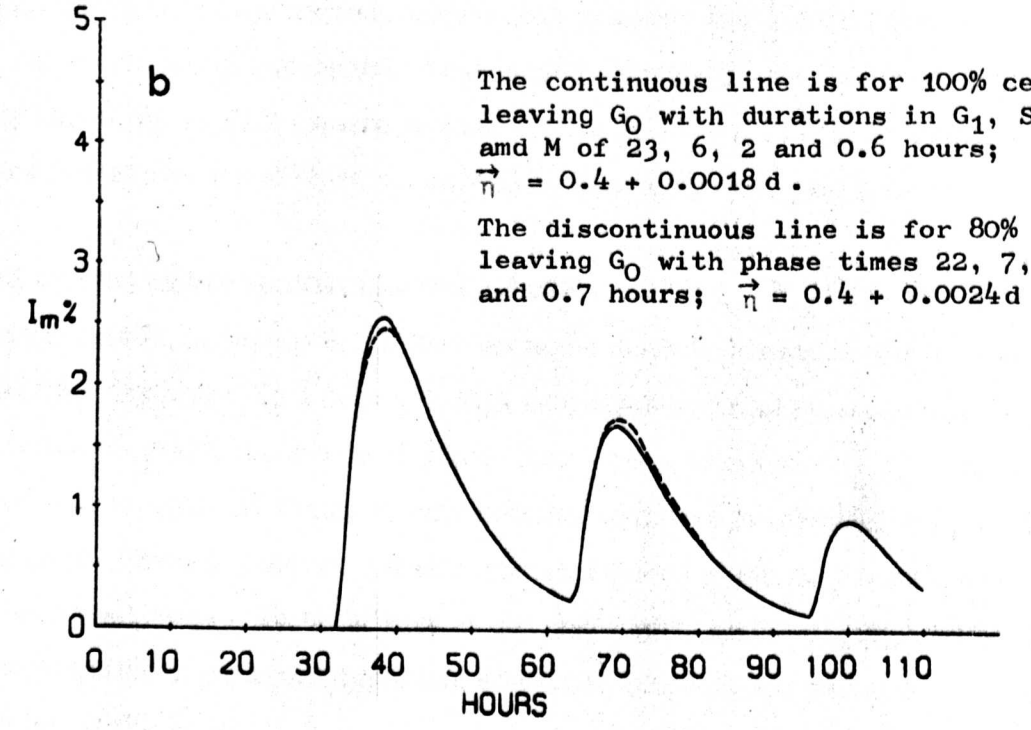
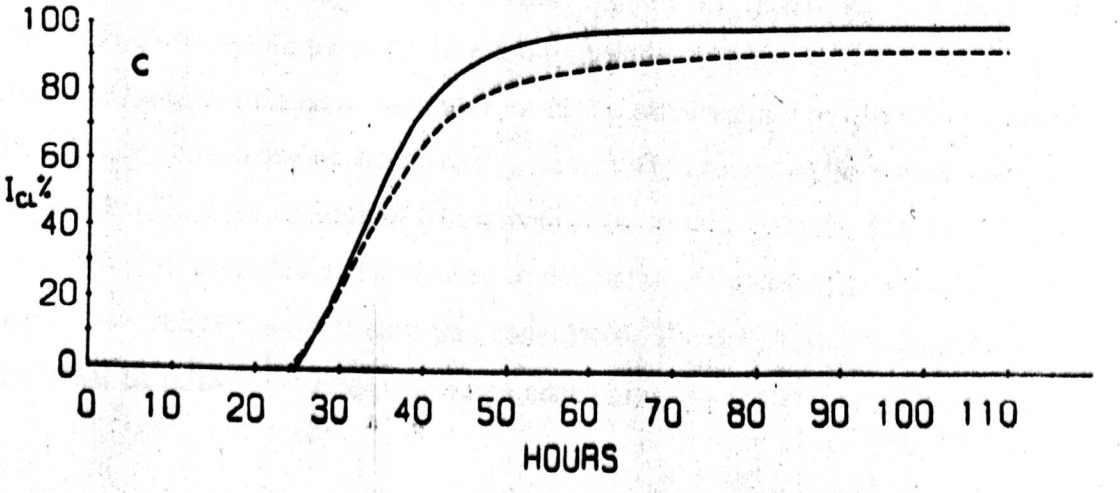


Figure 5.14. Output from two runs of the model in which the mean time in G_0 after stimulation was 14 hours.
(a) Labelling index
(b) Mitotic index
(c) Continuous labelling index.



The continuous line is for 100% cells leaving G_0 with durations in G_1, S, G_2 amd M of 23, 6, 2 and 0.6 hours; $\vec{n} = 0.4 + 0.0018d$.

The discontinuous line is for 80% cells leaving G_0 with phase times 22, 7, 2 and 0.7 hours; $\vec{n} = 0.4 + 0.0024d$.



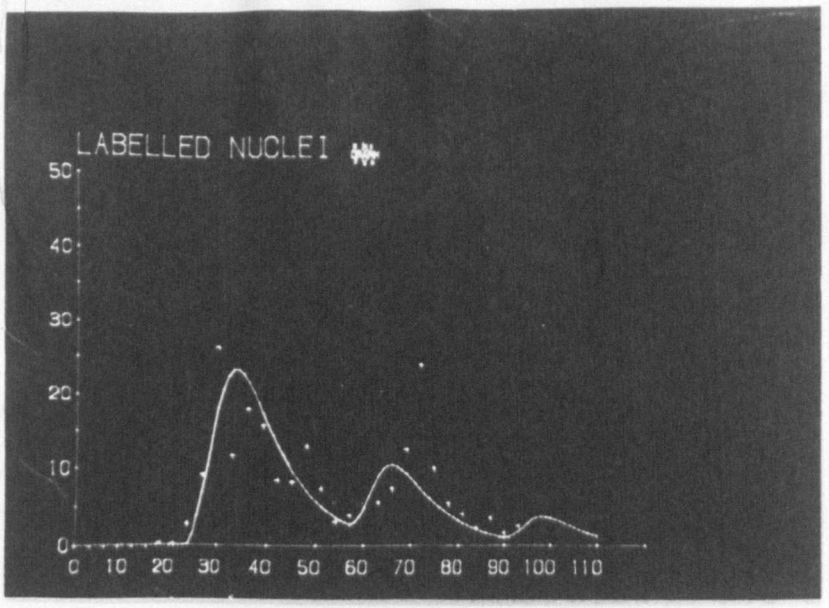
in fact the mean G_0 time was halved. This gives the distribution of times up to the first S phase a smaller variance and shows how the model is able to represent different degrees of synchrony. Secondly, it can be seen that the two curves in each of figures 5.14. (a) and (b) cannot be distinguished between as far as ability to represent a set of data is concerned, although the parameters used are quite different (see legend of figure 5.14.). The use of such curves therefore cannot adequately separate the effects of an increase in the number of cells participating in the cycle and a lengthening of t_s and t_m. The continuous labelling curves in figure 5.14. (c), however, are distinguishable, and so experiments giving these curves can be used as a basis for further investigations. The results given for the seminal vesicle must therefore be treated with caution until the outcome of further experimen tation is known.

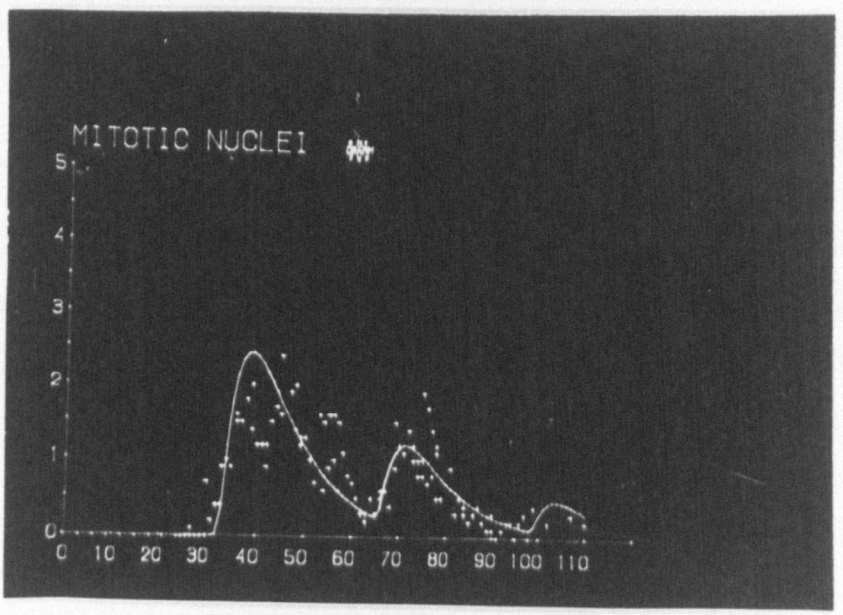
It is in this area that the model best demonstrates its usefulness. It can indicate which experiments would be most informative; it can determine at what stages of an experiment it might be important to take frequent observations to ensure that sudden fluctuations in I_L or I_m curves did not go undetected; it can decide when the growth fraction is likely to be steady enough for an FLM experiment to be carried out; and it is open to verification by suggesting further avenues of research, as well as estimating cell cycle parameters.

5.4.3. THE CASTRATE MOUSE COAGULATING GLAND

It was suggested in section 5.1.3. that the model described is applicable to many cases of induced DNA synthesis. This occurs through exogenous hormonal stimulation of a target organ in animals in which the source of trophic hormone has been removed, for example, stimulation of the uterus and vagina by administering oestrogen to castrated animals; and of adrenal cortex, thyroid gland and gonadal tissues deprived of pituitary trophic hormones. Further conditional renewal systems, such as are produced by partial hepatectomy or the induced proliferation of phytohaemagglutinin stimulated lymphocytes, may also be amenable to description by the model.

As preliminary evidence in support of this view, figure 5.15. (page 91) is presented, showing curves drawn by the model to fit data relating to the castrate mouse coagulating gland stimulated by administration of testosterone in the same manner as the seminal vesicle. Despite the different appearance of the I_L and I_m data in the two cases the model, as explained in section 5.1.2., is capable of fitting both sets. In this instance continuous labelling data was also available and was used to determine the percentage of cells leaving the G_0 compartment, before the other data was considered. The difficulties encountered in obtaining an exact fit raise interesting questions concerning the possibility of correlation





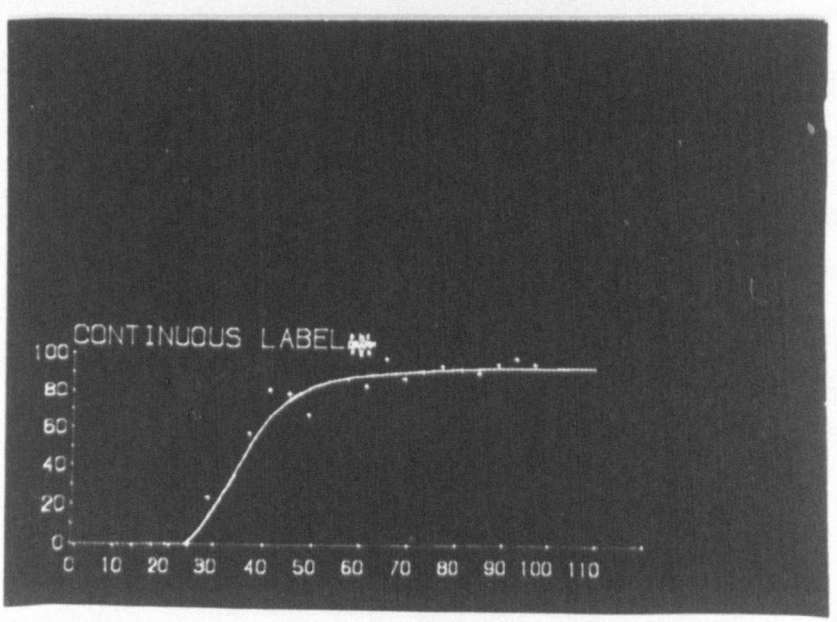


Figure 5.15. Fitting data for the castrate mouse coagulating gland.

between the phase times of a cell and its daughters, but much more extensive data would be required for an investigation.

Unlike the other computer-drawn graphs in this thesis, which were reproduced from hard-copy output, figure 5.15. was derived from photographs of the Computek screen.

5.5. THE VALUE OF AN INTERACTIVE MODEL

The increased efficiency and value of the interactive use of the cell cycle model in comparison with what could have been obtained in a batch environment is mainly due to the consequent ability to use the Computek graphical display equipment. To obtain the I_L, I_m and I_{CL} curves which each take about 10 seconds on the Computek requires about 20 minutes using the hard-copy graph plotter. It is true that this can be improved on by using a line-printer representation of the graphs, and this was done at one stage in the program development before the Computek was operational. The quality of the graphs, however, along with the increased convenience of terminal use, made the Computek much the more rewarding method. There comes a time when an otherwise feasible method becomes unworkable purely because of slowness and inconvenience, and without interaction this simulation would have passed into the latter category. Communication with the biologists whose data was used would have been intolerable had it involved studying successive sets of graph plotter or line printer output at widely separated intervals. Instead many examples could be discussed at the console. Comparison of several runs could be obtained in a form similar to that of fig. 5.11. as it was possible to superimpose pictures on the Computek. Unlike the genetic application, where the graphs were drawn directly on the Computek by the program, it was found appropriate to store the file produced, then to plot it and subsequently to return to the simulation. This procedure was necessary because one set of data gave rise to several graphs, the first of which might be wanted for a re-examination after the second and third had been viewed.

Such a procedure was specially valuable because when investigations into the system were begun very little information was available. Understanding of the system through frequent use of the model came rapidly, however, and it was soon possible to make use of the data for which the model had been built. As discussed in section 5.3.1., although residual sums of squares were calculated for fitted curves, the method is not ideally suited to the situation in question, and visual methods of curve fitting were adopted. This method would have been unthinkable without the interactive facilities.

To save unnecessary calculations being performed the program gives the user the option to suppress graphical output and retain only the 2741 print-out as shown in fig. 5.6. This was

most often done when a relationship between the decycling probability ($\vec{\eta}$) and the number of differentiated cells was chosen which led to $\vec{\eta}$ becoming greater than unity. Such a contingency arose not infrequently as it is likely that $\vec{\eta}$ will be close to 1 at the end of the simulated period.

6. COMPARISON OF THE THREE MODELS

6.1. SOPHISTICATION OF THE MODELS

Of the three models presented the cytokinetic one is by far the most sophisticated. It is equally capable of being used to provide insight into the system which it represents and of estimating the parameters of a particular manifestation of that system; as a research-orientated model this is critical since the system under investigation is naturally imperfectly known. Although most of the concepts involved in the formulation of the model are accepted by cell cycle kineticists this model is the first to place the emphasis for feedback control of differentiation on the decycling probability – which term has in fact been coined for the purposes of the model. The model has been carefully designed so that the results of several experimental investigations may be simulated, and it has suggested other experiments which could be used to check its validity. This is unfortunately not the case with the genetic model, which in its use of a differential equation to model the mode of inheritance of qualitative characteristics is at first sight so similar. Too many assumptions have to be made in the gathering of data for the latter model for it to be considered in the same light, and methods are not available to estimate sufficiently accurately the rate of migration or the effective population size. Nevertheless the genetic model is sufficiently faithful to reality to portray the gene flow in a population at a level suitable as an introduction to students. No impressions taken from the model will have to be unlearned; only the actual fitness values must be treated with caution.

The situation in the room usage model is rather different from the two biological ones. Here, the data collected was as far as could be ascertained an exact description of the system; but essentially that data was used to estimate a great many parameters, namely the proportion of students of any given registration code who would attend a particular lecture, and these proportions were extrapolated into the future. Also, the method used of estimating possible student numbers in succeeding years was crude in the extreme. Thus the model was very accurate for the year of the data collection but of unknown worth for subsequent years. It is, of course, possible to assess the fidelity of the model by carrying out its recommendations and observing how closely the system follows its predicted course.

6.2. DEGREE OF INTERACTION

It was pointed out (section 3.5.) in discussing the room utilisation model that interaction with the computer could take place in more than one way. Interacting with a program is a feature of the use of a terminal system, although even in the early days of computers an iterative sequence, for example, could be influenced by the user, who was able to interrupt

program execution, alter the contents of registers and restart the program. The new-found facility for such actions is however a product of time-sharing systems on third-generation computers. The next degree of interaction involves inspection of results of a completed program followed by loading and running the same or a different program, either choosing the program or the data for a rerun on the evidence of the new results. This has always been a common technique, but the advent of terminal systems has so improved turnround for short jobs with limited output that a new dimension is added to the process. In fact the advent of display screens for alphameric and graphical output has gone a long way towards lifting the restriction on the quantity of output which may conveniently be handled, and file-based systems like MTS place practically no upper bound on the amount of input which may be processed.

The speed of execution remains the critical factor in whether it is profitable to design a system for interactive use. Many models run quickly because the calculation involved in them is to some extent controlled by the application of Occam's razor. Certainly, remarkably accurate data would have to be available before a more complicated biological model could be studied than the cytokinetic one presented here.

Given that the models which have been described are suitable for interactive use it is appropriate to consider what type of interaction has been used in each and for what reasons. In the room usage model interaction performed two main functions. Firstly it made the system more accessible to users who had no computer training by leading them with questions and allowing them to choose from a set of instruction what the computer's next task should be. Secondly it enabled human decision making to be incorporated into the model, thus saving considerably on data collection and coding and easing the difficulty of having computer based decisions implemented. Interaction was mainly within one program, but opportunity was made of a logical hierarchy in the model to construct a suite of programs to save on the virtual memory required at any one time.

In the genetics teaching situation a computer model was used to avoid the need of presenting the underlying mathematics to the students. An interactive environment was chosen for running the model so that any set of parameters could be investigated quickly and student involvement be encouraged. Each of the three stochastic genetic models is used quite independently, but after each run of a program data is again requested and the parameters entered are chosen in the light of the graph which has just been drawn on the Computek.

A slight difference in approach was used with the cell cycle model out of necessity.

An interactive model was chosen again as a means of obtaining greater familiarity with the effects of parameter changes and as the best way of engendering interest in the model in the biologists with whom communication was maintained. Because several graphs were produced by each run of the program the output had to be stored, and viewed by a separate program. The interactive nature of the viewing program itself enabled graphs to be superimposed for easier comparison, a facility which could also have been used in the teaching application at the expense of such clear labelling of the individual graphs. One of the most important uses of interaction in the cell cycle model was in making feasible the non-automatic curve fitting procedure adopted.

It is impossible to say in which model "most" interaction was used or in which it played the most important role. Without interaction all three applications would have been qualitatively different and would have been less successfully completed.

6.3. VALUE OF THE MODELS

Since much of the research conducted in the investigation into the use of interactive models has been in fields other than computer simulation, it is pertinent to summarise the facts which have come to light in these other areas and to evaluate the worth of using interactive models by calculating the size of the contribution they have made in these disciplines.

In the administrative sphere the value of the model is, as yet, difficult to assess. It has indicated that the rooms available are capable of accommodating more students than was previously imagined possible, but this finding has still to be verified when student numbers are actually raised. Perhaps the main value of the room usage model is that it has demonstrated to the planners that such interactive simulations have a legitimate part to play in helping them to formulate strategies for the future.

The Monte-Carlo model of gene flow in populations did nothing to increase knowledge about the system it represented, but the stochastic model, in addition to furnishing a new derivation of the equilibrium distribution of gene frequencies at a multi-allelic locus under pressure, showed that it was not unreasonable to accept the present day distribution of ABO gene frequencies as resulting from selective and migratory pressures. It also demonstrated the value of using the diffusion equation model with fast graphical output as a method of teaching. This approach can be adopted in many fields and should prove especially valuable when there is no need for students to follow the mathematics involved but where they must be able to understand the effects of the parameters which influence the system.

It is in the application to research in cell cycle kinetics that the type of model investigated has provided the most interesting and useful results. This is not surprising since the system being investigated is not yet fully understood, and a model which can reproduce experimental results is liable to lead to a considerable gain in understanding of the system. It is clear that in the conditional renewal systems which were studied some cells divide more than once; the emphasis placed by the model on the concept of decycling probability is therefore valuable. On the hypothesis that this quantity depends on the number of differentiated cells the model was able to represent experimental data quite adequately. It therefore suggests that grain-count experiments be conducted to test the existence of such an association. Such experiments will themselves lead to extensions of the model so that their outcome can be predicted in the same way as the model already predicts the outcome of continuous labelling and DNA measurement experiments. The model increases its value by not only describing a possible mode of functioning of the simulated system but also encouraging and aiding further research.

7. THE USE OF SPECIAL LANGUAGES

7.1. GPSS and CSMP

The suite of programs which model room usage was written in ALGOL 60, a language which proved entirely satisfactory for the purpose, allowing convenient handling of data in arrays and, in the version used, permitting flexible input and clear output. Execution was sufficiently rapid to give the impression that responses occurred immediately.

The biological models were programmed in ALGOL W and FORTRAN as these were the only two languages which could be used in conjunction with the Computek. They also proved satisfactory in terms of programming facility and speed of execution. It is relevant to inquire whether a special purpose simulation language would have given any further benefits, and two of the candidates for consideration are IBM's GPSS, which is for simulations of discrete systems, and CSMP, the continuous system modelling program.

The preferred model for GPSS is of a queuing system, and, since this involves birth and death processes analogous to those in human populations, the language may clearly also be used in population genetics applications. As indicated in section 4.6.2., the programming is made very easy by the special language, although GPSS may be criticised on the score that certain calculations not performed automatically by the program may be tedious to carry out. However, but for the fact that the Monte-Carlo method is so unsuitable for interactive use, GPSS might have proved an attractive language in which to write a genetic model. It is, of course, unnecessary to make use of the GPSS random number generator; deterministic models of time-dependent systems may also be programmed conveniently in GPSS.

CSMP offers similar advantages in the simulation of the cytokinetic system. It could have been used to solve the differential equations for that system with very little expenditure on programming time being required of the modeller. It would also have been possible, had circumstances required it, to access the Computek from a CSMP program.

In an interactive environment, however, both of these simulation languages suffer from a considerable disadvantage which outweighs their seeming attractions. Their input/output capabilities are not at all well suited to terminal use. It would no doubt be possible to redesign these in such a way that parameters for each run

could be entered as data from a terminal and output made more concise, but at present the extent and format of the output are such that the packages are suitable only for batch use. The possibility of transmitting output to a file which can subsequently be inspected for the necessary figures is not appealing, because although some users would find this feasible the kind of person for whom the models were written cannot be expected to tolerate this imposition.

It is apparently the case, therefore, that simulation packages designed for batch use are not appropriate in an interactive environment, at least when one of the aims of the programmer is to enable users who are unfamiliar with the packages and with computers in general to attain facility in running the programs. Writing a specialised program with directly relevant output is more likely to be satisfactory than attempting to adapt a general system.

7.2. APL

APL is a programming language designed for conversational use. It is both elegant and easy to write, so that, having once used it, many programmers wish to write in no other language. It is necessary to ask whether they should in fact use APL for interactive modelling.

Because of the ease of programming in the language, APL is an extremely useful vehicle for testing the logic of a model. The model of the cell cycle which can be expressed in a closed form (constant decycling probability, section 5.2.4.) was programmed in APL for this reason before the model was converted to ALGOL W to run using the Computek. At the time the method was inconvenient as the APL output had to be edited before being used as input to a program giving graphical output, but when graph-plotting capabilities are added to APL, in the version of MTS operated, it will form a most convenient language for simulation. The decision whether to use it in any given instance will depend largely on the amount of calculation involved in the model, as speed of execution has, to some extent, been sacrificed for speed of programming and debugging. It is also important that a way be found around the difficulty of using APL at a CRT device,

which is brought about because the character set used by APL is peculiar to itself; recent developments in micro-programmed generators for CRT terminals seem sure to achieve what is needed.

8. CONCLUSION

Computer simulation of dynamic systems may advantageously be performed in an interactive environment if the execution time of the model is short enough to admit of frequent transfers of control to the user. Such considerations would appear to limit the value of Monte-Carlo simulations and to emphasise the utility of formulating deterministic and stochastic equation models.

In administrative, teaching and research contexts the use of interaction facilitates communication with those whose systems are being simulated, especially if fast alphameric and graphical output capabilities are available.

The familiarity with a system which is gained from an interactive simulation is particularly valuable in dealing with biological systems whose actions are imprecisely known.

REFERENCES

Because of the diversity of the subject matter covered, references in the text have been numbered independently in each chapter. The format of the thesis is such that almost all references occur in chapters 3, 4 and 5.

CHAPTER 1

1. Davies, R.G. "Computer Programming in Quantitive Biology"
Chapter 13 : "Models and Simulation".
Academic Press (1971) p.376.

2. Taylor, D.E.M. "Simulation studies using analog and hybrid computers".
In "Principles and Practices of Medical Computing".
ed. Whitby & Lutz.
Churchill Livingston (1971) p.289.

CHAPTER 3

1. Holz, R.E. "The GASP Manual". Massachusetts Institute of Technology (1963).

2. Hopkins, D.S.P. "On the Use of Large-Scale Simulation Models for University Planning". Review of Educational Research 41 (1971) p.467.

3. Longworth Smith, R. "Accommodating Student Demand for Courses by Varying the Classroom-Size Mix". Operations Research 19 (1971) p.862.

4. McDiarmid, C.J.H. "The Solution of a Timetabling Problem". Journal of the Institute of Mathematics and its Applications 9 (1972) p.23.

5. McNamara, J.F. "Mathematical Programming Models in Educational Planning". Review of Educational Research 41 (1971) p.419.

6. Scarborough, C.W. & Daniel, J.N. "Management Use of Simulation in Long-Range Planning for Colleges and Universities". Presented at TIMS/ORSA meeting. Peat, Marwick, Livingston & Co. (1968).

See also

Appleton, D.R. "An Interactive Computer Model of University Room Usage". Operational Research Quarterly. (In press.)

CHAPTER 4

1. Bernstein, F. "Fortgesetzte Untersuchungen aus der Theorie der Blutgruppen". Zeitschrift für Induktive Abstammungs und Vererbungslehre 56 (1930) p.233.

2. Brues, A. "Stochastic Tests of Selection in the ABO Blood Groups". American Journal of Physical Anthropology 21 (1963) p.287.

3. Cohen, B.H. & Sayre, J.E. "Further Observations on the Relationship of Maternal ABO and Rh Types to Fetal Death". American Journal of Human Genetics 20 (1968) p.310.

4. Crosby, J.L. "Teaching Genetics with an Electronic Computer". Heredity 16 (1961) p.255.

5. Evans, D. Personal Communication (1972).

6. Fisher, R.A. "On the Dominance Ratio". Proceedings of the Royal Society of Edinburgh 42 (1922) p.321.

7. Fraser, A.S. "Simulation of Genetic Systems by Automatic Digital Computer". In "Biometrical Genetics". ed. Kempthorne. Pergamon (1960) p.81.

8. Kimura, M. "Diffusion Models in Population Genetics". Journal of Applied Probability 1 (1964) p.177.

9. Kimura, M. "Stochastic Processes in Population Genetics, with Special Reference to Distribution of Gene Frequencies and Probability of Gene Fixation". In "Mathematical Topics in Population Genetics" ed. Kojima. Springer-Verlag (1970) p.178.

10. Li, C.C. "Population Genetics". University of Chicago Press (1955).

11. Livingstone, F.B. "Natural Selection, Disease, and Ongoing Human Evolution as Illustrated in the ABO Blood Groups". Human Biology 32 (1960) p.19.

12. Martin, F.G. & Cockerham C.C. "High Speed Selection Studies". In "Biometrical Genetics" ed. Kempthorne. Pergamon (1960) p.81.

13. Matsunaga, E. "Selektion durch Unverträglichkeit im ABO-Blutgruppensystem zwischen Mutter und Fetus". Blut II (1956) p.188.

14. Mourant, A.E., Kopec, C. & Domaniewska-Sobczak, K. "The ABO Blood Groups". Blackwell (1958).

15. Nassar, R.F. "Distribution of Gene Frequencies under the Case of Random Genetic Drift with and without Selection". Theoretical and Applied Genetics 39 (1969) p.145.

16. Roberts, J.A.F. "The Frequencies of the ABO Blood Groups in Southwestern England". Annals of Eugenics 14 (1948) p.109.

17. Vielrose, E. "Elements of the Natural Movement of Population". Pergamon Press (1965).

18. Wright, S. "Evolution in Mendelian Populations". Genetics 16 (1931) p.97.

19. Wright, S. "Adaptation and Selection". In "Genetics, Paleontology and Evolution" ed. Jepsen, Simpson and Mayr. Princeton University Press (1949) p.365.

See also

Appleton, D.R. "Interactive Use of Computer Models in Teaching Population Genetics". American Journal of Physical Anthropology. (In press.)

CHAPTER 5

1. Barrett, J.C. "A mathematical model of the mitotic cycle and its application to the interpretation of percentage labelled mitoses data". Journal of the National Cancer Institute 37 (1966) p.443.

2. Bresciani, F. "Cell proliferation in cancer". European Journal of Cancer 4 (1968) p.343.

3. Cairnie, A.B., Lamerton, L.F. & Steel, G.G. "Cell proliferation studies in the intestinal epithelium of the rat : theoretical aspects". Experimental Cell Research 39 (1965) p.539.

4. Cleaver, J.E. "Thymidine Metabolism and Cell Kinetics". North Holland (1967).

5. Gilbert, C.W. "The labelled mitoses curve and the estimation of the parameters of the cell cycle". Cell and Tissue Kinetics 5 (1972) p.53.

6. Howard, A. & Pelc, S.R. "Synthesis of desoxyribonucleic acid in normal and irradiated cells and its relation to chromosome breakage". Heredity 6 suppl. (1953) p.261.

7. Johnson, H.A. "Liver Regeneration and the "Critical Mass" Hypothesis". American Journal of Pathology 57 (1969) p.1.

8. Leroy, F., Galand, P. & Chrétien, J. "The mitogenic action of ovarian hormones on the uterine and vaginal epithelium during the oestrus cycle in the rat : a radioautographic study." Journal of Endocrinology 45 (1969) p.441.

9. McDonald, P.D.M. "Statistical inference from fraction labelled mitoses curve". Biometrika 57 (1970) p.389.

10. Morley, A.R., Wright, N.A. & Appleton, D.R. "Proliferation and differentiation in the castrate mouse seminal vesicle in response to testosterone propionate. I. Experimental observations". Cell and Tissue Kinetics. (In press.)

11. Pedersen, T. & Hartmann, N.R. "The kinetics of granulosa cells in developing follicles in the mouse ovary". Cell and Tissue Kinetics 4 (1971) p.171.

12. Price, D. "Comparative aspects of development and structure in the prostate". In "Biology of the Prostate and Related Tissues". National Cancer Institute Monograph 12 (1963) p.1.

13. Quastler, H. & Sherman, F.G. "Cell population kinetics in the intestinal epithelium of the mouse". Experimental Cell Research 17 (1959) p.420.

14. Steel, G.G. & Hanes, S. "The technique of labelled mitoses : analysis by automatic curve-fitting". Cell and Tissue Kinetics 4 (1971) p.93.

15. Takahashi, M. "Theoretical basis for cell cycle analysis. I. Labelled mitosis wave method". Journal of Theoretical Biology 13 (1966) p.202.

16. Takahashi, M., Hogg, J.D. & Mendelsohn, M.L. "The automatic analysis of FLM curves". Cell and Tissue Kinetics 4 (1971) p.505.

See also

Appleton, D.R., Morley, A.R., & Wright, N.A. "Proliferation and differentiation in the castrate mouse seminal vesicle in response to testosterone propionate. II. Theoretical considerations". Cell and Tissue Kinetics. (In press.)

APPENDIX 1

GLOSSARY OF GENETIC TERMS

ABO : a blood group system controlled by a triallelic locus whose alleles are deonoted A, B and O. A cells agglutinate with type B serum; B cells with type A serum; O cells with neither A nor B. There are therefore 4 phenotypes, namely O, A, B and AB.
A group A individual may have genotype AA or OA, and a group B individual may have genotype BB of OB. Genes A and B are codominant and dominant over O.

AGGLUTINATION : the clumping of cellular components in the presence of a specific immune serum.

ALLELE: one of an array of possible genes at a particular locus.

AUTOSOMAL : pertaining to any chromosome other than the sex chromosome.

BLOOD GROUP : a particular phenotype in a system of classification of blood, based on the occurrence of agglutination of red cells when bloods of incompatible groups are mixed. The ABO, Rhesus and MN are the most widely investigated systems.

CHROMOSOME : a thread of DNA and protein in the nucleus of a cell, so called because it may easily be stained and recognised by its colour.

CODOMINANCE : the situation in which both alleles of a pair are fully expressed in the heterozygote.

DIALLELIC : of a genetic locus having two alleles.

DOMINANT : pertaining to a gene which manifests itself equally in the heterozygote as in the homozygote. The other gene is recessive. Although these terms are well accepted they may be misleading, as dominance is dependent on the test by which the phenotype is determined.

EFFECTIVE POPULATION SIZE : the average number of individuals in a population which contributes genes to the next generation; a number which depends on mating pattern, population age and sex structure, and the history of previous effective population sizes. The concept is necessary because most models assume a constant population undergoing random mating.

FITNESS : the relative ability of an organism to survive and transmit its genes to the next generation; a quantification of this ability based on number of offspring. Fitness usually refers to a particular locus, all others being assumed to act independently.

GAMETE : a reproductive cell having a single set of chromosomes; two of these will combine to form a zygote.

GAMETIC SELECTION: the situation where one allele at a diallelic locus is superior to the other, so that the fittest genotype is the homozygote in this allele and the least fit is the homozygote in the other. (c.f. overdominance.)

GENE : a unit hereditary factor on the chromosome.

GENETIC DRIFT : the effect of random sampling of gametes (only 2 of the 4 parental gametes being transmitted to an offspring) which causes the frequency of a gene in a population to fluctuate. Without selection, mutation or migration genes would become lost in a population because of drift; this would be especially noticeable in small populations.

GENOTYPE : the genetic constitution of an individual, especially at a given locus.

HARDY-WEINBERG EQUILIBRIUM : the situation in which the ratios of the frequencies of the genotypes AA, Aa, aa are $p^2 : 2pq : q^2$.

HETEROZYGOTE : an organism having a pair of dissimilar alleles at a particular locus.

HOMOZYGOTE : an organism having two identical alleles at a particular locus.

LETHAL GENE : a gene which so influences development that the individual is rendered non-viable. Usually the case of interest is that of a lethal recessive, so that no individuals may be homozygous in the gene.

LOCUS : the position of a gene on a chromosome.

MATERNAL-FOETAL INCOMPATIBILITY : the situation in which, for example, an OO mother will have a deficit of OA compared to OO children by an OA father because of sponteneous abortion.

MIGRATION : in the usual constant population size models migration is represented by an exchange of individuals taken at random from the population with individuals from elsewhere whose gene frequency is known.

MUTATION : a rare process whereby an individual possesses an allele which he has not inherited from his parents.

OVERDOMINANCE : the situation in which the heterozygote at a diallelic locus has superior fitness to both homozygotes. (c.f. gametic selection.)

PANMIXIA : random mating.

PHENOTYPE : the detectable characteristics of an organism resulting from its genotype.

POLYMORPHISM : occurrence of different genes in individuals in the same population, at the same locus.

POPULATION GENETICS: the study of inheritance in populations by means of investigations into gene frequencies in these populations.

RECESSIVE : a gene possessed by an individual, whose effect is not expressed in his phenotype. (c.f. dominant.)

SELECTION : (used throughout for natural selection); the process determining the relative share allotted to individuals of different genotypes in the propagation of a population.

SELECTIVE COEFFICIENT: a measure (s) of the disadvantage of a given genotype (in a particular environment).
If on average 1 in 100 of that genotype fails to reproduce then $s = 0.01$.

SEMIDOMINANCE : the possession of an intermediate phenotype by individuals heterozygous in the genes concerned.

SEWALL WRIGHT EFFECT : genetic drift.

STABLE POLYMORPHISM : a polymorphism maintained in a population because the heterozygotes in the alleles under consideration have a higher fitness than either homozygote, with extension to multiallelic loci.

TRIALLELIC : of a genetic locus having three alleles.

ZYGOTE : a cell formed by the union of two gametes; most cells other than reproductive cells are zygotic.

APPENDIX 2

GLOSSARY OF CYTOKINETIC TERMS

ANAPHASE : the third of the four stages of mitosis.

AUTORADIOGRAPHY : a method of determining the position in organs or tissues of specific chemical substances by making them radioactive then recording their distribution on photographic film.

CELL CYCLE : the proliferative cycle of cells, consisting of the pre-synthetic phase, the (DNA–) synthetic phase, the post-synthetic phase and the mitotic phase.

CELL CYCLE PARAMETERS : Collectively t_{G_1}, t_S, t_{G_2}, t_m and T_C. Sometimes also the growth fraction.

CHEMOTHERAPY : in cancer, treatment with drugs of known chemical composition which are assumed to be directly toxic to the tumour cells; the drugs often act at a specific point in the cell cycle.

COAGULATING GLAND : that part of the prostatic complex in the mouse which secretes a substance capable of coagulating the seminal fluid.

CONDITIONAL RENEWAL SYSTEM : a system which normally has very low mitotic activity, but which may be stimulated into growth by trauma or administration of hormone.

CUMULATIVE LABELLING : a type of experiment in which all cells which are in the S-phase at any point during the experiment are labelled, the object being to determine the proportion of cells involved in the cell cycle.

Also CONTINUOUS LABELLING

CYTOKINETIC : pertaining to the movement of cells through the cell cycle.

DAUGHTER CELL : one of the two cells resulting from the division of a single cell.

DECYCLING PROBABILITY : the probability ($\vec{\eta}$) that a cell after undergoing mitosis will leave the cycle (e.g. differentiate). $\vec{\eta}$ is related to the distribution parameter (d) by $\vec{\eta} = \frac{1}{1+d}$.

DIFFERENTIATION : the process whereby descendants of a single cell achieve and maintain specialisations of structure and function.

DISTRIBUTION PARAMETER : the ratio (d) of the proportion ($\overleftarrow{\eta}$) of cells which re-enter G_1 and repeat the cell cycle after a mitosis to the proportion ($\vec{\eta}$) which leave the cycle. $\vec{\eta}$ is the decycling probability.

DNA : deoxyribonucleic acid. A stable nucleic acid component of cells, which consists structurally of two spirals linked transversely and constitutes a pattern or template for replication.

DOUBLING TIME : the time taken for a population of cells to double in number.

EXPONENTIAL GROWTH : a system is in exponential growth if its doubling time is constant.

FLM CURVE : the result of an experiment which measures as a function of time since administration of a label the fraction labelled of all mitoses. This is the most common means of estimating the cell cycle parameters.

G_0 : the symbol given to a phase (experimentally indistinguishable from G_1) in which cells are dormant, neither having differentiated nor continued round the cell cycle after a mitosis.

G_1 : the symbol given to the pre-synthetic phase of the cell cycle.

G_2 : the symbol given to the post-synthetic phase of the cell cycle.

GRAIN COUNT : the number of labels detectable on a cell nucleus in an autoradiographic experiment. Since there is a certain background of radioactivity only grain counts greater than, say, five are taken to indicate that a cell has been labelled. On division the label will be distributed binomially on the daughter cells; this dilution of label can cause difficulty in some types of experiment, but it can also be used to estimate the number of dividing cells.

GROWTH FRACTION : the fraction of cells in a population which are involved in the proliferative cycle.

HEPATECTOMY : removal of the liver.

HORMONE : a substance produced by cells which is necessary for the proper functioning of other cells in organs and tissues to which it is conveyed.

I_L : the labelling index.

I_m : the mitotic index.

INTERPHASE : the period between successive mitoses.

LABEL : a radioactive atom introduced into a molecule which, without altering its structure or function, allows it to be traced autoradiographically.

LABELLING INDEX : the proportion of cells in a population at any given time which have been labelled, indicating that at the time of labelling these cells or their parents were in an S-phase.

M PHASE : that part of the cell cycle in which the cell undergoes mitosis; it includes prophase, metaphase, anaphase and telophase.

MERISTEM : plant tissue which can form new tissue.

METAPHASE : the second of the four stages of mitosis.

MITOSIS : the production of two daughter cells each with the same nuclear content as the parent cell possessed at its birth.

MITOTIC INDEX : the proportion of cells in a population which are in mitosis at any given time.

NEOPLASM : a population of proliferating cells which are not governed by the usual limitations of growth.

OESTROGEN : a steroid hormone which stimulates uterine growth.

ORGAN : a part of an animal which forms a structural and functional unit.

PHASE DURATIONS : collectively t_{G_1}, t_S, t_{G_2} and t_m.

PLM CURVE : percentage labelled mitosis curve; same as FLM curve.

POST-SYNTHETIC PHASE : that part of the cell cycle after the S-phase but before mitosis during which the cell completes its preparations for mitosis.

PRE-SYNTHETIC PHASE : that part of the cell cycle after mitosis but before the S-phase.

PROLIFERATIVE COMPARTMENT : those cells, not necessarily spatially separate or distinguishable from the others, which are taking part in the cell cycle.

PROLIFERATIVE INDICES : I_m and I_L. These both provide estimates of the number of cells in the proliferative compartment if the phase durations are known.

PROPHASE : the first of the four stages of mitosis.

PROSTATE : the complex of hormonally sensitive organs around the neck of the bladder (involved in the production and transmission of seminal fluid).

RNA : ribonucleic acid. A nucleic acid which takes part in protein synthesis.

S PHASE : that part of the cell cycle during which DNA is synthesised.

SEMINAL VESICLE : part of the prostatic complex (see figure 5.2. page 66).

STEADY STATE : a cell population is in steady state when the rate of cell gain (by mitosis or immigration) is equal to the mean rate of cell loss (by differentiation, death or migration).

SYNCHRONY : a state of affairs in which cells in a population are not distributed uniformly round the cell cycle, but travel together. A population of cells which have been pulse-labelled are synchronous, but due to variation in the phase durations of individual cells the synchrony disappears in a diffusion process.

SYNTHESIS : when used without qualification this refers to the synthesis of DNA by a cell.

T_C : the time taken by a cell from its formation through mitosis to complete its own mitosis; the cell cycle time.

t_{G_1} : the duration of the pre-synthetic phase of the cell cycle.

t_{G_2} : the duration of the post-synthetic phase of the cell cycle.

t_m : the duration of mitosis.

t_S : the duration of the S-phase.

TARGET ORGAN (TISSUE) : the receptor organ (tissue) upon which a hormone has its effect.

TELOPHASE : the final part of mitosis when division is completed.

TESTOSTERONE PROPIONATE : testicular hormone $C_{19}H_{28}O_2$; a masculinising steroid hormone secreted by the testis.

THYMIDINE : a chemical used in DNA synthesis.

TISSUE : a population of cells of the same kind, performing similar function.

TRAUMA : injury to an organ or tissue.

TRITIATED THYMIDINE : thymidine made radioactive by the replacement of hydrogen by the radioactive isotope tritium, so that it provides a label for DNA synthesis.

TROPHIC : pertaining to nutrition; applied to hormones which influence the activity and growth of endocrine glands.

ALLOCATION PROGRAM

```
'BEGIN''BOOLEAN' CHAT,OVER,NONE,MANU,SELF;
      'INTEGER' N,I,J,K,N1,N2,C,COUNT,SUM,SMAX,HOLD,KEEP,SOURCE,DESTN,
                FIX,VAR,PER,TIM,ERROR,II,PAR1,PAR2,PAR3,POINT,CANWE;
      'REAL' X;

      'PROCEDURE' PRINT(DEV,FORM,X);
      'VALUE' DEV,FORM,X; 'INTEGER' DEV; 'REAL' FORM,X; 'CODE';

      'PROCEDURE' SPACE(DEV,N);
      'VALUE' DEV,N; 'INTEGER' DEV,N; 'CODE';

      'PROCEDURE' READ(PLEA,COND,CHAT,INT,MESS,FAIL,V1,V2);
      'COMMENT'  ASKS FOR DATA, READS IT IN AND, IF AN ERROR IN ENTERING IT IS
                 DETECTED, REPEATS THE QUERY;
      'BOOLEAN' COND,CHAT,INT; 'STRING' PLEA,MESS; 'LABEL' FAIL;
      'INTEGER' V1; 'REAL' V2; 'CODE';

      'PROCEDURE' REPL(N); 'INTEGER' N; 'CODE';
      'COMMENT'  TESTS THAT PROGRAM IS BEING RUN FROM A TERMINAL;

      'PROCEDURE' JUMP(STR,N1,N2);
      'COMMENT' THIS PROCEDURE MAY BE USED TO LIST THE DATA FILES AND RETURN TO THE
                PROGRAM.  IT HAS NOT BEEN IMPLEMENTED SO THAT THE USER WILL BE
                ENCOURAGED TO HAVE PRINT-OUTS OF THESE FILES AVAILABLE;
      'VALUE' N1,N2; 'STRING' STR; 'INTEGER' N1,N2;      ;

'INTEGER''PROCEDURE' SPOTMAX(A,N);
'VALUE' N; 'INTEGER' N; 'INTEGER''ARRAY' A;
'BEGIN''INTEGER' I,J,K;
      J:=A(/1/); K:=1;
      'FOR' I:=2 'STEP' 1 'UNTIL' N 'DO'
      'IF' A(/I/)>J 'THEN'
      'BEGIN' J:=A(/I/); K:=I;
      'END';
      SPOTMAX:=K;
'END';

'PROCEDURE' PERIOD(K);
'VALUE' K; 'INTEGER' K;
'BEGIN''INTEGER' A;
      A:=K-(K-1)'/'7*7;
      'IF' K<8 'THEN' OUTSTRING(1,'('MONDAY AT ')') 'ELSE'
      'IF' K<15 'THEN' OUTSTRING(1,'('TUESDAY AT ')') 'ELSE'
      'IF' K<22 'THEN' OUTSTRING(1,'('WEDNESDAY AT ')') 'ELSE'
      'IF' K<29 'THEN' OUTSTRING(1,'('THURSDAY AT ')') 'ELSE'
      OUTSTRING(1,'('FRIDAY AT ')');
      PRINT(1,3,'IF' A>4 'THEN' A-3 'ELSE' A+8); SYSACT(1,2,1);
'END';

'PROCEDURE' DAY(TIME,NO);
'INTEGER' TIME; 'LABEL' NO;
'BEGIN''INTEGER' J,SYM;
      ININTEGER(0,TIME);
      'IF' ABS(TIME-7)>5 | ABS(TIME-6.5)<2 'THEN''GOTO' NO;
      'FOR' J:=1,1 'WHILE' SYM=4 'DO' INSYMBOL(0,'('MTW F')',SYM);
      'IF' SYM<1 'THEN''GOTO' NO;
      'IF' SYM=2 'THEN'
      'BEGIN' INSYMBOL(0,'('XUZH')',SYM);
              'IF' ABS(SYM-3)¬=1 'THEN''GOTO' NO;
      'END';
      TIME:=7*(SYM-1)+('IF' TIME>4 'THEN' TIME-8 'ELSE' TIME+3);
'END';

'PROCEDURE' NUMROOMS(A,N,FAIL);
'VALUE' N; 'INTEGER''ARRAY' A; 'INTEGER' N; 'LABEL' FAIL;
'BEGIN''INTEGER' I,C,K,S,T;
      K:=1; SYSACT(0,14,2);
  RET: S:=T:=0;
      INSYMBOL(0,'('123456789F')',C);
      'IF' C<1 'THEN''GOTO' FAIL;
      'IF' C<10 'THEN'
      'BEGIN'T:=C;
            'FOR' I:=1 'WHILE' C<10 'DO'
            'BEGIN'INSYMBOL(0,'('123456789,')',C);
                  T:=10*T+C;
            'END';
            'IF' C¬=10 'THEN''GOTO' FAIL;
            A(/K/):=T'/'10-1; K:=K+1;
      'END''ELSE'
      'BEGIN'ININTEGER(0,T); ININTEGER(0,S);
            'IF' T>PAR2|S>PAR2|T>=S|S-T>N-K 'THEN''GOTO' FAIL;
            'FOR' I:=K 'STEP' 1 'UNTIL' S-T+K 'DO'
                  A(/I/):=T+I-K;
            K:=S-T+K+1;
      'END';
      'IF' K<N+1 'THEN''GOTO' RET;
'END';
```

```
        SYSACT(1,12,1); SYSACT(2,12,1); SYSACT(3,12,1);
        SYSACT(4,12,1); SYSACT(5,12,1); SYSACT(6,12,1);
        ININTEGER(2,PAR1); ININTEGER(3,PAR2); ININTEGER(4,N); ININTEGER(5,PAR3);
        'IF' N¬=PAR2 'THEN''GOTO' XXXX;
        SELF:='FALSE'; CHAT:='FALSE'; REPL(N); 'IF' N=0 'THEN' CHAT:='TRUE';

START:  READ(''('HOW MANY ROOMS ARE TO BE CONSIDERED?')',1<=N&N<=PAR2,CHAT,
                'TRUE',''('ILLEGAL NO. OF ROOMS')',CRUNCH,N,X);

        'BEGIN''INTEGER''ARRAY' ROOM,RCAP,CLASS,FULL,EMPTY(/1:N/),TABLE1,TABLE2(/1:35,1:N/),
                                CSIZE(/1:PAR1/),MASTR1,MASTR2(/1:PAR3/);

        'PROCEDURE' DISPLAY(J);
        'VALUE' J; 'INTEGER' J;
        'BEGIN'OUTSTRING(1,''('ROOM NO. ')'); PRINT(1,3,ROOM(/J/));
               OUTSTRING(1,''('   CAPACITY ')'); PRINT(1,3,RCAP(/J/));
               PRINT(1,6,CLASS(/J/)); OUTSTRING(1,''('  STUDENTS')');
               SYSACT(1,2,1);
        'END';

        'PROCEDURE' CONVERT(RM,NO);
        'INTEGER' RM; 'LABEL' NO;
        'BEGIN''INTEGER' I;
               'FOR' I:=1 'STEP' 1 'UNTIL' N 'DO'
                     'IF' ROOM(/I/)=RM 'THEN''GOTO' OK;
               'GOTO' NO;
          OK:  RM:=I;
        'END';

        'PROCEDURE' SUMEM(I,J,ANS);
        'VALUE' I,J; 'INTEGER' I,J,ANS;
        'BEGIN''INTEGER' K,REF1,REF2,REF3;
               REF1:=TABLE1(/I,J/); REF2:=TABLE2(/I,J/); ANS:=0;
               'IF' REF1¬=-1 'THEN'
               'BEGIN'
               RECUR: 'FOR' K:=1 'STEP' 1 'UNTIL' REF2 'DO'
                            ANS:=ANS+MASTR2(/K-1+REF1/)
                                 *CSIZE(/MASTR1(/K-1+REF1/)/);
                      'IF' MASTR1(/REF1+REF2/)¬=-1 'THEN'
                      'BEGIN' REF3:=REF1;
                              REF1:=MASTR1(/REF3+REF2/);
                              REF2:=MASTR2(/REF3+REF2/);
                              'GOTO' RECUR;
                      'END';
                      ANS:=0.01*ANS;
               'END';
        'END';

        'PROCEDURE' GETDATA;
        'BEGIN''INTEGER' I,J,K;
               'FOR' I:=1 'STEP' 1 'UNTIL' PAR1 'DO'
               'BEGIN' SYSACT(2,2,16);
                       ININTEGER(2,CSIZE(/I/));
               'END';
               'FOR' I:=1 'STEP' 1 'UNTIL' PAR3 'DO'
               'BEGIN'ININTEGER(5,MASTR1(/I/));
                      ININTEGER(5,MASTR2(/I/));
               'END';
               K:=0;
               'FOR' I:=1 'STEP' 1 'UNTIL' N 'DO'
               'BEGIN''IF' ROOM(/I/)¬=K+1 'THEN'
                       'BEGIN'SYSACT(3,14,ROOM(/I/)-K);
                              SYSACT(4,14,5*(RCOM(/I/)-K-1));
                       'END';
                       SYSACT(3,2,7); ININTEGER(3,RCAP(/I/));
                       'FOR' J:=1 'STEP' 1 'UNTIL' 35 'DO'
                       'BEGIN' ININTEGER(4,TABLE1(/J,I/));
                               ININTEGER(4,TABLE2(/J,I/));
                       'END';
                       K:=RCOM(/I/);
               'END';
        'IF' ROOM(/N/)¬=PAR2 'THEN' SYSACT(4,14,5*(PAR2-ROOM(/N/)));
        'END';
```

```
      'PROCEDURE' PUTC1(SOURCE,DESTN,PER);
      'INTEGER' SOURCE,DESTN,PER;
      'BEGIN''INTEGER' HOLD,KEEP;
            TABLE1(/PER,DESTN/):=TABLE1(/PER,SOURCE/);
            TABLE2(/PER,DESTN/):=TABLE2(/PER,SOURCE/);
            TABLE1(/PER,SOURCE/):=TABLE2(/PER,SOURCE/):=-1;
            HOLD:=CLASS(/SOURCE/); SUMEM(PER,SOURCE,CLASS(/SOURCE/));
            KEEP:=CLASS(/DESTN/); SUMEM(PER,DESTN,CLASS(/DESTN/));
            DISPLAY(SOURCE); DISPLAY(DESTN); SYSACT(1,2,1);
            'IF' PER¬=I 'THEN'
            'BEGIN' CLASS(/SOURCE/):=HOLD;
                  CLASS(/DESTN/):=KEEP;
            'END';
      'END';

      'PROCEDURE' EXCH1(SOURCE,DESTN);
      'INTEGER' SOURCE,DESTN;
      'BEGIN''INTEGER' HOLD,KEEP;
            HOLD:=TABLE1(/I,SOURCE/); KEEP:=TABLE2(/I,SOURCE/);
            TABLE1(/I,SOURCE/):=TABLE1(/I,DESTN/); TABLE2(/I,SOURCE/):=TABLE2(/I,DESTN/);
            TABLE1(/I,DESTN/):=HOLD; TABLE2(/I,DESTN/):=KEEP;
            HOLD:=CLASS(/SOURCE/); CLASS(/SOURCE/):=CLASS(/DESTN/); CLASS(/DESTN/):=HOLD;
            DISPLAY(SOURCE); DISPLAY(DESTN); SYSACT(1,2,1);
      'END';

      'IF' CHAT 'THEN' OUTSTRING(1,'('ENTER THEIR NUMBERS')'); SYSACT(1,2,1);
      ERROR:=0; NUMROOMS(ROOM,N,FAIL); ERROR:=0;
TIME: 'IF' CHAT 'THEN' OUTSTRING(1,'('AT WHAT TIME DO YOU WISH TO START?')');
      SYSACT(1,2,1); DAY(TIM,OOPS);
      'IF' CHAT 'THEN'
      'BEGIN' OUTSTRING(1,'('DO YOU WISH TO MAKE ALL ROOM CHANGES BY HAND?')');
            SYSACT(1,2,1); SYSACT(0,14,2); INSYMBOL(0,'('Y')',C);
            'IF' C=1 'THEN' SELF:='TRUE';
      'END';
      'GOTO' MORE;

FAIL: OUTSTRING(1,'('ERROR IN ENTERING ROOMS')'); SYSACT(1,2,1); SYSACT(0,2,1);
      ERROR:=ERROR+1; 'IF' CHAT&ERROR<2 'THEN''GOTO' START; 'GOTO' CRUNCH;
OOPS: OUTSTRING(1,'('ERROR IN ENTERING TIME')'); SYSACT(1,2,1);
      ERROR:=ERROR+1; 'IF' CHAT&ERROR<2 'THEN''GOTO' TIME; 'GOTO' CRUNCH;

MORE: GETDATA;
      'FOR' I:=TIM 'STEP' 1 'UNTIL' 35 'DO'

      'BEGIN' OVER:='FALSE'; MANU:='FALSE'; SMAX:=0;
            'FOR' J:=1 'STEP' 1 'UNTIL' N 'DO'
            'BEGIN' SUMEM(I,J,SUM);
                  CLASS(/J/):=SUM; FULL(/J/):=EMPTY(/J/):=-1;
                  'IF' SUM>RCAP(/J/) 'THEN'
                  'BEGIN''IF' ¬OVER 'THEN' PERIOD(I);
                        SPACE(1,4); DISPLAY(J); OVER:='TRUE';
                        FULL(/J/):=SUM;
                        'IF' SMAX<SUM 'THEN' SMAX:=SUM;
                  'END';
            'END';

            'IF' OVER & ¬SELF 'THEN'
            'BEGIN' SYSACT(1,2,1);
                  'FOR' J:=1 'STEP' 1 'UNTIL' N 'DO'
                  'IF' TABLE1(/I,J/)=-1 'THEN' EMPTY(/J/):=RCAP(/J/);
            SECND: POINT:=SPOTMAX(FULL,N);
                  CANWE:=SPOTMAX(EMPTY,N); 'IF' CANWE=1 & EMPTY(/1/)=-1 'THEN''GOTO' SECNE;
                  'IF' CLASS(/POINT/)<=RCAP(/CANWE/) 'THEN'
                  'BEGIN''IF' FULL(/POINT/)<0 'THEN''GOTO' CONT;
                        PUTC1(POINT,CANWE,I);
                        FULL(/POINT/):=EMPTY(/CANWE/):=-1;
                        'GOTO' SECND;
                  'END';
                  'IF' RCAP(/SPOTMAX(RCAP,N)/)<CLASS(/POINT/) 'THEN''GOTO' TOUGH;
            SECNE: 'FOR' J:=1 'STEP' 1 'UNTIL' N 'DO'
                  'IF' CLASS(/POINT/)<=RCAP(/J/) & CLASS(/J/)<=RCAP(/POINT/) 'THEN'
                  'BEGIN''IF' FULL(/POINT/)<0 'THEN''GOTO' CONT;
                        EXCH1(POINT,J); FULL(/POINT/):=-1;
                        'GOTO' SECND;
                  'END';
            TOUGH: MANU:='TRUE';
            'END';
```

```
        'IF'  OVER & MANU | OVER & SELF  'THEN'
        'BEGIN' SYSACT(1,2,1); OUTSTRING(1,'('THE FOLLOWING ROOMS ARE')');
                OUTSTRING(1,'(' EMPTY AND WILL TAKE THE LARGEST OF ')');
                OUTSTRING(1,'('THESE CLASSES')'); SYSACT(1,2,1);
                NONE:='TRUE';
                'FOR' J:=1 'STEP' 1 'UNTIL' N 'DO'
                      'IF' TABLE1(/I,J/)=-1&RCAP(/J/)>SMAX 'THEN'
                'BEGIN' DISPLAY(J); NONE:='FALSE';
                'END';
                'IF' NONE 'THEN'
                'BEGIN' OUTSTRING(1,'('NONE')'); SYSACT(1,2,1);
                        OUTSTRING(1,'('THE FOLLOWING ROOMS ARE LARGE ')');
                        OUTSTRING(1,'('ENOUGH')'); SYSACT(1,2,1);
                        'FOR' J:=1 'STEP' 1 'UNTIL' N 'DO'
                              'IF' RCAP(/J/)>SMAX 'THEN'
                        'BEGIN' DISPLAY(J); NONE:='FALSE';
                        'END';
                        'IF' NONE 'THEN'
                        'BEGIN' OUTSTRING(1,'('NONE')'); SYSACT(1,2,1);
                        'END'; SYSACT(1,2,1);
                'END';

                'IF' CHAT 'THEN'
                'BEGIN''SWITCH' PART:=CONT,STOP,FOCS,DOES,WHEN,HOWM,EXCH,
                                      PUTC,LIST,REPT,MOVE;
                        ERROR:=FIX:=0;
                  ACT:  OUTSTRING(1,'('WHAT ACTION IS REQUIRED?')');
                 ACT1:  SYSACT(1,2,1); SYSACT(0,14,2);
                        INSYMBOL(0,'('CSFDWHEPLRM')',C);
                        'IF' C=0 'THEN'
                        'BEGIN' ERROR:=ERROR+1;
                                'IF' ERROR<2 'THEN'
                                'BEGIN' OUTSTRING(1,'('RE-ENTER MESSAGE')');
                                        'GOTO' ACT1;
                                'END''ELSE'
                                'BEGIN' OUTSTRING(1,'('ILLEGAL MESSAGE')');
                                        SYSACT(1,2,1); 'GOTO' STOP;
                                'END';
                        'END';
                        ERROR:=0; 'GOTO' PART(/C/);

FOCS:   ININTEGER(0,FIX); CONVERT(FIX,WRONG); OUTSTRING(1,'('DONE. NOW ')');
        'GOTO' ACT;

DOES:   'IF' FIX=0 'THEN'
        'BEGIN' OUTSTRING(1,'('YOU HAVE NOT SAID WHICH ROOM WE ARE ')');
                OUTSTRING(1,'('DEALING WITH')'); SYSACT(1,2,1);
        'END''ELSE'
        'BEGIN' COUNT:=0; HOLD:=TABLE1(/I,FIX/);
                'FOR' K:=I+1 'STEP' 1 'UNTIL' 35 'DO'
                      'IF' TABLE1(/K,FIX/)=HOLD 'THEN' COUNT:=COUNT+1;
                'IF' COUNT>0 'THEN'
                'BEGIN' OUTSTRING(1,'('THE SITUATION OCCURS ANOTHER')');
                        PRINT(1,3,COUNT); OUTSTRING(1,'(' TIMES')');
                'END''ELSE'
                OUTSTRING(1,'('THE SITUATION DOES NOT OCCUR AGAIN')');
                SYSACT(1,14,2);
        'END';
        'GOTO' ACT;

WHEN:   'IF' FIX=0 'THEN'
        'BEGIN' OUTSTRING(1,'('YOU HAVE NOT SAID WHICH ROOM WE ARE ')');
                OUTSTRING(1,'('DEALING WITH')'); SYSACT(1,2,1);
        'END''ELSE'
        'BEGIN' HOLD:=TABLE1(/I,FIX/);
                'FOR' K:=I+1 'STEP' 1 'UNTIL' 35 'DO'
                      'IF' TABLE1(/K,FIX/)=HOLD 'THEN' PERIOD(K);
                OUTSTRING(1,'('DONE')'); SYSACT(1,14,2);
        'END';
        'GOTO' ACT;

HOWM:   ININTEGER(0,VAR); CONVERT(VAR,WRONG);
        DAY(PER,BOOB); SUMEM(PER,VAR,SUM);
        PRINT(1,3,SUM); SYSACT(1,2,1);
        'GOTO' ACT;
```

```
      PUTC: ININTEGER(0,SOURCE); CONVERT(SOURCE,WRONG);
            DAY(PER,BOOB);
            ININTEGER(0,DESTN); CONVERT(DESTN,WRONG);
            'IF' TABLE1(/PER,DESTN/)¬=-1 'THEN''GOTO' CANT;
            PUTC1(SOURCE,DESTN,PER);
            'GOTO' ACT;

      EXCH: ININTEGER(0,SOURCE); CONVERT(SOURCE,WRONG);
            ININTEGER(0,DESTN); CONVERT(DESTN,WRONG);
            EXCH1(SOURCE,DESTN);
            'FOR' J:=1 'STEP' 1 'UNTIL' N 'DO' FULL(/J/):=EMPTY(/J/):=-1;
            'GOTO' ACT;

      LIST: 'FOR' K:=1 'STEP' 1 'UNTIL' 5 'DO' INSYMBOL(0,'('SRC')',C);
            'IF' C<1 'THEN''GOTO' BOOB;
            'IF' C=1 'THEN'
            'BEGIN' OUTSTRING(1,'('STATUS OF ROOMS ON ')'); PERIOD(I);
                  'FOR' K:=1 'STEP' 1 'UNTIL' N 'DO' DISPLAY(K);
            'END''ELSE'
            'BEGIN' ININTEGER(0,N1); ININTEGER(0,N2);
                  'IF' N1>N2 'THEN''GOTO' BOOB;
                  'IF' C=2 'THEN' JUMP('('ROOM')',N1,N2)
                     'ELSE' JUMP('('COURSE')',N1,N2);
            'END';
            'GOTO' ACT;

      MOVE: II:=I; DAY(I,BANG); OUTSTRING(1,'('DONE')');
            SYSACT(1,2,1); I:=I-1; 'GOTO' CONT;

      REPT: I:=I-1; 'GOTO' CONT;

      CANT: OUTSTRING(1,'('ROOM NO. ')'); PRINT(1,3,ROOM(/DESTN/));
            OUTSTRING(1,'(' IS NOT EMPTY')'); SYSACT(1,2,1); 'GOTO' ACT;
     WRONG: OUTSTRING(1,'('WE ARE NOT INTERESTED IN THAT ROOM')');
            SYSACT(1,2,1); 'GOTO' ACT;
      BOOB: OUTSTRING(1,'('ILLEGAL INSTRUCTION')'); SYSACT(1,2,1);
            'GOTO' ACT;
       MIS: OUTSTRING(1,'('ERROR IN QUERY')'); SYSACT(1,2,1);
            'GOTO' ACT;
      BANG: OUTSTRING(1,'('NO SUCH TIME.  THE PROGRAM WILL CONTINUE')');
            SYSACT(1,2,1); I:=II; 'GOTO' CONT;

                  'END';
             'END';

CONT: 'IF' CHAT & I=35 'THEN'
      'BEGIN' SYSACT(1,14,2); OUTSTRING(1,'('WE HAVE COMPLETED THE WEEK')');
              SYSACT(1,2,1); OUTSTRING(1,'('DO YOU WISH TO LOOK AT AN EARLIER TIME?')');
              SYSACT(1,2,1); SYSACT(0,14,2); INSYMBOL(0,'('Y')',C);
              'IF' C=0 'THEN''GOTO' MORE;
              OUTSTRING(1,'('WHICH TIME?')'); SYSACT(1,2,1);
              DAY(I,BANG1); I:=I-1; 'GOTO' CONT;
       MORE: SYSACT(1,14,2); OUTSTRING(1,'('DO YOU WISH TO MAKE ANY ')');
             OUTSTRING(1,'('FURTHER CHANGES?')'); SYSACT(1,2,1); SYSACT(0,14,2);
             INSYMBOL(0,'('Y')',C); 'IF' C=0 'THEN''GOTO' STOP;
             OUTSTRING(1,'('YOU MAY MOVE ANY CLASS AT ANY PERIOD')'); SYSACT(1,2,1);
             OUTSTRING(1,'('STOP BY REFERRING TO CLASS IN ROOM 0')'); SYSACT(1,2,1);
      ALTER: ININTEGER(0,SOURCE); CONVERT(SOURCE,CONT);
             DAY(PER,CONT); ININTEGER(0,DESTN); CONVERT(DESTN,CONT);
             'IF' TABLE1(/PER,DESTN/)¬=-1 'THEN''GOTO' CONT;
             PUTC1(SOURCE,DESTN,PER); 'GOTO' ALTER;
      BANG1: OUTSTRING(1,'('NO SUCH TIME')'); SYSACT(1,2,1);
             'GOTO' STOP;
      'END';

      'END';
```

```
 STOP: 'IF' CHAT 'THEN'
       'BEGIN' SYSACT(1,14,2); OUTSTRING(1,'('DO YOU WISH TO SAVE THE NEW TIMETABLES?')');
               SYSACT(1,2,1); SYSACT(0,14,2); INSYMBOL(0,'('N')',C);
       'END';
       'IF' C=1 'THEN''GOTO' CRUNCH;
       SPACE(6,5); PRINT(6,3,PAR2); SYSACT(6,2,1);
       'FOR' C:=1 'STEP' 1 'UNTIL' PAR2 'DO'
       'BEGIN' J:=C; CONVERT(J,NOTIN);
               'FOR' I:=1 'STEP' 1 'UNTIL' 5 'DO'
               'BEGIN''FOR' K:=1 'STEP' 1 'UNTIL' 7 'DO'
                       'BEGIN' PRINT(6,-6,TABLE1(/7*(I-1)+K,J/));
                               PRINT(6,-3,TABLE2(/7*(I-1)+K,J/));
                       'END';
                       SYSACT(6,2,1);
               'END';
               SYSACT(4,14,5);
               'GOTO' NEXT;
       NOTIN: 'FOR' I:=1 'STEP' 1 'UNTIL' 5 'DO'
               'BEGIN''FOR' K:=1 'STEP' 1 'UNTIL' 7 'DO'
                       'BEGIN' ININTEGER(4,N1); ININTEGER(4,N2);
                               PRINT(6,-6,N1); PRINT(6,-3,N2);
                       'END';
                       SYSACT(6,2,1);
               'END';
 NEXT: 'END';

       'END'; 'GOTO' CRUNCH;
  XXXX: OUTSTRING(1,'('ERROR IN FILE PARAMETERS')'); SYSACT(1,2,1);
CRUNCH:SYSACT(0,12,0); SYSACT(1,12,0); SYSACT(2,12,0); SYSACT(3,12,0);
       SYSACT(4,12,0); SYSACT(5,12,0); SYSACT(6,12,0);
'END'
```

THE DIALLELIC LOCUS UNDER DRIFT

```
BEGIN INTEGER I,N,T,NN; REAL P,X,H,A,M,S,TT; STRING STR;
      LOGICAL BAT;

      LONG REAL PROCEDURE F(INTEGER VALUE A,B,C; REAL VALUE Z1,Z2);
      BEGIN COMMENT  PRODUCT OF TWO HYPERGEOMETRIC FUNCTIONS;
            INTEGER S,U; LONG REAL T,S1,S2,ZZ1,ZZ2;
            T:=S1:=S2:=ZZ1:=ZZ2:=1L; U:=-B-1;
            FOR S:=0 UNTIL U DO
            BEGIN T:=T*(A+S)*(B+S)/((C+S)*(S+1));
                  ZZ1:=ZZ1*Z1; ZZ2:=ZZ2*Z2;
                  S1:=S1+T*ZZ1; S2:=S2+T*ZZ2
            END;
            S1*S2
      END;

      REAL PROCEDURE DENSITY(REAL VALUE X,P,T);
      BEGIN COMMENT  TO CALCULATE THE GENE FREQUENCY DISTRIBUTION;
            INTEGER I; LONG REAL S,R; LONG REAL ARRAY Q(0::3);
            S:=0; R:=Q(1):=Q(2):=Q(3):=Q(0):=1;
            FOR I:=1 UNTIL 1000 DO
            BEGIN IF (ABS (R) <1'-6L) AND (ABS (Q(1)) <1'-6L)
                  AND (ABS (Q(2)) <1'-6L) AND (ABS (Q(3)) <1'-6L)
                  AND (ABS (Q(0)) <1'-6L) THEN GOTO EXIT;
                  Q(I REM 4):=R;
                  R:=I*(I+1)*(I+I+1)*F(I+2,1-I,2,X,P)
                      *LONGEXP(-0.25*I*(I+1)*T);
                  S:=S+R
            END;
      EXIT: IF S<0.01 THEN S:=0;
            S*P*(1-P)
      END;

      COMMENT  THE NEXT SEVEN PROCEDURES ARE LIBRARY PROCEDURES
               FOR THE GRAPH PLOTTER;
      PROCEDURE SCALE(REAL VALUE XS,YS,XO,YO);
      FORTRAN "SCALEM";
      PROCEDURE AXIS(INTEGER VALUE I; REAL VALUE X,Y,DU;
                     INTEGER VALUE N);
      FORTRAN "EGRID";
      PROCEDURE MOVE(INTEGER VALUE I; REAL VALUE X,Y);
      FORTRAN "EPLOT";
      PROCEDURE PLOTTEXT(STRING(256) VALUE ST; INTEGER VALUE NC;
                         REAL VALUE CS,ANG);
      FORTRAN "PLTEXT";
      PROCEDURE ENDPIC;
      FORTRAN "ENDPIC";
      PROCEDURE THICK(INTEGER VALUE I);
      FORTRAN "SETLIN";
      PROCEDURE DONOW(INTEGER VALUE I);
      FORTRAN "DONOW";
```

```
      PROCEDURE NUM(INTEGER VALUE Z; STRING(3) RESULT STR);
      BEGIN COMMENT  TO PRINT ANY POSITIVE INTEGER UP TO 999;
            INTEGER J,K; J:=0;
            FOR K:=100,10,1 DO IF (K=1) OR ((Z DIV K)>0) THEN
            BEGIN STR(J|1):=CODE(240+(Z DIV K));
                  Z:=Z REM K; J:=J+1
            END ELSE BEGIN STR(J|1):=" "; J:=J+1 END;
            IF (STR(0|1)¬=" ") AND (STR(1|1)=" ") THEN STR(1|1):="0"
      END;


      R_FORMAT:="A"; R_D:=2; R_W:=4;


      COMMENT  ENTER DATA FROM 2741;
      WRITE("RANDOM GENETIC DRIFT AT A DIALLELIC LOCUS"); WRITE(" ");
      WRITE("ONLINE OR BATCH PLOTTING?"); ICCONTROL(2);
      READ(STR(0|1)); BAT:=STR(0|1)="B"; WRITE(" ");
NEW:  WRITE("ENTER POPULATION SIZE, NUMBER OF GENERATIONS ELAPSED,");
      WRITE("AND INITIAL GENE FREQUENCY"); ICCONTROL(2);
      READ(N); IF N<=0 THEN GOTO STOP; READON(T,P);
      TT:=T/N; NN:=IF N<50 THEN 2*N ELSE 100;
      NN:=IF (TT<0.1) OR ((P-0.5)>0.4) THEN 0.5*NN; X:=1/NN;


      BEGIN REAL ARRAY V(1::NN-1);
            A:=M:=S:=0;
            FOR I:=1 UNTIL NN-1 DO
            BEGIN H:=V(I):=DENSITY(I*X,P,TT);
                  A:=A+H; M:=M+H*I; S:=S+H*I*I;
            END;
            M:=M*X/A; S:=SQRT(S*X*X/A-M*M); A:=1-A/(NN-1);
            IF A<0 THEN A:=0;
            WRITE("MEAN OF DISTRIBUTION    = ",M);
            WRITE("STANDARD DEVIATION      = ",S);
            WRITE("PROBABILITY OF FIXATION= ",A); ICCONTROL(2);

            IF ¬BAT THEN
            BEGIN WRITE("COMPLTEK OR PLOTTER?"); ICCONTROL(2);
                  READ(STR(0|1)); DONOW(IF STR(0|1)="O" THEN 1 ELSE 0)
            END;
```

```
            COMMENT  OUTPUT OF GRAPH WITH DESCRIPTION;
            SCALE(20,4,-0.1,-0.50);
            AXIS(0,0,0,0.05,20); AXIS(1,0,0,0.5,10);
            MOVE(1,X,V(1)); THICK(1);
            FOR I:=2 UNTIL NN-1 DO MOVE(2,I*X,V(I)); THICK(0);
            MOVE(1,0.5,3.5); MOVE(2,0.5,5); MOVE(0,1,5);
            MOVE(0,1,3.5); MOVE(0,0.5,3.5);
            STR:="POPULATION =     ";
            NUM(N,STR(13|3));
            MOVE(1,0.51,4.725); PLOTTEXT(STR,16,2,0);
            STR:="INITIAL P =      ";
            NUM(ROUND(100*P),STR(13|3)); STR(12|2):="0.";
            MOVE(1,0.51,4.35); PLOTTEXT(STR,16,2,0);
            STR:="GENERATIONS=     ";
            NUM(T,STR(13|3));
            MOVE(1,0.51,3.975); PLOTTEXT(STR,16,2,0);
            STR:="    % POPS. FIXED";
            NUM(ROUND(100*A),STR(0|3));
            MOVE(1,0.51,3.6); PLOTTEXT(STR,16,2,0);
            MOVE(1,-0.0333,-0.21);
            PLOTTEXT("0.0      0.2      0.4      0.6      0.8      1.0",
                     43,1.6667,0);
            FOR I:=0 UNTIL 5 DO
            BEGIN MOVE(1,-0.025,I-0.075); PLOTTEXT(CODE(240+I),
                        1,1.6667,0)
            END;
            MOVE(1,0.23,-0.46);
            PLOTTEXT("GENE FREQUENCY (P)",18,2,0);
            MOVE(1,0,5.6);
            PLOTTEXT("GENE  FREQUENCY  DISTRIBUTION  FOR",34,2,0);
            MOVE(1,0,5.35);
            PLOTTEXT("UNFIXED POPULATIONS AT A DIALLELIC",34,2,0);
            MOVE(1,0,5.1);
            PLOTTEXT("LOCUS  UNDER  RANDOM GENETIC DRIFT",34,2,0);
            ENDPIC;
            WRITE(" "); GOTO NEW;
        END;
STOP: ;
END.
```

THE DIALLELIC LOCUS UNDER SELECTION AND MIGRATION

```
BEGIN INTEGER I,NPOP,NN,MAX; REAL M,XI,ETA,XI1,ETA1,X,Y,XX,S1,S2,S3;
      LONG REAL H,A,N,S,MX,W; STRING(3) ST3; STRING(10) ST10;
      LOGICAL BAT;


      COMMENT  THE NEXT SEVEN PROCEDURES ARE LIBRARY PROCEDURES
               FOR THE GRAPH PLOTTER;
      PROCEDURE SCALE(REAL VALUE XS,YS,XO,YO);
      FORTRAN "SCALEM";
      PROCEDURE AXIS(INTEGER VALUE I; REAL VALUE X,Y,DU;
                     INTEGER VALUE N);
      FORTRAN "EGRID";
      PROCEDURE MOVE(INTEGER VALUE I; REAL VALUE X,Y);
      FORTRAN "EPLOT";
      PROCEDURE PLOTTEXT(STRING(256) VALUE ST; INTEGER VALUE NC;
                         REAL VALUE CS,ANG);
      FORTRAN "PLTEXT";
      PROCEDURE ENDPIC;
      FORTRAN "ENDPIC";
      PROCEDURE THICK(INTEGER VALUE I);
      FORTRAN "SETLIN";
      PROCEDURE DONOW(INTEGER VALUE I);
      FORTRAN "DONOW";

      PROCEDURE NUM(INTEGER VALUE Z; STRING(3) RESULT STR);
      BEGIN COMMENT  TO PRINT ANY POSITIVE INTEGER UP TO 999;
            INTEGER J,K; J:=0;
            FOR K:=100,10,1 DO IF (K=1) OR ((Z DIV K)>0) THEN
            BEGIN STR(J|1):=CODE(240+(Z DIV K));
                  Z:=Z REM K; J:=J+1
            END ELSE BEGIN STR(J|1):=" "; J:=J+1 END;
            IF (STR(0|1)¬=" ") AND (STR(1|1)=" ") THEN STR(1|1):="0"
      END;

      PROCEDURE FRACNO(REAL VALUE X; STRING(4) RESULT STR);
      BEGIN COMMENT  TO PRINT A FRACTION 0.DD;
            INTEGER Z; X:=X+0.0001; STR(1|1):=".";
            Z:=TRUNCATE(X); STR(0|1):=CODE(240+Z);
            X:=10*(X-Z); Z:=TRUNCATE(X); STR(2|1):=CODE(240+Z);
            STR(3|1):=CODE(240+ROUND(10*(X-Z)))
      END;

      R_FORMAT:="A"; R_D:=2; R_W:=4;
```

```
      COMMENT  ENTER DATA FROM 2741;
      WRITE("EQUILIBRIUM GENE FREQUENCY DISTRIBUTION FOR A");
      WRITE("DIALLELIC LOCUS UNDER SELECTION AND MIGRATION");
      WRITE(" "); WRITE("ONLINE OR BATCH PLOTTING?"); IOCONTROL(2);
      READ(ST3(0|1)); BAT:=ST3(0|1)="B"; WRITE(" ");
NEW:  WRITE("ENTER POPULATION SIZE, IMMIGRATION RATE");
      WRITE("AND GENE FREQUENCY OF IMMIGRANTS IF ANY"); IOCONTROL(2);
      READ(NPOP); IF NPOP=0 THEN GOTO STOP;
      READON(M); NN:=IF NPOP<50 THEN 2*NPOP ELSE 100;
      IF M¬=0 THEN READON(XI) ELSE XI:=0;
      XX:=1/NN; ETA:=1-XI;
      IF M¬=0 THEN
      BEGIN XI1:=1-4*M*NPOP*XI; ETA1:=1-4*M*NPOP*ETA END;
      WRITE("ENTER FITNESSES OF AA, AB AND BB"); IOCONTROL(2);
      READ(S1,S2,S3); A:=N:=S:=MX:=0;

      BEGIN LONG REAL ARRAY V(1::NN-1);
            FOR I:=1 UNTIL NN-1 DO
            BEGIN X:=I*XX; Y:=1-X;
                  W:=S1*X*X+2*S2*X*Y+S3*Y*Y;
                  H:=V(I):=IF M=0 THEN LONGEXP(2*NPOP*LONGLN(W))/(X*Y)
            ELSE LONGEXP(2*NPOP*LONGLN(W)-XI1*LONGLN(X)-ETA1*LONGLN(Y));
                  A:=A+H; N:=N+H*X; S:=S+H*X*X;
                  IF H>MX THEN MX:=H
            END;
            N:=N/A; S:=SQRT((S-A*N*N)/A);
            A:=(NN-1)/A; MAX:=5*(1+(TRUNCATE(1+MX*A)) DIV 5);
            FOR I:=1 UNTIL NN-1 DO V(I):=A*V(I);
            WRITE("MEAN OF DISTRIBUTION = ",N);
            WRITE("STANDARD DEVIATION   = ",S); IOCONTROL(2);

            IF ¬BAT THEN
            BEGIN WRITE("COMPUTER OR PLOTTER?"); IOCONTROL(2);
                  READ(ST3(0|1)); DONOW(IF ST3(0|1)="C" THEN 1 ELSE 0)
            END;

            COMMENT  OUTPUT OF GRAPH WITH DESCRIPTION;
            SCALE(20,20/MAX,-0.1,-0.1*MAX);
            AXIS(0,0,0,0.05,20); AXIS(1,0,0,1,MAX);
            MOVE(1,XX,V(1)); THICK(1);
            FOR I:=2 UNTIL NN-1 DO MOVE(2,I*XX,V(I)); THICK(0);
            MOVE(1,-0.0333,-0.042*MAX);
            PLOTTEXT("0.0     0.2     0.4     0.6     0.8     1.0",
                 43,1.6667,0);
            FOR I:=0 UNTIL 5 DO
            BEGIN MOVE(1,-0.07,0.2*MAX*(I-0.075));
                  NUM(ROUND(0.2*MAX*I),ST3);
                  PLOTTEXT(ST3,3,1.667,0)
            END;
      END;
```

```
      MOVE(1,0.23,-0.092*MAX);
      PLOTTEXT("GENE FREQUENCY (P)",18,2,0);
      MOVE(1,0,1.12*MAX);
      PLOTTEXT(" GENE FREQUENCY DISTRIBUTION  FOR",33,2,0);
      MOVE(1,0,1.07*MAX);
      PLOTTEXT(" A DIALLELIC LOCUS IN EQUILIBRIUM",33,2,0);
      MOVE(1,0,1.02*MAX);
      PLOTTEXT(" UNDER  SELECTION  AND  MIGRATION",33,2,0);
      MOVE(1,0.77,0.935*MAX); ST10:="  N =      ";
      NUM(NPOP,ST10(7|3)); PLOTTEXT(ST10,10,1.5,0);
      IF M=0 THEN GOTO NOM;
      MOVE(1,0.77,0.896*MAX); ST10:="N.M =      ";
      FRACNO(NPOP*M,ST10(6|4)); PLOTTEXT(ST10,10,1.5,0);
      MOVE(1,0.77,0.857*MAX); ST10:="  X =      ";
      FRACNO(XI,ST10(6|4)); PLOTTEXT(ST10,10,1.5,0);
NOM:  MOVE(1,0.77,0.819*MAX); ST10:=" S  =      ";
      FRACNO(S1,ST10(6|4)); PLOTTEXT(ST10,10,1.5,0);
      MOVE(1,0.77,0.78*MAX); FRACNO(S2,ST10(6|4));
      PLOTTEXT(ST10,10,1.5,0); MOVE(1,0.77,0.741*MAX);
      FRACNO(S3,ST10(6|4)); PLOTTEXT(ST10,10,1.5,0);
      IF M=0 THEN GOTO NON;
      MOVE(1,0.815,0.887*MAX); MOVE(2,0.83,0.887*MAX);
NON:  MOVE(1,0.82,0.819*MAX); PLOTTEXT("AA",2,0.75,0);
      MOVE(1,0.82,0.78*MAX); PLOTTEXT("AB",2,0.75,0);
      MOVE(1,0.82,0.741*MAX); PLOTTEXT("BB",2,0.75,0);
      MOVE(1,1.0,0.725*MAX); MOVE(2,1.0,0.975*MAX);
      MOVE(0,0.76,0.975*MAX); MOVE(0,0.76,0.725*MAX);
      MOVE(0,1.0,0.725*MAX); ENDPIC; WRITE(" "); GOTO NEW;
STOP: ;
END.
```

CELL CYCLE SIMULATION

```
BEGIN INTEGER NC,R,I,J,K,P,T,PAR,LAST,EXPTS,NL,NM,NCL,AX;
      REAL A,B,MU,TOP; LOGICAL HARD,REL,PC,ZERO,TEXT;
      STRING(4)COPY; STRING(1)CYC; STRING(4)STOP;

PROCEDURE NEWLINE(INTEGER VALUE N);
BEGIN INTEGER I;
      FOR I:=1 UNTIL N DO IOCONTROL(2)
END;

PROCEDURE PRINT(REAL VALUE X,F);
BEGIN R_FORMAT:="A";
      R_D:=ROUND(10*(F-ENTIER(F)));
      R_W:=ENTIER(F)+R_D+1;
      WRITECN(X)
END;

REAL PROCEDURE MAX(REAL ARRAY A(*); INTEGER VALUE M,N);
BEGIN INTEGER I; REAL L;
      L:=A(M);
      FOR I:=M+1 UNTIL N DO IF L<A(I) THEN L:=A(I);
      L
END;

LONG REAL PROCEDURE FACTORIAL(INTEGER VALUE K);
IF K<3 THEN K ELSE K*FACTORIAL(K-1);

REAL PROCEDURE F(REAL VALUE T,M; INTEGER VALUE K);
BEGIN REAL V; IF T<=0 THEN V:=0 ELSE
      BEGIN M:=K/M; K:=K-1;
            V:=M*(M*T)**K*EXP(-M*T)/FACTORIAL(K)
      END;
      V
END;

REAL PROCEDURE RAT(REAL VALUE A,B,C);
A+B*C;

REAL PROCEDURE FIT(REAL ARRAY A(*,*); REAL ARRAY B(*);
                   INTEGER VALUE N);
BEGIN INTEGER I; REAL SS,T1,T2; SS:=0;
      FOR I:=1 UNTIL N DO
      BEGIN A(3,I):=100*A(3,I)/A(2,I);
            T1:=0.01*A(3,I); IF T1<0 THEN T1:=0;
            T2:=0.01*B(ROUND(A(1,I))); IF T2<0 THEN T2:=0;
            SS:=SS+A(2,I)*(ARCSIN(SQRT(T1))
                  -ARCSIN(SQRT(T2)))**2
      END;
      SS
END;

REAL PROCEDURE ARCSIN(REAL VALUE X);
IF X<0.5 THEN ARCTAN(X/SQRT(1-X*X)) ELSE 0.5*PI-ARCTAN(SQRT(1-X*X)/X);
```

```
PROCEDURE GRAPH1(REAL ARRAY P(*,*); INTEGER VALUE N;
                 REAL ARRAY Q(*); INTEGER VALUE X1,X2,Y1,Y2;
                 REAL VALUE XS,YS; STRING TITLE);
BEGIN IF TEXT THEN PLOT(P,N,X1,X2,Y1,Y2,XS,YS);
      GRAPH(Q,X1,X2,Y1,Y2,XS,YS,TITLE)
END;

PROCEDURE GRAPH2(REAL ARRAY Q(*); INTEGER VALUE X1,X1A,X2,Y1,Y2;
                 REAL VALUE XS,YS);
BEGIN INTEGER I,J; STRING TITLE;
      I:=13; TITLE(0|13):="FLM LABEL AT ";
      NUM(X1,TITLE(13|3));
      GRAPH(Q,X1A,X2,Y1,Y2,XS,YS,TITLE)
END;

PROCEDURE PLOT(REAL ARRAY P(*,*); INTEGER VALUE N,X1,X2,Y1,Y2;
               REAL VALUE XS,YS);
BEGIN INTEGER I;
      SCALE(XS,YS,X1-2/XS,Y1-2/YS);
      FOR I:=1 UNTIL N DO
      IF (P(1,I)>=X1) AND (P(1,I)<=X2) THEN
      BEGIN MOVE(1,P(1,I),P(3,I));
            POINT(0)
      END;
      MOVE(1,X1-2/XS,Y1-2/YS)
END;

PROCEDURE GRAPH(REAL ARRAY Q(*); INTEGER VALUE X1,X2,Y1,Y2;
                REAL VALUE XS,YS; STRING TITLE);
BEGIN INTEGER I;
      SCALE(XS,YS,X1-2/XS,Y1-2/YS);
      IF TEXT THEN
      BEGIN AXIS(0,X1,Y1,10,-ENTIER(0.1*(X1-X2)));
            AXIS(1,X1,Y1,0.1*(Y2-Y1),10);
            LABEL(X1,X2,Y1,Y2,XS,YS);
            MOVE(1,X1,Y2+0.05*(Y2-Y1));
            PLOTTEXT(TITLE,16,3,0);
            GRAM(1.0)
      END;
      MOVE(-2,X1,Q(X1)); THICK(1);
      FOR I:=X1 UNTIL X2 DO MOVE(0,I,Q(I));
      ENDPIC
END;

PROCEDURE GRAM(REAL VALUE SIZE);
BEGIN REAL K; K:=0.0875*SIZE;
      PLOTTEXT("DRARM",5,SIZE,0);
      PLUMOVE(ROUND(-1029*K),ROUND(514*K));
      PLOTTEXT("N",1,SIZE,0);
      PLUMOVE(ROUND(-343*K),ROUND(-1029*K));
      PLOTTEXT("W",1,SIZE,0);
      PLUMOVE(ROUND(-800*K),ROUND(100*K));
      PLOTTEXT("M      G",7,0.5*SIZE,0);
      PLUMOVE(ROUND(-1200*K),ROUND(1029*K));
      PLOTTEXT("G      S",7,0.5*SIZE,0)
END;
```

```
PROCEDURE AXIS(INTEGER VALUE I; REAL VALUE X,Y,DU; INTEGER VALUE N);
FORTRAN "EGRID";

PROCEDURE POINT(INTEGER VALUE I);
FORTRAN "POINT";

PROCEDURE MOVE(INTEGER VALUE I; REAL VALUE X,Y);
FORTRAN "EPLOT";

PROCEDURE ENDPIC;
FORTRAN "ENDPIC";

PROCEDURE ABSMOVE(INTEGER VALUE IX,IY);
FORTRAN "ABSMOV";

PROCEDURE SCALE(REAL VALUE XS,YS,X0,Y0);
FORTRAN "SCALEM";

PROCEDURE THICK(INTEGER VALUE I);
FORTRAN "SETLIN";

PROCEDURE PLUMOVE(INTEGER VALUE IDX,IDY);
FORTRAN "RELMOV";

PROCEDURE PLOTTEXT(STRING(256) VALUE ST; INTEGER VALUE NC;
                   REAL VALUE CS,ANG);
FORTRAN "PLTEXT";


PROCEDURE LABEL(REAL VALUE X1,X2,Y1,Y2,XS,YS);
BEGIN INTEGER I,Z; STRING(3) STR;
      FOR I:=0 UNTIL 5 DO
      BEGIN Z:=ROUND(Y1+0.2*I*(Y2-Y1)); NUM(Z,STR);
            MOVE(1,X1-2/XS,Z-0.35/YS);
            PLOTTEXT(STR,3,2,0)
      END;
      FOR I:=0 UNTIL ENTIER(0.1*(X2-X1)) DO
      BEGIN Z:=ROUND(X1+10*I); NUM(Z,STR);
            MOVE(1,Z-1.4/XS,Y1-1.5/YS);
            PLOTTEXT(STR,3,2,0)
      END
END;

PROCEDURE NUM(INTEGER VALUE Z; STRING(3) RESULT STR);
BEGIN INTEGER J,K; J:=0;
      FOR K:=100,10,1 DO IF (K=1) OR ((Z DIV K)>0) THEN
      BEGIN STR(J|1):=CODE(240+(Z DIV K));
            Z:=Z REM K; J:=J+1
      END ELSE BEGIN STR(J|1):=" "; J:=J+1 END;
      IF (STR(0|1)¬=" ") AND (STR(1|1)=" ") THEN STR(1|1):="0";
END;
```

```
WRITE("ENTER CODE FOR TYPE OF GRAPHS"); NEWLINE(1); READ(COPY);
HARD:=COPY="HARD"; IF HARD THEN WRITE(COPY);
TEXT:=¬ (COPY="OVER"); ZERO:=(COPY="ZERO") OR (COPY="OVER");
WRITE("DURATION OF EXPERIMENT?"); NEWLINE(1); READON(LAST);
IF HARD THEN BEGIN PRINT(LAST,3); WRITEON("HOURS") END;
WRITE("MEAN TIME IN GO?"); NEWLINE(1); READON(MU);
IF HARD THEN BEGIN PRINT(MU,2.1); WRITEON("HOURS") END;
WRITE("PERCENTAGE OF CELLS LEAVING GO?"); NEWLINE(1); READON(NC);
IF HARD THEN PRINT(NC,5);
WRITE("PARAMETER FOR ERLANGIAN DISTRIBUTION?"); NEWLINE(1);
 READON(PAR); IF HARD THEN PRINT(PAR,2);
WRITE("NUMBER OF CYCLES POSSIBLE?"); NEWLINE(1); READON(K);
IF HARD THEN PRINT(K,2);
WRITE("DOES DECYCLING PROBABILITY (X) DEPEND ON TOTAL CELLS (T)");
WRITE("OR NUMBER (D) WHICH HAVE LEFT THE PROLIFERATIVE CYCLE?");
NEWLINE(1); READON(CYC); REL:=CYC="D"; IF HARD THEN WRITE(CYC);
WRITE("PARAMETERS FOR RELATIONSHIP OF X WITH ",CYC,"?"); NEWLINE(1);
READON(A,B); IF HARD THEN BEGIN PRINT(A,1.2); PRINT(B,1.5) END;
WRITE("NUMBER OF FLM EXPERIMENTS AND TIMES OF LABELLING?");
NEWLINE(1); READON(EXPTS); PC:=EXPTS>0;


BEGIN REAL ARRAY GO,N,S,M,D,X,CL,GF(-1::LAST); REAL ARRAY HOLD(1::4*K);
      REAL ARRAY W,PP(1::4,1::K); REAL ARRAY L(0::EXPTS,0::LAST);
      REAL ARRAY G(1::4,1::K,0::LAST);
      INTEGER ARRAY TAU(0::EXPTS);

IF PC THEN FOR I:=1 UNTIL EXPTS DO READON(TAU(I));
IF HARD THEN
BEGIN IF PC THEN
      BEGIN PRINT(EXPTS,1); WRITEON("EXPERIMENTS, LABELLING AT ");
            FOR I:=1 UNTIL EXPTS DO PRINT(TAU(I),3); WRITEON("HOURS")
      END ELSE WRITE("NO FLM EXPERIMENTS")
END;

WRITE("DURATIONS OF G1, S, G2 AND M FOR EACH CYCLE?"); NEWLINE(1);
FOR P:=1 UNTIL 4*K DO READON(HOLD(P));
IF HARD THEN FOR R:=1 UNTIL K DO
BEGIN FOR P:=1,2,3,4 DO PRINT(HOLD(P+4*(R-1)),5.1);
      NEWLINE(1)
END;
FOR P:=2 UNTIL 4*K DO HOLD(P):=HOLD(P-1)+HOLD(P);
FOR R:=1 UNTIL K DO FOR P:=1,2,3,4 DO W(P,R):=HOLD(P+4*(R-1));

GO(0):=N(0):=100; D(0):=S(0):=M(0):=0; X(-1):=X(0):=A;
IF PC THEN FOR I:=1 UNTIL EXPTS DO L(I,0):=C;
```

```
FOR T:=1 UNTIL LAST DO
BEGIN GO(T):=GO(T-1)-NC*F(T,MU,PAR);
M(T):=0; D(T):=D(T-1); N(T):=N(T-1); S(T):=S(T-1);
      IF PC THEN FOR I:=1 UNTIL EXPTS DO L(I,T):=0;
      X(T):=2*X(T-1)-X(T-2);
      FOR R:=1 UNTIL K DO
      BEGIN FOR P:=1,2,3,4 DO
            BEGIN PP(P,R):=1;
                  FOR J:=1 UNTIL R-1 DO IF T>W(P,R)-W(4,J) THEN
                  PP(P,R):=PP(P,R)*(1-X(ROUND(T-W(P,R)+W(4,J))));
                  G(P,R,T):=NC*(2**(R-1))*PP(P,R)*F(T-W(P,R),MU,PAR)
            END;
            D(T):=D(T)+2*X(T)*G(4,R,T);
            N(T):=N(T)+G(4,R,T);
            S(T):=S(T)+G(1,R,T)-G(2,R,T);
            M(T):=M(T)+0.5*(W(4,R)-W(3,R))*(G(3,R,T)+G(4,R,T));
            X(T):=RAT(A,B,IF REL THEN D(T) ELSE N(T));
            IF PC THEN FOR I:=1 UNTIL EXPTS DO FOR J:=1 UNTIL R DO
            L(I,T):=L(I,T)+0.5*(G(3,R,T)+G(4,R,T))*
              (IF T<TAU(I)+W(3,R)-W(2,J) THEN 0 ELSE
               IF T<TAU(I)+W(4,R)-W(2,J) THEN T-TAU(I)-W(4,R)+W(2,J)
               ELSE IF T<TAU(I)+W(3,R)-W(1,J) THEN W(4,R)-W(3,R) ELSE
               IF T<TAU(I)+W(4,R)-W(1,J) THEN TAU(I)+W(4,R)-W(1,J)-T
               ELSE 0)
      END
END;

FOR T:=0 UNTIL LAST DO
BEGIN IF PC THEN FOR I:=1 UNTIL EXPTS DO
      L(I,T):=IF M(T)=0 THEN 0 ELSE 100*L(I,T)/M(T);
      S(T):=100*S(T)/N(T);
      M(T):=100*M(T)/N(T);
      I:=ROUND(T-W(1,1)+W(1,2)-W(4,1));
      GF(T):=100*(1-(D(T)+GO(IF I>0 THEN I ELSE 0))/N(T));
      CL(T):=100*(1-GO(IF T<W(1,1) THEN 0 ELSE ROUND(T-W(1,1)))/N(T));
END;
```

```
WRITE("EXPERIMENTAL DATA?"); IOCONTROL(2);
READ(NL,NM,NCL);
BEGIN REAL ARRAY LI(1::3,0::NL); REAL ARRAY MI(1::3,0::NM);
      REAL ARRAY CLI(1::3,0::NCL);

IF NL>0 THEN
BEGIN FOR I:=1 UNTIL NL DO READON(LI(1,I),LI(2,I),LI(3,I));
      WRITE("RESIDUAL SUM OF SQUARES FOR LABELLING CURVE = ");
      PRINT(FIT(LI,S,NL),4.3)
END;
IF NM>0 THEN
BEGIN FOR I:=1 UNTIL NM DO READON(MI(1,I),MI(2,I),MI(3,I));
      WRITE("RESIDUAL SUM OF SQUARES FOR MITOSIS CURVE  = ");
      PRINT(FIT(MI,M,NM),4.3)
END;
IF NCL>0 THEN
BEGIN FOR I:=1 UNTIL NCL DO READON(CLI(1,I),CLI(2,I),CLI(3,I));
      WRITE("RESIDUAL SUM OF SQUARES FOR CONTINUOUS LABEL= ");
      PRINT(FIT(CLI,CL,NCL),4.3)
END;
WRITE("FINAL VALUES OF X, D AND T ARE"); NEWLINE(1);
PRINT(X(LAST),1.2); PRINT(D(LAST),5); WRITEON("AND");
      PRINT(N(LAST),5);
TOP:=IF NCL>0 THEN MAX(CLI(1,*),1,NCL) ELSE LAST;
WRITE("CUMULATIVE LABELLING INDEX AT "); PRINT(TOP,3.0);
WRITEON("HOURS IS "); PRINT(CL(ROUND(TOP)),3.1);
WRITEON("PER CENT"); NEWLINE(1);
WRITE("MAXIMUM VALUE OF ""GROWTH FRACTION"" IS ");
PRINT(MAX(GF,1,LAST),3.1); WRITEON("PER CENT"); NEWLINE(1);

WRITE("ENTER ""STOP"" IF GRAPHS NOT NOW REQUIRED"); NEWLINE(1);
READON(STOP); IF STOP="STOP" THEN GOTO FIN; ABSMOVE(0,0);
AX:=IF ZERO THEN 0 ELSE ENTIER(W(1,1))-5;
GRAPH1(LI,NL,S,AX,LAST,0,50,0.25,0.4,
       "LABELLED NUCLEI ");
IF ¬ ZERO THEN AX:= ENTIER(W(3,1))-5;
GRAPH1(MI,NM,M,AX,LAST,0,5,0.25,4,
       "MITOTIC NUCLEI  ");
IF ¬ ZERO THEN AX:= ENTIER(W(1,1))-5;
GRAPH1(CLI,NCL,CL,AX,LAST,0,100,0.25,0.1,
       "CONTINUOUS LABEL");
IF HARD THEN
BEGIN GRAPH(N,0,LAST,0,500,0.25,0.02,"POPULATION SIZE ");
      GRAPH(GF,0,LAST,0,100,0.25,0.1,"GROWTH FRACTION ");
      IF PC THEN FOR I:=1 UNTIL EXPTS DO
      GRAPH2(L(I,*),TAU(I),IF TAU(I)>W(3,1) THEN TAU(I) ELSE
        ENTIER(1+W(3,1)),LAST,0,100,0.25,0.1)
END;

FIN: END END
END.
```