SPECIFIC COMPUTER PROGRAMMES FOR GENETIC SELECTION

Spezifische Berechnungsprogramme für die genetische Selektion

Programas específicos de computación para selección genética

D. BURNELL *

Introduction

There are several possible approaches to the construction of selection simulation packages — possibly as many approaches as there are experimenters. We will examine some of the basic requirements that a general simulation system should meet and then compare two of the resulting systems. The two packages that will be discussed are the GSD-1 system developed at the University of California and a system designed by the author.

Among the basic requirements that a general simulation package should meet are ease of use, simplicity of modeling input and verification and easily used output systems. Another requirement, if the system is to bear a title including the word «general», is that of generality. The internally generality can be achieved in several ways. We can construct a program that is easily modified internally in terms of the programming language or we can include many user controlled options. The GSD-1 simulation package uses both of the above techniques. However, in the new system we've taken another approach. We use the input parameters to construct or generate a new program that meets the requirements of the simulation and, at the same time, allows the more sophisticated user the option of inserting segments of code for actual program requirements. Also the new system will allow the user considerably more freedom in handling the output data sets and the corresponding analysis.

* Department of Forestry, Washington State University, Pullman, Washington 99163, United States.

THE GSD-1 GENETIC SELECTION SIMULATION SYSTEM

The GSD-1 general simulation system is a computer program written at Davis, California during the early and mid 1960s. The program was written in Fortran IV and was implemented on some of the 7000 series of IBM machines. In 1972 the program was converted to the 360 series of machines.

The program's several virtues include ease of use, rapid execution and a

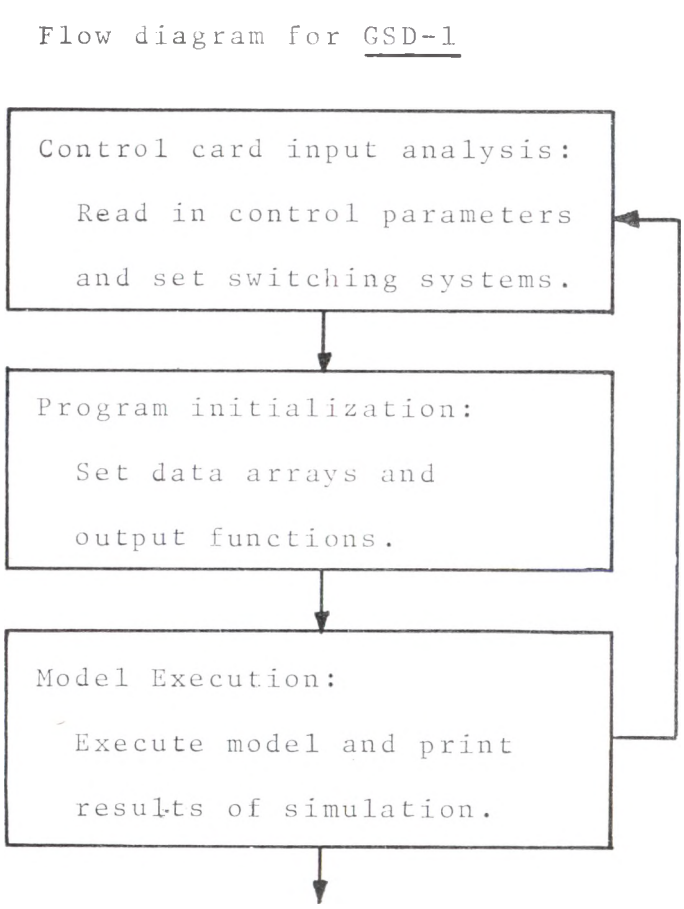

FIG. 1

variety of input and output options. Among its output capabilities are generation by generation output of varicus data including gene and zygote frequencies, chrosomal type frequencies and survival and mortility results. The program also sums over-rep results and plots most of the output variables where reasonable.

There are several liabilities of the system in terms of simulation or structure. The program will admit only a maximum number of genes or *loci* and bas a

population maximum of 1024. Also the display or output attributes of the system are limited and must be determined *before* the beginning of the simulation. Any modification of the program structure requires a more than minimal knowledge of the inner workings of the program. The only exception is if the modifications or changes can be made at the subprocedure level. Also the program is a complete module or package and the user carries along the complete system regardless of his input options.

In Figure I we note that the user sets up control cards to run the system and enters the program as nothing more than a data sourse and control manager. Certainly he does not have the option of determining the internal structure of the program.

GENSIM

The GENSIM system was designed to meet the requirements laid down in Section I. We accepted the fact that the notion of «ease of use» will vary according to the capabilities of the users of the system. The general philosophy was to allow a minimally prepared user to generate and use functional models using only the basic ideas of the system. At the same time the more sophisticated user will be able to generate complicated models by inserting in-line coding structures at appropriate times during model definition. Figure 2 is a flow diagram of the system with the break-down at major levels of the system.

DEFMOD: The model is defined using a simple replacement definition structure. The definition of the various parameters that are available is accomplished by the proper junction of simple arithmetic expressions and reserved words. There are several points in the system definition where a user may insert code that will replace and/or add to the ordinary code definition. These points include selection by fitness, evaluation of cross-over functions and sub-function insertion.

GENMOD: In GENMOD the model defined in DEFMOD is used to generate a Fortran IV source program and the resulting program is compiled or translated into executible code. As far as the user is concerned the program generated in GENMOD can be thought of as *the* model. The source program is available for perusal and modification at this point and the user can save the program for slight modification and re-runs.

EXECMOD: In this phase the model generated in GENMOD is executed and the output files that corresponding to the simulation of the model are stored on an appropriate system device. The storage is on a «named» file-sub-file basis that makes the files readily available during the analysis of the output.

DATANAL: The last phase of the system was designed as an open file analysis system to allow the user to examine the results of the simulation with the aid of various statistical analysis and display packages. These packages include:

a) plotting of various outputs,
b) listing,
c) regression analysis,
d) analysis of variance and
e) simple summaries.

Also there is a provision for the generation of «hard-copy» output.

Flow chart for GENSIM

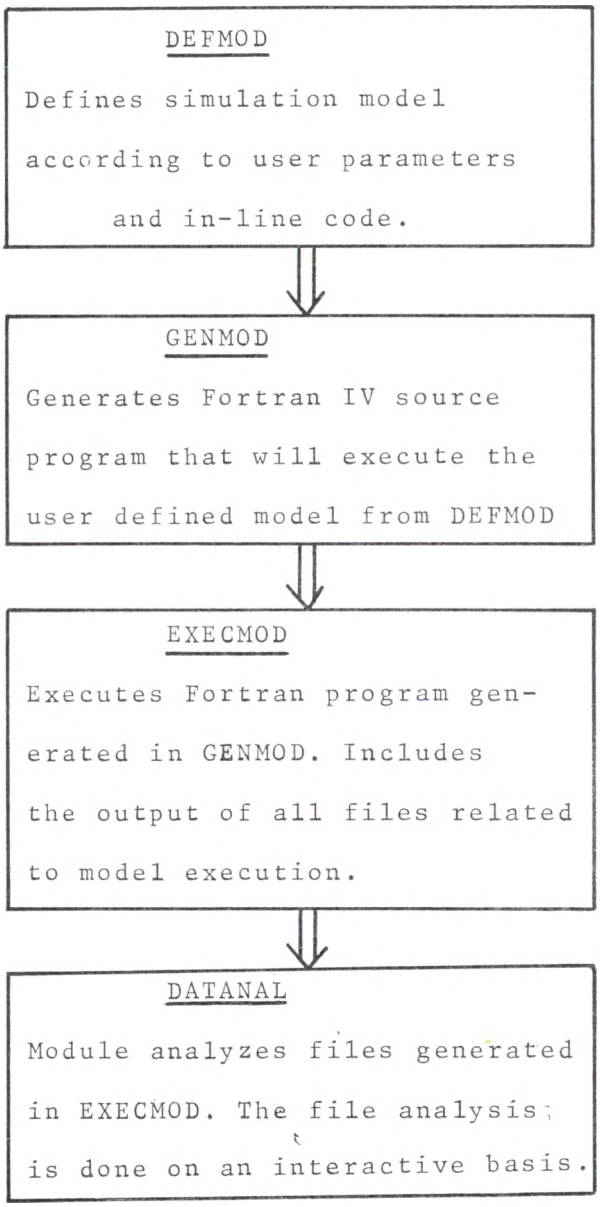

FIG. 2

Since the terminal used is expected to be a graphics terminal any information that is required on a permanent basis must be directed to the appropriate device.

ZUSAMMENFASSUNG

Meine Arbeit bespricht das allgemeine Problem der Zusammenstellung von Rechnertypen genetischer Systeme. Sie enthält auch eine kurze Erklärung des GSD-1 Programms und einen genaueren Umriss eines neuen Programms. Die Hauptbetonung der Arbeit liegt in der Entwicklung eines allgemeinen Systems, das nachstehende Vorstellungen einschliesst: 1) die dem Benutzer handgehabte Bequemlichkeit der Nutzbarmachung; 2) höchste Nutzbarmachung freier Struktur; und 3) Einfachheit des Programms, der Gültigkeitserklärung und der Analyse. Das GSD-1 System wurde zwecks Lehren und Forschen vom Autor und A. FRASER während der 60er Jahren entwickelt. Bis jetzt hat man es hauptsächlich dafür verwandt, neue Programme zu entwickeln und auszuwerten. Wie es so oft der Fall mit vielen dieser allgemeinen Systeme ist, hat das GSD-1 System mehrere Nachteile. Es ist unbiegsam in Bezug auf Input parameter und Output und lässt sich innerlich nur schwer abändern. Das neue System (GENSIM) ist das Ergebnis eines Versuchs, obengenannte Probleme durch die Aenderung der allgemeinen Struktur des Programms zu lösen. Das wird dadurch ausgeführt, dass man den Input benutzt, um ein Rechnerprogramm zu erschaffen, das nur die dafür notwendige Struktur einbezieht, das Programm auszuführen. Der ganze Output wird innerhalb eines Ordner-systems behalten, das es dem Benutzer mit Hilfe einer vorher erklärten Handlungsweise erlaubt, die Ergebnisse zu prüfen.

RESUMEN

Lo que se propone es discutir el problema general de adaptar la computación electrónica a la genética, explicar la agrupación GSD-1 de simulaciones y dar un tratamiento más completo de una nueva agrupación. El énfasis principal será sobre el desarrollo de un sistema general que incluye las nociones de facilidad de empleo, utilización máxima de estructura libre, y sencillez en la estructura, la verificación y el análisis de los modelos. El sistema GSD-1 fue desarrollado por el autor y el profesor A. FRASER durante la década de los 60 para emplearlo en investigaciones y en la enseñanza. El sistema ha sido utilizado principalmente en el desarrollo y en la evaluación de los modelos. Igual que otros muchos de estos sistemas generales, el sistema GSD-1 ofrece varias desventajas. Es inflexible tanto en la definición de los parámetros como en la producción, y es difícil modificarlo internamente. El nuevo sistema, GENSIM, es el resultado de una tentativa para eliminar los problemas, cambiando la estructura general de la simulación y del análisis de los datos. Esto se hace empleando los parámetros para producir un programa de computación eléctrica que incorpora sólo la estructura necesaria para efectuar la simulación. Toda la producción se guarda en un sistema catalogado que permite repasar los resultados con la ayuda de procedimientos ya existentes de gráficas y análisis.