

SIMULATION OF ECOLOGICAL SYSTEMS IN CSMP

Jan Goudriaan and Cornelis T. de Wit

Department of Theoretical Production Ecology

Agricultural University

WAGENINGEN

This text summarizes the course in simulation
at the Hebrew University of Jerusalem in
March 1973 and is not for publication.

CONTENTS

<u>Chapter</u>	<u>Page</u>
Introduction	1
Exponential growth	5
The growth of yeast	9
Time constant	13
Relational diagrams	15
Competition between plants	17
Competition between several plant species	22
Growth and competition in paramecia	24
Simulation of age dependent processes	33
Flow of heat in soil	39
Appendix	41
Literature	45
Index of described labels and functions	46

INTRODUCTION

Some concepts

L A model is a representation of an original. If the model would not be different from the original, there would be no need to make it. It agrees with the original on relevant points. The differences between model and original make the model simpler, easier to be handled and more lucid than reality. Simplification however means that relations are wrongly represented or just omitted. Sometimes this is done for easier handling, sometimes just because our knowledge of reality is incomplete. In a computational model for example simplifications should often be introduced to enable a solution. The analytical methods of solution demand more simplicity than the numerical methods using digital computers and simulation languages.

A simple example of a mathematical model is the relation between the distance (s) covered by a falling apple, the gravitational acceleration (g) and the time from the moment of release (t): $s = 0,5g t^2$.

There are many non-mathematical models:

A map is a simplification of the original that contains relevant information and allows measurements. Dependent on the purpose roads, railways, isohypses or soil types are represented.

A spatial model of a molecule represents the configuration of the composing atoms. This is especially important in organic chemistry, where the tetraeder is the main basic form.

A scale model of a ship enables measurements of the resistance in the water. The internal structure need not be copied. However, laws of scale must be satisfied, in the reduction from original to model to maintain the original relations between viscosity, density, velocity, length, etc.

ITEM A system is a limited part of reality with related elements. The set of relations is called the structure of the system. Both a model and a system have structure. It follows from the definition of a system that a model can be considered a system, but that in general the opposite is not true. A system M is a model of system O provided they have a partially overlapping structure. Which parts of O must be represented by M is determined by the requirements of relevance to be imposed on the model. Which parts should not be taken into M is dictated by convenience of handling.

Examples of a system are a cell, a plant, a cow, a field with a crop, a farm. The boundary between system and environment is preferentially chosen, such that the system is isolated. That means the behaviour of the system does not depend on its environment. If this is impossible, the boundary is chosen such that the environment admittedly influences the system, but the system itself not the environment. To achieve this it is often necessary to take the boundary wider than seems necessary for the problem. If for example the influence of temperature on the growth of plants is studied in a climate room, the climate room is part of the environment if the construction is so well done that temperature, moisture content and light intensity do not depend on the size of the plants. In badly constructed and conditioned climate rooms this requirement is not met so that the climate room itself is part of the system.

Speaking of the behaviour of a system implies that the system is dynamic, that means it changes with time. A simplified representation of a dynamic system is a dynamic model.

Simulation is the building of a model and the studying of its behaviour.

Simulation is useful if it increases the insight in reality by extrapolating and analogy, if it is the basis for the design of new experiments and if the model accounts for most relevant phenomena and contains no assumptions proved to be false. The latter requirement seems obvious. False assumptions are often made to enable an analytical solution. By the more recent simulation techniques limitations of this sort are often overcome so that the attention can turn from the solution technique to the study of the behaviour of model and system.

Analogous computers

Often electrical analogue models can be built. Such a model of a falling apple might consist of an apparatus with two condensers. The first one is charged with a current analogous to the gravitational acceleration and maintains a voltage analogous to the velocity. The second condenser is charged with a current proportional to the voltage of the first one and thus maintains a voltage analogous to the covered distance.

Question: Trace back by means of the mathematical relations if the above conclusions are right.

The integration in the condensers takes place continuously and simultaneously like in reality. Any moment the condition of the system is fully determined by the voltages of the condensers. The analogous computers that were developed on basis of this principle have proved to be quite useful in simulation of continuous processes. There are however some objections. The user should adapt the scale of the variable to the voltage range of the circuit elements and has to accept their inaccuracies. The resulting difficulties rapidly increase with increasing size and complexity of the systems to be simulated and sharper requirements of accuracy. These problems do not show up in simulation with digital computers. It is remarkable that the first languages to simulate continuous systems have been developed for digital machines to validate results of simulations with analogous machines and to facilitate the assessment of the scaling factors.

In further development of simulation languages for digital machines, these appeared to be preferable over analogous machines, if there are discontinuous processes and many empirical relationships in the model. They are however at a disadvantage if systems are simulated where changes occur at very different time scales.

In future the disadvantages of both machines may be eliminated and the advantages combined in hybrid computers where both types are combined.

Features of simulation languages

A digital computer, where all executions are discrete and take place in a sequential order, seems to be the most unsuitable instrument to simulate continuous systems where changes are parallel and continuous. The main feature of simulation languages is to overcome these two limitations. It is based on the axiom that changes of conditions in systems are not mutually dependent, but derived each one separately from the state of the system. In a mixture of yeasts the rates of growth do not depend upon each other, but each one depends separately upon the concentration of food and waste products, and on its own amount.

Thus in principle structural equations, that means n equations with n rates of changes, do not occur in the model. This gives the possibility to make a routine in the languages to sort the equations that describe the system. The sorting routine determines the order of executions during the process of calculation. All rates of change between time t and $t+\Delta t$ are

calculated from the condition at time t and if necessary data from the past. Only after the calculation of all the rates at the moment of the simulation, the changes are executed by integration over a small time interval. Integration is done semi parallel.

For practical reasons the time interval cannot be infinitely small. Finite intervals produce systematical errors but these can be kept within acceptable limits by the use of good numerical integration methods. The simplest one is the Eulerian or rectangular in which the new value of a state or integral equals the old value plus the product of the never changing time interval and the rate of change. There are however also integration methods in which the time interval depends upon the rate of change. If the rate of change is relatively large, small steps are taken, if the rate is small, large steps can be taken, always keeping the errors within acceptable limits.

CSMP

A commonly used simulation language is Continuous System Modeling Program, developed for IBM 360 and 370 machines.

The use of CSMP is explained in the rest of this manuscript employing examples of increasing complexity. These examples are available in CSMP, but it is advised to program the examples and self made variants.

In the margin of the text it is indicated what functions and constructions are being explained in the examples.

EXPONENTIAL GROWTH

The growth of bacteria and algae is often proportional to the amount of biomass that is present.

Question: What environmental conditions must be satisfied in any case?

This proportionality can be represented by the equation

$$dH/dt = RGR \times H$$

where H is the amount of organisms, measured in grams for example, T time in hours for example, dH/dt the rate of growth and RGR the relative growth rate.

Question: What are the dimensions of dH/dt and RGR?

Integration of the above differential equation yields the equation $H = IH \times e^{RGR \times T}$, where IH is the amount of organisms at time zero. Under these circumstances the amount increases exponentially with time.

Question: Calculate the value of H after 0, 2, 4, up to 10 hours for RGR equals 0.1 hour⁻¹ and IH equals 1 gram. Represent the results in a graph and connect the points by a smooth line.

For the execution of the above calculation one needs a slide rule or table of logarithms. If none of these is available the requested relation can also be found by the following stepwise solution.

If at a certain time T the amount of organisms equals H, the rate of growth at this moment equals RGRxH. During a short time interval DELT this rate of growth hardly changes so that at time T+DELT the amount of organisms equals approximately H+RGRxHxDELT. Using this new value of H the rate of growth at time T+DELT can be calculated and so the amount of organisms at time T+2xDELT.

Question: Calculate the values of H after 0, 2, 4, up to 10 hours for RGR equal to 0.1 hour⁻¹ and H set at 1 gram at time zero. Employ intervals of 2 hours and the following scheme:

TIME	H	RGRxH	RGRxHxDELT
0	1	0.1	0.2
2	1.2	etc.	
4			
6			
8			
10			

Represent the results in the graph and connect the points by straight lines.

A comparison of the stepwise and the analytical solution shows that the rate of growth is underestimated by the use of the stepwise solution. This is caused by the wrong assumption that the rate of growth remains the same during two hours, notwithstanding the increase of the amount of organisms during this period. It is to be expected that the deviation between the stepwise and the analytical solution decreases the shorter the time steps are.

Question: Execute the stepwise calculation also with intervals of 1 hour and 0.5 hour and represent the results after a growth of 0, 2, 4, ---- 10 hours in the same graph. Could you give a formula giving directly the size of H after n time intervals? What is the limit of this formula for DELT approaching zero?

A calculation with even shorter time steps is tedious and can better be done by using the simulation language CSMP, in which the problem reads:

```
TITLE EXPONENTIAL GROWTH
H=INTGRL (IH, RGRxH)
INCON IH=1.
PARAMETER RGR= 0.1
PRTPLT H
TIMER FINTIM=10., OUTDEL=0.5, DELT = 0.05
METHOD RECT
END
STOP
```

The first card mentions the title, which is repeated every page of output. On the second card it is stated that H equals IH at time zero, and is found later by integration of $RGR \times H$. The third and fourth card give the numerical values of the initial constant and the parameter. The next card indicates that the value of H should be plotted against time, whereby it is simultaneously given as a table.

IER

On the card beginning with `TIMER` it is said that the simulation should be pursued until 10 units of time are being elapsed (`FINTIM`), that a point must be plotted every 0.5 unit (`OUTDEL`) and that the time interval of the calculation is 0.05 unit (`DELT`).

Question: What parameter determines whether the 10 units of time mean 10 minutes, or 10 hours, or 10 days?

The `METHOD` card indicates that the integration should be executed employing the `RECT`angular method of Euler, that was also used in the previous examples of calculation by hand.

The `END` card indicates the end of the simulation program, and `STOP` of the whole calculation. If one would execute the calculations with an `RGR` of 0.2 as well it suffices to insert the cards `PARAMETER RGR=0.2` and `END` between the `END` and `STOP` mentioned above. This is called a rerun.

Question: Execute the simulation.

Make a graph of the amount of organisms after 10 days against the size of the time interval. Indicate the right answer by a thin horizontal line. If halving the time interval does not change this value by more than some 5 % the result of the simulation is acceptable in most situations, because the accuracy of the basic parameter is seldom better. What is the largest acceptable time interval when $RGR=0.1$?

How much larger or smaller is the acceptable time interval when $RGR=0.2$?

Finding the acceptable size of the time interval takes some time, and moreover the rates of change may be fast during some period of simulation and slow during another, so that the acceptable time interval is changing as well.

In CSMP integration methods may be used that adapt continually the size of the time intervals to the rates of change. The most universally applicable method (Runge-Kutta/Simpso) is called for by omitting the METHOD card.

Question: Does the time interval in the example decrease, increase, or
just stay equal with time?

THE GROWTH OF YEAST

Growth is only exponential if the factor that determines the rate of growth remains constant. This is the case when the amount of food is not limiting, and when accumulation of waste products does not occur or is not harmful. The mentioned constant factor in that case is the physiologically determined maximal rate of growth under optimal conditions.

Yeast, growing in a liquid solution with enough sugar as food, but in anaerob conditions, produces alcohol. An increasing concentration of alcohol hampers the formation of new buds, so that the growth will stop, even if there is still plenty of sugar.

Just as in the case of exponential growth it is assumed that the growth of the yeast will be proportional to the present amount, but in this case the growth rate will also be multiplied by a reduction factor that depends on the concentration of alcohol in the environment.

This is programmed by

$$Y=INTGRL (IY, RY)$$

where RY is expressed as

$$RY=RGRxYx (1.-RED)$$

Question: Why is the amount of sugar not mentioned?

Gause (in "The struggle for existence", 1934) stated in an experiment that the relative growth rate of the yeast *Saccharomyces* equals 0.218 hour^{-1} when alcohol is absent, and that bud formation is entirely prevented by an alcohol concentration of 1.5 % or more.

Assuming that the rate of bud formation and thus of growth decreases linearly with the alcohol concentration, it can be stated:

$$RED=ALC/LALC$$

The alcohol concentration itself is the integral of the rate of alcohol production:

$$ALC=INTGRL (0., ALCP)$$

starting with alcohol concentration zero.

Question: What is the dimension of ALCP?

Now it is assumed that the consumption of sugar and thus the production of alcohol is proportional to the rate of growth of the yeast:

$$ALCP = ALPF \times RY$$

Gause found for the proportionality factor:

PARAMETER ALPF=0.113

Question: Is this value only physiologically determined? Is it also valid for other similar experiments? (Notice the dimension).

Now finish the simulation program for the growth of yeast, and calculate the growth curve of the species *Saccharomyces* in case that the initial amount of yeast equals 0.45 units. Compare this with the observed data in table 1.

Try to choose sensible values of FINTIM and OUTDEL.

In the same table observed data of the species *Schyzosaccharomyces* are mentioned. Gause observed that the limiting alcohol concentration was the same for both species.

Question: Try to estimate a value for RGR and ALPF for the latter species and use these values for a simulation. Compare the results with the observations, repeat the estimation and simulation if needed.

Table 1: Measured volumes in indication marks on a centrifugation tube, for the species *Saccharomyces* and *Schyzosaccharomyces*. Mono-culture and mixed culture dependent on time (after Gause).

Age in hours	mono		mixed	
	Sacch.	Schizos.	Sacch.	Schizos.
0	0.45	0.45	0.45	0.45
6	0.37	-	0.375	0.291
16	8.87	1.00	3.99	0.98
24	10.66	-	4.69	1.47
29	12.50	1.70	6.15	1.46
40	13.27	-	-	-
48	12.87	2.73	7.27	1.71
53	12.70	-	8.30	1.84
72		4.87		
93		5.67		
117		5.80		
141		5.83		

If both species are growing in the same medium at the same time they influence each other by the production of alcohol. The combined alcohol production amounts:

$$ALC=INTGRL (0.,RY1 \times ALPF1+RY2 \times ALPF2)$$

The species are distinguished by adding a number to their names.

Question: Write a simulation program for the growth of both species.

Regulate the output by the following cards:

PRTPLT Y1 (RY1,RED,ALC)

PRTPLT Y2 (RY2,RED,ALC)

A graph will be given of the first variable. The other ones, enclosed by brackets, will be printed beside the graph.

Examine the output and compare the results with those of the experiment in table 1.

Growing together both species stops growth after some 50 hours. So it does not make sense to extend the simulation beyond this time. Superfluous computation can be prevented by inserting a finish condition:

FINISH ALC=FALC

FALC=0.99xLALC

These cards will stop the simulation at the moment that the alcohol concentration is one percent less than the limiting alcohol concentration.

Question: When will the simulation be finished if the card FINISH ALC=LALC is inserted rather than the above ones?

A comparison of simulated and real growth shows that the real growth of *Saccharomyces* in the mixture is slightly less than the simulated growth. It seems that the alcohol of the other species is more poisonous. Therefore it may be that *Schizosaccharomyces* produces another waste product beside ethanol, that is poisonous to the other species.

In the simulation programs it is supposed that the reduction factor is proportional to the alcohol concentration and it appears from a comparison between the real and the simulated results that this is an acceptable assumption. A more general assumption about the relation between alcohol concentration the reduction factor can be introduced by the expression:

RED=AFGEN (REDTB,ALC/LALC)

Here it is stated that the reduction factor RED is a function of the variable ALC/LALC. The shape of the function is given on a FUNCTION card.

In the mentioned case the function has the shape:

FUNCTION REDTB= (0.,0.),(1.,1.)

The first number of each pair is the value of the independent variable and the second one the corresponding value of the dependent variable. The values in between are obtained by linear interpolation.

Question: Plot the relation between the independent variable ALC/LALC and the dependent variable RED by hand for:

FUNCTION REDTB = (0.,0.),(0.25,0.5),(0.5,0.8),(0.75,0.9),(1.,1.)

What are the values of RED for ALC/LALC equal to 0.2,0.5,0.8, and 0.9?

If the reduction factor for the growth of yeast is proportional to the alcohol concentration the growth of yeast can be represented by the following differential equation:

$$dY/dt=RGRxYx(1.-Y/YM)$$

where YM is the maximal possible amount of yeast.

Question: Derive this differential equation by eliminating the alcohol concentration from the equations for the growth of yeast in the simulation program. Then calculate the values of RGR and YM for the two yeast species. Integrate the differential equation and calculate the course of growth for the species Saccharomyces. Compare the results with the simulated curve. Why does this differential equation hold only if the initial amount of yeast is sufficiently small?

The pattern of growth that follows from this differential equation is the logistic growth pattern.

When two species are growing in the same medium the rates of growth are often represented by the differential equations:

$$dY1/dt = R1xY1x(1.-A1xY1-B1xY2)$$

$$dY2/dt = R2xY2x(1.-A2xY1-B2xY2)$$

In general, this set of differential equations cannot be integrated into analytical expressions for Y1 and Y2 as dependent on time.

Question: Check that the simulation program for the mixed growth is practically covered by these differential equations and express the constants R1, R2, A1, A2, B1, B2 in the constants RGR1, RGR2, ALPF1, ALPF2 and LALC. Which constants are mutually equal? How do the constants in the differential equations change when one of the species is producing another waste product than ethanol that is poisonous for the other species?

Time constant

In dynamic systems the time constant characterizes on which time scale changes occur. It does not make sense to simulate the dynamic behaviour of a system employing a time interval DELT larger than the time constant of the system. If method RECT is employed, as a rule the time interval DELT should be smaller than one tenth of the time constant. Using RKS the variable value of DELT will turn out to be about half of the time constant.

To estimate the time constant of a system its differential equation should be approximated by:

$$\frac{dH}{dt} = \pm H/\tau$$

where τ is the time constant. If the first derivative has the same sign as the variable itself there is a positive feedback. This means that a once produced error is propagated more and more. In the opposite case, a negative feedback, an error will be damped out together with the variable itself. The behaviour of the relative error is however the same in both situations.

The simplest example of a positive feedback is the exponential growth:

$$\frac{dH}{dt} = RGR \times H$$

Obviously the time constant has the value $\tau = 1/RGR$.

Question: Mention a simple example with a negative feedback.

Often more complicated systems can be reduced to a differential equation of the form above. For the yeast the differential equation is

$$\frac{dH}{dt} = RGRxHx(1.-RED)$$

In the beginning RED may be neglected, so that the time constant is again $1/RGR$. Later on growth ceases so that the time constant becomes larger. DELT is however determined by the minimum value of the time constant $1/RGR$. It would be wrong to draw the conclusion that the time constant becomes infinite when H approaches its maximum. From what has been said on logistic growth it is possible to write the equation as:

$$\frac{dH}{dt} = RGRxHx(1.-H/HM)$$

When H approaches HM this is almost equal to

$$\frac{dH}{dt} = RGRxHMx(1.-H/HM) \quad \text{or}$$

$$\frac{dH}{dt} = RGRx(HM-H) \quad \text{or}$$

$$\frac{d(H-HM)}{dt} = -RGRx(H-HM)$$

Now, written for the new variable $H-HM$, there is a differential equation with a negative feedback and again a time constant $1/RGR$. At first the positive feedback becomes weaker and weaker, when H is $HM/2$ the feedback is zero and the time constant infinite, and subsequently the negative feedback becomes again stronger and stronger. This can be proved by taking the derivative of dH/dt to H, as this value represents the feedback of H on dH/dt . In the case of the logistic growth it follows:

$$\frac{d\left(\frac{dH}{dt}\right)}{dH} = RGRx\left(1.-\frac{2H}{HM}\right)$$

The time constant is the inverse of the absolute value of this coefficient.

Question: What is the time constant of a falling apple? What are the consequences for the time interval DELT? Are there similar consequences when METHOD RECT is being used? What is the source of this discrepancy?

RELATIONAL DIAGRAMS

Sometimes it is useful to indicate by means of a drawing how the relations are between parameters and variables occurring in a model. This can be done without mentioning the exact mathematical relationships. Often a scheme for representation is used, as developed by Forrester for industrial systems.



state variable or value of an integral



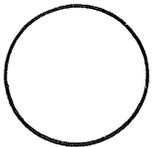
flow of material, into or from an integral



"flow" of information



rate of flow of material to an integral



variable, no rate or integral

name

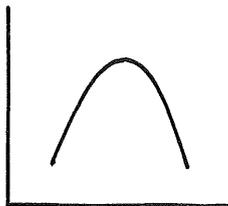
parameter



variable, derived from another relational diagram



time

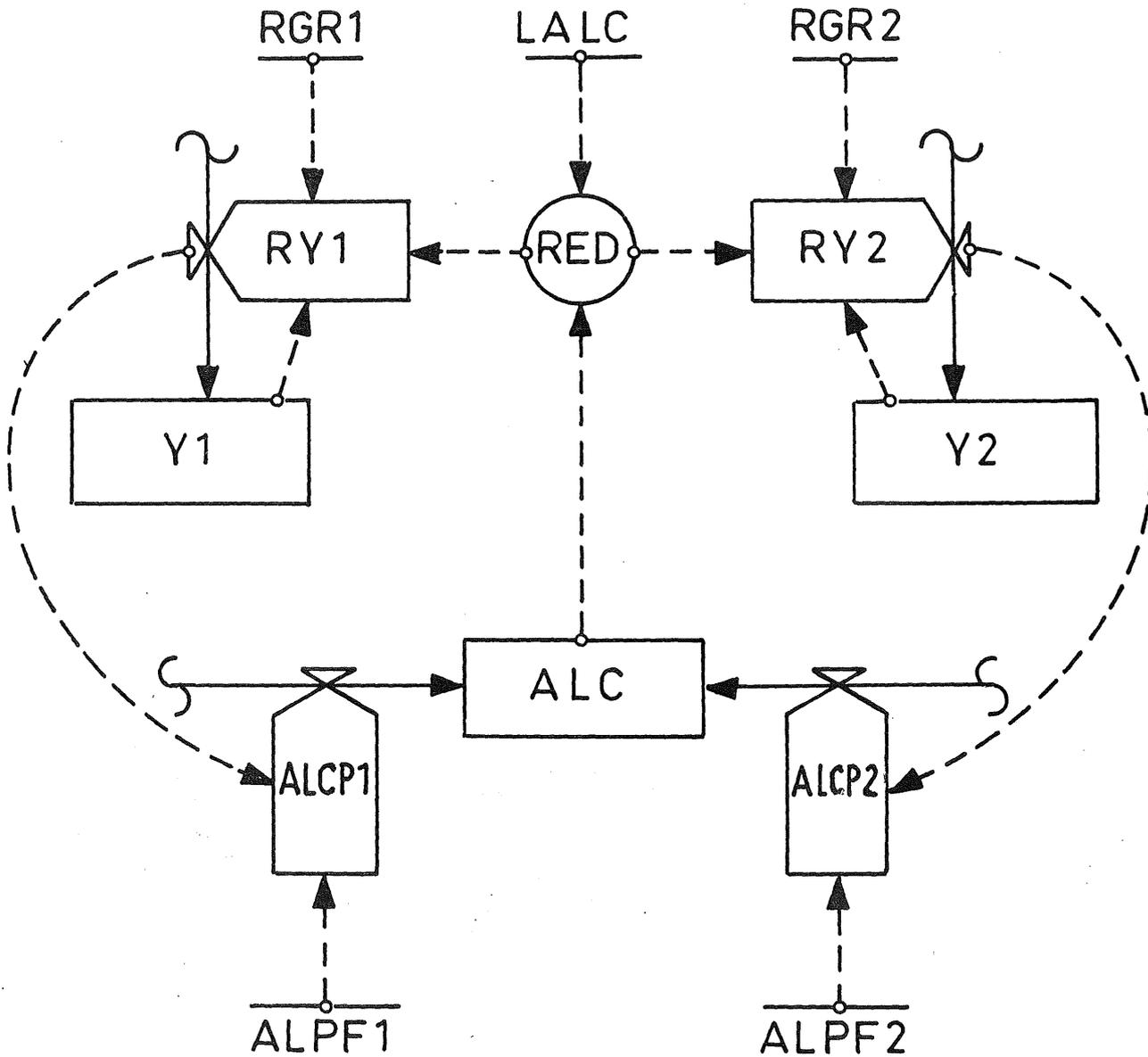


graphical relationship



sink or source, irrelevant to the model

The following relation diagram represents a mixed culture of two yeast species, that only interfere through the production of alcohol.

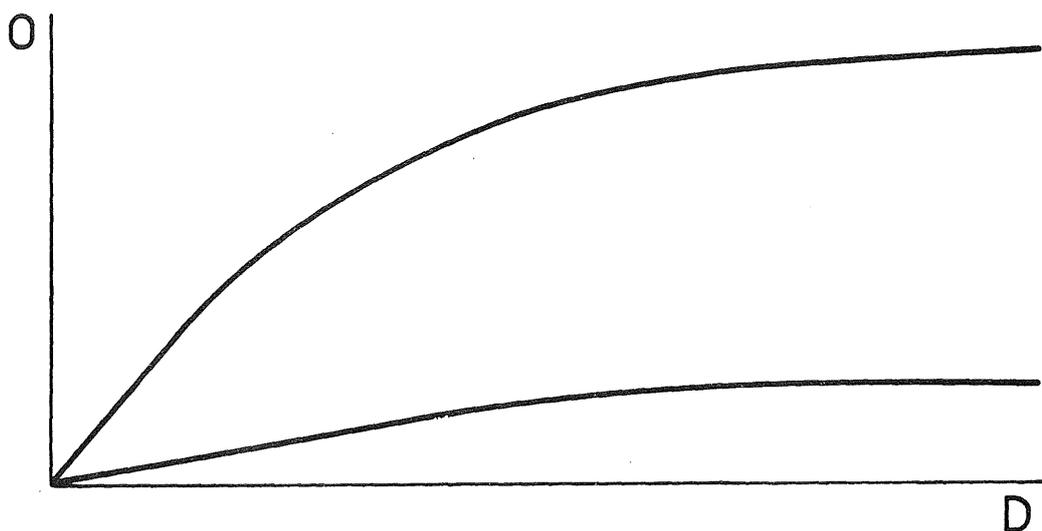


Question: Draw the relational diagram for a falling apple.

Which curiosity attracts the attention? Write the program as well.

COMPETITION BETWEEN PLANTS

The yield of a crop is related to the plant density and the length of the growing period. By yield is meant the total dry weight at any moment. In an early stage or at low plant density the yield per area is practically proportional to the plant density. In a later stage or in a high density, increase of plant density hardly gives any increase of yield. In the following figure the yield per area is represented as dependent on plant density at two different moments.



The drawn functions are reasonably represented by

$$O = \frac{BxD}{BxD+1} \times OM \quad (1)$$

where D is the density of sowing or planting, OM the maximal yield for large D, and B the space plant. The dimensions of B and D are inverse, i.e. m^2/plant and plant/m^2 resp. Both OM and B are a function of time, and independent of D. The maximum yield OM can be measured by periodic harvesting in a field with a very high plant density. The yield of a plant growing alone enables calculation of B as it equals $B \times OM$.

Sometimes it is more practical to work with rows rather than individual plants. The plant density is then expressed as number of rows per meter, and B as meter per row. This was done with two barley species Alasjmoen (1) and NHT (2). B and OM were found to be dependent on time as follows:

FUNCTION B1TB = (0.,0.), (30.,.03), (35.,.05), (40.,.09), (45.,.16), ...
(50.,.26), (55.,.38), (60.,.58), (65.,.88), (70.,1.02)

FUNCTION B2TB = (0.,0.), (30.,.07), (35.,.13), (40.,.18), (45.,.28), ...
(50.,.51), (55.,.82), (60.,1.30), (65.,1.75), (70.,1.94)

FUNCTION OM1TB= (0.,0.), (70.,0.56)

FUNCTION OM2TB= (0.,0.), (70.,0.56)

In these tables the independent variable is given in days after emergence, the dependent variable in the OM tables in kg dry matter/m² and in the B-tables in meter/row. The three dots at the end of a card mean that the expression or table is continued on the next card.

Question: Make acceptable that OM increases linearly with time and that the rate is the same for both species.

Employing formula (1) calculate the yield in kg/m² of a crop Alasjmoen and NHT sown at row distances of 10, 25 and 100 cm and harvested at 15, 30, 45 and 60 days after emergence.

Plot the results against density in row/m.

Calculate for both species the yield of a row per meter length sown wide apart from other rows.

Although the quantity B has the dimension space per plant or per row it can only be measured indirectly through formula (1). It is well possible that at 50 days a solitary row of the race NHT is not 51 cm wide. To obtain the right answer in the calculation, an effective width of 51 cm must be used, however.

The ratio O/D is the yield in kg per single plant or row. The ratio O/OM has an upper limit of 1 and is called the relative space occupied by the crop. It will be indicated by RO.

Question: The space occupied by a solitary plant is $\lim(RO/D)$ for D approaching zero. What is this limit?

By differentiating the expression for RO: $RO = \frac{BxD}{BxD+1}$

to time and eliminating D a differential equation in a well known form is arrived at:

$$\frac{dRO}{dt} = \frac{dB}{dt} \times \frac{1}{B} \times RO \times (1.-RO)$$

Question: Execute this differentiation.

This equation is very similar to the one for the logistic growth. The maximum amount here is one (R0 cannot exceed 1) and the relative growth rate is $\frac{dB}{dt} \frac{1}{B}$. The main difference is that this is no constant at all.

Question: In what situation is $\frac{dB}{dt} \frac{1}{B}$ a constant? What are the consequences?

The factor (1.-R0) reduces growth under influence of the occupied space. R0 need not be the space of the species itself, but may also be the space occupied by a competitor. This is true if the plants do not distinguish between the presence of plants of the same species or of the competitor.

$$\frac{dR01}{dt} = \frac{dB1/dt}{B1} \times R01 \times (1-SR0)$$

$$\frac{dR02}{dt} = \frac{dB2/dt}{B2} \times R02 \times (1-SR0)$$

$$SR0 = R01 + R02$$

These differential equations are the basis for a simulation program for the growth of two species in mixed culture. A possible formulation is:

TITLE COMPETITION BETWEEN TWO BARLEY SPECIES

x DAY IS NUMBER OF DAYS AFTER EMERGENCE

DAY = START + TIME

PARAMETER START = 10.

The asterix at the beginning of card means that the card contains comments only important for the reader and not for the computer.

x SUM OF THE RELATIVE SPACES

SR0 = R01 + R02

x GROWTH OF THE FIRST SPECIES

R01 = INTGRL (R01I,(DB1/B1)xR01x(1.-SR0))

B1 = AFGEN (B1TB, DAY).

DB1 = DERIV (DB1I,B1)

By this function the first derivative of the argument, here B1, is obtained with respect to time. The initial value at time zero must be given.

```
01 = R01xAFGEN (OM1TB,DAY)
x GROWTH OF THE SECOND SPECIES
R02 = INTGRL (R02I,(DB2/B2)xR02x(1.-SRO))
B2 = AFGEN (B2TB,DAY)
DB2 = DERIV (DB2I,B2)
02 = R02xAFGEN (OM2TB,DAY)
TIMER FINTIM = 60., PRDEL = 5., OUTDEL = 2.
PRINT R01, R02, 01, 02
PRTPLT R01, R02
INCON R01I =          ,R02I =          ,DB1I =          ,DB2I =
END
STOP
```

In this way every five days the values of the specified variables will be printed, and a plot be made of R01 and R02. To obtain a well-dimensioned graph OUTDEL should be between 1/25 and 1/50 of FINTIM; The functions to be used for both species are those for the monoculture that have been given before.

Question: What would have been the time interval DELT, if method RECT would have been used? Follow the procedure given in the paragraph on the time constant.

Question: Calculate the correct values of the initial constants, assuming that the two species are sown alternately in rows 25 cm apart. Execute the simulation with these values. Compare graphically the results of this program with the following results of the real experiment.

Days after emergence	30	40	50	60	70
Alasjmoen	.0100	.0250	.0600	.1050	.1250 kg/m ²
NHT	.0250	.1100	.1750	.2800	.3650 kg/m ²

Try to explain in words why NHT gains at the expense of Alasjmoen.

To avoid mistakes the initial values can better be calculated in the program itself. Therefore a reserved section can be used beginning by the

AL word INITIAL. The calculation is programmed as follows:

INITIAL

RO1I=B1/DIST1

B1=AFGEN(B1TB,START)

RO2I=B2/DIST2

B2=AFGEN(B2TB,START)

DB1I=AFGEN(B1TB,START+1.)-B1

DB2I=AFGEN(B2TB,START+1.)-B2

PARAMETER DIST1=0.5, DIST2=0.5

x DISTANCE BETWEEN ROWS OF THE SAME SPECIES IN METERS

DYNAMIC

MIC

The initial section is finished by the word DYNAMIC that indicates at the same time the beginning of the dynamic section. After this follows the program as discussed before.

Question: Which card must be removed?

Why may RO1I be set equal to B1/DIST1?

The instructions in the initial section will be executed at time zero only, and those in the dynamic section every time interval.

COMPETITION BETWEEN SEVERAL PLANT SPECIES

The program given in the last chapter can in principle be extended to simulate competition between several varieties. The only trouble is that the equations must be written for every variety. This encourages errors and does not make the program more surveyable.

In such a situation the MACRO feature may be used. In a macro a part of a process is described in general terms. Every time the macro is called for in the initial or dynamic section, CSMP writes the text using the indicated symbols, after which the whole is sorted. So a macro is not an order to compute something but to write a piece of text. When a macro is called for ten times, the text will be written ten times.

The macro for the growth of a crop in competition reads as follows:
MACRO O, RO=GROWTH (BTB,OMTB,SRO,DAY,DBI)
RO=INTGRL (ROI,(DB/B)xROx(1.-SRO))
B=AFGEN (BTB,DAY)
DB=DERIV (DBI,B)
O=ROxAFGEN (OMTB,DAY)
ENDMAC

In the first line, the macro declaration, is indicated which variables, functions and constants, that come out of the rest of the CSMP program, determine the variables O and RO. The ENDMAC card indicates the end of the macro. Within the macro GROWTH the equations of the last chapter are given under omission of the numbers 1 and 2.

Likewise, the relations of the initial section can be represented in a macro:

MACRO ROI,DBI=BEGIN (BTB,START,DIST)
ROI=B/DIST
B=AFGEN (BTB,START)
DBI=AFGEN (BTB,START+1.)-B
ENDMAC

The complete program for two varieties reads now:

TITLE COMPETITION OF TWO VARIETIES

At this place, directly after the title the macro definitions are inserted:

INITIAL

RO1I,DB1I=BEGIN (B1TB,START,DIST1)

RO2I,DB2I=BEGIN (B2TB,START,DIST2)

DYNAMIC

```
01,RO1=GROWTH (B1TB,OM1TB,SRO,DAY,RO1I,DB1I)
02,RO2=INTGRL (B2TB,OM2TB,SRO,DAY,RO2I,DB2I)
DAY=START+TIME
SRO=RO1+RO2
PARAMETER          DIST1=          ,DIST2=
FUNCTION (function definition as before)
TIMER FINTIM=70.,PRDEL=5.
PRINT 01,02
END
STOP
```

The program is now better readable.

Question: Write a simulation program for the growth of three varieties.

Assume that the maximal yield OM of the third variety grows as fast as for the other ones, but that the space per plant B reaches a maximum after 50 days and then stays constant. The varieties are sown in a ratio 1:1:1.

Compare the relative yield of the second and the third variety and explain the difference. Discuss also differences in relative yield RO of the first and second species compared with the previous programs.

GROWTH AND COMPETITION IN PARAMECIA

Paramecia are protozoa; unicellular organisms that live in water and feed on bacteria. Gause described an experiment with them in "The struggle for existence" in 1934, which will be explained by means of a simulation program. The species *Paramecium caudatum* and *Paramecium aurelia* are used in this experiment. They were grown in a test tube with 5 cm³ of Oosterhout's balanced physiological solution, buffered at pH 8.0. The medium was changed daily after centrifugation to separate the protozoa from the liquid. In this way accumulation of waste products was prevented. A standardized quantity of bacteria was added to the new solution as the daily food. Just before centrifugation the medium was carefully stirred and one tenth of the liquid was taken out, in which the number of protozoa were counted. At the beginning of every next day the number of protozoa was about 0.9 times the number at the end of the day before. Two series of experiments were executed, Gause speaks of the one loop experiment in case one standardized loop of bacteria was given, and of the half loop experiment in case a half loop was given daily. In both series the species were cultivated in monoculture and in mixed culture. The results are given in table 2 (next page). The given numbers are the numbers Gause counted in his sample of one tenth of the total volume.

Question: Plot the results. What might be the reason for the large scatter?

In the monoculture the population reaches a maximum in due course and stays there, just as in the case of the yeast. In the latter growth ceased due to the accumulated waste products. That cannot be the reason here in view of the careful centrifugation. The most probable reason is that the daily amount of food is limiting. In the equilibrium situation the food given is just sufficient to maintain the protozoa population and to replace the ten percent that is removed daily.

In the mixed culture one of the species vanished and the other survived at the same maximum level as in the monoculture. This fact must be explained, preferably with data derived from the monoculture only.

Assumptions on growth and death

Growth is proportional to the consumption of food. There is a constant ratio between the number of the newly grown protozoa and the amount of food

Table 2. Numbers as sampled by Gause.

Day of the experiment	Monoculture				Mixed culture			
	P.aurelia 0.5 cm ³		P.caudatum 0.5 cm ³		P.aurelia 0.5 cm ³		P.caudatum 0.5 cm ³	
	one loop	half loop	one loop	half loop	one loop	half loop	one loop	half loop
0	2	2	2	2	2	2	2	2
1	6	3	6	5	10	4	5	8
2	24	29	31	22	29	29	15	20
3	75	92	46	16	68	66	32	25
4	182	173	76	39	144	141	52	24
5	264	210	115	52	164	162	40	--
6	318	210	118	54	168	219	32	--
7	373	240	140	47	248	153	36	--
8	396	--	125	50	240	162	40	21
9	443	--	137	76	--	150	32	15
10	454	240	162	69	281	175	20	12
11	420	219	124	51	--	260	30	9
12	438	255	135	57	300	276	12	12
13	492	252	133	70	--	285	16	6
14	468	270	110	53	--	225	20	9
15	400	240	113	59	260	222	12	3
16	472	249	127	57	294	220	9	0

consumed. This ratio is conversion factor with the dimension number of protozoa per loop. It is assumed that there is a natural death rate besides the daily removal of one tenth of the population. This natural death rate is proportional to the population size with a ratio, the relative death rate, that is independent of the density. The rate of food consumption that determines the growth rate is proportional to the number of protozoa, to the density of the food and to the rate at which the protozoa search the water. Moreover there is an upper limit to the rate of food consumption, the maximal rate of digestion of food per animal.

Question: Determine the dimensions of the rates and factors mentioned above. The basic units are day, loop, protozo, volume of the test tube.

The program

As previously done in the program for the competition between plants, the dynamics of one species will be described in a MACRO. For each species the MACRO is invoked with the appropriate names. It has to be decided upon what names should appear on the MACRO declaration card. The output variables are the number of protozoa H, the rate of food consumption CNSN, and the size of the sample SPLE. The input variables are the rate of searching water RSW, the conversion factor CONVF, the maximum rate of digestion MRDIG, the relative death rate RDR and the initial population HI. The moment of sampling FDTIME and the amount of food FOOD are the same for both species so that they need not be included in the MACRO declaration.

The MACRO definition card now reads:

MACRO H,CNSN,SPLE=GROWTH (RSW,CONVF,MRDIG,RDR,HI)

The description of the structure of the MACRO follows:

H=INTGRL (HI,SN)

The rate of change SN is given by the difference between the growth rate GRSN and the rate of sampling RSAM

SN=GRSN-RSAM

The growth rate is the gross rate minus the "natural" death rate.

GRSN=CNSN \times CONVF-RDR \times H

In calculating CNSN saturation with food must be accounted for:

CNSN=H \times AMIN1 (MRDIG,RSW \times FOOD)

This function AMINI takes the smaller value of the arguments. If there is much food CNSN equals HxMRDIG, the upper limit. Below that limit the rate of food consumption is proportional to the amount of food and the rate of searching the water.

Question: Plot the relation between the amount of food FOOD and the growth rate GRSN for arbitrary values of the parameters. For which value of FOOD is the growth rate zero?

The calculation of the rate of sampling RSAM raises some difficulties. Up to this point only continuous processes have been considered. The sampling, however, occurs only once a day and is zero for the rest of the time. To let the sampled quantity vanish in one time step DELT, the rate of sampling must be defined as the size of the sample divided by DELT.

$$RSAM=SPLE/DELT$$

The size of the sample is

$$SPLE=FDTIME \times 0.1 \times H$$

FDTIME is defined in the main program outside the MACRO. It has the value 1 at the end of the day and is zero for the rest of the time. The MACRO is terminated with:

ENDMAC

In the main program the MACRO is called for by:

HA,CNSNA,SPLEA=GROWTH (RSWA,CONVFA,MRDIGA,RDRA,HIA)

HC,CNSNC,SPLEC=GROWTH (RSWC,CONVFC,MRDIGC,RDRC,HIC)

Then FDTIME is defined with:

$$FDTIME=IMPULS (1.,1.)$$

This function has the value 1. at the moment indicated by the first argument, and subsequently at moments with the second argument as intervals. It is zero for the rest of the time. This function is also used to replenish the daily food.

$$FEED=FDTIME \times (L-FOOD)/DELT$$

$$FOOD=INTGRL(L,FEED-CNSNA-CNSNC)$$

L is the amount of food given and is either 1. or 0.5.

Due to the discontinuity it is necessary to employ the rectangular method of integration.

METHOD RECT

TIMER FINTIM=16.,DELT=0.01,OUTDEL=0.1

For the purpose of comparing with Gause's data it is sufficient to have the sample size printed once a day. A simulation program enables a more thorough analysis by tracing more variables more frequently.

To complete the program the values of the properties must be given on parameter cards. Approximative values can be found by simplified calculations. More exact values can be found by repeated simulation runs, employing some criterion for the goodness of fit. Because there are 4 values to be adapted independently, it is necessary to know the order of magnitude beforehand.

Question: Why?

First estimation of the parameters

Gause observed that the liquid became transparent soon after the food was given, once the size of the population was not far from maximal. He concluded that the protozoa consumed all the bacteria, which was not the case in the beginning of the experiment, even at the end of the day. Once all the bacteria are eaten at the end of the day, it is in the monoculture of little importance how fast it happens. The food consumed and used for growth is anyway equal to the food given.

In the monoculture the maximum size of the population hardly depends on the rate of food consumption, or on the maximum rate of digestion MRDIG and the rate of searching the water RSW. It mainly depends on the food conversion factor CONVF and the relative death rate RDR. During one day about $H_x(RDR+0.1)$ dies and the newly grown amount is $CONVF \times L$. In the equilibrium situation these quantities are equal, so that

$$CONVF \times L = H_{eq} \times (RDR + 0.1)$$

This equation contains two unknowns: CONVF and RDR so that another equation is necessary to estimate their values. This equation is obtained at for instance the moment that three fourth of the maximum population size is reached. The growth rate GR may be estimated from the graph of the data and is also equal to:

$$GR = CONVF \times L - 0.75 \times H_{eq} \times (RDR + 0.1)$$

Combining both equations allows a first estimate of CONVF and RDR.

Question: Calculate these first estimates for both species and compare the results with the values given below.

From repeated simulation runs it appears that CONVFA is 3000 and CONVFB is 750, whereas RDR for both species is 0.45 day^{-1} .

The rate of searching the water RSW and the maximum rate of digestion MRDIG are estimated from the dynamics of the populations at the beginning of the experiment. During the early stages the number of protozoa is so small that the concentration of bacteria stays practically the same during the whole day. As can be seen from the data of the 0.5 and 1 loop experiment with P.aurelia the growth rate is about the same and this means that the maximum digestion rate is reached at the 0.5 loop concentration or that at this level

$$\text{MRDIG} = 0.5 \times \text{RSW}$$

When now, the availability of food constantly exceeds the saturation level, the relative growth rate, which can be read in the graphs is about

$$\text{RGR} = \text{MRDIG} \times \text{CONVF} - \text{RDR} - 0.1$$

These two equations allow a first estimate of RGR and MRDIG.

Question: Make these first estimates. Try to find as many reasons as possible why these first estimates may be very inaccurate.

The values found by simulation are considerably different and are listed below. It must be admitted that the data of the monoculture only were not sufficient to derive the values of these parameters, the data of the mixed cultures had to be used as well. This is not due to a fundamental reason but to the large scatter of the data. Moreover differences between species show up much more pronounced in mixed cultures.

Question: Why is this the case?

	<u>P.aurelia</u>	<u>P.caudatum</u>
Relative death rate RDR	0.45	0.45 day^{-1}
Conversion factor CONVF	3000.	$750. \text{ animal loop}^{-1}$
Saturation level MRDIG	$.56 \cdot 10^{-3}$	$2.24 \cdot 10^{-3} \text{ loop animal}^{-1} \text{ day}^{-1}$
Rate of searching water RSW	.006	$.006 \text{ volume animal}^{-1} \text{ day}^{-1}$

The relative death rate RDR and the rate of searching water RSW turn out to be the same for both species. MRDIG is assumed to be four times larger for P.caudatum as for P.aurelia.

Question: What could be a reason for this assumption?

Execute a sensitivity analysis for the parameters RSW, RDR, MRDIG and CONVF. For that purpose simulate the growth of the species for the parameter values given in the table and once for each parameter value separately 10 % larger.

Determine the influence of these changes on outputs you assume to be important. Do this for both the mono- and the mixed culture.

Which value would you have given to DELT?

Stochastic phenomena

Now the reasons for the scatter in Gause's data are considered. There are two stochastic phenomena that are accessible to calculation: the sampling and the death process. Up to this point we assumed that Gause was able to sample exactly one tenth of the population, and that during any sufficiently small time interval the death rate was exactly $H \times RDR$. Neither assumption is right.

As far as the death process is concerned, the rate is only an average value. Sometimes more protozoa die during a time interval than $H \times RDR \times DELT$, sometimes fewer. To simulate this the number of deaths must be drawn out of a probability distribution around the average death during a time interval. The normal distribution and the Poisson distribution may be used for this purpose. When each animal has an equal, constant and non-age dependent probability of dying it is the Poisson distribution that must be used.

Question: Give a good reason why the normal distribution function cannot be used.

The standard deviation of a Poisson distribution is entirely determined by its expectation value. Thus knowing the average a random choice of the actual number can be made according to the probability distribution. The random choice was programmed in a MACRO. In the main program the invocation to calculate the death rate reads:

```
AD = HxRDRxDELTA  
DRD = POISS(P,AD,DELTA)  
DR = DRD/DELTA
```

The average number of deaths during a time interval DELTA is HxRDRxDELTA which number is an input to the random selection of the number DRD that actually dies. To find the death rate it must be divided by DELTA. DR is the death rate and is used instead of HxRDR in the previous program. The first argument in the invocation of the MACRO is an odd, integer number required to start some pseudo random series, used for the random selection. The last argument indicates that random selection must be executed every DELTA.

Now the content of the MACRO will be given, but this is only understandable for those who know some FORTRAN.

```
MACRO N=POISS(P,MEAN,PERIOD)
```

Because DO-loops and IF-statements are used the statements must be in computational order. This is indicated by the card:

```
PROCEDURAL
```

If the time is not equal to n times PERIOD the drawing need not be executed and N equals zero:

```
N=0
```

```
IF(IMPULS(0.,PERIOD).LT.0.5) GO TO 1
```

whereby 1 is a CONTINUE statement at the end of the MACRO. In case the expectation value is larger than 25 the Poisson distribution is sufficiently approximated by a Gauss distribution with a standard deviation equal to the square root of the average.

```
IF(MEAN.LT.25) GO TO 2
```

```
N=GAUSS(P,MEAN,SQRT(MEAN))
```

```
GO TO 1
```

```
2 CONTINUE
```

The Gauss function is a CSMP function that executes the random choice out of a normal distribution. P can be any odd, integer number. The second and the third argument represent the average and the standard deviation respectively.

Below a number of 25 the deviation between the Poisson distribution and the Gauss distribution becomes too large. To execute the selection from the Poisson distribution a number is first drawn between 0 and 1 according to a standard uniform probability function. This is done by a

CSMP function:

```
LOT=RNDGEN(P)
```

P is again the odd, integer number.

Then this number is used to read the output from a cumulative Poisson distribution function. The cumulative Poisson distribution is obtained by a series development (Moroney, 1951). The probability of a number to be smaller than or equal to 0, 1, 2, 3..... is given by $e^{-z}(1 + z/1! + z^2/2! + z^3/3! \dots\dots\dots)$ where z is its average.

Question: Show that this series adds up to 1. Check if the following procedure is right.

SOM=1.
PROD=1.
EMINZ=EXP(-MEAN)
DO 4 J=1,100
IF (LOT.GT.SOMxEMINZ) GO TO 3
N=J-1
GO TO 1
3 CONTINUE
PROD=PRODxMEAN/J
SOM=SOM+PROD
4 CONTINUEZ
The MACRO is concluded by
1 CONTINUE
ENDMAC

Question: Program also a random sampling process by taking 1/10 of the solution at the beginning of every day.
Simulate now the growth of the species in a mixed culture for:
a. a random death process
b. a random sampling process
What is the main reason of the scatter in Gause's observations?

SIMULATION OF AGE DEPENDENT PROCESSES

When simulating the growth of protozoa, a constant relative death rate was assumed. This implies that death is independent of age. A similar situation might arise in a human society when most deaths are caused by traffic accidents. An entirely random character of death is assumed by using the Poisson or Gauss distribution to describe the process in small numbers.

The other extreme situation is that death occurs only and completely at a fixed age. In that case the growth rate of a population can be represented by:

$$dH/dt = RBR \times H - RBR \times H_{t-L}$$

Here, L is the duration of life of the organisms, H_{t-L} is the amount present at a time L earlier and RBR is the relative birth rate. A problem is that the age distribution at time zero must be described. If at time zero all organisms have the age 0.5L the initial population will die at time 0.5L. For this case the simulation program could read:

$$H = \text{INTGRL}(HI, GR)$$

$$GR = BR - DR$$

$$BR = H \times RBR$$

$$DR = \text{DELAY}(48, L, BR) + HI \times \text{IMPULS}(0.5 \times L, \text{FINTIM}) / \text{DELT}$$

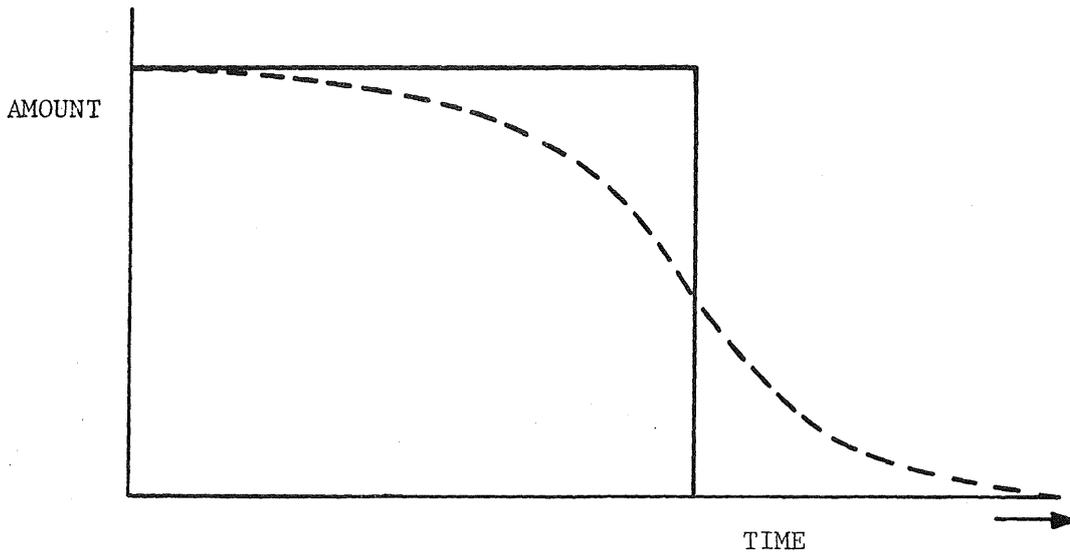
PARAMETER HI=1., RBR=0.2, L=24.

In this way the METHOD RECT must be used, because of the division by DELT. Another possibility is assuming that HI has a uniform age distribution.

$$DR = \text{INSW}(\text{TIME} - L, HI/L, \text{DELAY}(48, L, BR))$$

If the first argument of the INSW function is negative the output equals the rate HI/L, if it is positive the DELAY function will be used. Because there is no division by DELT anymore METHOD RKS may be used. The DELAY function delays its last argument over a period mentioned in the second argument. The integer number 48 means that 48 values are sampled during this period. The larger this number, the better the resolution power. The interval between two stored values must be about one tenth of the time constant of the concerned variable. The output of the DELAY function is zero before the delay time is reached.

The DELAY function gives a delay as given by the solid line in the figure below as a response to a stepwise change at time zero.



There is also a method to disperse the delay, as illustrated by the dashed line. This method is applied if some animals die earlier and some later than the average. It is called the method of the dispersed delay.

A 24 hour's period is divided in for instance 4 parts of 6 hours each. The amount in each of these groups is conserved in an integral. In each group there is an average residence time of 6 hours. This can be done as follows:

$$\begin{aligned} H_1 &= \text{INTGRL}(H_1, BR - H_1/6.) \\ H_2 &= \text{INTGRL}(0., (H_1 - H_2)/6.) \\ H_3 &= \text{INTGRL}(0., (H_2 - H_3)/6.) \\ H_4 &= \text{INTGRL}(0., (H_3 - H_4)/6.) \end{aligned}$$

 Question: Show that the average residence time in each group is 6 hours.

For this purpose set BR at zero and keep track of H1 in time.

The average residence time is

$$\frac{1}{H_1} \int_0^{H_1} L(H) dH$$

Here L is the actual residence time of dH.

 To calculate BR the total population H1+H2+H3+H4 must be multiplied by RBR.

Question: Execute the simulation with the DELAY function, and with the method of the dispersed delay, taking N=1, 2, 4 and 24 resp. Which simple, previously treated, situation is obtained when N=1? Plot the total population against time. Make also a simulation with RBR=0. Take FINTIM=48. and HI=100.

If N classes are employed to cover a lifetime L, the average residence time in each class is L/N. The number N is determined by the standard deviation of the lifetime. It can be derived that N is equal to

$$N=(L/\sigma)^2$$

in which σ is the standard deviation of the duration of life L. If the standard deviation is less than one fifth of the total lifetime, the number of classes must be so large that the computation time becomes prohibitive. Then it is better to use a so called boxcartrain.

The difference with the previous method is that there is not a continuous flow from one class to the next, but that once in every residence period all contents shift one place. To program this we need the IMPULS function, as in the sampling of the paramecia:

```
PUSH=IMPULS(1.,1.)
H1=INTGRL(HI1,BR - PUSHxH1/DELT)
H2=INTGRL(HI2, PUSHx(H1-H2)/DELT)
.
.
.
H24=INTGRL(HI24, PUSHx(H23-H24)/DELT)
```

The birth rate BR can be the total birth rate times the relative birth rate. It is also possible to introduce an age dependent fecundity by

$$BR=F1xH1+F2xH2+F3xH3+ + F24xH24$$

Here F is the fecundity of each class.

Programmed in this way death only occurs at the moment that all classes are passed, which is similar to the result of the DELAY function. It is also possible to introduce death in other age classes. The extended equations are:

```
H1=INTGRL(HI1,BR-RDR1xH1-PUSHxH1/DELT)
H2=INTGRL(HI2,-RDR2xH2+PUSHx(H1-H2)/DELT)
.
.
H24=INTGRL(HI24,-RDR24xH24+PUSHx(H23-H24)/DELT)
```

Both F and RDR may be functions of external conditions. A survey of the three discussed delay methods follows:

	DELAY function	dispersed delay	boxcartrain
Integration method	any	any	RECT
Course of the delay	sudden	gradual, dependent on the number of classes	sudden
Flexibility	poor	good	moderate

Question: Write a simulation program for the growth of the population in The Netherlands, based on data of 1968.

Population on December 31st 1968:

Age group	Number of men	Number of women
0 - 4	610736	581772
5 - 9	613494	586599
10 - 14	580261	552807
15 - 19	569490	542467
20 - 24	582563	553889
25 - 29	452267	419526
30 - 34	404872	379516
35 - 39	393384	380914
40 - 44	371039	377987
45 - 49	362217	376413
50 - 54	314031	330267
55 - 59	297444	323390
60 - 64	262211	297817
65 - 74	368998	452313
75 -	199953	259709
Total	6383000	6415000

Death per 1000 men and women per year

Age	Men	Women
0 - 1	15,6	11,4
1 - 4	1,0	0,7
5 - 14	0,5	0,3
15 - 29	1,0	0,4
30 - 44	1,8	1,2
45 - 64	10,9	5,5
65 - 74	41,7	23,9
75 -	117,6	95,7

Number of births in 1968 according to the age of the mother:

<15	15-19	20-24	25-29	30-34	35-39	40-44	45-49	>49
0	11825	75599	78787	43055	21039	6246	564	0

Ratio boys/girls 1.048

Consider carefully the choice of the age classes and the delay method. Ask for the composition of the population and the total number every five years. Simulate over 50 years. Consider the influence of postponing the first child by shifting the effective fecundity to a later age group.

The matrix method

This method is not a simulation method, but can be a useful additional tool. It provides the equilibrium age class distribution and the corresponding relative growth rate of a population. The principles of the method will be given.

In case that DELT in a program with the boxcartrain equals the average residence time in each class, the contents are shifted one place, diminished by the amount died, at every time interval. If the relative death and birth rate do not change in time (they may be dependent on age) a matrix method, introduced by Leslie in 1945, may be applied. Let the contents of the classes be the elements of a vector. If there are ten age groups, the vector is tendimensional. Then the age class distribution one time interval later is found by multiplying this vector by a matrix:

$$\begin{array}{|c|} \hline H1 \\ \hline \\ \hline \\ \hline \\ \hline \\ \hline \\ \hline H10 \\ \hline \end{array} \Big|_{t+DELTA} = \begin{array}{|c|} \hline F1 & F2 & \dots & \dots & \dots & F10 \\ \hline S1 & 0 & \dots & \dots & \dots & 0 \\ \hline 0 & S2 & 0 & & & \\ \hline & 0 & \diagdown & 0 & & \\ \hline & & 0 & \diagdown & 0 & \\ \hline & & & 0 & \diagdown & 0 \\ \hline & & & & 0 & \diagdown & 0 \\ \hline 0 & \dots & \dots & 0 & S9 & 0 \\ \hline \end{array} \times \begin{array}{|c|} \hline H1 \\ \hline \\ \hline \\ \hline \\ \hline \\ \hline \\ \hline H10 \\ \hline \end{array} \Big|_t$$

At the left hand side the vector is written as it is at time t+DELTA. It is found by multiplication of the vector at time t by the matrix. In the matrix, F1 is the relative number of births per time step in class I and S1 is the fraction of class I that passes to class I+1; in other words 1 minus the fraction that dies during a time interval.

From matrix algebra it is known that repeated multiplication by a matrix gives rise to a vector that has a constant relative composition in due course. Each time the length of the vector is increasing by a constant factor. These are called the dominant eigenvector and the corresponding eigenvalue. Standard programs are available to derive them from a matrix. The result of these mathematics agrees with the experience that any population approaches a stable age distribution and a constant relative growth rate in due course, provided the relative death and birth rates remain constant. This matrix method is usually cheap in terms of computation time, but requires more mathematical knowledge and is less flexible than simulation. It only provides information on the final stage. For that reason the method is useful if a certain distribution of death and birth rates will end up in a stable population or not. It will also show how the relative growth rate will change as influenced by a shift in births to another age group.

 Question: Is it possible to solve the questions on the growth of the Dutch population by means of the matrix method?

 Facultative question: Calculate the age distribution in equilibrium and the corresponding relative growth rate of the Dutch population by means of the matrix method on basis of the 1968 data.

FLOW OF HEAT IN SOIL

There is a great deal of similarity in simulating the age distribution of a population and physical diffusion processes in space and time. These similarities will be illustrated by a simulation program for the course of temperature in the soil as influenced by a cyclic variation of temperature at the surface.

To this end the soil is divided in horizontal layers, thus assuming that the only gradients are vertical. The thermal condition of the layers is characterized by heat content rather than by temperature.

In defining the initial condition the temperatures are used INITIAL.

$$HCI=TCOM \times VHCAP \times TEMPI$$

PARAM TCOM=0.02,VHCAP=1.E6,TEMPI=20.

x UNITS IN KG,M,S SYSTEM

Here TCOM is the thickness of the layers, VHCAP the volumetric heat capacity of the soil and TEMPI the initial temperature. The basic units are in the kilogram, meter, second system.

Derived units here are joule for energy, and joule per m³ per °C for the heat capacity.

The use of a consistent system of units may result in cumbersome numbers, but this is preferable above adaption of dimensions. It is a troublesome source of errors if the dimensions do not agree. Sometimes it is advantageous, however, to have input or output converted into a better known dimension. For instance it is often desirable to have time tabulated in hours besides seconds, like in this programme. That must be done in the dynamic section:

DYNAMIC

$$HOUR=AMOD(TIME/3600.,24.)$$

Time in hours is found by dividing the number of seconds by 3600. The AMOD function takes care that it is reset to zero every 24 hours. The integration is:

$$HCI=INTGRL(HCI,FL1-FL2)$$

.

$$HC15=INTGRL(HCI,FL15-FL16)$$

Question: What direction is positive for the flow of heat FL?

The driving force for the flow of heat is the temperature difference between adjacent layers.

$$FL1=(TMPS-TMP1) \times COND / (0.5 \times TCOM)$$

$$FL2=(TMP1-TMP2) \times COND / TCOM$$

.

.

$$FL15=(TMP14-TMP15) \times COND / TCOM$$

$$FL16=0.$$

The last and the first equation formulate the boundary conditions. The temperature at the surface TMPS will be given by:

$$TMPS=TAV+TAMPL \times \sin(6.2832 \times TIME / 86400.)$$

PARAM TAV=20.,TAMPL=10.

The temperature of the layers is calculated by:

$$TMP1=HC1 / (TCOM \times VHCAP)$$

.

$$TMP15=HC15 / (TCOM \times VHCAP)$$

Question: List the dimensions of all variables and parameters.

The timer card and the output cards may be:

PRINT TMPS,FL1,HOUR,TMP1,TMP2,TMP3,TMP4,,TMP15

TIMER FINTIM=172800.,PRDEL=1800.

Question: Compare the variation in temperature at a depth of 5 cm for a soil column of 30 cm and of 15 cm. Explain the difference.

Question: Express the value of the time constant in TCOM,COND,VHCAP.

What is the influence of doubling the number of layers on the computation time?

APPENDIX

In the last program and also in the program on population growth a list of similar equations had to be written. There are three methods available to reduce the amount of punching work involved.

a. Application of the MACRO feature

For the type of programs where layers or classes are employed the MACRO feature does not give much reduction of the work.

Question: Try this out.

b. Application of the INDEX feature

This feature is available on the computer in Nijmegen, Wageningen and Jerusalem. It enables to write a list of similar equations at once:

HC1=INTGRL(HCI,FL1-FL2)

.

.

HC15=INTGRL(HCI,FL15-FL16)

can be substituted by:

INDEX HC"1,15"=INTGRL(HCI,FL"1,15"-FL"2,16")

This INDEX-label generates the statements above. The pair of numbers following a variable indicates the first and the last number of the list. Within one statement a variable may be identified by numbers that are shifted, like FL in the statements above. The initial heat content HCI is the same for all layers so that it need not be numbered. The flow of heat may be specified by:

FL1=(TMPS-TMP1)xCOND/(0.5xTCOM)

INDEX FL"2,15"=(TMP"1,14"-TMP"2,15")xCOND/TCOM

FL16=0.

The first and the last flow have different formulations so that they must be defined separately. The INDEX feature can be combined with a convenient way of data-input. The problem of the growth of the Dutch population can be taken as an example. There are 15 age classes of men and women. The initial number of women in the 15 classes can be given as:

STOCK WOI(1-15)=581772.,586599.,.....,259709.

This is equivalent to:

PARAM WOI1=581772.,WOI2=586599.,.....,WOI15=259709.

c. The use of FORTRAN

Generally this is more complicated than the use of the INDEX method, but it cannot always be replaced by it, as in the case of the programming of the Poisson distribution in the growth of the paramecia.

FORTRAN can be intermixed freely with CSMP provided it is preceded by a NOSORT card that prevents the statements from being sorted by the computer. At the end of the FORTRAN section one should put a SORT card, that reactivates the sorting routine.

In this construction the sorting routine cannot sort the statements from both sides of a NOSORT section into one package.

That can be accomplished by using a PROCEDURE instead of a NOSORT section. In its use it must be stated what are the input and output variables that must be subjected to sorting.

```
PROCEDURE A,B,C=NAME(V,W,X,Y,Z)
computation
ENDPRO
```

The FORTRAN section is then concluded by ENDPRO. In this case there are three input variables and five output variables. Now the procedure is sorted as one block, behind the computation of the inputs and before the use of the output variables.

Once it is understood how the INDEX feature works, there are no problems with the DO-loop.

INDEX RH"2,24"=H"1,23"-H"2,24"

can be written in FORTRAN as:

```
DO 50 I=2,24
RH(I)=H(I-1)-H(I)
```

```
50 CONTINUE
```

I is the index of the array which runs from 2 to 24. It is a fixed or integer variable as specified on a FIXED card:

```
FIXED I
```

The number 50 is here the number of the DO-loop. It can be freely chosen by the programmer under the restriction it has not been used somewhere else as an identifier. Each DO-loop is started and terminated by the same number.

RH and H are arrays. The compiler is warned to reserve the required space in the memory by:

```
STORAGE RH(24),H(24)
```

Any number between the brackets of an array is integer so that it should be written without a decimal point.

DO-loops may be nested: within the mentioned DO-loop 50 another DO-loop may be put with its own number.

Sometimes a choice must be made between two computative procedures. This happened in the computation of the size of the sample out of the paramecia. The value of MEAN determined whether the GAUSS distribution or the Poisson distribution must be applied.

```
IF(MEAN.LT.25.) GO TO 2
```

```
  Computation of the Gauss distribution
```

```
  GO TO 1
```

```
2 CONTINUE
```

```
  Computation of the Poisson distribution
```

```
1 CONTINUE
```

If the condition MEAN less than 25. is satisfied, the computation leaps to 2. The expression within the parenthesis behind the word IF has to be written without blanks, and LT must be surrounded by two points.

The standard expressions for comparison are:

```
LT   less than
```

```
GT   greater than
```

```
EQ   equal to
```

```
NEQ  not equal to
```

Question: Write a program for the choice between Gauss and Poisson that leads to the same result, but employing the expression GT instead of LT.

It is not possible to have subscribed variables printed by means of the PRINT statement, but to request output by means of FORTRAN is cumbersome. Therefore, it is most convenient to convert the subscribed variables to regular ones by:

```
H1=H(1)
```

```
H2=H(2)
```

```
etc.
```

This can be accomplished by:

```
INDEX H"1,24"=H("1,24")
```

The obtained variables can be asked for on a PRINT card:

```
PRINT H1,H2, etc., H24 or by PRINT H(1-24).
```

LITERATURE

- Brennan, R.D.: Continuous System Modeling Programs, state-of-the-art and prospectus of development. In: Simulation Programming Languages, B.N.A. Uitgeversmij., Amsterdam.
- Brennan, R.D., C.T. de Wit, W.A. Williams and E.V. Quattrin: The utility of a digital simulation language for ecological modeling. *Oecologia* (Berl.), 4(1970), 113-132.
- I.B.M. S/360 Continuous System Modeling Program, Program Reference Manual.
- Forrester, J.W.: Industrial dynamics. MIT-Press, 1962.
- Franks, R.G.E.: Mathematical modelling in chemical engineering. Wiley, New York, 1967.
- Leslie, P.H.: On the use of matrices in certain population mathematics. *Biometrika*, 33(1945), 183-212.
- Moroney, M.S.: Facts from figures. Penguin books, 1951.
- Patten, B.C.: Systems Analysis and Simulation in Ecology. Academic Press, 1971.
- Usher, M.B. and M.H. Williamson; A deterministic matrix model for handling the birth, death and migration processes of spatially distributed populations. Vol. 26, no. 1, *Biometrics*, March 1970.

INDEX OF DESCRIBED LABELS AND FUNCTIONS

<u>Name</u>	<u>Page</u>
AFGEN	12
AMIN1	26
AMOD	39
DELAY	33
DELT	7
DERIV	19
DO	42
DYNAMIC	21
END	7
IF	43
IMPULS	27
INCON	7
INDEX	41
INITIAL	21
INTGRL	7
INSW	33
FINISH	11
FINTIM	7
GAUSS	31
MACRO	22
METHOD	7
NOSORT	42
PARAMETER	7
PRDEL	20
PRINT	20
PROCEDURE	42
RNDGEN	32
SORT	42
STOCK	42
STORAGE	43
STOP	7
TIMER	7
TITLE	7

There are more available. They are described in the FORTRAN and CSMP manual.