Mathematical structures for modeling pest populations

Phillip L. Shaffer

North Carolina State University

(Based on a talk given at the CIPM Modeling Workshop, Las Vegas, Nevada, 25 May 1983)

General approaches. There are three major modeling approaches - analytical, simulation, and optimization. Analytical models are those that consist of a group of equations for which a closed-form solution can be found. These models are often useful for long term analysis as for population stability or determination of steady state solutions. As such, they are not likely to be applicable to many agricultural pests, which have only a few generations in a year. Aphid and mite populations may reach a stable age distribution in one season, but still would not be in the stable age distribution for much of year (Taylor 1979). These models are often not biologically realistic because a system of equations complex enough to be realistic cannot generally be solved analytically. Thus most of these models cannot incorporate the effects of management and fluctuating environmental conditions. Familiar examples of analytical models are the Lotka-Volterra equations for predator-prey systems and the Nicholson-Bailey model of parasite-host systems.

Simulation models are composed of a system of equations which are solved numerically. Because there is no requirement for

analytical solution, these models may be as complex as necessary, and so can be biologically realistic. Time-varying effects may be incorporated, and transient systems may be modeled easily. These models often provide insight into organization and operation of modeled system, and may reproduce behavior of system over a range of conditions. Simulation models allow testing of effects of manipulation of system parameters, termed sensitivity analysis. This may be used to guide further research into the system.

Optimization models are designed to determine the values of management or environmental parameters which optimize some objective function, such as profit or yield. A variety of techniques such as linear programming, non-linear programming, and dynamic programming techniques may be used. The descriptive model on which the optimization is based may be an analytical or simulation model, with the optimization technique chosen placing constraints on model type and complexity. Optimization is perhaps everyone's goal, but it can be very difficult to determine the optimal solution from an extremely large number of alternatives given the complex equations that may be used to realistically describe a system's dynamics. Shoemaker (1981) reviews a number of optimization techniques. The relative merits of the three approaches are compared in Fig. 1, taken from Shoemaker (1980).

Choices for models. There are number of choices which must be made in the construction of any model of pest populations (Fig. 2). In general, the choices are determined by the goals of

the modeling effort, and by the biology and dynamics of the system to be modeled. Whether the model includes a single species or multiple species depends on how strong the interaction the component species of the system are. If the pest affects its host plant significantly or if the host plant varies in ways that affect the pest population dynamics, then the plant should be included in the model.

Whether age and/or stage structure should be included depends on what factors affect population dynamics. For most insects, the modeler can't ignore stages, as the biology differs so much among the stages, for example in mortality sources, host plants, susceptibility to pesticides, and habitat. Age structure is important in determining probability of transition between stages and fecundity of adults.

The choice between deterministic and stochastic models may be influenced by the relative importance of the different sources of variation: temporal or spatial variation in system inputs or driving variables; errors or variation in system parameters; and demographic stochasticity, i.e. differences among individuals in longevity and fecundity. Demographic stochasticity is often ignored in large populations, or modeled as a distribution for each attribute. For planning purposes, we would like to determine the probability distribution of outcomes. To do this it must be determined what factors vary and how variation in these influences the results. The relationship between variation in inputs or parameters and variation in outputs can be determined by Monte Carlo simulations using a stochastic model,

or by sensitivity analysis using a deterministic model. Markovian models of birth-death processes can be used to examine demographic stochasticity. For suitable systems, models using stochastic differential equations may be used to analytically determine probability distributions of the state of the system. The major problem with any truly stochastic approach is that to determine the distribution of outcomes, a large number of simulation runs are required.

Population processes are generally continuous in time and discrete in numbers, but are modeled as continuous in number and discrete in time. The choice between discrete and continuous models may be influenced by the nature of process, as by whether there are discrete or overlapping generations. If the population size is large, then approximating population size by a continuous value is not too bad an approximation. This choice is also influenced by the goal of model - an exploratory submodel may use a simulation of discrete individuals, and then be used to parameterize continuous equations.

For poikilotherms such as insects, temperature of the environment is usually the primary factor influencing developmental rate, longevity, and fecundity schedule, and must be incorporated in any meaningful model of such a population. This may be done by basing the model on chronological time, and incorporating physiological rates as explicit functions of temperature. This has the advantage that different processes, stages, or age groups may differ in their relationship with temperature. It is sometimes simpler to base the entire model on

a "physiological" time scale, such as degree-days, where all processes in the model are assumed to vary with temperature in the same way. This sacrifices realism for ease of implementation.

Implementation of models. I will now consider in more detail the implementation of a number of simulation or analytical population models which incorporate age structure or stage structure, or both (Fig. 3). I will discuss two matrix approaches to discrete models, the Leslie matrix and the Lefkovitch matrix. Simpler discrete models, based on difference equations of unstructured populations, are considered by May (1979). Continuous models may be specified as a system of differential equations, and then solved numerically by a continuous systems simulation language such as CSMP. However, these languages (to the best of my knowledge) do not handle partial differential equations, and so cannot model age-structured populations. Thus it is necessary to explicitly discretize partial differential equations, such as von Foerster equations, or to use the distributed delay approach.

The Leslie matrix, or age-projection matrix (Fig. 4) consists of two basic parts: the first row contains fecundities of each age class, while the sub-diagonal contains survivorship probabilities for each age class (Lewis 1942; Leslie 1945, 1948). The state vector of number of individuals in each age class at time t is multiplied by the matrix to obtain the vector of number at time t+1. This can also be expressed as a matrix equation, or as a series of difference equations. The "width" of an age class

is equal to the time between iterations of the model. The standard model has time-invariant fecundities and survivorships, so density-dependent effects or environmental effects are difficult to interpret. If standard matrix-handling routines are used, much computation time and storage space is wasted because the matrix is mostly zeroes. The advantages are that this approach is simple, and neatly parallels the standard demographic life table. An analytical solution can be found for growth rate when a stable age distribution has been reaches (growth rate is given by the eigenvalue and the age distribution is given by the eigenvector). Stability of the system may be examined by the values of the latent roots of the system. The Leslie matrix has been extended in a variety of ways (Usher 1972), but much of the analytical simplicity is lost with these extensions.

The Lefkovitch matrix (Fig. 5) replaces the age classes of the Leslie matrix with stages or size classes which may be of unequal age widths (Lefkovitch 1965). The top row remains fecundity, but each subdiagonal element represents the probability of moving from one stage to the next during the interval of an iteration (conditional on having been in that stage at the beginning of the interval). The diagonal elements are the probabilities of staying in the same stage in each time interval. Thus the probability of an individual dying during a time interval is equal to 1 minus the sum of the probabilities of either moving to the next stage or staying in the current stage during an interval. These probabilities act to create a distribution of times taken to pass through a certain number of

stages. As in the Leslie matrix, fecundities and probabilities are time-invariant. All individuals in a stage are assumed to be probabilistically identical, so no correlations are possible among stage-completion times. Also like the Leslie matrix, this approach is simple, and analytical solution is possible.

The distributed delay is a method of generating a distribution of stage-completion times, and has its origin in engineering (Manetsch and Park 1977). Each stage (or the entire life cycle) is divided into k discrete substages. The flow of numbers of individuals between substages is described by simple differential equations. These equations are then discretized over time and flow, and solved numerically. The series of substages results in a distribution of development times described by a gamma (or Erlang) function of order k (Fig. 6; in the figure, alpha = k). For k = 1, the distribution is exponential; as k becomes large, the distribution approaches a normal curve. This is equivalent to a stage-projection matrix with substages. While the value of k can be determined from mean and variance of distribution times, in practice k is usually chosen rather arbitrarily, as the shape of the curve also varies with k. A major problem with the distributed delay approach is that the substages do not represent biological groups; while they may be taken as approximating the fraction of development completed, mortality and fecundity must be specified in terms of the substages, and this may not be feasible or realistic. Distributed delays may be implemented using either fixed delays or time-varying delays.

Fixed delays are used by the PETE insect phenology model from Michigan State University (Welch et al. 1978). The differential equation giving the flow rate from each substage is:

$$dr_i/dt = 1/D_i \; [r_{i-1}(t) - r_i(t)]$$

where $r_i(t)$ is the flow from substage i at time t, and D_i is the delay of substage i. In this implementation, the discretized versions of this equation are iterated every degree day, thus the same developmental rate function is used for all stages of the insect modeled (i.e., degree days with the same threshold temperature). The advantages of this approach are the simplicity of the computations required, and the small number of parameters required: the overall threshold, and k and the number of degree days for development of each stage.

Time-varying delays are used in a model developed at Cornell University (IMP, by R. I. Carruthers, A. J. Sawyer, and T. Larkin). The basic differential equation for flow rate is:

$$dr_i/dt = 1/D_i \; \{r_{i-1}(t) - r_i(t)[1 + dD_i/dt]\}$$

The major difference is the addition of the derivative of the delay with respect to time. This implementation allows the use of separate non-linear developmental time functions for each stage (or a simple table look-up with interpolation, as in IMP, which can make direct use of developmental time versus temperature data). This model iterates at a fixed time interval in chronological time, but the discretization of the differential equations requires that this interval be small so that the changes in developmental rates between intervals are small. This probably would require iteration at least 12 times per day,

rather than 4 per day as implemented.

The von Foerster equations describe an age-structured population with age-dependent mortality and fecundity by two partial differential equations:

$$\partial n(a,t)/\partial t + \partial n(a,t)/\partial a = - \mu(a,t)\ n(a,t)$$

where $n(a,t)$ is number density at age a and time t, and $\mu(a,t)$ is mortality rate at age a and time t; and

$$B(t) = dN/dt\big|_{births} = n(0,t) = \int_0^\infty \lambda(a,t)\ n(a,t)\ da$$

where $B(t)$ is the birth rate at time t, $\lambda(a,t)$ is fecundity rate at age a and time t and N is the total number in population. The first equation describes the relationship between partial derivatives of number density with respect to age and time and the mortality rate; the second is a boundary condition describing the birth rate as an integral of fecundity rate times number density over age. Total number of individuals in a population is simply the integral of number density over ages. Here a is chronological age, so da/dt is always 1. The equations can be modified so that mortality and fecundity rates are functions of physiological age instead of chronological age, and the developmental rate, da/dt, is made a function of temperature T:

$$da/dt = r(T)$$

and temperature is a function of time:

$$T = \psi(t).$$

The equations are now:

$$\partial n/\partial t + r(\psi(t))\ \partial n/\partial a = - \mu(t,a)\ n(t,a)$$

$$n(0,t) = 1/r(\psi(t)) \int_0^\infty \lambda(a,t)\ n(a,t)\ da$$

Other variables describing important attributes of members of a

population, such as size or mass, can also be incorporated into the von Foerster equations; Streifer (1974) has an excellent introduction to these equations and their applications. A population can also be divided into stages, and the birth or emergence of each stage can be described as a function of the preceding stage. The loss rate of each stage must then include both death and emergence to the following stage. Curry, Feldman, and Smith (1978) developed this approach for two stages, immatures and adults. The advantages of using the von Foerster equations are that they are realistic and flexible; the cost of this is that many functions are needed to describe birth, death and emergence of each stage, and the simulation can be computationally expensive.

The von Foerster equations have been implemented in simulation models of pest populations in two ways: the finite differences method, and the iterative cohorts method. The finite differences approach has been used by Wang et al. (1977) and by Gutierrez et al. (1977). In this approach, discrete approximations are made for all equations in the model, and these equations are all iterated on the same small time step. This requires a large number of iterations. In the two papers cited, the models operated on physiological time (degree day) scales, thus forcing the approximation that all temperature-dependent aging processes occur at the same rates at any temperature. It would be possible to add maturity (physiological age) as a dimension to the model, and incorporate developmental rate as an explicit function of temperature.

The iterative cohorts method was developed by Curry, Feldman, and Smith (1978). In this approach, each stage is grouped into cohorts, i.e. individuals that were born or emerged during a single interval of time. The members of each cohort age according to a developmental rate function of temperature, and are assumed to be probabilistically identical at a given physiological age. Numbers in each stage and movement among stages are described by distributions of development times or longevity and fecundity or survival. The number of individuals in a cohort can only decline with time after it is born. The discretization of the partial differential equations is done on two time scales: the temperature-dependent functions such as developmental rate are updated frequently, e.g. once per hour, while the cohort numbers (i.e. mortality, fecundity, and emergence) are updated less frequently, e.g. once per day. The major assumption of this approach is that the transition functions depend only on physiological age, although other factors which vary along the slower time scale such as pesticide-induced mortality, can be incorporated.

I have recently developed a generalization of the iterative cohorts method which allows any number of stages, including overwintering stages. For each stage, four functions are specified, two of which are functions of temperature and two of which are functions of physiological age. The temperature dependent functions are developmental rate and production rate, which is the fraction surviving (and thus moving on to the next stage) for immatures, or the fecundity for adults. The

physiological-age dependent functions are cumulative fraction born or moved to the next stage, and cumulative fraction which have left the stage (mortality for adults, or mortality plus development to next stage for immatures). The physiological age scale is normalized so that at age one, one-half of the individuals of a stage have developed to the next stage (for immatures) or have died (for adults). Lack of data has forced the assumption that death and emergence functions for immatures are coincident (i.e. that all immatures which die do so at the time at which they would have moved to the next stage) but this is not inherent in the model. Again, effects of factors other than temperature are more difficult to implement, but it is possible to do so. The major difficulty is that a large number of functions must be specified, and then parameters must be estimated for these functions.

An implementation detail which saves considerable computational time and storage space is the use of linked lists rather than arrays for maintaining cohort data. The linked list is a data structure which consists of records containing a group of variable values (number, age, cumulative loss and cumulative emergence at the last iteration, in this model), and a pointer to another such record. These lists may vary in length as elements are inserted at the beginning, and deleted at any location in the list. When a cohort is "born", a new record is dynamically allocated (a feature of the programming language Pascal), and inserted at the head of a list maintained for the stage. Cohorts are updated by traversing the list from the head to the last

element, and if the number in a cohort drops below some small value, the cohort is considered to have died out, so the cohort is removed from the list, and the memory used to hold the variables is returned to a pool of storage for re-use. This method avoids having a large matrix with many zero values which must be re-examined at every iteration. The dynamic allocation and freeing of records also corresponds well with the birth and death processes which cohorts undergo.

Details of operation of the model's algorithm are shown in Fig. 7. There is an outer loop for each day, and inner loops for each hour for the temperature-dependent functions for each stage, and for the cohorts in each stage. As the cohorts are updated, contributions to the following stage (eggs "follow" adults) are summed; which stage follows is determined from a table read as data. This allows an overwintering stage to be distinguished from the same non-overwintering stage, while both can contribute to the same following stage. If sufficient numbers of a stage are generated in a day, a new cohort is created, and inserted in the list. Different generations of the same stage are not distinguished, unless they are explicitly included as separate stages. The unified treatment of all stages allows a simple structure in the model, and easy addition or elimination of stages as suits the biology of the system and the goals of the model.

An Annotated Bibliography

Curry, G. L., R. M. Feldman, and K. C. Smith. 1978. A stochastic

model of a temperature-dependent population. Theor. Pop. Biol. 13: 197-213. (development of iterative cohort model)

Curry, G. L., P. J. H. Sharpe, D. W. Demichele, and J. R. Cate. 1980. Towards a management model of the cotton-boll weevil ecosystem. J. Environ. Management 11: 187-224. (application of iterative cohort model)

Feldman, R. M., and G. L. Curry. 1982. Operations research for agricultural pest management. Oper. Res. 30: 601-618. (introduction to a variety of models and optimization)

Gutierrez, A. P., G. D. Butler Jr., Y. Wang, and D. Westphal. 1977. The interaction of pink bollworm (Lepidoptera: Gelichiidae), cotton, and weather: a detailed model. Can. Entomol. 109: 1457-1468. (application of finite difference approach)

Jones, J. W., H. D. Bowen, R. E. Stinner, J. R. Bradley, Jr., and J. S. Bacheler. 1977. Simulation of boll weevil population as influenced by weather, crop status and management practices. Trans. Am. Soc. Agric. Eng. 21: 121-125, 131. (a flexible discrete dynamics model)

Lefkovitch, L. P. 1965. The study of population growth in organisms grouped by stages. Biometrics 21: 1-18. (primary reference for stage-projection matrices)

Leslie, P. H. 1945. On the use of matrices in certain population mathematics. Biometrika 33: 183-212. (a primary reference on age-projection matrices)

Leslie, P. H. 1948. Some further notes on the use of matrices in population mathematics. Biometrika 35: 213-245.

Lewis, E. G. 1942. On the generation and growth of a population. Sankya 6: 93-96. (the first reference on age-projection matrices)

Manetsch, T. L., and G. L. Park. 1977. Systems analysis and simulation with applications to economic and social systems. Part II. Dept. of Electrical Engineering and Systems Science, Michigan State Univ., East Lansing, Mich. (development of distributed delay approach)

May, R. M. 1979. Simple models for single populations: an annotated bibliography. Fortschr. Zool. 25: 95-107. (a good reference on models without age structure)

Reynolds. J. F. 1979. Some misconceptions of mathematical modeling. What's new in plant pathology 10(11): 41-43. (some useful ideas to remember when modeling)

Shoemaker, C. A. 1973. Optimization of agricultural pest management. I: Biological and mathematical background. Math. Biosci. 16: 143-175. (a good introduction to pest management models)

Shoemaker, C. A. 1977. Mathematical construction of ecological models. pp. 75-114 in C. A. S. Hall and J. W. Day (eds.) Ecosystem modeling in theory and practice: an introduction with case histories. J. Wiley and Sons, Inc.

Shoemaker, C. A. 1980. Deterministic and stochastic analyses of the optimal timing of multiple applications of pesticides with residual toxicity. Based on a paper presented at the Pest Management Conference, Oct. 1979, IIASA, Laxenberg, Austria. 40 pp. Dept. of Environmental Engineering, Cornell

Univ., Ithaca, N.Y.

Shoemaker, C. A. 1981. Applications of dynamic programming and other optimization methods in pest management. IEEE Trans. Auto. Control AC-26: 1125-1132.

Stinner, R. E., R. L. Rabb, and J. R. Bradley. 1974. Population dynamics of Heliothis zea (Boddie) and H. virescens (F.) in North Carolina: a simulation model. Environ. Entomol. 3: 163-168. (a discretized continuous insect population model)

Streifer, W. 1974. Realistic models in population ecology. Adv. Ecol. Res. 8: 199-266. (an excellent introduction to von Foerster models)

Taylor, F. 1979. Convergence to the stable age distribution in populations of insects. Am. Nat. 113: 511-530.

Usher, M. B. 1972. Developments in the Leslie matrix model. pp. 29-60 in J. N. R. Jeffers (ed.), Mathematical models in ecology. Blackwell, Oxford. (extensions to the Leslie matrix)

Wang, Y., A. P. Gutierrez, G. Oster, and R. Daxl. 1977. A population model for plant growth and development: coupling cotton-herbivore interactions. Can. Entomol. 109: 1359-1374. (an application of the finite differences technique to von Foerster models)

Welch, S. M., B. A. Croft, J. F. Brunner, and M. F. Michels. 1978. PETE: an extension phenology modeling system for management of multi-species pest complex. Environ. Entomol. 7: 482-494. (application of distributed delay to a phenology model)

Figure 1. Advantages and disadvantages of different types of models

Type of Model	Changes in Parameters	Large Number of Variables	Large Number of Management Alternatives
Analytical	excellent	poor	fair
Simulation	fair	excellent	poor
Optimization	fair	fair	excellent

Figure 2. Choices in Model Construction:

Single- vs. Multiple-species

Age- or Stage-structured vs. No structure

Deterministic vs. Stochastic

Discrete vs. Continuous

Chronological vs. Physiological Time

Figure 3. Implementation of models
with age- or stage-structure

Discrete models:

Leslie matrix

Lefkovitch matrix

Discretized continuous models:

Distributed delay

von Foerster equations:

finite differences

iterative cohorts

Figure 4. Leslie Age-Projection Matrix

$$\begin{bmatrix} n_0 \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{bmatrix}_{t+1} = \begin{bmatrix} f_0 & f_1 & f_2 & f_3 & f_4 \\ p_0 & 0 & 0 & 0 & 0 \\ 0 & p_1 & 0 & 0 & 0 \\ 0 & 0 & p_2 & 0 & 0 \\ 0 & 0 & 0 & p_3 & 0 \end{bmatrix} \cdot \begin{bmatrix} n_0 \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{bmatrix}_t$$

$$\equiv$$

$$\underline{n}(t+1) = A\ \underline{n}(t)$$

$$\equiv$$

$$n_0(t+1) = \sum f_i\ n_i(t)$$

$$n_i(t+1) = p_{i-1}\ n_{i-1}(t)$$

Figure 5. Lefkovitch Stage-Projection Matrix

$$\begin{bmatrix} n_0 \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{bmatrix}_{t+1} = \begin{bmatrix} p_{0,0} & f_1 & f_2 & f_3 & f_4 \\ p_{0,1} & p_{1,1} & 0 & 0 & 0 \\ 0 & p_{1,2} & p_{2,2} & 0 & 0 \\ 0 & 0 & p_{2,3} & p_{3,3} & 0 \\ 0 & 0 & 0 & p_{3,4} & p_{4,4} \end{bmatrix} \cdot \begin{bmatrix} n_0 \\ n_1 \\ n_2 \\ n_3 \\ n_4 \end{bmatrix}_t$$

$p_{i,i}$ = probability of staying in stage i

$p_{i,i+1}$ = probability of moving from stage i to stage i+1

$1 - (p_{i,i} + p_{i,i+1})$ = probability of dying in stage i

(all probabilities are for iteration interval)

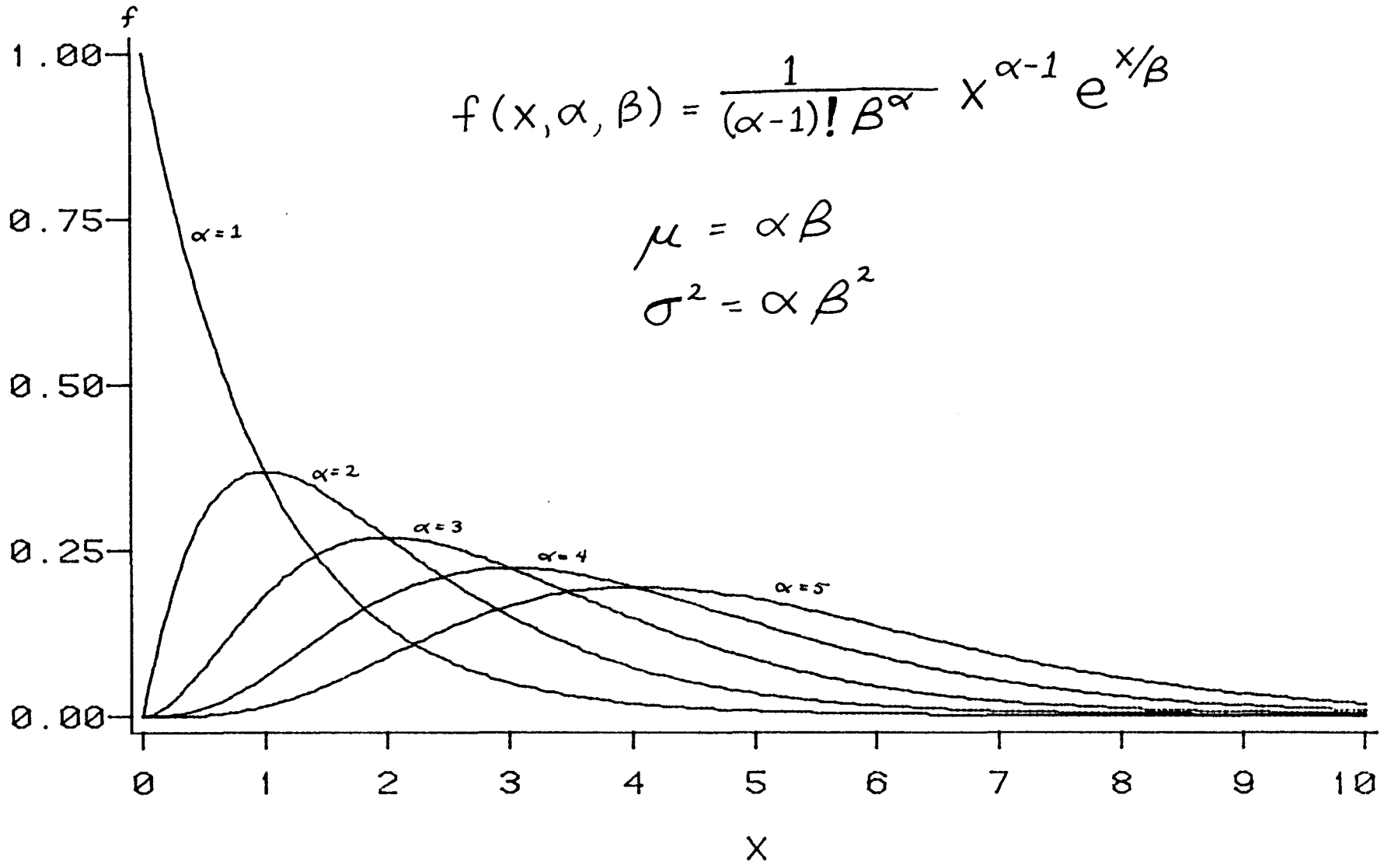

Figure 6. Gamma (Erlang) distribution for alpha = 1 to 5, beta = 1.

Figure 7. Generalization of Iterative Cohorts

Notation:

$N_{ij}(t,a)$ = number in cohort j of stage i at time t, and of physiological age a

$\psi(t)$ = temperature at time t (time in units of days)

$r_i(\psi)$ = developmental rate of stage i

$c_i(\psi)$ = production rate of stage i

$F_i(a)$ = loss (mortality) function of stage i ($0 \leq F \leq 1$)

$P_i(a)$ = birth or emergence function of stage i ($0 \leq P \leq 1$)

Computation:

For each day:

For each hour, for each stage i:

$$\Delta a_i = \Sigma\, r_i(\psi(t))\, \Delta t \quad \text{where } \Delta t = 1 \text{ hour}$$

$$c_i = 1/\Delta t \; \Sigma\, c_i(\psi(t))$$

For each stage i, for all cohorts j in i:

$$N_{ij}(t+1, a_{ij}+\Delta a_i) = N_{ij}(t, a_{ij}) \cdot [1 - F_i(a_{ij} + \Delta a_i)]/[1 - F_i(a_{ij})],$$

if $N_{ij} <$ min then delete cohort;

$$N_{k,new}(t+1,0) = c_i \; \Sigma\, N_{ij}(t, a_{ij}) \cdot [P_i(a_{ij} + \Delta a_i) - P_i(a_{ij})]/[1 - F_i(a_{ij})],$$

where k = stage following i.

For each stage i:

if $N_{i,new} >$ min then create new cohort

(more than 1 stage can contribute to a cohort)