Toward an Expert System for Development of Pest Simulation Models

JESSE A. LOGAN

Natural Resource Ecology Laboratory and Department of Entomology, Colorado State University, Fort Collins, Colorado 80523

Environ. Entomol. 17(2): 359–376 (1988)

ABSTRACT Computer modeling of pest population dynamics has assumed an important role in many integrated pest management programs. This approach would perhaps see even wider application if the existing modeling expertise were either more readily available or less expensive. Expert systems is a recently emerging technology that promises to reduce the expense associated with human expertise for some restricted problem domains. In this report I discuss development of a computer program, motivated from expert systems approaches, that automatically assembles a model describing insect population phenology. The ultimate goal of this research is to provide a system that generates an independent model comprising a complete life system description. Therefore, the current system represents only a first step toward meeting this goal. The current system is further restricted to the relatively straightforward case of building a phenological model based on laboratory data. Nevertheless, the system results in significant time saving for problems in this restricted domain, and its performance is judged to compare favorably with that of a human modeler.

KEY WORDS Insecta, phenology, simulation, pest management

APPLICATIONS OF computer modeling techniques have made important contributions to integrated pest management (IPM) research (for review articles see Ruesink [1976] and Getz & Gutierrez [1982]). This approach would perhaps see even wider application if the expertise necessary for implementation were more readily available or less costly, or both. One recently emerging technology which attempts to alleviate these two problems associated with human expertise is expert systems technology. An expert system can be described as "a computer program that uses expert knowledge to attain a high level of performance in a narrow problem area . . . to address problem areas that require years of special training and education for humans to master" (Waterman 1986). In other words, expert systems are computer programs that to some extent replace human experts. Expert systems have as their underlying principle the goal of making sophisticiated computer capabilities available without requiring the user to become a computer sophisticate (Klahr & Waterman 1986). The recent unrelenting increases in computing power, combined with new ideas in the design of computer languages and operating systems, have resulted in startling advances in the area of expert systems (Klahr & Waterman 1986, Waterman 1986).

Expert systems technology has been applied in diverse areas such as medical diagnosis, litigation, and international relations (see Waterman [1986] for a comprehensive bibliography of expert systems). Only recently has this technology begun to be applied in agricultural and entomological sciences (Holt 1985, Lemmon 1986). Stone et al. (1986) discuss potential applications of expert systems in entomological and IPM problems. However, they failed to mention the potential of this technology for the design of simulation models. Application of expert systems for design of complex physical systems has been highly successful (e.g., O'Connor [1984]), and by analogy, design applications in ecological problems are a potentially fruitful avenue for research.

The ultimate goal of research described in this paper is the development of a computer program capable of building an insect population model from life history information. As a first step toward meeting this goal, a computer program was developed that is capable of building a simulation model of population phenology. To remain consistent with the previous objective, the generated model must also provide a framework for a more detailed simulation that includes demographic variables, and the flexibility for embedding the generated model within a more complete description of community dynamics.

The basis for many, if not most, insect pest models is population phenology. This is a result of both the central position of timing to the adaptation of cold-blooded organisms in seasonal habitats, and the importance of accurate timing in design of efficient control strategies. A large number of phenological models for IPM applications have been based on laboratory experiments in which developmental observations are made for a series of constant-temperature experiments. The current model development system is restricted to these types of data, which may consist of either median developmental times (rates) or a more detailed description of percentile emergence times (rates). A sim-

0046-225x/88/0359-0376$02.00/0

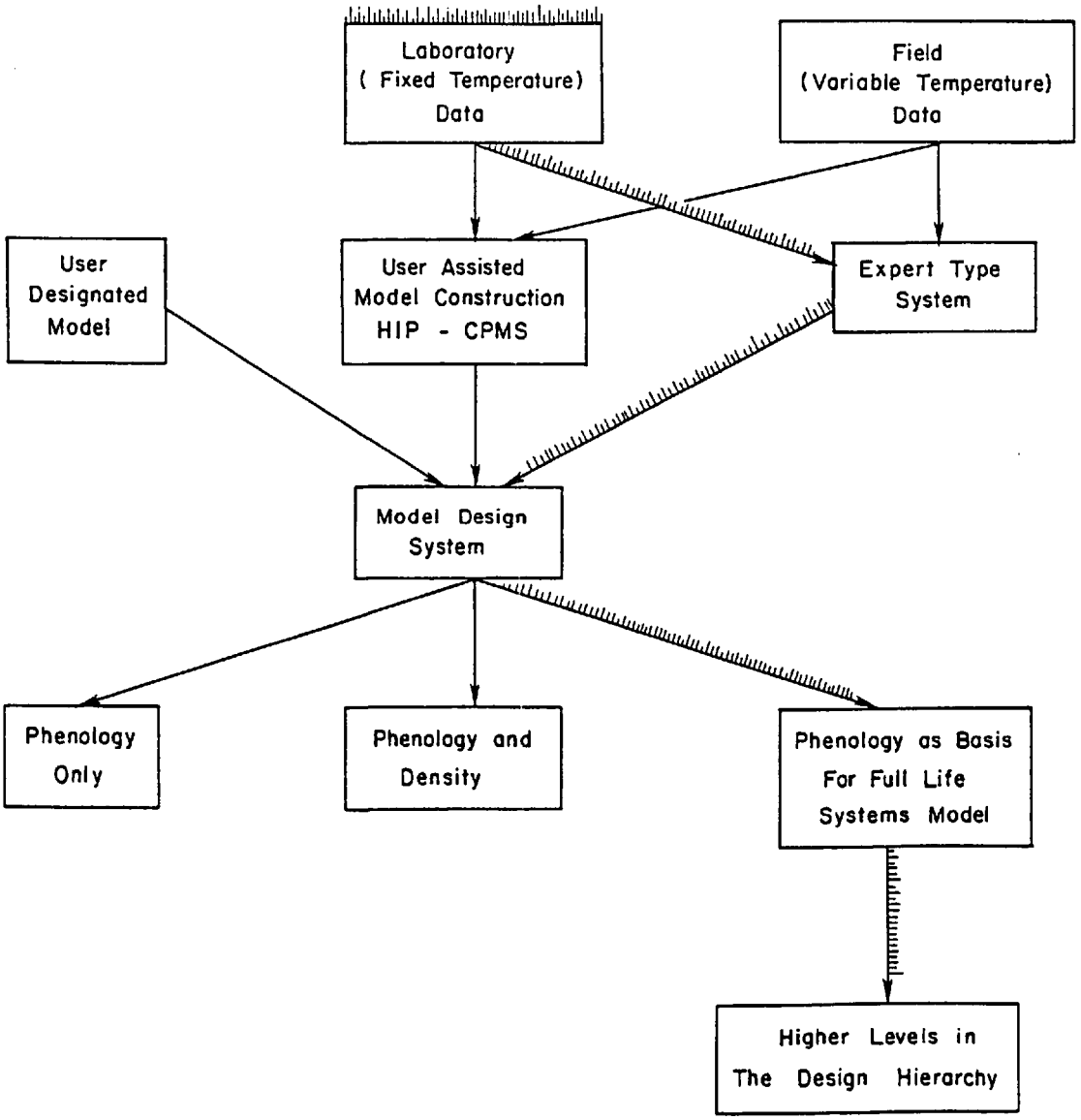

Fig. 1. Logical structuring of PMDS. The cross-hatched pathway is discussed in the text.

ilar model could, at least in theory, be developed from field observations of emergence times and a concurrent temperature record. However, the latter problem is more difficult, and it seems only prudent first to demonstrate the efficacy of the system for generation of models based on the more straightforward, and interpretable, laboratory data. In keeping with the long-term objective of this research, the acronym PMDS (Pest Model Design System) is used to reference this model development tool. PMDS is written in FORTRAN 77 and currently structured for execution on a VAX computer with the UNIX operating system. The model that it produces is an independent FORTRAN 77 program that simulates insect population phenology and includes the structure for simulating other demographic processes such as recruitment and mortality.

System Description

Overview of the Problem. A model description of insect phenology requires an adequate data base, the conceptual basis to interpret these data, and analytical tools necessary for implementation of the conceptual model. A diagrammatic representation of how these sources of knowledge fit together and interact is given in Fig. 1. Note that linking some components of this diagram can be accomplished without necessarily including inference capabilities. For example, the system could fit specified curves to data and build a user-designated model.

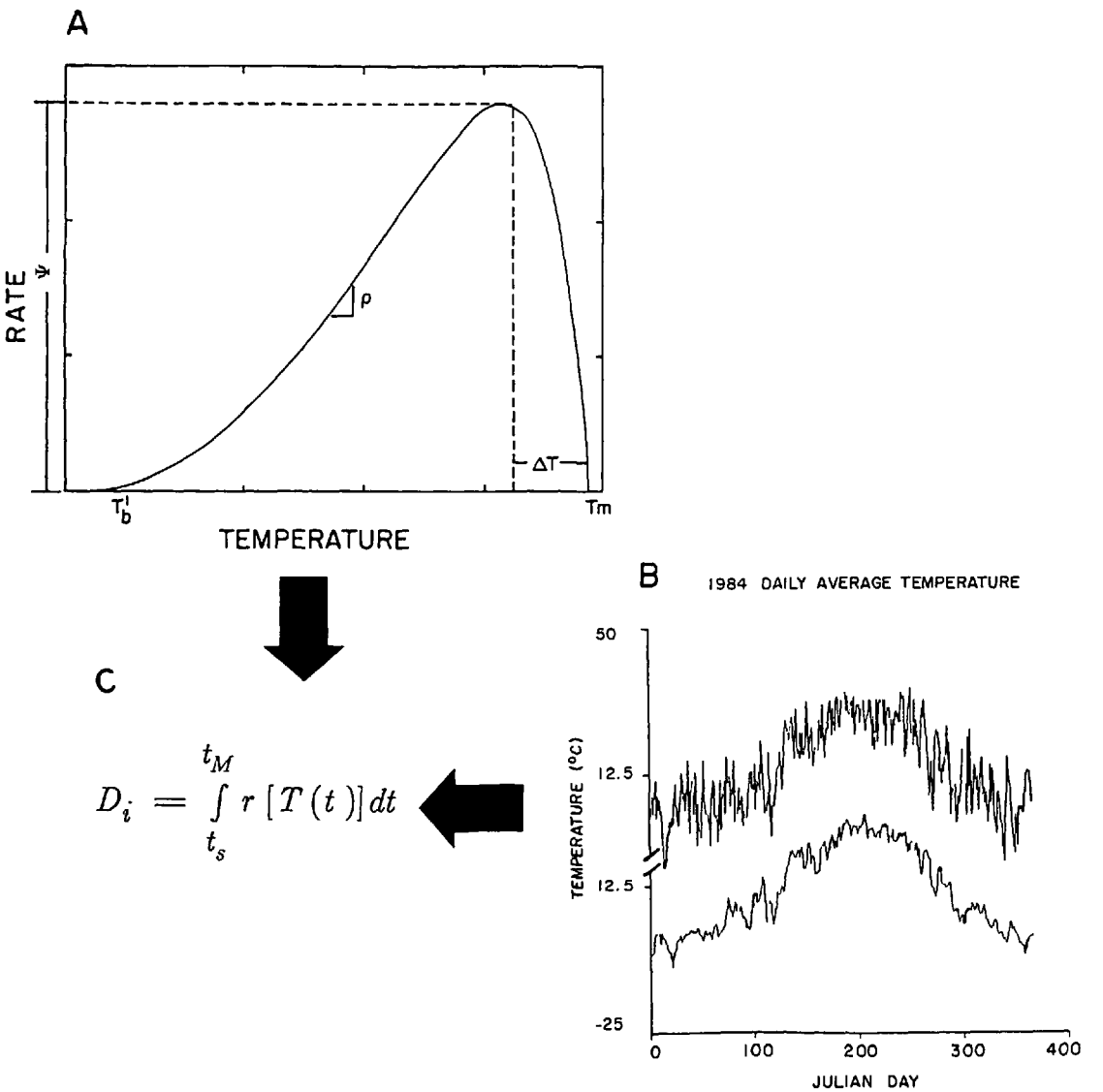

Fig. 2. (A) A generalized insect developmental rate curve. Descriptive parameters for this curve are: T_b, the base temperature below which development does not proceed; ψ, the maximum developmental rate; ρ, a measure of increasing rate with increasing temperatures; ΔT, the width of decline phase in developmental rate above optimum temperature; and T_m, the thermal maximum. (B) Annual curves for ambient air and 2.5-cm soil temperatures at the Central Plains Experiment Range, Nunn, Colo. Note discontinuous scales on the Y-axis. (C) The integral equation that is responsible for synthesis of the rate–temperature and temperature–time curves.

Although this application does not include the inference function of an expert system, it could result in significant time saving for a modeler.

Canonical Form for the Developmental Rate Curve. The basis for modeling insect phenology centers about the developmental rate curve, where developmental rate is defined as the inverse of the time required to complete an instar or life stage. A generalized developmental rate curve as a function of temperature is shown in Fig. 2A. This figure represents the shape of the curve over the full range of physiologically viable temperatures, which may or may not be experienced in a particular microhabitat. This generic form is well documented (Logan & Hilbert 1983, Wagner et al. 1984a) and generally accepted enough to be considered a canonical form. The five descriptive parameters listed on this figure are sufficient for the description of the canonical form of the curve. Because these parameters have descriptive interpretation, they are synonymous with the term "ecological parameters" used by Lamb et al. (1984).

Expression of the rate curve shown in Fig. 2A is mediated through the insects' microhabitat, as represented by the annual temperature cycles in Fig. 2B. Note that the noise associated with ambient

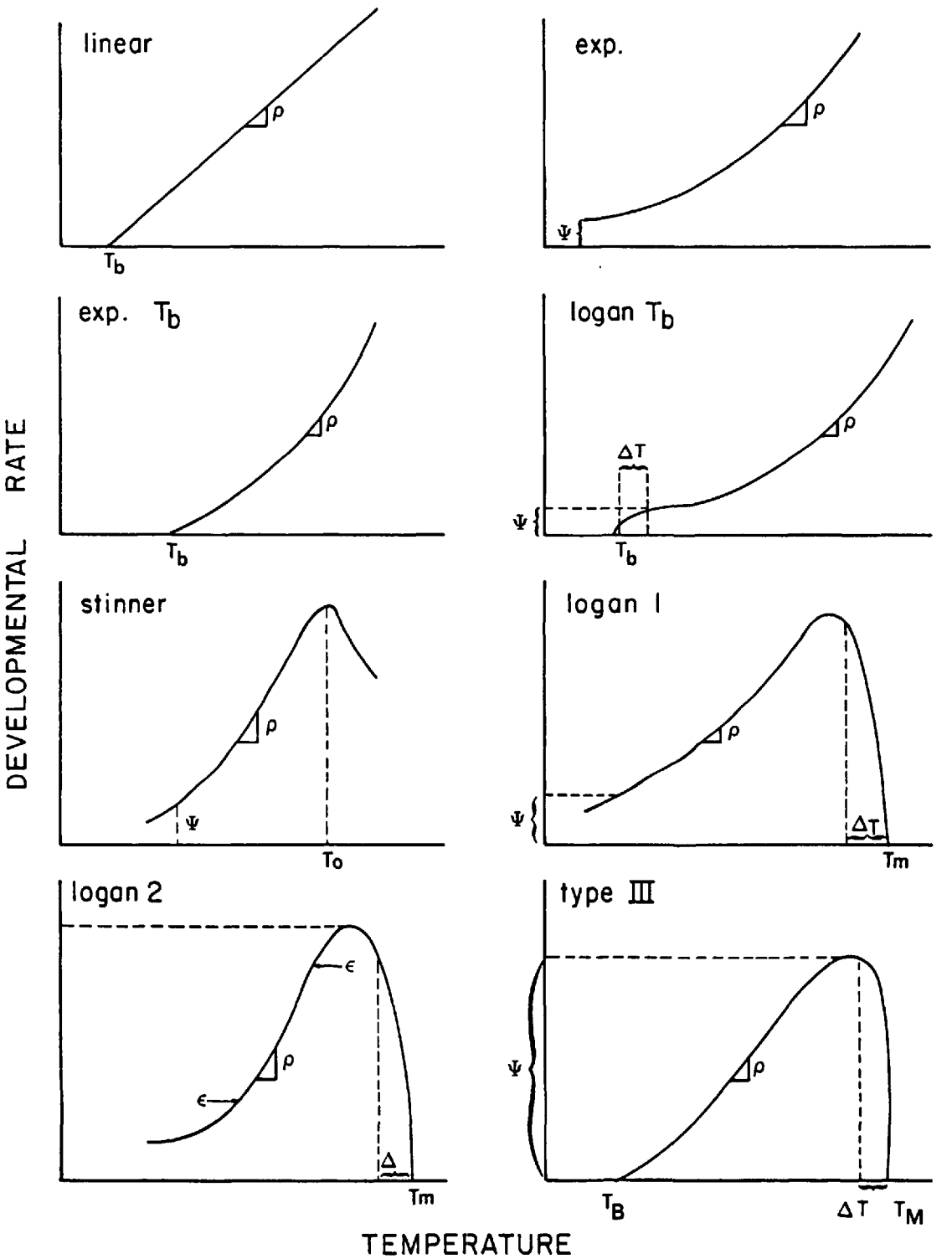

Fig. 3. Examples of functional relationships that have been used to represent the insect developmental rate curve. Reference names were derived either from descriptive properties of the function or from a published application. See Appendix I for reference and information on parameter estimation.

temperature is often reduced through an appropriate microhabitat filter, as represented by the accompanying soil temperature curve.

Integration, both figuratively and literally, is accomplished through application of the equation in Fig. 2C, where developmental rate (r) is a function of temperature (T), which in turn is a function of time (t). The equation in Fig. 2C can be interpreted in either individual or population terms. For the individual, a developmental index (D_i) of unity indicates that the individual has completed the life stage. For the population, a similar situation indicates that the median point has been reached in the distribution of developmental times. This equation embodies several assumptions, the most obvious of which is that developmental rates adjust instantaneously to changes in temperature, a reasonable assumption given the typically smooth, gradual nature of field temperature curves. Another assumption is lack of synergism due to fluctuating versus constant temperatures. This second assumption is more tenuous (Gregg 1982) but has been found to be approximately true for many situations.

Rate Curve Representation. Various representations of the developmental rate curve of Fig. 2A have been used in the past. A representative progression of such curves is shown in Fig. 3. Each of these curves has been used successfully as the physiological basis for modeling insect phenology. The degree of success in previous applications has depended largely on the particular ecological setting.

Three sources of information are used by PMDS to select the exact functional representation from the potential functions listed in Fig. 3. These sources, listed in priority order, are (1) objectives of the program user, (2) the best representation of the canonical form of the developmental rate function, and (3) data constraints imposed by the minimum degrees of freedom required for curve fitting. Various possible actions at decision junctures are determined both by the performance of the system prior to the juncture point, and the options remaining to the system beyond the juncture.

Ecological knowledge is entered into the system through dialogue with the user. For each life stage, the system prompts for an evaluation of the importance of the various ecological parameters listed on Fig. 2A. User response is a numerical evaluation ranging from 0 to 10. The magnitude of the response corresponds to the degree of importance of the ecological parameter to the particular life stage in question. A 0 response indicates the parameter is definitely not important for the ecological setting in question, a response of 5 indicates neutrality (the parameter may or may not be important), and a response of 10 indicates the user believes the parameter to be of critical importance to the developmental process under consideration. For example, in temperate zones, the description of life stages that occur in early spring or winter would most likely not emphasize representation of high temperature phenomena. Conversely, those in middle or late summer would probably not require low temperature thresholds.

Knowledge about the ecological setting for a particular life stage is combined with the best representation of the canonical developmental rate function through application of the attributes listed in Table 1. As indicated in this table, an initial ranking of functional forms to be fitted to data is computed by

$$IR_j = \sum_{i=1}^{5} I \cdot A_{ij} \qquad (1)$$

where IR_j is the initial ranking for the jth function, I is the relative importance of the ith descriptive attribute, and A_{ij} is how well the ith attribute is represented by the jth functional form. A_{ij} is typically null or unity (a candidate function either incorporates a particular attribute or it does not). However, note that this is not necessarily so, as with the Type III curve which in practice has not performed well for representation of low-temperature nonlinearity. The initial ranking obtained from Equation 1 is not necessarily unique; in fact, the Logan1 and Logan2 functions will always result in the same score. In case of a tie score, priority order will default to the function with the fewest parameters.

Once an initial ranking has been obtained, potential functions are tested for data constraints on two criteria—the minimum number of data points for parameter estimation (i.e., at least two points are required to fit a straight line), and information of declining developmental rate due to high temperature stress (Table 2). Constraints are imposed on the intitial ranking, and functional forms with unsatisfied constraints are eliminated. This process results in a final ranking of candidate functions to be considered for parameter estimation.

Model representation of the canonical form is accomplished through estimating parameters for developmental rate models in priority order. Curve fitting is accomplished either through linearization of the function and linear least squares regression, or through nonlinear least squares procedures. The nonlinear least squares routine used is based on Powell's (1965) function-minimizing algorithm. Past experience has demonstrated good convergence properties with this method in a wide variety of applications. Parameter fitting procedures are further discussed in Appendix 1.

Programs or procedures that perform various functions on command from an expert system are called demons (Coulson & Saunders 1987). In PMDS, a demon for parameter estimation is invoked once a final ranking of functional forms has been achieved. On return to the calling program, two possible conditions exist—either the demon was successful in parameter estimation or it wasn't. This information, combined with the existing knowledge base, allows several potential actions. These actions (goals) are listed, along with infer-

Table 1. Attribute table for potential representations of the canonical forms in Fig. 2. Application of this table by PMDS integrates domain- and meta-knowledge, resulting in an initial, priority-order ranking of developmental rate models

(1) Domain evaluation	(2) Descriptive property	(3) Descriptive value for development rate model							
		Linear	Exp	Exp Tb	Logan1 Tb	Stinner	Logan1	Logan2	Type III
0–10	Tb	1	0	1	1	0	0	0	1
0–10	Nonlinear	0	1	1	1	1	1	1	0.85
0–10	Rate decline $T \rightarrow To$	0	0	0	0	1	1	1	1
0–10	Asymmetric $T > To$	0	0	0	0	0	1	1	1
0–10	Tm	0	0	0	0	0	1	1	1
Initial ranking $\sum_{i=1}^{4}[(1)\cdot(3)]$									

ential pathways, in the dependency network of Fig. 4. The decision process is more complex and requires more detailed knowledge for an unsuccessful fit than for a successful fit. This is a reflection of the fact that failure cannot result from a successful fit; however, an unsuccessful fit does not necessarily imply failure. Both past history and future options must also be considered.

The inference procedure shown in Fig. 4 has "conflict resolution" as one of its goals. This goal is reached when fitting of curves in priority order results in ambiguity that must be resolved by other means. Because of the preceding model selection process, conflict resolution is not the typical statistical question of determining if one arbitrary function is better than another at some prechosen probability. However, it also does not seem fair to directly compare alternative functions (e.g., through R^2 values) when the degrees of freedom (i.e., the number of functional parameters) may be unequal. To resolve this dilemma, Kvalseth (1985) suggests application of an adjusted coefficient of determination in which

$$R_a^2 = 1 - (1 - R_i^2)[n/\{n\cdot(n - k)\}] \quad (2)$$

where n is the number of observations, k is the number of parameters in the ith function, and R_i^2 is computed as

$$R_i^2 = 1 - \sum_{j=1}^{n} (y_j - \hat{y}_{ji})^2 / \sum_{j=1}^{n} (y_j - \bar{y})^2$$

for y_j the jth observed median developmental rate and $\hat{y}_{ji}$ is the jth predicted developmental rate from the ith function.

Once conflicts have been resolved, PMDS provides the option of viewing a plot of all functions that were successfully fitted to data. Through this qualitative evaluation, the program user can choose either to accept the functional relationship selected by PMDS or to override the decision by choosing an alternative form. In any case, after successfully determining functions for each life stage in the model, PMDS prints summary tables of data, ecological knowledge represented, ranking of functional forms, and the final fit determined by PMDS (see Appendix II).

This procedure essentially guarantees that some functional form will be fit to data. Practically speaking, the worst-case scenario is a linear fit, which is in fact the basis of most phenological models in use today (i.e., degree-day models). The only way I foresee complete failure is if the data set consists of only one point, or if the slope is in the wrong direction (i.e., at low temperatures, developmental rate decreases with increasing temperature). In either of these cases, PMDS would advise the user to consult with a human IPM modeling expert.

Including Variation. To this point, the description of PMDS capabilities provides sufficient information for a purely deterministic model of insect phenology. The only way this description can be expanded to a population concept is to assume that simulation of the median individual captures the essence of population processes. This may or may not be the case, depending on the intended objectives of a specific model. Many important population processes, such as natural selection, include variation as an essential component. One approach to modeling variation that has been widely applied is that of Monte Carlo simulation. Monte Carlo approaches generate variation through multiple computer runs, each with new parameter values that are drawn from an appropriate sampling distribution. Limitations to Monte Carlo techniques include the cost of multiple computer runs,

Table 2. Data constraints for the functional forms in Fig. 2

	Linear	Exp	Exp Tb	Logan Tb	Stinner	Logan1	Logan2	Type III
Minimum no. points	2	2	2	4	3 or 4	4	5	5
$r(T_{max}) < r_{max}$	No	No	No	No	No	Yes	Yes	Yes

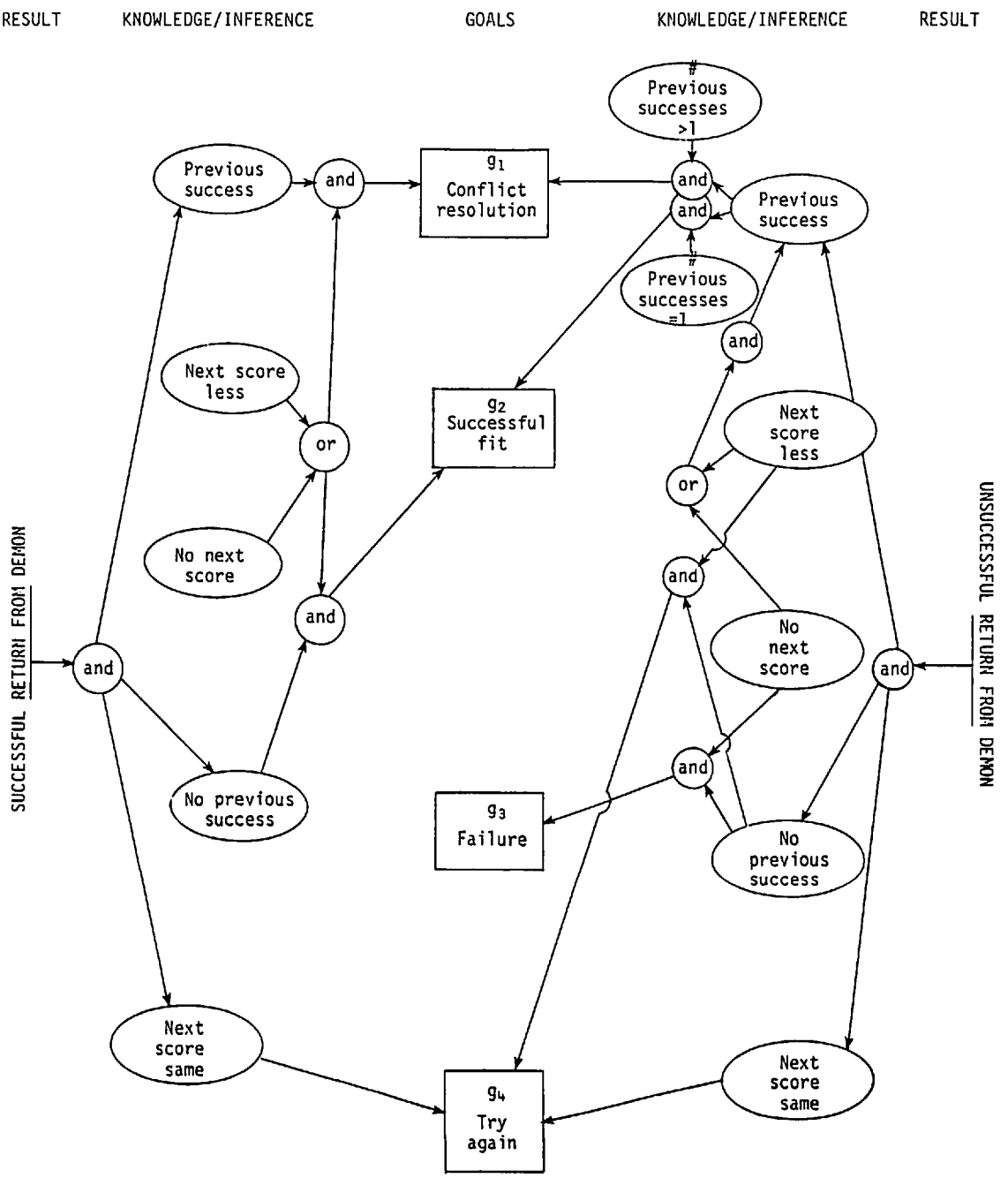

Fig. 4. Dependency network for program action based on the return status from the parameter estimation demon. See text for further explanation.

difficulty in estimating appropriate distribution functions for parameters, and the inability to capture some forms of variation (O'Neill 1979). An alternative to the Monte Carlo approach is one in which variation is explicitly included at a phenomenological level. In this approach, all variants about the median are simulated simultaneously. Such models are not truly stochastic, in that identical model prediction will always result from a given set of initial conditions, and they differ from typical deterministic models in that prediction includes more than that for the median individual. The term "distribution model" has been used to differentiate this approach from both stochastic and purely deterministic models (Logan & Amman 1986). If data are sufficient, PMDS proceeds to build a distribution model; if not, a purely deterministic model will be assembled. Because the detail of develop-

mental data for various life stages of a particular organism may vary, PMDS is capable of constructing a mixed model in which some life stages are represented deterministically and others by a distribution model.

The method PMDS uses to generate a distribution model is based on the same-shape approach derived by Sharpe et al. (1977) (further described in Wagner et al. [1984b]). The basic assumption of this technique is that the distribution of *normalized* developmental times (or rates) for the various fixed temperature experiments are approximately coincident. Therefore, when normalized, the various curves assume the same shape, and one cumulative probability distribution can be used to describe the aggregated data. This curve can be used subsequently to model developmental times under continuously varying temperatures. Other distribution approaches do not require the same-shape assumption (e.g., Régnière [1984]); however, these alternative approaches require more computer memory. In my opinion, there is currently insufficient evidence to warrant the additional computational overhead of relaxing the same-shape assumption.

The minimum data base required for including variation is the program user's best guess of the time required for the fastest and the slowest developing individuals to complete the life stage at any constant temperature. In this case, PMDS assumes that the distribution of developmental rates is approximately normal with finite end points A^* for the developmental rate of the slowest and B^* for that of the fastest. This distribution is then modeled by a hyper-power function (Stinner et al. 1975)

$$G^*(r^*) = (1 - z)^{6.192 \cdot z^{2.63}} \quad (3)$$

for

$$z = (B^* - r^*)/(B^* - A^*)$$

where G^* is the cumulative distribution of normalized developmental rates and r^* is normalized developmental rate. A best-guess model such as this would find primary application in answering "what if" types of questions or in preliminary sensitivity analysis, or both.

If more complete information is known (i.e., the developmental times of individuals in each constant temperature experiment), PMDS proceeds by: (1) normalizing the observed developmental times (rates) by multiplying (dividing) the observed values by the median developmental rate; (2) the resulting data sets are combined to provide data for estimating two distributions, one describing the cumulative distribution of developmental rates and the other for the cumulative distribution of developmental times; (3) the parameter estimation demon is called to fit three cumulative probability distributions to each data set; and (4) the best fit of the six distributions is determined by comparison of adjusted R^2 values (Kvalseth 1985). The three distributions that are fitted have all been used previously to model insect phenology. These distributions are the three-parameter Weibull distribution (Wagner et al. 1984b), the four-parameter hyper-power function of Stinner et al. (1975), and the two-parameter logistic (Régnière 1984). As in the previous case of determining the best model for developmental rates, PMDS provides the program user with graphic options to visually appraise performance and override choices made by PMDS.

Produced Model. The final task for PMDS is to take results for each life stage and construct a working simulation model describing population phenology. The general idea of a phenological model is the translation of insect physiological time to clock (calendar) time. An additional objective of PMDS is to provide a generalized modeling framework that can be easily modified to include demographic variables, such as life stage distribution, recruitment, and mortality. A PMDS-generated model should, therefore, perform two basic tasks. The first of these is translation of thermal to clock time. Second, an efficient and flexible bookkeeping scheme is necessary to keep track of demographic variables.

The same-shape model used by Sharpe et al. (1977) can be based either on the cumulative distribution of developmental rates $[G(r)]$ or on the cumulative distribution of developmental times $[F(t)]$. Normalization transforms each of these distributions to their corresponding same-shape distributions $[G^*(r^*)]$ and $[F^*(t^*)]$, where t^* is computed as the D_i of Fig. 2C, and $r^* = 1/t^*$. Because the distribution of either developmental times or developmental rates can serve equally as the basis of modeling phenology, PMDS determines which distribution to use purely on the empirical basis outlined in the previous section. The relationship between the normalized distributions of developmental rates and developmental times is shown in Fig. 5.

Perhaps the most widely used bookeeping algorithm for insect developmental models is based on a distributed delay approach (i.e., Erlang probability distributions). Distributed delay models often have worked quite well (e.g., Welch et al. [1978], Gutierrez et al. [1984]). However, there are difficulties with this technique. The underlying mathematical basis is arcane and difficult to explain to nonquantitative biologists. Distributive delay algorithms are also typically sensitive to changes in underlying rate functions and distribution relationships. A more intuitively straightforward and flexible approach is provided by capitalizing on a cohort concept of insect populations (Curry et al. 1978, Logan 1979, Shaffer & Gold 1985, Wagner et al. 1985, Logan & Amman 1986).

A cohort is defined as all individuals in a particular life stage that are of approximately the same chronological age; in practice, this is equivalent to all individuals that entered a particular life stage during the same simulation time step. Two pieces

of information are retained for each cohort—the initial number of individuals in the cohort (n) and their physiological age (a^*, defined as the summation of t^* since initiation of the cohort). To compute changes in demographic status attributable to phenology over some small time interval Δt, it is first necessary to update the physiological age of each cohort by computing the advancement of physiological time Δt^* that occurred during Δt. If the updated physiological age ($a^* + \Delta t^*$) for a cohort is greater than the physiological age of first emergence ($1/B^*$) but less than that of the slowest developing individual ($1/A^*$), then the number from the jth cohort that have completed the ith life stage during Δt (m_{ij}) is computed as

$$m_{ij} = [F^*(a^*_{ij} + \Delta t^*) - F^*(a^*_{ij})] \cdot n_{ij}, \quad (4)$$

where F^* denotes the standardized, cumulative emergence curve of Fig. 5.

A computed value $a^* + \Delta t^*$ that is greater than $1/A^*$ indicates that all individuals in the cohort have completed the life stage (e.g., $F^*(a^*_{ij} + \Delta t^*) = 1.0$). In this case m is computed as

$$m_{ij} = [1 - F^*(a^*_{ij})] \cdot n_{ij}, \quad (5)$$

and since all individuals in the jth cohort have completed the life stage, it can be dropped from the cohort vector. The updated number of cohorts comprising the ith life stage is then computed as the number of cohorts at the beginning of the time step (k_i) minus the number of cohorts in which $a^* + \Delta t^* > 1/A^*(h_i)$.

The number of individuals remaining in the ith life stage is then computed as

$$N_i = \sum_{j=1}^{k_i} [1 - F^*(a^* + \Delta t^*)] \cdot n_{ij}, \quad (6)$$

and the number that have progressed to the ith + 1 life stage is

$$c_i = \sum_{j=1}^{k_i} m_{ij}. \quad (7)$$

At the beginning of the next time step, the c_i individuals provide the initial density for a new cohort in the ith + 1 life stage. This new cohort has physiological age zero at time $t + \Delta t$, and the updated number of cohorts for the ith + 1 life stage is incremented by one. This updating procedure is diagrammatically represented in Fig. 6.

A PMDS model treats instars or life stages as a dual representation. For computation of demographic variables, every individual in a particular life stage is considered as equivalent; however, for computation of phenology, each life stage is considered a collection of cohorts. The duality of a PMDS model also meets the requirement for a generalized modeling framework, and both recruitment and mortality are included in the computational structure of the produced model (also see Shaffer & Gold [1985]).

Programming Considerations. The programs discussed in this paper comprise two distinct sets of code. The first is PMDS and the second is the model generated by PMDS. The generated model is coded entirely in ANSI FORTRAN 77. Every attempt has been made to produce transparent code that follows structured programming convention.

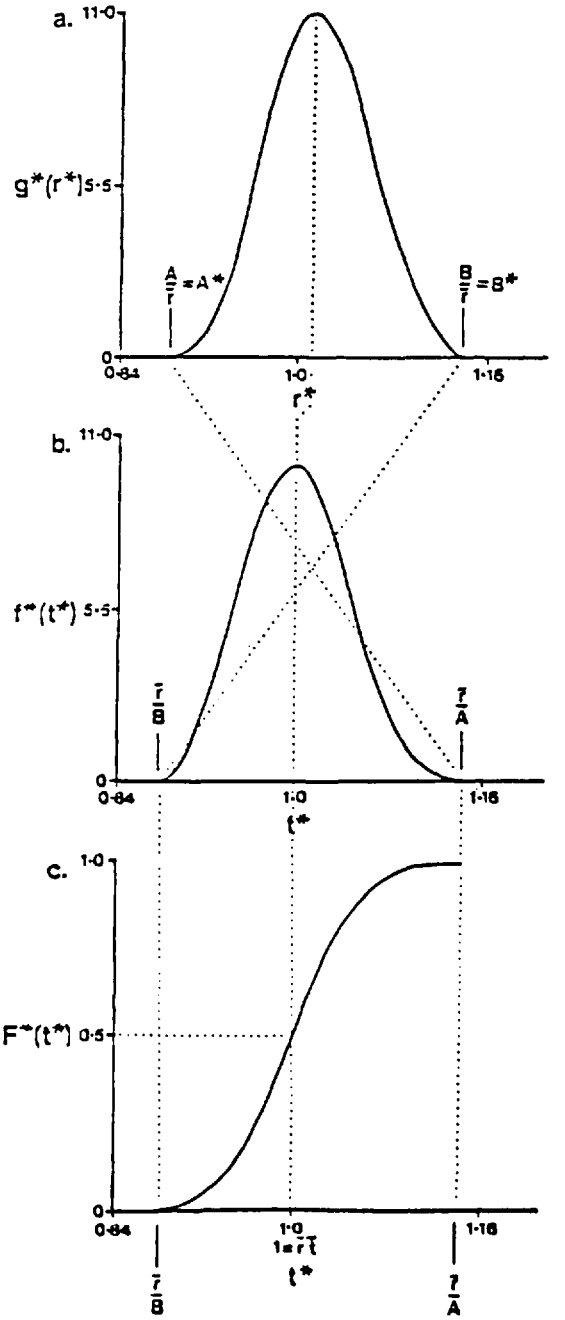

Fig. 5. Diagrammatic representation of normalized distributions. (a) Probability density of developmental rates [(g^*r^*)]. (b) Probability density of developmental times [$f^*(t^*)$]. (c) Cumulative density of developmental times [$F^*(t^*)$]. Curves are defined by the parameters $\bar{r}$ for the median developmental rate, A the rate of the fastest developing individual, B the rate of the slowest, and t^* computed from the developmental index (D_i) of Fig. 2C.

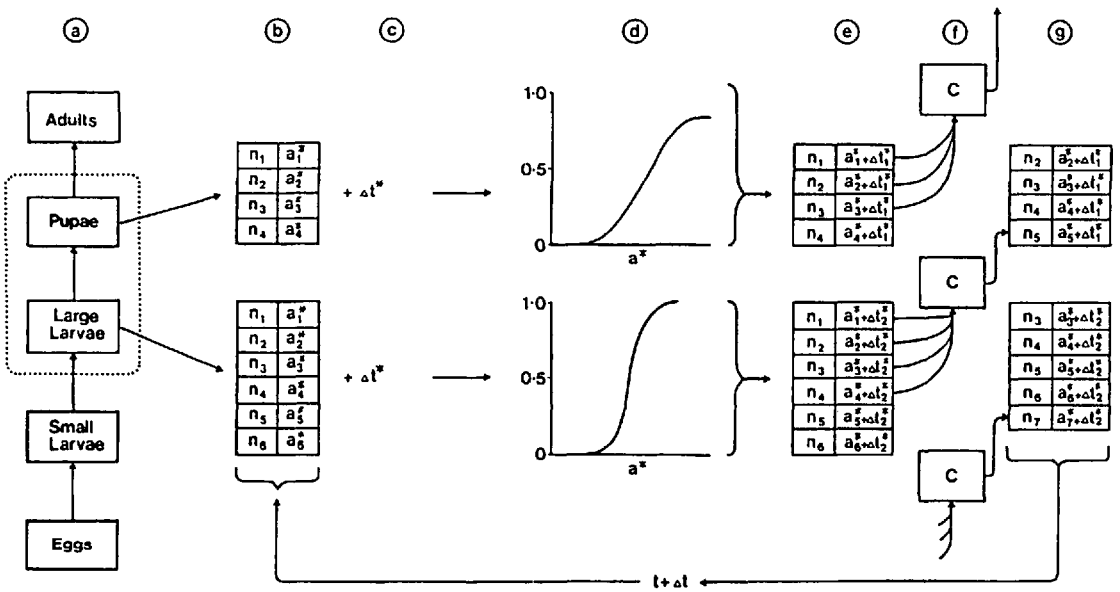

Fig. 6. Updating algorithm for a PMDS produced simulation of population phenology. The progression of population age structure is represented as (a) the flow of individuals through the various life stages. To accomplish this, each life stage is represented as (b) a vector of cohorts. Cohorts are updated over Δt by first computing Δt^* for each life stage, and (c) adding this to the a^* of each cohort. In (d), the stage-specific, cumulative emergence curve $[F^*(a^*)]$, is compared with (e) the updated physiological age. Depending on physiological age status (Equations 4 and 5), the proportion of each cohort that has completed the life stage is computed and in (f) a new cohort (c) is created from all individuals that have completed the previous life stage. Finally in (g), cohorts with $a^* \geq 1/A^*$ are dropped from the cohort vector and the c (Equation 7) from the previous stage are included as the last entry in the vector. The iterative nature of this process is indicated by the arrow from (g) to (b).

This has resulted in code that is highly portable. The produced model has been implemented on a wide variety of computers and operating systems, both in the United States and New Zealand, with minimal difficulties.

The programming environment used to develop PMDS had several requirements. Multitasking is the capacity for a computer to assign more than one concurrent computational task. Multitasking allows the executive program in PMDS to assign tasks conveniently to the appropriate demon. Efficient error trapping (detection of potentially fatal computational errors before they cause the program to crash) is required because fatal errors are not infrequent during execution of the numerical analysis procedures which are central to PMDS (e.g., nonlinear least squares programs). Finally, program development was facilitated by a mega-structuring approach which allowed program modules to be developed independently and linked later in an object oriented fashion. These three capabilities were sufficiently provided by the UNIX operating system (Berkeley version 4.3) currently in use on the Natural Resource Ecology Laboratory VAX 11/730 computer. This computer was used exclusively in the development of PMDS.

In the current version of PMDS, executable UNIX script is central to operation of the system and is used to link various FORTRAN 77 modules that actually perform the system's functions, for error trapping in computational demons, to accomplish graphics functions, and to provide the summary tables produced by the report writer demon (see Fig. A2.1, A2.2, and A2.5). Clearly, direct implementation of PMDS requires UNIX capabilities. This factor reduces the transportability of the system, although UNIX is essentially machine independent and available as either the principal or guest operating system on a wide variety of computers ranging from PCs to super computers (Crecine 1986).

Conclusions

Development of the FORTRAN 77 computer program that comprises the simulation model produced by PMDS predates the pest model development system by several years. Therefore, the code has already been applied for the description of a variety of insect systems. It is presently in use as the basis for simulation of mountain pine beetle populations in the western United States (Logan & Amman 1986; J.A.L., unpublished data); sitona weevil in New Zealand (Logan et al. 1985); Colorado potato beetle in Massachusetts (Voss & Ferro, personal communication) and New York (Naranjo & Shoemaker, personal communication); and spider mite–phytoseiid interactions in New Zealand (Hayes 1986) and Nebraska (Berry & Holtzer, personal communication). From these applications, it appears that the model structure is basically sound

and flexible enough to represent a wide variety of life systems (e.g., univoltine beetles as well as broadly overlapping generations of spider mites).

Experience for evaluation of PMDS itself is far more limited than for the produced model. However, when compared to my performance at the same task, the developmental rate model selection and curve fitting capabilities of PMDS are quite good. In fact, for some applications, I judge performance of the system to be better than I would have done in similar circumstances because PMDS tries options I probably would have passed by.

Model development by PMDS is extremely convenient. Work that previously took weeks can now be accomplished in a matter of hours. Therefore, widespread application of systems like PMDS should advance the conceptual development of IPM modeling approaches because of the convenience for modeling a large variety of problems. Likewise, this capability provides for objective comparison of different modeling approaches, a step that is currently seldom taken (Hochberg et al. 1986). With the dramatic reduction in overhead of model development, cost-effective model comparisons can be made. Performance evaluation of the knowledge base and inference components of the system can occur concurrently with the application of the system. Through this experience, modeling techniques with poor performance records can be discarded and new approaches conveniently implemented and evaluated.

It has been highly instructional to think about my own thought process and to attempt the design of a program that, for a restricted problem domain, emulates my modeling expertise. I am confident that systems such as PMDS will soon be in general use by quantitative biologists. Just how far such systems can go toward achieving the goal of a true expert system (i.e., that it can be effectively used by nonmodelers) remains to be seen. Such an evaluation will be possible only after widespread program application by scientists with diverse backgrounds in a variety of settings. Toward this end, copies of PMDS can be obtained at duplication cost from the author.

Acknowledgment

The author gratefully acknowledges the critical review and comments of D. L. Bartos, M. C. Coughenour, A. R. Grable, R. D. Moon, R. L. Olson, W. G. Ruesink, and two anonymous reviewers. Incorporation of many of their suggestions significantly improved an earlier version of this manuscript. Work reported in this manuscript was supported by a Cooperative Research Agreement between Colorado State University and the Mountain Pine Beetle Project, Forest Service Intermountain Research Station, U.S. Department of Agriculture, Ogden, Utah.

Appendix I. Parameter Estimation

Parameter estimation procedures for the curves shown in Fig. 3 are summarized for each curve. Curve fitting is attempted by the parameter estimation demon on command from the PMDS executive program.

Linear. The linear relationship is given in slope, x intercept form as,

$$r(T) = \rho(T - T_b), \qquad \text{(A1.1)}$$

where the slope (ρ) and x intercept (T_b) are estimated from standard linear regression procedures.

Exponential. The exponential curve is given as,

$$r(T) = \psi e^{\rho T}. \qquad \text{(A1.2)}$$

Parameter estimation is accomplished by first transforming the observed developmental rates to linear form by $r_i^T = \ln(r_i)$ for the r_i observed rates, and transformed temperature (T^T) computed as $T^T = T_i - T_{\min}$, where $T_{\min}$ is the minimum observed temperature. Parameters are then estimated from the linear regression of r_i^T on T_i^T.

Exponential T_b. An exponential form, modified for a low temperature threshold, is given by,

$$r(T) = e^{\rho(T - Tb)} - 1.0. \qquad \text{(A1.3)}$$

Parameter estimation is accomplished by first transforming Equation A1.3 to a linear form $r_i^T = \ln(r_i + 1.0)$, and regressing r_i^T on T_i.

Logan T_b. Functional form is given in Logan et al. (1979) as,

$$r(T) = \psi(e^{\rho T} - e^{-\tau}), \qquad \text{(A1.4)}$$

for $T = T_i - T_b$ and $\tau = T/\Delta T$. Parameters are estimated from nonlinear least squares regression, with initial estimates of: $\psi = T_{\min}$; ρ initialized as the slope of the line connecting the ordered pairs (T_n, y_1), (T_{n-2}, y_2), where $y_j = \ln(r_j/\psi)$; $\Delta T = T_3 - T_1$; and $T_b = T_{\min}$.

Stinner. The functional form for this model is given in Stinner et al. (1975) as,

$$r(T) = \begin{cases} T_0/(1.0 + e^{k_1 + k_2 T}) & \text{for } T \le T_0 \\ T_0/(1.0 + e^{k_1 + k_2(2T_0 - T)}) & \text{for } T > T_0 \end{cases} \qquad \text{(A1.5)}$$

where T_o is optimum temperature and k_1 and k_2 are empirical constants. Suggested starting values for nonlinear least squares parameter estimation are given in Stinner et al. (1975) as k_1 and k_2 from the slope and intercept respectively of the line connecting the ordered pairs (T_{min}, y_1) and (T_{max}, y_2), where $y_1 = \ln[(T_o/r_{\min}) - 1.0]$ and $y_2 = \ln[(T_o/r_{\max}) - 1.0]$, and $T_o = T_{rmax}$, for T_{rmax} the temperature at which occurred the maximum developmental rate.

Logan1. The functional form for this curve is obtained from Logan et al. (1976, equation 6) as,

$$r(T) = \psi(e^{\rho T} - e^{\rho \tau}) \qquad \text{(A1.6)}$$

for $\tau = (T_{\max} - T)/\Delta T$ where T is in degrees above an arbitrary base temperature ($T_{\min}$).

Initial conditions for nonlinear least squares parameter estimation are obtained by: $\psi = r_{\min}$; $\rho = \ln[(r_{\max} - r_1)/(T_{rmax} - T_1)]$; $T_M = T_{\max} - T_{\min}$; and $\Delta T = T_{\max} - T_{rmax}$.

A

DATA - m. p. b. Egg Developmental Rates	
Temperature	Rate
8.	.02966
10.	.03471
12.5	.04995
15.	.0767
17.5	.1074
15.	.07039
20.	.1353
22.5	.144
25.	.1807
30.	0.

B

DOMAIN KNOWLEDGE SUMMARY	
Canonical Attribute	Score
T_b	0
Non-linear	10
Rate Decline	10
Asymmetry	10
T_m	10

C

RANKING			
Initial		Final	
Fct.	Score	Fct.	Score
logan 1	40.	logan 1	40.
logan 2	40.	logan 2	40.
Type III	38.5	Type III	38.5
stinner	30.	stinner	30.
exp.	10.	exp.	10.
exp. Tb	10.	exp. Tb	10.
logan Tb	10.	logan Tb	10.
linear	0.	linear	0.

D

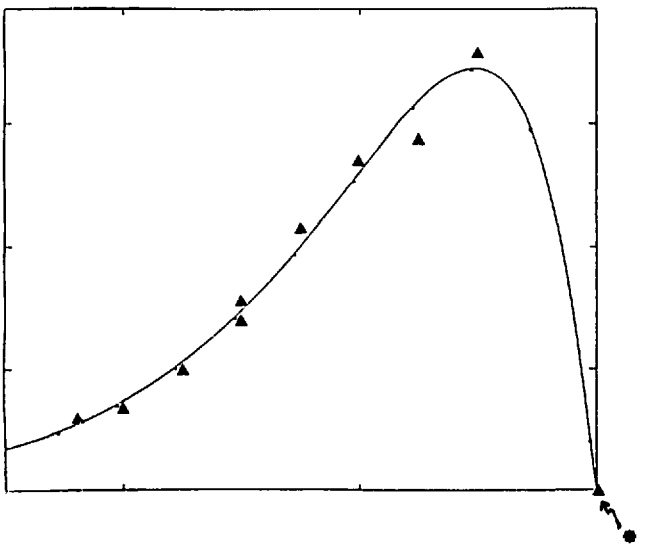

Fig. A2.1. Summary produced by the report writing demon of PMDS for the full data set from Logan & Amman (1985). (A) Data. (B) Summary table of the relative weights assigned to ecological parameters. (C) Initial and final ranking of potential developmental rate functions. (D) Plot of the best fit determined by PMDS.

Logan 2. The functional form is from Logan et al. (1976, equation 10), and is given by,

$$r(T) = \alpha[(1 + \kappa e^{-\rho T})^{-1} e^{-\tau}] \qquad \text{(A1.7)}$$

for T above an arbitrary base temperature $T_{\min}$, α and κ are empirical constants, and ρ and τ are defined as in Logan1.

Estimates of initial parameter are obtained as; $\alpha = r_{\max} + 0.2 \cdot r_{\max}$; $\kappa = [\alpha/(r_1 - 1.0)] \cdot 2.0$; $\rho = -\ln(0.2/\kappa)/(T_{r\max} - T_1)$; $T_M = T_{\max} - T_{\min}$; and $\Delta T = (T_{\max} - T_{r\max})/2.0$.

Type III. The functional form for this curve is found in Hilbert & Logan (1983), and is given by,

$$r(T) = \psi\left[\frac{T^2}{T^2 + D^2} - e^{-\tau}\right], \qquad \text{(A1.8)}$$

A

DATA - m. p. b. Egg Developmental Rates	
Temperature	Rate
8.	.02966
10.	.03471
12.5	.04995
15.	.0767
17.5	.1074
15.	.07039
20.	.1353
22.5	.144
25.	.1807

B

DOMAIN KNOWLEDGE SUMMARY	
Canonical Attribute	Score
T_b	0
Non-linear	10
Rate Decline	10
Asymmetry	10
T_m	10

C

RANKING			
Initial		Final	
Fct.	Score	Fct.	Score
logan 1	40.	N.A.	N.A.
logan 2	40.	N.A.	N.A.
Type III	38.5	N.A.	N.A.
stinner	30.	stinner	30.
exp.	10.	exp.	10.
exp. Tb	10.	exp. Tb	10.
logan Tb	10.	logan Tb	10.
linear	0.	linear	0.

D

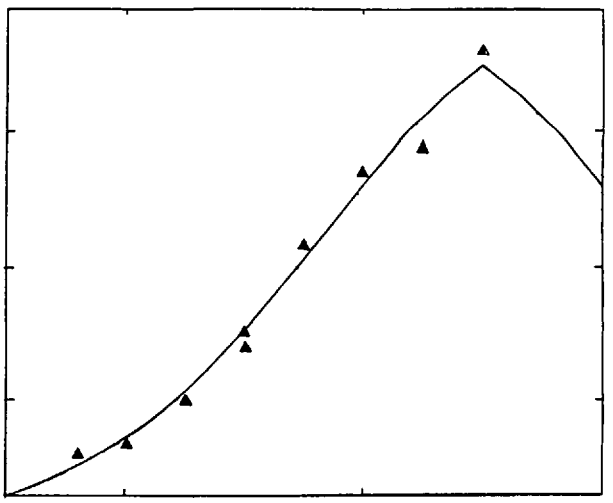

Fig. A2.2. Summary tables produced when the last datum from the previous data set is ignored. See text for further explanation.

for $\psi = r_{max}$, D an empirical constant, and τ defined as in Logan1.

Initial conditions for nonlinear least squares parameter estimation are obtained by: $\psi = 3.0 \cdot r_{max}$; $D = 2.0 \cdot (T_{max} - T_{min})$, and τ as in Logan1.

Appendix II. Example Problems

This Appendix contains example problems using data from an ongoing modeling project on the mountain pine beetle (Logan & Amman 1986). The

```
                    •
                    •
                    •

SUMMARY: Attempted fit of logan tm1      model

  A. Initial parameter values
       p( 1) =  2.9660000e-02
       p( 2) =  1.0629640e-01
       p( 3) =  2.2000000e+01
       p( 4) =  5.0000000e+00

  B. Initial Error Sum of Squares
         Error SS =  2.7476273e-02

  C. Final parameter values
       p( 1) =  5.4190651e-02
       p( 2) =  1.7959043e-01
       p( 3) =  2.2002514e+01
       p( 4) =  4.7109036e+00

  D. Final Error Sum of Squares
         Error SS =  4.0813157e-04


  E. The adjusted Coeficient of Determination =  9.7729099e-01


SUMMARY: Attempted fit of logan tm2      model

  A. Initial parameter values
       p( 1) =  2.1684001e-01
       p( 2) =  1.2621714e+01
       p( 3) =  2.4381508e-01
       p( 4) =  2.2000000e+01
       p( 5) =  2.5000000e+00

  B. Initial Error Sum of Squares
         Error SS =  1.6415570e-03
*** Illegal operand
*** Execution terminated

I have unsuccessfully attempted to fit the logan tm2 model to this data
set. Since I have been able to fit at least one other model with an equally
high score, I will proceed with the sequential evaluation of potential models


Are you interested in seeing a plot of logan tm1       model to data?

                    •
                    •
                    •
```

Fig. A2.3. PMDS program output for the problem summarized in Fig. A2.1. Note that the parameter estimation demon failed in its attempt to fit the Logan2 model; however, PMDS managed to accommodate this terminal computational error without catastrophic consequences.

→

Fig. A2.4. PMDS program output for first-instar larvae of the mountain pine beetle. Note that the two models with the highest ranking were successfully fitted to the data, and of these, the Logan2 model was determined to be superior on the basis of adjusted R^2 values.

```
SUMMARY: Attempted fit of logan tm1      model

  A. Initial parameter values
       p( 1) =  9.7864801e-03
       p( 2) =  1.7975035e-01
       p( 3) =  2.2500000e+01
       p( 4) =  2.5000000e+00

  B. Initial Error Sum of Squares
         Error SS =  1.6513248e-01

  C. Final parameter values
       p( 1) =  6.2838241e-02
       p( 2) =  2.0785698e-01
       p( 3) =  2.4059172e+01
       p( 4) =  4.2984319e+00

  D. Final Error Sum of Squares
         Error SS =  4.7634020e-03


  E. The adjusted Coeficient of Determination =  8.6819601e-01


SUMMARY: Attempted fit of logan tm2      model

  A. Initial parameter values
       p( 1) =  4.2766082e-01
       p( 2) =  8.5398293e+01
       p( 3) =  3.0283818e-01
       p( 4) =  2.2500000e+01
       p( 5) =  1.2500000e+00

  B. Initial Error Sum of Squares
         Error SS =  8.9175545e-02

  C. Final parameter values
       p( 1) =  6.2223601e-01
       p( 2) =  6.6121185e+01
       p( 3) =  3.0989936e-01
       p( 4) =  2.5409197e+01
       p( 5) =  4.4410644e+00

  D. Final Error Sum of Squares
         Error SS =  1.6235402e-03


  E. The adjusted Coeficient of Determination =  9.1015297e-01

Of the several models I attempted to fit to this data set, I have chosen the
logan tm2  model as the best fit.  This decision was based on the criteria
of maximum adjusted coeficient of determination (Kvalseth. 1985. Am. Stat. 39)

Are you interested in seeing a plot of logan tm1       model to data?
```

A

DATA - m. p. b. L1 Developmental Rates	
Temperature	Rate
5.	.0097
10.	.0641
15.	.1140
20.	.3345
25.	.3563
27.5	.2656

B

DOMAIN KNOWLEDGE SUMMARY	
Canonical Attribute	Score
T_b	0
Non-linear	10
Rate Decline	10
Asymmetry	7
T_m	3

C

RANKING			
Initial		Final	
Fct.	Score	Fct.	Score
logan 1	30.	logan 1	30.
logan 2	30.	logan 2	30.
Type III	28.95	Type III	28.95
stinner	20.	stinner	20.
exp.	10.	exp.	10.
exp. Tb	10.	exp. Tb	10.
logan Tb	10.	logan Tb	10.
linear	0.	linear	0.

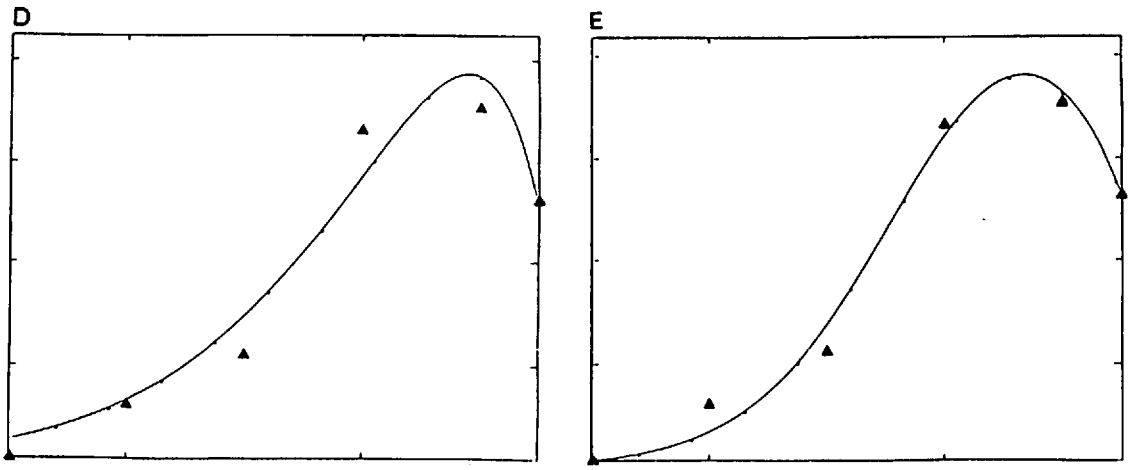

Fig. A2.5. Summary produced by the report writing demon for the first larval instar of the mountain pine beetle.

purpose of including these examples is twofold. First, they demonstrate some of the inferential capabilities of the program; and second, they illustrate the report writing capabilities of PMDS.

In the first example (Fig. A2.1), the full data set from Logan & Amman (1986) for mountain pine beetle egg development is used for parameter estimation. In particular, note that 30°C is an estimate of the thermal maximum for egg development. This single datum plays a crucial role in parameter estimation, because it is the only observation above the apparent high temperature threshold, and high temperatures were indicated to be the most ecologically important (Fig. A2.1, B). Fig. A2.2 illustrates the importance of this datum. In the example that produced this output, the same data set was used for parameter estimation, exclusive of the single observation above thermal maximum. Note that because this observation is lacking, the data set no longer satisfies the constraints required by the three models most capable of describing high temperature decline in development rates. However, the

same relative importance has been assigned to high temperature phenomena (Fig. A2.2, B). PMDS attempts to do the best it can to reconcile this apparent dilemma. It is smart enough to realize that if it assumes the best estimate of T_o is the temperature at which the maximum observed developmental rate occurred (T_{rmax}), then it can fix T_o in the Stinner function to T_{rmax} and subsequently invoke the parameter estimation demon to determine k_1 and k_2. This it proceeds to do, with results shown in Fig. A2.2, D.

Problem 3 (Fig. A2.3) is an example of PMDS action when the parameter estimation demon failed in its attempt to fit a model. In this example, the mountain pine beetle egg data set was used once again. PMDS attempted to fit both the Logan1 and Logan2 models. Although the initial sum-of-squares for the Logan2 model was less than that for the Logan1 model (indicating a potentially better fit), the nonlinear least squares routine crashed during execution. The resulting fit of the Logan1 model is shown in Fig. A2.1, D. Incidentally, the only way I was able to successfully fit the Logan2 model was by allowing the nonlinear least squares routine to treat T_h, which is arbitrary and therefore set to a constant (T_{min}) by the parameter estimation demon, as a parameter. This is an example of intuition, a characteristic that may always separate real from machine expertise (Dreyfus & Dreyfus 1986).

The final example (Fig. A2.4) illustrates a successful attempt by PMDS to fit more than one model to a data set (in this case mountain pine beetle first-instar larvae (J.A.L., unpublished data). This action, as summarized by the report writer, is shown in Fig. A2.5.

References Cited

Coulson, R. N. & M. C. Saunders. 1987. Computer-assisted decision-making as applied to entomology. Annu. Rev. Entomol. 32: 415–437.

Crecine, J. P. 1986. The next generation of personal computers. Science 231: 935–942.

Curry, G. L., R. M. Feldman & K. C. Smith. 1978. A stochastic model of a temperature-dependent population. Theor. Popul. Biol. 13: 197–213.

Dreyfus, H. & S. Dreyfus. 1986. Why expert systems do not exhibit expertise. IEEE-Expert 1: 86–90.

Getz, W. M. & A. P. Gutierrez. 1982. A perspective on systems analysis in crop production and insect pest management. Annu. Rev. Entomol. 27: 447–466.

Gregg, P. 1982. A simulation model of the development of *Chortoicetes terminifera* (Orthoptera: Acrididae) under fluctuating temperatures. Proceedings Australasian Conference on Grassland Invertebrate Ecology 3: 117–125.

Gutierrez, A. P., J. W. Baumgaertner & C. G. Summers. 1984. Multitrophic models of predator-prey energetics. Can. Entomol. 116: 923–963.

Hayes, A. 1986. Simulation models of the European red mite, *Panonychus ulmi* and its predator *Typhlodromus pyri* in New Zealand apple orchards. Ph.D. dissertation, University of Auckland, Auckland, New Zealand.

Hilbert, D. W. & J. A. Logan. 1983. Empirical model of nymphal development for the migratory grasshopper, *Melanoplus sanguinipes* (Orthoptera: Acrididae). Environ. Entomol. 12: 1–5.

Hochberg, M. E., J. Pickering & W. M. Getz. 1986. Evaluation of phenological models using field data: case study for the pea aphid, *Acyrthosiphon pisum*, and the blue alfalfa aphid, *Acyrthosiphon kondoi* (Homoptera: Aphididae). Environ. Entomol. 15: 227–231.

Holt, H. A. 1985. Computers in production agriculture. Science 228: 422–427.

Klahr, P. & D. A. Waterman. 1986. Expert systems: techniques, tools, and applications. Addison-Wesley, Reading, Mass.

Kvalseth, T. O. 1985. Cautionary note about R^2. Am. Statistician 39: 279–285.

Lamb, R. J., G. H. Gerber & G. F. Atkinson. 1984. Comparison of developmental rate curves applied to egg hatching data of *Entomoscelis americana* Brown (Coleoptera: Chrysomelidae). Environ. Entomol. 13: 868–872.

Lemmon, H. 1986. COMAX: an expert system for cotton crop management. Science 233: 29–33.

Logan, J. A. 1979. SIMBUG: a pest management simulation language I. Maintenance and reference manual. Melanderia 31: 1–17.

Logan, J. A. & G. D. Amman. 1986. A distribution model for egg development in mountain pine beetle. Can. Entomol. 118: 361–372.

Logan, J. A. & D. W. Hilbert. 1983. Modeling the effects of temperature on arthropod population systems, pp. 113–112. *In* W. K. Lauenroth, G. V. Skogerboe & M. Flug [eds.], Analysis of ecological systems: state-of-the-art in ecological modeling. Elsevier, Amsterdam.

Logan, J. A., D. J. Wollkind, S. C. Hoyt & L. K. Tanigoshi. 1976. An analytic model for description of temperature dependent rate phenomena in arthropods. Environ. Entomol. 5: 1133–1140.

Logan, J. A., R. E. Stinner, R. L. Rabb & J. S. Bacheler. 1979. A descriptive model for predicting spring emergence of *Heliothis zea* populations in North Carolina. Environ. Entomol. 8: 141–146.

Logan, J. A., E. R. Frampton & S. L. Goldson. 1985. A phenological model for *Sitona discoideus* Gyllenhal (Coleoptera: Curculionidae) and its application in Canterbury. Proceedings Australasian Conference on Grassland Invertebrate Ecology 4: 173–179.

O'Connor, D. E. 1984. Using expert systems to manage change and complexity in manufacturing, pp. 149–159. *In* W. Reitman [ed.], Artificial intelligence applications for business. Ablex, Norwood, N.J.

O'Neill, R. V. 1979. Transmutation across hierarchical levels, pp. 59–78. *In* G. S. Innis & R. V. O'Neill [eds.], Systems analysis of ecosystems. International Co-op. Publ. House, Fairland, Md.

Powell, M. J. D. 1965. A method for minimizing a sum of squares of nonlinear function without calculating derivatives. Comput. J. 7: 303–307.

Régnière, J. 1984. A method of describing and using the variability in development rates for the simulation of insect phenology. Can. Entomol. 116: 1367–1376.

Ruesink, W. G. 1976. Status of the systems approach to pest management. Annu. Rev. Entomol. 21: 27–44.

Shaffer, P. L. & H. J. Gold. 1985. A simulation model of population dynamics of the codling moth, *Cydia pomonella*. Ecol. Modell. 30: 247–274.

Sharpe, P. J. H., G. L. Curry, D. W. DeMichele & C. L. Cole. 1977. Distribution model of organisms development times. J. Theor. Biol. 66: 21–28.

Stinner, R. E., G. D. Butler, J. S. Bacheler & C. Tuttle. 1975. Simulation of temperature dependent development in population dynamic models. Can. Entomol. 197: 1167–1174.

Stone, N. D., R. N. Coulson, R. E. Frisbie & D. K. Loh. 1986. Expert systems in entomology: three approaches to problem solving. Bull. Entomol. Soc. Am. 32: 161–166.

Wagner, T. L., H. Wu, P. J. H. Sharpe, R. M. Schoolfield & R. N. Coulson. 1984a. Modeling insect development rates: a literature review and application of a biophysical model. Ann. Entomol. Soc. Am. 77: 208–225.

Wagner, T. L., H. Wu, P. J. H. Sharpe & R. N. Coulson. 1984b. Modeling distributions of insect development time: a literature review and application of the Weibull function. Ann. Entomol. Soc. Am. 77: 475–483.

Wagner, T. L., H. Wu, R. M. Feldman, P. J. H. Sharpe & R. N. Coulson. 1985. Multiple-cohort approach for simulating development of insect populations under variable temperatures. Ann. Entomol. Soc. Am. 78: 691–704.

Waterman, D. A. 1986. A guide to expert systems. Addison-Wesley, Reading, Mass.

Welch, S. M., B. A. Croft, J. F. Brunner & M. F. Michels. 1978. PETE: an extension phenology modeling system for management of multi-species pest complex. Environ. Entomol. 7: 487–494.

Received for publication 24 March 1987; accepted 28 October 1987.