UNIVERSITY OF WISCONSIN — MADISON
DEPARTMENT OF STATISTICS

Event-Driven Competing Risks

Bland Ewing[1]

Brian S. Yandell[2]

James F. Barbieri[3]

Robert F. Luck[4]

[1] Department of Entomology (Ret), University of California, Berkeley, CA
[2] Departments of Statistics and Horticulture, University of Wisconsin-Madison, WI
[3] Department of Entomology, University of California, Riverside, CA, and
Advanced Systems Development, NAWC-WPNS, China Lake, CA
[4] Department of Entomology, University of California, Riverside, CA

Technical Report # 1032

2 February 2001

Department of Statistics
University of Wisconsin-Madison
1210 West Dayton Street
Madison, WI 53706-1685

Phone: 608/262-2598
Fax: 608/262-0032
Internet: yandell@stat.wisc.edu

Event-Driven Competing Risks

Bland Ewing[1], Brian S. Yandell[2], James F. Barbieri[3], and Robert F. Luck[4]
Technical Report 1032, Department of Statistics, UW-Madison

Abstract

The non-homogeneous Poisson process with a competing risk structure can be used to simulate the interacting stochastic lives of individuals in an ecological community. Such a technique can help quantify the relationships among observed behaviors of individuals and describe the resulting coupling between interacting populations defined in state space descriptions commonly used in population biology. The event structure provides the dynamics that drives time, rather than the usual time-driven stochastic dynamic programming. We illustrate the ideas with the California Red Scale-*Aphytis* host-parasitoid system, although the method has wider applicability.

[1] Department of Entomology (Ret), University of California, Berkeley, CA
[2] Departments of Statistics and Horticulture, University of Wisconsin-Madison, WI
[3] Department of Entomology, University of California, Riverside, CA, and
Advanced Systems Development, NAWC-WPNS, China Lake, CA
[4] Department of Entomology, University of California, Riverside, CA

Theory of Competing Risks

The theory of competing risks arose initially in the study of mortality, with risks of death "competing" for an individual's life. The theory applies equally to reliability of machines and the study of illness processes, and is intimately connected with the study of life tables (Chiang 1968). It provides a mechanistic way to model how critical events evolve moment to moment, influenced by the surrounding environment. Competing risks and life table methods have been used regularly in ecological modeling (Caswell 1989). We reconsider the implications of theoretical developments in competing risks on the future of ecological modeling.

An individual alive at some time t has several "choices" about future events. It might die, for one of several possible reasons--the classical competing risks framework. Or it might eat, or reproduce, or migrate to another locale. The chance of the jth type of event occurring during the "instant" t is proportional to the competing risk, or hazard $m_j(t)$. If one assumes that at most one event can occur in any instant, then the risks add up and the chance of *any* event in an instant is $m(t) = \sum m_j(t)$.

One can model the effect of the environment and an individual's life stage on these competing risks in a variety of ways. Keyfitz (1966, 1968) and others developed complicated methods to incorporate birth and death processes into life tables indexed by age and gender. Chiang (1968, 1972) provided a simple unifying probability framework for life tables and competing risks, which tacitly assumes that events of interest occur at particular ages (e.g. every 5 years for humans) in order to make the process identifiable. The proportional hazards model (Cox 1972) and accelerated lifetimes model (Viertl 1988; Clarotti and Lindley 1988) have allowed researchers to consider smoothly varying competing risks. Recent efforts allow interdependence among risks (Fine

1999). This literature has been largely focused on inference for competing risks.

Fix and Neyman (1951) introduced an illness-death process, generalized by Chiang (1968), in which an individual may move between healthy and sick states, continually exposed to competing risks of death. However, one cannot identify competing risks with a single measurement per individual even if the illness is progressive (Clifford 1977). Yandell (1982) showed that independent competing risks cannot be identified even with periodic measurements of an individual's state unless these measurements correspond to the times of transition between states. Further, one can only identify the cumulative competing risks, or mean value functions $M_j(t) = \int m_j(t)dt$, rather than the competing risks $m_j(t)$ themselves

An alternative way to model competing risks involves potential lifetimes T_j for each type of risk j (see David 1974). Suppose several future events indexed by j are competing to become the next event for an individual. The observed time to next event T would be the minimum of these potential event times. Hence, the chance that no event occurs before time t is

$$\Pr\{ T > t \} = \Pr\{ \min(T_j) > t \} = \prod \Pr\{ T_j > t \} = \prod \exp(- M_j(t)) .$$

The product $\prod$ is justified if the potential event times, or equivalently the competing risks, are independent of one another. Tsiatis (1975) showed that without the assumption of independence the potential lifetimes T_j, and hence the sub-probabilities $\Pr\{ T_j > t \}$and mean value functions $M_j(t)$, cannot be uniquely determined from the data. Further, it is impossible to investigate the assumption of independence with these data. As a result, these potential lifetimes have fallen out of favor in the competing risks literature.

It is useful to examine how competing risks might depend on one another. The key issue is how competing risks $m_j(t)$ change over time. These changes may be due to discrete changes in state of other features of the environment--including states of other individuals--or to gradually changing environmental conditions. If one observes the changes in the environment, then the competing risk structure could be altered after each event. Similarly, predictable gradual changes in the environment could be readily incorporated into the competing risks structure. Both of these approaches have been used extensively in inference for hazards $m_j(t)$. We propose to build them into a modeling system based on potential lifetimes T_j and mean value functions $M_j(t)$. While we agree that inference cannot be based on these potential event times, they are extremely useful for simulation models of complex living systems.

Stochastic Models in Ecology

Stochastic models in ecology are designed to study ecological systems by simulating the underlying processes and then studying multiple realizations of a simulation model. We focus here on models of life history events. Interestingly, the life table approximation of dividing time into discrete "quanta" migrated early into stochastic models in ecology. The Lotka-Volterra model is typically developed in this way. Modern simulation studies have advanced by considering smaller and smaller time increments, but within this same fixed time step framework (Mangel and Clark 1988; Hutchinson and McNamara 2000).

The shortcoming of this approach is that while one must have finer scale time increments in order to capture more intricate events, more and more time is spent simulating no activity. Alternatively, one can develop models based on the actual time of transitions. The difficulty with

this shift in perspective is that events for individuals are no longer synchronized. Further, generations may not be synchronized, making life table summaries problematic.

The development below of a simulation structure for competing risks in a biological system is built upon the concept of potential lifetimes using the cumulative risks $M_j(t)$ as a basic building block. This approach necessitates detailed knowledge of the ecosystem under study, which is precisely what we want. The purpose here is to provide a framework for biologists to incorporate great detail about known and suspected aspects of an ecosystem. Biologists inherently recognize that their knowledge is incomplete and may even be wrong in part. However, it is extremely difficult for them to test their hypotheses about ecosystem-level and population-level properties that may depend on processes that affect individuals.

Work in complexity (Langton 1986) suggests that higher level structure can emerge from local structure. These models are promising, but to date suffer from the same quantization problem found with life-table derived methods.

Considerations for Living Systems

Consider a biological system as a stochastic process X that progresses through time t from one event to the next. An event is defined as a significant biological change that can be marked and counted, resulting in an instantaneous state change. That is, mathematically at a time t the process changes state to $X(t) = s$. The state s, from some space S of possible states, may be rather complicated, but it needs to be specified in some explicit or implicit fashion. It is typically assumed that the probability of two or more events occurring simultaneously is negligible. This allows probabilities for events to be characterized by a competing risk system. At any point in time t, the possible events are mutually exclusive, and hence their instantaneous probabilities, or risks $m_j(t)$, add up to the chance $m(t)$ for any event at t. These competing risks $m_j(t)$ may change over time, but are assumed to be fairly well-behaved. There is a companion process $\{N(t), t\geq 0\}$ that counts the number of events of the stochastic process $\{X(t), t\geq 0\}$. The mean number of events by time t, $E[N(t)]=M(t)$, is the integral of $m(t)$ from time $t=0$, with $M(0)=0$. $M(t)$ is usually assumed to be continuous and strictly increasing with derivative $m(t)$, or more generally to be right-continuous and non-decreasing with derivative $m(t)$ at all but a countable number of time points. The process N is a non-homogeneous Poisson process, as are the counting processes N_j for each type of risk.

A non-homogeneous Poisson process can be described as counting process with a time-varying intensity that satisfies certain properties, such as independent increments between distinct time intervals. The use of the non-homogeneous Poisson process to simulate a biological system forces some mathematical requirement onto the system not inherent in the biology. For one thing, it assumes that events can be identified, and thus counted. A straightforward use might be to include the whole system in X, but that would lead to a very complicated competing risk system and state space. Another approach involves letting X be a set of interconnected stochastic processes X_i indexed by individual i. This has typically been done with either independent processes or with limited interaction, e.g. based on nearest neighbors. The difficulty with these approaches is that they already have removed most of the biology, proscribing the types of interactions among individuals in an environment. We propose another way.

Stochastic processes are usually described as proceeding from time to time. This has led to a whole simulation industry in stochastic dynamic programs (Hutchinson and McNamara 2000), stepping through finer and finer time increments to approach "reality". We find it useful to think

instead of a process moving from event to event. In this way, time is defined implicitly through the sequence of realized events. From this perspective, imagine a stochastic process as a sequence of scheduled events, beginning in state s_0 at time t_0=0, with events, or state transitions, $s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow$ … that imply event times $t_1 < t_2 <$ …. If the structure of the process does not change, events can be scheduled as far forward as desired. For instance, one could schedule the next event for every individual in a population, and every event for each individual.

The investigator must decide up front what aspects of a biological system will be studied, based on measurable events and the focus of scientific inquiry. Once these have been determined, events can be defined within the context of the resolution and span of the model. Resolution is defined to be the smallest increment of time that contributes useful biological information to a simulation. Events over smaller time scales are assumed to occur instantaneously. Negligible knowledge can be gained about the process under immediate study by going to finer time scales, while such an effort will only increase the cost of simulation. Span is defined to be the largest amount of time the model can encompass. Aspects of a living system that occur at longer time intervals than the span of a model are considered to be essentially constant, or slowly varying in a smooth fashion, for the purposes of simulation. For example, with a resolution of one hour, an *Aphytis* feeding on a California red scale would appear to be instantaneous. On the other hand, if the same model has a span of six months, then a mature orange fruit could be considered static, not changing appreciably as a substrate for red scale.

Span and resolution of space must be established in tandem with span and resolution of time to be meaningful for a particular simulation. Further, spatial span and resolution determine the detail required for spatial interactions among individuals. It is important to note that if the resolution and span are changed then the model of the biological system will also change. In fact, it may be appropriate to model a biological system at several different spatial and temporal spans and resolutions.

In a biological system it is highly unlikely that the non-homogeneous Poisson process will be satisfied over even a few events, since the occurrence of an event will modify the probability of future events. Because of the requirements of the non-homogeneous Poisson process the structure of the event space must be carefully defined. The requirements of no events at time t=0 and of the probability of simultaneous events being vanishingly small put severe limitations on the biological system that can be suitably modeled using the non-homogeneous Poisson process. For instance, if these limitations are not addressed, then such processes as death due to predation and birth cannot be simulated. To avoid this limitation three special types of events are defined:

1. Future events are events that are scheduled to occur at some future time, such as reproduction, based on the competing risks system.
2. Immediate events handle multiple events that are not resolved individually at a particular resolution in the simulation. Such events--the birth of multiple offspring--appear to be coincident.
3. Pending events are events whose occurrence depend conditionally on other events or particular states of the model. This is especially designed to deal with predators and prey, or parasitoids and their hosts. It can also address events that depend on environmental changes such as fire.

Reproduction is a future event, which may have as its consequence the birth of a certain number of individuals, that for the purpose of the model may be considered as one immediate event. For instance, the birth of red scale crawlers would be considered to be instantaneous when viewed over the model span of six months with a model resolution of one hour. Another simulation at a finer time resolution (seconds) and span (days) might simulate the birth process in great detail. At birth, one or more new individuals enter the simulation, which modifies other future or pending events due to the search for food, etc.

Death may be a pending event, since it depends on locating and consuming an adequate amount of resources while avoiding predation. Notice that the focus is on "finding food" rather than "food availability"; it is not enough to model the presence of a resource without modeling the process of obtaining it. Pending events depend conditionally on other events and on particular states of the model. For instance, events that depend on cooperation of individuals would be a pending event. A pride of lionesses hunts as a group, depending on cues among members of the pride to successfully bring down a prey. Movements of one lioness depend on the actions of others in the vicinity. Pending events must be scheduled together. It is possible to chain such events, scheduling one, then the next conditional on the first, etc. For instance, one lioness may isolate a particular antelope, leading to two others charging in to attempt a kill. Another example of a pending event is germination of fire-sensitive seeds. One must first schedule a fire, and then schedule germination conditional on fire scorching seeds at a particular location. Pending events may have to be repeatedly rescheduled, while independent future events are only infrequently modified.

In a biological system it is highly unlikely that the assumptions of a non-homogeneous Poisson process will be satisfied over even a few events, since the occurrence of an event modifies the probability for future events. For this reason, the structure of the event space must be carefully defined in a manner to allow changes over time. The simulation of the biological system can be viewed as a sequence of non-homogeneous Poisson processes that are separated by future events. When a future event occurs, the simulation is stopped. Action is taken on any immediate events induced by the future event, and any pending events are modified as necessary. Once pending or immediate events are processed, the non-homogeneous Poisson structure is reconstructed containing only future events from the current time onward. The simulation proceeds to the next future event, at which time the simulation is again stopped.

With this in mind, it is only necessary to construct a new non-homogeneous Poisson structure containing future events. The simulation proceeds to the next future event, the one with the shortest time to occurrence. Suppose that a future event is scheduled to occur at some time t_i and that other future events are scheduled at t_{i+1}, t_{i+2}, …, after t_i. When the simulation reaches t_i, the future event is realized, immediate events are initiated and any pending events may be modified. This may change the competing risk structure of the model, which in turn may cause the non-homogeneous Poisson process to fail. The stochastic process must be rebuilt at this point, with the events scheduled at t_{i+1}, t_{i+2}, …, still being future events unless any immediate or pending event modified or rejected one or more of them.

The event structure is induced on the biological system by the requirements of the non-homogeneous Poisson process. Immediate or pending events can change the underlying structure of the model, requiring the rebuilding of the Poisson process, while future events change time in the model and control the competing risk sequencing. With this event structure, the biological system can be simulated by alternating between a sequence of future events satisfying a non-homogeneous Poisson process and a sequence of immediate or pending events that reconstruct that process. In most

situations, only a few competing risks are altered by immediate or pending events, requiring only modest changes. However, the competing risks for future events now depend on the past history of events. It is no longer sufficient to describe the process as simply a function of time t. Nevertheless the stochastic process is still predictable and hence is still well defined. This is dramatically different from the vast literature on ecological simulations that proceed from time to time waiting for events to occur. Here, time is a function of events, as opposed to events being a function of time.

Event Structure for an Individual

The structure of an individual can be divided into two interrelated parts. The first part consists of "static" properties that change as a function of time and are affected by the competing risks structure for future events, but are not directly a part of that competing risk system. Such properties as physical attributes and relationships with other individuals are included in this part. The second part of the total structure for an individual is the competing risk structure, or event structure. The event structure has within it the potential dynamics of each individual, and the dynamics of the simulation.

Viewed as a Poisson process over the time span $[a,b]$, a biological system may be represented by $\{X(t),\ t\in[a,b]\}$, in which time t marches over the span at a specified resolution. The values of the process X are states s in the state space S, and most of the time the state does not change.

However, the stochastic process under consideration is in reality a function of two arguments, time and the current state. When the state changes, it may alter the entire state space for the future of the stochastic process. One could imagine a progression of stochastic processes indexed by the current state,$\{X(t, s_j),\ t>t_j\ \}$, $j = 0, 1, \ldots$, with s_0 the initial state at $t_0=0$ and s_j the current state after the future event at time t_j. The realization of this stochastic process is equivalent to a sequence (t_0,s_0), (t_1,s_1), …, (t_j,s_j), …. While this can theoretically be accommodated within the framework of Markov processes if one extends the sample space to be arbitrarily large, it provides little insight into how to implement a simulation in practice.

For any one individual there are many events that can occur and hence there are many changes in the sample space during its lifetime. In a community of simulated individuals, the event structure would be frequently changing. Recasting the above stochastic process X for the whole community, it is clear that even for small simulations this structure can become cumbersome. At the occurrence of each event it would be necessary to rebuild the entire structure of the event space, which would be enormously expensive.

The following technique was developed to avoid this situation. Viewing the stochastic process as a biological system that changes with each new event, time T becomes an implicit random function of the current time and state and the next future event within the context of the current event structure. That is, if the current state is s_1 was realized at time t_1, then the future event $s_1\rightarrow s_2$ will happen at some random future time $t_2 = T(s_1\rightarrow s_2 \mid t_1,s_1)$. This property of the process allows a smooth transition between the time domain of the biological system and the event domain in modeling the biological system.

Measurements and the Sampling Process

If the biological system to be studied is sparse and satisfies the above assertions, then it is some sense "quantized" and is intrinsically dependent on measurements and the manner of sampling the biological system. If this modeling process is to be realistic, it is completely dependent on the data collected. Actually, model dependence on the sampling design may be more fundamental than is implied above. A member of a population is an integrated organism which functions as a unit, and this suggests that any attempt to explain the actions of that organism by analyzing its constituent will not lead to a complete description of the dynamics of that organism. In a sense, the reductionist techniques are limited in their applicability, hence one is faced with analyzing populations of whole organisms without further subdivisions. If a measurement is performed on a biological system, the researcher finds the members of that population in a number of particular, quantifiable states. The measurement process in some sense projects out the state of the system for that point in time. If the measurement is not performed, it is impossible to quantify the state of the population. In a sense, the totality of all of the potential states that the system may occupy are eliminated by the measurement process, leaving only the state that is measured.

We have approached the concept of a "measurement" from a mathematical point of view, and not from the point of view of a field biologist. The quality of what is called measurement is similar to the concept of a decomposition of a Hilbert space in which the measurement process and the observer are in some sense active contributors to the determination of the understanding of the system being studied. We are not proposing the existence of some sort of Heisenberg uncertainty principle of canonical variables. Rather we suggest that knowledge of the structure and dynamics of a particular system is uniquely determined by the measurement that a field biologist performs. Further the process of measurement and understanding of that system is itself a dynamic process. For the field biologist, the usefulness of any modeling technique is directly related to its ability to suggest clearer understanding of the data collected.

Time Depends on Events

The ethologist is faced with the task of defining probabilities for each observed event in such a way that those probabilities somehow reflect the observed data. In essence, this task involves the construction of life tables that provide links to age specific measured rates. Since events contain all the measured dynamics for the population, there is an advantage in transforming the simulation from a time domain to an event domain.

Let T be a random time to the next event for an individual. This time is assumed to have a left-continuous non-decreasing distribution function F defined on $[0,\infty)$

$$\Pr\{\ T \le t\ \} = F(\ t\ ) = 1 - \exp(\ -M(\ t\ ))$$

with the mean value function $M(t)$ being left-continuous, monotonic and non-decreasing, $M(t) = 0$ for $t \le 0$ but otherwise arbitrary. In general $F(\infty) < 1$, meaning that there is a finite probability that no event occurs. Scheduling this future event T can be viewed as drawing a random probability between 0 and 1 and mapping it to time rather than the other way around. The philosophical shift from marching through time to scheduling events by drawing a probability is crucial. By reformulating a time-dependent problem into an event-driven problem it is possible to carry the dynamics from

event to event using a competing risk structure. An event changes the state of the model at a future time t.

Now time becomes an implicit, dependent variable, and the event structure becomes the determining independent parameter. Processes that are continuous over a given span modify the scheduling of future events, but may not dynamically change the state s_j of the model. If the events are sparse, then a given simulation can be very efficient. The counting process assumes implicitly a categorization of the individuals into defined categories. However, for a biological system the categories must be arbitrarily defined.

Suppose one wished to model the structure and dynamics of a small population of bald eagles over a span of six years. One future event in such a study is reproduction, which usually occurs once a year. While one could set the time step at one year, this would not allow for other events such as hunting and migration. It would be more appropriate to consider a resolution of one day. A time-dependent simulation would require over 2000 daily increments per bald eagle, with most of the calculation being simple updates. In an event-driven simulation, it is possible to schedule the reproductive and other future events and proceed directly to from event to event when they reach the top of the event queue. Further, one can directly incorporate reproductive curves from field data. Finally, environmental effects such as pollution that can severely reduce the probability of *any* reproductive event, $F(\infty) < 1$, can be explicitly modeled.

While one may either specify the mean value function M or the distribution F, it seems more convenient to specify M as developed below. Consider drawing a random variate U, which is uniform over [0,1], and defining time T in terms of U as

$$M^{-1}(G(U)) = T$$

with $G(u) = -\log(1-u)$ and M^{-1} the inverse of M if it is continuous, or a suitably defined generalized inverse otherwise. Thus $M^{-1}(G(\cdot))$ maps from the [0,1] probability domain into the $[0,\infty)$ time domain. Figure 1 shows this mapping for a probability distribution F that has a plateau in the middle. The random variate $V = G(U) = M(T)$ has a standard exponential distribution

$$\Pr\{ V \le v \} = 1 - \exp(-v).$$

If the mean value function itself is the identity, $M(t)=t$, then $T=V$. Control over the shape of $M(\cdot)$ allows considerable flexibility to incorporate relevant knowledge of biological processes into the distribution of the scheduling time T. The mean value function M could be estimated from experimental data, or sketched based on partial knowledge and hunches.

Sampling a uniform U yields a future event time $M^{-1}(G(U)) = T$. On the other hand, sampling V avoids calculating logs, which are traditionally very expensive computationally. Thus, events may be scheduled in the following manner. Sample a standard exponential random variable V and construct the random time T as $M^{-1}(V)=T$, or $M(T)=V$. That is, $\Pr\{V \le M(t)\} = \Pr\{T \le t\} = F(t)$. Thus the mean value function $M(\cdot)$ transforms the exponential waiting time V based on constant risks to a biological time that may encompass the ongoing processes of an ecosystem.

However, this generic mean value function $M(\cdot)$ needs to be fine-tuned to each individual in a species, and to different species. Future event times may need to be adjusted based on individual histories and situations. In practice, many individuals may have similarly shaped mean value functions. Thus it is feasible to design a few such curves and then shift, stretch or otherwise modify

them to suit multiple needs. This can enhance the biologist's control over model simulations while keeping the decisions simple. The development below shows how this generic biological time can be easily modified to adjust to individual biological clocks.

Five-Dimensional Parameterization

For each individual *i* there exists a set of random potential times $\{T_{i\alpha}, \alpha = 1,2,\ldots\}$ when future events are predicted to occur. These random times are in general modified by changes in the population, changes in the environment, and changes in the individual over its life span. These changes may slow or delay the biological clock, or increase the risk of certain types of events. It is helpful to have an easily adaptable system to reschedule future event times. We have characterized this in terms of dispersion, location, intensity, truncation and rejection of events.

Modifications to the random times are handled using the following technique. A random time $T_{i\alpha}$ is defined by a five-dimensional transformation $T_{[a,b,c,d,e]}$ based on a five-element vector of non-negative real numbers that may depend on the individual *i* and on the future event α. This vector [*a,b,c,d,e*] may also be influenced by other individuals and by the current environment of the model system. The random variable $T_{i\alpha} = T_{[a,b,c,d,e]}$ schedules the time of the future event α of the process for individual *i* under the conditions in the model. The span and resolution of time determine how this five-parameter transformation can be used to modify scheduling of future events.

The five parameters in this vector have natural interpretations in terms of transformations of clock time into biological time, as defined below:

operation	**time**	**mean value**	**description**	**constraint**
$T_{[1,0,1,0,1]}$	$T = M^{-1}(V)$	$V=M(T)$	identity	
$T_{[a,0,1,0,1]}$	$T = aM^{-1}(V)$	$M(T/a)$	dispersion	$a>0$
$T_{[1,b,1,0,1]}$	$T = b + M^{-1}(V)$	$M(T-b)$	location	$b\geq 0$
$T_{[1,0,c,0,1]}$	$T = M^{-1}(V/c)$	$cM(T)$	intensity	$c>0$
$T_{[1,0,1,d,1]}$	$T = M^{-1}(V+G(d))$	$\max(0,M(T)-G(d))$	truncation	$0\leq d<1$
$T_{[1,0,1,0,e]}$	$T = M^{-1}(V)$ provided $F(T) < e$	$\min(M(T),G(e))$	rejection	$0<e\leq 1$

$T_{[1,0,1,d,e]}$ is censored if $F(T) \geq e$, or equivalently, $M(T) \geq G(e)$. Further, *T* is truncated if $M(T) < G(d)$.

The flexibility of this family of curves is illustrated in Figure 2, where each parameter except rejection is varied individually. Figure 3 shows how to achieve a modest reduction or extension in mean time to event by changing each of the five parameters, with markedly different results. These parameters in a sense adjust the biological clock of the individual with useful, intuitive biological interpretations. Dispersion can slow or speed the biological clock, or time to the next event. Location shifts probabilities of events uniformly to later times, postponing events as appropriate to changing conditions. Intensity can raise or lower the mean number of future events, while keeping the same shape M, corresponding to changes in the environment such as reduced food supply. Truncation and rejection have special roles in terms of immigration and emigration from the system under study, and are considered in some detail in the next section.

$$F_{[a,b,c,d,e]}(t) = 1 - \exp\left\{-\left(cM\left(\frac{t-b}{a}\right) - G(d)\right)\right\}$$

These transformations can be combined mathematically, leading to the sub-distribution provided that $F(t) < e$ and $t \geq aM^{-1}(G(d)/c)+b$. While this may be mathematically appealing, it appears rather complicated and is in fact incomplete. However, in practice it is only necessary to generate V and compute the scheduled future event time

$$T_{i\alpha} = aM^{-1}[\ (\ G(d) + V\ )\ /\ c\ ]\ +\ b.$$

The future event cannot be scheduled before time $aM^{-1}(G(d))+b$ due to truncation. In fact, truncation can shift the probability structure considerably. If $F(T_{i\alpha}) \geq e$, then the scheduling of this future event is rejected as it is beyond the meaningful lifetime of that individual for this simulation.

The importance of this technique in terms of computational efficiency cannot be overemphasized. This above parameterization can dramatically reduce the computer storage requirements for a simulation. In addition, the calculations involve no integration, reducing to an expontial random variate, a pair of linear transformations and an evaluation of $M^{-1}(\cdot)$. All that is needed in addition is an efficient computation of $M(\cdot)$ and $M^{-1}(\cdot)$, for instance using forward and backward cubic splines (e.g. the splines library in the R statistical package, www.r-project.org), and efficient generation of standard exponential pseudo-random numbers. Ewing and collaborators (see www.stat.wisc.edu/~yandell/ewing) have designed a prototype system to interactively design $M(\cdot)$ and $M^{-1}(\cdot)$ and use them and the five-parameter transformation in event-driven competing risks simulations of interacting individuals.

Truncation and Rejection

The first three parameters are fairly straightforward in interpretation. Dispersion is essentially Cox's (1972) proportional hazards, while intensity corresponds to accelerated lifetime models (Viertl 1988; Clarotti and Lindley 1988); location is merely a shift. Truncation allows for entry of an individual to a population from outside, conditional on some chance of prior events, while rejection allows for the removal of individuals, eliminating the scheduling of future events. In practical terms, individuals that are rejected may be removed from the remainder of the simulation, cleaning up the model considerably. Truncated individuals may enter the simulation carrying only vague knowledge of past history. An important future event for such a truncated individual may have happened before it entered the simulation, but the timing would be unknown; alternatively, there is a chance that future event could still be scheduled to occur in the near future. Truncation and rejection make it easy to simulate the population processes of immigration and emigration, respectively. Immigrants can move into an area and continue life processes based on imperfect information, while emigrants can leave an area with future events unknown and irrelevant.

As an example, consider an *Aphtis* arriving at an orange to attack red scale. This individual may enter the simulation with very little knowledge of its previous life history. The field biologist may know a small amount about its age, health, direction of travel. However, the biologist has no information concerning past history. The *Aphytis* is an immigrant, with a truncated life history. Eventually this invader may oviposit in red scale, laying eggs, and ultimately creating a new population of adult *Aphytis* that may either attack nearby red scale, or emigrate to another orange tree. Finally the original flying adult may leave the orange, never to return. While it, and its offspring, may attack other red scale, it is lost to the present simulation. Therefore at emigration its remaining life history is rejected, as its future is irrelevant.

Competing Risks Across Individuals

With these five parameters, it is possible to adjust the distribution functions for each individual in a biosystem to its individual situation. The basic distributions *F*, or equivalently *M*, are calculated once. Using the above five parameters, each individual is given its unique waiting time to each future event. For example, suppose we have a system involving $\beta = 50$ individuals, and suppose each individual can have six possible events occur to him. Defining the whole system's event space requires room to consider 300 events. However, only one event is scheduled next for each individual. This most immediate event will usually but not always be the most probable event, depending on random chance. Hence, we need only consider 50 events—one for each individual. For the whole system only one event is the most immediate. All that is necessary is a minimization of the event times. It is important to note that it is unnecessary to continually recalculate the distributions at each event point. By considering the five scaling parameters, this calculation is done implicitly.

For each individual *i*, there is a corresponding set of random times $\{T_{i\alpha}, \alpha=1,2,\ldots\}$for future events. The random times $T_{i\alpha}$ depend on the [*a,b,c,d,e*] vector, which in turn depends on the current state *s* of the system and the time of the most recent event. Thus the time may be altered by events for other individuals, or by changes in the environment. For each individual *i*, there is one event that will be the next event. By the linear properties of the non-homogeneous Poisson process, it is only necessary to calculate the minimum $T_i = \min\{T_{i\alpha}\ , \alpha = 1,2,3,...\}$.

The next event in a community of β individuals is found by calculating the minimum over all the individuals, $T = \min\{T_i,\ i=1,...,\beta\}$. If the simulation involves a high degree of structure, with many levels of events, then the minimization property given above becomes an efficient technique for finding the next event for the whole population.

The linearization property of the non-homogeneous Poisson process manifests itself in some intriguing ways. The distribution for the next event time *T* in the community is

$$F_T(t) = \Pr\{\ T \le t\ \} = 1 - \prod \Pr\{\ T_i > t\ \} = 1 - \prod \exp\{\ M_i(\ t\ )\ \} = 1 - \exp\{\ \Sigma M_i(\ t\ )\ \}$$

with M_i the mean number of events for individual *i*. Note that the mean number of events $M(t)$ is the sum $M(t) = \Sigma M_i(t)$. This linearization can be further refined for each individual across all future events by noting that $M_i(t) = \Sigma M_{i\alpha}(t)$. Note that we are not considering individuals to be independent per se. However, the competing risk structure *until the next future event* decomposes in this linear fashion as if the individuals were independent. At that time, $T=\min\{T_i\}$, the competing risk structure must be rebuilt, which may cause rescheduling of some future events.

Consider the situation where β predators of some species are hunting independently. The probability of an event occurring is certainly different than the situation in which one predator of that species is hunting. The above situation can be handled in the following manner. If β predators hunt independently with the same chance of success during a time interval, then the mean value function for the β predators is just $M(t) = \beta M_1(t)$.

The probability $F_T(t)=1-\exp\{\Sigma M_i(t)\}$is an example of the linearization process with respect to the non-homogeneous Poisson process. The fact that the product operation can be reduced to a summation operation for exponential functions is the critical step in the linearization of the competing risk structure.

Scheduling Immediate Events

There are occasions when it may be appropriate to schedule an event without knowing its outcome. At the time of that future event, its specific outcome may be predicted based on the environment and current state of the simulation. Which of the β predators made the kill? The future event scheduled may be based on the group, but the consequence of the event may affect certain predators (the prey, of course, is now dead). For instance, the dominant predator may get the biggest portion, which in turn might affect its future fitness, and may also affect the future fitness of other members of its group. This can be handled as an immediate event and consequent pending events.

The multinomial is an excellent way to schedule such immediate events. The future event of predation is scheduled by drawing from the distribution of T. Conditional on a predation event, one of the predators may be responsible for the kill. The immediate event deciding which is the primary predator involves drawing from the multinomial distribution with probabilities

$$p_i = M_i(t) / M(t),\ i = 1, \ldots, \beta.$$

The event i is processed, taking care to reschedule any pending events that might be affected by this. That is, the immediate events might affect other future events by modifying the five parameters for one or more individuals, based on the amount they are able to consume. Thus future events for individual i could be rescheduled at this time, changing the appropriate $M_i(t)$. Other individuals may need to reschedule events as well. Once this is completed, the process continues as before, but with a newly modified competing risk structure.

Again, consider California red scale being attacked by *Aphytis*. The parasitoid selects a host on which to oviposit by some random search. There may be β hosts in the vicinity, differing in terms of distance, size, and other factors that may affect the chance of being chosen. Once a host is selected, there is a further chance mechanism as to whether a male or female egg would be oviposited. Either way, the red scale is immediately rescheduled for death.

Immediate events that handle multiple events may need to decide how many events are to be created. The number of live births may be random, and hence can be drawn from an appropriate distribution. A natural choice would be to use a histogram based on experimental data. Since the distribution might be modified by environmental considerations (temperature), health of the individual, and actions of other individuals, one could construct modifications along similar lines to those considered for the probability distributions to schedule future events.

The Red Scale/Aphytis System

The California red scale *Aonidiella aurantii* is a major pest for California citrus growers and is responsible for significant financial loss for the growers. Though red scale does not actually kill the fruit, it effectively blemishes it so that the fruit is not marketable (Forester, L.D., Luck, R., and Grafton-Cardwell, E.E., 1998, “Life stages of California red scale and its parasitoids”, U.C. Div. of Agric. & Nat. Res. Pub. 21529). The normal method for controlling red scale is by spraying insecticide. However, red scale is showing an increased resistance to spraying. An alternative to spraying uses a biological control agent to augment chemical control. It is possible to release a small wasp, *Aphytis melinus* Debach, which effectively parasitizes red scale.

Environmental Factors

Yu and Luck (1988) found evidence that the developmental stages of red scale are extremely temperature dependent. Under certain environmental conditions, when the ambient temperature remains above a certain minimum threshold, the entire developmental stage may experience minimal successful parasitization. In addition, red scale developmental rates depend on the substrate, doing best on fruit, then leaves, and worst on stems and bark. *Aphytis* only parasitizes scale during daylight hours. However, scale can grow as long as the temperature stays above a certain threshold. Hence the number and quality of scale available as hosts depends on temperature, and the available host population is not even aged. Finally the entire process is seasonal.

In addition to the temporal description presented below, spatial distribution should be addressed. The migration pattern depends on the number of other red scale on the citrus. However, this dependency is local. Scale is normally not uniformly distributed on either the fruit or the leaves and branches; hence the searching mechanism used by *Aphytis* is non-trivial.

Red Scale Life History

Red scale life cycle begins as a crawler that is mobile long enough to find a suitable location on either a branch, leaf or fruit on which it can begin to feed. From this point on, red scale is immobile except for the adult males. Red scale displays two distinct immobile stages, an instar or feeding stage and a molt or dormant stage. Immature males complete a single molt while immature females complete two molts. Males and females feed during the instar stages but not during the dormant, molt stage. Since red scale development is temperature dependent, times between life stages are presented in degree-days, which are approximately the cumulative degrees above 11°C.

Males complete three distinct immature stages: the second instar male, the prepupal male, and the pupal male. After about 330 degree-days, the second instar males begin to pupate and in approximately 30 degree-days the prepupate male transforms into a pupate male and emerge as adult males approximate 20 degree-days later. Adult males depart the area to mate. Second instar females, on the other hand, enter a second molt which last approximately 50 degree-days. After that, females reinsert their rostrum to resume feeding as third instars. Third instar females may vary greatly in cover and size; mating occurs during this phase. Unmated third instar females continue to grow, but will not fully develop. A mated mature female is sealed inside the scale cover and stops feeding. In approximately 90 degree-days the mature female produces crawlers. The cycle from crawler to this final stage lasts approximately 650 degree-days.

***Aphytis* Life History**

Aphytis, an external wasp, is the primary natural enemy associated with red scale. *Aphytis* inserts its ovipositor through the scale cover and deposits its eggs on either the dorsal or ventral side of the insect body. Before laying its eggs, *Aphytis* permanently paralyzes the scale with a venom, leading to scale death scale even if eggs are not deposited. The food available to the developing *Aphytis* offspring is determined by the size of the scale body at the time it was paralyzed.

The life stages of *Aphytis* are as follows: egg stage, larval stage, prepupal stage, pupal stage and host feeding stage. Eggs may become damaged or flattened if suitable scale stages are not found, since the competition for suitable scale stages often results in more than one *Aphytis* parasitizing a single host (superparasitism). Superparasitized scale rarely yields healthy *Aphytis*. After two days,

health *Aphytis* eggs hatch into larvae and begin to feed. In approximately five days the larvae develop into prepupae. In approximately one more day, *Aphytis* completes pupation. Adult *Aphytis* emerge in approximately four to five days. The process of emergence of adult *Aphytis* once the egg is oviposited requires approximately thirteen days.

Not all scale stages are consistently available or of equal quality. *Aphytis* can only parasitize certain scale stages. In general, *Aphytis* prefer scale in one of the instar stages to those in the molt stage. During the instar stage the scale cover is free of the body and *Aphytis* can lay eggs on both the dorsal and ventral surfaces of the scale. During the molt and the mature female stage, the scale cover is rigidly fused to a hardened body, and *Aphytis* can only lay eggs on the ventral surface of the body. Since the scale is extremely tough, immature larvae find it difficult to feed. In fact, most immature *Aphytis* die if forced to develop in a molt stage.

From the perspective of *Aphytis*, red scale grow through a window of availability improving in quality as they grow larger, but suddenly become unavailable upon maturity. *Aphytis* prefer to oviposit during the scales third instar because of its large size. By paralyzing the scale before ovipositing on it, *Aphytis* choose the amount of food that will be available for it's offspring. The number of offspring is proportional to the size of the scale. The sex of the offspring is a function of scale size; most female *Aphytis* are produced from the third instar. In addition to scale stage and size, *Aphytis* seems to select second instar scale with large skirt areas over third instar with small skirts. Scale access may be more important than overall scale body size. In host selection, therefore, the probability that a particular scale is attacked is a function of the number of red scale and *Aphytis* in the area and of the stage, size and accessibility of each individual scale.

Female *Aphytis* usually mature their first batch of eggs, approximately 12% of its lifetime egg supply, within 24 hours of emergence using resources from their larval stage. They produce eggs during their entire adult lifetime, relying on periodic feeding on body fluids of small, immature hosts for sustenance (Opp and Luck,1986; Luck and Nunney 1999). Adult *Aphytis* host feed by probing the scale body more extensively than when ovipositing, feeding on the body fluids that ooze from the wound. *Aphytis* feed on small scale hosts while searching for larger scales to serve as suitable hosts to lay eggs. Host feeding kills a substantial percentage of California red scale beyond those killed through parasitism.

Within 12 to 18 hours of host feeding, the female develops approximately 1.3 eggs if it has not recently oviposited, or about 2.7 eggs if it has. Host feeding appears to provide both metabolic maintenance and to support egg production. Collier (1995) showed that *Aphytis* that do not have access to hosts for either oviposition or host feeding will re-absorb about one egg per day. However, egg re-absorption will not supply the metabolic needs of the wasp in the absence of honey or other carbohydrates.

Red Scale/*Aphytis* System Event Structure

Upon careful examination of the parasitization process displayed by *Aphytis*, the system is comprised of complex set of conditional probabilities that can be used to describe overall probability that certain events will occur, given a defined span and resolution.

1. Probability of successful search by *Aphytis* depends on: number of red scale in a local area, time of day, temperature, stage of development of an individual red scale, size of an individual red scale.

2. Probability of oviposit depends on: host stage of development, host size, host accessibility.
3. Probability of emergence depends on: host stage of development, host size, previous feeding and/or superparasitisim.

Thus there are three special types of events. In the red scale-*Aphytis* system the probability of a successful search is a *pending event*, while the probability of oviposit could be considered to be an *immediate event*, and finally the probability of emergence is a *future event*. In this system both parasitism and host feeding lead to the death of an individual scale. However, host feeding could affect the probability of emergence. It appears that *Aphytis* has exploited the system by selecting small sized red scale to feed upon, saving the larger red scale to oviposit on. The host's stage of development and size modifies *Aphytis* parasitizing strategy. *Aphytis*-red scale system exhibits a fairly complex set of conditional probabilities that ultimately determine the ability of *Aphytis* to reproduce successfully.

Simulation

A simulation system is being developed using the event-driven competing risks structure for quantitative population ethology as outlined above. The initial implementation is focused on the California red scale-*Aphytis* system, although the software module has no code specific to this system except for some details of handling of events. Simulated red scale "lives" on a degree-day basis (the integral of degrees above 52 Fahrenheit), while *Aphytis* follows a diurnal clock, inactive at night. When it is cold, Red Scale grows slowly and is readily parastized by *Aphytis*. However, when it is warm, Red Scale can continue to grow overnight, effectively escaping by maturing to gravid state. The simulation uses life history information from Forster et al. (1988 op. cit.). Times to future events are by default drawn from the exponential, but that can be tuned, as in Figures 1 and 2, using a graphical interface. The parasite search algorithm and life events for both species are being refined based on data collected by Luck. The software is written in the R language, which is graphical, extensible, and in the public domain (Venables and Ripley 2000; see www.r-project.org). Further details of the simulation and access to public domain software can be found at www.stat.wisc.edu/~yandell/ewing.

Acknowledgements

The authors would like to thank Dr. David L. Wood for his continual and unwavering encouragement over the long haul. We would like to thank David Baasch and Dr. William Waters for the contributions and encouragement they have provided in this effort, and Dr. Peter A. Rauch for his early efforts to organize this material. The authors would like to thank Drs. C. Huffaker, R. Smith, and J. Knox for their support in the development of this modeling technique. We would like to particularly thank Ms. Floy Worden who typed original manuscripts quickly and accurately in the 1970s, and Ms. Karen Denk who retyped these into a modern word-processor in the 1990s.

This work was performed in part under the auspices of the U.S. Atomic Energy Commission and supported in part by the National Science Foundation through a grant (NSF GB-34718) to the University of California; and by the Environmental Protection Agency through an interagency agreement (EPA-IAG-COV-4) with the Lawrence Livermore Laboratory. One of the authors (B.E.) was supported by the U. S. Forest Service in the initial phase of this effect. The findings, opinions

and recommendations expressed herein are those of the author(s) and not necessarily those of the University of California, the National Science Foundation, the Environmental Protection Agency, the U. S. Forest Service, the Atomic Energy Commission or the Lawrence Livermore Laboratory. This paper contains part of the intended Doctoral Dissertation of one of the authors (B.E.).

References

Caswell, H., 1989. Matrix Population Models, Sinauer: Sutherland, MA.

Chiang, C.L., 1968. Introduction to Stochastic Processes in Biostatistics, Wiley: New York, NY.

Chiang, C.L. 1972. On constructing current life tables. J. Amer. Statist. Assoc. 67, 538-541.

Clarotti, C.A., and Lindley D.V., ed., 1988. Accelerated Life Testing and Experts' Opinions in Reliability. North-Holland/Elsevier.

Clifford, P., 1977. Nonidentifiability in stochastic models of illness and death. Proc. Natl. Acad. Sci. USA 74: 1338-1340.

Collier, T.M., 1995. Host feeding, egg maturation, resorption, and longevity in the parasitoid *Aphytis melinus* (Hymenoptera: Aphelinidae). Ann. Entom. Soc. Amer., 88: 206-214.

David, H.A., 1974. Parametric approaches to the theory of competing risks. Reliability and Biometry, SIAM: Philadelphia, 275-290.

Fine, J.P., 1999. Analysing competing risks data with transformation models. J. Roy. Statist. Soc. B, 61: 817-830.

Fix, E., and Neyman, J., 1951. A simple stochastic model of recover, relapse, death and loss of patients. Hum. Biol. 23: 205-241.

Hutchinson, J.M.C., McNamara, J.M., 2000. Ways to test stochastic dynamic programming models empirically. Animal Behaviour 59: 665-676.

Keyfitz, N., 1966. A life table that agrees with the data. J. Amer. Statist. Assoc. 61: 305-312.

Keyfitz, N., 1968. A life table that agrees with the data: II. J. Amer. Statist. Assoc. 63: 1253-1268.

Langton, C.G., 1986. Studying artificial life with cellular automata. Physica 22D: 120-149.

Luck, R.F., and Nunney, L., 1999. A Darwinian view of host selection and its practical implications. In A. Hawkins and H. V. Cornell (eds). Theoretical Approaches to Biological Control. Cambridge University Press. Cambridge UK, pp. 283-303.

Mangel, M., and Clark, C.W., 1988. Dynamic Modeling in Behavioral Ecology, Monographs in Behavior and Ecology, ed. By JR Krebs and T Clutton-Brock, Princeton U. Press.

Opp, S.B., and Luck, R.F., 1986. Effects of host size on selected fitness components of *Aphytis melinus* DeBach and *A. lingnanensis* Compere (Hymenoptera: Aphelinidae). Ann. Entom. Soc. Amer., 79: 700-704.

Tsiatis, A., 1975. A nonidentifiability aspect of the problem of competing risks. Proc. Natl. Acad. Sci. USA 72: 20-22.

Venables, W.N., and Ripley, R.D., 2000. S Programming, Springer-Verlag, New York, NY.

Viertl, R., 1988. Statistical Methods in Accelerated Life Testing, Vandenhoeck and Ruprecht.

Yandell, B.S., 1982. Nonidentifiability of lethality in the survival experiment with serial sacrifice. Math. Biosci. 62: 1-6.

Yu, D.S., and Luck, R.F., 1988. Temperature dependent size and development of California red scale (Homopter: Diaspididae) and its effect on host availability for the ectoparasitoid, *Aphytis* melinus DeBach (Hymenoptera: Aphelinidae). Envir. Entom. 17: 154-161.

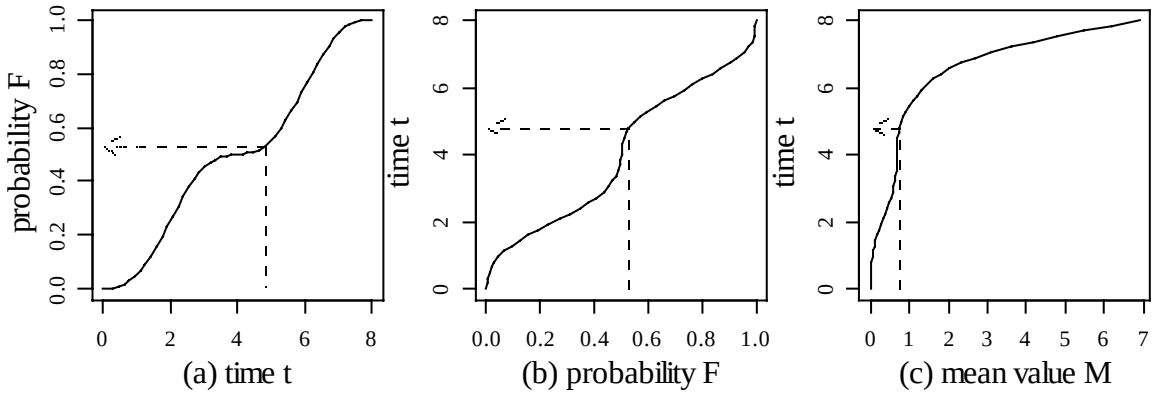

Figure 1. Time Depends on Events. (a) usual interpretation of probability F of event by time t; (b) inverse relation with t a function of F; (c) transform from F to mean value M shows how to map time by first randomly picking mean value. Curve is a cubic spline with 8 knots.

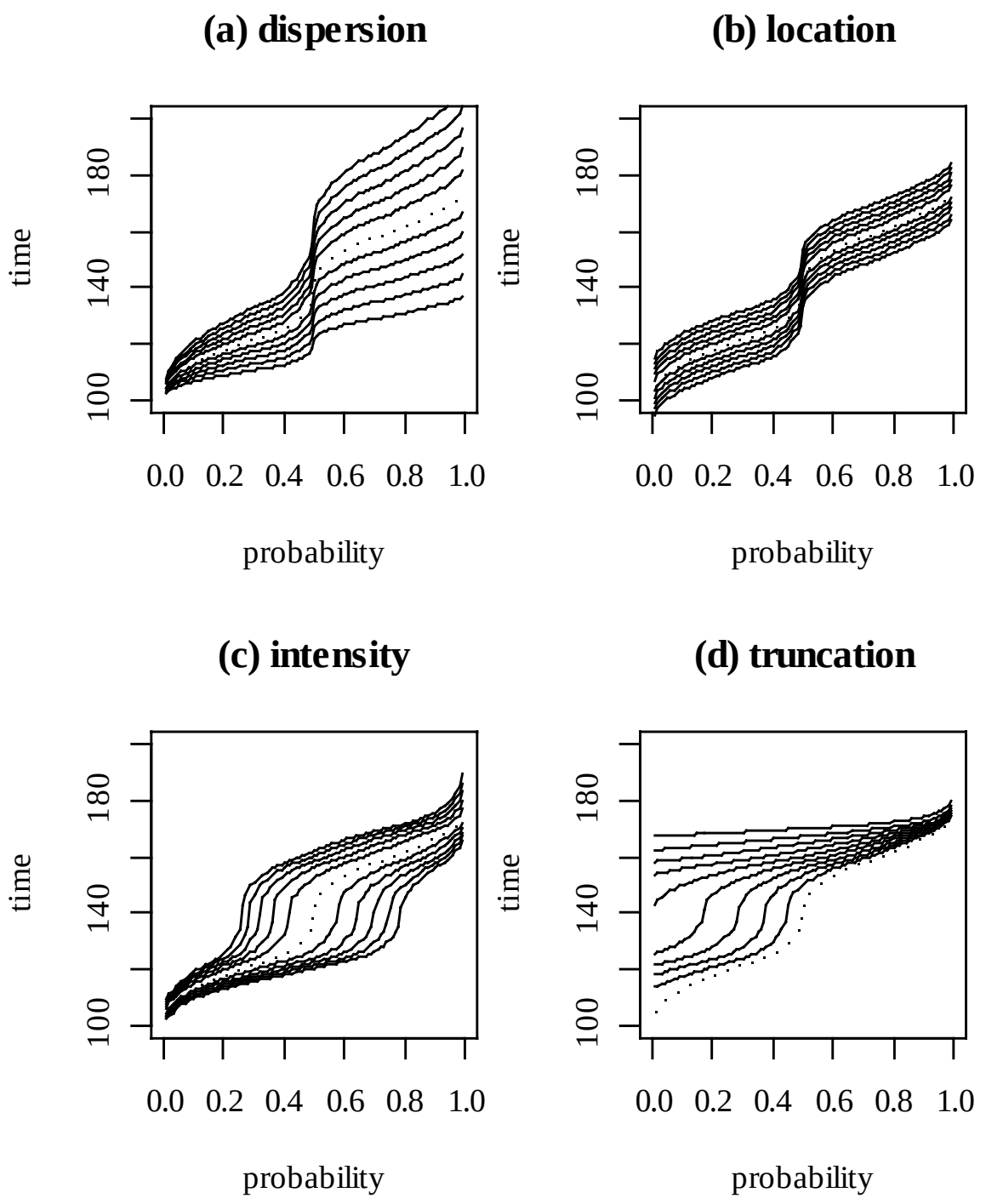

Figure 2. Four of the five parameters transform the probability to time mapping. Basic curve is from Figure 1 with dispersion=10 and location=100: (a) dispersion varies from 5 to 15; (b) location varies from 90 to 110; (c) intensity varies geometrically from 2.25 to 1/2.25; (d) truncation varies from 0 to 1.

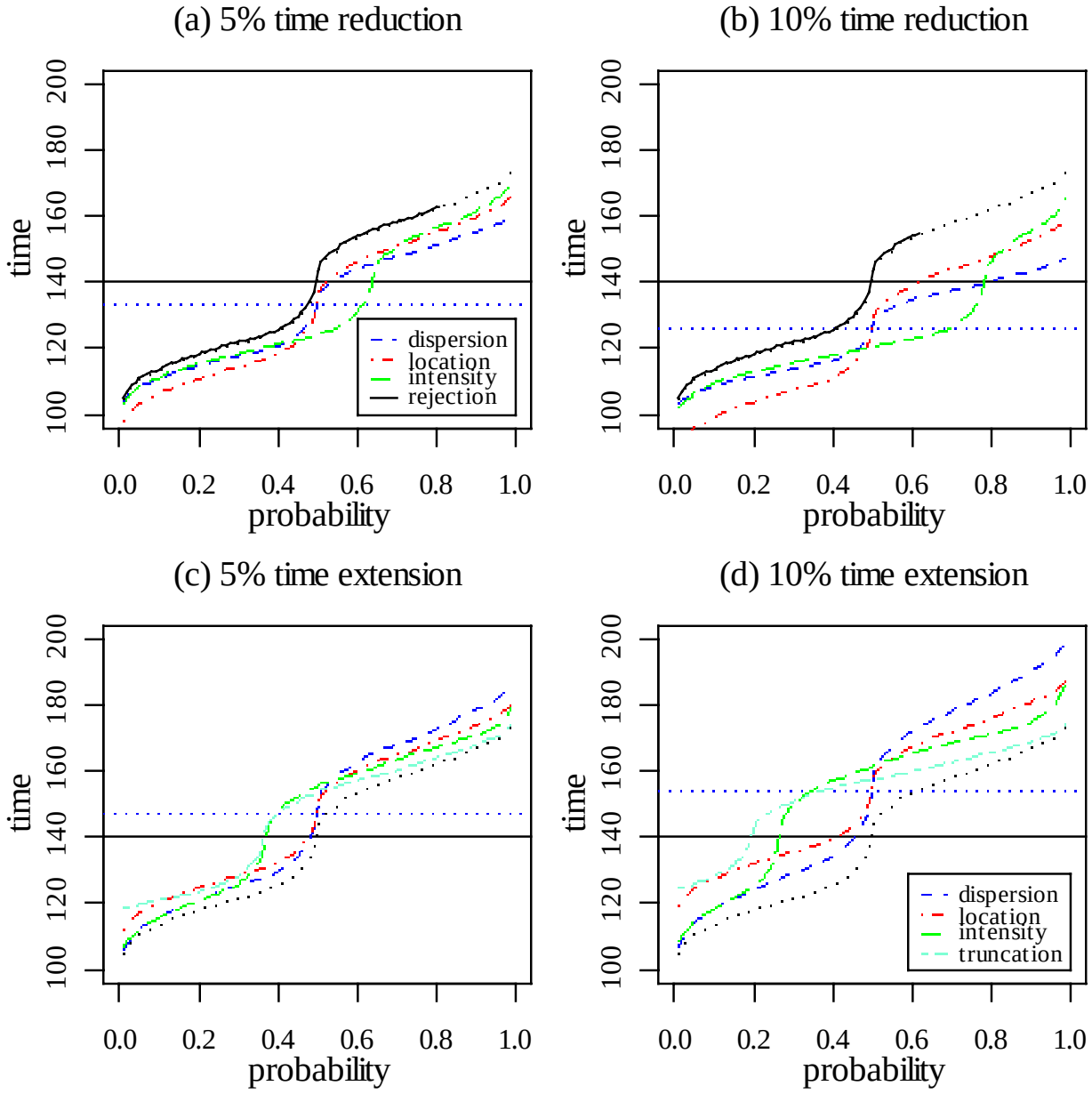

Figure 3. Modification of five parameters for time reduction or extension: (a) 5% and (b) 10% time reduction); (c) 5% and (d) 10% time extension. Original curve with mean of 140 days is black dotted line. Reduction or extension involves changing one parameter: dispersion (blue dash), location (red dot dash), intensity (green long dash), rejection (black solid) or truncation (aqua short-long dash). Horizontal black solid line is at mean time of 140 days; blue dashed line is at reduced or extended mean time.