Austrian Journal of Statistics
Volume 27 (1998), Numbers 1&2, 109-121

Canonical Modeling: A Link Between Environmental Models and Statistics

Eberhard O. Voit
Department of Biometry and Epidemiology
Medical University of South Carolina
Charleston, SC 29425-2203 U.S.A.

Abstract: The article describes three connections between statistics and modeling in environmental studies. As the first connection, typical exposure and disease models are derived mathematically from general principles of dynamical systems analysis. The second connection is developed between physiological and environmental processes on one hand and survival curves on the other. The third connection describes how dynamic processes affect distributions of random variables. Forming novel connections between different branches of environmetrics enhances our understanding of environmental phenomena and offers new avenues of analysis.

Zusammenfassung: Drei neue Beziehungen zwischen Umweltstatistik und mathematischer Modellierung werden vorgestellt. Zuerst werden typische Modelle der Risikoanalyse von allgemeinen Prinzipien der dynamischen Systemanalyse abgeleitet. Anschließend wird gezeigt, wie die Behandlung von komplexen physiologischen und umweltgesteuerten Prozessen durch kanonische Modelle direkt zu Funktionen führt, die aus der Überlebensanalyse bekannt sind. Als drittes Beispiel werden Trends in Verteilungen zeitabhängiger Zufallsvariablen diskutiert, die durch Systeme von Differentialgleichungen definiert sind. Die aufgezeigten Beziehungen vertiefen unser Verständnis einiger Prinzipien der Umweltstatistik und erweitern unser momentanes Repertoir an Analysemethoden.

Keywords: Canonical Model, Cox Model, Epidemiology, Exposure, Risk Assessment, S-systems.

1 Introduction

Statistics and modeling are often considered as two separate entities within environmental studies. The former primarily focuses on the characterization of associations between variables and uses methods of regression analysis and hypothesis testing, while the latter tries to discover or elucidate the mechanistic features of processes in the environment and typically uses differential equations. Both branches of environmetrics have created germane methods of mathematical experimentation and

analysis, and both have been very successful in their applications. Nonetheless, one must wonder whether the two approaches to solving environmental problems are independent of each other and, if not, what the links between them might be. Obviously, a comprehensive answer is out of reach at this point, but this article will show with three generic examples that a certain type of *canonical modeling* provides some natural links between mechanistic process descriptions, environmental statistics, epidemiology, and environmental health risk assessment.

The article begins with a brief review of the background and rationale of canonical modeling and identifies some prototypical environmental models as special cases. It then demonstrates three links between modeling and statistics. The first derives standard concepts of epidemiology from the dynamics of underlying disease models. The second link connects survival functions and statistical distributions with the dynamics of populations or individuals. Finally, the third link shows how dynamic changes in a random variable lead to time trends in the distribution of this random variable.

Several of the results presented here are not entirely new in themselves, but they have never been put in the present context. The article is thus more of a review of some recent literature that emphasizes the overall need for integrating statistical and modeling approaches rather than technical details, which can be found in the cited literature.

2 Canonical Models

A dogma of mathematical modeling states that one understands a system if one understands how each of its components (or *state variables*) changes over time, and possibly space. For phenomena without significant spatial aspects, the dogma directly leads to a system of ordinary differential (or difference) equations, of which each describes the temporal change in one of the state variables. If spatial variation cannot be ignored, as is the case in many exposure and transport models, the resulting equations are partial differential equations. However, for their mathematical analysis, space is often discretized, and the result is again a system of much simpler ordinary differential equations. Ordinary differential equations thus play a prominent role in modeling and are the exclusive focus of this article.

Without much loss of generality, a spatially homogeneous dynamical system may be formulated as

$$\dot{X}_i = \sum_{j=1}^{k_i} F_{ij}(X_1, X_2, \ldots, X_n, X_{n+1}, \ldots, X_{n+m}) \qquad i = 1, \ldots, n, \tag{1}$$

where the dotted variable on the left-hand side denotes the derivative of X_i with respect to time, X_1, ..., X_n are dependent (*state*) variables, and X_{n+1}, ..., X_{n+m} are independent (*external*) variables, which are considered constant during the time period of interest. Parameters are explicitly omitted, but implicitly part of the functions F_{ij}.

The k_i functions F_{ij} in the i-th equation themselves are usually ill defined, but we may assume, without restricting generality much, that they are differentiable and positive-valued. If so, approximation theory can be employed to capture the essence of

the model (1), at least to some degree. For this purpose, an *operating point* is chosen, which consists of the *nominal values* of all variables and often, but not necessarily, coincides with the steady-state point of the system. The simplest approximation of the functions F_{ij} is multi-dimensional linearization at the operating point. However, for complex environmental systems, linearization is not a good choice, since many natural phenomena simply are not linear. They are saturated, may exhibit stable oscillations, or may even be chaotic, and these types of responses are difficult, if not impossible, to model with linear functions.

For a useful alternative, one transports all variables X_i and all functions F_{ij} into logarithmic coordinates, linearizes the functions in this logarithmic space, and transports them back to Cartesian coordinates. The result of this procedure always is a representation of each F_{ij} in the form of a multivariate product of power-law functions:

$$F_{ij} \approx \gamma_{ij} \prod_{p=1}^{n+m} X_p^{q_{ijp}} \tag{2}$$

(e.g., Savageau, 1969ab, 1972, 1976; Voit, 1991).

If the original functions in (1) are replaced with these types of power-law approximations, the result is a *Generalized Mass Action* (*GMA*) system of the form

$$\dot{X}_i = \gamma_{i1} \prod_{j=1}^{n+m} X_j^{g_{1ij}} \pm \gamma_{i2} \prod_{j=1}^{n+m} X_j^{g_{2ij}} \pm \ldots \pm \gamma_{ik} \prod_{j=1}^{n+m} X_j^{g_{kij}} . \tag{3}$$

with positive *rate constants* γ and real-valued *kinetic orders* g.

A variant of special interest arises when one aggregates all functions F_{ij} in (1) that *increase* or *augment* the variable X_i into a single function F_i^+ and when one similarly aggregates all functions that *decrease* or *degrade* X_i into a single function F_i^-. Power-law approximation in this case leads to the ("synergistic") *S-system* model

$$\dot{X}_i = F_i^+ - F_i^- \approx \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \tag{4}$$

(e.g., Savageau, 1969b, 1976; Voit, 1991).

S-system models can be constructed even if information about the system is rather limited: One merely has to know the pattern of fluxes between variables and the pattern of signals that define which dependent or independent variables have a direct effect on a particular flux. The construction of the equations follows a simple rule. A variable X_j appears in a power-law term F_i^+ or F_i^- if and only if it has a direct influence on this particular term. If the influence is positive, the associated exponent is positive, and if the influence is negative, the exponent is negative. The result of this construction is a complete symbolic model, which sometimes already offers insight in some of the dynamic features of the system (e.g., Savageau, 1972, 1976; Voit, 1991). For numerical results, quantitative information is needed to specify rate constants and kinetic orders.

S-systems combine a full nonlinear repertoire of dynamic behaviors with unique analytical and logistic advantages. They often model complex systems with surprisingly high accuracy, sometimes over several orders of magnitude in variation, even though they are, by their nature, local approximations. Also, it was shown that all relevant

dynamic responses, including limit cycle oscillations and deterministic chaos, can be captured with S-systems (Savageau and Voit, 1987; Voit, 1993), which guarantees that this power-law structure is a framework rich enough for environmental phenomena.

In contrast to GMA systems and most other nonlinear models, steady states of S-systems, their local and structural stability, sensitivities, and responses to changes in input variables can usually be computed with analytical means (e.g., Savageau 1974, 1976; Voit, 1991). The algebraic nature of the steady states also facilitates the optimization of nonlinear systems in S-system form (Voit, 1992a; Torres et al., 1996, 1997). In general there are no analytical solutions to GMA or S-system differential equations, but their homogenous structure has been thoroughly exploited to devise very efficient numerical algorithms for their evaluation (e.g., Voit et al., 1989; Irvine and Savageau, 1990; Ferreira, 1992).

Canonical models in GMA or S-system form contain many mathematical structures as special cases. Numerous growth functions and physical laws can be represented in canonical form, and allometric scaling is uniquely supported by this type of model (Savageau, 1979b). Functions used for simple extrapolation in environmental risk assessment or for extrapolation with physiologically based, pharmacokinetic (PBPK) models are readily reformulated as canonical models, and so are some of the relevant cancer models (e.g., Voit and Schubauer-Berigan, 1998)

3 Derivation of Risk and Exposure Models from Canonical Models

One important goal of environmental studies is the characterization of health risks to humans or animals. Depending on the author's background in risk assessment or epidemiology, such risks are either formulated by means of a risk equation, a linear-logistic model, or Cox's proportional hazards model. The following demonstrates that all three can be directly derived from canonical models of the underlying exposure and disease processes (see also, Voit and Knapp, 1997; Voit and Schubauer-Berigan,1998).

We begin with the fundamental equation of risk assessment, which quantifies a health risk as a product of potency and dose:

$$Risk = Potency \times Dose. \tag{5}$$

The apparent simplicity of the equation is misleading, since the terms of the equation are aggregates of numerous contributing factors of often overwhelming complexity. The potency combines the entire toxicology of the risk agent, while the dose results from a plethora of transport and exposure phenomena. Fortunately, potency values for many compounds are published by agencies like the US EPA and need not be re-derived for every analysis.

Most models quantifying dose and exposure are ultimately sums of products. The sums account for different exposure routes, different exposure scenarios, or different microenvironments (cf. Ryan, 1991), while the products consist of contributing factors like concentration, duration, and body weight. A typical dose equation takes the form

$$Dose = \frac{C \times I \times T}{W \times L} \times F \tag{6}$$

where C is the concentration of the risk agent, I the intake, T the exposure time, W the body weight, L the lifetime, and F a factor accounting for conversions of units as well as for uncertainties (e.g., Hallenbeck, 1993: Ch. 3). For several exposure routes, such as inhalation and ingestion, products of this type are summed.

To see the relationship of these risk and exposure models to canonical models, suppose the concentration of the risk agent is coded as X_1 and the variables X_2, ..., X_n represent other physical features that contribute to the dynamics of X_1. For example, X_1 could represent a point concentration of benzene following a petroleum spill, and the other variables could reflect processes of dispersal, biodegradation, and fugacity. For simplicity of argument, but without restricting generality much, suppose further that X_1 does not affect its own production and that its intrinsic disappearance at the point of interest follows a first-order process. The appropriate GMA equation is thus

$$\dot{X}_1 = \gamma_{11} \prod_{j=2}^{n} X_j^{g_{11j}} \pm \gamma_{12} \prod_{j=2}^{n} X_j^{g_{12j}} \pm \ldots \pm \gamma_{1k} \prod_{j=2}^{n} X_j^{g_{1kj}} - \gamma X_1 . \tag{7}$$

At steady state, Eq. (7) is set equal to zero and can be divided by the positive rate constant γ. Bringing X_1 to the left-hand side results in a steady-state exposure equation that consists of a sum of products of power-law functions. If all variables affect dose or exposure in a linear fashion, all exponents g_{1ij} are either 0 or 1, and the steady-state equation of the GMA model (7) reduces to the traditional form of an exposure equation, as discussed above. Thus, by simplifying the full GMA model in such a way that it reflects traditional assumptions, it naturally produces the well-known results, quasi as a special case. The full GMA model (7), with kinetic orders not necessarily equal to 0 or 1, goes beyond the traditional model by accounting for nonlinear effects, synergisms, and feedbacks. The GMA model also describes the dynamics of the exposure process, which is traditionally ignored.

Epidemiologists approach health risks to populations in a different fashion, namely by studying the proportion P of the D diseased individuals within a population of size N: $P=D/N$. In the relevant case where not all individuals are healthy, a directly related, more familiar concept is that of the *odds of this proportion*, which is defined in terms of diseased and healthy (H) individuals as

$$\frac{P}{1-P} = \frac{(N-H)/N}{1-(N-H)/N} = \frac{D}{H} \tag{8}$$

(cf. Rothman, 1986: p.291). The proportion P is assumed to be related to risk factors Y_i via the *linear-logistic model*, which takes the form of a multi-variate logistic function:

$$P = \frac{1}{1 + e^{-(a_0 + a_1 Y_1 + a_2 Y_2 + \ldots + a_n Y_n)}} . \tag{9}$$

The coefficients a_1, ..., a_n quantify the impact of the respective risk factors on disease development, while the coefficient a_0 is not associated with a particular risk factor but

reflects the background risk. In epidemiological interpretation, it corresponds to the logarithm of the odds for disease among the unexposed.

Even though no rationale for (9) is given in the epidemiological literature, the linear-logistic model turns out to be widely applicable and surprisingly accurate. How can such a simple function possibly incorporate the enormous complexity that governs the dynamics of disease in a population? Suppose the disease process is formulated as a (potentially very large) S-system model. For bookkeeping purposes, assume that the two variables X_{n+1} and X_{n+2} of the S-system code for the numbers of diseased and healthy individuals, respectively. Solving the differential equations of the S-system numerically would produce the entire dynamics of disease development, and one could study the impact of the different risk factors and physiological determinants of disease. To reduce the complexity of such a modeling exercise, we restrict the analysis to a situation where the disease process has reached a steady state. Straightforward computation shows that D and H are products of power-law functions in some or all of the S-system variables,

$$\begin{aligned} D = X_{n+1} &= \gamma_{n+1} X_1^{e_1} X_2^{e_2} \ldots X_n^{e_n}, \\ H = X_{n+2} &= \gamma_{n+2} X_1^{f_1} X_2^{f_2} \ldots X_n^{f_n}, \end{aligned} \tag{10}$$

where the parameters γ, e, and f are sums and products of the original S-system parameters (Savageau, 1969b; Voit and Knapp, 1997). Substitution of this solution in the general formula (8) for the odds of the disease proportion results in

$$\frac{P}{1-P} = \frac{\gamma_{n+1}}{\gamma_{n+2}} X_1^{e_1-f_1} X_2^{e_2-f_2} \ldots X_n^{e_n-f_n}, \tag{11}$$

and renaming $a_0=\ln(\gamma_{n+1}/\gamma_{n+2})$, $a_j=e_j-f_j$, $Y_j=\ln(X_j)$ for $j=1,\ldots n$ yields

$$\frac{P}{1-P} = \exp(a_0)\exp(a_1 Y_1)\exp(a_2 Y_2)\ldots\exp(a_n Y_n)\,. \tag{12}$$

Elementary operations convert this result directly into the linear-logistic model (9). This means, that the linear-logistic model is a natural and necessary consequence of the formulation of the disease process as a dynamical model in S-system form. It is interesting that this result derives from the canonical model essentially without assumptions. The result has some practical implications. One can interpret the parameters in the linear-logistic model as weights of the contributing risk factors, as is done in epidemiology, but it is now also possible to interpret them within the framework of dynamical modeling, where they correspond to amplification or gain factors (e.g., Savageau 1972, 1976; Voit and Knapp, 1997).

If significant numbers of individuals are lost to follow-up or to unrelated death, the use of the logistic function becomes problematic, since it models cumulative incidence. Cox (1972) suggested instead to base risk estimates on person-years of survival and to compute an *incidence rate*, *incidence density*, or *hazard rate*. Cox's model in the typical case reads

$$\ln(\textit{incidence rate}) = b(t) + b_1 Y_1 + b_2 Y_2 + \ldots + b_n Y_n. \qquad (13)$$

The term $b(t)$ describes a time-dependent *baseline* effect, while the coefficients $b_1, \ldots, b_n$ account for the contribution of each risk factor to the overall disease incidence rate. Again, one asks where this particular mathematical form comes from. And again, the answer can be found in the corresponding canonical model. The analogue to the incidence rate in the canonical disease model is directly proportional to the α-term in the equation of the diseased, which are represented by X_{n+1}:

$$\dot{X}_{n+1} = \alpha_{n+1} \prod_{j=1}^{n+2} X_j^{g_{n+1,j}} - \beta_{n+1} \prod_{j=1}^{n+2} X_j^{h_{n+1,j}} . \qquad (14)$$

Just considering the α-term, which describes the overall instantaneous increase in the number of diseased individuals, taking its logarithm and recoding $Y_i = \ln(X_i)$ as before, one directly obtains

$$\ln(\textit{incidence rate}) \propto \ln(\alpha_{n+1}) + \Sigma\, g_{n+1,j}\, Y_j \qquad (15)$$

which is of the same form as Cox's model (13). The only apparent difference is the time dependence of $b(t)$ which, however, is readily explained. Suppose, only a few variables actually contribute to the disease process and are identified as risk factors Y_j, while the remaining variables represent physiological or environmental factors that drive the population dynamics, irrespective of whether the population is exposed to the risk factors in question. These remaining variables are time dependent, but since they are not individually listed in Cox's model, they are lumped with α_{n+1}, thus making this parameter a time dependent term that corresponds to $b(t)$. Again, these results follow directly from the S-system description of the disease process.

4 Canonical Survival Models

If dynamic models focus on death outcome rather than disease incidence, they become survival models. For simplicity of argument, suppose the general population in the above models is replaced with a cohort without replenishment. Its dynamics is described by a survival curve that starts at 100% and monotonically decreases toward 0. Survival phenomena are very complex, because they depend not only on time, but also on uncounted internal and external factors and processes that ultimately lead to survival or death. In light of this complexity, it is amazing that the overall appearance of observed survival curves is usually rather smooth and simple.

To explain this phenomenon, one may borrow an argument proposed by Savageau (1979a) in the context of growth. As a thought experiment, imagine that all processes contributing to survival could be captured in a comprehensive mathematical model and approximated as a multi-variate S-system, as described above in Section 2. The involved processes run at vastly different time scales. At one end of the spectrum, biochemical and molecular processes occur within seconds or minutes, while at the other extreme, evolutionary processes and global climate changes are very slow in comparison to the life span of an individual or population. The very fast processes reach

their steady states so quickly that they are essentially always at the steady state, while the very slow processes change so little that they are essentially constant throughout the lifetime of the individual or population. In either case, the time derivative of each associated variable is zero, and its defining equation becomes a constraint expressing the variable in question as an algebraic function of the remaining variables. This function can again be approximated as a power-law function and substituted in the S-system, which retains its mathematical structure but is reduced in size. Thus, if survival processes are dominated by just one or two dominating hazards, they can ultimately be represented with one- or two-variable S-systems. In the context of growth, Savageau (1979ab, 1980) supported the analogous conclusion by demonstrating that many growth laws found in the literature are indeed exact special cases of S-systems with one or two variables. Similarly, it was shown (Savageau, 1982, Voit, 1992b; Voit and Yu, 1994; Yu and Voit, 1995) that many statistical distributions and survival functions are well represented by a single S-system equation. This so-called *S-distribution* is expressed in terms of the random variable X and its cumulative distribution F and takes the form

$$\frac{dF}{dX} = \alpha(F^g - F^h) \qquad F(X_0) = 0.5. \tag{16}$$

X_0 is the median of the distribution, the positive parameter α determines its spread, and the real-valued powers g and h ($g<h$) characterize its shape. The S-distribution provides a novel shape-based classification of traditional and new distributions and a convenient tool for Monte-Carlo simulations in which input parameters have distributions whose mathematical structure is *a priori* unknown (Voit et al., 1995). In survival analysis, the random variable X represents time ($X=t$) and F represents the cumulative failure distribution, which is the complement of the survival function $S(t)$: $F(t) = 1 - S(t)$.

5 Trends in Distributions

The final example of a connection between statistics and dynamic modeling addresses the scenario in which a random variable changes over time. As an example, consider how the distribution of contaminant concentrations in an even-aged cohort of fish changes over time. Being constantly exposed, the fish accumulate the contaminant, and the mean and variance of the contaminant distribution increase. The actual contaminant concentration in a fish is driven by environmental exposure and by the dynamic processes of its metabolism, which include uptake of the contaminant, distribution throughout the body, chemical change, and excretion. In mathematical terms, the contaminant distribution is thus subject to complicated transformations over time.

An algebraic transformation of a random variable X is easily computed, if it satisfies some requirements of monotonicity and differentiability (e.g., Mood et al., 1983). For an invertible transformation, $Y = \varphi(X)$, the density of Y is computed from the density of X as

$$f_Y(y) = \left| \frac{d}{dy} \varphi^{-1}(y) \right| \; f_X(x). \tag{17}$$

To introduce the dynamic nature of the exposure process, consider how the dynamical process, during a given, fixed time period τ, shifts values of the random variable X to new values of Y. In the case of contaminant accumulation, one may study how much each concentration within the cohort of fish increases within one year. The accumulation process maps the concentration distribution of year i onto the concentration distribution of year i+1, and each member of the original distribution f_X becomes a unique member of the shifted distribution f_Y. This shift is analogous to the function $\varphi(X)$ in (17). It is not a function of time *per se*, but the magnitude of the time shift implicitly determines the numerical characteristics of this function.

In realistic situations, the dynamic process is governed by a set of differential equations, and it is necessary to characterize $\varphi(X)$ in this case. As a simple example, suppose the process is described by a single autonomous differential equation

$$\dot{X} = \Psi(X), \quad X_0 = X(t_0). \tag{18}$$

Within τ time units, each possible value of X at time t changes from $X(t)$ to $X(t+\tau)$, and this value corresponds to the desired Y. Y satisfies exactly the same differential equation as X, except that its time scale is shifted by τ time units:

$$\dot{Y} = \Psi(Y), \quad Y_0 = Y(t_0) = X(t_0+\tau). \tag{19}$$

Solving Eqs. (18) and (19) simultaneously yields two copies of the same process, shifted by τ time units.

According to Eq. (17), the transformation of the distribution of X into the distribution of Y requires the term dX/dY, which is readily obtained as

$$\frac{dX}{dY} = \frac{dX}{dt}\frac{dt}{dY} = \frac{\Psi(X)}{\Psi(Y)}. \tag{20}$$

Eq. (20) constitutes a new differential equation that describes how X changes as a function of Y. Multiplication of the term $\Psi(X)/\Psi(Y)$ with the probability density $f_X(x)$ thus yields $f_Y(y)$. Eq. (20) is initialized with $Y(t_0)$ as the initial value of the independent variable and $X(t_0)$ as the value of the state variable.

As an example, suppose the distribution of X (i.e., the contaminant distribution at time i) is formulated as an S-distribution (16). According to the derivation above, the distribution of Y is given as

$$\frac{dF}{dY} = \alpha (F^g - F^h) \frac{\Psi(X)}{\Psi(Y)}, \quad F(Y_0)=F_0, \tag{21}$$

which is solved simultaneously with the equations describing the shift, namely,

$$\begin{aligned} \frac{dX}{dY} &= \frac{\Psi(X)}{\Psi(Y)}, \quad X_0 = X(t_0), \\ \frac{dY}{dY} &= 1, \quad Y_0 = Y(t_0) = X(t_0+\tau). \end{aligned} \tag{22}$$

Technical details of the procedure were presented elsewhere (Voit, 1996; Voit and Sorribas, 1998), along with illustrative examples and an extension to cases in which the dynamic process and the density are represented not as single differential equations but as systems of differential equations (cf. Savageau, 1982; Voit and Rust, 1992).

The dynamic process can affect the original distribution of X in different ways. It may simply shift it without changes in variance or shape. A process of this type is the aging of a cohort without deaths. It may also change its variance, increase, decrease, or reverse skewness, and even lead to a bimodal distribution of Y. All such changes are frequently observed in nature (e.g., Ford, 1975; Mohler et al., 1978; Gates et al., 1983).

6 Conclusions

The environment presents the most complex systems science has attempted to understand. Like the five blind men in Chinese folklore, who try to describe and explain an elephant by each just feeling one body part, modeling and statistics have traditionally focused on different aspects of environmental phenomena. As in the Chinese tale, though, it appears that a full understanding requires all branches of science and mathematics not just to concentrate on one aspect, but to communicate and merge their findings in a concerted effort that integrates them in a holistic picture. This is not always easy, since scientific terminologies and jargons are different enough to impede communication and collaboration. Even among closely related branches, such as epidemiology and modeling, different usage of the same term sometimes creates an atmosphere of misunderstanding and suspicion.

A good example is the term *model* itself. For the classically trained epidemiologist, *model* refers to a regression model that associates disease outcomes with contributing factors and confounders. The model may be simple or quite complicated, but seldom reaches beyond the general concept of a static tool that helps to quantify an association. By contrast, the term *model* has quite a different meaning in the biomathematical community, where it usually refers to a dynamical system of differential or difference equations, while a regression model is an *ad hoc* function that has desired mathematical features but does not elucidate underlying mechanisms. The discrepancy in notion does not imply that biomathematicians consider regression models as useless or inferior. On the contrary, regression models are appreciated as a standard tool of parameter estimation and have met with more success than dynamical models in applied areas such as actuarial science or forest management.

Terminology strongly affects thinking, and different terminologies and methodologies readily lead to scientific segregation and isolation. It is necessary to counteract divisive trends, and since this is seldom achieved by dogmatic redefinition of terms, it may be more efficacious to elucidate similarities and connections between seemingly disparate scientific frameworks. The present article constitutes an attempt in this direction by deriving links between two branches of environmetrics that are probably closer to each other than most others. It demonstrates that dynamical models, if interpreted in particular ways, reduce to well-known statistical or epidemiological models. In the case of environmental health risk assessment, the constraining of a

dynamical disease model to the steady state leads directly to widely used exposure models, the linear-logistic model, and Cox's proportional hazards model. This connection provides new insights, because the exposure models and epidemiological models can now be reinterpreted within the framework of classical systems analysis. In cases where these models fail, their failure can be explained in some detail, and the models can be extrapolated, expanded, or reformulated with rigorous mathematical methods, based on the fuller dynamical models from which they were derived.

The simple appearance of survival curves is explained not by ignoring but by appreciating the complexity of the underlying processes. Formulating, in a *Gedanken Experiment*, the physiological and environmental processes leading to death or survival as a dynamical model, an approximation strategy developed by Savageau (1979a) can be employed, and the result is a greatly simplified dynamical model that can be reinterpreted as a survival function or statistical distribution.

Almost all variables in environmental systems are distributed and change over time, and it is thus necessary to consider the dynamics of distributions. Traditionally, statistics has dealt with transformations of variables only in the limited case of explicit, algebraic functions, and modeling has often ignored the distributed character of variables for reasons of simplicity. The third example of this article closes this gap between statistics and modeling by showing that the statistical textbook techniques for transforming random variables can be extended to dynamic models of much greater complexity.

One should expect that the examples presented here barely scratch the surface of a multitude of connections between modeling and statistics. It appears that environmental phenomena are so rich that they invite and, indeed, require the search for a deeper integration of these methodological frameworks. This integration will come from individuals or teams educated in both branches of environmetrics which, in turn, poses a challenge for educators in environmental studies.

References

D.R. Cox. Regression models and life tables (with discussion). *J. R. Stat. Soc. B.* **34**: 187-220, 1972.

A.E.N. Ferreira. Linked lists representation of S-systems: An alternative way to implement numerical algorithms (Abstract). Second S-System Symposium, Tampa, FL, 1992.

E.D. Ford. Competition and stand structure in some even-aged plant monocultures. *J. Ecology* **63**: 311-333, 1975.

D.J. Gates, R. McMurtrie, and C.J. Borough. Skewness reversal of distribution of stem diameter in plantations of *Pinus radiata*. *Australian For. Res.* **13**: 267-270, 1983.

W.H. Hallenbeck. *Quantitative Risk Assessment for Environmental and Occupational Health*. Lewis Publ., Boca Raton, 1993.

D.H. Irvine and M.A. Savageau. Efficient solution of nonlinear ordinary differential equations expressed in S-system canonical form. *SIAM J. Numer. Anal.* **27**: 704-735, 1990.

C.L. Mohler, P.L. Marks, and D.G. Sprugel. Stand structure and allometry of trees during self-thinning of pure stands. *J. Ecology* **66**: 599-614, 1978.

A.M. Mood, F.A. Graybill, and D.C. Boes. *Introduction to the Theory of Statistics* (3rd ed.), McGraw-Hill, New York, 1983.

P.B. Ryan. An overview of human exposure modeling, *J. Exp. Anal. Envir. Epidem.* **1(4)**: 453-474, 1991.

K.J. Rothman. *Modern Epidemiology*. Little, Brown and Company, Boston, 1986.

M.A. Savageau. Biochemical systems analysis, I. Some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor. Biol.* **25**: 365-369, 1969a.

M.A. Savageau. Biochemical systems analysis, II. The steady-state solutions for an *n*-pool system using a power-law approximation. *J. Theor. Biol.* **25**: 370-379, 1969b.

M.A. Savageau. The behavior of intact biochemical control systems. *Curr. Top. Cell Reg.* **6**: 63-130, 1972.

M.A. Savageau. Optimal design of feedback control by inhibition: Steady-state considerations. *J. Mol. Evol.* **4**: 139-156, 1974.

M.A. Savageau. *Biochemical Systems Analysis. A Study of Function and Design in Molecular Biology.* Addison-Wesley, Reading, Massachusetts, 1976.

M.A. Savageau. Growth of complex systems can be related to the properties of their underlying determinants. *Proc. Nat. Acad. Sci.* **76**: 5413-5417, 1979a.

M.A. Savageau. Allometric morphogenesis of complex systems: Derivation of the basic equations from first principles. *Proc. Nat. Acad. Sci. USA.* **76**: 6023-6025, 1979b.

M.A. Savageau. Growth equations: A general equation and a survey of special cases. *Math. Biosci.* **48**: 267-278, 1980.

M.A. Savageau. A suprasystem of probability distributions. *Biometr. J.* **24**: 323-330, 1982.

M.A. Savageau and E.O. Voit. Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Math. Biosci.* **87**: 83-115, 1987.

N.V. Torres, E.O. Voit, and C. H. Alcón. Optimization of nonlinear biotechnological processes with linear programming. Application to citric acid production in *Aspergillus niger*. Biotechn. Bioengin. **49**: 247-258, 1996.

N.V. Torres, E.O. Voit, C. Glez-Alcón, and F. Rodríguez. An integrated optimization method for biochemical systems. Description of method and application to ethanol, glycerol and carbohydrate production in *Saccharomyces cerevisiae*, Biotechn. Bioengin. **55(5)**: 758-772, 1997.

E.O. Voit, editor. *Canonical Nonlinear Modeling. S-System Approach to Understanding Complexity.* Van Nostrand Reinhold, New York, 1991.

E.O. Voit. Optimization in integrated biochemical systems. *Biotechn. Bioengin.* **40**: 572-582, 1992a.

E.O. Voit. The S-distribution. A tool for approximation and classification of univariate, unimodal probability distributions. *Biometr. J.* **34**: 855-878, 1992b.

E.O. Voit. S-system modelling of complex systems with chaotic input. *Environmetrics* **4**: 153-186, 1993.

E.O. Voit, W. L. Balthis, and R. A. Holser. Hierarchical Monte Carlo modeling with S-distributions: Concepts and illustrative analysis of mercury contamination in king mackerel. *Environment International* **21**: 627-635, 1995.

E.O. Voit, D.H. Irvine, and M.A. Savageau. *The User's Guide to ESSYNS.* Medical University of South Carolina Press, Charleston, 1989.

E.O. Voit and R.G. Knapp. Derivation of the linear-logistic model and Cox's proportional hazard model from a canonical system description, *Statistics in Medicine* **16**: 1705-1729, 1997.

E.O. Voit and P.F. Rust. Tutorial: S-system analysis of continuous univariate probability distributions. *J. Statist. Comput. Simul.* **42**: 187-249, 1992.

E.O. Voit and M.K. Schubauer-Berigan. The role of canonical modeling as a unifying framework for ecological and human risk assessment. In M.C. Newman and C. Strojan, editors, *Risk Assessment: Logic and Measurement.* Ann Arbor Press, Chelsea, MI, 1998.

E.O. Voit and A. Sorribas. Computer modeling of dynamically changing distributions of random variables, *Mathematical and Computer Modelling*, in press, 1998.

E.O. Voit, and S. Yu. The S-distribution. Approximation of discrete distributions. *Biometr. J.* **36**: 205-219, 1994.

S. Yu and E.O. Voit. A simple, flexible failure model. *Biometrical J.* **37**: 595-609, 1995.

Author's address:

Eberhard O. Voit, Ph.D.
Professor and Assistant Dean for Environmental Studies
Department of Biometry and Epidemiology
Medical University of South Carolina
Charleston, SC 29425-2203 U.S.A.
Tel: + 803 876-1122
FAX: + 803 876-1146
Electronic Mail: VoitEO@MUSC.edu