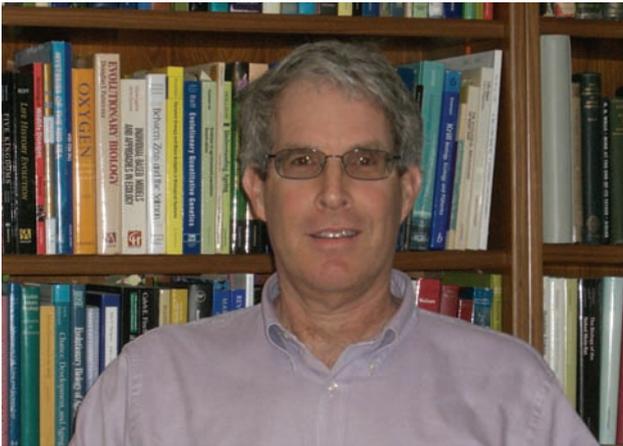


Uncertainty in Ecology: A Retrospective and Prospective

Marc Mangel



Introduction

In 1956, Elvis Presley's hit songs included *Heartbreak Hotel*, *Blue Suede Shoes*, *Hound Dog*, *Blue Moon*, and *Love Me Tender*. It was a good year for The King and we recently celebrated the 50th anniversary of those songs. 1956 was also a good year for Jerzy Neyman (a statistician), Thomas Park (an ecologist), and Elizabeth Scott (also a statistician) who published a remarkable paper (Neyman et al 1956) whose 50th anniversary was barely noted by anyone (and almost certainly not by most Elvis fans).

Two years before, Park (1954) had published a summary of his work on competition between *Tribolium confusum* (the confused flour beetle) and *Tribolium castaneum* (the red flour beetle). These experiments are in many ways the canonical example of competition between two species, as we teach our students through the study of the Lotka-Volterra equations for competition. Those equations begin with the reasonable assumption of logistic growth of each species in the absence of the other, with an additional term added to capture the interspecific competition. Depending upon the relative intensities of interspecific and intraspecific competition, it is well known that four outcomes are possible: a) both species persist and coexist; b) one species (e.g. *T. confusum*) always wins the competition; c) the other species (e.g. *T. castaneum*) always wins the competition; or d) the outcome of the competition depends upon the initial conditions, with each species winning some of the time. Today, we teach our students to understand these results

in terms of null-clines of the Lotka-Volterra competition equations, which can have an interior steady state that is either a stable node or a saddle point (Mangel 2006). In case d), we understand that the phase plane has a saddle point and that the deterministic trajectory that goes into the saddle point separates the phase plane into regions in which one species wins and regions in which the other species wins. That trajectory is called the separatrix.

However, what attracted Neyman et al were what Park considered to be 'zones of indeterminate outcome'. That is, Park could vary (in addition to composition of the flour) the temperature and humidity under which he conducted experiments. There were some initial conditions that lead to competitive exclusion of one of the species or the other of the species in all of the experimental replicates. But there were other initial conditions that lead to extinction of one of the species (say *T. confusum*) in only a fraction of the replicates. And this zone of indeterminate outcome varied with temperature and humidity (Figure 1).

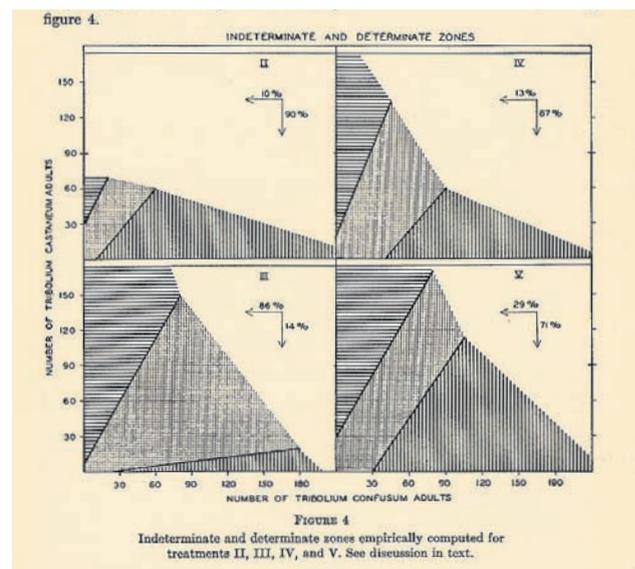


Figure 1. The figure of Neyman, Park, and Scott (1956) showing the zones of indeterminate outcome and the frequency of those outcomes, for four of the treatments that Park used. Reproduced by permission.

Thus the outcome varied with initial numbers and environmental conditions. This leads to the focal question: given the initial numbers of the two species, what will be the outcome of the competition? Writing about this many years later, Park said

In retrospect, it is perhaps informative that Park initially was not motivated by questions of stochastic events but,

rather, with the demonstration that certain reasonably realistic ecological factors could affect the survival of one species over another when both were competing. It is also amusing to record, again in retrospect, that Park's growing appreciation of indeterminacy was nurtured by *the data themselves*, by his association with the statisticians P. H. Leslie, J. Neyman and M. S. Bartlett, who were interested, *each in his own way*, in the *Tribolium* experiments; and, finally, that Park had an increasing awareness of the fact that ecological systems are indeed stochastic – that a population's survival, even in the absence of environmental stochasticity, is related intimately to a component which has recently, and meaningfully, been described as 'gambling for existence' (Mertz et al 1976, p1371, italics added)

Park's work inspired the great mathematical scientists V. D. Barnett (1962), M. S. Bartlett (Bartlett et al 1960), P. H. Leslie (Bartlett et al 1960, Leslie and Gower 1958) and these papers are still joys to read (and thanks to JSTOR, easy to access). Each of Barnett, Bartlett, and Leslie and Gower, set out to provide process-based models that would allow understanding Park's data.

One model for example, used extensively by all of the greats I've mentioned thus far, is based on a world view in which there are no underlying deterministic dynamics (as in the Lotka-Volterra competition equations) but the dynamics completely stochastic, due to random births and deaths. With such a view we might let $X_1(t)$ denote the number of adults of *T. castaneum* and $X_2(t)$ the number of adults of *T. confusum* at time t . We then characterize the transitions in population size from t to $t + \Delta t$ in terms of rates of birth and death that depend upon current population sizes. By simple use of conditional probability, and because we are interested solely in the outcome of the competition and not how long things take, one is able to focus just on the changes in population size and thus the computer memory requirements (not a big deal now, but very important when this work was first done) for simulating the dynamics of the competing species are rather minimal. Thus, this approach was suitable for rapid Monte Carlo simulation even 50 years ago. And with some more work, one can also show how Lotka-Volterra like equations are an emergent property of the full stochastic system (Mangel 1994).

An alternative model is one in which the deterministic competition equations really do underlie the dynamics, but are perturbed by noise. In that case, we replace the

deterministic Lotka-Volterra competition equations by stochastic differential equations (Mangel 2006), surround the separatrix by a band running parallel to it and introduce a probability $u(x_1, x_2) = \Pr\{\text{exit the band towards the } y\text{-axis given that } X_1(0)=x_1, X_2(0)=x_2\}$. It is relatively easy to derive the equation that this exit probability must satisfy, but much harder to find its solution. However, the struggle is worth it and in the end one is able to compute the contours of probability – that is, contours of indeterminacy – in the phase plane for extinction of one species or the other (Mangel and Ludwig 1977). This method has also been applied by Price et al (1993) to the study of peak shifts during correlated responses to selection.

How might we think about these classic experiments in the 21st century when our ability to compute is something that the greats who once worked on this problem are unlikely to have foreseen? To begin, we might at least symbolically write a pair of differential equations – in the style of the Lotka-Volterra equations – of the form

$$\frac{dX_1}{dt} = f_1(X_1, X_2, p_1) \quad \text{and} \quad \frac{dX_2}{dt} = f_2(X_1, X_2, p_2)$$

where $f_1(X_1, X_2, p_1)$ and $f_2(X_1, X_2, p_2)$ are functions (think of the classical Lotka-Volterra competition equations) of the population sizes and p_1 and p_2 are vectors of parameters (think of the maximum per capita rates of reproduction, the carrying capacities, and the interspecific competition parameters in the classical equations). But in the 21st century we can think broader and deeper than the classical Lotka-Volterra system. For example, we might draw the parameters from distributions and use the solution of the differential equations, with random parameters, to compute the zones of indeterminacy and update those distributions of parameters given experimental results. Or we might construct a Generalized Linear Model (GLM) for the parameters in terms of environmental (e.g. temperature and humidity) and genetic variables. Or, we might recognize that when the starting values of population size are around the separatrix, the population will spend a considerable amount of time in that vicinity so that we can let the data themselves tell us the form of $f_1(X_1, X_2, p_1)$ and $f_2(X_1, X_2, p_2)$ through Bayesian Nonparametric approaches (Munch et al 2005, Patil 2007).

More broadly, how should we think about uncertainty in ecology now that we are solidly in the 21st century? To begin, I'd say that most scientists are facultative Bayesians (Hilborn and Mangel 1997, Clark 2005, Mangel 2004). That is, they find frequentist statistics troubling because it often does not

Feature Article: Marc Mangel

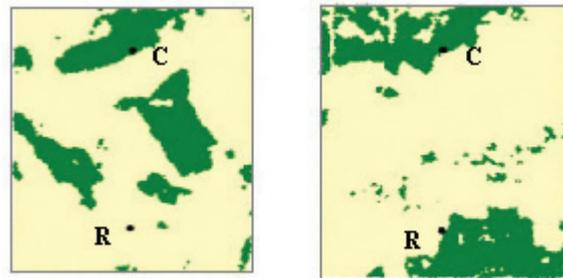
Bulletin of the British Ecological Society 2008 39:1

tell them what they really want to know (the probability of the hypothesis given the data) but something slightly altered (the probability of the data given the hypothesis). They are troubled by Bayesian statistics when the prior is described as “subjective,” with all that this charged word connotes; however the choice of $p=0.05$ is also subjective, if lending an air of objectivity to frequentist statistics. But above all, a scientist will do what is needed to understand nature. And, some statisticians are this way too. About two years ago, following a talk at Santa Cruz, my erstwhile UC Davis colleague Frank Samaniego told me “I am a Bayesian sympathizer”.

We should be careful about words and how we use them. What is noise to an engineer is the core variation on which natural selection acts to a biologist. At the BES meeting in Glasgow, Sam Berry [who was there for a meeting of former presidents of the society] reminded me that Bateson said ‘treasure your exceptions’. Berry himself once wrote ‘variation is the core of biology’. When doing our work, we must deal with epistemic uncertainty, which is uncertainty due to limited information and is a property of the analyst rather than the system. It is reducible and we tend to associate words such as ‘Ignorance’, ‘Observation Error’, or ‘Type B’ with it. We must also deal with aleatory uncertainty, which is uncertainty due to randomness in system. It is irreducible and we tend to associate words such as ‘Stochasticity’, ‘Variability’, ‘Process error’, ‘Type A’, or ‘Irreducible’ with it. It is a property of the system. Both kinds of uncertainty are important in ecology and in the 21st century we should not confuse them. I suggest that we move towards a common terminology in which we use 1) process stochasticity or variation to describe the inherent fluctuations in the system, 2) observation error to describe our inability to see things perfectly, and 3) model uncertainty to capture our lack of knowledge of the fundamental laws of biology (other than evolution by natural selection).

Ecological modeling is best done as an iterative process in which the form of the model depends on the question and the statistical analysis should inform what we do next. Now, statistics made the enormous progress it did because of a focus on modeling the data. Science makes its progress by a focus on modeling the processes that generate the data and in the 21st century we will begin to link these two kinds of models more and more. We must always know our data and stop and think about the choice of statistical distribution for modeling it. For example, whether two variables are correlated or not is a function of the statistical model and not

of the biology. When asking ‘is the normal distribution the right one to use for this problem?’ it does not matter if one is a Bayesian or frequentist. We should also remember that a statistical tool may be mathematically correct but biologically irrelevant. The two rectangle below (Figure 5 of Bélisle et al 2001; reprinted with permission) have the same amount of forest (dark green) and open space (pale) as measured by a statistical tool



But if we think of an animal released at R and moving to a capture point at C the amount of protection provided in the two panels differs enormously. The converse may hold true as well – things that are statistically different may be biologically identical because of the way the organism interacts with the resources of the environment (Roitberg and Mangel 1997).

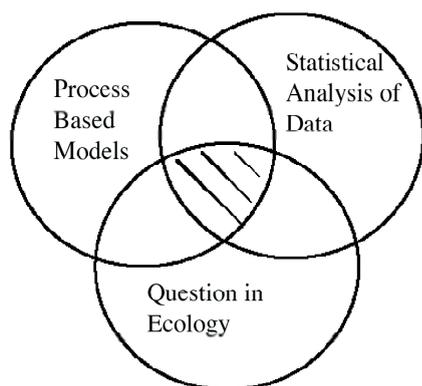
Bayesian nonparametric (BNP) methods, which provide probability distributions on functions rather than parameters (as in the usual Bayesian method), are tools for robust statistical inference (Munch et al 2005, Patil 2007). For example, a key component of population dynamics is the relationship between the number of mature adults, S , and individuals joining the population due to production by those adults, R . In fishery science this is called the spawner-recruit relationship and we have variety of offerings, often associated with names such as Ricker, Beverton-Holt, Shepherd attached (see Munch et al 2005 for details). In some cases (e.g. Ricker and Beverton-Holt) we even have mechanistic interpretations of those relationships (see Mangel 2006, Lucero 2007) and in the 20th century we spent much time arguing about which model is right, or at least which model fits the data the best (e.g. Myers et al 1995). But this is not what we should be doing in the 21st century. Rather, BNP methods allow us to put probability distributions on stock-recruitment functions and then let the data tell us the form of the function. These methods, and indeed all Bayesian methods in general, force us to be honest about uncertainty – when we go where there is no data (into the unknown unknowns), honesty requires that measures of confidence in our predictions are very weak.

Feature Article: Marc Mangel

Bulletin of the British Ecological Society 2008 39:1

Hierarchical models (e.g. Clark 2005, Clark and Gelfand 2006) provide a link between unknown parameters (θ), unobserved processes (Z) and observed data (Y) through application of conditional probability $P[y, z, \theta] = f(y|z)g(z|\theta)h(\theta)$. Hierarchical models work for both frequentist and Bayesian statisticians (once just has to 'believe' in conditional probability) and recognizing biological hierarchy is as important as doing statistical hierarchy.

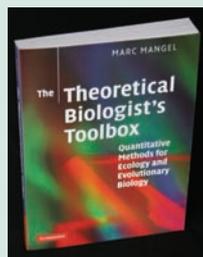
In summary, our future lies at the intersection of a problem in ecology, models based on process, and the analysis of data connecting the scientific question and the model.



Nature is indeed variable and complex. But much of that variability and complexity can be understood.

Acknowledgments. This piece is based on a closing talk that I gave at a Workshop on Uncertainty in Ecological Analysis at the Mathematical Biosciences Institute, Ohio State University in April 2006. I thank Noel Cressie for inviting me to do that and Alan Crowden for inviting me to write for the *Bulletin*.

Marc Mangel is Professor in the Department of Applied Mathematics and Statistics, and Center for Stock Assessment Research, University of California Santa Cruz, CA 95064. Marc's research group works in mathematical and theoretical biology, with a focus on ecology, evolution, and behavior and the broad goal of combining



first-rate basic science with important applied questions. Author and editor of several books, Marc's most recent volume is *The Theoretical Biologist's Toolbox. Quantitative Methods for Ecology and Evolutionary Biology* published by Cambridge University Press in 2006.

R E F E R E N C E S

- Barnett, V.D. The Monte Carlo solution of a competing species problem. *Biometrics* 18: 76-103.
- Bartlett, M.S., Gower, J.C., and P.H. Leslie. 1960. A comparison of some theoretical and empirical results for some stochastic population models. *Biometrika* 47:1-11.
- Bélisle, M., Desrochers, A., and M-J Fortin. 2001. Influence of forest cover on the movements of forest birds: a homing experiment. *Ecology* 82:1893-1904.
- Clark, J.S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8: 2-14.
- Clark, J.S. 2007. *Models for Ecological Data*. Princeton University Press, Princeton.
- Clark J S and Gelfand A E, eds (2006). *Hierarchical Modelling for the Environmental Sciences – statistical methods and applications*. Oxford University Press, Oxford.
- Hilborn, R. and M. Mangel. 1997. *The Ecological Detective. Confronting Models with Data*. Princeton University Press, Princeton.
- Leslie, P.H. and J.C. Gower. 1958. The properties of a stochastic model for two competing species. *Biometrika* 45: 316-330.
- Lucero, Y. 2007. Population Consequences of an Age-dependent Maternal Effect in Rockfish (genus *Sebastes*). PhD Dissertation, Ocean Sciences, University of California Santa Cruz.
- Mangel, M. 1994. Barrier transitions driven by fluctuations, with applications to ecology and evolution. *Theoretical Population Biology*. 45:16-40.
- Mangel, M. 2004. Weighing the evidence. A review of M. Taper and S. Lele (editors). *The Nature of Scientific Evidence. Statistical, philosophical, and empirical considerations*. Science 306:1478.
- Mangel, M. 2006. *The Theoretical Biologist's Toolbox. Quantitative Methods for Ecology and Evolutionary Biology*. Cambridge University Press, Cambridge.
- Mangel, M. and D. Ludwig. 1977. Probability of extinction in a stochastic competition. *Society for Industrial and Applied Mathematics (SIAM) Journal on Applied Mathematics* 33:256-266.
- Mertz, D., Cawthon, D.A. and T. Park. 1976. An experimental analysis of competitive indeterminacy in *Tribolium*. *Proceedings of the National Academy of Sciences* 73:1368-1372.
- Munch, S.B., Kottas, A. and M. Mangel. 2005. Bayesian nonparametric analysis of stock-recruitment relationships. *Canadian Journal of Fisheries and Aquatic Sciences* 62:1808-1821.
- Myers, R.A., Bridson, J., and N.J. Barrowman. 1995. *Summary of Worldwide Spawner and Recruitment Data*. Canadian Technical Report of Fisheries and Aquatic Sciences 2020, Fisheries and Oceans, Ottawa, Canada.
- Neyman, J., Park, T. and E.L. Scott. 1956. Struggle for Existence: The *Tribolium* Model: Biological and Statistical Aspects. Proceedings of the Third Third Berkeley Symposium on Mathematical Statistics and Probability, Volume IV, pp 41-79
- T. Park. 1954. *Competition: An Experimental and Statistical Study*. Iowa State University Press, Des Moines.
- Patil, A. 2007. Bayesian Nonparametrics for Inference of Ecological Dynamics. PhD Dissertation, Statistics and Stochastic Modeling, University of California Santa Cruz
- Price, T., Turelli, M. and M. Slatkin. 1993. Peak shift produced by correlated responses to selection. *Evolution* 47:280-290
- Roitberg, B.D. and M. Mangel. 1997. Individuals on the landscape: Behavior can mitigate landscape differences among habitats. *Oikos* 80:234-240