The Ecological Detective

CONFRONTING MODELS WITH DATA

RAY HILBORN AND MARC MANGEL

MONOGRAPHS IN POPULATION BIOLOGY • 28

The Ecological Detective

MONOGRAPHS IN POPULATION BIOLOGY

EDITED BY SIMON A. LEVIN AND HENRY S. HORN

Titles available in the series (by monograph number)

1. *The Theory of Island Biogeography*, by Robert H. MacArthur and Edward O. Wilson
6. *Stability and Complexity in Model Ecosystems*, by Robert M. May
10. *Geographic Variation, Speciation, and Clines*, by John A. Endler
14. *Some Adaptations of March-Nesting Blackbirds*, by Gordon H. Orians
16. *Cultural Transmission and Evolution: A Quantitative Approach*, by L. L. Cavalli-Sforza and M. W. Feldman
17. *Resource Competition and Community Structure*, by David Tilman
18. *The Theory of Sex Allocation*, by Eric L. Charnov
19. *Mate Choice in Plants: Tactics, Mechanisms, and Consequences*, by Mary F. Wilson and Nancy Burley
21. *Natural Selection in the Wild*, by John A. Endler
22. *Theoretical Studies on Sex Ratio Evolution*, by Samuel Karlin and Sabvin Lessard
23. *A Hierarchical Concept of Ecosystems*, by R. V. O'Neill, D. L. DeAngelis, J. B. Waide, and T.F.H. Allen
24. *Population Ecology of the Cooperatively Breeding Acorn Woodpecker*, by Walter D. Koening and Ronald M. Mumme
25. *Population Ecology of Individuals*, by Adam Lomnicki
26. *Plant Strategies and the Dynamics and Structure of Plant Communities*, by David Tilman
28. *The Ecological Detective: Confronting Models with Data*, by Ray Hilborn and Marc Mangel
29. *Evolutionary Ecology across Three Trophic Levels: Goldenrods, Gallmakers, and Natural Enemies*, by Warren G. Abrahamson and Arthur E. Weis
30. *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*, edited by David Tilman and Peter Kareiva

The Ecological Detective

Confronting Models with Data

RAY HILBORN
AND
MARC MANGEL

PRINCETON UNIVERSITY PRESS
PRINCETON, NEW JERSEY
1997

Published by PrincetonUniversity Press, 41 William Street,
Princeton, New Jersey 08540
In the United Kingdom: Princeton University Press,
Chichester, West Sussex

Library of Congress Cataloging-in-Publication Data

Hilborn, Ray, 1947–
The ecological detective : confronting models with data / Ray Hilborn and Marc Mangel.
p. cm. — (Monographs in population biology ; 28)
Includes bibliographical references and index.
ISBN 0-691-03496-6 (cloth : alk. paper). — ISBN 0-691-03497-4 (pbk. : alk. paper).
1. Ecology—Mathematics models. I. Mangel, Marc. II. Title. III. Series.
QH541.15.M3H54 1997 96-9638
574.5'0151—dc20 CIP

This book has been composed in Baskerville

Princeton University Press books are printed on acid-free paper and meet the guidelines for permanence and durability of the Committee on Production Guidelines for Book Longevity of the Council on Library Resources

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1
10 9 8 7 6
(Pbk.)

http://pup.princeton.edu
ISBN-13: 978-0-691-03497-3 (pbk.)

ISBN-10: 0-691-03497-4 (pbk.)

To Ulrike and Susan

No thanks would be too much

Contents

Preface

Beyond the Null Hypothesis

ABOUT THE TITLE

First, a word about the phrase "ecological detective," which we owe to our colleague Jon Schnute:

> I once found myself seated on an airplane next to a charming woman whose interests revolved primarily around the activities of her very energetic family. At one point in the conversation came the inevitable question: "What sort of work do you do?" I confess that I rather hate that question. . . . I replied to the woman: "Well, I work with fish populations. The trouble with fish is that you never get to see the whole population. They're not like trees, whose numbers could perhaps be estimated by flying over the forest. Mostly, you see fish only when they're caught. . . . So, you see, if you study fish populations, you tend to get little pieces of information here and there. These bits of information are like the tip of the iceberg; they're part of a much larger story. My job is to try to put the story together. I'm a detective, really, who assembles clues into a coherent picture." (Schnute 1987, 210)

As we began outlining the present volume, we realized that the phrase the "ecological detective" was most appropriate for what we are trying to accomplish. Some reviewers agreed, and some found it a bit too cute. After serious consideration, we decided to leave references to the ecological detective in the text, with apologies to readers who are offended. We find it preferable to "the reader."

It is our view that the ecological detective goes beyond the null hypothesis. As the revolution in physics in the twentieth century showed, there are few cases in science in which absolute truth exists. Models are metaphorical (albeit sometimes accurate) descriptions of nature, and there can never be a "correct" model. There may be a "best" model, which is more consistent with the data than any of its competitors, or several models may be contenders because each is consistent in some way with the data and none clearly dominates the others. It is the job of the ecological detective to determine the support that the data offer for each competing model or hypothesis. The techniques that we introduce, particularly maximum likelihood methods and Bayesian analysis, are the beginning of a new kind of toolkit for doing the job of ecological detection.

THE AUDIENCE AND ASSUMED BACKGROUND

In a very real way, this book began in October 1988, when we participated in an autumn workshop on mathematical ecology at the International Center for Theoretical Physics. Most of the participants were scientists who had been students in the two previous autumn courses. As these former students presented their work, we realized that although they had received excellent training in ecological modeling and the analysis of ecological models (cf. Levin et al. 1989), they were almost completely inexperienced in the process of connecting data to those models. For scientists in third-world countries, who will work on practical and important problems faced by their nations, such connections are essential, because real answers are needed. We decided then to try to provide the connection.

We envision that readers of this book will be third-year students in biology and upward. Thus, we expect the reader to have had a year of calculus, some classical statistics (typically regression, standard sampling theory, hypothesis testing, and analysis of variance) and some of the classical ecological models (logistic equation, competition equations) equivalent to the material in Krebs's (1994) textbook. Therefore, we will not explain either these classical statistical methods or the classical ecological models. Some readers of drafts took us to task, writing comments such as, "I took my last mathematics and statistics courses four years ago—how dare you expect me to remember any of it." Well, we expect you to remember it and use it. You should not expect to make progress with an attitude of "I learned it once, promptly forgot it, and don't want to learn it again." We worked hard to make the material accessible and understandable, but the motivation rests with you. The more effectively you can deal with data, the greater your contribution to ecology.

This book has equations in it. The equations correspond to real biological situations. There are three levels at which one can understand the equations. The first (lowest) level occurs when you read our explanation of the meaning of the equations. We have tried to do this as effectively as possible, but success can only really be guaranteed in that regard when there is interpersonal contact between student and teacher. The second (middle) occurs when you are able to convert the equation to words—and we encourage you to do so with *every* equation that you encounter. The third (highest) occurs when you explain the origin and meaning of the equation to a colleague. We also encourage you to strive for that.

COMPUTER PROGRAMMING

Computing is essential for ecological detection. We expect that you have access to a computer. Early drafts of the book, read by many reviewers, had computer programs (rather than pseudocodes) embedded in the text. Virtually all reviewers told us that this was a terrible idea, so we removed them. To really use the methods that we describe here, you must be computer-literate. It does not have to be a fancy computer language: Mangel does all of his work in TrueBASIC and Hilborn does all of his in QuickBASIC or Excel. We recommend that you become familiar with some computer environment that offers nonlinear function minimization, and that you program the examples as you go through the book. For complete neophytes, Mangel and Clark (1988) wrote an introduction to computer programming, focusing on behavioral ecology. In any case, this material will be learned much more effectively if you actually stop at various points and program the material that we are discussing. To be helpful, we give an algorithmic description, which we call a pseudocode, showing how to compute the required quantities. You cannot use these descriptions directly for computation, but they are guides for programming in whatever language you like. Understanding ecological data requires practice at computation, and if you read this book without trying to do any of the computations, you will get much less out of it.

REALISM AND PROFESSIONALISM

Each of the case studies we use to illustrate a particular point is a bona fide research study conducted by one of us. Even so, some readers of drafts accused us of the unpleasant and unprofessional, but too common (especially in evolutionary biology), behavior of setting up "strawpersons" just to knock them down or of misrepresenting opponent positions (see Selzer 1993 for an example). For example, we were told to

> treat each case study like a real research study and do not spend time rejecting obviously silly models. For example, no one should seriously try to fit a simple logistic equation to the data shown in Figure 8.1. Similarly, one would not need any formal analysis to reject the constant clutch model when presented with the data in Table 6.1. . . .

> This would get away from what our class came to call your "toy example" approach—illustrating models or techniques with silly examples, and then not explaining the hard decisions associated with the more interesting and complicated questions.

This charge is unfair. These apparently ridiculous models were *in fact proposed and used by pretty smart people.* Why? Because they had no alternative model. Our view is that the confrontation between more than one model arbitrated by the data underlies science. If there is only one model, it will be used, whether the questions concern management (as in the Serengeti example) or basic science (as in the insect oviposition example). Without multiple models, there is no alternative. Furthermore, in the case studies used here, the data are moderately simple and mainly one-dimensional. This allows us to "eyeball" the data and draw conclusions such as those given above. But in more complicated situations, this may not be possible.

Another side of professionalism is the development of a professional library. As described above, we consider this book a link between standard ecological modeling or theoretical ecology and serious statistical texts. After reading *The Ecological Detective,* the latter should be accessible to you. We consider that a good detective's library includes the following:

Efron, B., and R. Tibshirani. 1993. *An Introduction to the Bootstrap.* Chapman and Hall, New York.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 1995. *Bayesian Data Analysis.* Chapman and Hall, New York.

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models.* Chapman and Hall, New York.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1986. *Numerical Recipes.* Cambridge University Press, Cambridge, U.K.

ACKNOWLEDGMENTS

We owe lots of people thanks for careful reading, criticisms, and picking nits about the book. For comments on the entire volume, we thank Dick Gomulkiewicz, Peter Turchin, and

- Peter Kareiva and Zoology 572B, spring 1993, University of Washington: Three anonymous students, Beth Babcock, J. Banks, Tamre Cardoso, Kevin Craig, Bill Fagan, Ellen Gryj, Michael Guttormsen, Murdoch McAllister, Jennifer Ruesink (who provided a handy sketch of the ecological detective), Alice Shelly, Dave Somers, Ellen Smith, E. A. Steel (who provided some literature citations we did not know about), Angela B. Stringer, and Colin Wilson.
- Bill Morris and Biology/Zoology 295, fall 1993, Duke University: Sharon Billings, Kerry Bright, Sonia Cavigelli, Chiru Chang, Chris Damiani, Marc Dantzker, Brian Inouye, Brad Klepetka, Sarah May, Yavor Parashkevov, Nathan Phillips, Colleen Rasmussen, George Wilhere, and Neville Yoon.
- Dan Doak and Environmental Studies 270, winter 1995, University of California, Santa Cruz: Elizabeth Andrews, Dave Bigger, Geoff Bryden, Chris Cogan, Bruce Goldstein, Elaine Harding-Smith, David Hyrenbach, Suzy Kohin, Chris Lay, Michelle Marvier, Dan Monson, Jim Murphy, Michelle Paddack, Jacob Pollak, and Laura Trujillo.
- John Gillespie and Population Biology 298, winter 1995, University of California, Davis: Dave Cutler, Arleen Feng, Shea Gardner, Martha Hoopes, Barney Luttbeg, Steve Martin, Lance Morgan, Laura Peck, Rivera Perales, Rich van Buskirk, and Mike Zwick.
- The students in Ray Hilborn's graduate seminar, winter 1995: Eric Anderson, Patricia Dell'Arciprete, Nicholas

Beer, Claribel Coronado, Bill Driskell, Marianne Johnson, Alan Lowther, Mike McCann, Warren Schlechte, and Adrian Spidle.

We thank Jay Rosenheim and Tim Collier for comments on Chapters 5 and 6.

The Ecological Detective

CHAPTER ONE

An Ecological Scenario and the Tools of the Ecological Detective

AN ECOLOGICAL SCENARIO

The Mediterranean fruit fly (medfly), *Ceratitis capitata* (Wiedemann), is one of the most destructive agricultural pests in the world, causing millions of dollars of damage each year. In California, climatic and host conditions are right for establishment of the medfly; this causes considerable concern. In Southern California, populations of medfly have shown sporadic outbreaks (evidenced by trap catch) over the last two decades (Figure 1.1). Until 1991, the accepted view was that each outbreak of the medfly corresponded to a "new" invasion, started by somebody accidentally bringing flies into the state (presumably with rotten fruit). In 1991, our colleague James Carey challenged this view (Carey 1991) and proposed two possible models concerning medfly outbreaks (Figure 1.2). The first model, M_1, corresponds to the accepted view: each outbreak of medfly is caused by a new colonization event. After successful colonization, the population grows until it exceeds the detection level and an "invasion" is recorded and eradicated. The second model, M_2, is based on the assumption that the medfly has established itself in California at one or more suitable sites, but that, in general, conditions cause the population to remain below the level for detection. On occasion, however, conditions change and the population begins to grow in time and spread over space until detection occurs. Carey argued that the temporal and spatial distributions of trap catch indicate that the medfly may be permanently estab-

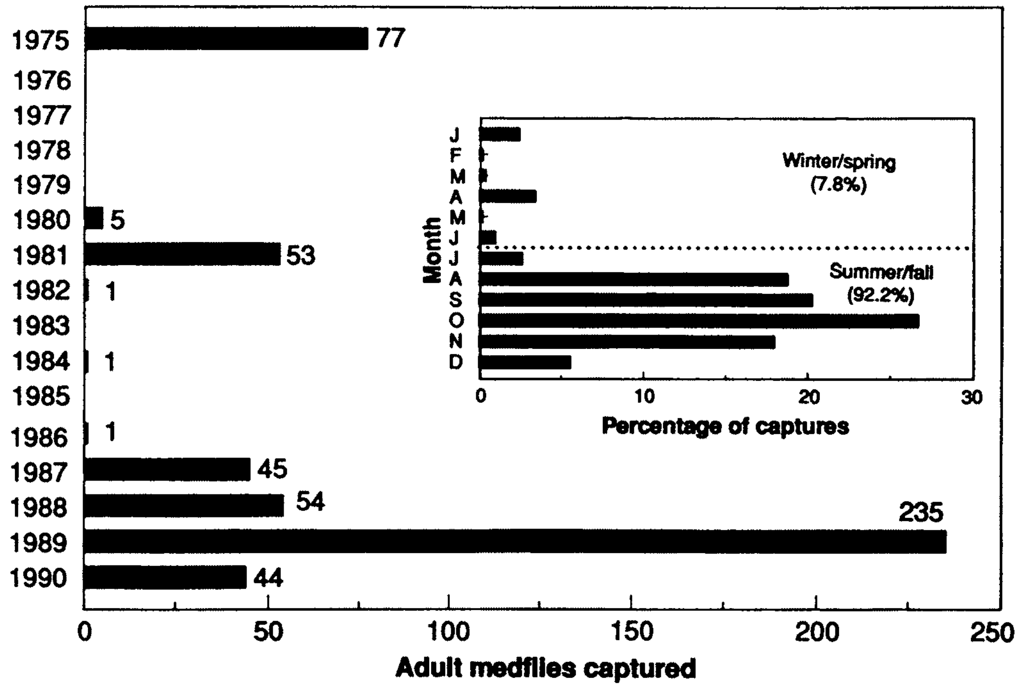

FIGURE 1.1. The capture of adult medfly in Southern California, from 1975 to 1990. (Reprinted with permission from Carey 1991. Copyright American Association for the Advancement of

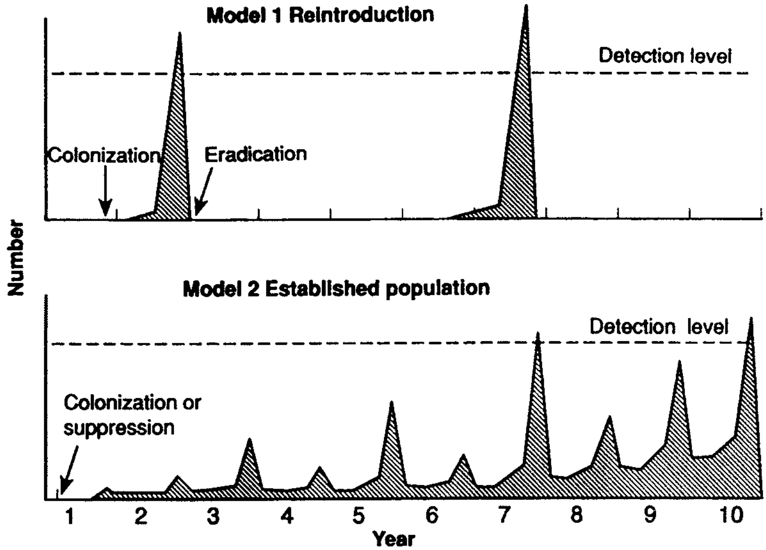

FIGURE 1.2. The outbreak of medfly can be described by two different methods. In model 1, we assume that there is continual reintroduction of the pest. After a reintroduction, the population grows until it exceeds the detection level. In model 2, we assume that the medfly is established, but that ecological conditions are only occasionally suitable for it to grow and exceed the detection threshold. (Reprinted with permission from Carey 1991. Copyright American Association for the Advancement of Science.)

lished in the Los Angeles area. Knowing which of these views is more correct is important from a number of perspectives, including the basic biology of invasions and the implications of an established pest on agricultural practices.

Determining which model is more consistent with the data is a problem in ecological detection. That is, if we allow that either model M_1 or model M_2 is true, we would like to associate probabilities, given the data, with the two models. We shall refer to this as "the probability of the model" or the "degree of belief in the model." How might such a problem be solved? First, we must characterize the available data, which are the spatial distribution of trap catches of medfly over time (Figure 1.3). We could refine these by placing

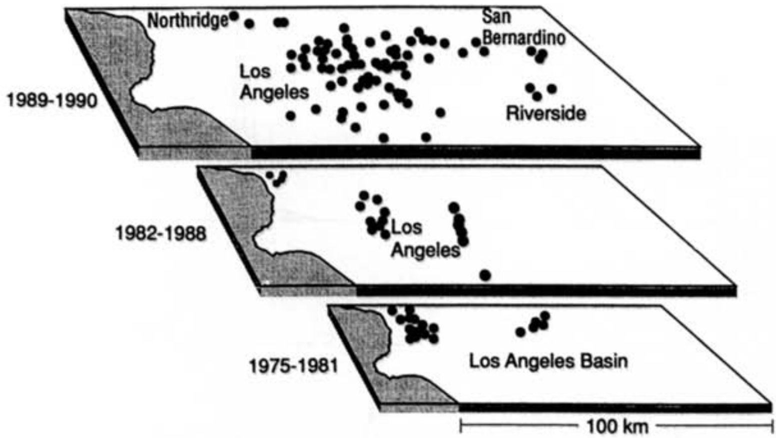

FIGURE 1.3. The data available for ecological detection in this case would be the spatial distribution of the catch of adult medfly over time. (Reprinted with permission from Carey 1991. Copyright American Association for the Advancement of Science.)

small grids over the maps and characterizing a variable that measures the number of flies that appear in cell i in year y. Second, we must convert the pictorial or verbal models shown in Figure 1.2 into mathematical descriptions. That is, some kind of mathematical model is needed so that the data can be compared with predictions of a model. Such models would be used to predict the temporal and spatial patterns in detected outbreaks; the mathematical descriptions would generate maps similar to the figures. The models would involve at least two submodels, one for the population dynamics and one for the detection process. Courses in ecological modeling show how this is done. Third, we confront the models with the data by comparing the predicted and observed results. At least three approaches can be broadly identified for such a confrontation.

Classical Hypothesis Testing. Here we confront each model separately with the data. Thus, we begin with hypotheses:

H_0: Model M_1 is true

H_a: Some other model is true

Here the alternate model might be that outbreaks are random over time and space. Using the mathematical descriptions of the models, we construct a "p value" for the hypothesis that M_1 is true. It might happen that we can definitely reject H_0 because the p value is so small (usually less than 0.05 or 0.01). Alternatively, we might not be able to reject H_0 (i.e., $p > 0.05$), but then might discover that the power of the statistical test is quite low (we assume that most readers are probably familiar with the terms "p values" and "power" from courses in elementary statistics, but we shall explain them in more detail in the following chapters). In any case, we use such hypothesis testing because it gives the "illusion of objectivity" (Berger and Berry 1988; Shaver 1993; Cohen 1994).

After we had tested the hypothesis that model 1 is true against the alternate hypothesis, we would test the hypothesis that model 2 is true against the alternate. Some of the outcomes of this procedure could be: (i) both models M_1 and M_2 are rejected; (ii) model M_1 is rejected but M_2 is not; (iii) model M_1 is not rejected but M_2 is; and (iv) neither model is rejected. If outcome (ii) or (iii) occurs, then we will presumably act as if model M_1 or M_2 were true and make scientific and policy decisions on that basis, but if outcome (i) or (iv) occurs, what are we to do? Other than collecting more data, we are provided with little guidance concerning how we should now view the models and what they tell us about the world. There is also a chance that if outcome (ii) or (iii) occurs, the result is wrong, and then on what basis do we choose the p level?

Likelihood Approach (Edwards 1992). In this case, we use the data to arbitrate between the two models. That is, given the data and a mathematical description of the two models, we can ask, "How likely are the data, given the model?" Details of how to do this are given in the rest of this book, but read on pretending that you indeed know how to do it. Thus, we first construct a measure of the probability of the observed data, given that the model is true—we shall denote this by Pr{data|M_i}. We then turn this on its head and interpret it as a measure of the chance that the model is the appropriate description of the world, given the data. This is called the likelihood and we denote it by $\mathscr{L}_i\{M_i|\text{data}\}$. We now compare the likelihoods of the two models, given the data. If $\mathscr{L}_1\{M_1|\text{data}\} \gg \mathscr{L}_2\{M_2|\text{data}\}$, then we would argue that model M_1 is a better description of the world; if $\mathscr{L}_1\{M_1|\text{data}\} \ll \mathscr{L}_2\{M_2|\text{data}\}$, then we would argue that model M_2 is a better description of the world; and if $\mathscr{L}_1\{M_1|\text{data}\} \sim \mathscr{L}_2\{M_2|\text{data}\}$, then we would argue that the data do not differentiate between the models. A smart decision maker would not act as if the most likely model were true, but would weigh the costs and consequences of each action against the relative probabilities of the alternative hypotheses. But what exactly is meant by "$\gg$," "$\ll$," or "$\sim$" in this approach?

In this book, we shall work out methods for determining when one likelihood is much larger than another, and what that means in terms of confronting models with data.

Bayesian Approach. Finally, we might have other information that allows us to judge a priori which model is more likely to be true. For example, we might know the ecology of invasion and establishment of medfly in other places. Or we might know that before certain outbreaks people had been caught bringing fruit into the country from places where medfly is established. This kind of information can be summarized in a "prior probability that model M_i is true," which we denote by p_i. If we allow only two models of the world (medfly are established or they reinvade), then $p_1 + p_2 = 1$. Now, given information consisting of trap catches and the mathematical model, we want to "update" these prior probabilities. That is, we want to evaluate a "posterior probability that model M_i is true, given the data." Procedures for doing this require an understanding of conditional probability and are generally called "Bayesian methods," named after the Reverend Thomas Bayes, who introduced such ideas. In biology and mathematics, one of the earliest modern proponents was Sir Harold Jeffreys (1948), who called the method "inverse probability." His goal was to find methods that allow us to combine prior information with the chance of observing the data to evaluate a posterior probability of different hypotheses, given a scenario associated with the prior information. Interestingly, although Jeffreys is most famous for his work in applied mathematics, astronomy, and geophysics, he was one of the earliest contributors to the *Journal of Ecology* (Sheail 1989). In this book, we shall illustrate how Bayesian methods can be developed and applied. They are particularly appropriate for cases in which studies cannot be replicated (e.g., Reckhow 1990) and for assessment of the risk and safety in various environmental settings in which "expert opinion" is sought (Emlen 1989; Apostolakis 1990; Bolt 1991). There are arguments that Bayesian reasoning is the only way to provide a unified and consistent approach to deterministic and statistical theories (Howson and Urbach 1989, 1991).

This ecological scenario illustrates three approaches that can be taken when confronting models with data. It is our opinion that the process of science consists of confronting more than one description of how the world works with data. Thus, in the rest of the book we spend little time on classical methods of hypothesis testing but focus on likelihood and Bayesian methods. Two recent special features in the journal *Ecology* contain a number of papers that deal with nonclassical approaches to the use of statistics in ecological problems (Carpenter 1990; Jassby and Powell 1990; Reckhow 1990; Walters and Holling 1990; Potvin and Roff 1993; Potvin and Travis 1993; Shaw and Mitchell-Olds 1993; Trexler and Travis 1993) or with particularities of ecological situations (Dutilleul 1993; Legendre 1993). They provide a good complementary background for this book.

CHAPTER ONE

THE TOOLS FOR ECOLOGICAL DETECTION

The modern ecologist usually works in both the field and laboratory, uses statistics and computers, and often works with ecological concepts that are model based, if not model driven. How do we make the field and laboratory coherent? How do we link models and data? How do we use statistics to help experimentation? How do we integrate modeling and statistics? How do we confront multiple hypotheses with data and assign degrees of belief to different hypotheses? How do we deal with time series (in which data are linked from one measurement to the next) or put multiple sources of data into one inferential framework? These are the kinds of questions asked and answered by the ecological detective.

Like all other forms of creative activity, ecological detection is a craft that requires the right tools as well as the skills and materials to use the tools. We envision four components.

Hypotheses are the first component. Notice the plural, which is essential to our viewpoint. Science consists of confronting different descriptions of how the world works with data, using the data to arbitrate between the different descriptions, and using the "best" description to make additional predictions or decisions. These descriptions of how the world might work are hypotheses, and often they can be translated into quantitative predictions via models. In Chapter 2, we review different kinds of models, the purposes of models, and how models are related to hypotheses.

Data are the second component. You cannot do good analysis if the data are not good. But what does "good" mean? Sometimes the role of analysis is to show that a set of data—at least within the context of a particular view of the world—is not as informative or as useful as one thought it would be. In Chapter 3, we stress that it is important to "Know Your Data" and we provide a sufficient review of probability and the stochastic processes that you will need to conduct the work of the ecological detective.

Goodness of fit is the third component. When the data are used to arbitrate between different hypotheses or models, we must have a measure to determine how well each description of the world fits the observations. In Chapters 5, 7, and 9, we describe a variety of measures of goodness of fit that can be used in the confrontation of models and data. We provide recommendations about when it is good to use a particular method.

Numerical procedures are the fourth component. Having a measure of goodness of fit between the model and the data is not enough—you must to be able to evaluate it quickly and efficiently and explore the goodness of fit of other models. Thus, in Chapter 11, we provide an introduction to numerical methods needed to assess goodness of fit and to find the best fit. There is a history of the use of numerical procedures in ecology (examples from a generation ago are given by Conway et al. 1970, Melzer 1970, and Marten et al. 1975), but it is the development of microcomputers that really allows the full richness of numerical procedures to be exploited by practicing ecologists.

Overarching these components are alternative views of the scientific method and the role of models in science, which we discuss in Chapter 2. There we present four of the major philosophies of science and show how two of them are closely connected to our work of ecological detection.

A final warning. We are practicing ecologists. We are not statisticians, numerical analysts, or philosophers, and the appropriate chapters will no doubt offend the appropriate experts. For this we make no apologies other than stressing that for the ecological detective the problem is paramount. Because of that, we bring to the problem whatever techniques—from wherever they come—needed to solve it. And if the techniques do not exist, then we must invent them.

CHAPTER TWO

Alternative Views of the Scientific Method and of Modeling

Science is a process for learning about nature in which competing ideas about how the world works are measured against observations (Feynman 1965, 1985). Because our descriptions of the world are almost always incomplete and our measurements involve uncertainty and inaccuracy, we require methods for assessing the concordance of the competing ideas and the observations. These methods generally constitute the field of statistics (Stigler 1986). Our purpose in writing this book is to provide ecologists with additional tools to make this process more efficient. Most of the material provided in subsequent chapters deals with formal tools for evaluating the confrontation between ideas and data, but before we delve into the methods we step back and consider the scientific process itself. No scientist can be truly "neutral." We all operate within a fundamental philosophical worldview, and the types of statistical tools we employ and the types of experiments we do depend on that philosophy. Here we present four such philosophies.

There is a commonly accepted model for the scientific process (and from it arose a well-developed body of statistics that is taught in nearly every university in North America). The basic view can be thought of as a learning tree of critical experiments, which was described by Platt (1964) as "strong inference," and consists of the following steps:

1. Devising alternative hypotheses
2. Devising a crucial experiment (or several of them) with

alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses
3. Carrying out the experiment so as to get a clean result
4. Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain, and so on (Platt 1964, 347)

Platt likens this to climbing a tree, where each fork of the tree corresponds to an experimental outcome, and we base the direction of the climb on the outcomes so far. It is especially interesting for us as ecologists that Platt associates a "second great intellectual revolution" with the "method of multiple hypotheses," and attributes some of the most original thinking in this area to the geologist T. C. Chamberlain who published at the end of the last century. In particular, Chamberlain stressed that we are guaranteed to get into trouble when we consider only a single hypothesis rather than multiple hypotheses. This is especially interesting because the similarities between the geological and ecological sciences are in some ways much greater than the similarities between the other physical and the ecological sciences. In both ecology and geology, experiments may be difficult to perform and so we must rely on observation, inference, good thinking, and models to guide our understanding of the world. In fact, ecology may be much more of an "earth science" than a "biological science" (Roughgarden et al. 1994). We include a reprint of Chamberlain's classic paper—first published in the 1890s—as the Appendix.

ALTERNATIVE VIEWS OF THE SCIENTIFIC METHOD

Platt's view is to a very great extent the logical extension of the work of Karl Popper (1979), who revolutionized the philosophy of science in the twentieth century by arguing that hypotheses cannot be proved, but only disproved

TABLE 2.1. Four philosophies of science.

Philosopher	Key word or phrase	Type of confrontation
Popper	Falsification of hypotheses	Single hypothesis is disproved by confrontation with the data.
Kuhn	Paradigms, normal science, scientific revolution	Single hypothesis used until there is so much contradictory information that it is "overthrown" by a "better" hypothesis.
Polanyi	Republic of science	Multiple views of the world allowed according to the different opinions of scientists. Confrontation between these views and the data judged on (i) plausibility, (ii) value, (iii) interest.
Lakatos	Scientific research program	Confrontation of multiple hypotheses with data as arbitrator.

(Table 2.1). The essence of Popper's method is to challenge a hypothesis repeatedly with critical experiments. If the hypothesis stands up to repeated experiments, it is not validated, but rather acquires a degree of respect, so that in practice it is treated as if it were true. Most "modern" scientific journals adopt this approach, even though there are difficulties in using it even under the best circumstances (e.g., Lindh 1993).

Coinciding with Popper's philosophical development was the statistical work of Ronald Fisher, Karl Pearson, Jerzy Neyman, and others, who developed much of the modern statistical theory associated with "hypothesis testing" (e.g., Kendall and Stuart 1979, 175 ff.). In hypothesis testing, we focus on a single hypothesis (called the "null hypothesis") and calculate the probability that the data would have been observed if the null hypothesis were true. If this probability is small enough (usually 0.01 or 0.05), then we "reject" the null hypothesis. To complete the calculation, we must also compute the statistical power associated with the test (Peterman 1990a,b; Greenwood 1993; Thompson and Neill 1993). The power is the probability that if the null hypothesis were actually false and we were given the same data, we would reject it.

For example, we might begin with the idea that larger flocks of birds forage more effectively than smaller flocks. The null hypothesis could be that there is no relationship between flock size and foraging efficiency. A typical application of hypothesis testing would be to use linear regression to test the null hypothesis by calculating the probability that the slope of a graph of flock size versus feeding efficiency is non-zero. If the probability that the data could have arisen from the null hypothesis (slope = 0) is greater than 0.05 (or 0.01), the null hypothesis is not rejected at the "5% level" (or the 1% level). In the case considered here, if the null hypothesis could not be rejected at the 5% or 1% level and the power were sufficiently high, then the real ecological hypothesis—larger flocks forage more efficiently—would effectively be rejected.

After testing the hypothesis that larger flocks forage more efficiently, we would continue to climb Platt's decision tree to another set of experiments, depending on whether the effect of flock size on foraging efficiency was or was not statistically significant. The key elements of this view of science are (1) the confrontation between a single hypothesis and the data, (2) the central idea of the critical experiment, and (3) falsification as the only "truth." Popper supplied the philosophy and Fisher, Pearson, and colleagues supplied the