QUANTITATIVE ECOTOXICOLOGY

Second Edition

Michael C. Newman

CRC Press
Taylor & Francis Group

QUANTITATIVE ECOTOXICOLOGY

Second Edition

QUANTITATIVE ECOTOXICOLOGY

Second Edition

Michael C. Newman

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20120615

International Standard Book Number-13: 978-1-4398-3565-4 (eBook - PDF)

To Peg, Ben, and Ian

… what we hope ever to do with ease we may learn first to do with diligence.

—Samuel Johnson (from Green 1984[*])

[*] Greene, D., ed., *Samuel Johnson*, Oxford, Oxford University Press, 1984.

Contents

Preface

IMPETUS FOR THIS BOOK

> It is, therefore, urged without reason, as a discouragement to writers, that there are already books sufficient in the world; that all topics of persuasion have been discussed, and every important question clearly stated and justly decided; … [However], whatever be the present extent of human knowledge, it is not only finite, and therefore in its own nature capable of increase; but so narrow, that almost every understanding may, by a diligent application of its powers, hope to enlarge it. It is, however, not necessary, that a man forbear to write, till he has discovered some truth unknown before; he may be sufficiently useful, by only diversifying the surface of knowledge, and luring the mind by a new appearance to a second view of those beauties which it had passed over unattentively before.
>
> **—Samuel Johnson (from Murphy 1836)**

While browsing the preface to Moriarty's 1983 book, *Ecotoxicology: The Study of Pollutants in Ecosystems*, in preparation for writing the preface for the first edition of this book, I was struck by the similarity of our intentions. He suggests, and I still concur as this second edition is being prepared, that an abundance of ecotoxicological data exists, but much of it is insignificant. Considering myself an unwitting, albeit minor, contributor to this dilute database and dismayed at the prospect of mediocrity as an inescapable theme in my professional career, I dedicated considerable thought to factors contributing to this situation. Certainly, ecotoxicology is not trivial, and as so is not characterized by practitioners lacking sufficient acumen or funding. For the most part, the professionals involved in the field are well trained, well intended, and well funded. Further, the prosaic argument that ecosystems are too complex to understand to any practical extent is inconsistent with the contrastingly rapid progress made in disciplines such as molecular genetics, immunology, computer/information sciences, and physics, which deal with equally complex and less tangible subjects. Further, substantial quantitative advancement in ecology has occurred in the last 70 years, but these advances have failed *en bloc* to permeate the ecotoxicology literature. Statistics demonstrating that the majority of publications in many sciences go practically uncited (Hamilton 1990) and are seldom read by scientists outside the specific field in which they appeared provided me with a broader appreciation of the problem but no conceptual tools to improve a condition seemingly more severe in ecotoxicology than in many other fields.

I had resigned myself to the fact that all fields go through a presynthetic or descriptive phase prior to maturation. I assigned ecotoxicology (and its predecessors such as aquatic and wildlife toxicology) to this unsatisfying status until I stumbled upon an article entitled "Strong Inference" (Platt 1964). Platt's arguments regarding qualities affecting the relative rates of advancement of various disciplines suggested that solidification of ecotoxicological principles and paradigms could be greatly accelerated by adapting a stronger inferential approach. In the process, we would also become much better environmental stewards. One goal of this book is to encourage a more rigorous inferential approach. This goal is shared with the companion books, *Fundamentals of Ecotoxicology* and *Ecotoxicology: A Comprehensive Approach*, in which Platt's strong inference approach was enlarged to a strongest possible inference approach by incorporating quantitative Bayesian inference.

Each chapter of this volume treats ecotoxicology as a scientific endeavor. As is the case with all sciences, the focus is "the organization and classification of knowledge on the basis of explanatory principles" (Nagel 1961). This book emphasizes the strongest possible inferential and quantitative themes. Consequently, many aspects relevant to regulatory activities, e.g., standard methods and environmental legislation, will not be presented in balance with their importance in addressing

ecotoxicological problems facing society today. By no means should this omission be considered a mute assignment of the regulatory aspects of ecotoxicology to a status inferior to those scientific. These critical topics are covered clearly and thoroughly in other volumes, such as Rand's 1995 *Fundamentals of Aquatic Toxicology*, and numerous publications by the U.S. Environmental Protection Agency (EPA), the Organization for Economic Cooperation and Development (OECD), and the United Nations (UN).

The topics covered in this book are arranged in order of increasing biological organization and scale. This being the case, early chapters may contain information that some workers may not consider sufficiently high in biological organization to be considered pertinent to ecotoxicology, i.e., not community or higher-level topics. My objection to this artificial limitation has been expressed often elsewhere. "Few ecologists would disagree that progress in ecology would have been slowed by exclusion of all but community and system level research. It seems illogical to assume that growth in this new field of ecology would not be similarly compromised by such a restriction" (Newman and McIntosh 1991). Regardless of the level examined, the intent in all chapters is to better understand the fate and effects of toxicants in the biosphere, and to frame this understanding in quantitative terms.

ORGANIZATION OF THIS BOOK

This book explores quantitative features of the science of ecotoxicology. It is revised to more neatly complement *Fundamentals of Ecotoxicology* (Newman 2010) and *Ecotoxicology: A Comprehensive Treatment* (Newman and Clements 2008). The first is an introductory textbook that provides a general description of ecotoxicology useful in graduate or upper-level undergraduate courses. The second is intended as a textbook for a more intensive or advanced course, and also as a reference book for professionals. Each is organized into chapters that move from biological themes emerging at lower to higher levels of organization. The overarching goal is to produce three interconnected books with complementary discussions of key concepts and approaches. Topics and organization of this second edition of *Quantitative Ecotoxicology*[*] overlap coverage of the aforementioned two textbooks but from the vantage of quantitative concepts and methods. The central role played by quantification in modern science mandates this third treatment of ecotoxicology.

> The most distinct and beautiful statement of any truth must take at last the mathematical form.
>
> **—Thoreau (from Walls 1999)**

> A precise statement can be more easily refuted than a vague one, and it can be better tested. This consideration also allows us to explain the demand that qualitative statements should if possible be replaced by quantitative ones….
>
> **—Popper (1972)**

> As we enter the twenty-first century, the statistical revolution in science stands triumphant. It has vanquished determinism from all but a few obscure corners of science.
>
> **—Salsburg (2001)**

The first chapters of *Quantitative Ecotoxicology* explore fundamental concepts and definitions essential to understanding the fate and effects of toxicants at various levels of biological organization

[*] The first edition had the longer title *Quantitative Methods in Aquatic Ecotoxicology*, reflecting an initial bias toward freshwater ecotoxicology.

as covered in the remaining chapters. Scientific ecotoxicology and associated topics are defined in Chapter 1. The historical perspective, rationale, and characteristics are outlined there for the quantitative approach taken later in this book. The second chapter discusses the general measurement process. It considers methodologies for defining and controlling variance, which could otherwise exclude valid conclusions from ecotoxicological endeavors. Ecotoxicological concepts at increasing levels of ecological organization are discussed in Chapters 3 through 7. Quantitative methods used to measure toxicant accumulation and effects are outlined in each of these chapters. The importance of establishing type II, in addition to type I, error rates, as emphasized in the revised Chapter 5, necessitated more discussion of design issues, especially sample size and power estimation (Appendix 27). The final chapter summarizes the book with a brief discussion of ecotoxicology from a nonregulatory vantage.

Michael C. Newman
A. Marshall Acuff Jr. Professor of Marine Science
College of William and Mary
Virginia Institute of Marine Science
Gloucester Point, Virginia

REFERENCES

Hamilton, D.P. Publishing by—and for?—the numbers. *Science* 250:1331–1332 (1990).
Moriarty, F. *Ecotoxicology: The study of pollutants in ecosystems*. Orlando, FL: Academic Press, 1983.
Murphy, A., ed. *The works of Samuel Johnson, LL.D.* Vol. 1. New York: George Dearborn Publishers, 1836.
Nagel, E. *The structure of science. Problems in the logic of scientific explanation*. New York: Harcourt, Brace and World, 1961.
Newman, M.C. *Fundamentals of ecotoxicology.* 3rd ed. Boca Raton, FL: Taylor & Francis/CRC Press, 2010.
Newman, M.C., and W.H. Clements. *Ecotoxicology: A comprehensive treatment.* Boca Raton, FL: Taylor & Francis/CRC Press, 2008.
Newman, M.C., and A.W. McIntosh. Preface. In *Metal ecotoxicology: Concepts and applications*, ed. M.C. Newman and A.W. McIntosh. Chelsea, MI: Lewis Publishers, 1991.
Platt, J.R. Strong inference. *Science* 146:347–353 (1964).
Popper, K.R. *Objective knowledge. An evolutionary approach*. Oxford: Clarendon Press, 1972.
Rand, G.M., ed. *Fundamentals of aquatic toxicology: Effects, environmental fate, and risk assessment*. Washington, DC: Taylor & Francis, 1995.
Salsburg, D. *The lady tasting tea: How statistics revolutionized science in the twentieth century*. New York: Henry Holt and Company, 2001.
Walls, L.D. *Material faith: Thoreau on science*. Boston: Houghton Mifflin Company, 1999.

Acknowledgments

If I have seen further it is by standing on the shoulders of giants.*

The sentiment of the above quote needs modification to accurately reflect my thoughts—perhaps to "If I have seen more, it is by standing on the tips of my toes to peer over the shoulders of some very remarkable people." I claim none of the concepts and methods in this book as mine. Most came out of decades of insightful and diligent work by others who are acknowledged clearly throughout the book. The author is extremely grateful to the many scientists, statisticians, and mathematicians who fashioned the fascinating concepts and valuable tools gathered together here. The intent is not to present my concepts and tools but simply to be "sufficiently useful, by only diversifying the surface of knowledge, and luring the mind by a new appearance to a second view of those beauties which it had passed over unattentively before" (see Samuel Johnson quote opening the preface). Of course, any errors based on misreading or misperception while peering over the shoulders of my predecessors I do claim as mine alone.

Along a more specific vein, Drs. James Oris and Mark Sandheinrich provided insightful guidance at the onset of this revision. Drs. Michael Hooper, Margaret Mulvey, and Paul Story provided valuable criticism of draft chapters. Mr. Jincheng Wang reviewed meticulously several equations and computer code applied in several illustrations of model behavior.

* Sir Isaac Newton penned this now hackneyed phrase in a 1676 letter to Robert Hooke. In keeping with the theme of this acknowledgment, it is instructive to note that Newton borrowed the phrase from an 1159 treatise by John of Salisbury, who in turn had borrowed it from Bernard of Chartres (Troy, S.D., ed., *Medieval Rhetoric: A Casebook*, Routledge Medieval Casebooks, vol. 36. New York: Routledge, 2004).

About the Author

Michael C. Newman is currently the A. Marshall Acuff Jr. Professor of Marine Science at the College of William and Mary's Virginia Institute of Marine Science, where he also served as Dean of Graduate Studies for the School of Marine Sciences from 1999 to 2002. Previously, he was a faculty member at the University of Georgia's Savannah River Ecology Laboratory. His research interests include quantitative ecotoxicology, environmental statistics, risk assessment, population effects of contaminants, metal chemistry and effects, and bioaccumulation and biomagnification modeling. In addition to more than 125 articles, he has authored 5 books and edited another 6 on these topics. The English edition and Mandarin and Turkish translations of *Fundamentals of Ecotoxicology* have been adopted widely as the textbook for introductory ecotoxicology courses. He has taught at universities throughout the world, including the College of William and Mary, University of California–San Diego, University of Georgia, University of South Carolina, Jagiellonian University (Poland), University of Antwerp (Belgium), University of Hong Kong, University of Joensuu (Finland), University of Koblenz–Landau (Germany), University of Technology–Sydney (Australia), Royal Holloway University of London (UK), Central China Normal University, and Xiamen University (China). He has served numerous international, national, and regional organizations, including the OECD, U.S. EPA Science Advisory Board, Hong Kong Areas of Excellence Committee, and the U.S. National Academy of Science NRC. In 2004, the Society of Environmental Toxicology and Chemistry awarded him its Founder's Award, "the highest SETAC award, given to a person with an outstanding career who has made a clearly identifiable contribution in the environmental sciences."

CHAPTER 1

Introduction

> Science is built up of facts, as a house is built of stones; but an accumulation of facts is no more a science than a heap of stones is a house.
>
> **—Poincaré (1952)**

> We speak piously of taking measurements and making small studies that will 'add another brick to the temple of science.' Most such bricks just lie around the brickyard.
>
> **—Platt (1964)**

1.1 ECOTOXICOLOGY AS A SCIENTIFIC DISCIPLINE

> The [ecotoxicology] literature is both enormous and, in large part, trivial.
>
> **—Moriarty (1983)**

Truhaut (1977) is credited as the first to use the term *ecotoxicology* to define the "natural extension of toxicology, the science of the effects of poisons on individual organisms, to the ecological effects of pollutants" (Moriarty 1983). Cairns and Mount (1990) similarly defined *ecotoxicology* as "the study of the fate and effect of toxic agents in ecosystems," a phrasing that adds the study of pollutant fate but removes the word *science*. It is puzzling to read these definitions if one accepts, as I do, that the aspirations of environmental scientists during the last half century have always been ecotoxicological in nature. What could be so lacking in the body of knowledge that had accumulated up to 1977 as to necessitate the implied reformation around a new field of ecotoxicology?

The impetus for defining this "new" approach seems to grow out of frustration with our continued inability to predict or, in many cases, clearly document effects at any but the lowest levels of ecological organization. Effective prediction remains an elusive goal, despite decades of sincere effort with ample funding. The frustration grows acute as the need for accurate description and prediction becomes more and more pressing (Clemmitt 1992). Discomfort is invoked periodically by statements such as Lederman's (1991), then president-elect of the American Association for the Advancement of Science, that "understanding … ecological and environmental issues and providing guidance to policymakers" is one of the major tasks facing U.S. scientists today. Or Al Gore's more recent statement, "… we are colliding with the planet's ecological system, and its most vulnerable components are crumbling as a result" (Gore 2006). Predictably, the ever-present banter about the relative virtues of applied versus basic science, standard versus nonstandard methods, field versus mesocosm versus laboratory studies, and reductionist versus holistic approaches took on Babelian proportions. As the din increased, an attempt emerged in the late 1970s to regroup under the new standard, ecotoxicology.

Such reformation provides the opportunity for great advances and equally great mistakes. This is particularly true in a melding of synthetic disciplines like ecology and toxicology (Maciorowski 1988). A series of insightful and timely papers that provide much needed perspective (e.g., Cairns 1989, 1991; Cairns and Mount 1990) was published several decades ago. A contextual framework also emerged for the application of ecotoxicological methods to environmental regulation and remediation (e.g., Connell 1987; Duffus 1986; Adams 1990). What still remains distinctly absent is a focused effort to provide an effective consilient scientific framework for this field. This absence of a consistent scientific framework, in my opinion, has slowed progress during the last 30 years. Basic principles are left to be pondered as afterthoughts as legitimate and immediate needs for standardization or for information on the next of a seemingly endless number of new toxicants are satisfied. This opinion seems to have been shared by Moriarty (1983), who stated in the last paragraph of his book, "I have tried to relate the problems of ecotoxicology to their ecological context. Failure so to do has led to much muddled thinking and to unreliable conclusions." Schwetz (in Clemmitt 1992) expressed a similar opinion that toxicology is "sometimes too much of an applied science. So most ideas come from other sciences."

The goal of this second edition is to further contribute to the development of "an organization and classification of knowledge on the basis of explanatory principles" (Nagel 1961) for the science of ecotoxicology. The emphasis will be on detailing quantitative methods because they lend themselves most readily to explicit formulation of conceptual models (hypotheses), falsification, and estimation of likelihood or statistical confidence for competing explanations. However, it should not be forgotten while reading this volume that "the mathematical box is a beautiful way of wrapping up a problem, but it will not hold the phenomena unless they have been caught in a logical box to begin with" (Platt 1964). General concepts of scientific inference will be discussed to aid in avoiding such recurrent logical errors. They will be extended to quantitative expressions where warranted. Further, explicit definitions fundamental to the discipline will be formulated at the onset to avoid confusion and ambiguity.

Another definition is required here to distinguish the science of ecotoxicology from the impressive body of information fulfilling essential regulatory or monitoring needs. Ecotoxicology is the science that attempts to organize knowledge about the fate and effects (including those to humans) of toxic agents in the biosphere based on explanatory principles (Newman and Clements 2008; Newman 2010). This definition is very similar to that of Jørgensen (1990) ("the science of toxic substances in the environment and their impact on the living organisms") but emphasizes a higher ecological level focus and several important qualities of scientific knowledge. The remainder of this chapter will be used to define and clarify the basic components of this definition: biosphere and ecosystem, toxicant effect, toxicant fate, and the organization of knowledge based on explanatory principles.

1.2 TOXICANTS AND BIOSPHERE

> Any ecosystem under study has to be delimited by arbitrary decision, but one has to remember always that the imposed boundaries are open.
>
> **—Margalef (1968)**

> The relationship between structure and functioning is a fundamental one in ecosystems science. Ecosystems, and indeed the global biosphere, are prototypical examples of complex adaptive systems, in which macroscopic system properties such as trophic structure, diversity-productivity relationships, and patterns of nutrient flux emerge from interactions among components, and feed back to influence the subsequent development of these interactions.
>
> **—Levin (1998)**

The functional unit of the biosphere in classical ecology is the ecosystem.* Many ecotoxicologists discuss only the biotic community residing in a defined area when dealing with ecosystem effects. However, an ecosystem includes the biotic community and its abiotic environment functioning together as a unit to direct the flow of energy and cycling of materials. The ecosystem approach embraces the concept of components functioning to maintain the system through a complex of feedback loops. Margalef (1968) offers the context that ecosystems are systems in which "individuals or whole organisms may be considered elements of interaction, either among themselves, or with a loosely organized environmental matrix."

It is important to keep in mind that the ecosystem concept is an artificial construct used to frame concepts and hypotheses. It is not a concept without limitations (Margalef 1968; Gutmann 1976). The ecosystem model should not be applied as a perfect depiction of reality despite its enormous usefulness. How closely the qualities of an operationally defined ecosystem conform to those of the abstract one depends on many factors, including spatial and temporal scale, distinctness of system boundaries, and the specific qualities under study. For this reason, comparison of the qualities of a speculatively impacted ecosystem to those of an idealized ecosystem may be a worthwhile mental exercise, but it could not be used to conclude definitively that an adverse effect occurred. The most appropriate comparisons would be to properties of a reference ecosystem or to the same ecosystem before contamination. Obviously, only temporal comparison seems possible for the global ecosystem, that is, the biosphere.

1.3 TOXICANT EFFECTS IN ECOSYSTEMS

1.3.1 Classification Based on the Stress Concept

> Everybody knows what stress is and nobody knows what it is.
>
> **—Selye (1973)**

1.3.1.1 Stress

Effects of toxicants on ecosystem components are often measured along a spectrum ranging from the molecular (e.g., induced detoxification proteins), to the whole ecosystem level (e.g., a shift in system respiration:production or nutrient cycling), to the entire biosphere (e.g., global warming or ozone thinning). Most often, a measurement of stress is implied for any significant shift in some quality regardless of the ecological level examined. The measured quality may be a primary (e.g., modified level of a hormone directly influenced by the toxicant) or higher-order (e.g., increased intensity of parasitic infection as a consequence of the debilitating effects of a toxicant) indicator of change.

But what is stress? Stress in the common vernacular is something that causes a tension that can alter a normal state, such as a heavy load might stress one's back. Unfortunately, this general definition is often confused with the more explicit biological definition of stress to individual organisms. Hans Selye (1956, 1973) developed the concept of stress and the associated general adaptation syndrome (GAS) as applied to individuals. "Stress is the state manifested by a specific syndrome which consists of all the nonspecifically induced changes within a biological system"

* From the vantage of a physicist or chemist, the phrasing "classical ecology" must seem an oxymoron. Ecology is a new and rapidly evolving science itself. "The word ecology was coined not much more than 100 years ago, and the oldest professional society, the British Ecological Society, is less than a century old" (May and McLean 2007). For this reason, some concepts in the first edition of this book needed modification to conform to revisions in modern ecology. This was especially needed as the spatial scale of ecological issues broadened and the associated tools became more convenient.

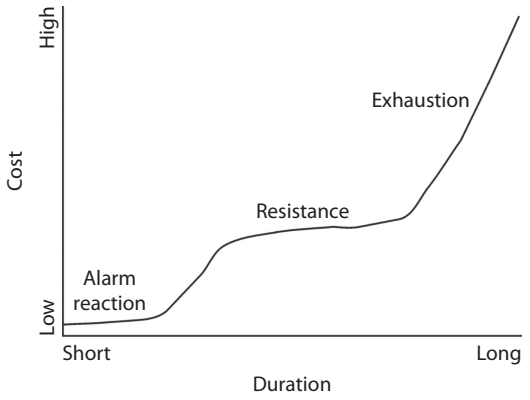

Figure 1.1 Cost associated with the three phases of Selyean stress.

(Selye 1956). He stated quite clearly that it is a *specific syndrome* in response to an external agent. It is a state achieved during any of a variety of activities, including many nondetrimental activities such as intense exertion. Further, it is a *distinctive or specific suite of responses that are beneficial or compensatory.* The stress response fulfills the purpose of turning the individual back toward the state of homeostasis. This stress response-achieved state approaching homeostasis is mediated by hypothalamic-pituitary-adrenal axis changes and was called heterostasis by Hans Selye. The associated general adaptation syndrome has three phases (Figure 1.1): the alarm reaction, adaptation (or resistance), and exhaustion components (Selye 1973). The alarm component involves immediate reactions, such as increased pulse rate and blood pressure. If the stressor continues to exert an effect on the organism, responses that stimulate tissue defense, such as enlargement of the adrenal cortex and shrinkage of the thymus, occur. After a period of exposure to the stressor, the individual enters a characteristic exhaustion phase, indicating that the organism's finite amount of adaptive energy has been exhausted. With continued stress, the individual will be unable to maintain itself and will die. The concerted expression of specific mechanisms acts to regain or resist deviation from homeostasis.

But what is stress in the context of ecotoxicology? The explicit definition originally given by Selye for individuals exposed to stressors and maintained in the medical literature has not been retained in studies of higher levels of ecological organization. The precise definition of stress depends on the level at which an effect is measured. Consequently, it is important to understand the various ecotoxicological meanings given to this concept in the literature.

Definitions vary, even at the individual level. Adams (1990) compiled the following definitions of stress, which focus on the level of the individual:

1. "The sum of all physiological responses that occur when animals attempt to establish or maintain homeostasis, the stressor being an environmental alteration and stress the organism's response."
2. "Adaptive physiological changes resulting from a variety of environmental stressors."
3. "A diversion of metabolic energy from an animal's normal activities."
4. "The sum of all the physiological responses by which an organism tries to maintain or reestablish normal metabolism in the face of chemical or physical changes."
5. "Alteration of one or more physiological variables to the point that long-term survival may be impaired."
6. "The effect of any environmental alteration that extends homeostatic or stabilizing processes beyond their normal limits."

Examining population level stress, including long-term, genetic consequences, the following definitions have been forwarded:

7. "An environmental change that results in reduction of net energy balance (i.e. growth and reproduction)…. Any reduction in production (somatic growth, reproduction or both) in response to an environmental change signifies reduced Darwinian fitness, and therefore represents a result of environmental stress" (Koehn and Bayne 1989).
8. "An environmental condition that, when first applied, reduces Darwinian fitness; for example, reduces survivorship (S) and/or fecundity (m) and/or increases time (t) between life-cycle events" (Sibly and Calow 1989).
9. "Anything which reduces growth or performance, it follows that, in a situation where a particular stress operates, there must be a reduction in fitness…. [If genotypes vary in fitness and stress is occurring consistently] evolutionary changes are to be expected" (Bradshaw and Hardwick 1989).
10. "A recent anthropogenic change in the environment affecting a population's reproductive reserve or reducing its environmentally controlled abundance limit" (Shuter 1990).

At the community or ecosystem levels, the following definitions have been advanced:

11. "A detrimental or disorganizing influence … negative responses to unusual external disturbances, or stressors of low probability to which a community of organisms is not preadapted" (Odum 1985).
12. "An external force or factor, or stimulus that causes changes in the ecosystem, or causes the ecosystem to respond, or entrains ecosystematic dysfunctions that may exhibit symptoms" (Rapport et al. 1985).
13. "A stressor is any condition or situation that causes a system to mobilize its resources and increase its energy expenditure. Stress is the response of the system to the stressor. Responses to stressors may include adaptation or functional disorder" (Lugo 1978).
14. "A perturbation (stressor) applied to a system (a) which is foreign to that system or (b) which is natural to that system but applied at an excessive amount" (Barrett et al. 1976).

Esch and Hazen (1978) provided the following definition that attempts to cover all levels of ecological organization.

15. "The effect of any force which tends to extend any homeostatic or stabilizing process beyond its normal limit, at any level of biological organization."

Hoffman and Parsons (1991), although focusing on population genetics, gave a similar definition that covers all levels of organization.

16. "The term 'stress' [is used] to represent an environmental factor causing change in a biological system which is potentially injurious."

Careful review of these definitions suggests that stress is used to identify either: (1) a response, (2) a characteristic or specific response, (3) an effect, or (4) an external factor causing a response or effect. In this book, the external factor is referred to as the stressor. The response or effect is stress. Inclusion of an effect that does not also constitute a response is contrary to a central theme of Selye's original concept. However, repeated omission of this theme in definitions necessitates the inclusion of nonresponse effects. It also necessitates establishment of classes of stress that clarify its meaning when used in ecotoxicology.

Four qualities are present in the above definitions regardless of the level of ecological organization. First, stress is a response to or effect of an external factor that is detrimental or disorganizing. Unlike the original concept of stress as advanced by Selye, stress does not include a response to or effect of a nondetrimental factor. Selye would have classified the body's response to extreme physical exertion as stress, although the exertion caused no detriment. Second, the detrimental or disorganizing factor is atypical or present at atypical levels. Implied here is the idea that the system has not adjusted itself previously to the specific stressor in such a way as to mediate its effects during predictable or highly probable exposures at a future time. Third, the system responds by or is characterized by a modification of energy flow or system structure. In the case of a response, the shift

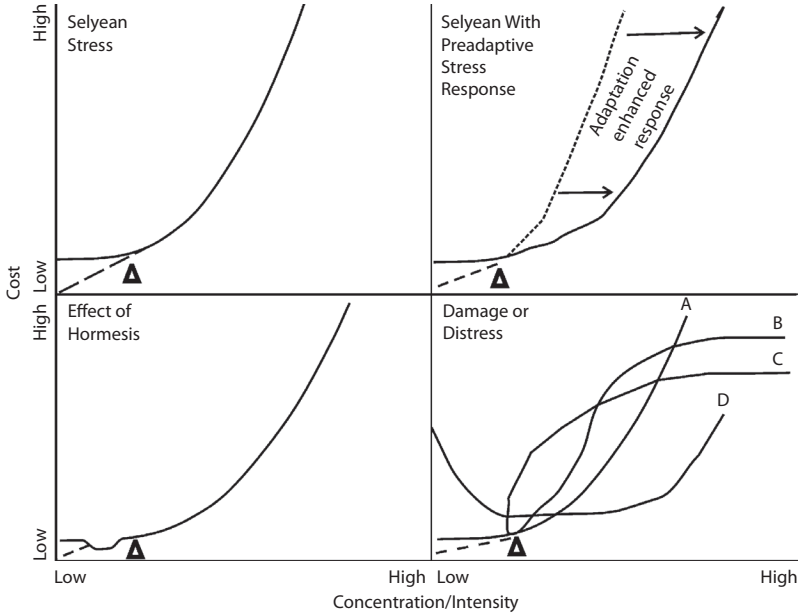

Figure 1.2 Examples of toxicant effects. The triangle denotes a possible threshold concentration. Below the threshold, the cost may be constant (solid line). Alternatively, it could increase with concentration (broken line). A simple Selyean and a Selyean effect with a superimposed preadaptation are illustrated in the top panels. An effect with hormesis is shown in the bottom left panel. Damage or distress effects can display a variety of trajectories. Examples shown here include the following: (A) zinc-induced gill damage (Hodson 1974), (B) DDT-induced fish mortality (King 1962), (C) oxygen consumption by gill tissue exposed to cadmium (Dawson et al. 1977), and (D) energy cost as influenced by salinity (Koehn and Bayne 1989). The curve shapes are products of the range of concentrations and units of effect as much as they are influenced by the type of effect.

acts to establish a condition similar to some norm or homeostasis, or to reestablish such a condition. Steady state is not an essential component of the response, although it is implied in several definitions. Fourth, temporal qualities are central to the concept of stress. Stress is a response to a recent stressor. In contrast to these four qualities common to the above definitions, the specific syndrome quality critical to Selye's stress concept has not been retained as an essential quality of stress at higher levels of organization.

With these common qualities identified, a clear, general definition of stress can be offered in the context of ecotoxicology: stress at any level of ecological organization is a response to or effect of a recent, disorganizing, or detrimental factor. A stress response represents an effort to maintain or reestablish a state approximating homeostasis, i.e., normal energy flow, material cycling, or system structure.

The following qualifiers are forwarded to differentiate between the various types of stress that can occur. The three categories of stress and three nonstress effects (Figure 1.2) defined here employ Selye's original stress concept as a yardstick for comparison. Selye's theory is a sensible and common yardstick bolstered by an enormous literature and refinement of ideas. This lends considerable justification to using Selye's concept as a touchstone.

1. Selyean stress is a specific or characteristic response to a recent, disorganizing, or detrimental factor. Its purpose is to maintain or reestablish a state similar to homeostasis (i.e., energy flow, material cycling, or system structure) within a defined norm. It is not characterized by any previous adaptation to the specific stressor. The increase in pulse rate or blood pressure associated with physical exertion is a typical example of this category of stress. If the stressor continues to elicit a response

over a longer period of time, other characteristics, such as those described above for individuals exposed to a stressor, may be expressed. Rapport et al. (1985) describe details of an ecosystem general adaption syndrome analogous to Selye's GAS. However, many examples cited suggest ecosystem effects of a type described in the next category.

2. Preadaptive stress is similar to Selyean stress, but it is characterized by a previous adaptation to the stressor, i.e., the system has specific information with which to mediate the effect of the stressor. The adaptation may be recent and transient (e.g., acclimation), or long term and relatively permanent (e.g., genetic adaptation). These responses will tend to be specific to the stressor. According to Rapport et al. (1985), this definition contains aspects of Selye's concept of eustress, a response to events that organisms or systems expect or anticipate. Induction of the cytochrome P-450 monooxygenase system by polycyclic aromatic hydrocarbons is a response to a specific class of toxicants resulting from genetic adaptation.* At the ecosystem level, preadaptive stress might be the type of response elicited by a regular or predictable stressor such as that associated with tides or seasonal fluctuations in soil moisture. To avoid confusion, it must be noted that this concept is not associated with that of genetic preadaptation.
3. Damage or distress is the adverse effect(s) of a stressor that is not a consequence of a system response. This category of effect is often defined as stress and, reluctantly, will be discussed as stress here. At lower levels of organization, the term is most commonly used to denote cell, tissue, or organ damage. Rapport et al. (1985) use the term *distress syndrome* when discussing this type of effect on ecosystems. A toxicant modifies the normal function or structure of a system without involving an active response by the system. For example, an effect might be measured if sufficient numbers of cells are damaged in a target organ. Similarly, signs of ecosystem distress would include a reduction in species diversity, increased nutrient loss, or a shift in the balance between productivity and biomass (Rapport et al. 1985; Levin 1998). In both cases, the effect is not an active response by the system to the toxicant: it is an adverse consequence of intoxication.

1.3.1.2 Nonstress Effects

What are some other effects of toxicants in ecosystems? Several types of effects are outside the concept of stress as defined above:

1. A hormetic effect (Figure 1.2) is a stimulatory effect exhibited upon exposure to low levels of some toxicants or physical agents. It is not normally characterized by a toxicant-specific system response (Stebbing 1982). Although seemingly counterintuitive, the effect of a toxic substance at a certain level of exposure can appear as beneficial. This general phenomenon is called hormesis. Southam and Ehrlich (1943) defined it as "a stimulatory effect of subinhibitory concentrations of any toxic substance on any organism." More recently, Calabrese (2008) defined it as "a biphasic dose-response phenomenon characterized by a low-dose stimulation and a high-dose inhibition." Recent treatments of this topic include effects of cadmium on growth of wood ducks (Brisbin et al. 1987) and radiation effects (Sagan 1987). Reviews of chemically induced hormesis were developed two decades ago (Calabrese et al. 1987) and more recently by Calabrese (2008).
2. A neutral effect is a measurable change that has no apparent impact (adverse or beneficial) on the system's overall qualities or probability of persistence. Although most measurable effects are likely to be positive or negative, it is illogical to reject the possibility of a neutral effect. Definition of this effect category can be particularly useful in formulation of null hypotheses for statistical or logical assessments of effect.
3. An ambiguous effect is a measured effect of undefined qualities relative to the degree of detriment/benefit, passivity, or preadaption. The present state of our understanding in ecotoxicology necessitates this category. Often effects measured at higher levels of organization fall into this category.

* In the unifying scheme of Selye, cytochrome P-450 monooxygenase induction is an example of a response controlled by catatoxic hormones, that is, by hormones designed to enhance stressor destruction. In contrast, syntoxic hormones, such as those orchestrating the Selyean stress response, facilitate the organism's ability to coexist with a stressor during exposure. Additional details can be found in Selye (1956) and Newman and Clements (2008, pp. 138–140).

Table 1.1 Categories of Ecotoxicological Effect

Category	Beneficial Response Deleterious Consequence Neutral Effect	Specific Characteristic to Response?	Response Involves Preadaptation to Stressor	Examples
Selyean stress	B	Yes	No	Increased pulse rate and blood pressure with exertion
Preadaptive stress	B	Yes	Yes	Metallothionein induction
Hormetic response	B	No	No	Stimulation of algal population growth
Damage/distress	D	NA	NA	Thinning of eggshell due to DDT exposure; decreased hunting efficiency of a predator due to intoxication
Neutral effect	N	NA	NA	Increase in toxicant concentration in a species with no adverse consequences
Ambiguous effect	?	?	?	

1.3.1.2.1 Balance between Beneficial and Adverse Effects

The discussion above has been focused on a restricted portion of the range of toxicant concentrations. For many toxicants or physical agents, the response curve can assume a shape similar to D in the bottom right-hand corner of Figure 1.2. Below certain concentrations, the effect becomes increasingly detrimental. Familiar examples include those associated with concepts such as Liebig's law of the minimum (e.g., phosphorus effects on crop yield) or Shelford's law of tolerance (e.g., salinity or temperature effects on marine species). An essential element can exert such a pattern of effect on individuals. Odum's push-pull model of stress suggests that disordering effects within a certain degree can be beneficial at the ecosystem level also. Odum's discussion of pulse stability in ecosystems (Odum 1969) presents such a beneficial effect in the context of a preadaptive stress. Lugo (1978) used Odum's push-pull or positive-negative effects model to describe numerous ecosystem level effects of stressors, including toxicant-associated effects.

1.3.1.3 Summary

The characteristics of effects based on the concept of stress are summarized in Table 1.1. Hypothetical diagrams of costs to a system with change in toxicant concentration are shown for each type in Figure 1.2. The responses are not necessarily exclusive of one another. For example, metallothionein induction may minimize cost at a low concentration of copper but, at a point of metallothionein saturation, "spillover" of significant amounts of metal to other cellular fractions occurs (see Klaverkamp et al. 1991). At that point, the Selyean stress response may become increasingly important. Prior to metallothionein induction, a hormetic response could have occurred. Damage to kidney tissues could have occurred during the exposure.

1.3.2 Classification of Effects Based on Other Criteria

Other classification systems of effects have been derived at various levels of biological organization. For example, a toxicant may be carcinogenic to an individual. At the population level, a toxicant may act to increase the risk of local population extinction. Community diversity might be decreased by a contaminant. A brief discussion of some common systems follows, and each will be discussed in detail in later chapters.

1.3.2.1 Temporal Context

Effects associated with individuals are frequently categorized in a temporal context, e.g., as acute or chronic. Often these definitions are used in discussions of toxicity testing methodologies. For example, an acute effect occurs immediately as a result of an intense exposure event. Casarett and Doull (1975) defined acute effects as "those that occur or develop rapidly after a single administration of a substance." They defined chronic effects as "those that are manifest after an elapse of time." Three decades later, Eaton and Gilbert (2008) defined acute in the latest version of the same classic toxicology textbook as "acute exposure is defined as exposure to a chemical for less than 24 hours … acute exposure usually refers to a single administration, [but] repeat exposures may be given." According to Rand and Petrocelli (1985), a chronic effect "may occur when the chemical produces a deleterious effect as a result of a single exposure, but more often they are a consequence of repeated or long-term exposures." The difference between these two types of effects is a matter of degree. To illustrate this point, Casarett explained that an acute exposure to a toxicant such as beryllium can produce an effect that will take some time to manifest itself. Finally, Suter et al. (1987) briefly discussed interpretations of the term *chronic effect* to mean those arising from exposure over "greater than 10% of the organism's lifespan." They suggested that all life stages and processes must be exposed to detect chronic effects. Although seemingly more applicable to nonhuman exposure situations than the above definitions, the proposed conventions of Suter et al. (1987) are no less arbitrary and do not clarify the intended distinction in the literature.

1.3.2.2 Lethality

Making the distinction between lethal and sublethal effects is also difficult (Moriarty 1983). Rand and Petrocelli (1985) discussed death or failure to produce viable offspring in the context of lethal effects. Sublethal effects include deleterious behavioral, anatomical, or physiological changes. Unfortunately, it is often impossible to say whether a sublethal effect (e.g., diminution of predator avoidance behavior) would or would not result in death (lethality) of an individual within a natural setting.

1.3.2.3 Site of Action

Most toxicological treatments (e.g., Casarett and Doull 1975; Rand and Petrocelli 1985) also distinguish between effects in the context of their sites of action. A systemic effect involves action on systems such as the central nervous, immune, or cardiovascular system. A local effect occurs at the primary site of damage,[*] such as a gill lesion caused by direct contact with the toxicant. Toxicants might also be classified according to their target organ. As examples, substances might be classified as hepatotoxicants, immunotoxicants, nephrotoxicants, neurotoxicants, or perhaps endocrine modifiers.

1.3.3 Summary of Toxicant Effects

As described in this section, toxicant effects, including system responses, can be discussed relative to the concept of stress. Six classes of effect (Selyean stress, preadaptive stress, hormetic response, damage/distress, neutral effect, and ambiguous effect) were described. Other classification schemes important in regulatory activities or traditional toxicology are based on time frame,

[*] Relative to the Selyean framework, stress-related changes can also be local or systemic (Selye 1956). In contrast to the GAS-related responses described above, local adaptation syndrome (LAS)-linked responses exist and include such responses as the change in immunological response in a local tissue experiencing inflammation.

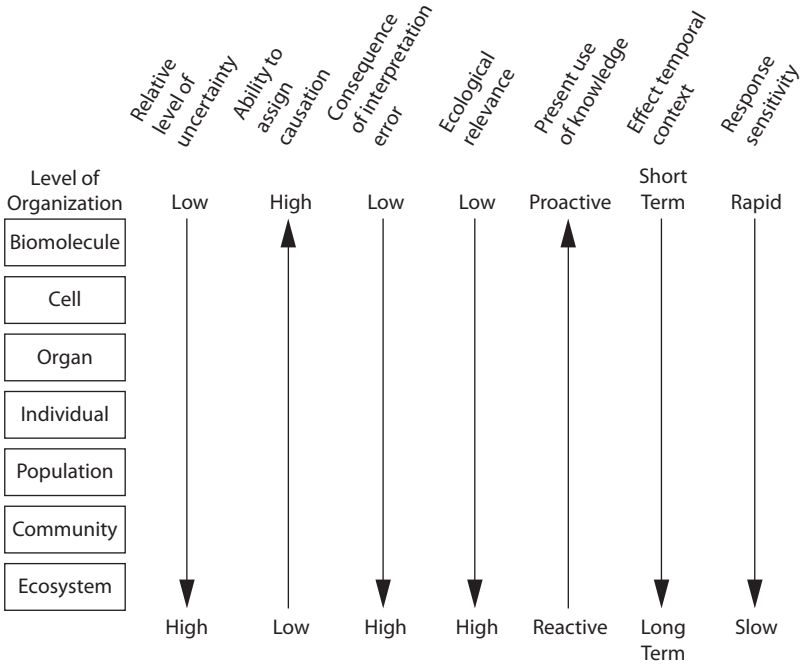

Figure 1.3 Features of ecotoxicological effects based on level of ecological organization. This figure is a composite derived from those of Haux and Forlin (1988, Figure 1), Adams et al. (1989, Figure 1), Chapman (1991, Figure 4), and Burton (1991, Figure 2).

lethality, or site of action criteria. Further, effects can be classified based on whether they exacerbate a preexisting condition, involve synergism with another agent, or represent an atypical reaction of a hypersensitive individual. Precise classifications of effects in the context of regulatory testing are detailed in sources such as Sprague (1969, 1971), Buikema et al. (1982), Suter et al. (1987), and Weber et al. (1989). Those classifications that focus on lower levels of organization will be discussed in later chapters.

What is the present state of knowledge relative to ecotoxicological effects? Figure 1.3 summarizes the present perception of our state of knowledge of effects along the ecological spectrum of organization. Our abilities to understand and assign causal relationships are best at the lower levels of organization and become poorer as the level of organization increases. The lower-level effects are generally believed to be more sensitive (i.e., manifested at lower toxicant concentrations) than effects at higher levels of organization. They often respond more rapidly to a toxicant. There is an associated advantage in that biochemical or physiological indicators can be used proactively (i.e., prior to irreversible or major ecological harm). When an ecosystem level effect is noted, it is most often used to document a degraded state. Such reactive documentation has little predictive value because the degradation has already occurred by the time an effect is seen. However, beyond documentation, this effect can be useful to establish a baseline for comparison as remedial actions are implemented. The responsiveness (rate at which the effect manifests itself after a toxic exposure occurs) and the temporal context (duration of time that the effect will be significant after removal of the toxic agent) of an effect will be shorter and more rapid at lower levels of organization. Chapman (1991) suggested that effects at lower levels tend to be more reversible than effects at higher levels of organization. At first glance, all of these qualities seem to indicate that effects at lower levels of organization are superior to higher-level effects as tools for managing ecosystem health. They respond quickly to change,

are more readily understood, are more easily assigned causation, and are more effectively used prior to permanent or significant ecological degradation. Further, "the immediate effects of pollutants are on individual organisms, by either direct toxicity or altering the environment" (Moriarty 1983).

Despite the virtues of lower-level responses, they have no overall superiority to higher-level responses. Moriarty (1983) continues the above quote by stating that although the immediate effects are at the individual level, "the ecological significance, or lack of it, resides in the indirect impact on the populations of species," that is, on the interacting populations in the ecological community. The goal of ecotoxicological stewardship is the protection of ecological systems, not biochemical moieties, physiological homeostasis, or even individuals, in most instances. The probability of falsely assigning an adverse ecotoxicological effect is increased when higher-level effects are neglected in favor of lower-level responses (Chapman 1991). Since our ability to relate lower-level responses to ecosystem degradation is limited, it follows that the ecological relevance of a lower-level response (e.g., a 50% decrease of total metal bound to metallothionein) is much more ambiguous than that of a higher-level effect (e.g., a 50% drop in species richness).

1.4 TOXICANT FATE IN ECOSYSTEMS

Rand and Petrocelli (1985) succinctly stated that toxicant fate is the "disposition of material in various environmental compartments (e.g., soil or sediment, water, air, biota) as a result of transport, transformation, and degradation." Concentration, distribution, speciation, and phase association of the toxicant in the various ecosystem components are considered, as well as toxicant sources and sinks (Figure 1.4). The toxicant will accumulate in both the biotic or abiotic components of the ecosystem under study.

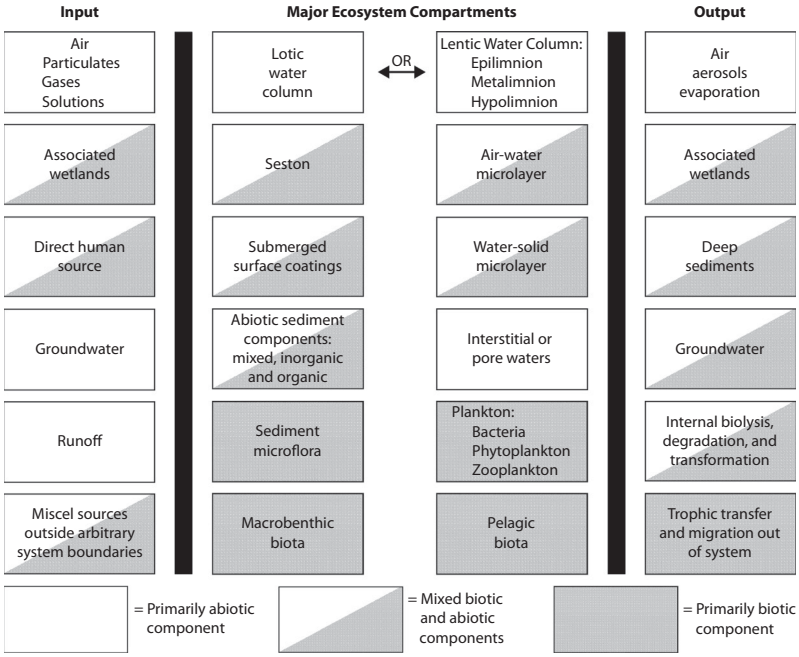

Figure 1.4 Compartments with significant ecotoxicological relevance to the fate of toxicants in aquatic ecosystems.

1.4.1 Fate in Biotic Components

Bioaccumulation is defined here as the accumulation of a toxicant in (or on, in certain situations) an individual. Results of bioaccumulation studies are very often extrapolated to make population or food web level implications. Studies may focus on taxonomic groups that accumulate relatively high concentrations of toxicants (e.g., metals in algae) or on groups more sensitive than others (e.g., DDT in nesting waterfowl). Alternatively, focus may be on assemblages associated physically with or linked functionally to components containing high concentrations of toxicant (e.g., infaunal species inhabiting contaminated sediments in the first case; scraper species grazing on metal-contaminated periphyton in the second case).

Accumulation can include direct uptake from air through lungs and associated pulmonary structures, from water through gills or epithelial tissues, and input from food via the gut. Usually, dietary sources include soil, detritus, living plant material, or animal tissue. For this discussion, the concept of bioaccumulation also explicitly includes accumulation of toxicants as a consequence of parasite-host, parent-egg/embryo/juvenile, or general symbiotic interactions. Some of these exchanges may not be direct trophic interactions as commonly envisioned. Further, the accumulation of toxicants from items processed coincidentally with food items is also included as a component of bioaccumulation. As an important example, pica (intentional or unintentional ingestion of soil) can be a substantial source for some terrestrial herbivores. Bioaccumulation will be discussed in more detail in Chapter 3.

As noted above, the transfer of toxicant from one individual to another during trophic interactions involves bioaccumulation. Biomagnification is said to occur if the toxicant concentration increases during successive trophic transfers. This trophic transfer of toxicants will be discussed in Chapter 7.

1.4.2 Fate in Abiotic Components

The intimate association of biotic and abiotic components of ecosystems is emphasized in Figure 1.4 using examples of significant inputs, outputs, and locations of toxicants in ecosystems. Indeed, any discussion of toxicant fate within ecosystems that ignores either abiotic or biotic components would be incomplete. Not only are these components in close physical contact, but they modify one another in such ways as to facilitate or inhibit toxicant exchange or transformation. For example, bioturbation may enhance a lipophilic toxicant movement between sediments and overlying waters, and chemical speciation of a metal dissolved in water will influence bioaccumulation.

Physical and chemical mechanisms influence toxicant fate in and exchange between ecosystem components. Equilibrium partitioning may determine the distribution of organic pollutants. Redox reactions and dissolution/precipitation or coprecipitation phenomena can have strong influences on metal or radionuclide movement within ecosystems. Complexation, photolysis, and adsorption/desorption also play key roles. Examples of physical processes include soil weathering and erosion, sedimentation, atmospheric fallout, bulk water movement, and bottom scouring.

1.5 ORGANIZATION OF KNOWLEDGE BASED ON EXPLANATORY PRINCIPLES

> While ecologists have sought to elucidate the nature of these "parallel effects" of [human-induced] stress on ecosystems, they have perhaps more often followed what Dyson identified as the predominant methodology of the biological sciences: a preoccupation with description of the diversity of phenomena, by and large to the exclusion of consideration of unifying themes.
>
> **—Rapport et al. (1985)**

> It has also been argued that ecology is largely ideographic (explains particulars) rather than nomothetic (derives universal laws). For a certain class of questions about ecosystems it is true that only particular explanations are needed (e.g., grass dominates this system because of frequent fires, a historical event) but the same is true in physics when known laws are applied to a given situation. This does not preclude the existence of laws.
>
> **—Loehle (1988)**

> Well, there are two kinds of biologists, those who are looking to see if there is one thing that can be understood, and those who keep saying it is very complicated and that nothing can be understood…. You must study the simplest system you think has the properties you are interested in.
>
> **—Cy Levinthal (quoted in Platt 1964)**

1.5.1 Introduction

The ecotoxicological literature is replete with statements that a working knowledge of ecosystems is impossible due to ecosystem complexity. Yet, such statements are inconsistent with the rapid growth of knowledge in equally complex disciplines. A casual glance at the intricate charts of metabolic pathways or the human genome should bring such statements immediately to question. The complex of metabolic networks functioning within each individual is staggering, but taken *en masse*, our working knowledge of such systems is impressive. A similar impression can be brought away after browsing the molecular genetics literature: entire genomes have been mapped for several species, including humans.

What then is the reason for our limited knowledge base in ecotoxicology? The belief is forwarded here that a significant impediment to the growth of ecotoxicological knowledge lies in the approach by which such knowledge is sought, not the complexity of the system under study. Although some approaches are adopted of necessity, others are selected out of habit. Some were learned from a parent discipline. The predisposition for description in ecology (see Loehle and Rapport et al. quotes above) and the propensity for single-species endpoint toxicity tests transplanted from traditional toxicology are two significant pieces of evidence supporting this argument. Other habits are more pervasive in life sciences (see Levinthal quote above). Regardless of their origins, all habits are not equally fruitful and, consequently, should be subject to review, modification, replacement, or rejection. Based on this premise, the remainder of this chapter will examine scientific habits worth considering as ecotoxicology matures as a science.

1.5.2 The Structure of Scientific Knowledge

1.5.2.1 Historical Perspective

Chamberlin (1897) suggested that the format of scientific inquiry has changed throughout history. Initially, knowledge was so limited that it appeared to be within the abilities of learned individuals to develop ruling theories that explained all phenomena. No sooner was a phenomenon presented than a ruling theory was used to explain it. This process built a large knowledge structure by repeated application alone. Chamberlin referred to this habit as precipitate explanation, the immediate and sufficient application of a theory to explain an observation. The ruling theory approach slowly was replaced by the working hypothesis approach. "Under the ruling theory, the stimulus is directed to the finding of facts for the support of the theory. Under the working hypothesis, the facts are sought for the purpose of ultimate induction and demonstration, the hypothesis being but a means for the more ready development of facts and their relations" (Chamberlin 1897). The working hypothesis approach questions theory but still retains a propensity toward precipitate

explanation. When a theory or hypothesis is put forward, it tends to be given favored status in testing. It was Chamberlin's thesis that, although its roots are no longer considered valid, the habit of precipitate explanation continues into modern scientific inquiry. He suggested the method of multiple hypotheses to minimize such bias. The method of multiple working hypotheses considers all potential hypotheses simultaneously. Equal amounts of effort are spent on the hypotheses. In this manner, the tendency to unintentionally favor one hypothesis is lessened.

Karl Popper (1965, 1968) argued that scientific inquiry should test hypotheses by a formal process of falsification. No theory can be proven true, but it can be shown to be false through observation or experimental challenge. Repeated survival of a theory or hypothesis through a rigorous falsification process confers a favored status to it. Its strength is enhanced if it continues to be "corroborated by past experience." Regardless, it is never deemed true. Acknowledging that no scientist is totally objective, Popper referred to the testability of a hypothesis as "subject to intersubjective testing." A good theory is one that can be tested by scientists, each of whom approaches it with his or her own prejudices. A process that Popper likened to natural selection occurs: the fittest theory survives.

> The empirical basis of objective science has thus nothing "absolute" about it. Science does not rest upon solid bedrock. The bold structure of its theory rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or "given" base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry structure, at least for the time being.
>
> **—Popper (1968)**

Survival of intersubjective testing alone does not constitute corroboration and consequent enhanced status for a theory. The testing must be rigorous. Some tests have higher powers to falsify based on logic alone. A classic illustration of this point involves Einstein's theory of relativity. The explanation of eccentricities in planetary orbits was less powerful in supporting Einstein's theory than the observation that light was attracted by gravity because it had more alternative explanations than the latter observation (see Popper 1965, pp. 35–36, for more details of this example). The theory of relativity was at higher risk of rejection during the second test. Rousseau (1992) suggests that avoidance of high-risk testing is one of three symptoms of pathological science, that is, science practiced without objectivity. A tradition of low-risk testing of theories in any field has an associated danger because "our habit of believing in laws is a product of frequent repetition" (Popper 1965). The ability to separate dogma from paradigm becomes impaired if low-risk testing is common in a discipline.

Regardless of the logical power of a test, a test with insufficient measurement precision or accuracy is valueless in the process of falsification (see discussion of condensation bounds in Popper 1968, pp. 123–127). It may even slow progress due to the confusing or ambiguous nature of associated conclusions. For this reason, Popper (1968) advanced the opinion of "superiority of methods that employ measurements over purely qualitative methods." This is the reason why emphasis is placed on quantitative methods in this book.

This very brief discussion of the history of scientific inquiry will end here. Certainly, cessation does not suggest that Popper provides the capstone for scientific logic. Bayesian theory extends many associated topics pertinent to our discipline. The interested reader is referred to Howson and Urbach (1989) for an excellent presentation of the Bayesian approach. A Bayesian expansion of this approach will be sketched below and revisited throughout later chapters. Regardless, this discussion has served to identify three habits to be avoided in ecotoxicology: precipitate explanation, low-risk hypothesis testing, and imprecise or inaccurate measurement.

1.5.2.2 Strong Inference

> Strong inference is just the simple and old-fashioned method of inductive inference that goes back to Francis Bacon…. The difference comes in their systematic application.
>
> **—Platt (1964)**

As mentioned in the preface, Platt (1964) observed that scientific disciplines progress at very different rates as a consequence of the general approach taken to extract and organize knowledge. Some lack a tradition of strong inference, that is, rigorous hypothesis formulation and falsification procedures. Strong inference includes the well-known steps of hypothesis formulation, execution of experiments designed to falsify, alternate hypothesis generation, and continued testing until one hypothesis remains corroborated. However, Platt suggested that the distinction between disciplines lies in the value placed on rigorous and consistent application of such techniques.

Some fields focus too much on "surveys, taxonomic, design of equipment, systematic measurements and tables, theoretical calculations—all [of which] have their proper and honored place, provided they are parts of a chain of precise induction of how nature works" (Platt 1964). He argued that such preoccupation is taught by example to students: it is not inherent to the field.

Some disciplines are characterized by a focused effort on rigorous testing of hypotheses. Others show a meandering tendency toward precipitate explanation. Experiments are inattentively designed to support favored hypotheses. Platt advocated the use of formal experimental inference methods (the scientific method), coupled with the method of multiple hypotheses, to minimize this problem.

He recommended the following practices to foster strong inference in a discipline.

1. Apply methods of inductive inference consistently and systematically.
2. Formulate hypotheses such that they are amenable to falsification. Use the "logic of exclusion" when possible.
3. State all reasonable alternate explanations of observations when presenting results.
4. Use the method of multiple hypotheses.
5. "Be explicit and formal and regular about it, to devote a half hour or an hour to analytical thinking every day, writing out a logical tree and the alternatives and crucial experiments explicitly in a permanent notebook."
6. After hearing a scientific explanation, ask two questions. Is there an experiment which could disprove it? What explanation does the present explanation exclude?

1.5.2.3 Selection of Hypotheses

Hypothesis or theory selection is a subjective process. The Bayesian treatment presented by Howson and Urbach (1989) supports this statement using the familiar "all ravens are black" argument. They explain that, if approached objectively, the possible number of theories to be tested is equal to the number of ravens in existence (n) raised to the number of possible color patterns (m) or n^m. Unless some level of subjective experience is used to select profitable hypotheses for testing, the inferential process would become impossibly cumbersome.

How then are hypotheses selected? A variety of criteria have been advanced. Popper would favor hypotheses that are easily falsifiable (practice 2 above). Loehle (1990) suggested that an optimum region (Medawar zone) exists relative to the tractability of the hypothesis and payoff for solving the problem. Loehle (1988) also recommended that ecologists should be concerned with theory reduction during the selection process. Theory reduction strives to explain one theory on the basis of another. This enhances parsimony and gives "two levels of explanation for the same phenomena." Linked also to parsimony is the application of Occam's razor to theory or hypothesis selection. Popper (1968) favors simple hypotheses as "they tell us more; because their empirical content is

greater and because they are better testable." A strong, quantitative argument favoring parsimonious hypotheses can also be developed from the Bayesian vantage (Jeffreys and Berger 1992).

Chamberlin (1897) advocated a process in which a series of plausible hypotheses are considered equally and simultaneously. He observed that the movement from ruling theory to working hypotheses still remains biased toward a central hypothesis. The multiple working hypotheses approach remains subjective but lessens the tendency for a single hypothesis to become the controlling idea. "In developing the multiple hypotheses, the effort is to bring up into view every rational explanation for the phenomenon in hand and to develop every tenable hypothesis relative to its nature, cause or origin, and give all of these as impartially as possible a working form and a due place in the investigation. The investigator thus becomes the parent of a family of hypotheses; and by his parental relations to all is morally forbidden to fasten his affections unduly upon any one." Admittedly, phrases such as "rational explanation" and "tenable hypothesis" are permeated with subjectivity. Regardless, bias is lessened in the process of falsification, and a habit of thoroughly considering all hypotheses is fostered. The multiple working hypotheses approach has one additional advantage. It lessens the tendency to stop inquiry when a single "cause" is found. It increases the probability of detecting multiple or complex causes by evenly distributing effort between a set of hypothetical causes. The approach fosters thoroughness as well as lessening bias. As powerful as the multiple hypothesis approach is, its formal application in a falsification process is logically compromised in many situations that scientists confront. However, the approach about to be described extends the approach described so far in a way that minimizes this difficulty.

1.6 BAYESIAN INFERENCE

> It is occasionally possible to encapsulate a method of science as a recipe. The most satisfying is that based on multiple competing hypotheses, also known as strong inference. It works only on relatively simple processes under restricted circumstances and particularly in physics and chemistry, where context and history are unlikely to affect the outcome.
>
> **—E.O. Wilson (1998)**

Although Wilson confuses the strong inference and multiple hypotheses concepts in the above quote, he is correct about how difficult it can be to apply Platt's approach to situations with several plausible explanations. This difficulty is periodically invoked as justification for resorting to weak inference. However, there is a solution to the difficulty that combines the strong inference approach with quantitative Bayesian methods, that is, inference favoring explanations or hypotheses that are most probable based on evidence in hand. Newman (Newman and Clements 2008) refers to this straightforward extension of Platt's strongest inference as the strongest possible inference approach. The strongest possible inference approach assumes four conditions: (1) Chamberlin's multiple working hypothesis context is the best available, (2) two or more working hypotheses have been identified to explain a set of data or evidence, (3) the most probable hypothesis based on the existing evidence is the most credible, that is, the most likely of the candidate hypotheses to be true, and (4) any favored causal hypothesis is a working hypothesis subject to reevaluation as new information emerges. Formally, that hypothesis with the highest (posterior) probability is said to be the most credible one, i.e., most worthy of belief based on the available data (Woodworth 2004). Evidence-based, conditional levels of credibility for the candidate explanations are assigned based on probabilities. If additional evidence emerges in the future, the probabilities are recalculated and levels of belief/credibility in the candidate multiple working hypotheses adjusted accordingly.

Platt's strong inference approach is an ideal special case in the strongest possible inference approach. Platt emphasized rigorous hypothesis formulation and strong testing based on dichotomous

falsification decisions, that is, a logic tree with branching associated with reject/do not reject decisions. Every effort should be expended to achieve the testing design he describes, but realistically, there are some stages in the process of inquiry and some situations in which Platt's strong inference is not possible. Also, as illustrated in Example 1.1, type I and II errors complicate application of even this simplest of approaches. The strongest possible inference approach expands the strong inference approach to include less discriminating testing situations that are inevitably part of most scientific inquiry schemes. Less discriminating testing situations are especially common at early stages of investigation in a research program. As Wilson states above, many situations do not permit a strong, discerning single test of the plausibility of two competing explanatory hypotheses. Although Platt's strong inference is the undeniable ideal to strive toward,* the broader framework provided by the strongest possible inference approach is required.

The strongest possible inference is essentially quantitative abductive inference, that is, inference that favors the most probable explanation or hypothesis.† This syllogism from Josephson and Josephson (1996) presents abductive inference in more formal terms:

D is a collection of data about a phenomenon,
H explains the data collection, *D*,
No alternate hypothesis (H_A) explains *D* as effectively as *H* does,
∴ *H* is probably true.

Bayesian equations facilitate abductive inference quantification by more explicitly defining the qualifiers, "as effectively as" and "probably true," in the above syllogism. Using the assessment of a single hypothesis (*H*) as a simple example, the Bayesian context to estimate the amount of support, credulity, belief, or plausibility warranted by data (*D*) is the following (Howson and Urbach 1989):

D provides support for *H* if $P(H|D) > P(H)$.
D draws support away from *H* if $P(H|D) < P(H)$.
D provides neither undermining nor supportive information if $P(H|D) = P(H)$

where $P(H)$ = probability of *H* being true before any consideration of the data, and $P(H|D)$ = probability of *H* being true given *D*. At this point, the task becomes estimating the probabilities in this scheme. Evidence (*D*) is combined with a prior probability of *H* being true to produce a statement of (posterior) probability given the evidence—a new probability of an explanation being true is established. If more evidence (D_{NEW}) was then collected during an inquiry, the newly established probability can be used as the new prior probability‡ and combined with (D_{NEW}) to calculate a new post probability reflecting the plausibility of *H*, given *H* and D_{NEW}. Bayes's theorem (Equation 1.1) can be used to estimate $P(H|D)$ in this case,

$$P(H \mid D) = \frac{P(H)P(D \mid H)}{P(D)} \tag{1.1}$$

where $P(D|H)$ = the (prior) probability of getting *D* if the hypothesis were true, and $P(D)$ = the probability of getting *D* whether or not *H* is true. The resulting $P(H|D)$ can become the new prior probability ($P(H_{\text{NEW}})$) with the collection of additional data, D_{NEW}.

* This is true only if false positive and negative error rates are handled properly, as is often not the case during the associated statistical testing of hypotheses (Newman 2008).
† This explanation of the strongest possible inference comes directly from Chapter 36 in Newman and Clements (2008).
‡ The probability is a "prior probability" relative to the collection of the new data.

$$P(H \mid D_{NEW}) = \frac{P(H_{NEW})P(D_{NEW} \mid H)}{P(D_{NEW})} \tag{1.2}$$

This process can be repeated with the addition of data until the associated probability is sufficient to make an evidence-based judgment about hypothesis plausibility. The $P(H|D_{NEW})$ can be recalculated as more data are collected.

This same process can also be applied to judging any hypothesis against its negation ($\sim H$), a single alternate (H_A), or several alternate hypotheses (e.g., H_{A1}, H_{A2}, H_{A3} …). Equation (1.3) illustrates how the posterior odds for H versus H_A being true ($P(H|D)/P(H_A/D)$) can be calculated from the prior odds ($P(H)/P(H_A)$) and likelihood ratio ($P(D|H)/P(D|H_A)$):

$$\frac{P(H \mid D)}{P(H_A \mid D)} = \frac{P(H)}{P(H_A)} \frac{P(D \mid H)}{P(D \mid H_A)} \tag{1.3}$$

Equation (1.4) estimates the probability of a hypothesis from its prior ($P(H)$), the prior of its negation ($P(\sim H)$), $P(D|H)$, and $P(D|\sim H)$ (e.g., the null hypothesis is true versus the null hypothesis is untrue) (Hacking 2001):

$$P(H \mid D) = \frac{P(H)P(D \mid H)}{P(H)P(D \mid H) + P(\sim H)P(D \mid \sim H)} \tag{1.4}$$

Equation (1.5) is a generalization in which the probability of the ith hypothesis or explanation (H_i) of n hypotheses/explanations is true given the information (D),

$$P(H_i \mid D) = \frac{P(D \mid H_i)P(H_i)}{\sum_{i=1}^{n} P(D \mid H_i)P(H_i)} \tag{1.5}$$

So, the relative credulities for a set of alternate explanatory hypotheses can be judged quantitatively using these evidence-based probabilities. Lane et al. (1987), Lane (1989), and Hutchinson and Lane (1989) provide convincing demonstrations in which this approach greatly improved causal assessment in medical diagnostics. Equally important, estimates of belief warranted by evidence can be recalculated as evidence accumulates through time. These calculations can easily accommodate the most discriminating accept/reject context described by Popper (1959) and advocated by Platt (1964) in his strong inference approach. They can also be applied in more complex contexts characterized by higher uncertainty, as referred to by Wilson (1998) and encountered often at different stages of ecotoxicological investigations.

Example 1.1

It ain't what you don't know that gets you in trouble. It's what you know for sure that just ain't so.

—Attributed to both Artemis Ward and Mark Twain

Rizak and Hurdey (2006) posed a hypothetical, but very realistic, question to a large number of environmental professionals in an attempt to understand how accurate one should expect his or her responses to be during an actual situation requiring a sound decision. The question was the following:

Monitoring evidence for a(n) [Australian] city has indicated that in treated drinking water, a pesticide, say "atrazine," is truly present above the recognized standard methods detection limit once every 1,000 water samples from consumers' taps. The analytical test for the pesticide has the following characteristics:

- *95% of tests will be positive for detection when the contaminant is truly present above the detection limit, and*
- *98% of the tests will be negative for detection when the contaminant is truly not present above the detection limit.*

With these characteristics, given a positive result (detection) on the analytical test for the pesticide in the drinking water system, how likely do you think this positive result is true?

Most of the 352 respondents were highly experienced with these kinds of environmental issues. They were split into two general categories, water professionals and environmental engineering/science professors, although both categories produced similar responses. Of the water professionals, 32% believed the pesticide was almost certainly (0.95–1.00 probability) present and another 36% believed the pesticide was very likely (0.80–0.95 probability) present. At the other extreme, only 7% believed it was extremely unlikely (0.00–0.05 probability) or very unlikely (0.05–0.20 probability) to be present. The overall judgment of the professionals was a 0.80 probability that the pesticide was truly present given a positive test. (This book's author obtained very similar results after conducting a survey with this question while teaching two large classes in southern China and another in Vietnam, a seminar in Virginia, and giving keynote talks in Seville, Spain, and Kerala, India.)

The information provided by Rizak and Hrudey (2006) can be incorporated into Equation (1.4) to estimate the probability for the hypothesis of the pesticide being truly present given a positive test ($^{+}Test$), that is, $P(Pesticide|^{+}Test_1)$.

$$P(Pesticide \mid {}^{+}Test_1) = \frac{P(Pesticide)P({}^{+}Test_1 \mid Pesticide)}{P(Pesticide)P({}^{+}Test_1 \mid Pesticide) + P({}^{\sim}Pesticide)P({}^{+}Test_1 \mid {}^{\sim}Pesticide)}$$

$$P(Pesticide \mid {}^{+}Test_1) = \frac{(0.001)(0.95)}{(0.001)(0.95) + (0.999)(0.02)} = \frac{0.00095}{0.00095 + 0.01998} = 0.0454$$

The probability that the pesticide is truly present given a positive test is 0.045, not 0.80. Despite the opinions of most polled professionals, the plausibility is very low for the hypothesis that the pesticide was present. The basic error made often in these kinds of judgments is paying too little attention to the base rate (1 in 1,000 in this case).

Revision of such a posterior probability ($p = 0.0454$) based on more evidence can be illustrated by extending this fictitious example. Let us assume that a positive test for a tap water sample resulted automatically in it being subjected to further testing with a more sensitive, accurate, and precise analytical method. The probability of a positive test if the pesticide is truly present is 0.99, and the probability of a negative test when the pesticide is truly not present is 0.98. Assume that a positive test results from this second analysis and a new posterior probability is estimated,

$$P(Pesticide \mid {}^{+}Test_2) = \frac{(0.0454)(0.99)}{(0.0454)(0.99) + (0.9546)(0.02)} = 0.702$$

Positive results for the first and the follow-up test result in a new posterior probability of 0.702 for the hypothesis that the pesticide is truly present. Now the plausibility of the pesticide being present in the tap water, as calculated correctly from the evidence in hand, is high enough to warrant concern and consequent action. Perhaps this moderately high plausibility is sufficient to then prompt a second sampling of the tap from which the original positive sample was taken, and this second sample also produces a positive test. The new posterior probability can be calculated as follows:

$$P(Pesticide \mid {}^{+}Repeat\ Sample / Test_1) = \frac{(0.702)(0.95)}{(0.702)(0.95) + (0.298)(0.02)} = 0.991$$

A positive test from the first sample, a positive confirmation test for that same first sample, and a positive test for a second follow-up sample result in an estimated 0.991 probability that the pesticide is present in the water from that tap. Most decision makers would assume correctly from this information that the pesticide was present in this water source and take appropriate action.

Formal application of these equations is not always needed to apply quantitative abductive inference. In fact, it is often easier to communicate insights about such situations by using diagrams of natural frequencies (Gigerenzer 2002). This approach is illustrated in Figure 1.5 for the first two steps of the above abductive inference exercise.

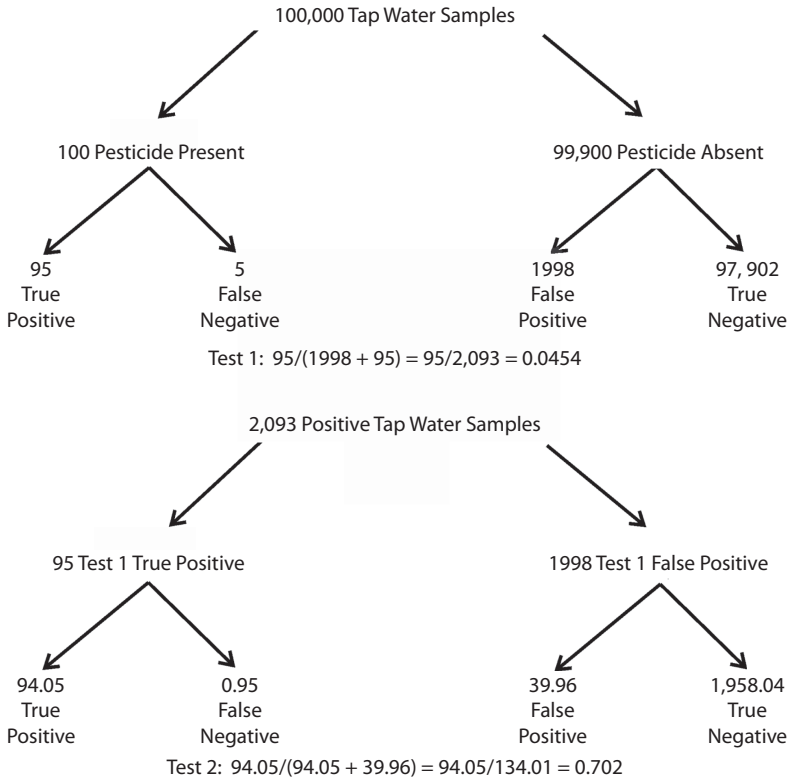

Figure 1.5 Diagrams of natural frequencies, instead of Bayesian equations, are used here to estimate the probability that a pesticide is present in a sample for which a positive test was obtained for an initial test (test 1, top of figure) and then a second, follow-up test (test 2, bottom of figure). In this illustration, 100,000 tap water samples were chosen as the initial number of samples analyzed for the pesticide. Frequencies of true positive and false positive samples were estimated using the given 1 in 1,000 prior frequency of the pesticide being present, and test 1 false negative and false positive rates of 0.05 and 0.02, respectively. Only 95 of the total number of positive tests (2,093) were true positive tests, so the posterior probability of the pesticide being present given test 1 produces a positive result is 95/2,093, or 0.0454. For 100,000 tap water samples, 2,093 will produce a positive test 1 result and be tested further with test 2. The frequency-based method of estimating the posterior probability of the pesticide being present if test 2 also shows a positive result is shown at the bottom of the figure. The posterior probability of the pesticide being present after positive results from the two tests are obtained for a tap water sample is now 94.05/134.01, or 0.702.

1.7 TOWARD STRONGEST POSSIBLE INFERENCE AND CLEAR ECOLOGICAL RELEVANCE

Emphasis is placed in this chapter on the development of ecotoxicology as a science. The discipline is characterized by a strong need for prediction at all levels of biological organization, but it is generally lacking in sufficient knowledge for making such predictions, especially at higher levels of organization. This condition is unfortunate because ecological relevance is highest for effects at the higher levels.

Increased complexity was rejected as the explanation for our lack of understanding at higher levels. Instead, the opinion is forwarded that the approaches employed in ecotoxicology, especially at higher levels of organization, do not foster rapid growth of knowledge (Figure 1.6). A strongest possible inferential approach, the best case of which is Platt's strong inference scheme, is advocated to alleviate some of this difficulty. Strongest possible inferential methods are applicable to all levels of organization, although powerful techniques with more explicit probabilities such as random assignment experiments or even quasi-experiments are logistically easier at lower levels. Criticisms based on the relative values of field versus mesocosm versus laboratory, holistic versus reductionist,

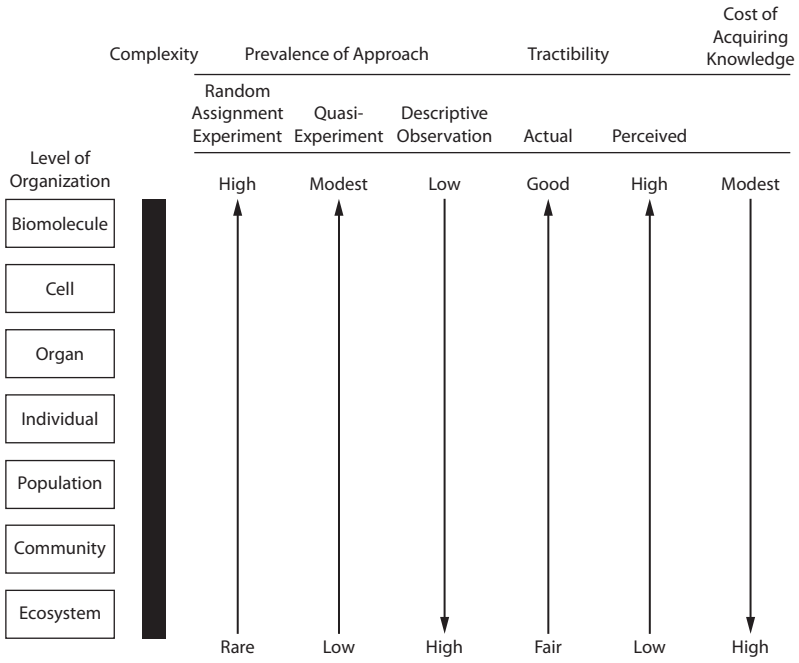

Figure 1.6 Elements of ecotoxicological knowledge and predominant approaches used to increase knowledge at each level of organization. This figure, used with Figure 1.3, qualitatively defines our present state of ecotoxicological inquiry. General approaches listed here include experiments with random assignment of experimental units to treatments, quasi-experiments (treatments; some measure of outcome and experimental units are present, but there is no random assignment of experimental units to treatments), and descriptive observation studies (see Cook and Campbell 1979 for more details). All are valuable approaches, but the values given to them relative to implying causal relationships are random assignment experiments > quasi-experiments > descriptive observation. The decrease in tractability (ability to extract knowledge) with increasing level of organization is not a consequence of system complexity. It is a consequence of the predominant approaches customarily taken and differences in cost of inquiry at the various levels of organization.

or standard versus nonstandard approaches become unfounded if strongest possible inference is an integral theme in all approaches. The decreased tractability of study at higher levels of organization or in the field must be counterbalanced by increased efforts and intensified inferential structure. Some difficulties can be offset to a degree by using the microcosm/mesocosm or large research team approaches, such as those taken by Likens and coworkers (1977). Contrary to common belief, stronger inferential techniques are most valuable at higher levels of organization where ecological relevance is high but costs can be prohibitive. An analogy would be the intense effort invested in planning crucial and decisive experiments employing limited beam time on particle accelerators.

Strongest possible inference is advocated as the most effective means of obtaining useful knowledge in the emerging and socially obligated field of ecotoxicology. It is characterized by the systematic and consistent application of standard methods of inference, especially Bayesian inference techniques. It includes the concept of multiple hypotheses as implemented by abductive methods as a means of minimizing subjectivity in selection and comparison of hypotheses. Methods for selecting hypotheses should be influenced by amenability to falsification/abductive computations and quantitative formulation, and parsimony, including theory reduction. To be most effective, the measurement process must be demonstrably precise and unbiased. Hypotheses should be readily amenable to abductive analysis such that statistical rejection criteria or evidence-based plausibilities are defined in terms of probability.

Based on the materials discussed above, 12 habits are suggested to enhance the rate at which ecotoxicological knowledge is acquired.

1. Be aware of and avoid the habits of precipitate explanation and informal abduction.
2. While recognizing the value of such approaches, avoid an excessive preoccupation with "surveys, taxonomic, design of equipment, systematic measurements and tables, [and] theoretical calculation."
3. Systematically and consistently apply inductive/abductive inference techniques.
4. Favor experiments or observations with a high risk of logical falsification or producing evidence with the most influence on estimated posterior probabilities.
5. Favor quantitative methods under rigorous precision/accuracy control.
6. Preferentially formulate hypotheses amenable to falsification or the strongest testing possible.
7. Apply the principle of multiple hypotheses.
8. Favor hypotheses that are easily falsifiable or subject to the strongest possible testing.
9. Favor hypotheses that are tractable and have good probability of solving the problem.
10. Favor parsimonious hypotheses.
11. Favor hypotheses that enhance theory reduction.
12. Recognize negative results as a critically important component of the process, not a consequence of the worker's failure to pick "the right question."

REFERENCES

Adams, S.M., ed. Biological indicators of stress in fish. *Am. Fisheries Soc. Symp.* 8:1–8 (1990).

Adams, S.M., K.L. Shepard, M.S. Greeley Jr., B.D. Jimenez, M.G. Ryon, L.R. Shugart, and J.F. McCarthy. The use of bioindicators for assessing the effects of pollutant stress on fish. *Mar. Environ. Res.* 28:459–464 (1989).

Barrett, G.W., G.M. Van Dyne, and E.P. Odum. Stress ecology. *Bioscience* 26(3):192–194 (1976).

Bradshaw, A.D., and K. Hardwick. Evolution and stress—Genotypic and phenotypic component. *Biol. J. Linn. Soc.* 37:137–155 (1989).

Brisbin, I.L., Jr., K.W. McLeod, and G.C. White. Sigmoid growth and the assessment of hormesis: A case for caution. *Health Phys.* 52(5):553–559 (1987).

Buikema, A.L., Jr., B.R. Niederlehner, and J. Cairns Jr. Biological monitoring part IV—Toxicity testing. *Water Res.* 16:239–262 (1982).

Burton, G.A., Jr. Assessing the toxicity of freshwater sediments. *Environ. Toxicol. Chem.* 10:1585–1627 (1991).

Cairns, J., Jr. Will the real ecotoxicologist please stand up? *Environ. Toxicol. Chem.* 8:843–844 (1989).

Cairns, J., Jr. Restoration ecology: A major opportunity for ecotoxicologists. *Environ. Toxicol. Chem.* 10:429–432 (1991).

Cairns, J., Jr., and D.I. Mount. Aquatic toxicology. Part 2 of a four-part series. *Environ. Sci. Technol.* 24(2):154–161 (1990).

Calabrese, E.J. Hormesis. Why is it important to toxicology and toxicologists? *Environ. Toxicol. Chem.* 27:1451–1474 (2008).

Calabrese, E.J., M.E. McCarthy, and E. Kenyon. The occurrence of chemically induced hormesis. *Health Phys.* 52(5):531–541 (1987).

Casarett, L.J., and J. Doull. *Toxicology: The basic science of poisons*. New York: Macmillan Publishing Co., 1975.

Chamberlin, T.C. The method of multiple working hypotheses. *J. Geol.* 5:837–848 (1897).

Chapman, P.M. Environmental quality criteria. What type should we be developing? *Environ. Sci. Technol.* 25(8):1353–1359 (1991).

Clemmitt, M. Public, private health concerns spur rapid progress in toxicology. *Scientist* 6(4):1–7 (1992).

Connell, D.W. Ecotoxicology—A framework for investigations of hazardous chemicals in the environment. *Ambio* 16(1):47–50 (1987).

Cook, T.D., and D.T. Campbell. *Quasi-experimentation design and analysis issues for field settings*. Boston: Houghton Mifflin Company, 1979.

Dawson, M.A., E. Gould, F.P. Thurberg, and A. Calabrese. Physiological response of juvenile striped bass, *Morone saxatilis*, to low levels of cadmium and mercury. *Chesapeake Sci.* 18(4):353–359 (1977).

Duffus, J.H. 1986. Interpretation of ecotoxicity. In *Toxic hazard assessment of chemicals*, ed. M. Richardson. London: Royal Society of Chemistry, Burlington House, pp. 90–116.

Eaton, D.L., and S.G. Gilbert. 2008. Principles of toxicology. In *Casarett and Doull's toxicology. The basic science of poisons*, ed. C.D. Klaassen. 7th ed. New York: McGraw-Hill Co., chap. 2.

Esch, G.W., and T.C. Hazen. Thermal ecology and stress: A case history for red-sore disease in largemouth bass. In *Energy and environmental stress in aquatic systems,* ed. J.H. Thorp and J.W. Gibbons. CONF-771114. Springfield, VA: National Technical Information Service, 1978, pp. 331–363.

Gigerenzer, G. *Calculated risks. How to know when numbers deceive you*. New York: Simon & Schuster, 2002.

Gore, A. *An inconvenient truth. The planetary emergency of global warming and what we can do about it*. New York: Rodale, 2006.

Guttman, B.S. Is "levels of organization" a useful biological concept? *Bioscience* 26(2):112–113 (1976).

Hacking, I. *An introduction to probability and inductive logic.* Cambridge: Cambridge University Press, 2001.

Haux, C., and L. Förlin. Biochemical methods for detecting effects of contaminants on fish. *Ambio* 17(6):376–380 (1988).

Hodson, P.V. *The effect of temperature on the toxicity of zinc to fish of the genus Salmo*. PhD thesis, University of Guelph, Ontario, 1974.

Hoffman, A.A., and P.A. Parsons. *Evolutionary genetics and environmental stress*. Oxford: Oxford University Press, 1991.

Howson, C., and P. Urbach. *Scientific reasoning. The Bayesian approach*. La Salle, IL: Open Court Publishing Company, 1989.

Hutchinson, T.A., and D.A. Lane. Assessing methods for causality assessment of suspected adverse drug reactions. *J. Clin. Epidemiol.* 42:5–16 (1989).

Jeffreys, W.H., and J.O. Berger. Ockham's razor and Bayesian analysis. *Am. Sci.* 80:64–72 (1992).

Jørgensen, S.E. 1990. *Modelling in ecotoxicology*. New York: Elsevier.

Josephson, J.R., and S.G. Josephson. *Abductive Inference. Computation, philosophy, technology*. Cambridge: Cambridge University Press, 1996.

King, S.F. *Some effects of DDT on the guppy and the brown trout*. U.S. Fish and Wildlife Service Scientific Report 399. Washington, DC: U.S. Fish and Wildlife Service, 1962.

Klaverkamp, J.F., M.D. Dutton, H.S. Majewski, R.V. Hunt, and L.J. Wesson. Evaluating the effectiveness of metal pollution controls in a smelter by using metallothionein and other biochemical responses in fish. In *Metal ecotoxicology, concepts and applications*, ed. M.C. Newman and A.W. McIntosh. Chelsea, MI: Lewis Publishers, 1991, pp. 33–64.

Koehn, R.K., and B.L. Bayne. Towards a physiological and genetical understanding of the energetics of the stress response. *Biol. J. Linn. Soc.* 37:157–171 (1989).

Lane, D.A. Subjective probability and causality assessment. *Appl. Stoch. Model Data Anal.* 5:53–76 (1989).

Lane, D.A., M.S. Kramer, T.A. Hutchinson, J.K. Jones, and C. Naranjo. The causality assessment of adverse drug reactions using a Bayesian approach. *Pharmaceut. Med.* 2:265–283 (1987).

Lederman, L.M. *Science: The end of the frontier?* A supplement to *Science*. Washington, DC: AAAS, January 1991, p. 20.

Levin, S.A. Ecosystems and the biosphere as complex adaptive systems. *Ecosystems* 1:431–436 (1998).

Likens, G.E., F.H. Bormann, R.S. Pierce, J.S. Eaton, and N.M. Johnson. *Biogeochemistry of a forested ecosystem*. New York: Springer-Verlag, 1977.

Loehle, C. Philosophical tools: Potential contributions to ecology. *Oikos* 51:97–104 (1988).

Loehle, C. A guide to increased creativity in research—Inspiration or perspiration? *Bioscience* 40:123–129 (1990).

Lugo, A.E. Stress and ecosystems. In *Energy and environmental stress in aquatic systems*, ed. J.H. Thorp and J.W. Gibbons. Springfield, VA: National Technical Information Service CONF-771114, 1978, pp. 62–101.

Maciorowski, A.F. Populations and communities: Linking toxicology and ecology in a new synthesis. *Environ. Toxicol. Chem.* 7:677–678 (1988).

Margalef, R. *Perspectives in ecological theory*. Chicago: University of Chicago Press, 1968, p. 13.

May, R.M., and A.R. McLean. 2007. *Theoretical ecology. Principles and applications*. Oxford: Oxford University Press.

Moriarty, F. *Ecotoxicology: The study of pollutants in ecosystems*. Orlando, FL: Academic Press, 1983.

Nagel, E. *The structure of science. Problems in the logic of scientific explanation*. New York: Harcourt, Brace and World, 1961.

Newman, M.C. "What exactly are you inferring?" A closer look at hypothesis testing. *Environ. Toxicol. Chem.* 27:1013–1019 (2008).

Newman, M.C. *Fundamentals of ecotoxicology.* 3rd ed. Boca Raton, FL: CRC Press, 2010.

Newman, M.C., and W.H. Clements. *Ecotoxicology: A comprehensive treatment.* Boca Raton, FL: CRC Press, 2008.

Odum, E.P. The strategy of ecosystem development. *Science* 162:262–270 (1969).

Odum, E.P. Trends expected in stressed ecosystems. *BioScience* 35:419–422 (1985).

Platt, J.R. Strong inference. *Science* 146:347–353 (1964).

Poincaré, H. *Science and hypothesis*. New York: Dover Publishing, 1952. English translation of 1905 *La Science Et l'Hypothèse* by G.B. Halsted.

Popper, K.R. *The logic of scientific discovery*. London: Routledge, 1959.

Popper, K.R. *Conjectures and refutations. The growth of scientific knowledge*. New York: Harper and Row, 1965.

Popper, K.R. *The logic of scientific discovery*. London: Hutchinson and Company, 1968.

Rand, G.M., and S.R. Petrocelli. *Fundamentals of aquatic toxicology*. New York: Hemisphere Publishing Corporation, 1985.

Rapport, D.J., H.A. Regier, and T.C. Hutchinson. Ecosystem behavior under stress. *Am. Nat.* 125(5):617–640 (1985).

Rizak, S.N., and S.E. Hrudey. Misinterpretation of drinking water quality monitoring data with implications for risk management. *Environ. Sci. Technol.* 49(17):5244–5250 (2006).

Rousseau, D.L. Case studies in pathological science. *Am. Sci.* 80:54–63 (1992).

Sagan, L.A. What is hormesis and why haven't we heard about it before? *Health Phys.* 52(5):521–525 (1987).

Selye, H. *The stress of life*. New York: McGraw-Hill Book Company, 1956.

Selye, H. The evolution of the stress concept. *Am. Sci.* 61:692–699 (1973).

Sibly, R.M., and P. Calow. A life-cycle theory of responses to stress. *Biol. J. Linn. Soc.* 37:101–116 (1989).

Shuter, B.J. Population-level indicators of stress. *Am. Fisheries Soc. Symp.* 8:145–166 (1990).

Southam, C.M., and J. Ehrlich. Effects of extract of western red cedar heartwood on certain wood-decaying fungi in culture. *Phytopathology* 33:517–524 (1943).

Sprague, J.B. Measurement of pollutant toxicity to fish. I. Bioassay methods for acute toxicity. *Water Res.* 3:793–821 (1969).

Sprague, J.B. Measurement of pollutant toxicity to fish. III. Sublethal effects and "safe" concentrations. *Water Res.* 5:245–266 (1971).

Stebbing, A.R.D. Hormesis—The stimulation of growth by low levels of inhibitors. *Sci. Total Environ.* 22:213–234 (1982).

Suter, G.W., A.E. Rosen, E. Linder, and D.F. Parkhurst. Endpoints for responses of fish to chronic toxic exposures. *Environ. Toxicol. Chem.* 6:793–809 (1987).

Truhaut, R. Ecotoxicology: Objectives, principles and perspectives. *Ecotoxicol. Environ. Saf.* 1:151–173 (1977).

Weber, C.I., W.H. Peltier, T.J. Norberg-King, W.B. Horning II, F.A. Kessler, J.R. Menkedick, T.W. Neiheisel, P.A. Lewis, D.J. Klemm, Q.H. Pickering, E.L. Robinson, J.M. Lazorchak, L.J. Wymer, and R.W. Freyberg. *Short-term methods for estimating the chronic toxicity of effluents and receiving waters to freshwater organisms.* 2nd ed., EPA/600/4-89/001. Cincinnati, OH: EMSL U.S. Environmental Protection Agency, 1989.

Wilson, E.O. *Consilience. The unity of knowledge*. New York: Random House, 1998.

Woodworth, G.G. *Biostatistics. A Bayesian introduction*. Hoboken, NJ: John Wiley & Sons, 2004.

CHAPTER 2

The Measurement Process

Boswell: "Sir Alexander Dick tells me, that he remembers having a thousand people a year to dine at his house; that is reckoning each person as one, each time he dines there." Johnson: "That is about three a day." Boswell: "How your statement lessens the idea." Johnson: "That, Sir, is the good of counting. It brings everything to a certainty, which before floated in the mind indefinitely." Boswell: "But Omne ignotum pro magnifico est:* one is sorry to have this diminished." Johnson: "Sir, you should not allow yourself to be delighted with error." Boswell and Glover

—Samuel Johnson (1750)

2.1 GENERAL

2.1.1 Overview

Rousseau (1992) defines three symptoms of pathological science (the excessive loss of scientific objectivity). The first is an aversion to crucial experiments that could disprove a favored theory. The second is a disregard for prevailing ideas and theories. Traditional theories are given inadequate consideration as the researcher becomes more and more enamored with a new discovery. These first two symptoms should seem familiar to the reader as they were discussed in Chapter 1 as low-risk testing and precipitate explanation. The third symptom was mentioned only very briefly in that chapter. Rousseau suggests that the last symptom of pathological science often begins with an effect that is "at the limits of detectability or has very low statistical significance…. Once the investigator [is] convinced that something new and important has been discovered, the fact that all of the parameters involved … are not under control is viewed as having little consequence" (Rousseau 1992). The improperly controlled or poorly understood measurement process acts as the seed from which increasingly biased behavior grows. Fortunately, the means are available for minimizing such misinterpretations of the results of most measurement processes. Several methods for controlling measurement difficulties in or defining the limitations of the measurement process are outlined here at the beginning of this book because, in their absence, later methods would be useless.

Although sometimes believed to be pertinent primarily to chemical analyses, the techniques described in this chapter for assessing such qualities as accuracy and precision are amenable to and necessary for any measurement process. For example, if a plankton net is towed, the measured numbers of individuals of each species from that tow have associated limits of detection. (Indeed, this point will be discussed again in Chapter 7 regarding species abundance.) Further, questions regarding precision and accuracy of measurements must be answered prior to any meaningful data analysis. In this example, precision may be quantified with replicate tows of the same net or one tow of

* "The unknown always passes for the marvelous."

two identical nets coupled by a common yoke. Precision of the enumeration process might involve having several skilled individuals count the same plankton sample in the laboratory. Accuracy during enumeration could be estimated with species "spikes" to a portion or aliquot of a sample.

2.1.2 The Necessity of Controlled Measurement

It is difficult to find a sound reason why this crucial aspect of ecotoxicology was given such low priority for many years. Although formal methods for implementing quality control have been available since the 1920s (Grant 1964; Shewart 1986), quality control was followed informally with varying degrees of commitment until environmental regulations mandated otherwise. Inexplicably, it remains underemphasized in college course work outside of statistics, applied chemistry, and engineering. Indeed, programs implemented to ensure controlled measurement can still elicit extreme responses. "Some of the scientists seem to consider quality assurance an insult to their professional integrity, an obstruction to 'real work' and an inference that scientists will cheat or falsify data or results" (Zimmerman 1990). Such an immoderate attitude may be a remnant of concepts abandoned early in the development of science (see Chapter 1 or Chamberlin 1897). Regardless, intersubjective testing as practiced today requires clear documentation of measurement conditions (Ayala et al. 1989). Failure to do so inhibits progress and improvement of skills (Taylor 1987; Rayl 1991), decreases the effectiveness of the decision-making process (Keith et al. 1983; Palca 1991), fosters self-delusion (Rousseau 1992), and inhibits our ability to detect the infrequent occasion of fraud (Koshland 1987; Culliton 1988).

2.2 REGIONS OF QUANTITATION

2.2.1 Overview

Keith et al. (1983) and Taylor (1987) provide lucid explanations of quantitation regions; consequently, their presentations are condensed into this overview with only minor modification. The interested reader is urged to examine the original materials for enriching details.

The certainty of a measured value can be gauged relative to the standard deviation (s_o) for samples with concentrations near 0, i.e., signals near the baseline noise of the measurement process. The relative uncertainty is most commonly used for this purpose (Taylor 1987).

$$\text{Relative Uncertainty } (\%) = 100\,\frac{z\sqrt{2}}{N} \tag{2.1}$$

where z = z statistic at a confidence level of $100(1 - \alpha)\%$, and N = measured value of an observation expressed as a multiple of s_o.

For example, the relative uncertainty of a value three times larger than s_o at a 95% confidence level is $100[(1.960\sqrt{2})/3]$, or ±92%. Such an observation has a measurement uncertainty nearly as large as its mean value. A more acceptable measurement uncertainty might be that associated with a $10s_o$ value, i.e., $100[(1.960\sqrt{2})/10]$, or ±28%.

The information needed to calculate the relative uncertainty (mean value and s_o) can be generated in several ways. The sample signal (S_t) may be used to estimate the mean value if there is no significant blank signal (S_b). If the blank signal is measurable then the difference between the sample and blank signals ($S_t - S_b$) may be used. Next, one of two methods can be used to estimate the standard deviation of the measurement process (s_o). As recommended by Taylor (1987), s_o can be

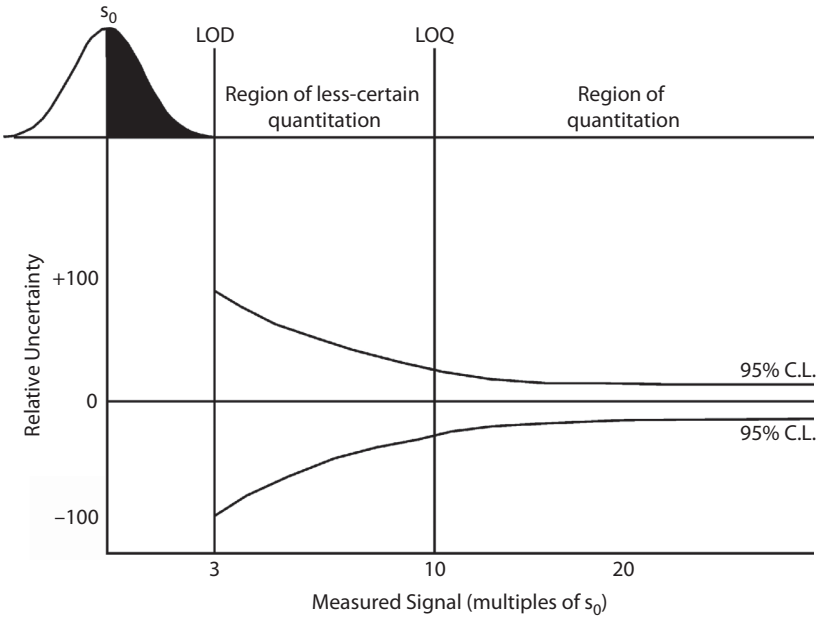

Figure 2.1 Regions of quantitation. (Modified from Keith, L.H. et al., *Anal. Chem.,* 55, 2210–2218, 1983, and Taylor, J.K., *Quality Assurance of Chemical Measurements*, Lewis Publishers, Chelsea, MI, 1987.)

estimated by plotting the standard deviation of signal versus concentration for a series of replicate standards (or samples), including one set close to a concentration of 0. The Y-intercept of the plot is then used to estimate s_o. A minimum of three concentration levels with a total of at least seven measured signals is recommended for estimating s_o. Although this linear extrapolation method is preferred by Taylor (1987), s_o can also be estimated with signals from replicates at one concentration near the limit of quantitation. Taylor (1987) recommends using a concentration close to approximately $20s_o$ in this single concentration approach, although no explanation is given for choosing this concentration. The author uses replicates of samples at or near that of the background.

Using the information generated above, each measurement can now be assigned to a region of quantitation (Figure 2.1). Again, the measured values are expressed as multiples of s_o. At $3s_o$ above the baseline signal (the limit of detection (LOD)), the measured value is estimated to be within nearly ±100% of the true value with a 95% confidence level. Values below the LOD are often reported as below the detection limit (<DL or <LOD) or not detected (ND). Above the limit of detection but below $10s_o$ is the region of less certain quantitation (also called the region of qualitative analysis by Currie (1968)). In this region, a measurement is detectable but associated with a sufficiently large uncertainty (greater than approximately ±30%) to render it semiquantitative. Above $10s_o$ (the limit of quantitation (LOQ)) but below the point at which linearity of the standard curve ends (limit of linearity (LOL)) is the region of quantitation.

A major goal in design of quantitative samplings and procedures should be to generate a data set with all observation values within the region of quantitation. Indeed, recommendations have been forwarded that only data in the region of quantitation should be used to make quantitative decisions (Keith et al. 1983). Unfortunately, constraints such as available instrumentation, regulation-mandated methodologies, temporal variation in concentrations, and incorporation of control treatments or uncontaminated sites often produce data sets with observations in all or several measurement regions.

2.2.2 Data Sets with Below Detection Limit Observations

2.2.2.1 Definitions

Extending the above discussion, a straightforward definition of the limit of detection can be stated. The limit of detection is defined as the signal level $3s_o$ above the baseline signal expressed in the units of interest, e.g., concentration units. Below the limit of detection, the measurement uncertainty is approximately equal to or greater than the value itself. Note that specific applications may necessitate modification of this general definition, e.g., radiological methods assuming a Poisson distribution for signal variability. Estimation for nonlinear calibration methods (Schwartz 1983), comparison of instruments (Arellano et al. 1985), estimation using chromatographic techniques (Synovec and Yeung 1985), measurement of radioactivity (Currie 1968; Donn and Wolke 1977), and measurement of mixture analytes such as Aroclor 1242 (Alford-Stevens 1987) or dioxins (Helsel 2010) are a few important examples that might require refinement of the above definition.

Further, several categories of limits of detection can be defined depending on the intended use of the resulting statistic. An instrument detection limit (IDL) measures the signal-to-noise ratio associated with the measurement equipment. The method detection limit (MDL) includes variation in all measurement steps leading to estimation of s_o. For example, the mass of sample used, extraction consistency, and other procedural steps contributing to signal variation may be incorporated during estimation of a MDL.

2.2.2.2 Reporting

The limit of reporting (LR) is "a limit above which data values are reported without qualification by an analytical laboratory" (Helsel 2005). The LR may be the LOD or even the LOQ (Keith et al. 1983; Helsel and Gilliom 1986). Most often, data sets are reported with all observations assigned numerical values except those below the LOD. Observations with values less than the LOD are noted as below the LOD (BDL, <DL, <LR, or ND). The laudable intent of this practice is to report only values that are above the noise of the measurement process by a statistically defined amount. The resulting data set that contains a subset of observations with no assigned numerical value below a certain point is said to be censored. The data can be further defined as left censored because censoring at the LOD involves low values from the left portion of the sample distribution. Such an approach can cause interpretation problems later because it gives a false impression to unwary data end users that the quality of the information associated with reported values from the region of less certain quantitation is equivalent to that of values from above the limit of quantitation.

Left censoring of data sets creates many problems. Censoring precludes use of valuable information by the decision maker. Fortunately, arguments against universal censoring are increasing in frequency (Gilbert and Kinnison 1981; Gilliom et al. 1984; Gilbert 1987; Porter et al. 1988; Newman et al. 1989). According to these authors, it could be more effective in some situations to report values for all observations along with the associated measurement uncertainty. For example, the relative uncertainty could be estimated using Equation (2.1) if s_o was reported. Then the end user could examine and manipulate the data set as appropriate for his or her particular need. For example, a rank order test may extract valuable information more effectively from a data set with a moderate proportion of observations below the LOD or LOQ. Such an approach would require considerable—perhaps unrealistic—levels of diligence during information extraction from data sets in order to avoid equally serious compromises as those imposed by the current practice of censoring.

2.2.2.3 Estimating Mean and Standard Deviation for Censored Data

2.2.2.3.1 Deletion and Substitution Methods

Regardless of arguments against censoring, left censoring of data sets remains a widespread practice. Commonly used, informal methods of coping with censored data sets are deletion or substitution techniques (Newman et al. 1989). In the deletion procedure, values below the reporting limit are not used in estimation of mean and standard deviation. In the substitution procedure, some value such as 0, ½LOD, or the LOD is substituted for all values below the reporting limit when the mean and standard deviation are to be calculated. Although substitution of ½LOD for a single or few LOD observations in a large data set may not cause major problems, the general application of these statistically unsubstantiated techniques biases estimates of mean and standard deviation. For example, deletion and substitution methods will produce biased estimates of the mean. In Figure 2.2A, two distributions with identical means but dissimilar standard deviations are used to illustrate this bias. The estimate of the mean for the distribution with the wider standard deviation will be biased upward more than that of the distribution with the narrower standard deviation if observations below the LOD are deleted, or if the LODs are substituted for <DL observations. If 0 or ½LOD is substituted for the censored values, the estimated mean for the broader distribution will be smaller than that of the narrow distribution.

Similarly, standard deviations estimated with deletion or substitution methods are biased (Figure 2.2B). For example, deletion of or substitution of the LOD for censored values will bias the

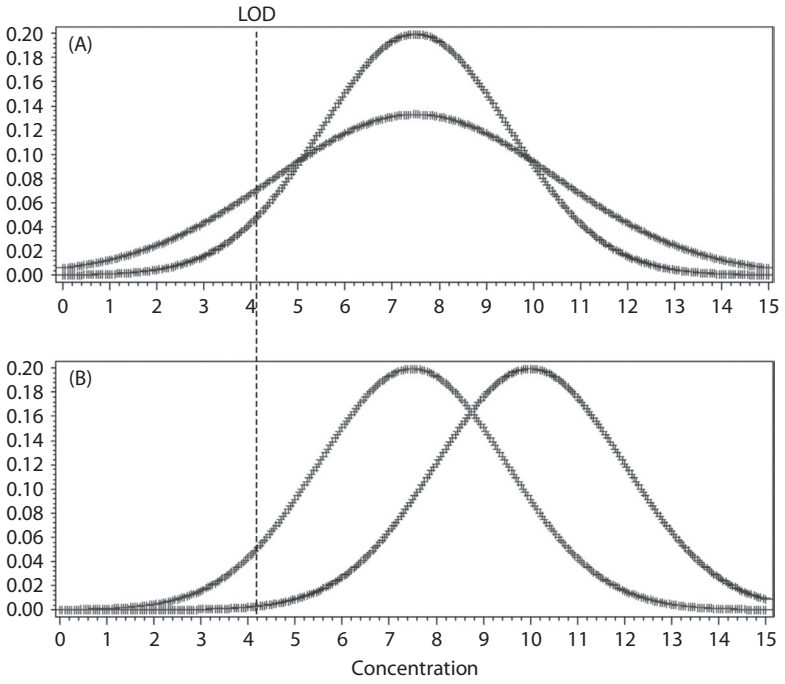

Figure 2.2 Problems associated with deletion or substitution techniques used to estimate means and standard deviations in left-censored data sets. Panel A shows two normal distributions with identical means but different standard deviations. Panel B shows two normal distributions with different means but identical standard deviations. The points censored (below the LOD) are those in the curve areas to the left of the vertical dashed line. (Modified from Newman, M.C., and P.M. Dixon, *Am. Environ. Lab.,* 4/90, 27–30, 1990.)

standard deviation for the distribution with the smaller mean more than that with the larger mean. Substitution of 0 will have the opposite effect.

2.2.2.3.2 Winsorized Mean and Standard Deviation

Winsorized estimates of mean and standard deviation can be used for censored data sets. The assumption is made that the data are distributed symmetrically. The <DL values are replaced by the smallest observation value above the LOD; however, the same number of largest values are also replaced by the value of the next smallest observation. The mean and standard deviation are calculated for this modified data set. The simple, arithmetic mean for the modified data (Winsorized mean) is unbiased (Equation 2.2). However, Equation (2.3) is needed to provide an unbiased estimate of the Winsorized standard deviation.

$$\overline{x}_w = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.2}$$

$$s_w = \frac{s(n-1)}{v-1} \tag{2.3}$$

where s = standard deviation of the modified data set, n = the total number of observations, and v = the number of observations not modified.

Example 2.1

The following sulfate concentrations (mg/L) were measured during a routine water quality survey of the Savannah River (South Carolina) (n = 21, mean = 5.1, standard deviation = 2.1):

1.3 ("<2.5")	3.5	5.2	6.5
2.3 ("<2.5")	3.6	5.6	6.9
2.6	4.0	5.7	7.1
3.3	4.1	6.1	7.7
3.5	4.5	6.2	7.9
			9.9

For this illustration, let's assume that the lowest two observations were censored; i.e., we have an embarrassing detection limit of only 2.5 mg/L. The smallest two observations are replaced by the next largest value (<2.5 and <2.5 become 2.6 and 2.6). Next, the two largest values are replaced by the value for the next smallest observation (7.9 and 9.9 become 7.7 and 7.7). The modified data set is now the following:

2.6	3.5	5.2	6.5
2.6	3.6	5.6	6.9
2.6	4.0	5.7	7.1
3.3	4.1	6.1	7.7
3.5	4.5	6.2	7.7
			7.7

The Winsorized mean ($\overline{x}_w$) is 5.08. The standard deviation of the modified data set (s) is 1.79. The number of observations (n) is 21 and the number of unmodified observations (v) is 17 (or n

– 4). The Winsorized standard deviation (s_w) is estimated with Equation (2.3) to be [1.79(21 – 1)]/(17 – 1), or 2.24. Rounding off, the Winsorized mean (5.1) for the censored data set is the same as that estimated for the uncensored data. The Winsorized standard deviation is slightly higher (2.2) than that estimated for the uncensored data (2.1).

The SAS code below performs the Winsorization just described and also the similar approach of trimming data from the two distribution tails. A 95% confidence interval for the mean (4.04 to 6.12) is estimated. In this case, a trimming of 0.952 from each tail results by intent in two observations being removed from each tail.

```
DATA ALL;
INPUT SO4 @@;
DATALINES;
1.3 2.3 2.6 3.3 3.5 3.5 3.6 4.0 4.1 4.5 5.2 5.6 5.7 6.1 6.2 6.5 6.9 7.1
7.7 7.9 9.9
;
RUN; /* WINSORIZATION OF TWO LOWEST AND TWO HIGHEST VALUES */
PROC CAPABILITY WINSOR=2(TYPE=TWOSIDED ALPHA=.05);
VAR SO4;
RUN; /* TRIMMING OF TWO LOWEST AND TWO HIGHEST VALUES */
PROC CAPABILITY TRIMMED=.0952;
VAR SO4;
RUN;
```

Winsorization or trimming is useful if the distribution is symmetrical. In Example 2.1, the assumption of a normal distribution is made after careful examination of the data set. The technique can also be used for some skewed data if a transformation is performed that produces a symmetrical distribution. For example, a symmetrical (normal) distribution is produced if a set of lognormal observations is log transformed. However, a backtransformation bias would then have to be estimated for log-transformed data, as will be described below. Further discussion of Winsorized estimates can be found in Dixon and Massey (1969), Sokal and Rohlf (1981), Gilbert (1987), and Berthouex and Hinton (1991).

A Winsorized t can be calculated that allows a confidence interval for the Winsorized mean to be generated (Dixon and Massey 1969),

$$t = \frac{v-1}{n-1} t_w \tag{2.4}$$

where t_w = a t from a conventional table or software function using a specified probability (p) and df of v – 1. The Excel® function TINV(1 – confidence level,df) can be used and, in this example, produces TINV(0.05,16) = 2.12. The confidence interval is then constructed,

$$\bar{x}_w \pm \frac{n-1}{v-1} t_{0.975(v-1)} \frac{s_w}{\sqrt{n}} \tag{2.5}$$

In this example, the 95% confidence interval is 5.08 ± 1.03, that is, 4.05 to 6.11. Notice in the PROC CAPABILITY statement given in Example 2.1 that a two-sided 95% confidence interval was requested during Winsorization. The estimated confidence interval from SAS was 4.04 to 6.12.

2.2.2.3.3 Probability Plotting

Perhaps the most straightforward method for estimating the mean and standard deviation of censored data sets is probability plotting. The sulfate data from Example 2.1 are used in Figure 2.3

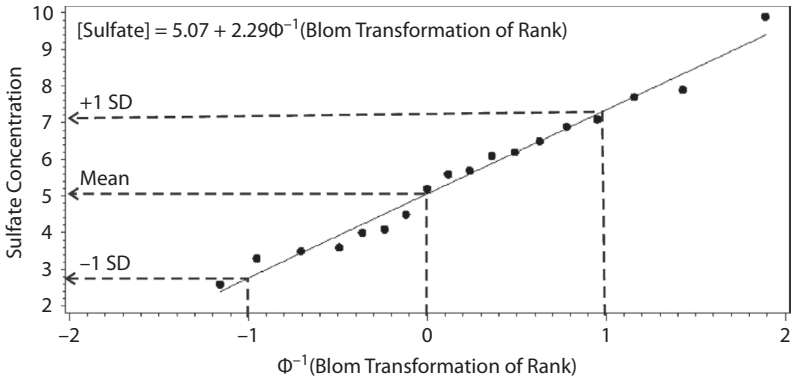

Figure 2.3 An example of estimating mean and standard deviation by probability plotting. The two lowest of the 21 sulfate concentrations are treated as less than the LOD. It is clear that, with the points above the censored values, a reasonable estimate of the mean and standard deviation can be obtained using concentrations corresponding to the regression line intercept and slope.

to illustrate this approach. Again, the lowest two observations are assumed to be below the LOD. The observations above the LOD are ranked from smallest to largest. The ranks for the 19 observations begin at 3 and progress to 21 because the censored observations occupied the first two lowest ranks of 1 and 2. The ranks can then be converted to proportions using one of several approximations, the van der Waerden (Equation 2.6), Tukey (Equation 2.7), or Blom (Equation 2.8) scaling transformations. Then these ranks are converted to their corresponding inverse normal scores (Φ^{-1}(proportion)), which express the proportions in units of standard deviations from the mean. The inverse normal score can be obtained readily with functions of numerous software packages such as Excel NORMINV(0.XX,0,1), where 0.XX is the estimated proportion and 0,1 specify a unit normal curve with a mean of 0 and standard deviation of 1.

$$\Phi^{-1}\ \frac{Rank}{n+1} \tag{2.6}$$

$$\Phi^{-1}\ \frac{Rank-\frac{1}{3}}{n+\frac{1}{3}} \tag{2.7}$$

$$\Phi^{-1}\ \frac{Rank-\frac{3}{8}}{n+\frac{1}{3}} \tag{2.8}$$

The Φ^{-1}(rank transformation) values are then plotted against their corresponding concentrations for all observations above the LOD (Figure 2.3). A linear regression model might be performed to produce a line intercept and slope that are estimates of the mean and standard deviation, respectively. The following code was used to generate Figure 2.3 and the regression estimates given in that figure. The scaling transformation preferred by this author (NORMAL = BLOM) was specified but the van der Waerden (NORMAL = VW) or Tukey (NORMAL = TUKEY) transformation might have been chosen instead.

```
GOPTIONS RESET=ALL BORDER;
DATA ALL;
INPUT SO4 @@;
IF SO4<2.5 THEN SO4=2.4; RANKSO4=SO4;
DATALINES;
1.3 2.3 2.6 3.3 3.5 3.5 3.6 4.0 4.1 4.5
5.2 5.6 5.7 6.1 6.2 6.5 6.9 7.1 7.7 7.9 9.9
;
RUN;
PROC RANK TIES=MEAN NORMAL=BLOM OUT=NEW;
VAR RANKSO4;
RUN;
DATA NPLOT;
SET NEW;
IF SO4 NE 2.4;
RUN;
PROC GLM;
MODEL SO4=RANKSO4/CLPARM;
OUTPUT OUT=PRED PREDICTED=PSO4;
RUN;
SYMBOL1 INTERPOL=JOIN VALUE=NONE COLOR=BLACK;
SYMBOL2 VALUE=DOT COLOR=BLACK;
PROC GPLOT DATA=PRED;
PLOT (PSO4 SO4)*RANKSO4/OVERLAY;
RUN;
```

The mean and standard deviation as estimated from the intercept and slope, respectively, were 5.07 and 2.29. The 95% confidence limits are also readily generated for the intercept and slope by including CLPARM in the GLM MODEL statement. The estimates are similar to the mean and standard deviation for the uncensored data set (5.1 and 2.1). This approach is referred to by Helsel (2005) as the fully parametric regression on order statistics approach, and is judged less useful than the robust regression on order statistics described below. It is worth noting that Antweiler and Taylor (2008) agree with Helsel's assessment. With the robust method (robust regression on order statistics) (Helsel 2005), the regression model is used to generate predicted concentrations for the censored observations and the predicted values for these observations are added to the uncensored data. The mean and standard deviation are then estimated with this filled-in data set. In this way, only the fill-in observation values require any assumption of a particular distribution to produce and the original values are used for the above LOD observations. Helsel often advocates for the assumption of a lognormal distribution (that is, log transformation of observations prior to method application) when generating the fill-in observation values because it is difficult to determine distribution type for many environmental data sets and his simulations have shown that the lognormal distribution is the most robust assumption to make. Helsel provides a SAS macro on his webpage (http://www.practicalstats.com/nada/) that implements this procedure. His macro was applied to the sulfate data with the following SAS code:

```
%MACRO cros(f,rem,value);
PROC PRINT;
RUN;
DATA SO4;
INPUT value rem $ @@;
DATALINES;
```