Glenn R. Guntenspergen *Editor*

Application of Threshold Concepts in Natural Resource Decision Making

Springer

Application of Threshold Concepts in Natural Resource Decision Making

Glenn R. Guntenspergen
Editor

Application of Threshold Concepts in Natural Resource Decision Making

Springer

Editor
Glenn R. Guntenspergen
USGS Patuxent Wildlife Research Center
Laurel
Maryland
USA

ISBN 978-1-4899-8040-3 ISBN 978-1-4899-8041-0 (eBook)
DOI 10.1007/978-1-4899-8041-0
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2014930313

© Springer Science+Business Media, LLC 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the PublisherâŁ™s location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Chapter 1
Threshold Concepts: Implications for the Management of Natural Resources

Glenn R. Guntenspergen and John Gross

Abstract Threshold concepts can have broad relevance in natural resource management. However, the concept of ecological thresholds has not been widely incorporated or adopted in management goals. This largely stems from the uncertainty revolving around threshold levels and the post hoc analyses that have generally been used to identify them. Natural resource managers have a need for new tools and approaches that will help them assess the existence and detection of conditions that demand management actions. Recognition of additional threshold concepts include: utility thresholds (which are based on human values about ecological systems) and decision thresholds (which reflect management objectives and values and include ecological knowledge about a system) as well as ecological thresholds. All of these concepts provide a framework for considering the use of threshold concepts in natural resource decision making.

Keywords Natural resource management · Non-linear · Regime shift · Time series

Natural resource managers face a complex decision-making environment that is not adequately addressed by traditional natural resource planning and decision-making processes. This situation can be partly attributed to changes in the dominant ecological paradigms used in natural resource management. In the past, habitat management has implicitly assumed that ecologists and managers are able to identify a "desired state" for ecosystems and that resource managers are then able to implement actions that can achieve and maintain the desired state. This philosophical strategy, aptly termed "command and control" (Holling and Meffe 1996), has been only partly successful and works best with problems that are relatively simple in terms of cause and effect (Knight and Meffe 1997). Historically, natural resource managers believed that the best way to achieve a "natural" state was to leave an area alone, or if it was

---

G. R. Guntenspergen (✉)
USGS Patuxent Wildlife Research Center,
12100 Beech Forest Road, Laurel, MD, 20708, USA
e-mail: glenn_guntenspergen@usgs.gov

J. Gross
Inventory and Monitoring Program, U.S. National Park Service,
1201 Oakridge Drive, Fort Collins, CO, 80525–5596, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making* DOI 10.1007/978-1-4899-8041-0_1,

disturbed, to simply remove the source of disturbance. In theory, a somewhat linear sequence of successional changes would eventually result in formation of a stable climax state (sensu Clements 1936).

With the recent appreciation of the complexity of ecosystem dynamics, the uncertainty associated with management actions, and the adoption of ecosystem management concepts (Grumbine 1994; Link et al. 2002; Tallis et al. 2010), ecologists and managers alike have embraced more quantitative methods and sophisticated models to guide management actions. These contemporary models accommodate a broader range of dynamics, and they often discard simple linear trajectories for those with non-linear behaviors and multiple possible outcomes.

The ideas of non-linear responses, tipping points, and regime shifts are now recognized as more likely the rule than the exception in ecological systems (Folke et al. 2004). Indeed there is growing evidence for strong non-linearities in the shaping of population dynamics (e.g., Stenseth et al. 1999) and the structure of ecosystems (e.g., Carpenter 2003). As a result, the concept of complex non-linear physical, chemical, and biological interactions and feedbacks is now generally accepted (Pielke et al. 2003; Scheffer and Carpenter 2003; Groffman et al. 2006). These ideas are at the core of the ecological threshold concept. This concept originates with the ideas of multiple ecological stable states (Holling 1973) and non-equilibrium systems (DeAngelis and Waterhouse 1987).

Threshold concepts can also have broad relevance to natural resource management. In this context, they are often viewed as triggers that prompt the need for specific actions to maintain a desired condition or keep a specific state variable within a desired range (Eaton et al., Chap. 5). Operational definitions of thresholds and their use by ecologists and managers have been an important area of focus (Briske et al. 2006; Bestelmeyer 2006; Groffman et al. 2006). Groffman et al. (2006) described three non-exclusive definitions of thresholds. The first definition describes abrupt and dramatic "shifts in ecosystem state." This is perhaps the most common use in the ecological community. A second definition describes "critical loads," which more specifically applies to levels of pollutant inputs that result in unacceptable ecosystem responses. The third definition describes "extrinsic factor thresholds" where cross-scale interactions lead to abrupt changes. This final use falls within the conceptual framework of hierarchy theory, where broad-scale factors constrain system dynamics (Allen and Hoekstra 1992).

Bestelmeyer (2006) offers contrasting definitions of thresholds, focusing on use of threshold concepts in rangelands and identifies ambiguities related to ecological scale, pattern, and process. To address the need for a unifying framework, Bestelmeyer (2006) proposed a classification of thresholds consisting of pattern thresholds, process thresholds, degradation thresholds, and a more synthetic set of classification thresholds based on either preventative management or restoration of rangeland. This framework accommodates many of the requirements for rangeland managers and places an emphasis on broadening the attributes used to define thresholds.

Others have proposed a more general definition of thresholds that include "a defined target level or state based on the avoidance of unacceptable outcomes or an ecologically defined shift in system status" (Polasky et al. 2011). Martin et al. (2009) distinguish between three broad threshold concepts that are relevant for natural resource managers and add the concept of decision and utility thresholds to that of ecological thresholds. Decision thresholds represent values of a state variable that when exceeded should elicit management action. "Utility thresholds" are derived from management objectives and indicate where "small changes in environmental conditions produce substantial improvements in management outcomes. . . " (Samhouri et al. 2010). These alternative concepts are not easily reconciled with the identification of ecological thresholds, nor do they provide a general conceptual basis that fully integrates our understanding of thresholds into a comprehensive decision-making process (Martine et al. 2009; Polasky et al. 2011).

Increasingly, the importance of understanding interactions between and among biotic and abiotic factors in ecosystems and how these interactions lead to complexities are factored into resource management actions (Huggett 2005; Groffman et al. 2006; Bestelmeyer 2006; Andersen et al. 2009; Suding and Hobbs 2009; Hobbs and Suding 2009). However, the widespread acceptance of threshold concepts in ecological models—"ecological thresholds"—has not been followed by their widespread adoption and incorporation into management goals (Hobbs and Suding 2009). The ability to move from theory to application and make threshold concepts a problem-solving tool for natural resource management remains a daunting challenge. One of these impediments involves confusion over the appropriate use of threshold concepts in natural resource decision-making processes. Bennetts et al. (2007) described seven concepts widely used by natural resource management agencies in identifying points or zones of interest to managers and that could be used to inform the management of natural resources. In addition to ecological thresholds, these include: critical loads, regulatory or policy standards, management thresholds, desired condition, range of natural variation, and thresholds of potential concern. The typical implementation of these concepts ranges from precisely defined quantities to more qualitative descriptions; and each of these concepts contributes to our broader understanding of the use of threshold concepts in natural resource management. These seven concepts encompass the three types of thresholds proposed by Martin et al. (2009) as relevant for natural resource decisions: ecological thresholds, utility thresholds, and decision thresholds. Decision thresholds have often been referred to as management thresholds, and utility thresholds can in certain cases coincide with ecological thresholds (Samhouri et al. 2010). The other concepts identified by Bennetts et al. (2007) can be used to develop utility and decision thresholds. When regulatory thresholds like water or air quality standards or critical loads are exceeded, the responses may be clearly dictated by law, with little latitude for local decision making. However, for many natural resource management situations, the use of desired condition, range of natural variation, and thresholds of potential concern may result in a variety of reasonable responses when attributes approach or exceed a (sometimes arbitrarily defined) value.

So, beyond agreeing that ecological thresholds may be common and sometimes important, there is no shared understanding or agreement on the role or appropriate use of this concept in natural resource management in spite of the fact that there is a rich literature that addresses the concept of ecological resilience and alternative stable states. Likely, this largely stems from the uncertainty revolving around threshold levels and the post hoc analyses that have generally been used to identify them. Advances in and new applications of statistical techniques (Andersen et al. 2009; Ficetola and Denoel 2009) have greatly enhanced our ability to detect the locations of thresholds once they have been crossed, but most techniques still rely on long-term temporal series of observations (e.g., Carpenter and Brock 2006, Andersen et al. 2009). Identifying the level at which threshold behavior occurs may be possible if we can accumulate a large body of empirical observations. Otherwise, new work in identifying generic early warning metrics may provide our best opportunity for identifying the probability of such an event occurring (Scheffer et al. 2009; Biggs et al. 2009; Scheffer et al 2012).

Natural resource managers have a need for new tools and approaches that will help them assess the existence and detection of conditions that demand management actions. This book addresses several of the issues that have profoundly affected the use of thresholds in natural resource management—uncertainty, different types of thresholds, appropriate use of thresholds in decision making, and the development of a comprehensive decision framework as a unifying approach for threshold concepts.

The first set of chapters in this book provide a conceptual framework for threshold concepts in natural resource management and conservation based on the theory of structured decision making. Risk analysis (Suter 2007), decision theory (Morgan et al. 1990), and structured decision making (Martin et al. 2009) have all been promoted as a means to advance natural resource management decisions. These approaches provide a structured process that enables natural resource decision makers to identify interventions that can lead to improvement or to avoid future problems. Each of these frameworks has three elements—a clear statement of the problem and objectives, a list of discrete management actions, and quantitative scientific information in the form of one or more models that can be used to predict the outcome of different management actions. Nichols et al. (Chap. 2) provide a conceptual framework (Structured Decision Making) for the use of threshold concepts in natural resource decision making and discuss the important distinctions between utility, decision, and ecological thresholds. Runge and Walshe (Chap. 3) provide a more expanded description of identifying objectives and alternative actions needed to frame a natural resource decision problem. Williams and Nichols (Chap. 4) then describe the role of optimization in providing an objective approach for deciding which potential action to take. Finally, Eaton et al. (Chap. 5) illustrate an application of the various classes of thresholds introduced by Nichols et al. (Chap. 2) and their use in structuring a decision process for the management of human recreational activities and the impact of nesting Golden Eagles in Alaska's Denali National Park.

The next four chapters discuss the role of monitoring for threshold-dependent decisions and the evaluation of bioassessment designs. Smith et al. (Chap. 6) review the literature on monitoring for threshold-dependent management decisions and compare adaptive management with targeted monitoring with the sequential evaluation of resource condition with surveillance monitoring. They further build on the prior section by examining the threshold concepts of ecological change, utility value, and decision threshold in resource management and how these concepts are incorporated into resource management and monitoring. Bowker et al. (Chap. 7) use case studies from the dryland ecosystems of the Colorado Plateau to illustrate how state and transition models can be used to identify transition and triggers likely to be detectable by monitoring programs. Symstad and Jonas (Chap. 8) examine how our understanding of the natural range of variation for plant communities can be used in developing decision thresholds when ecological thresholds are unknown or do not exist. Snyder et al. (Chap. 9) illustrate how simulation techniques may be used to optimize bioassessment decision thresholds and sampling designs with a case study of benthic macroinvertebrate communities in a US National Park. Finally, Mitchell et al. (Chap. 10) use ongoing monitoring data by the US National Park Service Vital Signs Program to illustrate how threshold detection can be used in establishing ecological assessment points and how the concept of ecological integrity can be reported to resource managers and decision makers. They describe and illustrate how concepts of ecological integrity, thresholds, and reference conditions (natural range of variability) can be integrated into a research and monitoring network.

Field data are being explored with new statistical and graphical techniques, and more sophisticated models are being used in the monitoring and management of ecosystems and the detection of response patterns. The final series of chapters in this book describe different quantitative approaches to estimate ecological thresholds. King and Baker (Chap. 11) describe how a new method Threshold Indicator Taxa Analysis (TITAN) uses ecological community data for estimating community thresholds. They use a case study that examines macroinvertebrate community response to a phosphorus gradient in the Everglades, a large subtropical wetland in the southern USA. Carstensen (Chap. 12) introduces a statistical inferential approach based on generalized additive models to examine ecosystem trajectories during degradation and recovery phases using observations from four monitoring programs of phytoplankton communities in northeastern European coastal waters. Washington-Allen and colleagues (Chap. 13) used biophysical models Normalized Difference Vegetation Index (NDVI) from a time series of Landstat images of the Mojave Desert of the western USA to examine the hypothesis that changes in the variance, as a threshold is approached, may provide an early warning signal of change. The concluding chapter by James Pirri et al. (Chap. 14) illustrates how threshold concepts can be used by managers to evaluate responses to restoration activities or describe the overall condition of salt marsh ecosystems along the northeastern Atlantic coast of the USA. They use multivariate methods to illustrate how shifts in the characteristics of vascular plant and nekton communities can be used as ecological thresholds upon which decision thresholds for natural resource managers can be used.

The threshold concept has become a major theme in ecology, and advocates suggest that it can also play a key role in natural resource management, restoration, conservation, and land policies. Like many issues and concepts, threshold concepts can mean different things to different people. The discussion of thresholds in the literature has largely emphasized the identification of ecological thresholds and their role as components of ecological models in predicting system responses to management actions, but has not always been clear about the distinctions among different threshold types. Managers and scientists are not necessarily limited to the ideas and concepts of ecological thresholds when considering the management of natural systems. Increasingly, utility thresholds (which are based on human values about ecological systems) and decision thresholds (which reflect management objectives and values and include ecological knowledge about a system) have also been promoted (Martin et al. 2009). The chapters and case studies in this book illustrate how these different threshold concepts can be applied in conservation and land management decisions.

References

Allen, T. F. H., and T. W. Hoekstra. 1992. *Toward a unified ecology*. New York: Columbia University Press.

Andersen, T., J. Carstensen, E. Hernandez-Garcia, and C. M. Duarte. 2009. Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evolution* 24:49–57.

Bennetts, R. E., J. E. Gross, K. Cahill, C. McIntyre, B. Bingham, A. Hubbard, L. Cameron, and S. Carter. 2007. Linking monitoring to management and planning: assessment points as a generalized approach. *George Wright Forum* 64:59–77.

Bestelmeyer, B. T. 2006. Threshold concepts and their use in rangeland management and restoration: The good, the bad, and the insidious. *Restoration Ecology* 14:325–329.

Biggs, R., S. R. Carpenter, and W. A. Brock. 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Sciences* 106:826–831.

Briske, D. D., S. D. Fuhlendorf, and F. E. Smeins. 2006. A unified framework for assessment and application of ecological thresholds. *Rangeland Ecology and Management* 59:225–236.

Carpenter, S.R. 2003. *Regime shifts in lake ecosystems: Pattern and variation*. Oldendorf: International Ecology Institute.

Carpenter, S. R., and W. A. Brock. 2006. Rising variance: A leading indicator of ecological transition. *Ecology Letters* 9:311–318.

Clements, F. E. 1936. Nature and structure of the Climax. *Journal of Ecology* 24:252–284.

DeAngelis, D. L., and J. C. Waterhouse. 1987. Equilibrium and nonequilibrium concepts in ecological models. *Ecological Monographs* 57:1–21.

Ficetola, G. F., and M. Denoel. 2009. Ecological thresholds: An assessment of methods to identify abrupt changes in species-habitat relationships. *Ecography* 32:1075–1084. doi:10.1111/j.600-0587.2009.05571.x.

Folke, C., S. Carpenter, B. Walker, M. Scheffer, T. Elmqvist, et al. 2004. Regime shifts, resilience, and biodiversity in ecosystem management. *Annual Review Ecology, Evolution, and Systematics* 35:557–581.

Groffman, P., J. Baron, T. Blett, A. Gold, I. Goodman, L. Gunderson, B. Levinson, M. Palmer, H. Paerl, G. Peterson, N. Poff, D. Rejeski, J. Reynolds, M. Turner, K. Weathers, and J. Wiens. 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? *Ecosystems* 9:1–13.

Grumbine, R. E. 1994. What is ecosystem management. *Conservation Biology* 8:27–38.

Hobbs, R. J., and K. N. Suding, eds. 2009. *New models for ecosystem dynamics and restoration*. Washington, D.C.: Island Press.

Holling, C. S. 1973. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics* 4:1–24.

Hollings, C. S., and G. K. Meffe. 1996. Command and control and the pathology of natural resource management. *Conservation Biology* 10:328–337.

Huggett, A. J. 2005. The concept and utility of 'ecological thresholds' in biodiversity conservation. *Biological Conservation* 124:301–310.

Knight, R. L., and G. K. Meffe. 1997. Ecosystem management: Agency liberation from command and control. *Wildlife Society Bulletin* 25: 676–678.

Link, J. S., J. K. T. Brodziak, S. F. Edwards, W. J. Overholtz, D. Mountain, J. W. Jossi, T. D. Smith, and M. J. Fogarty. 2002. Marine ecosystem assessment in a fisheries management context. *Canadian Journal of Fisheries and Aquatic Sciences* 59:1429–1440.

Martin, J., M. C. Runge, J. D. Nichols, B. C. Lubow, and W. L. Kendall. 2009. Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications* 19:1079–1090.

Morgan, M. G., M. Henrion, and M. Small. 1990. *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge: Cambridge University Press.

Pielke, R. A., H. J. Schellnhuber, and D. Sahagian. 2003. Non-linearities in the earth system. *Global Change Newsletter* 55: 11–15.

Polasky, S., S. R. Carpenter, C. Folke, and B. Keeler. 2011. Decsion-making under great uncertainty: Environmental management in an era of global change. *Trends in Ecology and Evolution* 26: 398–404. doi:10.1016/j.tree.2011.04.007.

Samhouri, J. F., P. S. Levin, and C. H. Ainsworth. 2010. Identifying thresholds for ecosystem-based management. *PLoS One* 5: e8907. doi:10.1371/journal.pone.000890.

Scheffer, M., and S. R. Carpenter. 2003. Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends in Ecology and Evolution* 18:648–656.

Scheffer, M., J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk, and G. Sugihara. 2009. Early-warning signals for critical transitions. *Nature* 461:53–59.

Scheffer, M., S. R. Carpenter, T. M. Lenton, J. Bascompte, W. Brock, V. Dakos, J. van de Koppel, I. A. van de Leemput, S. A. Levin, E. H. van Nes, M. Pascual, and J. Vandermeer. 2012. Anticipating critical transitions. *Science* 338: 344–348.

Stenseth, N. C., K. S. Chan, H. Tong, R. Boonstra, S. Boutin, C. J. Krebs, E. Post, M. O'Donoghue, N. G. Yoccoz, M. C. Forchhammer, and J. Hurrell. 1999. Common dynamic structure of Canada Lynx populations within three climatic regions. *Science* 285:1071–1075.

Suding, K. N., and R. J. Hobbs. 2009. Threshold models in restoration and conservation: A developing framework. *Trends in Ecology & Evolution* 24:271–279.

Suter, G. W. 2007. *Ecological Risk assessment*. Boca Raton: CRC Press.

Tallis, H., P. S. Levin, M. Ruckelshaus, S. E. Lester, K. L. McLeod, D. L. Fluharty, and B. S. Halpern. 2010. The many faces of ecosystem-based management: Making the process work today in real places. *Marine Policy* 34:340–348.

Chapter 2
Thresholds for Conservation and Management: Structured Decision Making as a Conceptual Framework

James D. Nichols, Mitchell J. Eaton and Julien Martin

Abstract A conceptual framework is provided for considering the threshold concept in natural resource management and conservation. We define three kinds of thresholds relevant to management and conservation. *Ecological thresholds* are values of system state variables at which small changes bring about substantial or specified changes in system dynamics. They are frequently incorporated into ecological models used to project system responses to management actions. *Utility thresholds* are components of management objectives and are values of state or performance variables at which small changes yield substantial changes in the value of the management outcome. *Decision thresholds* are values of system state variables at which small changes prompt changes in management actions in order to reach specified management objectives. Decision thresholds are derived from the other components of the decision process. We advocate a structured decision making (SDM) approach within which the following components are identified: objectives (possibly including utility thresholds), potential actions, models (possibly including ecological thresholds), monitoring program, and a solution algorithm (which produces decision thresholds). Adaptive resource management (ARM) is described as a special case of SDM developed for recurrent decision problems that are characterized by uncertainty. We believe that SDM, in general, and ARM, in particular, provide good approaches to conservation and management. Use of SDM and ARM also clarifies the distinct roles of ecological thresholds, utility thresholds, and decision thresholds in informed decision processes.

Keywords Adaptive management · Decision threshold · Ecological threshold · Structured decision making · Utility threshold

---

J. D. Nichols (✉)
U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, MD, USA
e-mail: jnichols@usgs.gov

M. J. Eaton
Southeast Climate Science Center, U.S. Geological Survey,
127H David Clark Labs, NCSU, Raleigh, NC27695, USA

J. Martin
Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute,
100 8th Ave SE, St Petersburg, FL, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_2,

Introduction

Thresholds and their relevance to conservation are widely discussed by ecologists, conservation biologists, managers, and policy makers (Burgman 2005; Bestelmeyer 2006). These discussions are certainly useful in many respects, but they can also lead to confusion about how thresholds should be used in the conduct of conservation. In this chapter, we provide a conceptual framework for thresholds that we hope will be useful to those involved in conservation and management. We define three general classes of thresholds. Our purpose in doing so is not simply to introduce new vocabulary to a subject area already rich in terminology, but rather to draw distinctions among thresholds that have specific, yet different, uses in conservation programs. Our focus on the use of thresholds in decision processes requires a description of such processes, as they provide the framework required for our discussion.

Structured decision making (SDM; Clemen and Reilly 2001) is a logical and transparent process that requires breaking a decision into its component parts. This decomposition insures that discussions among stakeholders with different opinions are properly focused and helps to clarify points of agreement and disagreement. The components identified in SDM also serve to clarify roles of different participants in the decision process. Some components focus on values and require substantive input from all relevant stakeholders, whereas other components focus on system dynamics and are addressed primarily by managers and scientists. Most relevant to this chapter, adoption of SDM leads naturally to consideration of definitions and roles of different kinds of thresholds in the conservation process.

We will structure this chapter by first defining three types of thresholds relevant to conservation decisions. We then describe the components of the SDM process, emphasizing the position and role of each type of threshold with respect to these components. We next describe adaptive resource management (ARM) as a special case of SDM developed for recurrent decisions characterized by uncertainty. Finally, we provide a discussion of this threshold framework and advocate its use with SDM for conservation decision making.

Thresholds

Ecological Thresholds

Three kinds of thresholds are relevant to making decisions in conservation : *ecological, utility,* and *decision* thresholds (Martin et al. 2009a). *Ecological thresholds* have been defined in many ways, but common to most definitions is a point or zone at which there is a sudden change in the condition or dynamics of a biological system (e.g., Fahrig 2001; Huggett 2005; Pascual and Guichard 2005; Groffman et al. 2006; Bennetts et al. 2007). We operationally define an ecological threshold as a value (or set of values) of a state variable, environmental variable, or rate parameter of a system at which small changes either (1) produce changes in system dynamics of specified

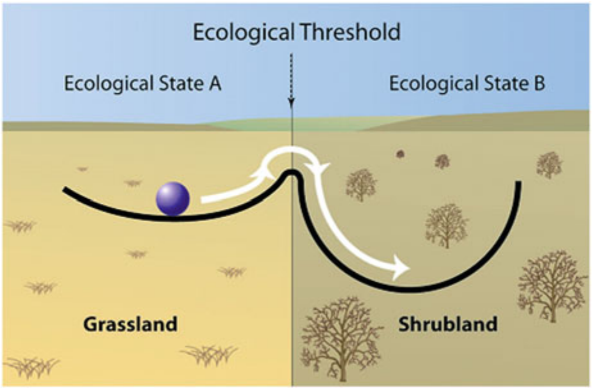
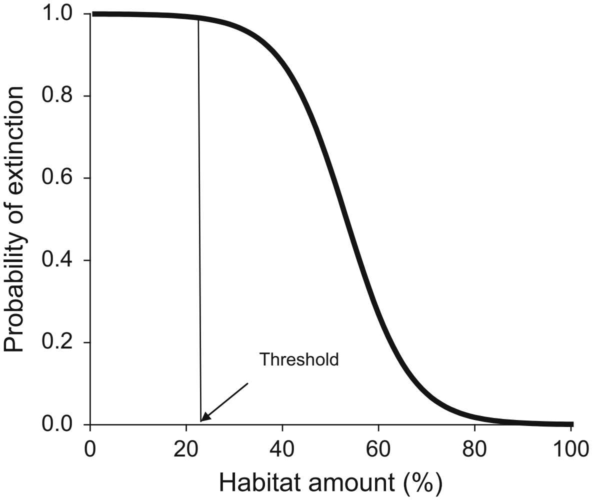


Fig. 2.1 Example of an ecological threshold. In this example a small change in the amount of precipitation (environmental variable) leads to a substantial change in system state from grassland (*ecological state A*) to shrubland (*ecological state B*). The *ball* and *valleys* provide an illustration of the tendency to remain in the same ecological state, or with the possibility to switch to another ecological state. (Reproduced from Bennetts et al. 2007)

magnitude (typically large or ecologically substantial changes) or (2) cause system state variables or rate parameters to attain certain specified values. An example of the first kind of ecological threshold can be found in vegetation communities of the Chihuahuan Desert (Fig. 2.1). Precipitation is a key environmental variable of this system, and an ecological threshold is the level(s) of precipitation at which small changes induce a shift from grass- to shrub-dominated communities and vice versa (Brown et al. 1997; Groffman et al. 2006). An example of the second kind of ecological threshold is Lande's (1987) concept of extinction threshold for metapopulation systems. In this case, the proportion of potentially available habitat that is suitable for the focal species is an important system state variable. The extinction threshold is the proportion of suitable habitat at which probability of metapopulation extinction becomes one (Fig. 2.2; see Lande 1987; Fahrig 2001; Benton 2003).

We have no strict views about the functional forms of ecological thresholds, as illustrated by two examples of thresholds from Martin et al. (2009a). A step function corresponds closely to most views of the threshold concept. For example, Fig. 2.3a depicts an ecological threshold as a value of a state variable (1,500 units of water in a wetland) at which a vital rate (rate of patch colonization) increases from 0 to 0.1. The threshold concept can also apply to regions of a functional relationship at which small changes in one variable produce large changes in another. Figure 2.3b depicts such a case, where changes in water levels within a particular region (600–1,250 units of water) produce large changes in probability of patch extinction. Some discussions of ecological thresholds focus on shifts of state variables to an absorbing state (e.g., permanent extinction) from which transition is not possible (Lande 1987). Discussions of ecological thresholds frequently include other terms relevant to system change and dynamics. The concept of "resilience" (Holling 1973; Gunderson 2000) concerns the magnitude of perturbation required to induce a substantive change in system state. "Elasticity" (Bodin and Wiman 2007) refers to aspects (e.g.,

Fig. 2.2 Probability of metapopulation extinction as a function of the amount of suitable habitat remaining. The extinction threshold is the proportion of suitable habitat at which probability of metapopulation extinction becomes one (or very close to one). (Based on Lande 1987; Fahrig 2001)



time elapsed) of transient dynamics following a perturbation as a system returns to equilibrium.

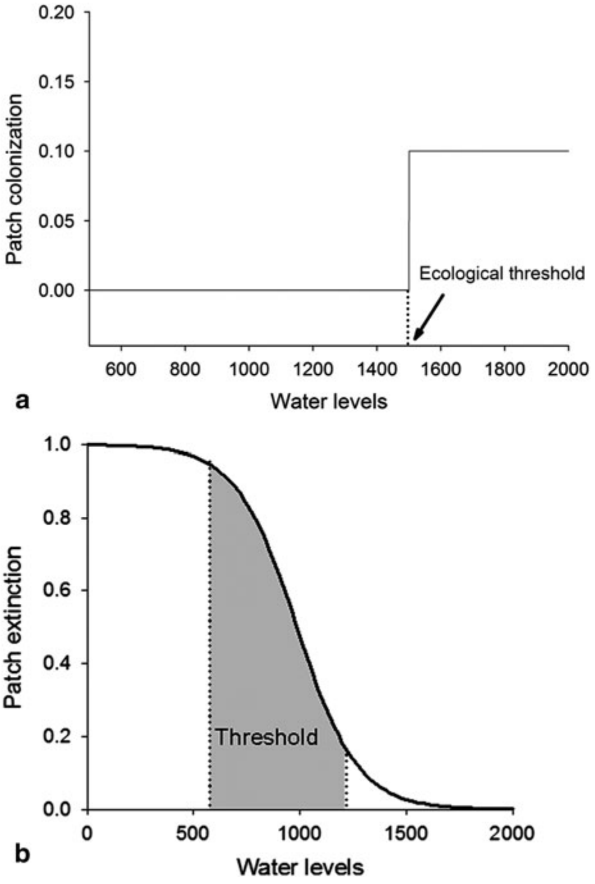
Our definition of ecological threshold is thus very general, and we acknowledge that discussions of related concepts can be very wide-ranging. However, the role of ecological thresholds in management and conservation is very specific: They are components of models used to predict system responses to management actions. Ecological models need not include thresholds, as threshold concepts may not be relevant to the dynamics of all ecological systems. However, when ecological thresholds are relevant to system dynamics and response to management, they are incorporated in the functional relationships of ecological models (Martin et al. 2009a; see also Conroy et al. 2003; Bestelmeyer 2006).

Utility Thresholds

We define *utility thresholds* as values of state or performance variables at which small changes yield substantial changes in the value of the management outcome. For example, we might specify that an objective of management for a particular species in a national park is that the population size should remain above some level, say N^*. Unlike ecological thresholds, which are part of the pattern and process of nature, utility thresholds are determined by human values. In many cases, utility thresholds have some ecological basis; for example, they are frequently based on historical observations of system state variables (e.g., Runge et al. 2006; Martin et al. 2011). But there is no necessary link between utility thresholds and ecology; instead, utility thresholds provide explicit statements of what managers value.

Statements of management objectives need not include utility thresholds. For example, a management objective might be to minimize the probability that an endangered species becomes extinct over a specified time horizon. Utility thresholds

Fig. 2.3 Illustration of two types of ecological threshold based on the example from Martin et al. (2009a). **a** The diagram depicts an ecological threshold as a value of a state variable (1,500 units of water in a wetland) at which a vital rate (rate of patch colonization) increases from 0 to 0.1. **b** The graph depicts a threshold zone where changes in water levels within a particular region (600–1,250 units of water) produce large changes in probability of patch extinction



are frequently used in objective functions that include competing objectives. For example, in Chap. 5 (Eaton et al.) we describe management of potential disturbance by hikers and tourists to golden eagles in Denali National Park (see also Martin et al. 2009b; Martin et al. 2011). Park managers seek to provide a rewarding experience to hikers, but also want to maintain a healthy breeding population of golden eagles. The objective function for this specific decision problem is to minimize the number of eagle nesting territories at which hiker access is restricted, while maintaining the occupancy of potential territories above a specified utility threshold (e.g., 0.8).

Decision Thresholds

We define *decision thresholds* (sometimes referred to as management thresholds, see Bennetts et al. 2007) as values of system state variables that should prompt specific management actions. Decision thresholds are thus conditional on, and derived from,

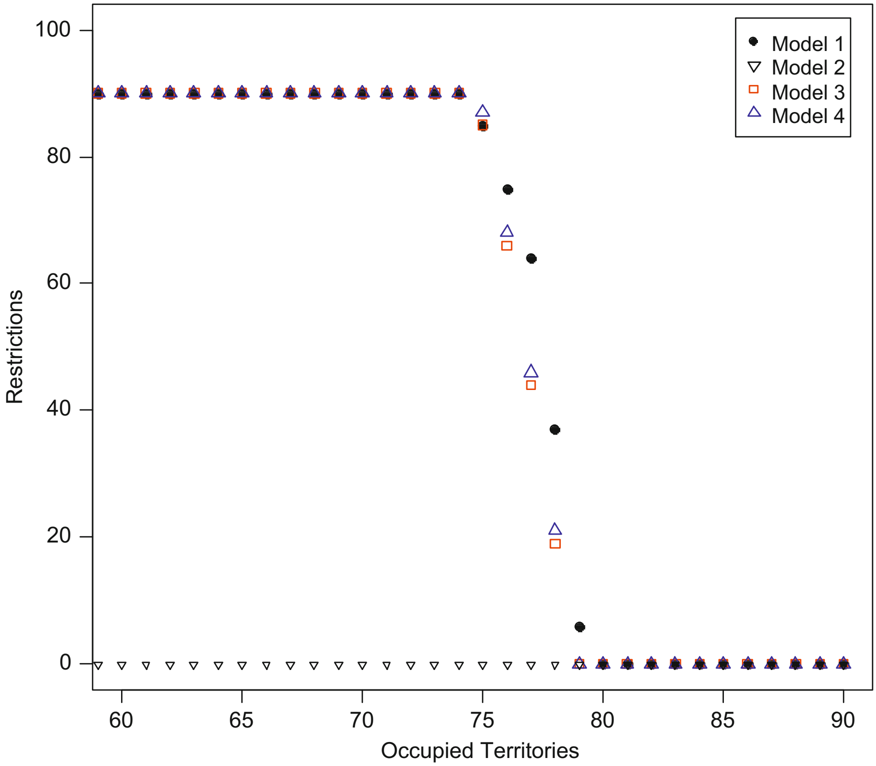


Fig. 2.4 Policy matrix showing the optimal number of restricted territories as a function of the number of eagle territories that are occupied. (From Eaton et al., Chap. 5)

ecological and utility thresholds. In the example of Denali golden eagles and hikers, golden eagle occupancy proportion of potential nest sites is potentially affected by hiker disturbance. The management decision is whether to close hiker access to potential territories. Because of the desire to minimize restrictions to hikers, if projected eagle occupancy is sufficiently high relative to the utility threshold, hikers will not be restricted. However, as current eagle occupancy reaches levels that are sufficiently low that projections indicate a good possibility of dropping below the utility threshold, the optimal action will be to restrict hikers. The value of the state variable(s) (proportion of potential territories that are occupied) at which the recommended action shifts from no hiking restrictions to restrictions can be viewed as a decision threshold.

An example policy matrix for the Denali golden eagle example presented in Chap. 5 (see also Martin et al. 2011) is shown in Fig. 2.4. While the detailed analysis of Martin et al. (2011) focused on 25 out of 93 territories that were believed to have the potential to be disturbed by hikers, Eaton et al. (Chap. 5) focused their analysis on a hypothetical 90 nesting sites, all with the potential for closure. Specifically, the management decision is, "How many of these sites should be closed to hikers in order to minimize closures while keeping the projected number of occupied eagle territories above a utility threshold based on historic data?" A stochastic dynamic programming algorithm (Bellman 1957) was originally implemented using the software of Lubow (1995) to derive the optimal policy (Fig. 2.4). The decision policy is based on the number of these 90 sites that are occupied. The vertical axis in Fig. 2.4 represents the management decision at any level of system state, specified as number of territories restricted. Under any of the four proposed dynamic models, the optimal number of restrictions is 0 sites if the number of occupied sites is between 80 and 90, so there is no decision threshold for these values of the state variable. However, if the number of occupied territories drops to 79, then the optimal number of restricted sites (under one hypothesis of occupancy dynamics) shifts from 0 to 6. This change in number of occupied territories from 80 to 79 thus represents a decision threshold, because different actions are recommended for these two different values of the state variable.

Sources of Confusion

Discussions of thresholds and their role in conservation have not always been clear, especially with respect to the distinctions among the three types of thresholds that we have identified. For example, it is common for managers to equate utility and decision thresholds. One approach to management under the declining population paradigm (Caughley 1994) is to view a finite rate of population increase (λ) of 1 simultaneously as a utility and a decision threshold. A declining population ($\lambda < 1$) is viewed as undesirable, such that $\lambda = 1$ is a utility threshold. The manager periodically tests for a negative trend in abundance (e.g., using monitoring data and statistical models and inference procedures). If a "significant" negative trend is detected, then management actions are taken, so $\lambda = 1$ is also viewed as a decision threshold.

Management under the SDM approach that we advocate tends to produce decision thresholds that are more conservative than this trend-detection approach. If $\lambda = 1$ is our utility threshold, then under optimal management, actions typically occur before the population is actually declining, in an effort to keep $\lambda \geq 1$. Indeed, the trend-detection approach has been criticized as leading to unnecessary delays in management actions (Maxwell and Jennings 2005; Nichols and Williams 2006). In addition, the usual approach of placing trend detection in a hypothesis-testing framework invites discussion about type I and II error rates (e.g., arbitrary α for hypothesis testing) and the relative risks associated with these errors (see Field et al. (2004) for a discussion of this topic). Use of SDM and treatment of decision processes as optimization problems, rather than as problems of hypothesis testing, produce decision thresholds that frequently differ from utility thresholds.

Synthesis

Ecological thresholds may characterize the dynamics of managed ecological systems. When this is true, and when they can be identified (this can be difficult), they should be incorporated into the models used by managers in the decision process. Utility

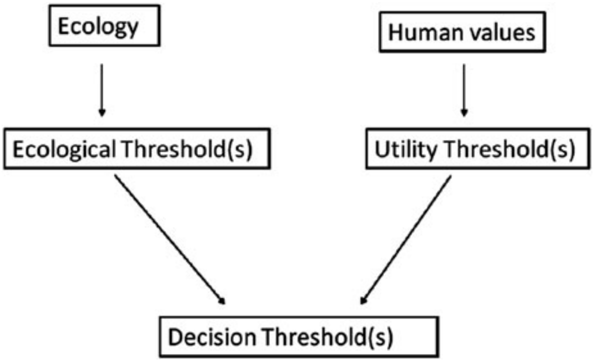


Fig. 2.5 Relationships among ecological, utility, and decision thresholds. (Modified from Martin et al. 2009a)

thresholds reflect human values about ecological systems and may be included in management objectives. Decision thresholds are derived from the ecological and utility thresholds or, more generally, from management objectives, available actions, and models of system dynamics and responses to management. These relationships among the different types of thresholds are depicted in Fig. 2.5.

Structured Decision Making (SDM)

SDM is a formal decision process employed to identify decisions that are optimal with respect to specified objectives. SDM is rooted in decision theory, which provides a useful framework for making decisions about the management of virtually any kind of system (Bellman 1957; Intriligator 1971; Williams et al. 2002; Burgman 2005; Halpern et al. 2006). SDM has been used in a variety of fields, including engineering, economics, and natural resource management (e.g., Johnson et al. 1997; Clemen and Reilly 2001; Miranda and Fackler 2002; Halpern et al. 2006). In the context of conservation, the elements of the decision-making problems often include the following components: objectives, potential management actions, model(s) of system behavior (specifically, models that predict how system state is expected to change with application of each different management option), a monitoring program to provide estimates of system state variables, other variables related to management returns, system vital rates, and finally a method to identify the solution (Williams et al. 2002; Dorazio and Johnson 2003; McCarthy and Possingham 2007). Two of these components, model(s) and estimates of system state, are typically characterized by substantial uncertainties that must be accommodated in the optimization process.

Objectives and Management Actions

The specification of objectives is a critical component of any decision-making process. Objectives should reflect the values of relevant stakeholders and constitute specific statements of what is to be achieved by implementing management actions. Objectives provide the currency by which alternative decision options are judged (Clemen and Reilly 2001; Conroy and Moore 2001). Examples of objectives relevant to conservation include maximizing species diversity in a natural area or minimizing the probability of quasi-extinction of a threatened species (Kendall 2001). As noted above, objectives may be stated as utility thresholds, such as maintaining a population size at or above some specified value.

In cases involving multiple stakeholders with competing interests, utility thresholds are often used as a means of providing constraints on competing objectives. In the example of Denali golden eagles (Martin et al. 2011; Eaton et al., Chap. 5), competing objectives were a desire to permit hikers to fully enjoy Denali National Park and a desire to maintain a healthy breeding population of golden eagles. The hypothesis that disturbance by hikers may limit occupancy and/or reproductive success of golden eagles at potential nesting sites leads to a consideration of trade-offs between objectives. In this case, the objective was expressed as minimizing the number of sites at which hiker access was restricted, subject to the constraint that predicted golden eagle occupancy or successful reproduction exceeded a specified utility threshold (Martin et al. 2011; Eaton et al., Chap. 5). Thus, utility thresholds may be used to specify simple objectives or to serve as constraints for problems with competing objectives.

Objectives (including associated constraints) should generally be determined through discussions among stakeholders (Kendall 2001). This determination can be one of the most difficult steps in a decision process, especially in the common case where different stakeholder groups have competing values and interests. Formal techniques are sometimes used to elicit values and select appropriate objectives (see Clemen and Reilly 2001; Burgman 2005). Once objectives and constraints have been selected, they can be formalized mathematically into an objective function. The objective function quantifies the benefit (or return) obtained by implementing specific decisions at each time step, accumulated over the time horizon of the decision problem (Lubow 1995; Williams et al. 2002; Fonnesbeck 2005).

The other component of SDM that is driven primarily by human values is the selection of the set of management actions to be considered. Frequently in conservation settings, the set of available actions is very small. Actions can include regulations that restrict harvest or various activities that cause human disturbance to a natural area (boating, hiking, using snowmobiles). Actions can also include various forms of habitat management, land acquisition, translocation of animals, etc. Sometimes, actions (e.g., predator control) that may be potentially useful and cost-effective are viewed as unacceptable based on human values. In summary, objectives and the set of potential management actions are not established by managers and scientists alone, but should be based on the values of all relevant stakeholders. Objectives and available actions are extremely important in SDM as they effectively drive the entire decision process.

Model(s) of System Behavior

Informed decisions require some basis for predicting effects of the different actions under consideration. Absent the ability to predict consequences of management actions, such actions might be determined by virtually any random process, but terms such as "management" and "conservation" do not really apply to such uninformed manipulation of a system. Models can be viewed as structures that provide predictions based on hypotheses about how the focal system "works" or, more specifically, how it responds to management actions. Models may reside in the heads of wise managers, or they may be mathematical, perhaps incorporated into computer code. Models that project the consequences of management actions should generally be developed by scientists and managers familiar with both the managed system and general principles of system dynamics. Although input from knowledgeable stakeholders is welcome, stakeholders are generally not as important to model development as they are to determining the value-driven components of SDM (objectives and actions).

Models used in SDM typically incorporate relationships between management actions and either (1) the vital rates that determine state variable dynamics (e.g., Fig. 2.3) or, less frequently, (2) the state variables themselves. These relationships may include ecological thresholds (Fig. 2.3). In the case study of Denali golden eagles (Martin et al. 2011; Eaton et al., Chap. 5), the management action (closure of a nesting site to hikers) is believed to increase the probability of a site making the transition from any state to the desired state of "occupied." However, scientists and managers are uncertain about the importance of disturbance to occupancy by eagles at a site. For this reason, several competing models are considered in the decisions for the Denali golden eagles. The example presented by Eaton et al. (Chap. 5) posits four hypotheses regarding the impact of disturbance and the availability of a particular prey species on eagle occupancy dynamics. Competing models differ in the hypothesized effects of management and prey level on parameters governing occupancy and include one model that incorporates an ecological threshold for prey abundance and another that assumes no effect of prey level or disturbance (and therefore of site closure to hikers) on golden eagle occupancy.

In order to incorporate this uncertainty (four models reflecting very different hypotheses about the effects of management) into the decision process, we must specify the relative influence of each model on the decision. Relative influence should be determined by the relative degree of faith we have in the predictive abilities of the models. We can specify the influence of each model on the decision using model "weights" or "credibility measures." These weights lie in the interval [0, 1] and sum to one for the members of the model set. In our Denali case with four models, for example, we might begin by assigning a weight of one fourths to each model (e.g., if we had no prior information as to which models were better predictors). These weights would indicate that we have equal faith (or equal uncertainty) in each model in the set. There are multiple reasonable ways to determine initial model weights if some prior information exists, including analysis of historical data and expert opinion. In recurrent decision problems, the ability to monitor effects of management actions provides an opportunity to learn. For recurrent decisions, a formal approach can be used to update model weights over time based on their relative predictive abilities as revealed through monitoring (see ARM).

Monitoring Program

Monitoring is important for informed decision making and SDM in providing estimates of system state for making state-dependent decisions. Many decisions will be state dependent as actions are likely to be very different depending on whether the system is judged as being near objectives (e.g., near a utility threshold) or far from them. In addition to state dependence of decisions, monitoring data are also used to assess the success of management. In the case of recurrent decisions, monitoring serves two additional roles: (1) providing the ability to learn by comparing model-based predictions with estimates of system state and related variables and (2) providing a means of obtaining updated estimates of key model parameters for periodic model revisions (Yoccoz et al. 2001; Nichols and Williams 2006; Lyons et al. 2008).

We note that these explicit roles of monitoring data in SDM suggest development of a monitoring program tailored as a specific component of SDM. Omnibus monitoring programs (developed to be generally useful, but not tailored to a specific purpose) are frequently claimed to be useful for informing management, but in reality they usually are inadequate or at least suboptimal for use in SDM (Nichols and Williams 2006). Monitoring is usually based on survey methods that yield some sort of count (of individual animals, of species, of sites occupied by a species, etc.). Good monitoring programs deal with two important sources of variation in such counts, geographic variation and detectability (Yoccoz et al. 2001; Williams et al. 2002). Geographic variation concerns the spatial variation found in most state variables and the frequent inability to conduct counts over the entire area of interest. Dealing with geographic variation requires selection of sites at which counts are conducted, in such a way as to provide inference about sites not selected (e.g., Thompson 2002). Detectability refers to the fact that even in sites where we do conduct our counts, we virtually never detect all individual animals (or species or occupied sites) that are actually present. This source of variation requires that we estimate the probability of detection in order to use count data for inference about the actual state variable(s) of interest, and a variety of methods has been developed for this purpose (Seber 1982; Williams et al. 2002; Borchers et al. 2003).

Solution Algorithm

The components described above, objectives, potential actions, models, and monitoring, provide the information needed to make an informed decision. However, taking this information and using it to develop an optimal, or even good, decision is frequently a nontrivial task. Often in natural resource management, a manager will examine available information and use common sense, intuition, or some other kind of thought process to decide on what action to take. However, a variety of optimization methods can be used to determine optimal decisions for well-defined problems in natural resource management (Walters and Hilborn 1978; Williams 1982, 1989, 1996, 2009; Williams et al. 2002; Williams and Nichols Chap. 4). The advantage of optimization approaches is that they yield the best possible decision recommendations with respect to the other SDM components. Optimization approaches result in policy or decision matrices (Fig. 2.4) that specify the optimal action for each possible system state (for each combination of system state variables). As noted above, decision thresholds represent locations in state space at which a change in state leads to a change in the optimal action. Objective solution algorithms (such as optimization algorithms) usually produce unambiguous policy matrices, reinforcing a previous point that decision thresholds are derived from the other components of the SDM process, including utility thresholds that are incorporated into objectives and any ecological thresholds that may be found in the system models (Fig. 2.5).

Adaptive Resource Management (ARM)

SDM is a general approach that can be used for virtually any kind of decision problem. Many problems in natural resource management entail recurrent decisions, in the sense that management decisions for a system are made at various points over time, as with annual decisions about harvest regulations or habitat management, for example. Because of the need to deal adequately with system dynamics, solution (e.g., optimization) algorithms for recurrent problems can be more difficult than those developed for a single time step. Specifically, the optimization must account for the fact that a decision this year influences the state of the system at the time of next year's decision. So a decision this year will influence the decisions that are available (and wise) next year. Thus, optimization based on a single time step can result in suboptimal decision policies, and the optimization algorithm must deal with the entire sequence of decisions for the time horizon of the process. Stochastic dynamic programming (Bellman 1957; Lubow 1995; Williams et al. 2002) is a powerful approach to optimization when dealing with recurrent decision problems.

Many (most) decision problems in natural resource management are characterized by substantial uncertainty. *Environmental variation* and resulting variation in system dynamics are well-known sources of uncertainty to all ecologists and wildlife managers. *Partial observability*, the inability, to observe nature directly, is also well known to those who study natural systems as we must almost always rely on inference methods that include sampling variation or error of estimation. *Partial controllability* refers to the indirect and/or imprecise application of management actions, as when our actions dictate hunting regulations rather than the precise rate of hunting mortality to be imposed on a managed population. *Structural uncertainty* refers to our typically inadequate understanding of managed systems and how they respond to management (i.e., uncertainty about system dynamics). For example, we may wish to incorporate in the decision process multiple hypotheses about system response to management. These forms of uncertainty constrain most problems in natural resource management, limiting the effectiveness of management to varying degrees (Williams 1997).

ARM (Holling 1978; Walters 1986; Williams et al. 2002; Williams et al. 2007) was developed for use with recurrent decision problems characterized by uncertainty. In addition to producing difficult optimization problems, recurrent decisions provide an opportunity to learn and to reduce structural uncertainty. Specifically, uncertainty is reduced by comparing predictions (from models) against observations (from monitoring) of system response. This reduction in uncertainty and corresponding increase in understanding are then used in adaptive management to increase the effectiveness of management over time. To summarize, ARM was developed for recurrent decision problems characterized by uncertainty. Efforts to simultaneously manage in the present and reduce uncertainty for better management in the future are definitive of ARM. The adaptive management process includes two phases, a deliberative phase and an iterative phase.

Deliberative Phase

The deliberative or "setup" phase of adaptive management (Williams et al. 2007) entails developing and assembling all of the SDM components. The development of a clear objective statement and the decision about what management alternatives to consider require input from all relevant stakeholders. One of the most common factors underlying failure of decision processes is stakeholder groups that do not believe they have had adequate input to the process. Even reasonable objectives will be criticized if stakeholder groups perceive that their input has not been solicited or has been ignored. Stakeholder involvement will frequently require joint meetings, and facilitation is sometimes useful. It is very useful to have some meeting participants who are accustomed to developing precise objective statements from general opinions and value statements. Regardless of the exact approach used to develop objectives and select potential management actions, these two SDM components essentially drive the entire decision process, and their importance should not be underestimated. Utility thresholds are frequently used in the development of objectives, especially as constraints in objective functions that include competing objectives.

The deliberative phase also requires development of initial models of system dynamics and response to management actions. Model development is driven by the selection of objectives and potential management actions, as model output must minimally include the response variables that are relevant to objectives (and that are thus used to value different outcomes) and provide predictions about responses of key system variables to the different management actions. Uncertainty about system response can be incorporated using multiple discrete models or by including a very general model with uncertainty characterizing a key parameter. Ecological thresholds may be included in system models if they are thought to characterize system dynamics and responses.

A monitoring program should also be established during the deliberative phase, and the characteristics of the program should be driven by the other decision process components, objectives, available actions, and models. The monitoring must provide estimates of key variables that reflect system state as such estimates are needed for state-dependent decisions (i.e., for establishment of decision thresholds). The monitoring must provide estimates of variables relevant to objectives, so that management success can be judged. Because adaptive management involves recurrent decisions, the monitoring program must provide estimates of variables and rate parameters that can be used to assess model adequacy and, periodically, to update model parameters.

A solution algorithm or approach should be identified as well. There must be some method of integrating and using the other SDM components to develop a management recommendation (select the "best" action). We have emphasized optimization approaches (Williams and Nichols, Chap. 4) as these are readily defended, but other approaches may be used as well. For example, a common approach involves simulation-based projections of consequences of different sequences of management actions, an approach that yields the best set of actions among those considered (e.g., McGowan et al. 2011). Sometimes, the decision is simply made by individuals without the benefit of computations of any sort. Although this latter approach is sometimes difficult to defend, it is commonly used. The adaptive management process requires some means of selecting the appropriate action based on the decision process components, but that approach does not have to involve optimization.

Iterative Phase

The iterative phase of adaptive management uses the SDM components assembled during the initial deliberative phase to make management decisions. The decisions involve selection of a management action from those available, and the periodicity is dictated by the decision process. Some decisions (e.g., establishment of waterfowl-hunting regulations) are made annually whereas other decisions may involve longer time periods and/or irregular intervals between decisions. The decision itself is obtained using the selected solution algorithm in conjunction with the specified objectives, the available actions, and the current state of knowledge about the system. That knowledge includes the system models and their associated credibility measures, as well as the current state of the system as estimated via the monitoring program.

Once the decision has been made, the selected action is applied to the system. The decision is based on the predicted system response to the different actions, as indicated by the different models. The action combines with relevant environmental variables to drive the system state to a new position, which is then identified by the monitoring program. Each system model also makes a prediction about system state following application of the management action. This comparison of predicted and estimated system state leads naturally to the updating of model weights or credibility measures, with increased weights for models that predict well and decreased weights for models that predict poorly.

Specifically, this updating can be accomplished using Bayes' Theorem (e.g., Williams et al. 2002)

$$p_i(t+1) = \frac{p_i(t) \times P_i(x_{t+1}|x_t, d_t)}{\sum_{i=1}^{n} p_i(t) \times P_i(x_{t+1}|x_t, d_t)} \tag{2.1}$$

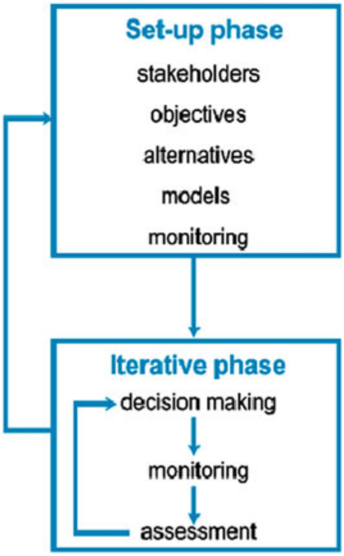
where $p_i(t)$ is the credibility measure (weight) for model i at time t, n is the number of models in the model set, and $P_i(x_{t+1}|x_t, d_t)$ is the probability of the observed system state at $t+1$ under model i, given that the system was in state x_t at time t and that decision d_t was implemented. $P_i(x_{t+1}|x_t, d_t)$ can be computed based on the monitoring data, for example, using standard likelihood-based models (Nichols 2001; Williams et al. 2002). Updating is thus a function of the model weight or prior probability at time t, reflecting knowledge accumulated until t, and the new information about how well the model predicted the most recent state transition between t and $t+1$. These updated probabilities then become the new model weights (or new priors) for the next decision and set of predictions (Kendall 2001; Nichols 2001; Williams et al. 2002).

At the next decision point, the above process is repeated, with some components remaining unchanged, specifically the objectives, available actions, models, and solution algorithm. However, knowledge of the system and its dynamics is updated as the new decision utilizes the current estimate of system state and the updated model weights, thus emphasizing in the decision process those models that have performed best over the accumulated history of the decision process. This use of multiple models with associated weights that evolve through time provides a formal approach to learning and is definitive of adaptive management (Walters 1986; Williams et al. 2002; Williams et al. 2007). Provided that reasonable models have been included in the model set, this iterative process should lead to the identification (high model weights) of models that provide good predictions. Thus, the adaptive process provides decision thresholds at each decision point that reflect the current state of knowledge about system response to management actions. The process leads to improved knowledge of the ecological system and its response to management, including any ecological thresholds that characterize system behavior.

Integration of Phases and Double-Loop Learning

The usual progression of an adaptive management project is to begin with the obligatory deliberative phase and to then implement the iterative phase. The deliberative phase produces the needed decision components, and the iterative phase then uses them to produce informed decisions at each decision point. The term "double-loop learning" (Lee 1993, p. 148; Williams et al. 2007) has been used to describe the process of revisiting the components of the initial deliberative (setup) phase, based

Fig. 2.6 Schematic diagram of double-loop learning in adaptive management. Within the setup phase, utility thresholds may occur in *objectives,* and ecological thresholds may occur in *models*. Within the iterative phase, decision thresholds will typically be used in the *decision making* step. (From Williams et al. 2007)



on experience with the process. For example, periodic input from stakeholders may indicate that objectives themselves should be changed and/or management actions should be modified or expanded (Runge et al. 2006). If none of the models in the model set provides consistently good predictions (e.g., as indicated by model weights that fluctuate, but do not accumulate for one or two models), then the models themselves should be revisited. Frequently, examination of the directions of differences between predictions and estimates of state variables may offer clues to the modification of models. When monitoring programs provide imprecise estimates or otherwise weak inferences about relevant variables or parameters, then these programs should be revised to correct these deficiencies. Finally, computing research may lead to improved solution algorithms that merit consideration and possible use.

Any of the above reasons provides a motivation to move out of the iterative phase and back into the deliberative phase of adaptive management (Fig. 2.6). Such movement typically occurs at a time scale that is longer than that of the iterative decision process phase. Nevertheless, such double-loop learning provides an important mechanism for learning and adaptation that extends beyond the evolution of model weights to every component of the decision process, including stakeholder views and institutional changes.

Discussion

This chapter has focused on the threshold concept as relevant to management and conservation. Certainly, much has been written about ecological thresholds, and we have contributed little to this discussion and literature. Rather we have tried to draw distinctions among three kinds of thresholds relevant to conservation, to clarify their origins (how does each threshold arise), and to describe the specific role of each in decision processes. The basic distinctions and origins are depicted in Fig. 2.5. Ecological thresholds, regardless of their detailed definitions, are relevant to decisions as components of ecological models that are used to predict system responses to management actions. Ecological thresholds arise from our attempts to provide simplified descriptions of natural systems and reflect our knowledge of such systems and their behaviors. Utility thresholds are included in statements of process objectives and are especially useful when objectives include multiple, competing objectives (e.g., maximally exploit resource x while maintaining resource y above some minimum level, the utility threshold). Utility thresholds arise not only from our understanding of managed systems, but also from our judgments about reasonable goals of system management. Utility thresholds are thus based on human values, and their development requires input from all relevant stakeholders. Decision thresholds are then derived from ecological and utility thresholds, in the sense that they are determined by management objectives, available actions, system models, and the decision solution algorithm.

Because these distinctions and definitions are imbedded within a management context, we described one approach to informed management, structured decision making (SDM). SDM is perhaps not the only approach to informed management, but it is logical, conceptually simple, and thus worthy of description and emphasis. Adaptive resource management (ARM) was then described as a form of SDM applied to recurrent decisions in the face of uncertainty. In particular, ARM provides a mechanism for learning and thus reducing uncertainty for the purpose of better management in the future. Our definitions and distinctions for types of thresholds are consistent with our descriptions of SDM and ARM decision processes and thus provide a coherent framework for viewing thresholds in the context of conservation and management. The management problem described in Eaton et al. (Chap. 5) is intended to illustrate these concepts of thresholds and informed decision processes. It is our hope that these chapters will promote use of SDM and ARM processes as logical ways to approach serious conservation decisions.

Acknowledgments We thank Bill Kendall, Bruce Lubow, Mike Runge, and Ken Williams for sharing ideas about decision making and thresholds. The manuscript benefitted from constructive comments by Glenn Guntenspergen.

References

Bellman, R. 1957. *Dynamic programming*. Princeton: Princeton University Press.

Bennetts, R. E., J. E. Gross, K. Cahill, C. McIntyre, B. B. Bingham, A. Hubbard, L. Cameron, and S. L. Carter. 2007. Linking monitoring to management and planning: Assessment points as a generalized approach. *George Wright Forum* 24:59–77.

Benton, T. G. 2003. Understanding the ecology of extinction: Are we close to the critical threshold? *Annales Zoologici Fennici* 40:71–80.

Bestelmeyer, B. T. 2006. Threshold concepts and their use in rangeland management and restoration: The good, the bad, and the insidious. *Restoration Ecology* 14:325–329.

Bodin, P., and B. L. B. Wiman. 2007. The usefulness of stability concepts in forest management when coping with increasing climate uncertainties. *Forest Ecology and Management* 242:541–552.

Borchers, D. L., S. T. Buckland, and W. Zucchini. 2003. *Estimating animal abundance*. New York: Springer-Verlag.

Brown, J. H., T. J. Valone, and C. G. Curtin. 1997. Reorganization of an arid ecosystem in response to recent climate change. *Proceedings of the National Academy of Sciences of the United States of America* 94:9729–9733.

Burgman, M. 2005. *Risks and decisions for conservation and environmental management*. Cambridge: Cambridge University Press.

Caughley, G. 1994. Directions in conservation biology. *The Journal of Animal Ecology* 63:215–244.

Clemen, R. T., and T. Reilly. 2001. *Making hard decisions with decision tools*. Pacific Grove: Duxbury.

Conroy, M. J., and C. T. Moore. 2001. Simulation models and optimal decision making in natural resource management. In *Modeling in natural resource management: valid development, interpretation and application*, ed. T. M. Shenk and A. B. Franklin, 91–104. Washington, D.C.: Island Press.

Conroy, M. J., C. R. Allen, J. T. Peterson, L. Pritchard, and C. T. Moore. 2003. Landscape change in the southern Piedmont: Challenges, solutions, and uncertainty across scales. *Conservation Ecology* 8:3.

Dorazio, R. M., and F. A. Johnson. 2003. Bayesian inference and decision theory—A framework for decision making in natural resource management. *Ecological Applications* 13:556–563.

Fahrig, L. 2001. How much habitat is enough? *Biological Conservation* 100:65–74.

Field, S. A., A. J. Tyre, N. Jonzen, J. R. Rhodes, and H. P. Possingham. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters* 7:669–675.

Fonnesbeck, C. J. 2005. Solving dynamic wildlife resource optimization problems using reinforcement learning. *Natural Resource Modeling* 18:1–39.

Groffman, P., J. Baron, T. Blett, A. Gold, I. Goodman, L. Gunderson, B. Levinson, M. Palmer, H. Paerl, G. Peterson, N. Poff, D. Rejeski, J. Reynolds, M. Turner, K. Weathers, and J. Wiens. 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? *Ecosystems* 9:1–13.

Gunderson, L. H. 2000. Ecological resilience—in theory and application. *Annual Review of Ecology and Systematics* 31:425–439.

Halpern, B. S., H. M. Regan, H. P. Possingham, and M. A. McCarthy. 2006. Accounting for uncertainty in marine reserve design. *Ecology Letters* 9:2–11.

Holling, C. S. 1973. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics* 4:1–23.

Holling, C. S. 1978. *Adaptive environmental assessment and management*. New York: Wiley.

Huggett, A. J. 2005. The concept and utility of 'ecological thresholds' in biodiversity conservation. *Biological Conservation* 124:301–310.

Intriligator, M. D. 1971. *Mathematical optimization and economic theory*. Englewood Cliffs: Prentice-Hall.

Johnson, F. A., C. T. Moore, W. L. Kendall, J. A. Dubovsky, D. F. Caithamer, J. R. Kelley, and B. K. Williams. 1997. Uncertainty and the management of mallard harvests. *Journal of Wildlife Management* 61:202–216.

Kendall, W. L. 2001. Using models to facilitate complex decisions. In *Modeling in natural resource management: valid development, interpretation and application*, ed. T. M. Shenk and A. B. Franklin, 147–170. Washington, D.C.: Island Press.

Lande, R. 1987. Extinction thresholds in demographic-models of territorial populations. *American Naturalist* 130:624–635.

Lee, K. N. 1993. *Compass and gyroscope: Integrating science and politics for the environment*. Washington, D.C.: Island Press.

Lubow, B. C. 1995. SDP: Generalized software for solving stochastic dynamic optimization problems. *Wildlife Society Bulletin* 23:738–742.

Lyons, J. E., M. C. Runge, H. P. Laskowski, and W. L. Kendall. 2008. Monitoring in the context of structured decision-making and adaptive management. *Journal of Wildlife Management* 72:1683–1692.

Martin, J., M. C. Runge, J. D. Nichols, B. C. Lubow, and W. L. Kendall. 2009a. Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications* 19:1079–1090.

Martin, J., C. L. McIntyre, J. E. Hines, J. D. Nichols, J. A. Schmutz, and M. C. MacCluskie. 2009b. Dynamic multistate site occupancy models to evaluate hypotheses relevant to conservation of Golden Eagles in Denali National Park, Alaska. *Biological Conservation* 142:2726–2731.

Martin, J., P. L. Fackler, J. D. Nichols, M. C. Runge, C. McIntyre, B. L. Lubow, M. G. McCluskie, and J. A. Schmutz. 2011. An adaptive management framework for optimal control of recreational activities in Denali National Park. *Conservation Biology* 25:316–323.

Maxwell, D., and S. Jennings. 2005. Power of monitoring programmes to detect decline and recovery of rare and vulnerable fish. *Journal of Applied Ecology* 42:25–37.

McCarthy, M. A., and H. P. Possingham. 2007. Active adaptive management for conservation. *Conservation Biology* 21:956–963.

McGowan, C., D. R. Smith, J. A. Sweka, J. Martin, J. D. Nichols, R. Wong, J. E. Lyons, L. J. Niles, K. Kalasz, J. Brust, M. Klopfer, and B. Spear. 2011. Multi-species modeling for adaptive management of horseshoe crabs and red knots in the Delaware Bay. *Natural Resource Modeling* 24:117–156.

Miranda, M. J., and P. L. Fackler. 2002. *Applied computational economics and finance*. Cambridge: MIT Press.

Nichols, J. D. 2001. Using models in the conduct of science and management of natural resources. In *Modeling in natural resource management: development, interpretation and application*, ed T. M. Shenk and A. B. Franklin, 11–34. Washington, D.C.: Island Press.

Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. *Trends in Ecology & Evolution* 21:668–673.

Pascual, M., and F. Guichard. 2005. Criticality and disturbance in spatial ecological systems. *Trends in Ecology & Evolution* 20:88–95.

Runge, M. C., F. A. Johnson, M. G. Anderson, M. D. Koneff, E. T. Reed, and S. E. Mott. 2006. The need for coherence between waterfowl harvest and habitat management. *Wildlife Society Bulletin* 34:1231–1237.

Seber, G. A. F. 1982. *The estimation of animal abundance and related parameters*. New York: MacMillian Press.

Thompson, R. L. 2002. *Sampling*. New York: Wiley.

Walters, C. J., and R. Hilborn. 1978. Ecological optimization and adaptive management. *Annual Review of Ecology and Systematics* 9:157–188.

Walters, C. J. 1986. *Adaptive management of renewable resources*. New York: Macmillan.

Williams, B. K. 1982. Optimal stochastic control in natural resource management – framework and examples. *Ecological Modelling* 16:275–297.

Williams, B. K. 1989. Review of dynamic optimization methods in renewable natural resource management. *Natural Resource Modeling* 3:137–216.

Williams, B. K. 1996. Adaptive optimization of renewable natural resources: Solution algorithms and a computer program. *Ecological Modelling* 93:101–111.

Williams, B. K. 1997. Approaches to the management of waterfowl under uncertainty. *Wildlife Society Bulletin* 25:714–720.

Williams, B. K., and J. D. Nichols. In press. Optimization in natural resources conservation. In *Thresholds for conservation*, ed. G. Guntherspergen and P. Geissler: Wiley, New York.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. *Analysis and management of animal populations: Modeling, estimation, and decision making*. San Diego: Academic Press.

Williams, B. K., R. C. Szaro, and C. D. Shapiro. 2007. *Adaptive management: The U.S. Department of the Interior technical guide*. Washington, D.C.: U.S. Department of the Interior.

Williams, B. K. 2009. Markov decision processes in natural resources management: Observability and uncertainty. *Ecological Modelling* 220:830–840.

Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology & Evolution* 16:446–453.

Chapter 3
Identifying Objectives and Alternative Actions to Frame a Decision Problem

Michael C. Runge and Terry Walshe

Abstract In this chapter, we discuss the role of objectives and alternative actions in framing a natural resource management decision problem, with particular attention to thresholds. We outline a number of considerations in developing objectives and measurable attributes, including when *utility thresholds* may be needed to express the decision-makers' values. We also discuss the development of a set of alternative actions, and how these might give rise to *decision thresholds*, particularly when the predictive models contain *ecological thresholds*. Framing of a decision problem plays a central role in decision analysis because it helps determine the needs for a predictive ecological model, the type of solution method required, and the value and structure of a monitoring system.

Keywords Utility threshold · Decision threshold · Ecological threshold · Decision analysis · Means objectives

Introduction

Collectively, the chapters in this volume discuss the concept of thresholds in the context of decision-making for natural resources. Nichols et al. (Chap. 2) lay out the foundations of this approach, defining several types of thresholds, and describing a structured approach to decision-making. One of the central tenets of structured decision-making is values-focused thinking (Keeney 1996), the recognition that all decisions are ultimately expressions of the values of the decision-maker, and thus, articulation of the objectives should have a primary role in framing a decision. The nature of the decision itself is best captured by consideration of the alternative choices

M. C. Runge (✉)
USGS Patuxent Wildlife Research Center, 12100 Beech Forest Road,
Laurel, MD 20708, USA
e-mail: mrunge@usgs.gov

T. Walshe
Australian Centre of Excellence for Risk Analysis, School of Botany,
The University of Melbourne,
Parkville, VIC 3010, Australia

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_3,

the decision-maker faces; thus, development of the alternatives also plays a primary role in decision analysis. Taken together, these two elements—articulating the objectives and developing the alternatives—frame a decision problem.

In this chapter, we discuss the roles of objectives and alternative actions in a decision analysis, with particular attention to thresholds. We outline a number of considerations in developing objectives and measurable attributes, including when *utility thresholds* may be needed to express the decision-makers' values. We also discuss the development of a set of alternative actions, and how these might give rise to *decision thresholds*, particularly when the predictive models contain *ecological thresholds*. We argue that decision framing plays a central role in decision analysis, leads to specification of the needs for a predictive model, identifies the solution method, and justifies the structure and value of a monitoring program.

Objectives

Biological interactions are complex and management problems are typically ill-structured. This complexity may not readily lend itself to formulation under traditional optimization techniques, or simple intuitive problem solving. Most problems in natural resource management involve trade-offs among multiple objectives. For example, the management of fisheries and forests routinely deals with tensions arising from simultaneous objectives of maximizing yields and maximizing conservation outcomes. Those with a stake in the management of a resource inevitably advocate a policy position or alternative that implicitly assigns greater (or exclusive) weight to the objective that aligns with their interests. It is not surprising that stakeholders (and management agencies) occasionally become entrenched in advocating a single policy or alternative that is diametrically opposed by other stakeholders (or agencies). *Alternatives*-focused thinking is prey to anchoring, myopia, and entrenched stakeholder conflict. To insulate against the poor decisions associated with entrenched conflict, Keeney (1996) emphasizes the identification of objectives as the basis for *value*-focused thinking.

Types of Objectives

Keeney (2007) describes four types of objectives:

- *Strategic objectives:* objectives influenced by all of the decisions made over time by the organization or individual facing the decision at hand.
- *Fundamental objectives:* the objectives used to describe the consequences that essentially define the basic reasons for being interested in the decision.
- *Means objectives:* objectives that are important only for their influence on achievement of the fundamental (ends) objectives.
- *Process objectives:* objectives concerning how the decision is made rather than what decision is made.

Strategic objectives assist in defining the frame or context of a decision, but they are generally too vague to be used directly. Effective decision support deals with fundamental objectives. A key step is to disentangle means objectives from fundamental objectives. Process objectives govern the methods by which the decision will be made, who will be included, and how the decision will be documented and communicated; often the process objectives are influenced by agency or institutional policy. The achievement of process objectives rests substantially on the analyst's capacity to assist decision-makers and stakeholders structure the decision problem appropriately.

In initial brainstorming discussions with decision-makers, the objectives articulated will often comprise a confusing mix of means objectives, fundamental objectives, strategic and process objectives, together with associated attributes, constraints, targets, or alternatives. Organizations with a strong emphasis on science or evidence-based decision-making are commonly distracted by the drivers of system dynamics rather than fundamental objectives and key value-based trade-offs. For example, let us say a government agency is responsible for managing a commercial fishery and is considering regulating the kind of fishing gear the commercial fleet can use in its operations. The agency generates the list of objectives below in consultation with in-house scientists and key stakeholders.

The following list contains substantial redundancy, which if left untreated will result in double-counting and other errors of aggregation:

- Minimize environmental impact
- Minimize bycatch
- Maximize conservation
- Maintain nutrient dynamics
- Maximize turtle population
- Change fishing gear
- Maximize yield
- Maximize profitability
- Maintain fish stock
- Maximize public acceptability

Minimizing environmental impact is a strategic objective that offers little insight to the decision at hand. *Maximizing conservation* might be the fundamental objective to which the strategic objective alludes, at least in this particular decision context. *Minimizing bycatch* is a means of achieving the fundamental end of *maximizing conservation*. The *maintenance of nutrient dynamics* could have been suggested by the organization's scientists. It may be a good indicator of system function and could usefully be incorporated into predictive models of consequences under various alternatives, but it is not a fundamental objective itself. When we ask *why* is it that nutrient dynamics are important, the answer may be for conservation values and the maintenance of the fishery's stock levels. So again, nutrient dynamics might be a means objective that contributes to the fundamental ends of conservation and production. *Maximizing the turtle population* may be a socially relevant and measurable attribute for describing conservation outcomes. *Changing fishing gear* is an alternative rather than an objective. *Maximizing yield* is a means towards the fundamental objective of

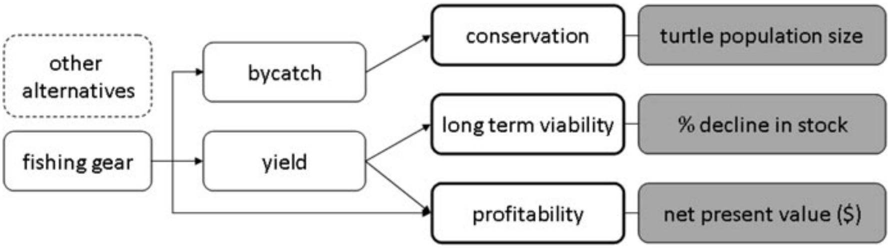


Fig. 3.1 Means-ends diagram for management of a hypothetical commercial fishery. *Nodes with bold outlines* indicate fundamental (ends) objectives. *Grey nodes* are notional attributes that can be used to describe the performance of one or more alternatives against fundamental objectives. Ends are influenced by the means objectives of minimizing bycatch and maximizing yield. The type of fishing gear used by the industry can be seen as an alternative that affects all objectives. The diagram may be used to stimulate additional creative alternatives

maximizing the fishery's profitability. Profit will also be influenced by a change in fishing gear (and probably any other alternatives), depending on who bears the costs of any changes. *The maintenance of fish stock* is a constraint which can be included as an objective and traded off against others, or left as a strict requirement whereby any alternative that fails to satisfy the constraint is omitted from detailed consideration. *Maximizing public acceptability* is a process objective. Sound formulation of the decision problem via the value-focused thinking outlined here will contribute significantly toward its achievement.

A good set of fundamental objectives will be (Keeney 2007):

- Complete—all of the important consequences of alternatives in a decision context can be adequately described in terms of the set of fundamental objectives.
- Non-redundant—the fundamental objectives should not include overlapping concerns.
- Concise—the number of objectives should be minimal.
- Specific—each objective should be specific enough so that consequences of concern are clear and attributes can readily be selected or defined.
- Understandable—any interested individual knows what is meant by the objectives.

Graphs are an effective way of capturing and organizing the elements of an ill-structured decision problem (Montibeller et al. 2008). Influence diagrams can be used to communicate the relationship between means objectives and fundamental objectives (Keeney 1996, Failing et al. 2007). Figure 3.1 is an illustrative example using our hypothetical commercial fishery.

Can Learning be a Fundamental Objective?

During the objective-generation phase, learning is often cited as a fundamental objective, especially in science-based organizations. In our experience, for applied decision-making settings, learning is best viewed as either a means objective or as a strategic objective. When scientists or decision-makers are asked why learning is an objective, the response is typically, “because it will help us manage better in the future.” Manage for what? Manage for some other fundamental objective. If that other fundamental objective is part of the decision at hand, then learning is a means to it. This distinction is important because it suggests that the decision-maker does not want to pursue learning for its own sake, but only insofar as it helps achieve the fundamental objective. Learning can be expensive, both in direct costs and in terms of lost opportunities. For learning to be warranted, these costs should be outweighed by improved long-term achievement of the fundamental objectives.

Sometimes, the fundamental objective whose achievement would be improved through learning is not part of the decision at hand, but part of some other decision. In this case, the decision-maker is recognizing that learning in one setting may help decision-making elsewhere. Thus, learning is a strategic objective for the organization. In this case, it might be appropriate to treat learning as if it were a fundamental objective for the decision at hand, but careful consideration of the weight given to this objective is needed.

Attributes

Attributes are used to judge the performance of alternatives against fundamental objectives. That is, an attribute is a measurable scale that reflects achievement (or expected achievement) of an objective. Ideally, attributes are measured on natural scales—scales that can be measured without subjective judgment and which correspond directly with the intent of the objective they represent. Often, however, other types of scales, subjective scales or proxy scales, need to be substituted for practical reasons. Clarity in fundamental objectives and their attributes focuses the elicitation of expert opinion, data capture, and detailed causal modeling used to estimate consequences under each alternative.

Desirable attributes are (Keeney and Gregory 2005):

- Unambiguous—a clear relationship exists between an objective and description of consequences under each alternative using the attribute.
- Comprehensive—the attribute levels cover the range of possible consequences for the corresponding objective under all alternatives, and value judgments implicit in the attribute are reasonable.
- Direct—the attribute levels directly describe the consequences of interest.
- Operational—in practice, information to describe consequences can be obtained and value trade-offs can reasonably be made.
- Understandable—consequences and value trade-offs made using the attribute can readily be understood and clearly communicated.

Using the example of our hypothetical fishery, some measurable attributes on natural scales for our fundamental objectives include (Fig. 3.1): sea turtle population size as an attribute for our conservation objective; decline in the fish stock as an attribute for

Table 3.1 Examples of measurable attributes for some fundamental objectives that might arise in natural resource management. The third column shows the type of scale the attribute is measured in

Fundamental objective	Measurable attribute	Type of scale
Maximize long-term viability of a population	Probability of extinction	Natural
Minimize regulatory burden on the public	Regulated area-years (product of the regulated area and the expect length of time it would be under regulation)	Natural
Maximize long-term consumptive harvest	Sustained annual yield	Natural
Maximize non-consumptive viewing opportunities	Bird population size	Proxy (population size does not exactly equate to viewing opportunities, but it is likely to be strongly correlated)
Minimize adverse public health consequences	Number of individual health concerns that result in a visit to a doctor	Proxy (this is not the only aspect of public health consequence, and it does not consider the severity of the condition, but it is likely to correlate with total public health concern)
Preserve cultural and religious heritage of a natural resource	Five-point scale, ranging from (1) complete loss of culturally important resources to (5) full long-term preservation of culturally important resources	Subjective (requires judgment of an expert panel)
Maximize legal defensibility of a rule-making	Three-point scale reflecting probability of rule being upheld in a legal challenge, with three categories: 0–25 %, 25–75 %, 75–100 %	Subjective (requires judgment of an expert panel of lawyers familiar with the setting)

the long-term viability of the fish stock; and net present value of long-term harvests as an attribute for profitability of the fishery. Note that other tempting measurable attributes, like the magnitude of the bycatch, do not reflect fundamental objectives. Some other examples of measurable attributes for fundamental objectives that might arise in natural resource management settings are shown in Table 3.1.

Single-Objective Utility

Attributes provide a scale on which to measure achievement of objectives, but how the decision-maker cares about the outcome may not be a linear function of the attribute. Rather, we can speak of the *utility* of the outcome to the decision-maker,

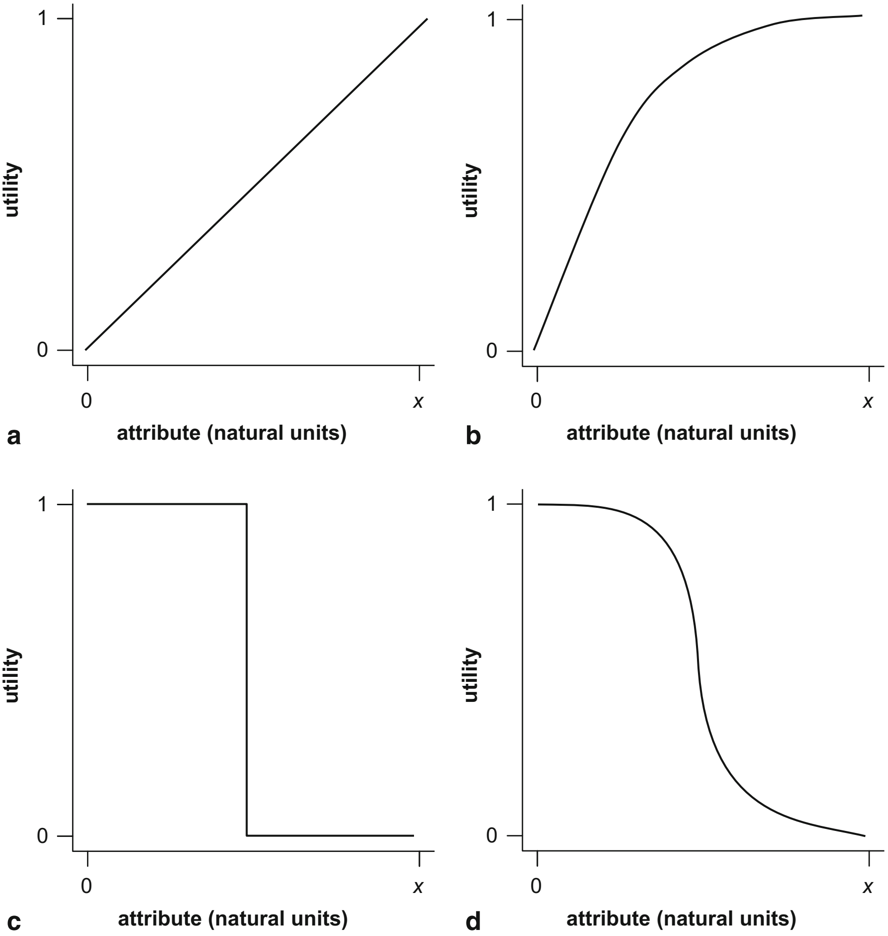


Fig. 3.2 Examples of single attribute utility functions: **a** linear; **b** logarithmic; **c** step; and **d** logistic. The step and logistic utility functions contain utility thresholds: small ranges on an attribute scale over which the utility changes abruptly

and this measure reflects the magnitude of well-being associated with an outcome on a particular attribute scale. (For decision theory purists, note that we are using the term utility to refer generically to both *value* and *utility*, and are not making the traditional distinction between the two. See von Neumann and Morgenstern 1944.)

A marginal utility function is a utility function for any single attribute in isolation. Figure 3.2 illustrates four of an infinite number of possible forms. For a linear utility function (Fig. 3.2a), the attribute scale is a direct expression of the degree to which the decision-maker cares about the outcome. The linear function may only be appropriate over limited ranges of a particular attribute. For larger ranges, the rate of utility increases commonly diminishes with increasing acquisition of the attribute of interest (assuming more of the attribute is desired), as in the logarithmic utility function (Fig. 3.2b). The step function (Fig. 3.2c) indicates maximal utility below a certain threshold and zero utility beyond that threshold (i.e., less of the attribute is desired). The dramatic change in utility described is unlikely to be a natural representation of any decision-maker's value judgments, but it might be appropriate for objectives that correspond to constraints (for example, fixed budget constraints or clear legal requirements). In our fisheries example, such a function might be appropriate where the management agency can tolerate a defined magnitude of decline in stock, but no more.

Utility thresholds occur when the utility changes abruptly over a small range in the measurable attribute. The step function (Fig. 3.2c) is an extreme case—here the utility changes instantaneously above and below a single threshold value. The logistic function (Fig. 3.2d) is slightly more tempered—the utility changes abruptly over a narrow range of values in the measurable attribute. But both of these functions act in a similar manner; they greatly favor outcomes below the threshold, and strongly discourage outcomes above the threshold.

Multiple-Objective Utility

As noted earlier, most problems in natural resource management involve multiple objectives. Even after understanding the marginal utility of each objective separately, the decision analyst needs to understand how to capture appropriately the values of a decision-maker (or multiple decision-makers) when combining those objectives. The field of multi-criteria decision analysis (MCDA) addresses these challenges. If marginal utilities can be considered preferentially independent (i.e., the utility of one attribute is unaffected by the amount of another attribute) then compensatory methods of aggregation (like weighted sums) are appropriate. Otherwise non-compensatory techniques should be used. Three common approaches are discussed here.

Single objective with multiple constraints. In this first approach, one of the objectives is primary, while the other objectives are treated as constraints. In essence, all of the alternatives that do not meet the constraints are eliminated, and the remaining alternatives are evaluated against the single objective. This implies a deep importance and non-negotiability to the constraints—the constraints are utility thresholds like Fig. 3.2c. One analytical method used to solve such problems is linear programming; this technique has been used extensively, for example, in forestry problems.

As an example, consider a forestry management decision. Suppose the fundamental objectives are to maximize stand yield (measured in board-feet), while using only existing staff and equipment. The latter two objectives serve as constraints: staff investment needs to remain below, say, 400 person-hours; and there are only six portable saws and two skidders available. Any alternatives that meet those constraints are admissible; the preferred alternative is the admissible alternative with the highest yield.

Non-compensatory multiple-objectives. In a more complicated scenario, the objectives may be non-compensatory (in that there are threshold utilities or preferential dependence among objectives), but it may not be as simple as using step functions to eliminate alternatives. In this case, two approaches are possible: development of a single composite utility function, or "outranking" MCDA methods. The former is discussed here; regarding the latter, the interested reader is referred to the MCDA literature (Vincke 1992, Stewart and Losa 2003).

A composite utility function seeks to combine mathematically the multiple objectives into a single expression that captures the decision-makers values. As an example, the utility function currently used to manage mid-continent mallards in North America is

$$\sum_{t=0}^{\infty} H_t U(N_{t+1}), \text{where } U(N_{t+1}) = \begin{cases} {}^{N_{t+1}}\!/_{8.8} & \text{if } N_{t+1} < 8.8\,\text{million} \\ 1 & \text{otherwise} \end{cases}.$$

This utility function actually combines three fundamental objectives: maximize annual harvest, ensure long-term persistence of the population, and keep the population size above the North American Waterfowl Management Plan goal of 8.8 million. It achieves the first objective by maximizing annual harvests (H_t). It achieves the second objective by maximizing the sum of those harvests over an infinite-time horizon; such a maximum can only occur if the population persists indefinitely. Finally, the third objective is achieved by devaluing the harvest in any year if the projected population size in the next year is below the goal. Note that the third objective is a threshold utility, but not a step function.

Compensatory multiple-objectives. If the objectives are all compensatory, that is, they can be traded off against each other, then a composite utility function can be readily formed. The most common method is a simple weighted summation of the marginal utilities. The composite utility function U over two or more relevant attributes, x_i, can be written as

$$U(x_1, x_2, ..., x_n) = \sum_{i=1}^{n} w_i v_i(x_i)$$

where w_i are the weights and v_i are marginal utility functions (Bedford and Cooke 2001). Where fundamental objectives satisfy the properties listed in the section "Types of objectives," there is a strong case for use of simple weighted summation.

In the simple weighted summation model for a decision-maker's value function over multiple objectives, the preference for one objective over another is described by the weights. Weights can be elicited using a variety of techniques (Hajkowicz et al. 2000), not all of which are credible. The weight assigned to any single fundamental objective needs to consider two elements; (1) the inherent importance of the objective, and (2) the range of the consequences estimated across all alternatives. A very common mistake in assigning weights is to ignore the range of consequences (Keeney 2002, Steele et al. 2009).

Alternative Actions

The alternatives from which a decision-maker may choose are, of course, central to any decision. One of the evolved tendencies of human decision-makers is the inclination to immediately jump to consideration of alternatives (Keeney 1996). But in a structured, values-focused decision-making process, alternatives are identified *after* the objectives have been articulated. There are two primary reasons for this: first, any decision is an expression of the decision-maker's values, so those values should retain primacy and drive the other aspects of the decision analysis; second, when alternatives are generated absent objectives, they tend to be a narrow set, anchoring typically on the status quo and perhaps a few other options. When the objectives are identified first, they can be used to craft a wide range of creative alternatives that seek to fulfill the objectives.

What is a Decision?

In generating alternative actions, it is important to recognize the nature of the decision. A decision is an "irrevocable allocation of resources… not a mental commitment to follow a course of action but rather the actual pursuit of the course of action" (Howard 1966). In this spirit, it is helpful to think about what true action underlies the decision. In natural resource management, actions may have direct or indirect effects. Direct effects result from many on-the-ground actions, like modification of habitat, movement of animals, captive breeding, and immunocontraception, among many others. But many actions, especially those taken by government agencies, have indirect effects through the responses of others: enactment of regulations, issuance of permits, and tax incentive schemes all work in this manner. Nevertheless, in both direct and indirect cases, a committed step is taken.

Natural resource management entities often conduct exercises in prioritization. For example, they will prioritize species or habitat types for conservation, land parcels for acquisition, or invasive plants for control. But while a prioritized list may precede a decision, it is not itself a decision, as it involves no irrevocable allocation of resources. The decision arises out of how that prioritized list is used. If a prioritized list of land parcels is used to select a subset of parcels for acquisition, then the decision concerns the set of land parcels to acquire, and the alternatives are all the possible sets of land parcels that could be acquired. A focus on what action will actually be taken helps to appropriately identify the true alternatives.

Is monitoring an action? Decision-makers will often identify learning or monitoring as an option, but it is important to understand what lies behind this suggestion. If monitoring is suggested as a preferred alternative without consideration of actions that influence the resource under management, then the decision setting is not fully framed, and the monitoring is serving as a delay in implementing any action. A full analysis should consider the option of taking action without learning first. In this way, the decision-maker can ask whether the costs of delay are offset by the expected benefits of learning. This question can be viewed as a linked decision

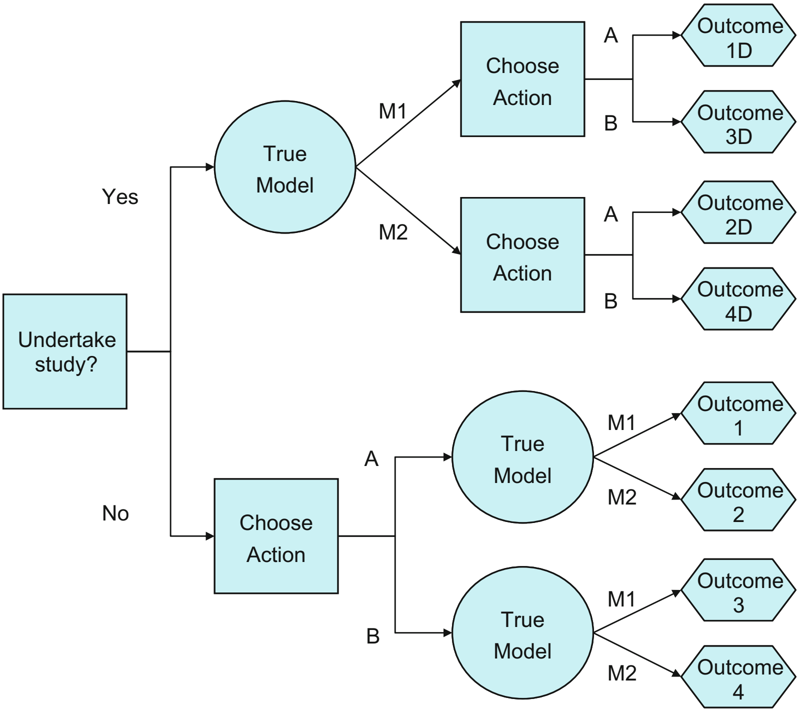


Fig. 3.3 Decision tree for evaluating the value of information. The outcomes are a function of the action taken (*A* or *B*) and the true state of the world (*M1* or *M2*); in addition, if a study is undertaken, there are costs of the delay (*D*), both in terms of direct costs of the study or monitoring, and indirect costs in lost opportunity. Whether or not to undertake the study or proceed directly to action depends on whether those costs are offset by the improved management that arises from the information the study will gather

(Fig. 3.3). If action is delayed so that a study or monitoring can be undertaken, then information about how the world works is acquired before the management action has to be chosen. All other things equal, this should improve the expected performance of the decision. But the direct costs of the study and the opportunity costs of delaying the decision may not be offset by the value of the information. The value of learning arises out of the interaction between the objectives (as measured by the outcomes, which may include utility thresholds), the ecological models (which may include ecological thresholds), and uncertainty. The point here is that "learning," as an alternative action, requires special treatment.

Types of Actions

Sets of alternative actions can take a number of forms. In the simplest case, the decision-maker is choosing from a *discrete set* of alternative actions. The set may be small (perhaps as small as two!) or large, but is composed of unique, distinct alternatives. For example, the manager of a grassland may be considering burning, mowing, or haying as alternative disturbance regimes to maintain the field in an early stage of succession; these alternatives are seen as distinct choices, with little opportunity for intergradations.

A special type of discrete set called *portfolios* arises when subactions are assembled into collections, and frequently are associated with prioritization problems. For example, a review panel deciding how to allocate research funding among a number of proposals is making a decision about which portfolio of proposals to fund. The alternative portfolios are all the possible combinations of proposals. It is important to note here that the commitment of resources is to the portfolio that is funded, so the alternative actions are all the possible portfolios, not the individual proposals themselves.

In other cases, the decision-maker is choosing from a *continuous set* of alternative actions, with a decision variable taking any value within a given range. For example, the harvest rate for a fishery can take on continuous values between zero and the intrinsic rate of increase for that population. Other decision variables that might take on continuous values include: the amount of herbicide to apply to a field, the length of time to continue a reintroduction program, the volume of water to release from a dam, the number of hectares of habitat to acquire through easements, and the amount of effort (person-hours) to expend in an eradication program.

Typically, sets of alternative actions are even more complex. They might be *combination sets* that include both discrete and continuous elements. For example, control of an invasive plant may include the discrete choice among types of herbicide, as well as the continuous choice among levels of herbicide to apply, perhaps even in combination with a discrete choice involving the season of application. *Strategy tables* are a technique for developing these complex alternatives; they involve identifying all the sub-elements of an alternative, and articulating a large (if not exhaustive) set of combinations of these sub-elements (Skinner 1999).

Decision Thresholds

Decision thresholds arise when the preferred alternative changes abruptly over a narrow range in the state variables. Thus, as discussed in Chap. 6 (Williams and Nichols), decision thresholds are identified at the optimization stage of a decision analysis. They arise, if at all, out of the interaction between objectives (possibly with utility thresholds), alternative actions, and models (possibly with ecological thresholds).

As described in Chap. 1 (Guntenspergen and Gross), ecological thresholds are situations where the system dynamics change abruptly over a small range of a state variable. These may or may not give rise to decision thresholds, and when they do, the ecological and decision thresholds may not correspond. For example, consider a particular ecological threshold—an Allee effect such that the reproductive rate of a population falls below replacement level when the population size falls below a particular value. This is a stark and important threshold because once a population

Table 3.2 Optimal regulatory strategy for mid-continent mallards for the 2010 hunting season (USFWS 2010)

Bpop (in millions)	Ponds (in millions)									
	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
≤4.5	C	C	C	C	C	C	C	C	C	C
4.75–5.75	R	R	R	R	R	R	R	R	R	R
6	R	R	R	R	R	R	R	R	M	M
6.25	R	R	R	R	R	R	M	M	M	L
6.5	R	R	R	R	M	M	M	L	L	L
6.75	R	R	R	M	L	L	L	L	L	L
7	R	M	M	M	L	L	L	L	L	L
7.25	M	L	L	L	L	L	L	L	L	L
7.5	L	L	L	L	L	L	L	L	L	L
≥7.75	L	L	L	L	L[a]	L	L	L	L	L

The two state variables are the breeding population size (Bpop) and the number of ponds in prairie Canada (ponds)

Regulatory packages: *C* closed, *R* restrictive, *M* moderate, *L* liberal

[a] The regulatory prescription for 2010

falls below that population size, it cannot naturally recover; over a small range of population size, the attraction point abruptly switches from the carrying capacity to zero. Suppose further that this population, which normally is well above the Allee threshold, is subject to harvesting. In a full optimization analysis, a decision threshold will arise—to cease hunting when the population size falls below some level. But the decision threshold will be considerably higher than the ecological threshold, to guard against the risk of inadvertently pushing the population below the tipping point.

Decision thresholds can arise from both discrete and continuous sets of actions. In the discrete case, a decision threshold is natural, because if the preferred alternative changes over some range of the state space, it necessarily must do so abruptly, even if there is no utility or ecological threshold. In the continuous case, decision thresholds can arise as a result of utility or ecological thresholds.

In the USA, duck-hunting regulations are set through a decision-analytical process, which includes a formal optimization step (Table 3.2). There are numerous decision thresholds. For example, if there are 3.5 million ponds in prairie Canada, the optimal regulatory package changes from restrictive to liberal over a very small range of mallard breeding population size (6.25–6.75 million). The decision thresholds in this strategy arise for a number of reasons. At a population size of 4.5 million mallards, there is a sharp decision threshold between a closed and restrictive hunting season; this decision threshold arises because it was imposed directly as a closure constraint. Thus, it corresponds to a utility threshold. Without this utility threshold, closed seasons would occur more often. There is another utility threshold at 8.8 million mallards (see the section "Multiple-objective utility"), but this does not have a direct effect in producing a decision threshold. The most prominent decision threshold—the diagonal "knife-edge" over which regulations change abruptly from restrictive to liberal—does not arise from either a utility or ecological threshold. Instead, it is a consequence of the optimization, and would occur even if the decision variable was continuous (e.g., harvest rate).

Decision Framing as an Iterative Process

Together, the specification of the objectives and the articulation of the alternatives frame the decision problem, and lead to the subsequent steps in a structured decision analysis. The framing helps determine what is needed in a predictive ecological model by specifying both the inputs (alternatives) and the outputs (measurable attributes for the objectives). It also determines the type of solution method required to identify the preferred alternative. Finally, because the objectives specify the scales on which performance is measured, and because the actions (and subsequent decision thresholds) determine the possible management solution, framing is central in identifying the value and structure of a monitoring system, both to evaluate achievement of objectives and to reduce uncertainty. Thus, decision framing plays a central role in decision analysis.

But framing is challenging. Decision-makers often cannot clearly articulate their objectives on the first pass, and need time wrestling with the decision problem to fully construct their objectives (Keeney 1996; Bond et al. 2008). Developing novel alternatives is often a creative process that requires a fresh perspective. Alternatives themselves often lead to identification of further objectives as decision-makers iteratively work through the problem and identify elements that better distinguish the overall performance of candidates. For example, management actions that entail undesirable side-effects can prompt the inclusion of an objective to minimize those side-effects.

Tremendous insight can arise from a prototype analysis. An initial decision framing leads to an initial prediction of consequences and analysis of preferred alternatives. If the prototype analysis is intuitively unappealing, an examination of the components of the analysis, especially the framing, can suggest improvements. Are there other objectives that were not captured? Is it possible to construct new alternatives from the insights gained from the prediction of consequences? Sometimes constraints arise in the analysis—a lack of data for the proposed model, or the absence of an analytical solution method for the problem as framed; these constraints can also lead to a reframing of the problem.

An iterated cycle of prototyping, sensitivity analysis, reflection, and revision is a healthy way to develop a decision analysis. It allows for feedback from the decision-maker at multiple points, and favors development of decision analyses that are only as complex as they need to be.

References

Bedford, T., and R. Cooke. 2001. *Probabilistic risk analysis. Foundations and methods*. Cambridge: Cambridge University Press.

Bond, S. D., K. A. Carlson, and R. L. Keeney. 2008. Generating objectives: Can decision makers articulate what they want? *Management Science* 54:56–70.

Failing, L., R. Gregory, and M. Harstone. 2007. Integrating science and local knowledge in environmental risk management: A decision-focused approach. *Ecological Economics* 64:47–60.

Hajkowicz, S. A., G. T. McDonald, and P. N. Smith. 2000. An evaluation of multiple objective decision support weighting techniques in natural resource management. *Journal of Environmental Planning and Management* 43:505–518.

Howard, R. A. 1966. Decision analysis: Applied decision theory. In *Proceedings of the Fourth International Conference on Operational Research,* ed. D. B. Hertz and J. Melese, 55–71. New York: Wiley.

Keeney, R. L. 1996. Value-focused thinking: Identifying decision opportunities and creating alternatives. *European Journal of Operational Research* 92:537–549.

Keeney, R. L. 2002. Common mistakes in making value trade-offs. *Operations Research* 50: 935–945.

Keeney, R. L. 2007. Developing objectives and attributes. In *Advances in decision analysis. From foundations to applications,* ed. W. Edwards, R. F. Miles Jr., and D. von Winterdfeldt. Cambridge: Cambridge University Press.

Keeney, R. L., and R. S. Gregory. 2005. Selecting attributes to measure achievement of objectives. *Operations Research* 53:1–11.

Montibeller, G., V. Belton, F. Ackermann, and L. Ensslin. 2008. Reasoning maps for decision aid: An integrated approach for problem-structuring and multi-criteria evaluation. *Journal of the Operational Research Society* 59:575–589.

Skinner, D. 1999. *Introduction to decision analysis*. 2nd ed. Gainesville: Probabilistic Publishing.

Steele, K., Y. Carmel, J. Cross, and C. Wilcox. 2009. Uses and misuses of multi-criteria decision analysis (MCDA) in environmental decision-making. *Risk Analysis* 29:26–33.

Stewart, T. J., and F. B. Losa. 2003. Towards reconciling outranking and value measurement practice. *European Journal of Operational Research* 145:645–659.

United States Fish and Wildlife Service (USFWS). 2010. *Adaptive harvest management: 2010 hunting season*. Washington, DC: U.S. Department of Interior.

Vincke, P. 1992. *Multicriteria decision-aid*. Chichester: Wiley.

von Neumann, J. and O. Morgenstern. 1944. *Theory of games and economic behavior*. Princeton: Princeton University Press.

Chapter 4
Optimization in Natural Resources Conservation

Byron K. Williams and James D. Nichols

Abstract The previous three chapters of this book have been devoted to specific components of informed decision processes: objectives, potential actions, model(s) predicting system change and response to potential actions, and monitoring to provide estimates of system status. The final component of an informed decision process is a solution algorithm, providing a means for deciding which potential action to take. Optimization algorithms provide an objective and transparent approach to select the action that will do the best job of meeting objectives. Static optimization provides a solution to decision problems that are not iterative, and we provide examples for one or more decision variables (variables that are components of potential actions). Many decision problems in natural resource management are best viewed as dynamic, in that they are iterative and require decisions that are repeated through time. In dynamic decision problems, decisions made at one point in time are expected to influence system state of the next time step, thus influencing the state-dependent decision at that time. For any specific decision, dynamic optimization algorithms must thus consider all subsequent time steps for the time horizon of the decision problem. In addition to being dynamic, most decision problems in natural resource management are characterized by substantial uncertainty, and dynamic optimization algorithms have been extended to deal with several sources of uncertainty. An important source is uncertainty about how the system responds to management actions, and we may develop multiple models to characterize this uncertainty. Adaptive dynamic optimization algorithms provide solutions that deal not only with objectives, but with the anticipated reduction in uncertainty that will characterize future decisions. The output of an optimization algorithm is frequently a graph or table of recommended actions for specific values of system state variables. *Decision thresholds* are thus defined by the optimization algorithm and are simply locations in state space where a small change in the value of a state variable produces a change in the optimal or recommended management action.

J. D. Nichols (✉)
Patuxent Wildlife Research Center, USGS, Laurel, MD 20708, USA
e-mail: jnichols@usgs.gov

B. K. Williams
Cooperative Research Units, USGS, Reston, VA 20192, USA

The Wildlife Society, 5410 Grosvenor Lane, Suite 200, Bethesda, MD, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_4,

Keywords Adaptive management · Conservation · Markov decision process · Optimization · Uncertainty

Introduction

Among other things, natural resources conservation is about informed decision-making. Conservation builds on our understanding of natural resource systems and the influence of drivers and stressors on them. In many instances it relies on tracking and monitoring programs, which provide the information needed to determine resource states and conduct investigations about patterns of change. It is (or should be) responsive to partners and stakeholders, who often have different perspectives and values about resource conditions and the decisions needed to achieve those conditions. In combination with information management and assessment, these features play an important role in natural resources conservation. Simply put, conservation is about making decisions in ways that incorporate understanding and information to achieve management goals and objectives (Holling 1978; Walters 1986; Williams et al. 2002, 2007; Allen et al. 2011).

In this chapter, we describe conservation in terms of smart decision-making, that is, choosing the best decision from a set of alternatives according to some measurement criterion. The idea is that decision-makers select a particular decision from a range of possible decisions, with the selection influencing the resources under consideration. The decision choice is guided by goals and objectives that incorporate the production of services sought by the decision-maker, and it (often) includes future resource conditions. The concept of goals and objectives is very general, and may include minimization of risk to a resource in the face of uncertainty just as readily as maximization of some ecosystem service. The intent is to make decisions that are optimal with respect to desired objectives, given the available options.

An important class of resource problems involves decision-making that accounts for the future consequences of decisions. A focus on the future in turn requires at least some understanding of the influence that decisions can have on resource dynamics, which is captured with models that describe that involve resource conditions. This is a natural framework for resource problems involving iterative decision-making through time, where decisions made at any given time have the potential to influence future resource status, and thus, future decision-making (Walters 1986; Williams et al. 2002, 2007; Nichols and Williams 2012).

Optimization produces the decision thresholds described in Chap. 2 (Nichols et al., Chap. 2; also see Martin et al. 2009). Decisions at any point in time are typically state-specific, that is they depend on the state of the system at the time at which the decision is made. A decision threshold is simply a value of a state variable where a small change in that variable leads to a different decision. For example, endangered species management might focus on animal abundance as the relevant state variable, with predator control being the optimal decision for a local population with fewer than 70 individuals and no predator control being optimal for 70 or more animals.

This location in state space (70 animals) would represent a decision threshold, with different actions occurring on either side of the threshold. Any utility thresholds (Nichols et al., Chap. 2; Runge and Walshe, Chap. 3) included in the objective function and ecological thresholds included in the system models are important components of input to the optimization algorithm. In summary, optimization uses most of the other components of the decision process, which may include utility and ecological thresholds, to develop decision thresholds specifying actions that are optimal with respect to the decision problem.

Note that the use of optimization methods for making decisions in natural resource management is not the current norm. Rather, it is typical for managers to make decisions with no recourse to optimization, but instead to base decisions on the ability to integrate relevant decision components (objectives, possible actions, model(s) of system response to actions, current state of system) and to develop the optimal decision mentally. Even if humans were able to carry out this integration and make wise decisions without using quantitative or graphical methods, optimization has the advantage of being transparent and defensible. But of course, humans frequently make poor decisions, providing an additional, strong argument for the use of formal optimization methods.

In this chapter, we provide an introduction to optimization methods, emphasizing their role as one component within a larger decision process. More specifically, optimization is the solution algorithm that provides the decision thresholds used in informed decision-making. We begin by establishing the general optimization problem, emphasizing natural resource management. We then describe static (not iterative) optimization for one and then two or more decision variables. This is followed by a discussion of dynamic optimization, a common situation in natural resource management, in which the action at any time influences future system states, requiring an accounting of the consequences of any action throughout the time frame of the management program. Next, we discuss important sources of uncertainty that characterize most decision problems in natural resource management. Approaches to optimization that appropriately incorporate these sources of uncertainty are then described, with emphasis on dynamic decision problems. Extensions to the described approaches are discussed and include different ways of expressing management objectives by focusing on extreme (e.g., small population sizes) rather than on average values. Other extensions are required to deal with nonstationary system dynamics, as might be expected, for example, in the face of climate change. The final section of the chapter then returns to linkages between these various optimization topics and the threshold framework that forms the basis for this book.

The General Optimization Problem

In broadest terms, the generic optimization problem in natural resources is simply to choose among a set of decision options so as to maximize some resource objective that is expressed in terms of the decision choices. Four basic elements typically are required:

First, a range of decision options is needed, from which an option can be selected. The options might focus on resource exploitation, as in harvest rates or amounts; or resource enhancement, as in stocking rates or amounts; or restrictions on certain uses of a resource for recreation or economic production; or any combination of these and other actions that could be taken to influence resource outputs and conditions. In what follows we refer to the options in terms of *decision variables*, and the range of alternatives as a *feasibility set*.

Second, utilities associated with key resource inputs, outputs, and services must be identified. The utilities may be based on the costs of material and energy inputs, or the output of waste products, or the economic benefits of valuable outputs, or ecological features of the system, or aggregates of these and other attributes. In some cases utilities are tied to resource conditions at the time the decision is made, and they may be time-specific. Resource-based utilities may include thresholds indicating, for example, that one component of the objective is to maintain resource state variables (e.g., population size) above some specified level (the utility threshold).

Third, an objective must be specified that aggregates attribute utilities, possibly in a time-dependent way. Objectives often are expressed in terms of minimizing costs, or maximizing benefits, or maximizing benefits net of costs, or other forms that can be linked to the aggregation of utilities. In what follows, we will refer to an *objective function*, to emphasize its relationship to the resource condition and/or the decision that is selected.

Fourth, decision-making for natural resources, especially renewable natural resources, usually requires some accounting for the effect of decisions on resource status and condition. The potential consequences of a decision might be immediate, as in harvest that reduces the size of a population, or longer term, as in the effect a decision has on moving a population toward some desired population status (e.g., decisions aimed at achieving a targeted resource state in minimal time). The projection of the future consequences of present actions is described with a *resource model*. Such models may include ecological thresholds, or locations in state space where small changes in state variables can lead to large changes in system dynamics and responses to management.

Examples of optimization with biological systems might include the following kinds of problems:

- Manage the habitat (and thus the competition coefficients) of three competing populations, so as to minimize a function of the population equilibrium states.
- Choose a fixed harvest rate that maximizes total harvest of a population over some discrete time frame.
- Allocate limited resources among recruitment, survivorship, and other management needs, so as to maximize long-term biological productivity of a population.
- Manage a population to attain a given stock size (utility threshold) while minimizing costs associated with both population size and its rate of change over time.
- Choose a sequence of harvest rates to maximize accumulated harvests as a function of population size and harvest rate.

- Choose a function of stocking rates to attain a given population size (utility threshold), while minimizing expected accumulated costs for a population subject to stochastic influences over time.

In each of these examples, there is at least an implied statement of the objective to be pursued through decision-making, a specification of the decision options that are available to achieve that objective, and an explicit or implied indication of the resource consequences of decision-making.

In optimization with natural resources, it is especially the case that biological complexity begets analytic complexity. For example, single-species systems can be analyzed for optimal patterns much more easily than multi-species assemblages. Populations with only a single age or size or geographic cohort are easier to handle than populations that require an age, size, or spatial structure. An important complicating factor is decision-making through time, with actions taken at any time influencing system behavior at subsequent times.

Static Optimization—A Single Decision Variable

A common form of optimization considers non-iterative optimal decision-making. A statement of the problem is to choose a single decision from a set of decision alternatives that will optimize an objective function that varies with the decision choice.

Discrete alternatives. The simplest and most intuitive form of such decision-making is expressed with a discrete set of decision alternatives. An example might include light, medium, and heavy grazing, with an objective of maximizing sustainable biological production. Assuming production is density-dependent and influenced by grazing pressure, the problem reduces to a simple comparison of sustainable yield for the three grazing levels, with a selection of the level corresponding to the highest yield.

This simple example illustrates a generic approach to static optimization when alternatives range over a discrete set of values. Thus, one essentially enumerates the alternatives, determines the value of the objective for each, and chooses the alternative with the highest (for maximization) or lowest (for minimization) value. This relatively straightforward approach is applicable to a wide range of problems. Examples might include the selection of a policy from a discrete set of policy options; or selection of a method of habitat improvement from a set of alternatives such as seeding, mowing, and fertilization; or the choice of a survey design from a limited set of alternative designs. The key is to be able to express the value of the objective in terms of each alternative, so as to facilitate comparison and ranking of the alternatives. One complicating factor is that if the utilities in the objective function depend on the condition of the resource, then the optimal decision will also vary with resource condition, with different decisions recognized as optimal over different ranges of resource condition.

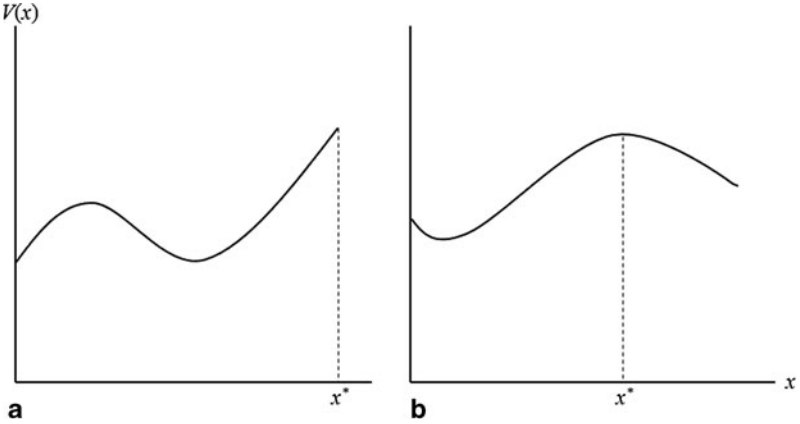


Fig. 4.1 Optimization involving a continuous decision variable. **a** Maximum value is obtained at a point on the boundary of the decision set. **b** Maximum value is obtained at an interior point

In what follows, we restrict attention to maximization, recognizing that a minimization problem can be recast in terms of maximization by the simple expedient of negating the objective function. Then the optimal choice of an option for the redefined maximization problem coincides with that of the original minimization problem, and the optimal values of the two objective functions are reciprocals of each other.

Continuous alternatives. A somewhat more complicated situation is described when the decision alternatives vary over a continuous range. Examples include exploitation at rates that range over the interval from 0 to 1, or application of fertilizer in amounts that can range from 0 to some upper limit, or stocking at levels that can be any value between 0 and some upper limit. Here, the problem is to select an alternative from a continuous range of options, so as to maximize (or minimize) an objective that is a function of the decision option.

For any given problem, there are two possibilities for maximization. One is that a maximum value for the objective function is found at a boundary of the range of feasible options (Fig. 4.1a). The other is that a maximum value is found in the interior of the feasibility set (Fig. 4.1b). Assuming a continuous range of decision alternatives and an objective function that changes smoothly, an optimal interior decision corresponds to a point where the marginal change in the objective function vanishes. Mathematically that means that the derivative of the objective function with respect to the decision variable is 0.

One often-cited example of this is maximum sustainable harvest for a wildlife population with population change given by the logistic equation with harvest,

$$N(t+1) = N(t) + rN(t)\left[1 - \frac{N(t)}{K}\right] - H(t)$$

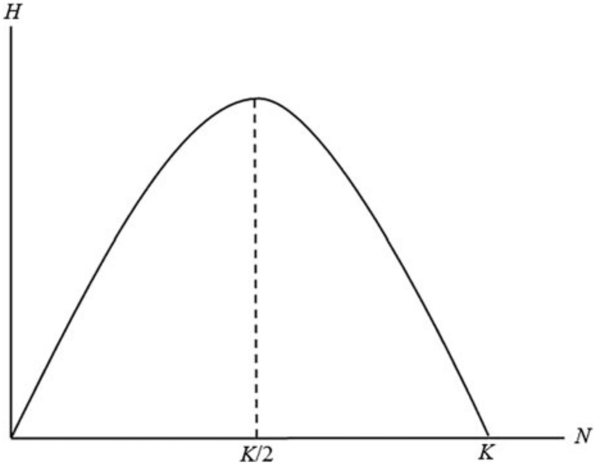


Fig. 4.2 Harvest level for a logistic population with post-recruitment harvest under equilibrium conditions. *K* is the carrying capacity for the population

where r is the intrinsic population rate of growth, K is the "carrying capacity" or maximum sustainable population size, and *H(t)* is the harvest at time t. A simple re-writing of this equation allows us to express harvest H (Fig. 4.2) for a given equilibrium population size N as

$$H = rN\left[1 - \frac{N}{K}\right].$$

An objective function of sustainable harvest can be maximized by selecting the appropriate equilibrium population size, which in turn is given by setting the harvest derivative to 0:

$$\frac{dH}{dN} = r - 2rN/K = 0$$

or

$$N^* = K/2$$

with a corresponding optimal harvest of

$$H^* = rK/4.$$

In general, optimization with a continuous decision variable extends the basic idea of finding the decision alternative in a feasibility set for which an objective function is maximum. In terms of a continuous decision variable x, this means finding a value x^* for which $f(x^*) \geq f(x)$ for every other value x in the feasibility set. A condition that must be met for an interior optimum at x^* is that the first derivative of the objective function vanishes, i.e., $df/dx^* = 0$. If the second derivative at x^* is negative, then local optimality at x^* is guaranteed (Miller 2000).

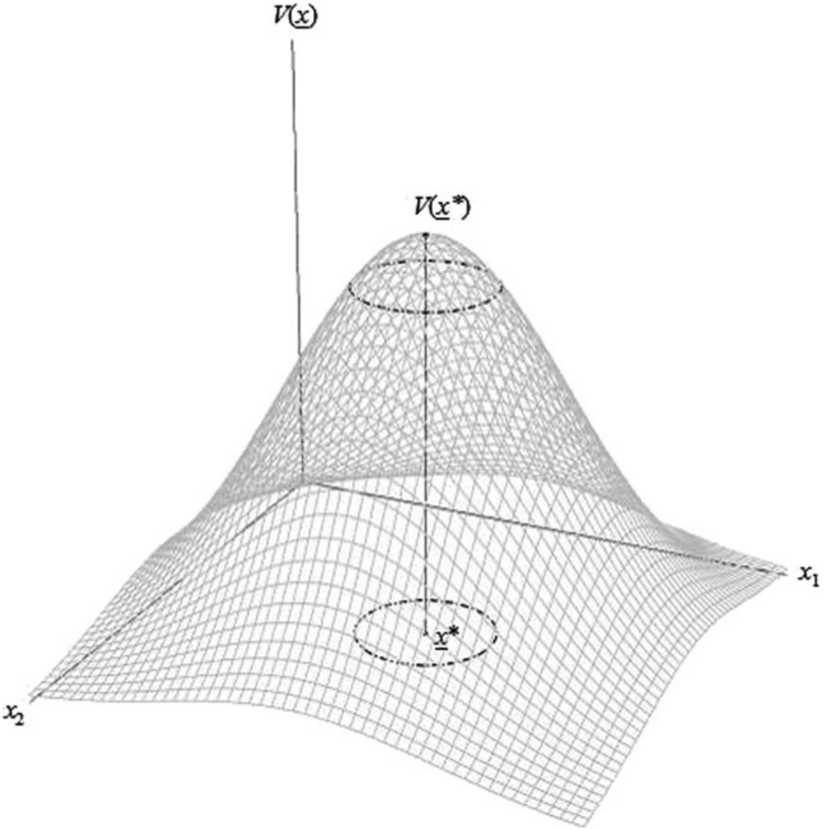


Fig. 4.3 An objective function for a decision problem with two continuous decision variables. $V(\underline{x})$ is the value of the objective function for $\underline{x} = (x_1, x_2)$

Static Optimization—Two or More Decision Variables

The complexity of the optimization problem increases if decision-making involves two or more decision variables. The issue now is to find a combination of values for multiple variables that maximize the objective function. Assume, for example, that a problem can be stated in terms of two decision variables in an objective function. In this situation, three different optimizations are possible. The first two consist of an optimization of one variable holding the other fixed at a particular value, in the manner described above. Since there are two decision variables, either can be optimized conditional on a value assigned to the other. The third optimization concerns the optimal choice of both decision variables, in which the best combination is sought to maximize the objective function (Fig. 4.3).

As above, the optimal combination of continuous decision variables can lie in the interior of the set of feasible values, or at a boundary of that set. For an interior solution, the partial derivatives of the objective function with respect to each of the decision variables must be 0. This condition is not required for optimal solutions at a boundary of the feasibility set.

A simple example of a problem with multiple decision variables is illustrated in Williams et al. (2002), in which a population is subject to exponential post-harvest growth and constant harvest H for a specified amount of time. The equation (Williams et al. 2002) for population size N after T years is

$$N = (1 + r)^T (N_0 - H/r).$$

The population size N clearly depends on the initial population size N_0 and harvest level H. Assuming an objective of maximizing a quadratic expression in both total harvest and the cost of stock removal at the end of the time frame,

$$V(H, N) = (TH)^2 - N^2,$$

the problem is to choose optimal values of the decision variables H and N_0. Partial derivatives of the objective function with respect to the variables yields two equations in H and N_0, and the solution of these equations are the optimal values of these variables.

Optimization in two or more variables extends the idea of finding the value of a decision variable in a feasibility set for which an objective function is maximum. The conditions of a vanishing first derivative and negative second derivative that guarantee an interior optimum also can be extended, to include the vanishing of the partial derivatives of the objective function with respect to each of the decision variables, along with a somewhat complicated negativity condition involving the second-order partial derivatives (Miller 2000).

Problems involving multiple decision variables can become considerably more complicated if there are equality or inequality constraints. Then it is necessary to account for the possibility that a constraint is active, i.e, that an optimal solution is found at a constraint boundary. In a few instances with simple problems (see Fig. 4.1) a boundary solution can be recognized by inspection. More generally, it is necessary to account for constraints in the problem formulation and solution methodology. A common approach utilizes the so-called Kuhn-Tucker conditions, which essentially involve the inclusion of additional decision variables in the problem, one for each constraint. This can complicate the solution approach and analysis of results considerably (see, e.g., Luenberger 1984; Miller 2000; and Hillier and Lieberman 2001 for a detailed discussion of the Kuhn-Tucker conditions).

Linear programming. The complexity introduced by constraints is greatly reduced when both the objective function and other constraints are linear. A problem is amenable to linear programming if its objective function is linear,

$$V(\underline{x}) = c_1 x_1 + c_2 x_2 + \ldots + c_n x_n,$$

constraints on the decision variables are linear, i.e.,

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{im} x_m \leq b_i$$

or

$$a_{i1} x_1 + a_{i2} x_2 + \ldots + a_{im} x_m = b_i,$$

and non-negativity constraints apply,

$$x_i \geq 0.$$

There is a very efficient approach to solving optimization problems of this kind, the well-known Simplex solution algorithm (Luenberger 1984). The algorithm can be used to solve problems with a virtually unlimited number of decision variables, as long as the linearity conditions are met.

Since they are special cases of constrained optimization, linear programming problems also must satisfy the Kuhn-Tucker conditions. These conditions translate into a "dual" linear programming problem, the solution of which expresses the sensitivities to marginal changes in the constraint constants $b_i$in the original problem (Hillier and Lieberman 2001; Williams et al. 2002).

Dynamic Optimization

The level of complexity in optimization increases substantially when decision-making is iterative and a decision at any given time influences future system behaviors. In this situation, a decision made at the beginning of the time frame potentially influences future resource conditions (and therefore future decision-making) over the remainder of the time frame. A second decision made at a later time also influences future resource conditions and future decision-making. In like manner, the resource condition (and decision to be made) at any given time will be influenced by all the decisions made previously, and in turn will influence the resource conditions and decisions made from that point forward. Obviously, this presents a real challenge in deciding what combination of actions to take over time.

A large number of problems in ecology and natural resources are fundamentally dynamic in nature. One traditional approach is to assume fixed decision-making and equilibrium conditions (i.e., the resource condition is unchanging through time), and interpret the patterns in the steady-state resource conditions. By imposing an equilibrium condition on the system state and seeking optimal decisions under that condition, the decision-making problem is essentially treated in terms of static optimization (see the logistic population example above).

When it is not appropriate to assume steady-state conditions, dynamic optimization approaches must be used, with resource conditions as well as decisions potentially changing through time. To properly describe this dynamic optimization problem, it is useful to introduce notation that distinguishes resource conditions from decision variables. Thus, in what follows, *state variables* representing the resource condition or state at a given time in the timeframe will be denoted by $x(t)$. Similarly, *decision variables* representing the decision to be made at that time will be denoted by $a(t)$. Resource change then can be expressed with a *transition equation*

$$x(t+1) = x(t) + f(x, a)$$

that links state and decision variables together, and expresses change in terms of single-step transitions from one time to the next. More generally, multidimensional transitions can be expressed by the transition equation

$$\underline{x}(t+1) = \underline{x}(t) + \underline{f}(\underline{x}, \underline{a})$$

where the underlines indicate more than one measure of resource condition is tracked and/or more than one decision variable is considered at any given time. This formulation of the dynamic decision problem is expressed in discrete time, in which the time index takes a discrete set of values, usually regularly spaced over a timeframe from some starting time t_0 to an endpoint T. System dynamics can also be expressed in continuous time, as

$$d\underline{x}/dt = \underline{f}(\underline{x}, \underline{a})$$

where $d\underline{x}/dt$ represents the instantaneous rate of change of the resource.

The other part of the dynamic resource problem concerns objectives. Assume a utility function $U(x, a)$ that measures the utility (e.g., benefits net of costs) associated with action a when the resource is in state x. Then an objective function

$$\sum\nolimits_{t_0}^{T} U(a(t), x(t))$$

can be defined by accumulating utilities as actions are taken and resource conditions respond over the timeframe.

With this additional notation, we can describe the optimization problem as

$$\underset{A_{t_0}}{\text{maximize}} \sum\nolimits_{t_0}^{T} U(a(t), x(t))$$

subject to

$$x(t+1) = x(t) + f(x, a)$$

and

$$x(t_0) = x_0$$

where A_{t_0} represents a strategy of state-specific and time-specific actions from t_0 to T. For problems in continuous time, the summation in the objective function is replaced with integration, so that the optimization problem is

$$\underset{\mathrm{A_{t_0}}}{\text{maximize}} \int_{t_0}^{T} U(a(t), x(t))dt$$

subject to

$$dx/dt = f(x, a)$$

and

$$x(t_0) = x_0.$$

Many variants of this problem statement are possible (Bryson and Ho 1975). One that is especially relevant to natural resources includes in the objective function a salvage value $F_1[x(T)]$ that depends on the state of the resource at the end of the time frame. For example, a value can be recognized in leaving a cohort of a harvested population at the end of a timeframe, to sustain the population into the future. Another variant includes an assumption that the future is less important than the present, so that utilities are time-discounted in the objective function. Including these elements gives an objective function of the form

$$\sum\nolimits_{t_0}^{T} \lambda^t U(a(t), x(t)) + F_1[x(T)],$$

where λ^t discounts the time t utility back to present value.

The mathematics needed to solve such problems are elegant but complicated, and comprehensive treatments can be found in the corpus of literature on systems analysis and optimal control. The theory and approaches for optimal control of dynamic systems, including the calculus of variations, Pontryagin's Maximum Principle, and dynamic programming, constitute one of the great triumphs of modern applied mathematics (Sage and White 1977; Stengel 1994; Bertsekas 1995).

Uncertainty and Resource Dynamics

Natural resources management almost always involves uncertainties as to the consequences of management actions. We focus here on four uncertainty components that occur frequently in natural resource decision-making, namely environmental variation, partial controllability, partial observability, and structural uncertainty. Each component influences natural resources management in different ways and at different points in a resource system (Fig. 4.4). Taken separately or in combination, they can limit our ability to identify useful management strategies. We discuss each of the components below, along with approaches for including them in dynamic optimization.

Environmental variation. Environmental conditions can be viewed as external factors that influence, but are not influenced by, resource conditions and dynamics. They can directly and indirectly influence the ecological and physical processes that determine resource dynamics. They vary randomly over time, so that future conditions cannot be predicted with certainty. The resultant uncertainty in resource dynamics complicates decision-making.

In the context of optimal decision-making, environmental variation is often treated by the artifice of including a random component $z(t)$ in the transition equation governing system dynamics:

$$x(t+1) = x(t) + f(x, a, z).$$

The net effect of adding a random component is to induce randomness in the transitions, so that the state at each time is predictable only probabilistically. This

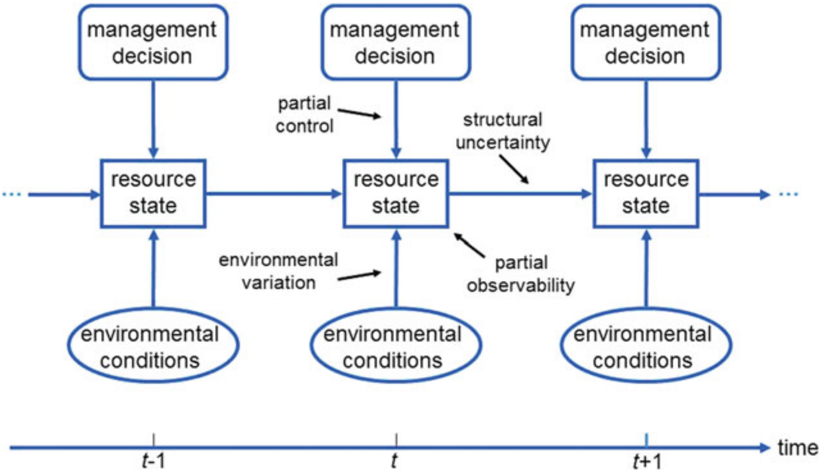


Fig. 4.4 Change in a resource system as influenced by fluctuating environmental conditions and management actions. Different uncertainty factors affect resource dynamics and management in different ways

randomness is often handled by modeling resource transitions as if they constitute a Markovian decision process (Puterman 1994), with transitions between states described by a decision-specific transition probability structure:

$$P(x(t+1)|x(t), a(t)).$$

In words, this expression specifies a probability that the resource will be in state $x(t+1)$ at time $t+1$, conditional on resource state being $x(t)$ at time t and assuming action $a(t)$ is taken.

Randomness in resource transitions through time induces randomness in the utilities that depend on resource conditions. The objective function inherits this randomness, and is itself random. Under such circumstances an appropriate objective simply utilizes an average or expected value,

$$E\left\{\sum\nolimits_{t_0}^{T} U(a(t), x(t))\right\},$$

with the expectation essentially connoting an averaging over the possible resource trajectories.

The pattern of transition above is sometimes called a first-order Markovian decision process. Higher-order processes also are possible. For example, a second-order Markovian process is defined by a probability structure in which the transition to a new resource state $x(t+1)$ depends on both $x(t)$ and $x(t\text{-}1)$. Although sometimes a more realistic portrayal of resource dynamics, optimization under these conditions becomes much more difficult (Williams 2007).

Partial controllability. Partial controllability refers to an imprecise linkage between decisions that are made and the actions that are actually implemented. Unintended outcomes are often a result of management decisions implemented by indirect means. For example, hunting permits may be used as an indirect means to attain a desired rate of ungulate harvest, or forestry regulations may be used to limit logging-related impacts on wildlife. The net effect of partial controllability is that the intended outcome of a management decision is only partially accomplished by the action that ultimately is taken.

A natural way to model this uncertainty is to assign a probability distribution to the decision variable, wherein a particular decision corresponds to any of a range of possible actions along with a probability structure describing the likelihood that each action will actually be taken. In this way, partial controllability becomes yet another random factor like environmental variation. In combination, partial controllability and environmental variation lead to less-precise predictions of the transitions between resource states. As a practical matter, they often are expressed as a single random component that includes both effects.

Partial observability. Partial observability expresses an inability to observe completely a resource system that is being managed. Natural resources are almost always partially observed. For example, only a part of the area where a fish population occurs can be monitored, and a sampling strategy needs to allow inferences over the whole area on the basis of the observation of only a part of it. Observability is further complicated by the fact that individual organisms often escape detection, even in areas that are intensively monitored. In combination, incomplete geographic coverage and incomplete detectability mean that observations collected in the field are associated with—but not the same as—actual system states. By obscuring the resource status on which effective management depends, partial observability reduces management effectiveness, even if environmental variation is minimal and management actions are precisely controlled.

Partial observability is commonly measured by sampling variation, which occurs when field data are collected and analyzed. One way to reduce partial observability is by designing field sampling efforts efficiently, for example with more intensive sampling, optimal geographic design of the sampling effort, and the use of standard survey principles such as randomization, replication, and control. However, partial observability can rarely be eliminated, no matter what the design and sampling intensity.

There are several ways of dealing with partial observability in decision-making. One is to estimate resource status with field data, and then treat the estimate as if it accurately represents resource conditions. Another is to state the uncertainty about resource status explicitly with probabilities for possible resource states, and incorporate these probabilities directly into the decision-making process (Williams 2009). The first approach is far more common in natural resource management. Of course, the most straightforward way to address partial observability is to reduce it as much as is practicable with well-designed monitoring.

Structural uncertainty. Structural uncertainty is a result of a lack of understanding (or lack of agreement) about the processes that control resource dynamics. In virtually all cases there is some degree of uncertainty about the forms and functions—i.e., the

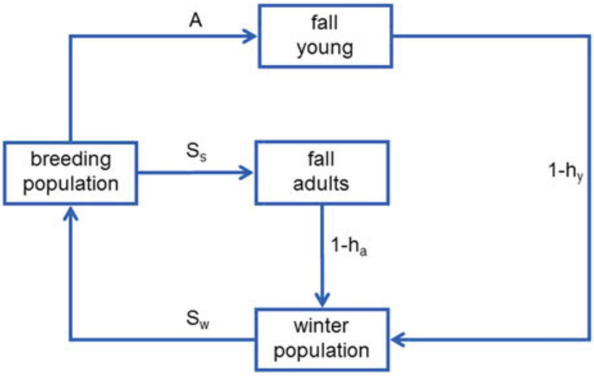


Fig. 4.5 Conceptual model of annual cycle of waterfowl population dynamics. Model includes survival rates for spring-summer (S_s) and fall-winter (S_w), along with harvest rates for young (h_y) and adults (h_a) and age ratio (A) for reproduction/recruitment.

structure—of natural processes. Structural uncertainty can limit our ability to manage resources effectively and efficiently, even if monitoring is exact, management actions are rigorously controlled, and environmental variation is minimal.

Differing views about how natural processes work and how they respond to management are examples of structural uncertainty. These views can be framed as hypotheses about system processes and responses and then embedded in models, which in turn can be used to make testable predictions. Examples of uncertainty about resource form and function include hypothesized associations among different attributes of the resource; or relationships between controls and resource elements; or connections between environmental conditions and resource states; or parameterizations of these relationships. The hypothesized forms and parameterizations can be incorporated in different models, and structural uncertainty then is expressed in terms of uncertainty about which model (and its imbedded hypothesis) best represents resource dynamics.

An example that highlights many of these points is the modeling framework for adaptive harvest management of waterfowl (Williams and Johnson 1995; Williams et al. 2002), which uses a simple model to represent associations among fall harvest, seasonal survivorship, and spring reproduction for waterfowl (Fig. 4.5). Contrasting hypotheses about the impact of harvest on annual survivorship are easily incorporated into different versions of the model, by describing different functional relations between harvest rates and post-harvest survival. In addition, contrasting hypotheses about the importance of density dependence in recruitment are incorporated by describing recruitment in terms of spring population size. In combination, these hypotheses define different models, each with its own predictions about harvest impacts.

Structural uncertainty is often quantified with measures of confidence in the ability of competing models to predict resource dynamics. A common mathematical approach to updating these measures is Bayesian updating, which combines confidence values and monitoring data at each point in time to generate new confidence values for the next point in time (Lee 1989). In this way confidence increases for models that forecast resource conditions accurately, and confidence declines for models that do not make accurate forecasts.

Structural uncertainty, like the other forms of uncertainty, has a tendency to obscure the effects of management and reduce effectiveness. However, it differs from environmental variation and partial controllability in its point of influence (Fig. 4.4) and the manner in which it is treated. Structural uncertainty can be reduced by applying management strategies to affect the measures of model confidence. In contrast, environmental variation (and in some cases partial controllability) are effectively uncontrolled.

Optimization in the Face of Uncertainty

Given the prevalence of uncertainty, treating it in a context of optimal decision-making constitutes an important challenge in natural resource management. The development of theory and methods to account for uncertainty is a fast growing area, with decision science being advanced by contributions coming directly from the natural resources field. In what follows we characterize the problem of optimal decision-making in the face of uncertainty, and suggest some ways to address each of the sources of uncertainty mentioned above.

Environmental variation and partial controllability. Both environmental variation and partial controllability can be incorporated into an optimization problem in fairly straightforward ways. Environmental variation can be handled by including a random environmental variable in the transition equation, as discussed above. And partial controllability can be incorporated by simply allowing the actions implemented for a given decision to be random. The basic structure of the optimization problem remains intact with both these adjustments, but the transitions become stochastic. One result is that the optimization objective must be based on an average of accumulated utilities over the possible resource trajectories

$$\underset{\mathrm{A}_{t_0}}{\text{maximize}}\ E\left\{\sum\nolimits_{t_0}^{T} U(a(t), x(t))\right\}$$

with the expectation representing an averaging over the trajectories. An economical expression of the problem describes it as a Markovian decision process, with the transitions specified in terms of Markovian probabilities $P(x(t+1)|x(t), a(t))$.

Partial observability. A well-recognized challenge in natural resources is to find efficient and effective ways to incorporate partial observability into the decision-making architecture. Research on partially observable Markov decision processes has been ongoing for many years in operations research, artificial intelligence, and other fields, and this work offers a useful way to deal with partial observability (Kaelbling et al. 1998). Thus, one simply acknowledges that the resource state is not known with certainty at any given time, and must be characterized with probabilities assigned to the possible states. Then the state-specific objective function is replaced by one that is averaged over all possible states, based on a distribution of state probabilities. Since the averaged objective function depends on these probabilities, it is a function of the distribution rather than a single state, as would be the case for an optimization problem in the absence of partial observability.

A key challenge under partial observability is that the state probability distribution evolves through time, as the system responds to management and information is collected through monitoring. In this sense the distribution itself is a dynamic attribute, with changes that must be tracked and incorporated into the objective function to allow optimization through time. This increases the complexity of the optimization problem.

Structural uncertainty. Finally, there is structural uncertainty, which presents a different challenge for optimization. Here, different models are required to express alternative views about how the resource system works and how it responds to management

$$x(t+1) = x(t) + f_k(x, a, z)$$

where the subscript k is used to denote one of several different models. Furthermore, time-specific measures $q_k(t)$of model confidence are required, one for each model that expresses the level of confidence in that model in representing resource dynamics. As the system responds to management through time and resource conditions change, monitoring information can be used to update the model confidence values.

Since state trajectories over the timeframe depend on which model is used, the aggregation of utilities will also. To account for variation across models, the objective for this optimization problem averages the aggregations with the confidence measures

$$\underset{\mathrm{A}_{t_0}}{\text{maximize}} \sum\nolimits_k q_k(t) E\left\{\sum\nolimits_t U_k(a(t), x(t))\right\}.$$

Because the averaged objective function is based on the particular likelihoods $q_t(k)$ at each decision point, it is necessary to account not only for the state of the resource system, but also the model measures themselves. In this way the optimization problem becomes larger and more complex, with the objective a function of both the resource state and the confidence measures.

A key challenge here is that the distribution of model probabilities evolves through time as the system responds to management. Thus, the objective function effectively changes through time, and the anticipated change must be included in identifying optimal decisions. As the most appropriate model to represent resource dynamics is revealed through time, decision-making becomes more responsive to the system characterization that best reflects actual system changes. The adjustment of management strategy as one learns about resource responses through time is the essence of adaptive decision-making (Walters 1986; Williams 2011).

Extensions

The framework for optimization can be expanded to include other forms for decision-making objectives and other types of uncertainty. Here we mention a few examples that are especially pertinent to natural resources decision-making.

Minimax and maximin. In the development above, we utilized averaging over a range of possible responses to management, with objectives that are weighted averages over distributions of random variables. For example, the objective for a problem with environmental variation was described in terms of the average or expected value of aggregated utilities over the range of possible trajectories of system states.

Rather than focusing on average responses, it sometimes is useful to emphasize the extremes in resource dynamics. As an example, consider a problem involving environmental variation in which one seeks the smallest value an objective can take for each decision, with an intent to select the decision with the maximum of those minima. Such an approach is indifferent to large values of the objective, but very sensitive to small values. This emphasis is known as "maximin" optimization, and it is appropriate for problems that are cast in terms of maximizing the minimum gain among decision alternatives. The idea is to seek decisions that essentially protect "away from" small values of aggregate utility. An example might be protecting against species extinction by choosing the conservation option with the maximum of the smallest population size among the decision alternatives.

On the other hand, one could instead consider the largest value the objective can take for each decision, and seek the decision that *minimizes* that maximum. Here the approach is indifferent to small values of the objective, but is very sensitive to large values. This emphasis is known as minimax, and it is appropriate for problems that are cast in terms of minimizing the maximum loss among decision alternatives. The idea is to seek decisions that protect "away from" large values of aggregate loss. An example might be ensuring against the invasion of alien species by choosing the option with the minimum of the largest population size among the decision alternatives.

A useful approach to finding solutions of minimax and maximin problems involves computer simulations to explore objective values corresponding to a large number of potential trajectories for each of the feasible decisions. A sort through the trajectories for each decision to find upper or lower limits for the objective can identify candidates to consider for solutions of the problem.

Non-stationary change and deep uncertainties. In most of the above development, we have assumed that the future, even if not known with certainty, can be anticipated stochastically. For example, randomness in environmental conditions induces random resource changes, but stability in environmental patterns over time preserves a recognizable probability structure for future resource changes. This stochastic structure can be used to good effect to compare and contrast decision alternatives in the process of decision-making. Approaches to system analysis and control have traditionally rested on the assumption that patterns of fluctuation in system features are stable over time.

However, it is becoming increasingly clear that for a great many resource systems, the ecological structures and processes controlling resource dynamics are changing in new and sometimes unpredictable ways. The scale and scope of these changes are large, complex and deeply uncertain, in that it is not possible to forecast the resource changes, even probabilistically.

Of particular importance is that environmental conditions, and the ecological processes influenced by them, are exhibiting directional patterns of change. An obvious example is climate change, in which the environment is seen as evolving directionally in terms of temperature, precipitation and other variables. Though some overall trends in some climate metrics are becoming clear, what is not at all clear is the rapidity and extent of change, or its impacts on natural resources at regional and local scales. An important challenge for resource management is to include such non-stationary trends, and the deep uncertainties associated with them, in natural resource planning and management (Milly et al. 2008; Nichols et al. 2011).

One approach to this problem is to develop environmental scenarios with different patterns of directional change, and try to design acceptable management strategies that account for uncertainties among the scenarios. Scenario analysis is gaining currency in natural resource assessment and management, as we adapt to large-scale and uncertain climate and land-use changes (Fahey and Randall 1998). Another approach is to use distribution-free techniques such as info-gap (Ben-Haim 2002) and robust design methods (Lempert and Collins 2007) to identify strategies that appear to meet minimal performance requirements across multiple response criteria.

Non-stationarity and the deep uncertainties associated with it present a new and serious challenge to resource decision-making, one for which we need new approaches that go beyond the standard ways of framing and conducting decision analysis. At a minimum it is necessary to look for directional trends in environmental conditions and systematic changes in resource structures and functions, and consider ways to accommodate them.

Optimization and Thresholds

This chapter began with a description of the generic optimization problem, and outlined different approaches to problems under different structural assumptions. The generic problem included specification of objectives, available actions, and models that make predictions about utilities and changes in system state expected to result from each possible action. An objective of management may be to maintain values of key variables above or below certain specified values. We refer to such values as utility thresholds (Nichols et al., Chap. 2; Runge and Walshe, Chap. 3), and these are commonly found in objective functions for decision problems with multiple objectives. Models of managed natural resource systems frequently include locations in state space within which small changes in environmental drivers or management actions bring about large changes in state variables. We refer to these locations as ecological thresholds (Martin et al. 2009). Conditional on objectives (which may include utility thresholds) and models (which may include ecological thresholds), optimization produces recommended optimal actions. In many cases, these recommendations will be state-specific. Decision thresholds (Martin et al. 2009; Nichols et al., Chap. 2) are locations in state space where a small change in the value of a state variable produces a change in the optimal or recommended management action. In summary, a framework of optimal decision-making may well include ecological, utility, and decision thresholds. The purpose of the optimization is to produce recommended actions for different locations in state space. These recommendations then define the decision thresholds to be used by managers.

References

Allen, C. R., K. L. Pope, and J. J. Fontaine, eds. 2011. Adaptive management for natural resources. *Journal of Environmental Management* 92:1339–1428.

Ben-Haim, Y. 2002. *Info-gap decision theory: Decisions under severe uncertainty*. 2nd ed. New York: Elsevier.

Bertsekas, D. P. 1995. *Dynamic programming and optimal control: Volumes I and II*. Belmont: Athena Scientific.

Bryson, A. E., and Y.-C. Ho. 1975. *Applied optimal control: Optimization, estimation, and control*. Bristol: Taylor and Francis.

Fahey, L., and R. M. Randall. 1998. *Learning from the future: Competitive foresight scenarios*. New York: Wiley.

Hillier, F. S., and G. J. Lieberman. 2001. *Introduction to operations research*. 7th ed. New York: McGrw-Hill.

Holling, C. S., ed. 1978. *Adaptive environmental assessment and management*. New York: Wiley.

Kaelbling, L. P., M. L. Littman, and A. R. Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial Intelligence* 101:99–134.

Lee, P. M. 1989. *Bayesian statistics: An introduction*. London: Edward Arnold.

Lempert, R. J., and M. T. Collins. 2007. Managing the risk of uncertain threshold response: Comparison of robust, optimum, and precautionary approaches. *Risk Analysis* 27:1009–1026.

Luenberger, D. G. 1984. *Linear and nonlinear programming*. 2nd ed. Reading: Addison-Wesley.

Martin J., M. C. Runge, J. D. Nichols, B. C. Lubow, and W. L. Kendall. 2009. Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications* 19:1079–1090.

Miller, R. E. 2000. *Optimization*. New York: Wiley.

Milly, P. C. D., J. Betancourt, M. Falkenmark, R. M. Hirsch, Z. W. Kundzewicz, D. P. Lettenmaier, and R. J. Stouffer. 2008. Stationarity is dead: Whither water management? *Science* 319:573–574.

Nichols, J. D., M. J. Eaton, and J. Martin. In press. Thresholds for conservation and management: Structured decision making as a conceptual framework. In *Thresholds for conservation*, ed. G. Gunterspergen and P. Geissler. New York: Wiley.

Nichols, J. D., M. D. Koneff, P. J. Heglund, M. G. Knutson, M. E. Seamans, J. E. Lyons, J. M. Morton, M. T. Jones, G. S. Boomer, and B. K. Williams. 2011. Climate change, uncertainty and natural resource management. *Journal of Wildlife Management* 75:6–18.

Nichols, J. D., and B. K. Williams. 2012. Adaptive management. In *Encyclopedia of Environmetrics*. 2nd ed. (pp. 1–6), New York: Wiley.

Puterman, M. L. 1994. *Markov decision processes: Discrete stochastic dynamic programming*. New York: Wiley.

Runge, M. C., and T. Walshe. In press. Identifying objectives and alternative actions to frame a decision problem. In *Thresholds for conservation*, ed. G. Guntenspergen and P. Geissler. New York: Springer.

Sage, A. P. and C. C. White. 1977. *Optimum systems control*. 2nd ed. Englewood Cliffs: Prentice-Hall.

Stengel, R. F. 1994. *Optimal control and estimation*. New York: Dover Publishers.

Walters, C. J. 1986. *Adaptive management of renewable resources*. Caldwell: Blackburn Press.

Williams, B. K. 2007. Optimal management of non-Markovian biological populations. *Ecological Modelling* 200:234–242.

Williams, B. K. 2009. Markov decision processes in natural resources management: Observability and uncertainty. *Ecological Modelling* 220:830–840.

Williams, B. K. 2011. Adaptive management of natural resources: Framework and issues. *Journal of Environmental Management* 92:1346–1353.

Williams, B. K., and F. A. Johnson. 1995. Adaptive management and the regulation of waterfowl harvests. *Wildlife Society Bulletin* 23:430–436.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. *Analysis and management of animal populations*. San Diego: Academic Press.

Williams, B. K., R. C. Szaro, and C. D. Shapiro. 2007. Adaptive management: The U.S. department of the interior technical guide. Washington, DC: Adaptive Management Working Group (U.S. Department of the Interior).

Chapter 5
Application of Threshold Concepts to Ecological Management Problems: Occupancy of Golden Eagles in Denali National Park, Alaska

**Mitchell J. Eaton, Julien Martin, James D. Nichols, Carol McIntyre,
Maggie C. McCluskie, Joel A. Schmutz, Bruce L. Lubow
and Michael C. Runge**

Abstract In this chapter, we demonstrate the application of the various classes of thresholds, detailed in earlier chapters and elsewhere, via an actual but simplified natural resource management case study. We intend our example to provide the reader with the ability to recognize and apply the theoretical concepts of utility, ecological and decision thresholds to management problems through a formalized decision-analytic process. Our case study concerns the management of human recreational activities at Alaska's Denali National Park, USA, and the possible impacts of such activities on nesting Golden Eagles, *Aquila chrysaetos*. Managers desire to allow visitors the greatest amount of access to park lands, provided that eagle nesting-site occupancy is maintained at a level determined to be acceptable by the managers themselves. As these two management objectives are potentially at odds, we treat minimum desired occupancy level as a *utility* threshold which, then, serves to guide the selection of annual management alternatives in the decision process.

M. J. Eaton (✉) · J. Martin · J. D. Nichols · M. C. Runge
Patuxent Wildlife Research Center,
U.S. Geological Survey, Laurel, MD, USA
e-mail: mitchell.eaton@usgs.gov

M. J. Eaton
Southeast Climate Science Center U.S. Geological Survey
127H David Clark Labs, NCSU Raleigh, NC 27695, USA

J. Martin
Florida Fish and Wildlife Conservation Commission, Fish and Wildlife
Research Institute, 100 8th Avenue SE, St. Petersburg, FL 33701, USA

C. McIntyre · M. C. McCluskie
National Park Service, 4175 Geist Road, Fairbanks, AK 99709, USA

J. A. Schmutz
Alaska Science Center, U.S. Geological Survey,
4210 University Drive, Anchorage, AK 99508, USA

B. L. Lubow
Natural Resource Ecology Laboratory,
Colorado State University, Fort Collins, CO 80523, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_5,

As human disturbance is not the only factor influencing eagle occupancy, we model nesting-site dynamics as a function of both disturbance and prey availability. We incorporate uncertainty in these dynamics by considering several hypotheses, including a hypothesis that site occupancy is affected only at a threshold level of prey abundance (i.e., an *ecological* threshold effect). By considering competing management objectives and accounting for two forms of thresholds in the decision process, we are able to determine the optimal number of annual nesting-site restrictions that will produce the greatest long-term benefits for both eagles and humans. Setting a utility threshold of 75 occupied sites, out of a total of 90 potential nesting sites, the optimization specified a *decision* threshold at approximately 80 occupied sites. At the point that current occupancy falls below 80 sites, the recommended decision is to begin restricting access to humans; above this level, it is recommended that all eagle territories be opened to human recreation. We evaluated the sensitivity of the decision threshold to uncertainty in system dynamics and to management objectives (i.e., to the utility threshold).

Keywords Golden Eagles · *Aquila chrysaetos* · Utility threshold · Ecological threshold · Decision threshold · Occupancy modeling · Structured decision-making · Adaptive management · Uncertainty · Wildlife disturbance

Introduction

Structured Decision-Making and Thresholds

Thresholds, in the context of management decisions, have recently received attention in the conservation and ecological literature (Martin et al. 2009c; Samhouri et al. 2010; Andersen et al. 2009). In this volume, Nichols et al. (Chap. 2) have provided clear guidelines to distinguish among classes of threshold and, at the same time, have offered a logical conceptual framework for considering the role and appropriate application of threshold types in structuring a decision process for management problems. Here, we illustrate this conceptual framework by describing the formal inclusion of thresholds into a process of structured decision-making (SDM). Our example focuses on the management of recreational activities near nesting Golden Eagle (*Aquila chrysaetos*) territories in Alaska's Denali National Park (Denali NP). We used a simplified version of an actual case study (Martin et al. 2011) to illustrate the relationship among different types of thresholds when applying SDM to natural resource management. Our objectives for this chapter are to describe the formulation of the management problem in an SDM framework and explore in detail the process of testing for and incorporating thresholds in the SDM framework.

SDM is an analytic framework that aids decision-makers in coping with complexity and uncertainty by deconstructing the problem into components, identifying the sources of uncertainty and impediments to the decision, and then finding the optimal solution by integrating the components (Clemen and Reilly 2001). Essential elements of the SDM process include a clear statement of objectives that are expressed as quantitative measures and are used to evaluate the success of management decisions, a set of discrete actions that form the basis of the decision, one or more models of the system dynamics that predict the outcome of each potential management action in terms of the measurable objectives, and an optimization method that identifies the action that is most likely to achieve the objectives given the expected outcomes and effects of uncertainty (Clemen 1996). When decisions are made repeatedly over time, SDM can become an adaptive process if it includes a targeted monitoring program that is used to reduce uncertainty in system behavior and feed information back to managers. Monitoring of this form is specifically designed to provide information on the state of the system to (1) allow state-dependent decisions to be made, (2) evaluate progress towards objectives following the implementation of a management action, and (3) improve future management decisions by comparing observations of system response to predictions generated by competing models, thereby reducing the uncertainty of future predictions (Lyons et al. 2008; McCarthy and Possingham 2007; Williams et al. 2002). As natural resource management decisions are often made in the context of thresholds—in the form of triggers that prompt the need for specific actions to be taken or a desire to keep a focal state variable above or below a specified level—clarifying threshold categories and their roles is essential to improving our decision-making abilities.

Three types of thresholds—ecological, utility, and decision—have been identified as being relevant to natural resource management (Nichols et al., Chap. 2; Martin et al. 2009c). Ecological thresholds, arguably the type most commonly encountered in the literature, are considered as boundaries between alternative ecological regimes and represent values of system state where substantial changes in the dynamics of one or more elements of the system are observed (e.g., Fahrig 2001), or where system state variables or rate parameters reach certain levels (Nichols et al., Chap. 2). For example, in the context of a predator species, prey abundance level may be regarded as a relevant state variable, such that attainment of some level (ecological threshold) brings about dramatic increases or decreases in local rates of colonization or extinction. Alternatively, ecological thresholds can be viewed as values of state or other variables at which vital rates attain specified values. For example, the concept of extinction threshold (Lande 1987) concerns the proportion of suitable habitat potentially available to a metapopulation. The extinction threshold is that proportion of patches containing suitable habitat at which the probability of metapopulation extinction is equal to one.

Utility thresholds, in contrast, are formulated from management objectives and defined as the point where small changes in system state or performance level result in significant improvements (or declines) in the return (utility) of decision outcomes (Martin et al. 2009c). Utility thresholds are derived from value judgments of stakeholders and most often pertain to desired ecosystem states or functions. Correspondence between ecological and utility thresholds is possible, but establishment of a utility threshold can be independent of the existence of ecological thresholds. Samhouri et al. (2010) provide an example of how a utility threshold might coincide with an ecological threshold: If the desired system state of a freshwater lake is clear water, then the utility threshold value may correspond naturally with an ecological threshold where small changes in nutrient input result in dramatic changes to water clarity. In the context of Golden Eagles, a utility threshold might be based on the desire of managers to ensure that some minimum number of eagle territories is occupied each year (Martin et al. 2011). Such an objective could arise from the values of the protected area manager, from observations on nesting numbers before significant human disturbance was recorded in the park, or from a population viability analysis (PVA) suggesting that sustained occupancy above this threshold level will maintain the risk of local extinction at a desired low level. The latter possibility represents a utility threshold coinciding with one form of ecological threshold.

Finally, decision thresholds are defined as changes in state variable values that result in changes in the optimal management action recommended to meet management objectives. As such, a decision threshold is the product of the SDM process itself, conditional on any ecological threshold(s) included in the predictive models and on the utility threshold(s) included in the objective function. In the case of managing eagles in Denali NP, a decision threshold would be represented by a change in management policy (e.g., from few to many restrictions imposed on human recreational activities) resulting from predicted changes in the number of occupied nesting territories (a state variable) and the desired occupancy level (utility threshold). Changes in management decisions, therefore, will be a product of the model(s) of system dynamics and the objective function specified by managers. The model(s) of eagle occupancy dynamics predicts the impacts of human activities, accounting for any hypothesized ecological thresholds, while the objective function contains any specified utility thresholds.

Golden Eagles and the Impacts of Recreational Activities on Nesting in Denali National Park

Denali NP, Alaska, contains the highest-reported nesting density of Golden Eagles in North America, with approximately 80 breeding pairs monitored since 1988 (C. McIntyre, personal communication; Kochert et al. 2002). In a 1,800-km^2 study area within the park, eagles nest exclusively on cliffs and rock outcroppings. Denali Golden Eagles are migratory, returning to the park each March to lay 1–3 eggs. Eggs are incubated for approximately 40 days, with hatching occurring in June and young eaglets fledging by early August. Managers in Denali NP are concerned that back-country hiking, airplane tours, and other recreational activities may negatively impact the occupancy of Golden Eagles in potential nesting sites and, therefore, reduce overall breeding success in the study area. Martin et al. (2009a) suggested that Golden Eagle occupancy and breeding success may be influenced by human disturbance and the abundance of snowshoe hares (*Lepus americanus*), a principal prey item of nesting eagles. Human recreational activities have been implicated as a significant factor in wildlife disturbance, including negative effects on raptor nesting success and stress or reduced productivity in other nesting species (McGowan and Simons 2006; Morse et al. 2006; Steidl and Anthony 2000; Swarthout and Steidl 2003). Managers, however, are also mindful of the role of the NP and are mandated to provide as many recreational opportunities to human visitors as possible without threatening or causing undue disturbance to habitat or wildlife. Thus, the management decision in this problem is to what extent the park should restrict human recreation activities in eagle nesting territories. The potential for human disturbance to affect the occupancy of Denali eagles is unknown and represents one source of uncertainty. In our formulation of the decision structure for this management problem, we also recognize that the form of the relationship between prey abundance and eagle occupancy represents additional uncertainty, and therefore consider alternative hypotheses to describe and test this relationship. Martin et al. (2009a) used multistate site occupancy models (unoccupied, occupied, occupied with breeder) to evaluate the effect of disturbance and hare abundance on parameters that govern the occupancy and breeding dynamics of eagles. For our current emphasis on the role of thresholds in decision-making, we simplified the example of Martin et al. (2009a) by using two-state occupancy models (i.e., unoccupied or occupied) described by MacKenzie et al. (2006). We extend a similar two-state occupancy analysis developed by Martin et al. (2009b) to include model covariates of hare abundance and human disturbance. Finally, we describe a monitoring program that could be implemented to reduce uncertainty in model confidence through an adaptive management approach (Williams et al. 2002, 2007).

Methods

Defining an Objective Function with Utility Threshold Constraints

Management objectives embody the fundamental desires of the decision-maker and can, and in most cases should, represent the values of all stakeholders. Objectives, then, become the basis for assessing the success of alternative management decisions. The objective function, a mathematical formulation of management objectives and constraints (Williams et al. 2002), is the formal means to quantify the management outcome (return) of implementing any particular decision at a given time. If the decision-maker must consider several objectives simultaneously, it is often useful to convert one or more objectives into constraints and include them in the objective function as utility thresholds. Management objectives for Denali NP are to maximize recreational opportunities for human visitors, while at the same time minimizing the effects of recreation on site occupancy levels of Golden Eagles in nesting territories. To reconcile these seemingly competing goals, we convert the second objective to a constraint and include it as a utility threshold in the objective function. The utility threshold, like the objective function in general, is a value judgment and is decided on by the decision-maker. In this case, park personnel provided expert opinion and concluded that using the average number of occupied nesting territories observed over the last 20 years was an appropriate minimum threshold for management objectives. This threshold value is incorporated in decision-making by way of a penalty parameter that devalues the current return on a particular management action, given the expected system response (Martin et al. 2011):

$$\alpha = \begin{cases} 0, \; E_i\left(N^{\mathrm{O}}_{t+1}\right) < \tau \\ 1, \; E_i\left(N^{\mathrm{O}}_{t+1}\right) \geq \tau \end{cases}, \tag{5.1}$$

where α is the penalty factor, $E_i(N^{\mathrm{O}}_{t+1})$ is the expected number of occupied nesting sites in year $t+1$, following management action i, and τ is the utility threshold value. As specified by this equation, if occupancy is expected to fall below τ after the implementation of management action i, the value returned by the objective function is multiplied by the penalty factor and, thus, reduced to 0, i.e., $\alpha = 0$. If expected occupancy is equal to or greater than τ, the return produced by the objective function retains full value, i.e., $\alpha = 1$.

The full utility function can then be defined as

$$U_t\left(N^{\mathrm{O}}_t, r_t\right) = \left(N^{\mathrm{tot}} - r_t\right) \times \alpha, \tag{5.2}$$

where the utility value, U_t, is a function of the number of occupied territories (N^{O}) and the number of territories restricted to human activity (r) at time t. N^{tot} is the total number of nesting sites available. By minimizing the number of restricted territories, we maximize the function ($N^{\mathrm{tot}}_t - r_t$), but only so long as expected nesting-site occupancy remains above τ (i.e., $\alpha \neq 0$).

We then select a sequence of management actions, from the present (t) to some future time (T) that maximizes our objective function with respect to expectations under random environmental variation

$$\max_{r_t} E \sum_{t}^{T} [U_t(N^{\mathrm{O}}_t, r_t)]. \tag{5.3}$$

Specifying Alternative Management Actions

The annual decision for Denali NP managers is how many potential nesting sites to restrict to park visitors. The nesting area believed to be affected by human recreation contains 90 potential nesting territories. We have simplified the problem such that the number of sites restricted in any year (r_t) can take an integer value from 0 to 90. We do not consider the spatial location or arrangement of territories in determining the optimal level of restrictions, but recognize that it may not be realistic to restrict access to one territory independently of adjacent territories (i.e., trails might pass through multiple territories and, if closed, would naturally affect access to all territories they cross).

Developing Dynamic Models of System Behavior

We use a two-state occupancy model, simplified from previous analyses of this population (Martin et al. 2009a, 2011), to describe eagle dynamics in their nesting sites. The model links territory transition probabilities (extinction and colonization rates) with hypotheses about the effects of management actions on these dynamics. The number of occupied territories in a given year can be modeled as a Markov process:

$$N_{t+O}^{\mathrm{O}} = (N_t^{\mathrm{U}} \times \gamma) + \left[N_t^{\mathrm{O}} \times (1 - \epsilon)\right], \tag{5.4}$$

where N^{O} is the number of occupied territories, N^{U} is the number of unoccupied territories, γ is the probability that an unoccupied territory will be occupied the next year (colonization), and ϵ is the probability that an occupied territory will be unoccupied in the next year (local extinction). Simply put, this model states that the number of occupied sites in time $t+1$ depends on the number of unoccupied sites in year t that are colonized, plus the number of occupied sites that do not go extinct between year t and $t+1$. We modify the basic occupancy model to link the predicted impacts of our management actions to eagle occupancy dynamics (see Martin et al. (2011) for the three-state version of this model):

$$\begin{aligned} N_t^{\mathrm{O}} = {} & \frac{N_{t-1}^{\mathrm{U}}}{N^{\mathrm{tot}}}\left[r_t \gamma_R + \left(N^{\mathrm{tot}} - r_t\right)\gamma_{NR}\right] \\ & + \frac{N_{t-1}^{\mathrm{O}}}{N^{\mathrm{tot}}}\left[r_t\left(1 - \epsilon_R\right) + \left(N^{\mathrm{tot}} - r_t\right)\left(1 - \epsilon_{NR}\right)\right], \end{aligned} \tag{5.5}$$

where N^{tot} is the total number of available territories, r_t is the number of territories which are restricted to recreation, subscripts R and NR correspond to the anticipated effects of restricting and not restricting territory sites, respectively, on the probabilities of colonization and extinction. As we do not consider the spatial configuration of territories or location of restrictions, Eq. 5.5 makes the assumption that once the number of site restrictions is determined, they are applied without regard to the occupancy status of a territory. This is a simplified but realistic approach because we assume that decisions on the number of site restrictions will often have to be made prior to ascertaining the occupancy status of territories in the study area.

Predicting nesting-site occupancy in Eq. 5.5 is contingent on estimating occupancy transition parameters, γ and ϵ. Martin et al. (2009a) estimated nesting and reproductive transition probabilities for eagles using 20 years of nest survey data. They tested for the expected effects on eagle occupancy dynamics of disturbance and environmental variables such as nesting-site elevation and snowshoe hare (*L. americanus*) abundance. Here, we extend this work by considering the possibility that a specific level of snowshoe hare abundance may constitute an ecological threshold related to patch extinction or colonization probabilities. As system dynamics are not known with certainty, we account for this uncertainty by presenting multiple hypotheses regarding the functional relationship between hare abundance, disturbance to nesting sites, and the parameters that govern eagle occupancy. To simplify the problem for illustrative purposes, we offer four a priori models to represent possible relationships between environmental variables (prey abundance), management alternatives (minimizing disturbance through restricting access), and conservation objectives (maintain site occupancy and recreational opportunities). In this case, our initial model (*Model 1*) hypothesizes a negative relationship between hare abundance and the probability of local patch extinction and a positive, additive effect of both hare abundance and reduced disturbance on patch colonization. *Model 2* represents a no-effect model predicting that neither extinction nor colonization probabilities are influenced by snowshoe hare abundance or restricting access to eagle territories. *Model 3* hypothesizes that hare abundance has no effect on occupancy dynamics, but human disturbance negatively affects the probability of colonization. *Model 4* posits the existence of an ecological threshold, where values of colonization and extinction are predicted to differ above and below a given hare abundance level. While the structure of *Model 4* could take many forms, we offer one hypothetical example where transition parameters are as follows:

$$\begin{aligned} logit\,(\gamma) &= \alpha + \beta_1 \times HareTH + \beta_2 \times Disturb, \qquad \text{and} \\ logit(\epsilon) &= \alpha + \beta_1 \times HareTH, \end{aligned}$$

where α's are intercepts and β's are slope parameters describing the relationship between covariates and probabilities of colonization (γ) and extinction (ϵ). As in *Models 1–3*, γ and ϵ are modeled as linear-logistic functions and converted to linear functions via the logit link (MacKenzie et al. 2006). In *Model 4*, the logit of colonization is modeled as a linear combination of human disturbance (where $Disturb = 0$ when access to a site is restricted, and 1 otherwise) and hare abundance relative to a given threshold, which is modeled as a binomial outcome (*HareTH*). Extinction probability is modeled as a function of threshold hare abundance only. *HareTH* is a dummy variable that takes the form

$$HareTH = \begin{cases} 0, & hare\,index \leq \tau_h \\ 1, & hare\,index > \tau_h \end{cases},$$

where *hare index* is a covariate for hare abundance measured annually and relevant to all sites, and τ_h is an ecological threshold value for the hare abundance index. In our example, we arbitrarily set $\tau_h = 0.07$. We simplify the analysis by modeling *hare index* as a random variable following a distribution based on expert opinion and observed hare fluctuations (mean = 0.12, SD = 0.11), but hare abundance can be modeled in a more realistic manner (see Martin et al. 2011).

We use the 20-year data set collected on nesting-site occupancy to provide initial measures of credibility (weights that sum to 1 for all members of the model set) for the four models. Occupancy modeling is implemented in PRESENCE 2.4 (Hines 2008), and model selection is based on Akaike information criterion (AIC; Burnham and Anderson 2002). AIC weights are then used as relative measures of confidence in each candidate model when determining the optimal management decision for any level of site occupancy (see next section). A directed monitoring program allows us to reduce the structural uncertainty represented by competing models and adjust model weights as empirical evidence supports one or more models over the others (Williams et al. 2002).

Optimal Decision-Making and Simulations

The aim of this analysis is to select the optimal number of nesting sites to restrict each year in order to meet management goals defined through our objective function: maximizing recreational opportunities while achieving a minimum threshold level of site occupancy. We use the general expression for system dynamics (Eq. 5.5) to discriminate among all possible management alternatives at each time step and select the number of nesting sites to restrict each year that is expected to provide the optimal long-term benefit given the current state of the system. We consider this a Markov decision process because annual occupancy state is modeled as dependent on the state in the previous year. Uncertainty in system dynamics must be accounted for in decision-making and is represented here by differences in the predictions of competing models (*Models 1–4*). The optimal, state-dependent decision is then obtained by means of a passive optimization algorithm, which accounts for the uncertainty (weight associated with each model) via weighted model averaging. Initial model weights are based on AIC values from the model selection process and used to average the expected return from each of the four models. In an actual management situation, monitoring would follow the decision at each time step and provides the ability to learn about the system by confronting model predictions with observations. Model weights would then be updated via Bayes' theorem to reflect the new confidence in one or more models, resulting in improved predictions and better management decisions (see Williams et al. 2002). This approach is considered one of passive adaptive management, as the evolution of model weights is not accounted for over the time horizon of the optimization (Williams et al. 2002). We calculated the optimal sequence of state-dependent decisions using stochastic dynamic programming, based on the Principle of Optimality (Bellman 1957) and implemented in ASDP v3.2 (Lubow 2001). Stochastic dynamic programming iterates backwards from some future time and aggregates long-term benefits to the current return obtained by the decision made in the present time step (Williams et al. 2002). We ran the dynamic model for a maximum of 350 iterations, until a stable decision policy was reached and maintained over 15 consecutive iterations.

We simulated annual eagle occupancy levels predicted through implementation of the optimal decision policies under each of the four models as representing the "true" behavior of the system. To assess the value of selecting optimal annual restriction levels, we compared this policy to alternative suboptimal decision scenarios including a fixed policy of no management and that of restricting all sites to recreation. Under the belief that each model, in turn, represents the best hypothesis of system dynamics, we also simulated the evolution in model weights to explore the reduction of uncertainty over time.

Decision Thresholds and Sensitivity

As described earlier, decision thresholds are the products of the SDM process, resulting from interactions between the objective function (including utility thresholds), the predictions of system dynamics models (including identified ecological thresholds), the set of decision alternatives available, and the optimization procedure. Thus, a decision threshold is a location in state space where the optimal management action shifts from one alternative to another. This change occurs as a function of the predicted effects of management decisions on those state variables and the desired outcome as expressed through the objective function. Uncertainty in system dynamics, and therefore, in the response of the system to management, reduces the returns expected to result from optimal decisions because those decisions are made with incomplete understanding of the system. In order to assess the importance of uncertainty to management decisions, we can investigate the sensitivity of the optimal decision to the uncertainty inherent in our models. If the competing models all lead to the same management actions for a point or region of state space, then the decision is said to be "robust" to uncertainty (Regan et al. **?** . In this situation, there is no advantage to try to reduce structural uncertainty. In addition to structural uncertainty related to models and to possible ecological thresholds, we also evaluate the sensitivity of decision thresholds to our selection of utility threshold values.

Results

Occupancy Dynamics of Golden Eagles in Denali National Park

Using the simplified set of four competing models describing the dynamics of eagle occupancy, *Model 3* (no hare effect; human disturbance influences colonization) best explained the process underlying 20 years of nesting-site observations in Denali NP with an AIC weight of 0.74 (Table 5.1). *Models 1* and *4*, hypothesizing that hare abundance influences both colonization and extinction probabilities either linearly or beyond an ecological threshold, both received some support in the model selection process ($w = 0.14$ and 0.11, respectively; Table 5.1) and, therefore, should be considered as plausible models for explaining system dynamics. The no-effect model (*Model 2*) received virtually no support. Parameter estimates for model coefficients were in the expected directions for covariables with disturbed (unmanaged) sites showing reduced colonization probability and increased hare abundance enhancing colonization and reducing extinction probabilities (Table 5.2). Using the coefficient estimates from the best-supported model (*Model 3*), we slightly modified parameter values for the remaining models such that equilibrium occupancy ($\psi^* = \gamma/[\gamma + \epsilon]$) was approximately equal across all models under conditions of an undisturbed site

Table 5.1 Occupancy model selection for Golden Eagles in Denali National Park, USA, using Akaike information criterion

Model		ΔAICc	*w*	*K*
Model 3	$\psi(1)\epsilon(.)\gamma(\mathit{Disturb})p(t,.)$	0	0.74	13
Model 1	$\psi(1)\epsilon(\mathit{Hare})\gamma(\mathit{Disturb}+\mathit{Hare})p(t,.)$	3.32	0.14	15
Model 4	$\psi(1)\epsilon(\mathit{HareTH})\gamma(\mathit{HareTH}+\mathit{Disturb})p(t,.)$	3.74	0.11	15
Model 2	$\psi(1)\epsilon(.)\gamma(.)p(t,.)$	15.6	0.01	12

Model parameters included probability that a site was occupied in the first study season $[\psi(1)]$, the probability of a site becoming unoccupied if occupied in the previous year (extinction, ϵ), the probability of an unoccupied site becoming occupied in the following year (colonization, γ), and the probability of detecting nesting eagles, conditional on the site being occupied (p). We estimated initial occupancy as constant; extinction and colonization probabilities were modeled variously as functions of human recreational activity at the nesting site (*Disturb*), of prey availability as related linearly (*Hare*), of prey availability functioning as an ecological threshold (*HareET*), or as constant (.). Detection probability was modeled as varying among years but constant within a given year *p(t,.). AICc*: Akaike's information criterion corrected for small sample sizes, *ΔAICc*: for the *i*th model is computed as $AICc_i - \min(AICc)$; *w*: AICc weight; *K*: number of parameters

Table 5.2 Coefficient estimates for covariate parameters included in occupancy dynamic models

Parameter	Coefficient	Model 1	Model 2	Model 3	Model 4
γ	α_0	– 0.880	– 0.85	– 0.740	– 0.770
	$\beta_{1(\mathit{Hare,\ HareTH})}$	1.000	–	–	0.150
	$\beta_{2(\mathit{Disturb})}$	– 1.390	–	– 1.380	– 1.378
ϵ	α_0	– 2.770	– 2.854	– 2.843	– 2.822
	$\beta_{1(\mathit{Hare,\ HareTH})}$	– 1.040	–	–	– 0.085
	$\beta_{2(\mathit{Disturb})}$	–	–	–	–

The structure for the four models is provided in Table 5.1. Coefficients for linear predictors of colonization (γ) and extinction (ϵ) probabilities include an intercept (α_0), prey abundance (β_1), and human disturbance at a nesting site (β_2). Under *Model 4* the beta coefficient for *HareTH* relates to a dummy variable signaling prey abundance above or below a threshold of 0.07

at average hare abundance. Applying coefficient values from Table 5.2, a graphical representation of the four competing models is provided in Fig. 5.1.

Optimization

We set the utility threshold to $\tau = 75$ out of 90 potential nesting sites, based on the objectives (values) of the decision-makers in Denali NP and on historical occupancy levels. Assuming, sequentially, that each of the four models approximates "truth," we determined the optimal decision in a given year for each value of the occupancy state variable (Fig. 5.2a). Differences in the recommended decision at any point in the state space, depicted in Fig. 5.2, demonstrates the relevance of structural uncertainty to the optimal decision. As expected, the optimal decision under the no-effect model (*Model 2*) is to restrict none of the sites at any level of occupancy because limiting human disturbance has no impact on future occupancy. For the remaining models,

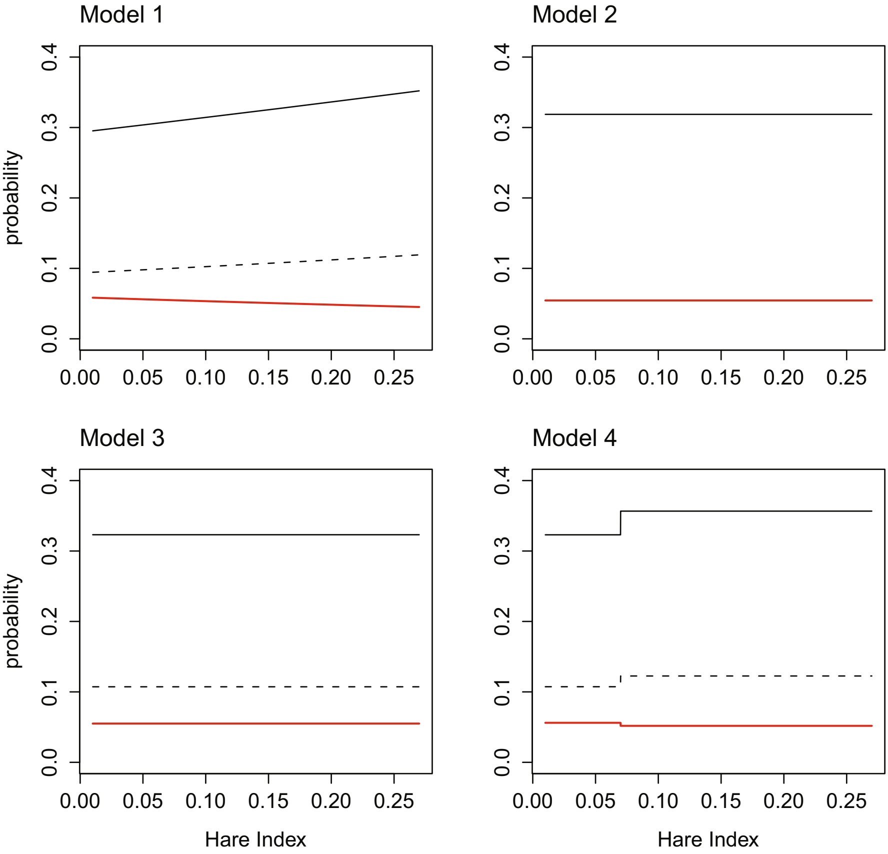


Fig. 5.1 Hypothetical models representing alternative functional relationships between snowshoe hare abundance, management, and occupancy dynamics for Golden Eagles at potential nesting sites. *Black lines* are local colonization probabilities, with *dashed lines* representing unmanaged sites and *solid lines* representing sites at which disturbance is reduced by restricting hiker access to nesting areas. *Red lines* are local extinction probabilities. *Model 1* is considered the global model, predicting that both human disturbance and snowshoe hare abundance influence colonization probability and that hare abundance affects extinction. *Model 2* represents a no-effect model in which neither hare abundance nor disturbance affects site occupancy. *Model 3* hypothesizes that disturbance influences colonization, but that hare abundance has no effect on occupancy dynamics. *Model 4* depicts a hare index value of 0.07 as an ecological threshold, with different colonization and extinction probabilities above and below this hare index. This model also includes a human disturbance effect

restrictions begin at occupancy levels below 80 territories and quickly increase until all sites are restricted to human access as soon as site occupancy falls below the threshold value (Fig. 5.2a). Note that for any level of site occupancy, the optimal management decision under *Model 1* or *Model 4* (both of which incorporated hare abundance as a random variable) is to restrict more sites than under *Model 3* (which included no stochastic component). Management decisions made under models that

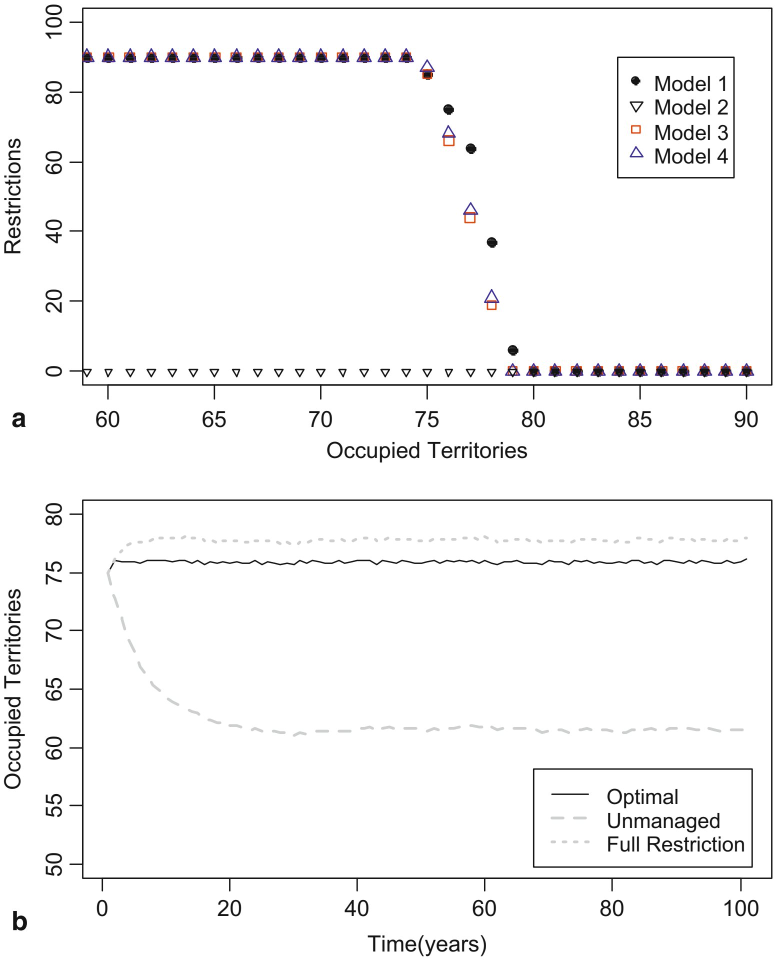


Fig. 5.2 **a** A decision rule depicting the optimal state-dependent management decision at each level of the state variable (nesting-site occupancy) for each of the four models under evaluation. *Model 2* predicts no influence of disturbance on occupancy dynamics and, therefore, recommends no management action taken at any level of occupancy. Decision thresholds are strongly influenced by the utility threshold in this scenario ($\tau = 75$ occupied sites) and occur over the range of 75–79 occupied territories for *Models 1*, *3*, and *4*. **b** Simulations (average of 10 iterations) under *Model 4* of Golden Eagle occupancy levels following optimal (*solid line*) and suboptimal (*dotted* and *dashed lines*) decisions. Suboptimal decisions included restricting hiking in all 90 eagle territories each year and, alternatively, opening all nesting territories for human recreation access

incorporate stochastic elements are expected to be more conservative than if based on deterministic models. The optimization algorithm anticipates the expected loss in return from periodically falling below the utility threshold and, therefore, recommends greater site restrictions in order to maintain average occupancy levels above those predicted by a deterministic model.

Figure 5.2b illustrates simulated levels of occupancy predicted under a single stochastic model (*Model 4*) when following the optimal decision policy as compared to maintaining fixed, suboptimal policies of no management and, alternatively, restricting human access to all eagle nesting territories. Equilibrium occupancy was approximately 0.66 under a policy of no park management; this is compared to average occupancy of 0.84 by following the optimal decision policy. The added benefit to eagles resulting from a policy of restricting all sites to human activities was determined to be minimal under *Model 4*, with the probability of occupancy increasing only 2 %, to 0.86. The costs (i.e., expenditure of resources and denying recreational benefits to park visitors) of such intensive management obviously outweigh the slight gains in eagle occupancy. Indeed, using the current objective function (Eq. 5.2) to quantify the return of implementing each of these management policies, we determined that both of the suboptimal approaches resulted in significantly lower annual utility values ($\bar{U}_t \cong 0$) than the optimal policy under *Model 4* ($\bar{U}_t = 20.7$, SD $= 8.0$).

Model Uncertainty

We evaluated the relevance of structural uncertainty to system dynamics by simulating the predicted eagle response to optimal decision policies under each of the four models and comparing the outcomes. Although "truth" was represented by a single model in each simulation, optimal decisions at each time step were determined by incorporating structural uncertainty (represented by the distribution of model weights) in the optimization. Beginning with an initial occupancy of 75 out of 90 nesting territories, we simulated the sequence of decisions and predicted consequences over a 100-year period (Fig. 5.3a, b). The variability observed in both occupancy and restriction policy predicted by *Models 1* and *4* is attributable to fluctuating prey populations and illustrates the influence of environmental variation on decision processes (Fig. 5.3a, b). The uncertainty of environmental variation (random variation in hare abundance under *Models 1* and *4*) produces higher average occupancies than deterministic *Model 3*, which is held at the utility threshold value. Put another way, in order to maintain eagle occupancy above the utility threshold in the face of environmental variation, the optimal policy in a stochastic system is to manage for somewhat higher occupancy levels in order to avoid declines below the utility threshold in years of low hare abundance. Occupancy under deterministic *Model 2* is unaffected by hare abundance or management actions and remains at an equilibrium of 0.85 (76.1 sites occupied; Fig. 5.3a).

Management policies (temporal variation in number of sites restricted) were a function of initial model weights, the time required to "learn" which model was most appropriate for the system, and the predicted occupancy state. As annual decisions were a function of predicted occupancy state, the average decision policies under the two stochastic models were slightly less conservative than *Model 3* due to the higher levels of occupancy maintained under stochastic dynamics (Fig. 5.3b). Occupancy under *Model 1* was more variable and observed to fall below the threshold more

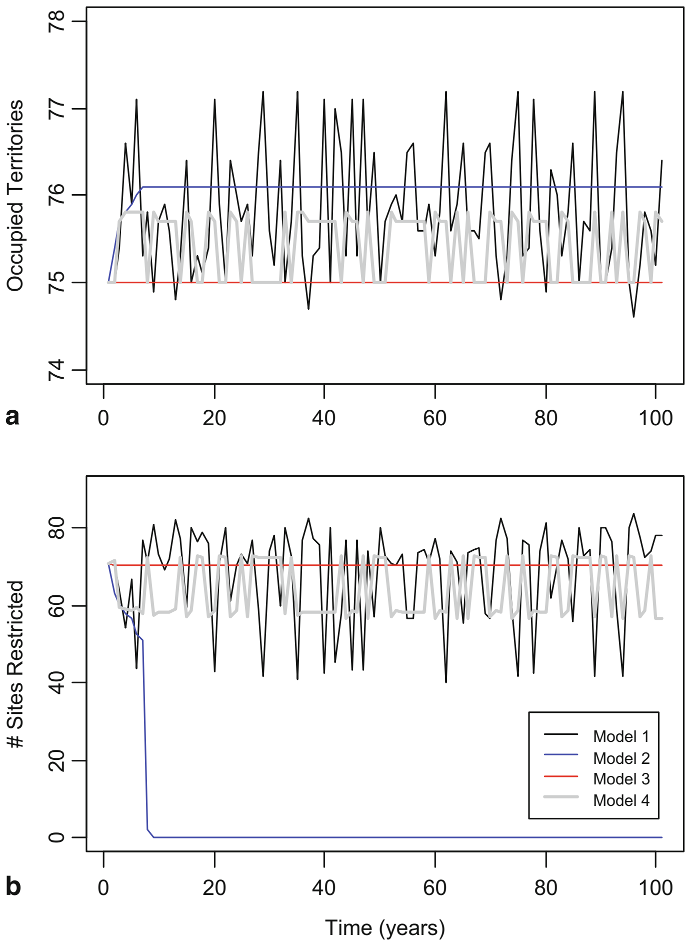


Fig. 5.3 A 100-year simulation (single realization) of **a** predicted system response (occupancy) and **b** the optimal decision policies under each of four competing models. All simulations were performed with model weight based on the AIC model selection (Table 5.1), allowing model weights to evolve over time, but where each model, in turn, represented "truth." *Model 3* is held at the utility threshold because it predicts occupancy dynamics as deterministic and settles on a stable decision policy almost immediately. Stochastic *Models 1* and *4* must maintain occupancy levels above the threshold value to reduce the chances that occupancy falls below the desired level

frequently than under *Model 4*. The decision policy, therefore, under *Model 1* showed greater variability and was more conservative (mean annual restrictions = 68.4 sites) than under *Model 4* (mean annual restrictions = 63.8 sites). The initial weight of *Model 2* was very low (initial $w = 0.01$), and thus updating of model weights required several years before the influence of the other models on the annual decision finally abated and the appropriate action (no restriction) under this model was selected (Fig. 5.3b). We illustrate the evolution of model weights more directly in Fig. 5.4, in which system dynamics were simulated under *Model 4*. We note that the weight

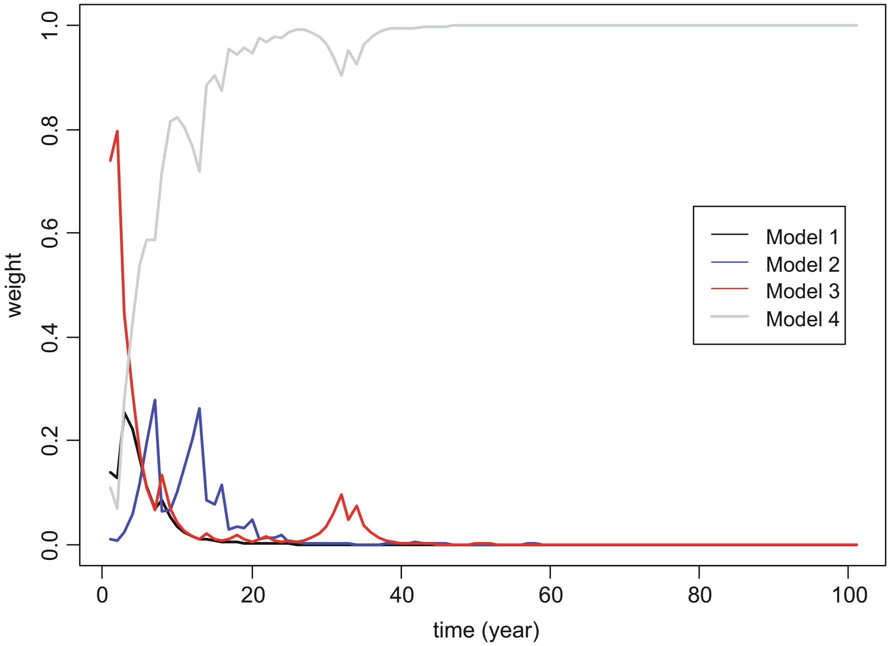


Fig. 5.4 Simulation (single realization using a passive adaptive optimization) of the change in model weights over time for the four models included in the model set, under the assumption that *Model 4* (threshold model) represented "truth."

for the initial highest-ranking model (*Model 3*) rapidly drops and that weights for *Models 1* and *3* share similar evolutions due to their comparable structure (Table 5.1).

Structural Uncertainty and Sensitivity of Decision Thresholds

While the sensitivity of decision thresholds to structural uncertainty under *Models 1*, *3*, and *4* is relatively low (Fig. 5.2a), the magnitude of differences in decisions made under each of these models in the simulation was significant, suggesting that reducing uncertainty in the structural dynamics of eagle occupancy would be valuable for managing the species (Fig. 5.3b). The impact of uncertainty on decision-making is most apparent when considering the potential for differences in optimal management response under the "no-effect" model (i.e., if *Model 2* is determined to be closest to "truth," no sites are restricted) relative to the level of restriction under the other models in our model set (Figs. 5.2a and 5.3b). Although managers will account for uncertainty at any point in time by weighting the consequences predicted under each model by its relative degree of support and selecting that management action determined to be optimal (results not shown), the range of possible decision thresholds under the four

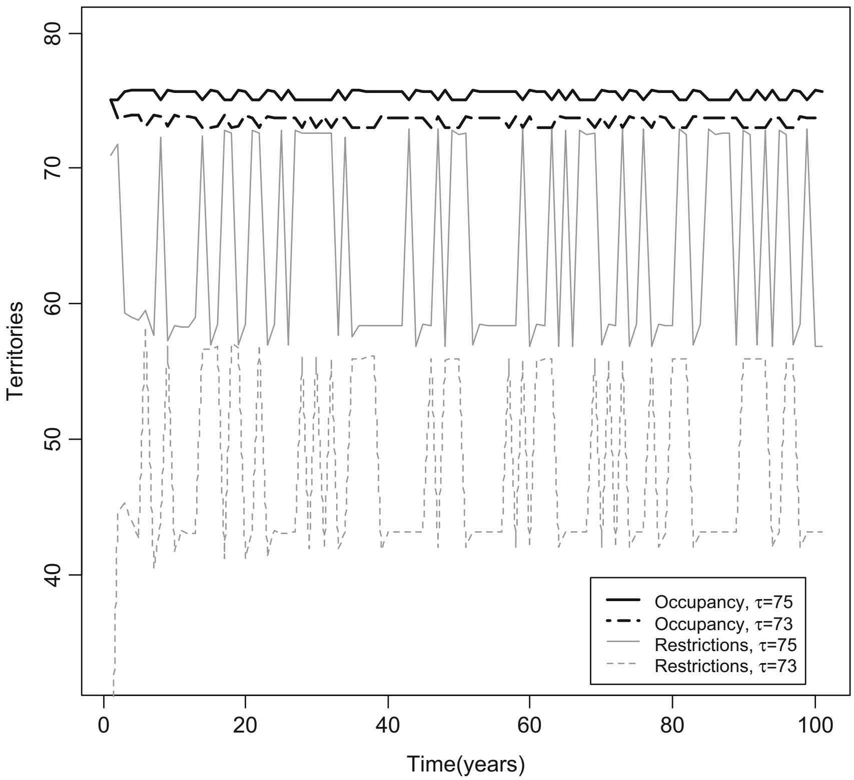


Fig. 5.5 Simulations (single realization) demonstrating the sensitivity of the decision threshold to utility thresholds ($\tau = 75$ and 73). Simulations included all models (weight distribution based on AIC model selection and allowed to evolve over time), but data depict the differences in occupancy and decision policies under *Model 4* as representing "truth." A reduction of the utility threshold from 75 to 73 resulted in a slight decline in average occupancy (*dark lines*) but a much larger reduction in average management actions (*gray lines*)

models in our set demonstrates the degree to which management could be improved by resolving this uncertainty.

In addition to evaluating the sensitivity of decision thresholds to model uncertainty, we can also examine the impact on decisions resulting in changes to the utility threshold. Evaluating changes in optimal management decisions when small changes are made to utility threshold values may be useful if management objectives are expected to change or evolve over time. For example, by reducing the utility threshold of desired occupancy level from 75 to 73 territories, we observe only a slight decline in average occupancy under *Model 4*, whereas the decision threshold was highly sensitive to this change. With this small change in utility threshold, we observed a substantial reduction in the average optimal number of sites to restrict, dropping from 63.9 (SD = 7.3) to 47.4 (SD = 7.1) territories per year (Fig. 5.5).

Discussion

The example we presented here originated from an actual case study, but was simplified to illustrate the three types of thresholds—ecological, utility, and decision—and to demonstrate how such thresholds might be included under an SDM framework in the management of natural resources. Following the operational definitions outlined by Martin et al. (2009c) and Nichols et al. (Chap. 2), utility thresholds were derived from the values of managers or stakeholders and can be incorporated explicitly into the management objectives via an objective function. Utility thresholds specify which values of state variables are viewed as desirable and undesirable and can result in changes to management when the system state approaches undesirable levels. Ecological thresholds, as the name suggests, represent biological phenomena and are the values of system state variables or environmental drivers where small changes result in either substantial changes to system dynamics or cause state variables or other parameters to reach specified levels. As such, ecological thresholds are important when considering the predictions of system response to management actions (or other changes in state variables) and should be included in system models. Decision thresholds are a product of the decision-making process and can formally be derived from the objective function, which may include utility thresholds, and from the models of system dynamics, which may include ecological thresholds.

In our example concerning human disturbance and nesting Golden Eagles in Denali NP, Alaska, we developed an objective function that accommodated two competing objectives: permitting recreational opportunities for human visitors to the Park, while concurrently maintaining what is believed to be an appropriate level of eagle nesting-site occupancy. We treated one objective, eagle occupancy, as a utility threshold which acted as a constraint on the remaining objective of recreational opportunities. Specifically, the objective function sought to maximize the number of potential eagle nesting sites at which hiking was permitted, subject to the constraint that eagle occupancy was maintained above the level specified by the utility threshold.

The concept of an ecological threshold is illustrated in our example by a single hypothesis describing the relationship between nesting-site occupancy dynamics and the abundance of a specific eagle prey item, snowshoe hares. This hypothesis, with its corresponding threshold, is incorporated into our set of potential models and, thus, represents uncertainty in system dynamics that can be confronted with data and reduced over time via an adaptive management strategy (Williams et al. 2002). By incorporating ecological threshold hypotheses into competing models that are relevant to the predicted effects of management actions on system dynamics, we focus our attention on those biological hypotheses that are most applicable to our stated management objectives. We used a model selection process to evaluate whether human disturbance and hare abundance are likely to influence the colonization and extinction probabilities of nesting-site occupancy. After the no-effect model (*Model 2*), which received virtually no support, the threshold model (*Model 4*) received the least amount of support ($w = 0.11$; Table 5.1). In adaptive management, however, optimal decisions are not based solely on the top-ranking model, but instead consider the predictions of all plausible descriptions of system dynamics, weighted by our relative confidence in each model (i.e., multimodel decision-making). In our example, weights were derived from a prior analysis, but they can also be based on "expert opinion," by consensus of stakeholder groups or other means. We then make the best decision recognizing that uncertainty exists. As we select management actions and monitor the response of the system, we learn about system dynamics and update our relative confidence in each model. Such a process leads to the accumulation of knowledge based on science and allows us to concurrently improve decision-making.

The decision thresholds in the Golden Eagle example were found to be highly sensitive to the value of the utility threshold. Lowering the utility threshold by only two sites resulted in a significant reduction in the average number of management restrictions imposed each year. Conducting such a sensitivity analysis provides decision-makers a tool with which to analyze the consequences of value judgments and evaluate the costs and benefits of their decision policies. Except in the case of the no-effect model (*Model 2*), the decision thresholds were moderately insensitive to the uncertainty associated with the remaining system dynamics models although the potential benefit from resolving this uncertainty may be significant. By directing our attention explicitly to those areas of uncertainty that have the greatest impacts on the management decision, our analysis allows us to reduce the complexity of the problem (and unnecessary impediments to decision-making) by choosing to ignore the many additional uncertainties that have little or no influence on decision thresholds. Although the decision thresholds were relatively robust to uncertainty in the three effect-models (*Models 1*, *3*, and *4*), the simulated optimal decision policies anticipated by these models were affected by treating the hare index as a stochastic random variable. Martin et al. (2011) discuss in greater detail the various approaches to handling environmental covariables, such as prey abundance, and possible consequences to the decision optimization.

Misconceptions about, or failure to distinguish among, utility, ecological, and decision thresholds has likely obstructed efforts to understand the roles and impact of thresholds on decision-making in conservation. The SDM framework, as we have outlined here, appears to serve as a natural and appropriate mechanism for clarifying and applying specific threshold concepts in the context of natural resource management. We hope that our example encourages managers to think carefully about their objectives and to be explicit when considering the incorporation of thresholds into their decision-making process.

References

Andersen, T., J. Carstensen, E. Hernandez-Garcia, and C. M. Duarte. 2009. Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evolution* 24:49–57.

Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multi-model inference: A practical information-theoretic approach*. 2nd ed. New York: Springer.

Bellman, R. 1957. *Dynamic programming*. Princeton: Princeton University Press.

Clemen, R. T. 1996. *Making hard decisions: an introduction to decision analysis*. Pacific Grove: Duxbury Press.

Clemen, R. T., and T. Reilly. 2001. *Making hard decisions with decision tools*. Pacific Grove: Duxbury Press.

Fahrig, L. 2001. How much habitat is enough? *Biological Conservation* 100:65–74.

Hines, J. E. 2008. PRESENCE2—Software to estimate patch occupancy and related parameters. USGS-PWRC. http://www.mbr-pwrc.usgs.gov/software/presence.html. Accessed 25 Jan 2012.

Kochert, M. N., K. Steenhof, C. L. McIntyre, and E. H. Craig. 2002. Golden Eagle (*Aquila chrysaetos*). In *The birds of North America*, eds. A. Poole and F. Gill. Philadelphia: The Birds of North America Inc.

Lande, R. 1987. Extinction thresholds in demographic-models of territorial populations. *American Naturalist* 130:624–635.

Lubow, B. C. 2001. *Adaptive stochastic dynamic programming (ASDP) v3.2*. Fort Collins: Colorado State University.

Lyons, J. E., M. C. Runge, H. P. Laskowski, and W. L. Kendall. 2008. Monitoring in the context of structured decision-making and adaptive management. *Journal of Wildlife Management* 72:1683–1692.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, J. E. Hines, and L. L. Bailey. 2006. *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. San Diego: Elsevier.

Martin, J., J. D. Nichols, C. L. McIntyre, G. Ferraz, and J. E. Hines. 2009b. Perturbation analysis for patch occupancy dynamics. *Ecology* 90:10–16.

Martin, J., C. L. McIntyre, J. E. Hines, J. D. Nichols, J. A. Schmutz, and M. C. MacCluskie. 2009a. Dynamic multistate site occupancy models to evaluate hypotheses relevant to conservation of Golden Eagles in Denali National Park, Alaska. *Biological Conservation* 142:2726–2731.

Martin, J., M. C. Runge, J. D. Nichols, B. C. Lubow, and W. L. Kendall. 2009c. Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications* 19:1079–1090.

Martin, J., P. L. Fackler, J. D. Nichols, M. C. Runge, C. McIntyre, B. Lubow, M. C. McCluskie, and J. A. Schmutz. 2011. An adaptive-management framework for optimal control of hiking near golden eagle nests in Denali National Park. *Conservation Biology* 25:316–323.

McCarthy, M. A., and H. P. Possingham. 2007. Active adaptive management for conservation. *Conservation Biology* 21:956–963.

McGowan, C. P., and T. R. Simons. 2006. Effects of human recreation on the incubation behavior of American Oystercatchers. *Wilson Journal of Ornithology* 118:485–493.

Morse, J. A., A. N. Powell, and M. D. Tetreau. 2006. Productivity of Black Oystercatchers: Effects of recreational disturbance in a national park. *Condor* 108:623–633.

Nichols, J. D., M. J. Eaton, and J. Martin. In press. Thresholds for conservation and management: Structured decision making as a conceptual framework. In *Application of Threshold Concepts in Natural Resource Decision Making,* ed. G. Gunterspergen New York:Springer.

Regan, H. M., Y. Ben-Haim, B. Langford, W. G. Wilson, S. J. Andelman, and M. A. Burgman. 2005. Robust decision-making under severe uncertainty for conservation management. *Ecological Applications* 15:1471–1477.

Samhouri, J. F., P. S. Levin, and C. H. Ainsworth. 2010. Identifying thresholds for ecosystem-based management. *PLoS ONE* 5:e8907. doi:10.1371/journal.pone.000890.

Steidl, R. J., and R. G. Anthony. 2000. Experimental effects of human activity on breeding Bald Eagles. *Ecological Applications* 10:258–268.

Swarthout, E. C. H., and R. J. Steidl. 2003. Experimental effects of hiking on breeding Mexican spotted owls. *Conservation Biology* 17:307–315.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. *Analysis and management of animal populations: modeling, estimation, and decision making*. San Diego: Academic Press.

Williams, B. K., R. C. Szaro, and C. D. Shapiro. 2007. *Adaptive management: The U.S. Department of the Interior Technical Guide*. Washington, D.C.: U.S. Department of the Interior.

Chapter 6
Monitoring for Threshold-Dependent Decisions

David R. Smith, Craig D. Snyder, Nathaniel P. Hitt and Paul H. Geissler

Abstract Management decisions are threshold-dependent if actions or determinations change when monitoring data indicate that a resource crossed a specified value (e.g., reference vs impaired conditions). In this chapter, we review the literature on monitoring for threshold-dependent decisions and illustrate how uncertainty and prior knowledge about resource condition may affect such decision thresholds. A critical consideration is whether monitoring is linked to specific management actions and models are available to predict the consequences of those actions on the resource condition. This consideration leads to a split between two different management and monitoring approaches; adaptive management with *targeted monitoring* or sequential evaluation of resource condition with *surveillance monitoring*. We compare and contrast these two types of monitoring with regard to threshold concepts, objectives, use of models, and incorporation of uncertainty. Both types of monitoring are being applied to natural resource management, and we cannot conceive of a time when all monitoring will be of only one type. The best strategy, in our view, is to be familiar with when and how to apply both.

Keywords Targeted monitoring · Surveillance monitoring · Ecological change · Utility value · Decision threshold

Introduction

As stated simply by Gerber et al. (2005), “monitoring is the systematic acquisition of information over time,” and the role of monitoring is to “gain information needed for management decisions.” It is common that management decisions depend on thresholds when actions or determinations change if the measurement of the resource crosses a specified value. Such *decision thresholds* define boundaries between states of resource condition, e.g., reference and impaired. So, we view threshold-dependent decisions to be synonymous with state-dependent decisions (Lyons et al. 2008; Martin et al. 2009).

D. R. Smith (✉) · C. R. Snyder · N. P. Hitt
U.S. Geological Survey, Leetown Science Center, 11649 Leetown Road, Kearneysville, WV, USA
e-mail: drsmith@usgs.gov

P. Geissler
684 Homestead Street, Lafayette, CO, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_6,

Several recent papers have presented important overviews of monitoring in natural resource management, which are particularly relevant to monitoring for threshold-dependent decisions within natural areas (Nichols and Williams 2006; Bennetts et al. 2007; Lyons et al. 2008; Fancy et al. 2009; Lindenmayer and Likens 2009). Many of the concepts and terminology are held in common by these presentations, but there are differences. Commonalities include the importance of clear objectives and the use of conceptual models to guide monitoring program design. However, important differences involve whether and how information gained from monitoring links to specific a priori management actions.

The critical difference is whether there are specific management actions coupled with models to predict the consequences of the actions on the resource condition. This difference defines two management/monitoring approaches; either adaptive management with *targeted monitoring* or sequential evaluation of resource condition with *surveillance monitoring* (Fig. 6.1). In adaptive management, specific management actions have been identified, models have been developed to predict the consequences of those actions, and monitoring is targeted to acquire information necessary to make and test those predictions (Nichols and Williams 2006; Lyons et al. 2008). In sequential evaluation of resource condition, management actions are unspecified or listed as potential (Bennetts et al. 2007), and surveillance monitoring periodically assesses metrics that are indicative of the desired resource condition (Fancy et al. 2009).

Both types of management and monitoring are being applied to natural resource management (Fancy et al. 2009; Lyons et al. 2008), and we cannot conceive of a time when all monitoring will be of only one type. Targeted monitoring is tied to adaptive management and the frequency of its application is increasing (Williams et al. 2007; Lyons et al. 2008; Martin et al. 2009). However, surveillance monitoring is quite common (Lindenmayer and Likens 2009), and a majority of applications within National Parks including the national I&M networks employ surveillance monitoring (Fancy et al. 2009). The best strategy, in our view, is to be familiar with both types of monitoring and to understand when and how to apply both. Towards that end, we review the recent papers to cross reference terminology and identify distinct requirements of both monitoring types.

Comparing and Contrasting Monitoring Types

Both targeted and surveillance monitoring begins with clearly stated objectives and conceptual models to describe the resources to be conserved and the important factors affecting those resources, such as environmental stressors. In addition, adaptive management requires an a priori list of management actions that can directly affect resource condition and analyses (models) to predict those effects (Martin et al. 2009). For example, regulating creel limits in a trout fishery may directly affect population abundance, and adaptive regulations can be based on optimizing sustainable harvest (Bain 1987; Bytnerowicz et al. 2002). In this example, monitoring would be targeted at the state variable, which is population abundance. Surveillance monitoring does not

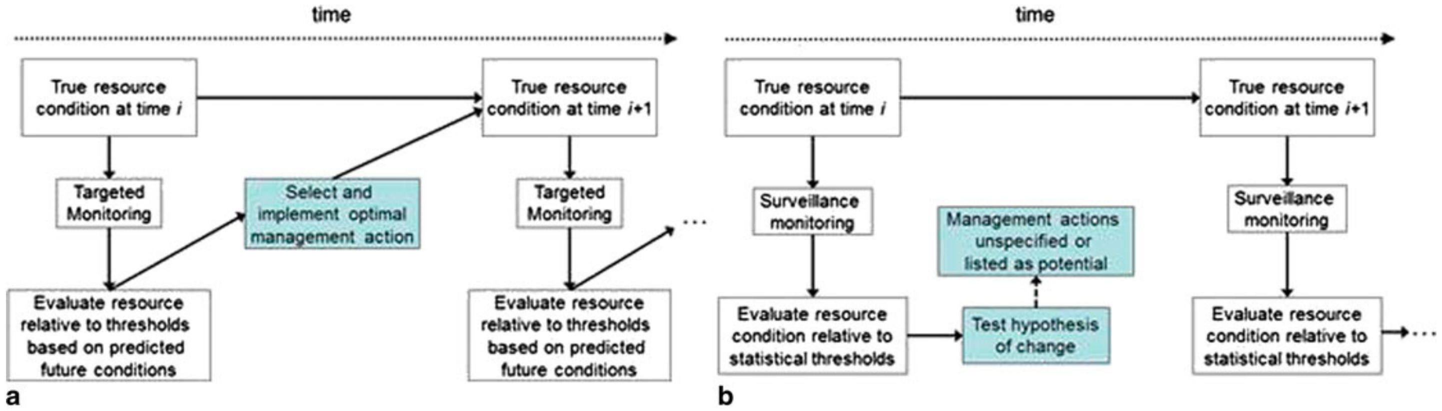


Fig. 6.1 Schematic of implementation of monitoring types through time. Monitoring entails sampling of the resource following a statistically valid design, and use of the data to evaluate resource condition. **a** Targeted monitoring leads to model-based predictions of future conditions and specific actions. Actions could affect future condition, and learning occurs by comparing model-based predictions to monitoring at the next time step. **b** Surveillance monitoring leads to inference from either a hypothesis test or updated probabilities of resource condition, followed by unspecified actions

require a priori management actions and predictive models. For example, detection of air quality changes is not linked directly to management actions (Bytnerowicz et al. 2002).

A basic timeline for monitoring calls for periodic evaluation of the resource (e.g., estimate abundance) and comparison to decision thresholds (Fig. 6.1). If the resource condition has crossed a decision threshold, then management actions are implemented depending on management and monitoring type. Targeted monitoring is integrated into an adaptive management framework, so that the action that is predicted to be best at meeting the management objectives is implemented and that action is expected to be optimal for maintaining desired future resource condition (Fig. 6.1a). In contrast, when conducting surveillance monitoring, periodic evaluation does not necessarily lead to implementation of any specific actions. Rather crossing statistical decision thresholds in surveillance monitoring could trigger a process to determine the best action, but is not linked to particular management actions. Bennetts et al. (2007) refer to these events (i.e., monitoring observation crossing a statistical threshold) as assessment points and argues for flexibility in implementation of management actions. In surveillance monitoring, there might be a list of potential actions, but there is not an a priori analysis to determine what the best action is to be taken given the circumstances. Early warning is emphasized in surveillance monitoring because managers need time to figure out what to do, which provides the motivation for multiple assessment points, as recommended by Bennetts et al. (2007). In contrast, adaptive management automatically incorporates an "early warning" feature by setting the decision threshold based on system response to management actions, which tends to increase the likelihood that the resource remains in the desired condition (Martin et al. 2009).

Unifying Threshold Concepts

Martin et al. (2009) identified three threshold concepts in resource management: ecological change, utility value, and decision threshold. These concepts are shared by the two management and monitoring types, but are incorporated into management differently (Table 6.1). Ecological change could include an ecological threshold, which is an abrupt change in resource over a relatively small change in stress load (Andersen et al. 2009; Swift and Hannon 2010). It is not required by either management type that an ecological threshold exists; however, ecological change in stressors is often assumed to underlie change in resource condition. In surveillance monitoring, ecological change is modeled conceptually. In targeted monitoring, ecological change is modeled quantitatively to predict response to specific adaptive management actions.

Utility values arise by quantifying management objectives (Keeney 1992). High utility value is placed on the desired resource condition, and utility value decreases when the resource condition is altered from the desired state. If resource condition is categorical (i.e., reference vs impaired), then utility value will change abruptly as the condition changes resulting in a utility threshold (cf. Martin et al. 2009). However,

Table 6.1 Unifying threshold concepts shared by approaches to manage and monitor natural resources

Threshold concept	Surveillance monitoring for sequential evaluation of resource condition	Targeted monitoring for adaptive resource management
Ecological change	Ecological change modeled conceptually to identify vital signs (Bennetts et al. 2007; Fancy et al. 2009)	Ecological change modeled quantitatively to predict consequence of management action (Martin et al. 2009)
Utility value	Value derived by quantifying the objectives, which depend on the risk of errors in classifying resource condition (Field et al. 2004)	Value derived by quantifying the objectives, which are linked to management action through predictive model(s) (Nichols and Williams 2006; Martin et al. 2009)
Decision or management threshold	Boundary used to classify resource condition: analogous to assessment points (Bennetts et al. 2007) or statistical thresholds (Field et al. 2004)	Boundary defining threshold-dependent management action. Determined by predicting consequences of management action and optimizing the expected utility value (Nichols and Williams 2006; Martin et al. 2009)

a utility threshold is not required. Utility value can change linearly or nonlinearly with measurable attributes depending on risk attitude (Keeney 1992). In sequential evaluation with surveillance monitoring, utility value can be based on the risk of errors in classifying resource condition (Field et al. 2004). In practice, utility value should be elicited from decision makers so as to accurately reflect their values and risk attitudes.

A decision or management threshold is common to both management and monitoring types, but is derived using different techniques (Table 6.1). In sequential evaluation with surveillance monitoring, the decision threshold is derived from Type I and II error rates for classifying resource condition (Field et al. 2004). This is consistent with our interpretation of the assessment point concept as outlined by Bennetts et al. (2007). In their discussion of how assessment points are determined, they consider "the level of uncertainty regarding the resource condition, and how conservative . . . to be in detecting a point of concern" (Bennetts et al. 2007). Decision thresholds or assessment points in sequential evaluation depend on uncertainty in measures of resource condition and the utility that managers put on risk of committing classification errors. Field et al. (2007) analyzed decision thresholds by minimizing cost of Type I and II errors. Also, Snyder et al. (Chap. 9) provide an example of determining decision thresholds by optimizing expected utility associated with Type I and II errors.

Compliance monitoring is a special case of surveillance monitoring where the decision threshold is set through regulatory or policy processes. However, these criteria can be consistent with a statistical threshold approach. For example, the Environmental Protection Agency (EPA) recommends use of a Data Quality Objectives Process to determine criteria for decision making that relies on specifying probability limits for false rejection and false acceptance decision errors, i.e., Type I and II error rates (EPA 2006).

In targeted monitoring for adaptive management, determination of decision thresholds is integrated with other components of a structured decision process (Lyons et al. 2008; Martin et al. 2009). First, objectives are identified, which reflect the value (i.e., utility) that the decision makers place on resource conditions. Second, management actions that could affect the resource condition are specified. Third, predictive models are developed from current understanding of ecological change, and uncertainty is incorporated into the models. Only at this point, with other components in place, can decision thresholds be determined by finding best management actions for a given resource condition. "Best" is often defined by maximizing the utility value, so that a management action that is predicted to result in the highest utility value compared to all other actions would be in that sense, the best. The role of monitoring is to provide information to evaluate the resource condition for comparison to the decision threshold and to the model predictions to learn about causes of ecological change (Fig. 6.1). Adaptive management incorporates all of the threshold concepts (Table 6.1), which was the main point made by Martin et al. (2009).

Objectives

Both management and monitoring types can share common objectives. Proponents of both targeted monitoring (Nichols and Williams 2006) and surveillance monitoring (Fancy et al. 2009) state that supporting and improving decision making is a primary objective. Nichols and Williams (2006) state that targeted monitoring is "integrated into conservation practice with design and implementation based on a priori hypotheses and associated models of system responses to management." However, a direct linkage to management actions is missing in surveillance monitoring, and possible management responses are described in general terms (Bennetts et al. 2007; Fancy et al. 2009). Although learning to improve management effectiveness is a commonly stated objective, the absence of a direct link to management actions causes surveillance monitoring to be inefficient at learning compared to targeted monitoring (Nichols and Williams 2006).

Use of Models

If improving decision making is of primary importance, then models must be integral to the decision process (Starfield 1997; Nichols and Williams 2006; Fancy et al. 2009). Models identify what needs to be monitored, and data gathered through monitoring can be compared to model predictions to improve decision making and facilitate learning. Model development provides an opportunity for collaboration between managers and researchers (Lindenmayer and Likens 2009).

In the development of a surveillance monitoring program, Fancy et al. (2009) described how conceptual models were used to identify the candidate vital signs for monitoring. In that application, conceptual models were designed to summarize the ecosystem function and structure with emphasis on the effects of stressors on resource condition. Bennetts et al. (2007) illustrate how the frequency of assessment points can depend on a model of ecological change through time. In that illustration, the frequency of assessment points would increase as the measure of resource condition approached the predicted ecological threshold. The framework presented by Bennetts et al. (2007) implicitly requires a model to predict the probability of the ecological threshold, so that as the predicted probability increases, so does the frequency of assessment points.

In the development of a targeted monitoring program, modeling and monitoring interact as components of a structured decision process (Lyons et al. 2008). The objectives of the decision problem determine both what is predicted by the model and what is monitored. For example, Lyons et al. (2008) present an example of wetland management. The objectives included maximizing waterbirds; thus, waterbird abundance was both predicted and monitored. The decision structure used in adaptive management sets up an efficient process for learning. Uncertainty in the form of multiple hypotheses regarding ecological change can be represented by multiple predictive models (Nichols and Williams 2006). Observations from monitoring can be compared to predictions from multiple models. The relative closeness or likelihood of model-based predictions to the observations from targeted monitoring provides evidence to differentiate between the underlying hypotheses. Although proponents of both surveillance and targeted monitoring discuss the role of models and monitoring in learning about the dynamics of ecological systems, it is only adaptive management with targeted monitoring that offers a specific process for learning.

Incorporating Uncertainty

Information from monitoring is subject to uncertainty from sampling errors resulting from the sampling process, model uncertainty resulting from competing hypotheses describing the response to ecological change, and natural variation resulting from environmental and demographic stochasticity (Regan et al. 2002). These sources of uncertainty influence both monitoring types. However, the manner and degree to which uncertainty is incorporated into the decision process varies between the monitoring types.

Much has been written about statistical sampling designs to reduce sampling error in monitoring programs (Jackson and Resh 1988; Overton and Stehman 1996; Vos et al. 2000; Yoccoz et al. 2001; Pollock et al. 2002; Stevens Jr. and Olsen 2004). The importance of probabilistic sampling to account for spatial and temporal variation, methods to estimate detectability, and hierarchical structure (cluster sampling, longitudinal sampling, and multiple stage sampling) in logistically feasible sampling designs have all been emphasized. Incorporating panel structures of sampling units to reduce spatial variation when estimating temporal trends have been recommended (Stevens Jr. and Olsen 1999). Spatial balance has recently emerged as a critical feature of effective sampling designs (Stevens Jr. and Olsen 2004). These issues apply regardless of monitoring type.

Sampling error and natural variation are variance components of the distribution of resource condition (Gould and Nichols 1998). Precision along with magnitude of ecological change determines sensitivity of surveillance monitoring to detect ecological change given a specific decision threshold. The traditional approach to optimizing design for surveillance monitoring has focused on sensitivity or statistical power (Jackson and Resh 1988; Taylor and Gerrodette 1993). In that approach, the decision threshold is derived from the reference distribution after first specifying a type I error rate (α). Then, the effect on sensitivity of other design aspects, such as sample size, is examined over an ecological change model. The design that maximizes sensitivity within allowable costs is found through statistical power analysis and then implemented. Statistical power analysis is a special case of maximizing utility value when all value is placed on correct classification of the impaired resource conditions and utility associated with the reference condition is given no value. In a power analysis, the utility associated with correct or incorrect classification of reference condition is implicit to the specification of type I error rate and thus the decision threshold is wholly a function of type I error rate. An alternative approach is to include the decision threshold as a design factor and optimize utility (or minimize loss) over the full range of resource conditions (Field et al. 2004).

In surveillance monitoring, a decision threshold (T_α) is the critical value for a resource measure (e.g., abundance, diversity) under the reference condition. In other words, the probability is α that a resource measure from the reference condition would fall below the decision threshold (T_α) in a sample of size n, i.e., $\alpha = F(T_\alpha \,|\Theta_0, n\,)$, where Θ_0 denotes the parameters for the distribution of the resource measure under reference condition. In general, the probability that a resource measure falls below the decision threshold (T_α) depends on the resource condition and sample size, i.e., $\Pr(y = 1) = F(T_\alpha \left|\Theta_{ij}, n\right.)$, where the variable y takes values of 0 or 1 to indicate the resource measure is above or below the threshold, Θ_{ij} denotes distribution parameters for the i^{th} resource condition at time j and $F(\cdot)$ is the cumulative distribution function. The distribution parameters, Θ_{ij}, are a function of ecological change. The utility value, $U(y, \Theta_{ij})$, depends on the observation (y) and on the resource condition (Θ_{ij}). An optimal design would be determined by the type I error rate and sampling design that maximizes the expected utility,

$$E[U] = \sum_i \sum_j [U(y = 0, \Theta_{ij}) \Pr(y = 0) + U(y = 1, \Theta_{ij}) \Pr(y = 1)]. \tag{6.1}$$

We present an example of optimizing utility value in a surveillance monitoring design based on freshwater bioassessment (Snyder et al. this volume). We used simulation to predict the probability that a metric of macroinvertebrate diversity would fall above or below a decision threshold over a 20-year period under three models of ecological change. Sample sizes included 100, 300, and 1,500 taxa identified per year. The

Table 6.2 Scenarios of future change to compute expected utility. Scenarios represent probabilities distributed among three change models: no change, gradual change, and threshold change

Scenario	No change	Gradual change	Threshold change
1: Likely to remain in reference	0.70	0.20	0.10
2: 1-to-1 odds of impairment	0.50	0.25	0.25
3: 2-to-1 odds of impairment	0.33	0.33	0.33
4: 3-to-1 odds of impairment	0.25	0.50	0.25

models of ecological change included (1) a no change model that remained in reference condition, (2) a gradual change model (linear), and (3) an abrupt change model (nonlinear). We analyzed two cases for assigning utility to resource classification. The first case assigned equal utility for classifying reference and impaired conditions; maximum value (1) was assigned to correct classification, and minimum value (0) was assigned to incorrect classification. The second case assigned a reduced value to correctly classifying reference condition relative to impaired condition; utility for correctly classifying reference was half (0.5) the utility for correctly classifying impaired condition. Otherwise, utilities for the other outcomes were the same for the first and second cases. We averaged expected utility (Eq. 6.1) across time and change models to calculate expected utility for a range of sample sizes and type I error rates. The average across change models was weighted to incorporate uncertainty among future change models (Table 6.2). The scenarios were selected to illustrate the concept of incorporating uncertainty in expected utility and do not represent any empirically based expected future change.

In this example, type I error rate ≤ 0.10 was only justified when the probability remaining in reference was high (Fig. 6.2a, e). Otherwise expected utility was maximized at type I error rates in excess of 0.10 and often ≥ 0.20. The type I error rate where expected utility was maximized increased as the probability of impairment increased among the scenarios of future change (cf. top to bottom panels in Fig. 6.2). The effect of sample size also increased as the probability of impairment increased. Expected utility increased with sample size (number of taxa identified per year), and expected utility was maximized at a higher type I error rate for lower sample sizes (Fig. 6.2). In general, low sample size should be compensated for by a high type I error rate.

In adaptive management with targeted monitoring, natural variation can be incorporated into the models used to predict consequences of management actions. Decision thresholds are based on these predictive models. Martin et al. (2009) show a case where inclusion of natural variation in predictive models results in more conservative thresholds than would be the case if uncertainty was ignored. Uncertainty in ecological change is incorporated as competing models for predicting response to management actions. These competing models are analogous to multiple hypotheses and provide for efficient learning about management effectiveness and underlying ecological relationships (Nichols and Williams 2006). Learning about which model or hypothesis is better at prediction and therefore deserving greater support or belief is accomplished by applying Bayes' theorem to update model weights (Dorazio and Johnson 2003; Martin et al. 2009).

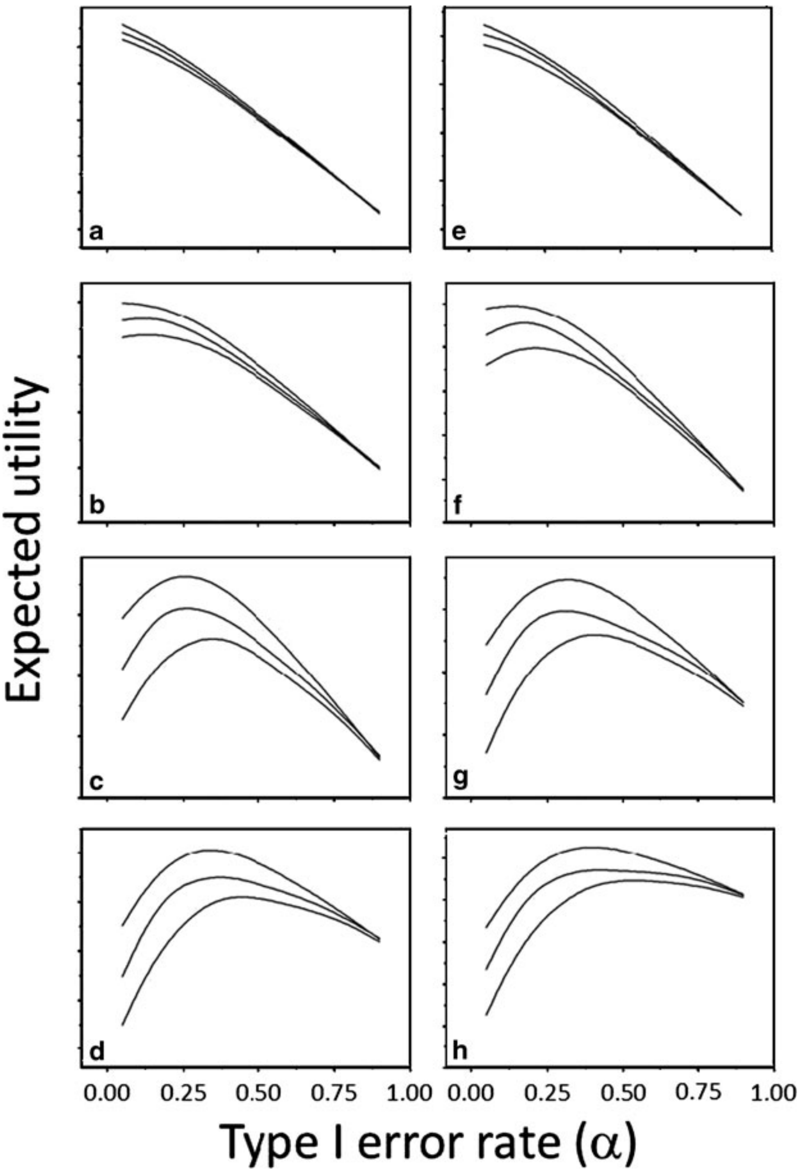


Fig. 6.2 Example of expected utility for two factors in the design of a surveillance monitoring program. The statistical decision threshold to classify the resource as reference or impaired was determined by type I error rate. Probabilities of falling above or below the decision threshold were based on freshwater bioassessment (Snyder et al. this volume). Two cases of utility are shown: (**a**–**d**) maximum utility (*1*) for correct classification as reference or impaired and (**e**–**h**) utility for correct classification of reference was half (*0.5*) the utility for correct classification of impaired condition. In both cases, utility was minimized (*0*) for incorrect classification of either reference or impairment. Expected utilities were averaged across time and scenarios of future change, which distributed probability across three possible change models (cf. Table 6.2). Plots **a** and **e** correspond to scenario 1, **b** and **f** to scenario 2, **c** and **g** to scenario 3, and **d** and **h** to scenario 4. *Lines* within plots are by sample size (taxa identified); *bottom line* is $n = 100$, *middle line* is $n = 300$, and *top line* is $n = 1{,}500$

Because surveillance monitoring lacks a direct connection to management actions (Fig. 6.1), a typical sequential evaluation of resource condition will result in a hypothesis test to determine if the resource measure is above or below a threshold (Jackson and Resh 1988). From one year to the next, in a hypothesis-testing approach, the resource measure could switch from above to below the threshold and back again even when the resource condition is degrading. An alternative approach applies Bayes' theorem to update prior probabilities of resource condition, which are specified at the beginning of the surveillance monitoring. This updating approach would accumulate evidence for or against change.

Let y_t denote the evaluation of resource condition in time t, such that $y_t = 1$ when the resource measure falls below the decision threshold, which occurs with probability $\Pr(y_t = 1) = F(T_\alpha \mid \Theta_t, n)$ and depends on type I error rate α and sample size n. The likelihood of observing y_t must be assumed or estimated, $\Pr(y_t \mid \Theta_t, \alpha, n)$. Then the updated probability of resource condition is

$$\Pr(\Theta_{t+1} \mid y_t, \alpha, n) = \frac{\Pr(y_t \mid \Theta_t, \alpha, n) \Pr(\Theta_t)}{\sum\limits_i \Pr(y_t \mid \Theta_i, \alpha, n) \Pr(\Theta_i)}. \tag{6.2}$$

An example of updating probabilities based on the freshwater bioassessment (Snyder et al. this volume) is presented in Fig. 6.3. There were three resource conditions: reference, early warning, and impaired. The decision threshold was based on the type I error rate of 0.2. The top panel shows evaluations of resource condition over 20 years resulting from a comparison of an index of macroinvertebrate diversity to the statistical decision threshold. At each year, the evaluation results in a binomial random variable indicating reference or nonreference condition. The middle panel shows updated probabilities for the three resource conditions (reference, early warning, and impaired). The initial prior probabilities were 0.9, 0.07, and 0.03, respectively. Even though year-specific evaluations in years 11, 13, and 14 indicated reference condition (i.e., the measure was above the threshold in each year), the updated probability of reference condition remained low because previous observations had indicated nonreference conditions. Thus, in the updating approach, the past observations inform the current evaluation. The bottom panel shows the expected probabilities over 1,000 replications of a simulated ecological change. The ecological change, which was a function of increasing landscape alterations, is also shown in the bottom panel.

Summary and Conclusions

Threshold-dependent decisions are common in natural resource management and may be significantly influenced by monitoring design considerations. In this chapter, we reviewed two fundamental approaches to monitoring, targeted and surveillance monitoring, and evaluated their conceptual and statistical distinctions. We do not intend to be neutral regarding the relative value of these two management and monitoring types. Adaptive management with targeted monitoring has greater potential

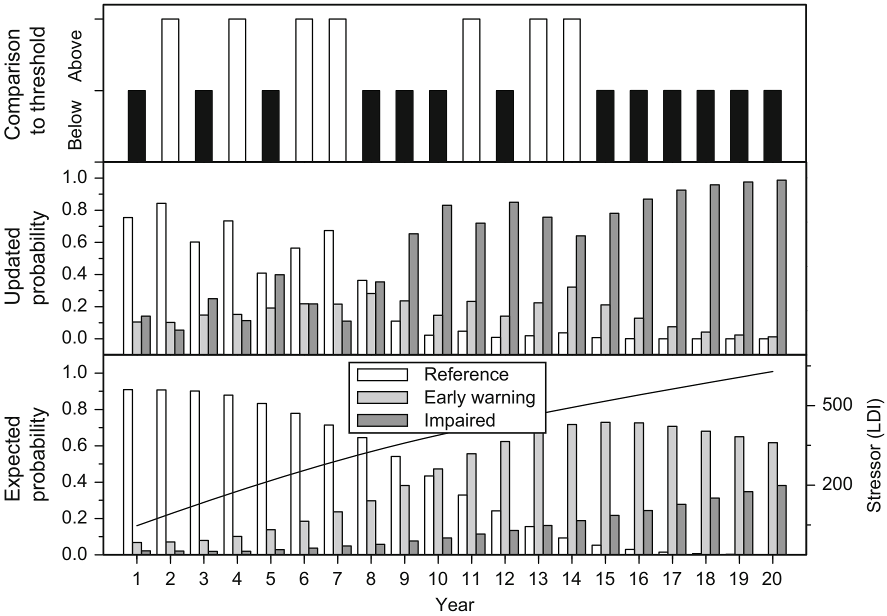


Fig. 6.3 Example of updating probabilities of resource condition. Resource conditions are based on level of stressor, in this case a Landscape Development Index (LDI). Conditions are *reference* (LDI < 50), *early warning* ($50 \leq$ LDI < 400), or *impaired* (LDI ≥ 400). *Top panel* shows sequential evaluation of resource condition over 20 years. Each year a bioassessment metric is compared to a decision threshold; above the threshold indicates reference and below indicates impaired. *Middle panel* shows sequential updated probability of resource condition based on observations from the *top panel*. The initial probabilities were *reference* (0.90), *early warning* (0.07), and *impaired* (0.03). *Bottom panel* shows expected probability of resource condition if observations were repeated a large number of times each year. The underlying ecological change model is also shown in the *bottom panel*

value to natural resource management than sequential evaluation with surveillance monitoring because management actions are linked directly to resource condition and to information from monitoring. As a result, adaptive management is better at informing management than the alternative. However, surveillance monitoring is commonly applied and there are instances when management actions cannot be specified. Moreover, surveillance monitoring may be appropriate when management actions are unlikely to affect the resource state under investigation without a substantial time lag (e.g., climate change). Thus, it is our contention that the best strategy is to understand when and how to monitor to support the different management types.

Our review leads to the following conclusions:

- The unifying concepts of ecological change, utility value, and decision threshold are common to the different monitoring types. Ecological change should be monitoring at least conceptually if improving decision making is of primary importance. Ecological thresholds, abrupt change across small changes in resource condition, are a special case of ecological change, although change could be found empirically to be gradual or linear. Utility value is associated with resource condition to reflect value of desired condition and relative costs of management. Decision thresholds are present in both management types, but are derived differently.

- The critical difference between the two monitoring types is whether management actions have been specified a priori and models are available to predict the consequences of those actions on resource condition. Specific management actions and predictive models are integrated into adaptive management but are not for surveillance monitoring.
- The objectives of supporting and improving management decisions are held in common. However, adaptive management is integrated into a structured decision process, and thus is well positioned to meet those objectives. Although surveillance monitoring can be designed to provide information for management decisions, there is no a priori mechanism to translate that information into better decision because the future consequences of those decisions are not predicted or compared.
- Both types of management and monitoring are confronted with the same sources of uncertainty. Adaptive management incorporates uncertainty into all components of a structured decision process, especially in the predictive step. Surveillance monitoring incorporates uncertainty mainly through setting the decision threshold based on the type I error rate. Using power analysis to design aspects of surveillance monitoring is a special case of maximizing utility value.

References

Andersen, T., J. Cartensen, E. Hernandez-Garcia, and C. M. Duarte. 2009. Ecological thresholds and regime shifts: Approaches to identification. *Trends in Ecology and Evolution* 24:49–57.

Bain, M. B. 1987. Structured decision making in fisheries management: trout fishing regulations on the Au Sable River, Michigan. *North American Journal of Fisheries Management* 7:475–481.

Bennetts, R. E., J. E. Gross, K.Cahill, C. McIntyre, B. B. Bingham, A. Hubbard, L. Cameron, and S. L. Carter. 2007. Linking monitoring to management and planning: Assessment points as a generalized approach. *The George Wright Forum* 24:59–77.

Bytnerowicz, A., M. Tausz, et al. 2002. Summer-time distribution of air pollutants in Sequoia National Park, California. *Environmental Pollution* 118(2):187–203.

Dorazio, R., and F. Johnson. 2003. Bayesian inference and decision theory—A framework for decision making in natural resource management. *Ecological Applications* 13(2):556–563.

EPA (U.S. Environmental Protection Agency). 2006. Guidance on systematic planning using the data quality objectives process. EPA QA/G4. http://www.epa.gov/quality/qs-docs/g4-final.pdf.

Fancy, S. G., J. E. Gross, and S. L. Carter. 2009. Monitoring the condition of natural resources in US national parks. *Environmental Monitoring and Assessment* 151:161–174.

Field, S. A., A. J. Tyre, N. Jonzen, J. R. Rhodes, and H. P. Possingham. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters* 7:669–675.

Gerber, L. R., M. Geger, M. A. McCarthy, and H. P. Possingham. 2005. A theory for optimal monitoring of marine reserves. *Ecology Letters* 8:829–837.

Gould, W., and J. Nichols. 1998. Estimation of temporal variability of survival in animal populations. *Ecology* 79(7):2531–2538.

Jackson, J. K., and V. H. Resh. 1988. Sequential decision plans in monitoring benthic macroinvertebrates: Cost savings, classification accuracy, and development of plans. *Canadian Journal of Fisheries and Aquatic Sciences* 45:280–286.

Lindenmayer, D. B., and G. Likens. 2009. Adaptive monitoring: A new paradigm in long-term studies. *Trends in Ecology and Evolution* 24:482–486.

Lyons, J. E., M. C. Runge, H. P. Laskowski, and W. L. Kendall. 2008. Monitoring in the context of structured decision-making and adaptive management. *Journal of Wildlife Management* 72:1683–1692.

Keeney R. L. Value-Focused Thinking. 1992. Harvard University Press, Cambridge, Massachusetts.

Martin, J., M. C. Runge, et al. 2009. Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications* 19:1079–1090.

Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. *Trends in Ecology and Evolution* 21:668–673.

Overton, W. S., and S. V. Stehman. 1996. Desirable design characteristics for long-term monitoring of ecological variables. *Environmental and Ecological Statistics* 3(4):349–361.

Pollock, K., J. Nichols, et al. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* 13(2):105–119.

Regan, H. M., M. Colyvan, et al. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12(2):618–628.

Starfield, A. M. 1997. A pragmatic approach to modeling for wildlife management. *Journal of Wildlife Management* 61:261–270.

Stevens Jr., D., and A. Olsen. 1999. Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics* 4(4):415–428.

Stevens Jr., D., and A. Olsen. 2004. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* 99(465):262–279.

Swift, T. L., and S. J. Hannon. 2010. Critical thresholds associated with habitat loss: A review of the concepts, evidence, and applications. *Biological Reviews* 85:35–53.

Taylor, B., and T. Gerrodette. 1993. The uses of statistical power in conservation biology: The vaquita and northern spotted owl. *Conservation Biology* 7(3):489–500.

Vos, P., E. Meelis, et al. 2000. A framework for the design of ecological monitoring programs as a tool for environmental and nature management. *Environmental Monitoring and Assessment* 61(3):317–344.

Williams, B. K., R. C. Szaro, and C. D. Shapiro. 2007. *Adaptive management: The U.S. Department of the Interior technical guide*. Washington, D.C.: U.S. Department of the Interior.

Yoccoz, N. G., J. D. Nichols, et al. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16:446–453.

Chapter 7

Applying Threshold Concepts to Conservation Management of Dryland Ecosystems: Case Studies on the Colorado Plateau

Matthew A. Bowker, Mark E. Miller, Steven L. Garman and Travis Belote

Abstract Ecosystems may occupy functionally distinct alternative states, some of which are more or less desirable from a management standpoint. Transitions from state to state are usually associated with a particular trigger or sequence of triggers, such as the addition or subtraction of a disturbance. Transitions are often not linear, rather it is common to see an abrupt transition come about even though the trigger increases only incrementally; these are examples of threshold behaviors. An ideal monitoring program, such as the National Park Service's Inventory and Monitoring Program, would quantify triggers, and be able to inform managers when measurements of a trigger are approaching a threshold so that management action can avoid an unwanted state transition. Unfortunately, both triggers and the threshold points at which state transitions occur are generally only partially known. Using case studies, we advance a general procedure to help identify triggers and estimate where threshold dynamics may occur. Our procedure is as follows: (1) Operationally define the ecosystem type being considered; we suggest that the ecological site concept of the Natural Resource Conservation Service is a useful system, (2) Using all available a priori knowledge to develop a state-and-transition model (STM), which defines possible ecosystem states, plausible transitions among them and likely triggers, (3) Validate the STM by verifying the existence of its states to the greatest degree possible, (4) Use the STM model to identify transitions and triggers likely to be detectable by a monitoring program, and estimate to the greatest degree possible the value of a measurable indicator of a trigger at the point that a state transition is imminent

---

M. A. Bowker (✉)
School of Forestry, Northern Arizona University,
200 East Pine Knoll Drive, Flagstaff, AZ, USA
e-mail: matthew.bowker@nau.edu

M. E. Miller
National Park Service, Southeast Utah Group, 2282S. West Resource Blvd., Moab, UT, USA

S. L. Garman
U.S. Geological Survey, GeoSciences and Environmental Change Science Center,
Box 25046 Denver Federal Center, Denver, CO, USA

T. Belote
The Wilderness Society, Northern Rockies Regional Office,
503W. Mendenhall, Bozeman, MT, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_7,

(tipping point), and values that may indicate when management intervention should be considered (assessment points). We illustrate two different methods for attaining these goals using a data-rich case study in Canyonlands National Park, and a data-poor case study in Wupatki National Monument. In the data-rich case, STMs are validated and revised, and tipping and assessment points are estimated using statistical analysis of data. In the data-poor case, we develop an iterative expert opinion survey approach to validate the degree of confidence in an STM, revise the model, identify lack of confidence in specific model components, and create reasonable first approximations of tipping and assessment points, which can later be refined when more data are available. Our goal should be to develop the best set of models possible given the level of information available to support decisions, which is often not much. The approach presented here offers a flexible means of achieving this goal, and determining specific research areas in need of study.

Keywords Monitoring · State and transition model · Tipping point · Expert opinion · Alternative stable state · Dryland · Ecosystems · Assessment points · Delphi method

Introduction

Threshold concepts are used in research and management of ecological systems to describe and interpret abrupt and persistent reorganization of ecosystem properties (Walker and Meyers 2004; Groffman et al. 2006). Abrupt change and the progression of reorganization can be triggered by one or more interactive disturbances such as land-use activities and climatic events (Paine et al. 1998). Thresholds occur when feedback mechanisms that typically absorb forces of change are replaced with those that promote development of alternative equilibria or states (Suding et al. 2004; Walker and Meyers 2004; Briske et al. 2008). The alternative states that arise have reduced ecological integrity and value in terms of management goals relative to the original or reference system. Alternative stable states with some limited residual properties of the original system may develop along the progression after passing a threshold; an eventual outcome may be the complete loss of prethreshold properties of the original ecosystem. Reverting to the more desirable reference system becomes increasingly difficult and expensive along the progression gradient and may eventually become impossible. Ecological-threshold concepts have been applied as a heuristic framework and to aid in the management of rangelands (Bestelmeyer 2006; Briske et al. 2006, 2008), aquatic (Scheffer et al. 1993; Rapport and Whitford 1999), riparian (Stringham et al. 2001; Scott et al. 2005), and forested ecosystems (Allen et al. 2002; Digiovinazzo et al. 2010). They have been applied in contexts varying from ecological restoration (Hobbs and Norton 1996; Whisenant 1999; Suding et al. 2004; King and Hobbs 2006) to ecosystem sustainability (Herrick 2000; Chapin et al. 1996; Davenport et al. 1998) to assessment of natural resource impairment (USDI-NPS 2003).

Achieving conservation management goals requires the protection of resources within the range of desired conditions (Cook et al. 2010; Symstad and Jonas (Chap. 8).

The goal of conservation management for natural resources in the US National Park System is to maintain native species and habitat unimpaired for the enjoyment by future generations. Achieving this goal requires, in part, early detection of system change and timely implementation of remediation. The recent National Park Service Inventory and Monitoring program (NPS I&M) was established to provide early warning of declining ecosystem conditions relative to a desired native or reference system (Fancy et al. 2009). To be an effective tool for resource protection, monitoring must be designed to alert managers of impending thresholds so that preventive actions can be taken. This requires an understanding of the ecosystem attributes and processes associated with threshold-type behavior, how these attributes and processes become degraded, and how risks of degradation vary among ecosystems and in relation to environmental factors such as soil properties, climatic conditions, and exposure to stressors. In general, the utility of the threshold concept for long-term monitoring depends on scientists' and managers' ability to detect, predict, and prevent the occurrence of threshold crossings associated with persistent, undesirable shifts among ecosystem states (Briske et al. 2006). Because of the scientific challenges associated with understanding these factors, the application of threshold concepts to monitoring designs has been very limited to date (Groffman et al. 2006). As a case in point, the monitoring efforts across the 32 NPS I&M networks were largely designed with the knowledge that they would not be utilized to their full potential until the development of a systematic method for understanding threshold dynamics and methods for estimating key attributes of state changes.

This chapter describes a generalized approach we implemented to formalize understanding and estimating of threshold dynamics for terrestrial dryland ecosystems in National Parks of the Colorado Plateau. We provide a structured approach to identify and describe degradation processes associated with threshold behavior, and to estimate indicator levels that characterize the point at which a threshold crossing has occurred or is imminent (tipping points), and points where investigative or preventive management action should be triggered (assessment points). We illustrate this method for two case studies in National Parks included in the Northern and Southern Colorado Plateau I&M Networks, where historic livestock grazing, climatic change, and invasive species are key agents of change. The approaches developed in these case studies are intended to enhance the design, effectiveness, and management relevance of monitoring efforts in support of conservation management in dryland systems. They specifically enhance NPS capacity for protecting park resources on the Colorado Plateau, but have applicability to monitoring and conservation management of dryland ecosystems worldwide.

Background: Threshold and State-and-Transition Concepts

Salient features among frameworks of ecological thresholds include concepts of reference conditions, feedback dynamics, threshold triggers, properties of the progression after a threshold crossing, and changes in restoration potential along this progression. Native or reference conditions, typically, are the desired state for conservation management, and consist of community phases and transitions among phases due to natural disturbances and climate variability. Negative feedback of the reference system confer system resilience and maintain the community phases within a characteristic range of variability. For instance, a negative feedback that inhibits shrub dominance in some grasslands is the interaction between amount of grass cover and fire return interval. Given sufficient grass cover, wildfire events are frequent and large enough to maintain grassland structure due to the selective elimination of fire-intolerant woody plants. Phases comprising the natural range of reference conditions differ in their vulnerability to crossing a threshold. Phases with degraded resilience are more vulnerable and may be described as "at risk" of a persistent transition to an alternative state (Briske et al. 2008). Identifying the patterns that increase vulnerability to change and reasons for these patterns can define preventive-management goals (Bestelmeyer 2006).

Both biotic and abiotic mechanisms may trigger state changes (Beisner et al. 2003; Briske et al. 2006). Biotic mechanisms include altered biotic structure and interactions, such as plant–herbivore interactions. Abiotic mechanisms (e.g., extreme soil erosion) can result in threshold behavior through the modification of inherent site characteristics. A single trigger may initiate a state change, or the temporal order or spatial convergence of multiple triggers may be critical. For example, drought or intensive livestock grazing alone may not trigger a state change, but the two factors in combination or in sequence may trigger such a change through adverse effects of one stressor on ecosystem resilience to the other stressor (Scheffer et al. 2001). Triggers result in conditions that exceed the resilience of the reference system, and lead to an increasing dominance of positive, destabilizing feedback. Triggers often initiate changes in the pattern or spatial structure of an ecosystem (e.g., decreased vegetation cover or increased patchiness) with subsequent and often nonlinear changes in processes (e.g., soil erosion, nutrient cycling; Peters et al. 2007).

The progression resulting from a state change is characterized by increasing dominance of positive feedback, and changes in pattern and processes (Briske et al. 2008). Along this progression is the continual loss of properties of the reference condition. Multiple alternative states, each with their own set of varying community phases, can occur along this threshold gradient with some becoming stable as negative feedback of the alternative state confers resilience. Progression can lead to a degraded state where features of the reference condition are effectively no longer present. Degraded states may no longer afford provision of services such as water, livestock forage production, or desirable recreational opportunities, and may no longer support the biodiversity of native systems.

The potential for restoration to prethreshold conditions is determined by the amount of residual properties of the reference condition and the resilience of the new, alternative state (Suding and Hobbs 2009). Where extensive site preparation and reintroduction of native species are required for conversion to prethreshold conditions, the costs may effectively prohibit restoration. In some cases, complete restoration to native conditions may never be possible due to the extinction of native biota (i.e., species, genomes), or the loss of inherent properties (e.g., soil fertility) necessary to support reference conditions.

Focused study and interpretation of threshold processes and consequences benefit from using conceptual models of ecosystem dynamics. State-and-transition models

(STMs) are a type of conceptual model that have become prominent in rangeland management, and are used to illustrate reference conditions of an ecosystem, ecosystem responses to natural and anthropogenic drivers, and the mechanisms of transition among distinctive assemblages or states of an ecosystem (Bestelmeyer et al. 2003, 2009). These models also provide a basis for discerning levels of system properties indicative of both the risk and occurrence of transition among states (Briske et al. 2008).

Identifying indicator levels indicative of an impending state change is a critical component for the design of effective monitoring. Monitoring efforts should result in alerting land managers of indicator levels in advance of a state change to account for lag-time in decision making and uncertainty in the effectiveness of remediation actions. From a statistical perspective, the number and frequency of monitoring observations to provide an early warning is dependent on the difference between the current status of the indicator, the early-warning status level, and the inherent spatial and temporal variability of the indicator. Realistically, given uncertainty in early-warning levels and inherent variability of indicators, monitoring resources are likely insufficient to statistically detect a declining trend within a time period sufficient for decision making (Field et al. 2004). Bennetts et al. (2007) have proposed the use of management-assessment points along a continuum of indicator values to safeguard against uncertainties in estimates of thresholds, in indicator variability, and in the efficacy of a monitoring or sampling design. Ecosystem progression, where monitored attributes reach an assessment point does not necessarily warrant immediate remediation action, but instead motivates close scrutiny. Assessment points ideally are based on management goals and concerns, including understanding risks (Nichols et al. Chap. 2). However, a fundamental component for establishing assessment points is a credible estimate of resource and environmental conditions indicative of impending state changes.

A General Approach to Applying Threshold Theory to Management

We developed a general approach for identifying properties of thresholds to inform estimates of management-assessment points in a long-term monitoring context. Our approach relies on using conceptual models of threshold dynamics, and various sources of information to verify the conceptual model, and to make informed estimates of state changes and associated indicator values:

1. *Identification of target ecosystems:* We adopted the US Department of Agriculture Natural Resource Conservation Service's (USDA-NRCS) ecological site concept as a spatial framework for ecosystem classification and model development. Ecological sites are land units differentiated by (a) physical attributes including inherent soil properties (texture, depth, and horizonation), geomorphic setting, and climate; (b) the potential (rather than current) vegetation associated

with these physical attributes within a specific ecoregion, and (c) characteristic dynamics in response to climate, management, and other driving factors (Herrick et al. 2005; Bestelmeyer et al. 2009).

2. *Conceptual models of system dynamics:* We developed STMs to organize current knowledge or hypotheses about the dynamics and community phases of specific ecological sites, the key alternative states representative of degradation pathways, and the transitions that are possible among these states. Possible triggers of transitions among alternative states, and pattern and process indicators of specific degradation pathways are identified or hypothesized based on published literature, unpublished expert knowledge of an ecological site, or general ecological principles. Identifying triggers is most useful since observations of their occurrence could initiate preventative management actions. This process- and theory-based focus in the construction of the STM, contrasts with pattern-based efforts, which seek to define states based upon classification of multivariate community structure data (e.g., Allen-Diaz and Bartolome 1998). These data-driven approaches offer the credibility of being based upon real data, but assume that a dataset is likely to capture all of the important states that are possible within a given ecological site, and that the identified states are fundamentally and functionally distinct (Bestelmeyer et al. 2003). Rather, we advocate using available data to test specific elements of process-based conceptual STMs, as a means of calibrating and validating the model.
3. *Model calibration:* Model building is an iterative process, and it is important to include a calibration step. Calibration includes testing the concepts presented in the model using available datasets, or subjecting them to the scrutiny of an expert panel. This enables an opportunity to revise the model, identify new transitions and associated triggers, processes and indicators, and allows an estimation of our confidence that the revised model is reasonable.
4. *Identification of key transitions and estimation of tipping points:* The calibrated model is used to identify the most likely transitions that might be detected by a monitoring program, emphasizing those known to be of concern to management, such as the persistent conversion of perennial grasslands to ecosystems dominated by invasive annuals or woody plants. The values of key indicators at the point of a state change—when one state abruptly transitions to another—are estimated. We refer to these as tipping points; they are roughly equivalent to restoration thresholds (*sensu* Bestelmeyer 2006). Because abrupt transitions in progress are seldom observed, statistical methods are used to model the tipping points in indicator values using sample representative of discrete states. In data-sparse situations, these estimates are derived from expert knowledge rather than statistical modeling. The assessment points are another set of indicator values which trigger management action prior to observing a tipping point, so that the undesired transition can be avoided. These values occur chronologically before tipping points and allow managers sufficient response time. They are based upon the range of natural variability in the reference or less-degraded state when data are available, or upon opinions from an expert panel when data are lacking.

Case Studies

We now present two case studies that illustrate different methods for identifying assessment points based on contrasting scenarios of data availability. The case studies specify two ecological sites that occur in NPS units on the Colorado Plateau, where the general monitoring goal is to provide early warning of system decline in sufficient time for management actions to avert impending undesirable changes.

A. Data-Rich Case Study: Semidesert Sandy Loam Ecological Site, Canyonlands National Park

Ecological Site Characteristics The semidesert sandy loam (SDSL) ecological site is widely distributed throughout the Colorado Plateau region of North America and is significant for its past and current use for livestock grazing (USDA-NRCS major land resource area 35, ecological site 035XY215UT). This ecological site occurs on flat to gently sloping landforms at 1,310–2,010 m elevation and receives 20–30 cm mean annual precipitation. Soils are formed in moderately deep to very deep (from 50 to greater than 150 cm) aeolian and alluvial deposits from sandstone and are moderately alkaline with sandy loam or loamy sand texture. In relatively undisturbed settings, the vascular plant community typically has a grassland aspect and is characterized by a mixture of perennial C_3 (*Hesperostipa comata* and *Achnatherum hymenoides*) and C_4 (*Sporobolus* spp.) bunchgrasses, C_4 rhizomatous grasses (*Pleuraphis jamesii* and *Bouteloua gracilis*), shrubs, and annual herbaceous species. In contrast with many dryland ecosystems, most common shrubs (e.g., *Krascheninnikovia lanata* and *Atriplex canescens*) are palatable to livestock and shrub-dominated communities can occur with long-term absence of livestock grazing. Plant nomenclature here and throughout follows USDA-NRCS 2010. Biological soil crust (Belnap 2003) is another biotic functional type that is a characteristic component of relatively undisturbed SDSL sites (Kleiner and Harper 1972; Bowker and Belnap 2008). Biological soil crusts have yet to be widely incorporated in conceptualizations of dryland ecosystem dynamics despite evidence of their functional significance for soil stabilization (Belnap 1995; Warren 2003), nutrient cycling (Evans and Lange 2003), hydrologic processes (Warren 2003), and mediation of plant establishment (Belnap et al. 2003; Escudero et al. 2007). Biological soil crusts are also notable for their lack of resistance to surface disturbances which can result in long-term reductions in spatial continuity, biological diversity, physical structure, and functionality (Belnap and Eldridge 2003; Miller 2008).

Management Goals and Land-Use History General NPS goals for management of natural resources are (1) to preserve and restore the natural abundance, diversity, and dynamics of native plant and animal populations and the communities and ecosystems in which they occur, and (2) to minimize human impacts on native plant and animal populations, communities, ecosystems, and the processes that sustain them (USDI-NPS 2006). Canyonlands National Park preserves regionally significant examples of SDSL ecosystems that remain relatively undisturbed by human activities exclusive of anthropogenic atmospheric changes. Within Canyonlands National Park, however, there also are extensive examples of SDSL ecosystems with persistently degraded composition, structure, and function attributable to impacts of past livestock grazing (e.g., Neff et al. 2005; Belnap et al. 2009). Domestic livestock were introduced to this area in the late 1880s and portions of Canyonlands were grazed by livestock until 1974. Livestock grazing remains an important economic activity on adjacent lands outside Canyonlands. Unlike many semiarid grasslands, neither fire nor frequent grazing by herds of large mammals are characteristic natural disturbances associated with the SDSL site. Thus, grazing and associated surface disturbances by livestock represent novel disturbances in this system.

Data Availability Three general types of data characterize structural and functional attributes of the SDSL ecological site for Canyonlands National Park and surrounding areas: (1) poorly replicated in space and time (Kleiner and Harper 1972; Neff et al. 2005), (2) well replicated in time, poorly replicated in space (Belnap et al. 2009; S.M. Munson unpublished data), and (3) well replicated in space, poorly replicated in time (Miller et al. 2011). Of these options, the first two provide many insights into ecosystem dynamics but only the third type provides the necessary replication for the statistical estimation of tipping or assessment points, or are broad enough to characterize the variability within states. The third type of data is derived from a broad-scale ecosystem inventory project purposefully designed to characterize ranges of variability in key compositional and structural attributes of dryland ecosystems in Canyonlands National Park and on adjacent lands currently used for livestock grazing (Miller et al. 2011). These inventory data were collected over a 3-year time period and thus do not quantify temporal transitions among states. However, through a combination of targeted sampling and extensive spatial replication (substituting space for time) with random sampling, this data set documents current ranges of variability for the SDSL and provides a relatively rich basis for estimating tipping points and associated assessment points. The data set quantified variability among 72 SDSL plots on a single soil type (Begay series) on the basis of live cover of biological crusts and vascular plants, ground cover, and indicators of erosion resistance including soil aggregate stability, spacing between perennial plant canopies, and spacing between perennial-plant bases (Miller et al. 2011; sampling methods followed Herrick et al. 2005). Sampling was conducted both within and outside Canyonlands National Park to ensure that the data set spanned a wide range of ecosystem conditions.

Methods: Building a State-and-Transition Model and Estimating Tipping Points with Rich Inventory Data Field observations, published literature (Kleiner and Harper 1972; Neff et al. 2005; and Belnap et al. 2009) and an existing USDA-NRCS ecological site description (USDA-NRCS ecological site 035XY215UT) provided the basis for developing an STM articulating hypotheses about system dynamics, degradation pathways among alternative states, and associated ecosystem patterns, processes, and feedback.

The conceptual model identifies four ecosystem states based on persistent differences in the relative abundance of biotic functional types (Fig. 7.1; Tables 7.1 and 7.2). Two states are dominated by biological crusts and are distinguished from one another by the absence (S1) or presence (S2) of functionally significant invasive exotic annuals (e.g., *Bromus tectorum* or *Salsola* sp.). An invaded state (S3) is characterized by the replacement of biological crust by bare ground and a vascular plant community dominated by perennial grasses (S3P1) or palatable shrubs (S3P2) with significant levels of invasive annuals. The fourth state (S4) is characterized by persistent dominance by invasive annual grasses or forbs. The first state represents the desired condition relative to NPS management goals, whereas states two through four represent increasing degrees of degradation to be avoided or mitigated.

We used a logical quantitative process to analyze the inventory data set to examine evidence for our STM. It consists of construct validation of the STM, and determination of quantitative classification rules of state membership. To validate the existence of the states proposed in our a priori STM, fuzzy cluster analysis (Equihua 1990) was applied to four state properties including biological crust cover, bare ground cover, combined cover of perennial grasses and palatable shrubs, and relative cover of invasive exotic annuals based upon a Bray–Curtis distance matrix. Fuzzy clustering methods offer more flexibility than hierarchical clustering when attempting to group elements which may overlap or have vague boundaries, such as states. Following cluster identification, classification tree modeling (De'ath and Fabricius 2000) was used to derive quantitative decision rules for differentiating clusters (Fig. 7.3). While these methods may or may not arrive at the same clustering of data, their utility is somewhat different. Starting with the root node (composed of the entire dataset), classification trees iteratively and dichotomously partition the data set into increasingly homogenous groups, producing a dendritic pattern of terminal nodes. Each partition is based upon the values of a single predictor variable. This property of classification trees makes them useful for isolating the single variable(s) most informative in determining node/cluster membership, and provide a decision rule based on that predictor (e.g., $\geq 28\,\%$ relative exotic cover = State 3). These values represent classification thresholds (*sensu* Bestelmeyer 2006) for clusters or nodes rather than actual functional or degradation thresholds for the SDSL ecological site; however, they provide a reasonable first approximation of tipping points in empirical measurements of key functional indicators.

The cluster analysis distinguished three clusters analogous to states S2–S4 in the conceptual model (Miller et al. 2011), and provided no evidence for states not included in the model. The classification analysis splits cluster S4 from clusters S2 and S3 at 28.3 % relative cover of invasive exotic annuals (Fig. 7.2a). Clusters S2 and S3 are split from one another at 30.3 % bare ground (Fig. 7.2a).

Implications for Monitoring This case study applies a conceptual model of ecosystem dynamics, a relatively rich set of inventory data, and multivariate data analysis techniques to derive monitoring-assessment points. Despite the fact that some pristine sites were included in the dataset, examples of S1 (partially defined by a lack of exotics) were not located. Thus, the management goals ought to detect and avoid the

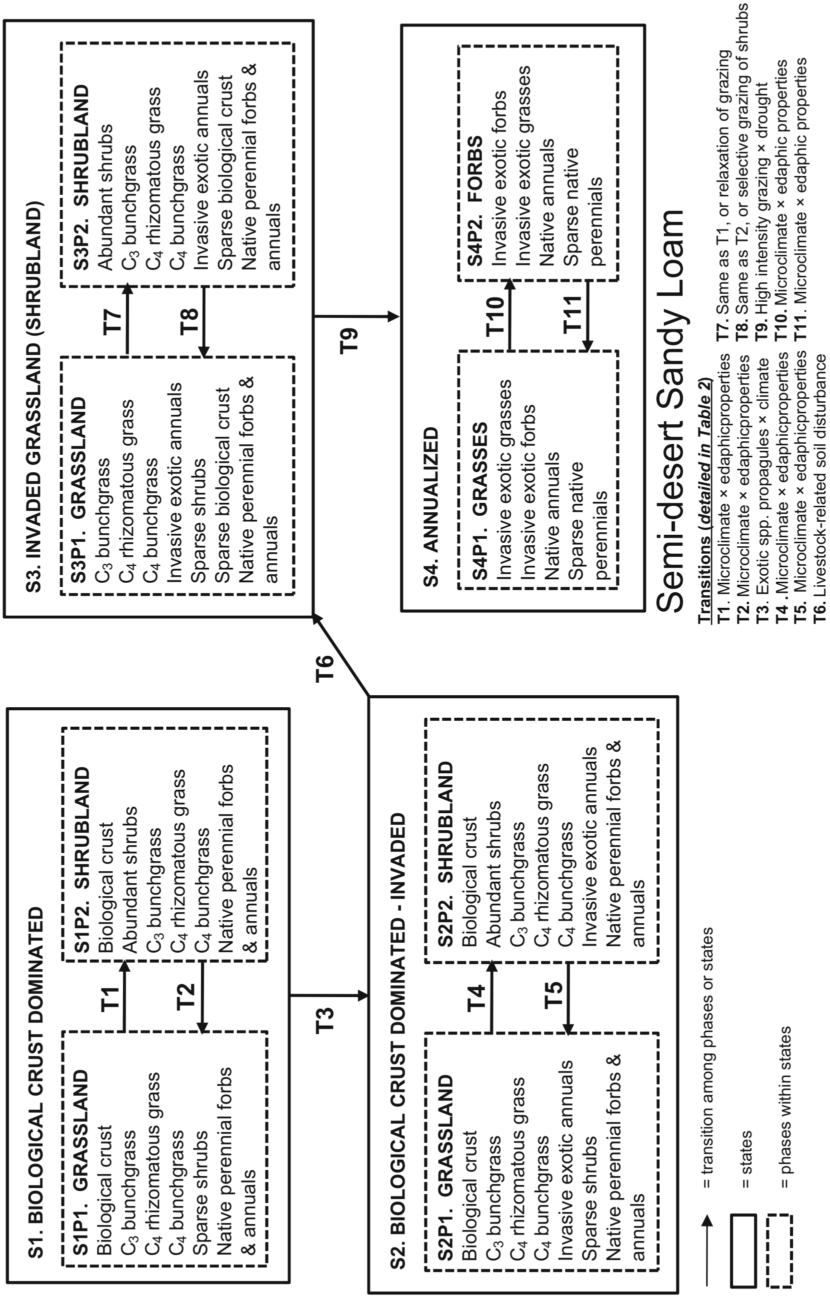


Fig. 7.1 State and transition diagram for semidesert sandy loam

Table 7.1 Catalog of states and phases in semidesert sandy loam

State	Phase	Structural properties	Functional properties	Feedback
S1. Biological crust dominated	P1. Grassland	Biological crust dominant or codominant relative to vascular plants; perennial grasses abundant relative to shrubs, with variable grass composition due to climate fluctuations, soil variability, and site history. High degree of soil-surface roughness	High biological crust cover maintains high capacity for resource capture and retention (including nutrients, water, litter, and seeds) even with fluctuations in plant cover[a,b]	High resource retention promotes plant community resilience to climatic fluctuations and natural disturbance
	P2. Shrubland	Similar to S1P1 but with palatable shrubs abundant relative to perennial grasses	Same as S1P1	Same as S1P1
S2. Biological crust dominated—invaded	P1. Grassland	Similar to S1P1 but invasive exotic annuals present. Cover of invasive annuals fluctuates with climate	Similar to S1P1, but presence of invasive annuals can cause greater climate-driven fluctuations in cover and production relative to S1P1	Same as S1P1
	P2. Shrubland	Similar to S1P2 but invasive exotic annuals present. Cover of invasive annuals fluctuates with climate	Similar to S1P2, but presence of invasive annuals can cause greater climate-driven fluctuations in cover and production relative to S1P2	Same as S1P2
S3. Invaded grassland (or shrubland)	P1. Grassland	Biological crust replaced by bare ground; otherwise similar to S2P1. Major decline in soil-surface roughness relative to S1P1 and S2P1	Loss of stability and roughness associated with biological crust result in major decline in site capacity for resource capture and retention; accelerated losses of soil, nutrients, water, litter, and seeds occur[a,b]	Accelerated losses of soil resources and seeds contribute to declines in plant community resilience to climatic fluctuations and to declines in vegetative cover and production, which result in further declines in site resistance to erosion and resource loss

Table 7.1 (continued)

State	Phase	Structural properties	Functional properties	Feedback
	P2. Shrubland	Biological crust replaced by bare ground; otherwise similar to S2P2. Major decline in soil-surface roughness relative to S1P2 and S2P2	Similar to S3P1, although rates of resource loss may be greater in shrub-land due to relative lack of perennial grass cover	Same as S3P1
S4. Annualized	P1. Grasses	Dominated by invasive exotic annual grasses (e.g., *Bromus*). Native annuals may be present, but perennials sparse	Dominance by annuals results in high fluctuations in cover due to climate, with corresponding high (and potentially extreme) fluctuations in resource loss/erosion[b]	Same as S3P1, but greater. Potential spiraling declines in resource availability and site productivity[a]
	P2. Forbs	Dominated by invasive exotic annual forbs (e.g., *Salsola*). Native annuals may be present, but perennials sparse	Same as S4P1	Same as S4P1

[a] Neff et al. 2005

[b] Belnap et al. 2009

Table 7.2 Key to transitions in Fig. 7.1 (semidesert sandy loam)

Transition	Trigger(s)	Associated process(es)	Relevant indicator(s)
T1	Climate variability, perhaps interacting with landscape position and inherent soil properties (shrubland phase may be more common on older geomorphic surfaces with greater inputs of late-Pleistocene loess, greater silt content, and greater pro- file development relative to grassland phase.	Plant population processes (reproduction, recruitment, mortality)	Absolute and relative cover of perennial grasses and palatable shrubs (or shrub:grass ratio)
T2	Similar to T1, but favoring opposite relative dominance of plant functional types	Same as T1	Same as T1
T3	Establishment of invasive exotic annuals, facilitated by favorable climatic conditions	Seed dispersal and plant population processes	Density, frequency, and/or cover of invasive exotic annuals
T4	Same as T1	Same as T1	Same as T1
T5	Same as T2	Same as T1	Same as T1
T6	Repeated soil disturbance (trampling), typically associated with livestock grazing	Destruction of biological crusts due to trampling; increased connectivity of bare-ground patches; decreased soil-surface roughness and capacity for capturing/retaining litter, seeds, aeolian dust inputs, and runoff; accelerated erosion	Absolute cover of biological crust; cover of biological crust relative to bare ground and vascular plants; soil-surface roughness; percent bare ground; size and connectivity of bare ground patches; soil aggregate stability
T7	Similar to T1, but also may be facilitated by a sustained reduction in grazing pressure on palatable shrubs where previous herbivory by livestock has suppressed shrubs relative to perennial grasses.	Plant population processes; shrub regrowth following reduction in grazing pressure	Same as T1
T8	Similar to T2, but facilitated by heavy grazing pressure and selective herbivory on palatable shrubs	Selective herbivory and competitive suppression of palatable shrubs relative to perennial grasses; plant population processes	Same as T1

Table 7.2 (continued)

Transition	Trigger(s)	Associated process(es)	Relevant indicator(s)
T9	Sustained high-intensity grazing and associated soil-surface disturbance (trampling), perhaps in combination with drought	Selective herbivory and reduction of perennial grasses and palatable shrubs through effects on physiological vigor, resistance/resilience to drought, competitive relations, seed production, and replenishment of the soil seed bank; facilitation of invasive exotic plants through soil-surface disturbance and reduced competitive vigor of grazed perennials	Absolute cover of perennial grasses and palatable shrubs; absolute cover of invasive exotic annuals; relative cover of perennials and invasive annuals; soil aggregate stability
T10	Climate variability that favors exotic annual forbs relative to exotic annual grasses; relative dominance of exotic annual forbs and exotic annual grasses also may vary along elevation and/or topo-edaphic gradients through effects on soil moisture	Plant population processes (reproduction, recruitment, mortality)	Absolute and relative cover of exotic annual grasses and forbs
T11	Similar to T10, but favoring opposite relative dominance of plant functional types	Same as T10	Same as T10

initiation of transitions from S2 to S3. Likewise, for sites already in S3, management should strive to detect and prevent transition to S4. Current monitoring conducted by NPS is well designed to detect changes in key indicators of these transition sequences for the SDSL ecological site, including the relative cover of invasive exotic plants and percent bare ground. Because we are able to provide rough estimates of tipping points based on these data, the necessary prerequisites for establishment of assessment points are established. We reason that an assessment point for a given transition must lie between the estimated tipping point and the mean value of the relevant indicator in the state at risk of transition. Its actual position is determined subjectively based upon management goals and adaptively refined based upon success as a decision support tool. Some reasonable management-assessment points, ordered from most conservative to most liberal, include: the at-risk state node mean ± SE, the upper or lower bound of 95 % confidence interval of the at-risk node mean, the midpoint between the at-risk mean and the tipping point, the upper or lower bound of 95 % confidence interval of the tipping point, and the tipping point ± SE (Fig. 7.2b).

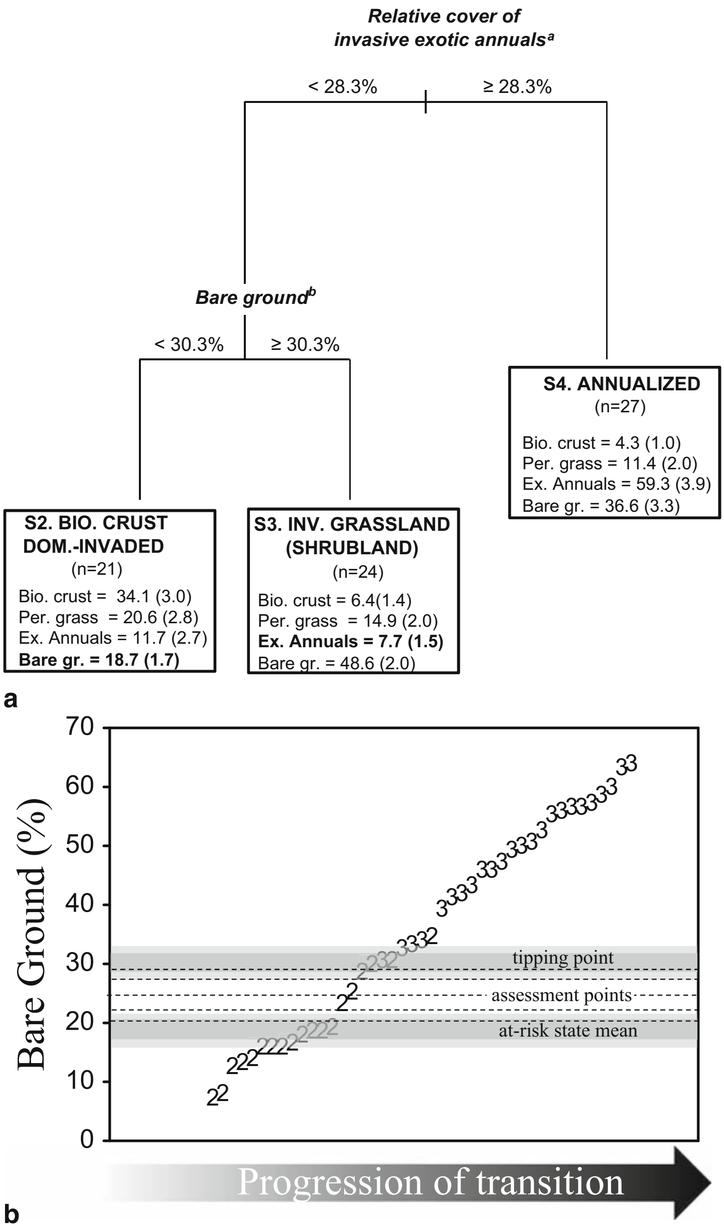


Fig. 7.2 Tipping and assessment points in the semidesert sandy loam case study: **a** Classification tree diagram depicting classification thresholds separating three states of the semidesert sandy loam ecological sites. The figure, from *top* to *bottom*, classifies samples into groups based upon values of indicators using a sequential dichotomous splitting procedure. The indicator used to make a split is in *bold italics*. Its critical values appear below it; these values are initial approximations of tipping points. End nodes are represented as *boxes* which correspond well to hypothesized states. Indicator

Current sampling is not designed to characterize or detect changes in the spatial configuration or connectivity of invasive exotic plants or bare ground. Spatial connectivity (or the length of connected pathways) in dryland ecosystems is increasingly recognized as an important structural indicator of processes such as accelerated soil erosion, overland flow, and wildfire (Okin et al. 2009). Current NPS monitoring of the SDSL ecological site on the Colorado Plateau includes measurements of gaps between perennial plant canopies and bases as indicators of resistance to erosion by wind and water (Herrick et al. 2005; Okin 2008). But no data are collected to characterize the connectivity of bare ground patches (or biological crust patches, alternatively) in the spaces between perennial plant canopies or bases. In circumstances when an assessment point is prompted by increasing levels of bare ground, measures of surface patch (intact biological crust and/or bare ground) connectivity may provide additional insights regarding degradation risks related to erosional processes.

B. Data-Sparse Case Study: Limy Uplands Ecological Site, Wupatki National Monument

Ecological Site Background Limy uplands are an ecological site represented in Wupatki National Monument and surrounding areas, situated atop fairly level basalt flows, receiving 15.2–25.4 cm of rainfall per year (USDA-SCS 1983). The soil is weathered from the underlying basalt, and from later cinder deposits due to regional volcanism. The surface is gravelly due to high-surface cinder coverage. Grassland vegetation is most common, and is dominated by C4 rhizomatous or stoloniferous grasses including *Pleuraphis jamesii* and *Bouteloua eriopoda*; C3 grasses may have been somewhat diminished due to past grazing. Savannah vegetation is less common and is characterized by an overstory of *Juniperus monosperma* of varying density and an understory of perennial grasses (Jameson 1962; Ironside 2006; DeCoster and Swan 2009).

Management Goals The primary management goals of the National Monument are to protect and preserve over 2,000 catalogued archeological sites, including structures, and agricultural fields of the ancient ancestral Hopi cultures, and to provide interpretive and educational experiences for park visitors (USDI-NPS 2002). In addition to these primary goals, NPS management goals for natural resources are the same as those summarized earlier for the semidesert sandy loam ecological site. Cattle grazing was permitted in portions of the Monument until 1989 when livestock were removed and a boundary fence was constructed (USDI-NPS 2002). The

---

means and standard errors are presented along with each node. *a* approximate tipping point corresponds to T9 in Table 7.2 and Fig. 7.1. *b* approximate tipping point corresponds to T6 in Table 7.2 and Fig. 7.1. **b** Percentage bare ground in rank order as a basis for establishment of assessment points. Point symbols represent state membership. Five alternative definitions of assessment points are derived from tipping points and at-risk state means estimated using a classification tree

Monument highlights the presence of a rare, large, ungrazed grassland as one of its significant resources, and NPS staff are concerned that increasing tree densities in Monument grasslands are attributable to a decrease in fire frequency since the late nineteenth century caused by diminished fine fuels due to grazing (Cinnamon 1988; USDI-NPS 2002; Ironside 2006). Currently, the wildfire management plan calls for suppression of fires, but retains the option of prescribed fire (USDI-NPS 2005).

Data Availability Relevant vegetation data for this ecological site either are well replicated and incomplete, or modestly replicated and reasonably complete. In aggregate, these data may not represent a sufficient range of the possible states, nor the ideal time series data capturing a transition in action to validate an STM, and may lack measurements of some potentially useful indicators. There is no single complete dataset, for validation of an STM or estimation of tipping and assessment points. Hassler (2006) likely conducted some sampling of *Juniperus* density, growth rate, and fire mortality on limy uplands. In a remote sensing-based vegetation mapping project, Hansen et al. (2004) sampled numerous accuracy assessment releves in limy uplands that qualitatively identify community type. Miller et al. (2007) developed and tested monitoring techniques at seven plots. DeCoster and Swan (2009) summarize the first years of the I&M program and contains the most purposefully collected monitoring dataset for limy uplands, but is limited to ten sites. The randomly selected study design may fortuitously capture recovery from fire gradients (1 plot in 1995 “North fire”, 3–4 plots in the 2002 “Antelope fire”; USDI-NPS 2005). The data include detailed information on vegetation structure and ground cover, including some metrics of juniper density, but lacks direct indices of connectivity of fine fuels.

Methods: Building a State-and-Transition Model and Estimating Assessment Points with Sparse Data Due to the incomplete nature of the available data, we pursued an alternative strategy for the validation of the states and dynamics delineated in STM. Our approach has much in common with the Delphi technique of engaging expert opinion panels, in that, it is a multiphase, iterative approach, employs a “straw-document” as a starting point, and engages participants individually so that outputs are not disproportionately affected by dominant personalities (Linstone and Turoff 1975; Oliver 2002). This approach has proven to be useful when “the problem does not lend itself to precise analytical techniques but can benefit from subjective judgements on a collective basis (Linstone and Turoff 1975).” We constructed email-based questionnaires in two stages: (1) model calibration, (2) estimation of tipping and assessment points in indicators which enable detection of proximity to threshold crossings. Based on literature findings and past experience, we drafted an STM including a catalog of states, phases, and transitions. We identified a list of potential expert consultants from the authors of relevant literature, and from professional interactions. We initially contacted selected experts by email to gauge interest. Of eight people contacted, five were willing to participate. The format of the model calibration survey included: (1) a paragraph-length overview of STM concepts, (2) a brief description of the target ecological site, (3) a draft STM including a diagram and verbal catalog, and (4) a questionnaire. The questionnaire consisted of four required questions and six optional ones. The required questions asked respondents to identify any states, phases, or transitions which should be removed from or added to the model. For additions, respondents were prompted to identify: structural and functional properties and stabilizing negative feedback of states and phases, and triggers (including their characteristic scale) and appropriate indicators of transitions. Our questionnaires specifically employed estimates of confidence in responses, an important measure of uncertainty. In the Phase 1 questionnaire, respondents were asked to estimate their confidence in a revised model, which took into account their proposed changes (a subjective scale taking any value from 0 to 100 %, where 0 % = "It's anyone's guess, this model is no better than any other model," 50 % = "Because this model is reasonable, I would tend to believe it until evidence to the contrary is presented," 100 % = "The model is so well supported by evidence and accumulated knowledge, that I am certain it is correct."). The same information was requested for each individual model component (states, phases, and transitions). These confidence estimates are hereafter known as "C-own." As a complementary question, respondents were also asked to estimate their confidence in the model generated by a theoretical "best qualified" person, to help gauge their confidence in a survey-based procedure for developing STM (hereafter known as "C-best"). We received four surveys with an average response time of 9 days (we had requested return within a week). We revised the model, according to all respondents' comments. We also calculated an aggregate confidence value. First, the C-best values were used to correct optimistic or pessimistic tendencies in respondents' estimation of C-own. For example, if a respondent's C-best value was 20 % less than the mean C-best value, their C-own value was adjusted up by 20 % to account for their greater than average pessimism. The adjusted C-own values were averaged across all respondents, and calculated for the entire model and for each model component.

The second phase of the survey was more focused on thresholds associated with a key transition (T6 from reference grasslands to savannized ecosystems, see later). This survey consisted of the following parts: (1) a revised STM with aggregate confidence values, (2) a background section regarding resiliency concepts, tipping and assessment points, and (3) a questionnaire. In the questionnaire portion, respondents were presented with a set of indicators and their characteristic units, and were asked to estimate tipping and assessment points for each. As in the previous survey, we required C-own and C-best values for all indicators overall, and respondents were invited to provide them for each individual indicator. We emailed the Phase 2 surveys to the four respondents who had previously returned Phase 1, in addition to one new respondent and several previous candidates who had not been able to respond. We received six of them back with an average response time of 20 days. To calculate aggregate estimates of assessment and tipping points, we adjusted each respondent's C-own values using their C-best values, using the same procedure described earlier. The adjusted C-own values were then used to compute a weight for a given estimate of a given respondent by dividing the adjusted C-own of the respondent, divided by the sum of all respondents' adjusted C-own values, yielding a proportion. This proportion was used in a weighted averaging procedure to calculate the group's estimates.

Final Resilience-Based State-and-Transition Model We acknowledge a prehistory of volcanism and occupation by agricultural societies, and subsequent depopulation (Sullivan and Downum 1991), but omit detail on these states for brevity. We emphasize states, phases and transitions within the current management sphere (Fig. 7.3; Tables 7.3 and 7.4).

The survey-based approach proved to be quite useful, but perhaps not fully satisfactory. On one hand, they proved to be an excellent tool for calibration of STMs, as new states, phases, and transitions were identified, and differing levels of confidence emerged in different portions of the model, identifying the greatest research needs (e.g., potential for transition to woodland, and recovery of grass dominance after shrub dominance; Fig. 7.3). The overall aggregate confidence in the model was quite high on a subjective scale (71 %), indicating that despite the lack of data, survey respondents tended to believe that this model was the correct model of ecosystem dynamics. We were able to provide quantitative approximations of tipping and assessment points based upon subjective rather than empirical data for only three of ten indicators (based on estimates given by a minimum of three respondents; Table 7.5). This was because respondents were reticent to offer estimates about subjects for which they did not feel knowledgeable (less than about 20 % confidence), thus for indicators related to livestock or native grazer activity and connectivity of fuels we obtained little information. However, each respondent did suggest at least one additional indicator resulting in a total of seven additional indicators that could be folded into a monitoring program. Data gaps could probably be ameliorated with a larger sample size of surveys when possible; however, our approach has the inherent limitation that there are a small pool of respondents with knowledge of the target site, and even fewer available to respond to surveys.

Implications for Monitoring Expert opinion surveys resulted in a highly useful model of ecosystem dynamics and seven suggestions of indicators which should be investigated further for their potential to indicate change, several of which could be derived from the data currently being collected. Most respondents tended to believe that the transition to savannahs is fire regulated. As a result, we were able to establish rough first approximations of tipping points in some related indicators to aid in the establishment of assessment points (Table 7.5). These estimates should be confirmed based upon data when possible, but illustrate that even when data are lacking, an operational tipping point can be established. Compared to the data-rich case, there is less available information to establish assessment points; for example we do not know the distributions of indicator values within the at-risk state. However, the weighted average of survey respondents' assessment points provides a reasonable starting point.

Survey products suggested several ways to learn about this ecosystem. For example, the two leading hypotheses regarding savannization, that the process is fire-limited, and that the process is favored by wet climate periods, could be tested using monitoring data. Currently, the NPS I&M sampling strategy within Wupatki's limy uplands is well designed for detecting changes in vegetation structure such as increasing relative abundance of woody plants. However, the design could be

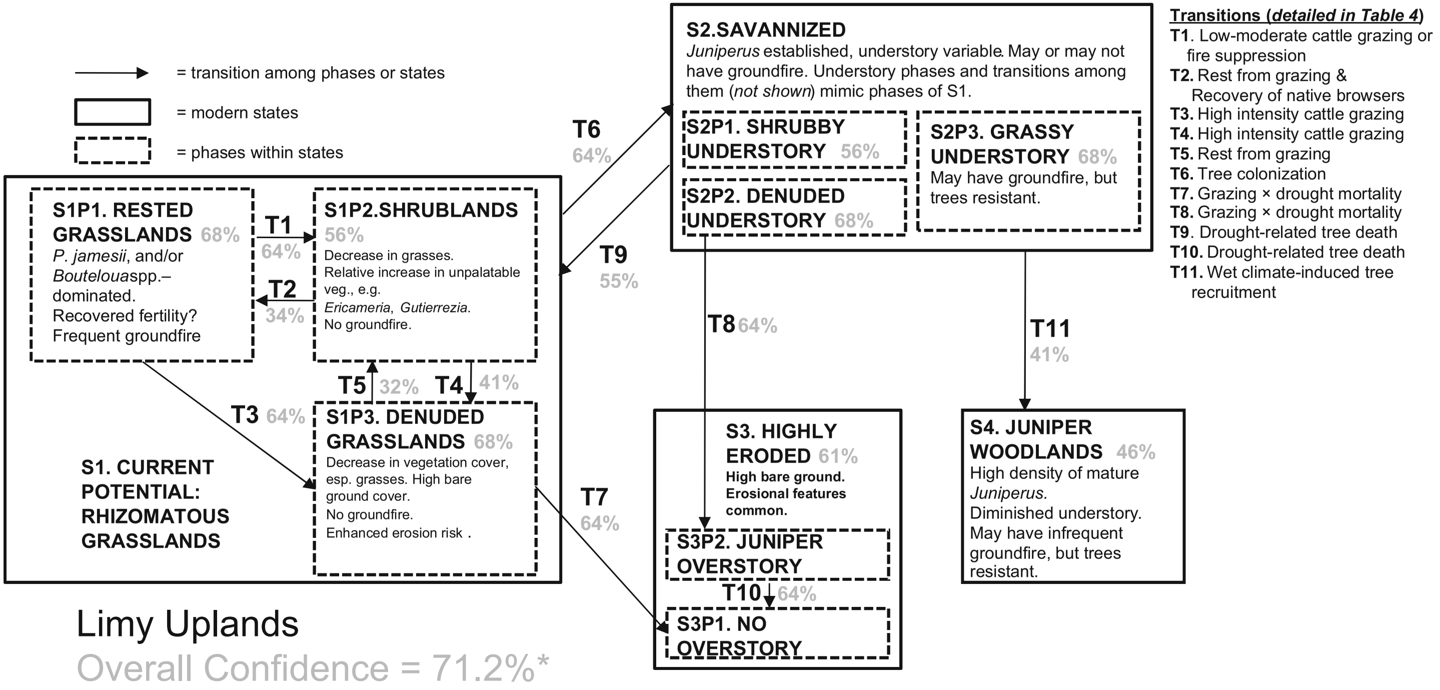


Fig. 7.3 State-and-transition diagram for limy uplands. Overall confidence values also apply to any model component (state, phase, and transition) for which no confidence estimate is provided (*gray text*)

Table 7.3 Catalog of states and phases in limy uplands

Phase	Structural properties	Functional properties	Feedback
P1. Rested grassland	Grassland: *P. jamesii*, and/or *Bouteloua* Bouteloua spp., *H. comata* well represented[a,b]	Presumed recovered productivity equal or greater than Pre1; possibly recovered soil fertility; otherwise similar tp Pre2	Frequent ground fires (15–20 year return),[a,c] resprout of rhizomatous grasses, and browsing by *Antilocapra americana* constrain woody plant abundance
P2. Shrubland	Relative increase in unpalatable shrubs (*Ericameria, Gutierrezia, Artemisia*) or cattle-grazing tolerant grasses (e.g., *Bouteloua gracilis*)[d,e]	Frequent fire cycle of SIP1 interrupted due to loss of connectivity or amount of fine fuels[c,e]; at-risk of state transition; otherwise similar to S1P1	Resprout of rhizomatous grasses, after cattle grazing confers resilience[b,f] improved forage for *A. americana* promotes transition back to grass dominance
P3. Denuded grassland	Relative increase in unpalatable shrubs (*Ericameria, Gutierrezia*), or cattle grazing tolerant grasses (e.g., *Bouteloua gracilis*)[d]; increased bare ground (may be extreme)[e] *Juniperus* may begin colonizing[e]	Frequent fire cycle of SIP1 interrupted due to extreme loss of connectivity and amount of fine fuels[c,e]; at-risk of state transition; otherwise similar to S1P1	Resprout of rhizomatous grasses, rapid colonization of shrubs, after cattle grazing confers resilience[b,f]
P1. Shrubby understory	Understory similar to S1P2; *Juniperus* established in site[a,b,c]	Frequent fire cycle of SIP1 interrupted due to loss of connectivity and amount of fine fuels[c,e]	Same as S1P2 in understory
P1. Denuded understory	Understory similar to S1P3; *Juniperus* established in site[a,c]	Frequent fire cycle of SIP1 interrupted due to extreme loss of connectivity and amount of fine fuels[c,j]	Same as S1P3 in understory
P2. Grassy understory	Understory similar to S1P1, *Juniperus* established in site[a,c]	Recovered connectivity and amount of fine fuel in understory; Except for overstory functionally similar to S1P1	Frequent ground fires (15–20 year return)[a,c] and browsing by *Antilocapra americana* prevent new woody plant colonization, but does not cull extant *Juniperus*[c]
P1. Highly eroded—no overstory	Low vegetation and high bare ground cover	Productivity too low to temper erosivity, declining soil fertility, erosional features apparent	Lack of vegetation allows erosion, erosion prevents recolonization

Table 7.3 (continued)

Phase	Structural properties	Functional properties	Feedback
P2. Highly eroded—Juniper overstory	Same as S3P2, except *Juniperus* established in site	Same as S3P2 in understory	Lack of vegetation allows erosion, erosion prevents recolonization
n.a.	Increased frequency, cover of *Juniperus*,[c] decreased understory due to shading and litter deposition	May have less frequent ground fire, but mature trees not culled	*Juniperus* reduces fire susceptibility, which favors *Juniperus*

[a] Cinnamon 1988, [b] USDA-SCS 1971, [c] Hassler 2006, [d] Jameson 1962, [e] Sullivan and Downum 1991, [f] Stone and Downum 1999

improved in terms of its ability to detect changes in fire susceptibility, since fire occurrence is a resilience mechanism. We recommend refinement and implementation of indicators focused directly on fine fuels connectivity (e.g., combustible patch length, interspace length (devoid of combustible materials)). While the total amount of fuels is important, fuel arrangement in space may be equally informative. A site-specific fire susceptibility model, using these same indicators, would be a highly useful tool to predict the effects of monitorable variables upon site resiliency, which is based upon the fire return cycle. Fire susceptibility may function as a more anticipatory indicator than vegetation structure alone. Such a model could provide a simulation-based confirmation of transition dynamics, and assessment/tipping point estimates, and some degree of forecasting ability, such as the most probable location of the next fire. The role of periods of above-average precipitation in the savannization phenomenon should also be investigated both retrospectively, and using simulation modeling of future climate.

This case study is an example of a situation where monitoring can be applied for scientific or learning processes (Nichols and Williams 2006). As understanding of this ecosystem advances, the monitoring program could move towards a focused tool for decision making.

Discussion

Our operational approach to evaluating threshold dynamics for upland ecological sites in dryland systems offers a variety of advantages:

1. State-and-transition models for individual ecological sites specifically articulate hypotheses regarding reference conditions and ecosystem dynamics in the context of goals for management and monitoring. Attributes of alternative states help to identify biophysical features that may be indicators of an impending transition (threshold crossing). Listing known or hypothesized mechanisms and processes underlying transitions among alternative states and phases also aids in identifying indicators to be monitored. This helps guide quantitative and qualitative estimation of tipping points,

Table 7.4 Key to transitions in Fig. 7.3

Transition	Trigger(s)	Associated process(es)	Relevant indicator(s)
T1	Introduction of persistent light to moderate cattle grazing, associated reduction of native browsers; fire suppression	Reduced amount/connectivity of fine fuels (e.g., grass) leading to interrupted fire cycle	Stocking rate, cowpie density, *A. americana* pellet density, total or basal cover (incl. litter), shrub: grass cover, bare and combustible patch size, time since fire
T2[a]	Cessation/reduction of cattle grazing fire—wild or controlled *Antilocapra americana* browsing	Recovered amount/connectivity of fine fuels (grasses) leading to restored fire cycle	Rest period length, total or basal cover (incl. litter), *A. americana* pellet density, shrub: grass cover, bare and combustible patch size, time since fire
T3[a]	High intensity cattle grazing with little rest (similar to pre-Taylor Grazing Act), associated reduction of native browsers	Strong reduction in amount/connectivity of fine fuels leading to interrupted fire cycle	Stocking rate, cowpie density, *A. americana* pellet density, total or basal cover (incl. litter), bare and combustible patch size
T4[a]	Same as T3	Same as T3	Same as T3
T5[a]	Cessation of cattle grazing	Recolonization of vegetation, including resprouting shrubs and grasses or persistent wet conditions	Rest period length, pellet density, total or basal cover (incl. litter), bare and combustible patch size
T6[b]	Tree colonization (linked to T1, T3, T4)	If seed source exists, *Juniperus* may establish due to lack of fire	Frequency/density of trees, tree height
T7[a]	Sustained high-intensity grazing possibly in concert with drought	Vegetation loss allows erosion, high erosion rates prevent recolonization	Rills, gullies, terracettes total plant cover
T8	Same as T7	Same as T7	Same as T7
T9[c,d]	Interaction of extreme drought, high temperatures, edaphic/physiographic stressors	Hydraulic failure of trees, loss of overstory	Percent of tree mortality
T10	Same as T9	Same as T9	Same as T9
T11[d]	Climate change-linked prolonged wet period	Major recruitment and establishment of *Juniperus*	Same as T6

[a] Cinnamon 1988
[b] USDA-SCS 1971
[c] Hassler 2006
[d] Jameson 1962

Table 7.5 Estimates of tipping and assessment points based upon expert opinion surveys in the limy uplands case study

Indicators	Assessment point[a]	Mean adj. confidence	Maximal adj. confidence	Number of respondents	Tipping point	Mean adj. confidence	Maximal adj. confidence	Number of respondents
Time since fire (y)	22.4	37 %	48 %	4	28.0	40 %	48 %	3
Total plant cover (%)	18.8	40 %	51 %	6	7.7	39 %	51 %	5
Basal cover incl. litter (%)	17	38 %	51 %	5	9.7	37 %	51 %	4
Interspace length (cm)	66.7	42 %	42 %	4	–	–	–	–
Average tree height (m)	0.86	42 %	42 %	3	1.62	27 %	42 %	3

– indicates that fewer than three respondents supplied an estimate, thus these estimates are omitted

[a] Value represents the mean assessment point supplied by survey respondents, and functions only as a guide for where managers would place such a subjective value or values

and establishment of assessment points for monitoring purposes. In dryland systems, resource managers use ecological sites to stratify sampling in monitoring programs due to the likelihood that dynamics will vary among ecological site types (e.g., Herrick et al. 2005, 2006; O'Dell et al. 2005; Thomas et al. 2006). Applying STMs and associated threshold-related assessments to individual ecological sites provides results specific to individual ecosystems and their unique management challenges.

2. This approach enables monitoring for focused management decision making, by narrowing the breadth of information to monitor. Theoretically, the number of possible threshold triggers affecting an ecological site and resulting pathways can be unlimited. In developing an STM, there is a natural rendering of this unlimited number to those known to occur from past observation, or perceived to be highly plausible based on logic and inductive reasoning (i.e., experience with other dryland systems or ecological site types). This more limited and practical domain is more understandable by managers, and preventative and remediation actions can be prescriptive for specific conditions and alternative states. Furthermore, explicit consideration of key-change agents and associated management actions in STMs promotes monitoring for management decision making (qv. Nichols and Williams 2006). A major barrier to monitoring for active conservation is a lack of explicit representations of hypotheses about ecosystem responses to management actions, climate, and other drivers of ecosystem dynamics. Formalizing current system knowledge in STMs is an initial and critical step for focused discussion and understanding of useful indicators for monitoring, and for designing responsible and efficient monitoring efforts to inform management actions.

3. We provide a quantitative approach to estimate tipping and assessment points using data. An ideal dataset for the estimation of assessment and tipping points would consist of a well-replicated experimental manipulation of stressors where quantitative sampling of multiple key indicators in a time series would capture the progression of a transition. Such data resources are the minority, whereas data employing space-for-time replacement tend to be much more available. Within one or a few points in time, samples are obtained that represent spatially discrete examples of different states and phases. Since the transitions are not actually documented in the data, it is assumed that the hypothesized states and transitions articulated in the STM are the correct model of ecosystem dynamics; observed degraded states are assumed to have transitioned in the past from other states due to the model-specified mechanisms. Statistical assessments relying on cluster analysis and the quantification of differences among clusters defines state membership, and indicator values, most useful for distinguishing among states, represent operational tipping points. Assessment points for the identified indicators can be specified on the basis of the natural variation in the less-degraded state. Identifying key indicators and status associated with vulnerable phases or threshold crossings enables managers and scientists to ascribe meaningful and useful assessment points to ensure detection of a changing resource, and to provide sufficient response time to prevent resource degradation or loss. This approach can be applied to the majority of cases for which there are available data; the basic requirements are hypothesized ecosystem dynamics and datasets which are able to capture multiple ecosystem states.

4. We provide a nonempirical, partially quantitative approach to modeling ecosystem dynamics and estimating tipping and assessment points in the absence of data. We developed a practical, qualitative approach to developing STMs and describing system dynamics where empirical data are sparse or lacking. This may be the dominant data-availability scenario in dryland ecological sites of the Colorado Plateau. To accommodate these situations, we developed a Delphi-like protocol to use expert opinion and experience of resource managers and scientists to develop an STM, and to begin to identify system attributes of impending thresholds and of alternative states after a threshold crossing. The Delphi method is based on the principle that group judgment is more accurate than individual judgment. Delphi methods attempt to estimate an unknown quantity (e.g., probability of an event occurring) by asking an anonymous expert panel their opinions in isolation (Linstone and Turoff 1975; Oliver 2002). Multiple iterations allow respondents to change their answer, based upon the anonymous responses of other members, until convergence is achieved on a single value or a narrower range of values. We used some of the principles of this approach, but did not seek convergence. We used the respondents' confidence in their own responses as weights in a procedure analogous to model averaging. In this way, we arrived at quantitative estimates of both assessment points and tipping points in a few indicators along a transition sequence in only one iteration. We found this method to be reasonably efficient, requiring only 2 months and two surveys; however, it was difficult to obtain sufficient information on most indicators. Further, rather than seeking consensus, confidence estimation provides an additional product measuring respondents' self-assessed level of uncertainty about an issue and identifies the most pressing needs for evidence.

Critics of similar expert-opinion methods suggest that such approaches only serve to boost confidence in respondents' ignorance. However, the dominant practice in resource conservation tends to be based on the experiential knowledge of individuals, rather than high-quality data or organized group judgment (Cook et al. 2010). We present our expert-opinion protocol as an improvement over the experiential knowledge of individuals that can be applied to identify critical indicator levels in monitoring any ecosystem. This approach can be applied more quickly and cheaply than a scientific study, giving it much utility when time or funds are limiting. Weighted averages of group assessment and tipping point estimates provide an intermediate level of quantitative data quality, higher than individual judgment and lower than quantitative field and experimental data. We do not consider a model produced using this procedure to be final, rather it is a first iteration of a useful model which should be refined as more information becomes available. Estimates of model parameters can serve to inform prior information in later Bayesian estimation using data.

Concluding Remarks Monitoring efforts by the NPS I&M networks are unlikely to attain their full potential without a clear understanding of vulnerable conditions and tipping points associated with ecological thresholds; however, the strength of these monitoring efforts is that they anticipate the development of this understanding. Scientific research and synthesis must provide the missing information. The two approaches we used in this chapter have the potential to provide a credible basis for establishing assessment points for these monitoring efforts. Estimates of assessment point values are surprisingly rare in the literature (but see Digiovinazzo et al. 2010), yet they seem crucial to the goal of applying threshold concepts to management problems. This goal is consistent with application of a preventive threshold: Attaining an assessment point of one or more indicators could trigger regulation of "changes to patterns that make systems vulnerable to deterministic or event-driven change" so that the undesired transition never occurs (Bestelmeyer 2006). In conservation and resource management, decisions must often be made regardless of the level of confidence in our knowledge of ecosystems (Soulé 1985; Cook et al. 2010). Our goal should be to develop the best set of models possible given the level of information available to support decisions. The approach presented here offers a flexible means of achieving this goal, and determining specific research areas in need of study.

Acknowledgments This work was supported by funding from the US Geological Survey (Status and Trends of Biological Resources Program and Southwest Biological Science Center), the NPS Northern and Southern Colorado Plateau I&M programs, and The Nature Conservancy of Utah, with additional assistance and cooperation from Canyonlands National Park and the Bureau of Land Management. We thank Lisa Thomas and Dusty Perkins of the NPS for their interest and support of this effort. We especially thank Steve Cinnamon formerly of the NPS, Billy Cardasco of Babbit Ranches, Dr. Christoper McGlone of the Ecological Restoration Institute, Dr. Matthew Loeser of Yakima Valley Community College, and Kirsten Ironside, Monica McTeague, and Dr. Kathryn Thomas of the US Geological Survey for devoting valuable time to expert opinion questionnaires. Drs. Chris Lauver, Brandon Bestelmeyer, Jayne Belnap, and Craig D. Allen contributed original ideas and feedback on this project. We thank Glenn Guntenspergen for constructive comments that improved this chapter.

References

Allen-Diaz, B., and J. W. Bartolome. 1998. Sagebrush-grass vegetation dynamics: Comparing classical and state-and-transition models. *Ecological Applications* 8:795–804.

Allen, C. D., M. Savage, D. A. Falk, K. F. Suckling, T. Schulke, T. W. Swetnam, P. B. Stacey, P. Morgan, M. Hoffman, and J. T. Klingel. 2002. Ecological restoration of southwestern ponderosa pine ecosystems: A broad perspective. *Ecological Applications* 12:1418–1433.

Beisner, B. E., D. T. Haydon, and K. Cuddington. 2003. Alternative stable states in ecology. *Frontiers in Ecology and the Environment* 1:376–382.

Belnap, J. 1995. Surface disturbances: Their role in accelerating desertification. *Environmental Monitoring and Assessment* 37:39–57.

Belnap, J. 2003. The world at your feet: desert biological soil crusts. *Frontiers in Ecology and the Environment* 1:181–189.

Belnap J, and D. J. Eldridge. 2003. Disturbance and recovery of biological soil crusts. In *Biological soil crusts: Structure, function, and management*, ed. J. Belnap, and O. L. Lange, 363–383. 2nd ed. Berlin: Springer-Verlag.

Belnap, J., R. Prasse, and K. T. Harper. 2003. Influence of biological soil crusts on soil environments and vascular plants. In *Biological soil crusts: Structure, function, and management*, ed. J. Belnap, and O. L. Lange, 281–300. 2nd ed. Berlin: Springer-Verlag.

Belnap, J., R. L. Reynolds, M. C. Reheis, S. L. Phillips, F. E. Urban, and H. L. Goldstein. 2009. Sediment losses and gains across a gradient of livestock grazing and plant invasion in a cool, semi-arid grassland, Colorado Plateau, USA. *Aeolian Research* 1:27–43.

Bennetts, R. E., J. E. Gross, K. Cahill, C. McIntyre, B. Bingham, A. Hubbard, L. Cameron, and S. L. Carter. 2007. Linking monitoring to management and planning: assessment points as a generalized approach. *The George Wright Forum* 24:59–77.

Bestelmeyer, B. T. 2006. Threshold concepts and their use in rangeland management and restoration: The good, the bad, and the insidious. *Restoration Ecology* 14:325–329.

Bestelmeyer, B. T., J. R. Brown, K. M. Havstad, G. Chavez, R. Alexander, and J. E. Herrick. 2003. Development and use of state-and-transition models for rangelands. *Journal of Range Management* 56:114–126.

Bestelmeyer, B. T., A. J. Tugel, G. L. Peacock, D. G. Robinett, P. L. Shaver, J. R. Brown, J. E. Herrick, H. Sanchez, and K. M. Havstad. 2009. State-and-transition models for heterogeneous landscapes: A strategy for development and application. *Rangeland Ecology and Management* 62:1–15.

Bowker, M. A., and J. Belnap. 2008. A simple classification of soil types as habitats of biological soil crusts on the Colorado Plateau, USA. *Journal of Vegetation Science* 19:831–840.

Briske, D. D., S. D. Fuhlendorf, and F. E. Smeins. 2006. A unified framework for assessment and application of ecological thresholds. *Rangeland Ecology and Management* 59:225–236.

Briske, D. D., B. T. Bestelmeyer, T. K. Stringham, and P. L. Shaver. 2008. Recommendations for development of resilience-based state-and-transition models. *Rangeland Ecology and Management* 61:359–367.

Chapin III, F. S., M. S. Torn, and M. Tateno. 1996. Principles of ecosystem sustainability. *The American Naturalist* 148:1016–1037.

Cinnamon, S. K. 1988. The vegetation community of Cedar Canyon, Wupatki National Monument as influenced by prehistoric and historic environmental change. MS Thesis, Northern Arizona University.

Cook, C. N., M. Hockings, and R. W. Carter. 2010. Conservation in the dark? The information used to support management decisions. *Frontiers in Ecology and the Environment* 8:181–186.

Davenport, D. W., D. D. Breshears, B. P. Wilcox, and C. D. Allen. 1998. Viewpoint: Sustainability of pinyon-juniper ecosystems—A unifying perspective of soil erosion thresholds. *Journal of Range Management* 51:231–240.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees; a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178–3192.

DeCoster, J. K., and M. C. Swan. 2009. Integrated upland vegetation and soils monitoring for Wupatki National Monument: 2008 summary report. Natural Resource Data Series NPS/SCPN/NRDS—2009/022. National Park Service, Fort Collins, Colorado.

Digiovinazzo, P., G. F. Ficetola, L. Bottoni, C. Adnreis, and E. Padoa-Schioppa. 2010. Ecological thresholds in herb communities for the management of suburban fragmented forests. *Forest Ecology and Management* 259:343–349.

Equihua, M. 1990. Fuzzy clustering of ecological data. *Journal of Ecology* 78:519.

Escudero, A., I. Martinez, A. de la Cruz, M. A. G. Otalora, and F. T. Maestre. 2007. Soil lichens have species-specific effects on the seedling emergence of three gypsophile plant species. *Journal of Arid Environments* 70:18–28.

Evans, R. D., and O. L. Lange. 2003. Biological soil crusts and ecosystem nitrogen and carbon dynamics. In *Biological soil crusts: Structure, function, and management*, eds. J. Belnap, and O. L. Lange, 263–279. 2nd ed. Berlin: Springer-Verlag.

Fancy, S. G., J. E. Gross, and S. L. Carter. 2009. Monitoring the condition of natural resources in US national parks. *Environmental Monitoring and Assessment* 151:161–174.

Field, S. A., A, J. Tyre, N. Jonzén, J. R. Rhodes, and H. P. Possingham. 2004. Minimizing the cost of environmental management decision by optimizing statistical thresholds. *Ecology Letters* 7:669–675.

Gitlin, A. R., C. M. Sthultz, M. A. Bowker, S. Stumpf, K. L. Paxton, K. Kennedy, A. Munoz, J. K. Bailey, and T. G. Whitham. 2006. Mortality gradients within and among dominant plant populations as barometers of ecosystem change during extreme drought. *Conservation Biology* 20:1477–1486.

Groffman, P., J. Baron, T. Blett, A. Gold, I. Goodman, L. Gunderson, B. Levinson, M. Palmer, H. Paerl, G. Peterson, N. Poff, D. Rejeski, J. Reynolds, M. Turner, K. Weathers, and J. Wiens. 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? *Ecosystems* 9:1–13.

Hansen, M., J. Coles, K. A. Thomas, K. Cogan, M. Reid, J. Von Loh, and K. Schulz. 2004. USGS-NPS National Vegetation Mapping Program: Wupatki National Monument, Arizona, Vegetation Classification and Distribution. U.S Geological Survey, Southwest Biological Science Center.

Hassler, F. 2006. Dynamics of juniper invaded grasslands and old growth woodlands at Wupatki National monument, Northern Arizona, USA. MS Thesis, Northern Arizona University.

Herrick, J. E. 2000. Soil quality: An indicator of sustainable management? *Applied Soil Ecology* 15:75–84.

Herrick, J. E., J. W. Van Zee, K. M. Havstad, L. M. Burkett, and W. G. Whitford. 2005. Monitoring manual for grassland, shrubland and savanna ecosystems. Vol I: Quick start. Las Cruces: USDA-ARS Jornada Experimental Range.

Hobbs, R. J., and D. A. Norton. 1996. Towards a conceptual framework for restoration ecology. *Restoration Ecology* 4:93–110.

Ironside, K. 2006. Climate change research in national parks; paleoecology, policy, and modeling the future. MS Thesis, Northern Arizona University.

Jameson, D. A. 1962. Effects of burning on a Galleta-Black Grama Range invaded by Juniper. *Ecology* 43:760–763.

King, E. G., and R. J. Hobbs. 2006. Identifying linkages among conceptual models of ecosystem degradation and restoration: Towards an integrative framework. *Restoration Ecology* 14:369–378.

Kleiner, E. F., and K. T. Harper. 1972. Environment and community organization in grasslands of Canyonlands National Park. *Ecology* 53:299–309.

Linstone, H. A., and M. A. Turoff, eds. 1975. The Delphi method: Techniques and applications. Reading: Addison-Wesley.

Miller, M. E. 2008. Broad-scale assessment of rangeland health, Grand Staircase-Escalante National Monument, USA. *Rangeland Ecology & Management* 63:249–262.

Miller, M. E., D. L. Witwicki, R. K. Mann, and N. J. Tancreto. 2007. Field evaluation of sampling methods for long-term monitoring of upland ecosystems on the Colorado Plateau. USGS Open File Report 2007-1243.

Miller, M. E., R. T. Belote, M. A. Bowker, and S. L. Garman. 2011. Alternative states of a semiarid grassland ecosystem: implications for ecosystem services. *Ecosphere* 2:art55.

Neff, J. C., R. L. Reynolds, J. Belnap, P. Lamothe. 2005. Multi-decadal impacts of grazing on soil physical and biogeochemical properties in southeast Utah. *Ecological Applications* 15:87–95.

Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. *Trends in Ecology & Evolution* 21:668–673.

O'Dell, T., S. Garman, A. Evenden, M. Beer, E. Nance, D. Perry, R. DenBleyker, D. Sharrow, K. Wynn, J. Brown, M. Miller, and L. Thomas. 2005. Northern Colorado Plateau Inventory and Monitoring Network, Vital Signs Monitoring Plan, National Park Service, Inventory and Monitoring Network, Moab, UT.

Okin, G. S. 2008. A new model of wind erosion in the presence of vegetation. *Journal of Geophysical Research* 113:F02S10.

Okin, G. S., A. J. Parsons, J. Wainwright, J. E. Herrick, B. Bestelmeyer, D. C. Peters, and E. L. Fredrickson. 2009. Do changes in connectivity explain desertification? *BioScience* 59:237–244.

Oliver, I. 2002. Introduction to an expert panel based approach for assessment of vegetation condition within the context of biodiversity conservation. *Ecological Management and Restoration* 3:227–229.

Paine, R. T., M. J. Tegner, and E. A. Johnson. 1998. Compounded perturbations yield ecological surprises. *Ecosystems* 1:535–545.

Peters, D. C., R. A. Pielke Sr., B. T. Bestelmeyer, C. D. Allen, S. Munson-McGee, and K. M. Havstad. 2007. Cross-scale interactions, nonlinearities, and forecasting catastrophic events. *Proceedings of the National Academy of Sciences of U S A* 101:15130–15135.

Rapport, D. J., and W. G. Whitford. 1999. How ecosystems respond to stress: Common properties of arid and aquatic systems. *BioScience* 49:193–203.

Scheffer, M., S. H. Hosper, M. L. Meijer, B. Moss. 1993. Alternative equilibria in shallow lakes. *Trends in Ecology and Evolution* 8: 275–279.

Scheffer, M., S. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413:591–596.

Scott, M. L., A. M. D. Brasher, E. Reynolds, A. Caires, and M. E. Miller. 2005. The structure and functioning of riparian and aquatic ecosystems of the Colorado Plateau: Conceptual models to inform long-term ecological monitoring. U.S. Geological Survey, Fort Collins Science Center, Fort Collins, CO.

Soulé, M. E. 1985. What is conservation biology? *Bioscience* 35:727–734.

Stone, G. D., and C. E. Downum. 1999. Non-Boserupian ecology and agricultural risk: Ethnic politics and land control in the arid Southwest. *American Anthropologist* 101:113–128.

Stringham, T. K., W. C. Krueger, and D. R. Thomas. 2001. Application of non-equilibrium ecology to rangeland riparian zones. *Journal of Range Management* 54:210–217.

Suding, K. N., R. J. Hobbs. 2009. Threshold models in restoration and conservation: A new framework. *Trends in Ecology and Evolution* 24:271–279.

Suding, K. N., K. L. Gross, G. R. Houseman. 2004. Alternative states and positive feedbacks in restoration ecology. *Trends in Ecology and Evolution* 19:46–53.

Sullivan, A. P., and C. E. Downum. 1991. Aridity, activity, and volcanic ash agriculture: A study of short-term prehistoric cultural-ecological dynamics. *World Archaeology* 22:271–287.

Thomas, L. P., M. N. Hendrie, C. L. Lauver, S. A. Monroe, N. J. Tancreto, S. L. Garman, M. E. Miller. 2006. Vital signs monitoring plan for the Southern Colorado Plateau Network, Natural Resource Report NPS/SCPN/NRR-2006/002, National Park Service, Flagstaff, AZ.

USDA-NRCS. 2010. The PLANTS database (21 July 2010). U.S. Department of Agriculture, Natural Resource Conservation Service, Baton Rouge, LA. http://plants.usda.gov. Accessed 21 July 2010.

USDA-SCS. 1971. Soil survey and range site and condition inventory. Wupatki National Monuments, Arizona: A special report. US Department of Agriculture, Soil Conservation Service, Washington, DC.

USDA-SCS. 1983. Soil survey of Coconino County area, Arizona, Central Part. US Department of Agriculture, Soil Conservation Service, Washington, DC.

USDI-NPS. 2002. Final environmental impact statement general management plan. Wupatki National Monument, Arizona. US Department of the Interior, National Park Service, Flagstaff, AZ.

USDI-NPS. 2003. Interim technical guidance on assessing impacts and impairment to natural resources. U.S. Department of Interior, National Park Service, Natural Resource Program Center, Fort Collins, CO.

USDI-NPS. 2005. Fire Management Plan Environmental Assessment/Assessment of effect. Department of the Interior. Flagstaff Area National Monuments, Flagstaff, AZ.

USDI-NPS. 2006. Management policies 2006: The guide to managing the National Park System. U.S. Department of Interior, National Park Service, Washington, DC.

Walker, B., and J. A. Meyers. 2004. Thresholds in ecological and social-ecological systems: a developing database. *Ecology and Society* 9:3. http://www.ecologyandsociety.org/vol9/iss2/art3/. Accessed 22 June 2010.

Warren, S. D. 2003. Synopsis: Influence of biological soil crusts on arid land hydrology and soil stability. In *Biological soil crusts: Structure, function, and management*, ed. J. Belnap, and O. L. Lange, 349–360. 2nd ed. Berlin: Springer-Verlag.

Whisenant, S. G. 1999. *Repairing damaged wildlands: A process-oriented, landscape-scale approach*. Cambridge: Cambridge University Press.

Chapter 8
Using Natural Range of Variation to Set Decision Thresholds: A Case Study for Great Plains Grasslands

Amy J. Symstad and Jayne L. Jonas

Abstract Natural range of variation (NRV) may be used to establish decision thresholds or action assessment points when ecological thresholds are either unknown or do not exist for attributes of interest in a managed ecosystem. The process for estimating NRV involves identifying spatial and temporal scales that adequately capture the heterogeneity of the ecosystem; compiling data for the attributes of interest via study of historic records, analysis and interpretation of proxy records, modeling, space-for-time substitutions, or analysis of long-term monitoring data; and quantifying the NRV from those data. At least 19 National Park Service (NPS) units in North America's Great Plains are monitoring plant species richness and evenness as indicators of vegetation integrity in native grasslands, but little information on natural, temporal variability of these indicators is available. In this case study, we use six long-term vegetation monitoring datasets to quantify the temporal variability of these attributes in reference conditions for a variety of Great Plains grassland types, and then illustrate the implications of using different NRVs based on these quantities for setting management decision thresholds. Temporal variability of richness (as measured by the coefficient of variation, CV) is fairly consistent across the wide variety of conditions occurring in Colorado shortgrass prairie to Minnesota tallgrass sand savanna (CV 0.20–0.45) and generally less than that of production at the same sites. Temporal variability of evenness spans a greater range of CV than richness, and it is greater than that of production in some sites but less in other sites. This natural temporal variability may mask undesirable changes in Great Plains grasslands vegetation. Consequently, we suggest that managers consider using a relatively narrow NRV (interquartile range of all richness or evenness values observed in reference conditions) for designating a surveillance threshold, at which greater attention to the situation would be paid, and a broader NRV for designating management thresholds, at which action would be instigated.

A. J. Symstad (✉)
Northern Prairie Wildlife Research Center, U.S. Geological Survey,
26611 U.S. Highway 385, Hot Springs, SD 57747, USA
e-mail: asymstad@usgs.gov

J. L. Jonas
Department of Forest and Rangeland Stewardship, Colorado State University,
Fort Collins, CO, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_8,

Keywords Natural range of variation · Great Plains grasslands · Prairie · Plant species richness · Plant species evenness · Diversity · Temporal variability · Surveillance threshold · Management threshold

Introduction

Ecological thresholds are an appealing concept for natural resource management because they provide dramatic, drastic pictures of the consequences of mismanagement to a broad audience. Unfortunately, predicting the conditions that precede the crossing of an ecological threshold is notoriously difficult (Scheffer and Carpenter 2003; Thrush et al. 2009; Hastings and Wysham 2010). Conversely, not all ecosystems exhibit threshold behavior (Stafford Smith 1996; Bagchai et al. 2012), or ecological thresholds may occur far outside the range of conditions maintained by management, as is often the case in protected natural areas like national parks. These conditions do not preclude the establishment of decision thresholds— values of ecosystem state variables that prompt changes in management actions. Instead, managers can establish decision thresholds, also known as action thresholds (Ford et al. 1999), management thresholds (Bennetts et al. 2007), or action-assessment points (Mitchell et al., Chap. 10), based on the natural range of variation (NRV) of the ecosystem attributes they are monitoring.

NRV is a concept with many names, including "range of natural variation," "historical range of variation," "natural variability," and "reference variability," with some authors preferring "historic" over "natural" because it implies that the effects of indigenous people on ecosystems are included, and because it avoids the ambiguity of the term "natural" (Egan and Howell 2001b). A similar concept is "reference condition," used frequently in assessing the ecological integrity of streams and wetlands, but with less emphasis on range and variation than the other terms (Stoddard et al. 2006). We use "natural range of variation" to be consistent with Mitchell et al. (Chap. 10). Regardless of its exact name, the concept was developed to recognize that ecosystems are dynamic, but that their dynamics operate within bounds that remain relatively consistent over time (Morgan et al. 1994). To some, it also implies that the ecosystem is self-sustaining within the range of these bounds, but outside the range the system becomes unrecognizable (Egan and Howell 2001b) and, presumably, difficult to return to its original condition. In this context, NRV is related to ecological thresholds, in that it assumes there is a point beyond which an ecosystem will shift to a different state (Groffman et al. 2006). The driver behind this shift can be natural (e.g., a strong hurricane) or anthropogenic (e.g., nutrient enrichment of water bodies from agricultural runoff). Either way, there is no inherent assumption in the concept that the small step from inside to outside a system's NRV will result in a large, abrupt change in an ecosystem quality, property, or phenomenon (Unnasch et al. 2009).

On the other hand, NRV can also be a useful concept when an ecosystem property is of management interest, but the quality may not change substantially even as the ecosystem as a whole crosses a threshold to a different state. For example, water yield from a watershed may increase dramatically after a fire kills all of the trees in the watershed. Although the forest may have shifted to a long-lived grassland state as a result of the fire, the dramatic increase in water yield may be short lived, returning to prefire conditions in just a couple of years when the grasses become well established. Water managers would be interested in the NRV in this ecosystem property across the ecological threshold, and the temporary increase would simply be a part of that variation.

Plant diversity in the grasslands of North America's Great Plains is an ecosystem quality important to managers in this region, but its behavior is not well understood with respect to specific management practices or ecological thresholds. Consequently, NRV is a pragmatic approach for determining decision thresholds in this situation. In this chapter, we first outline the general process for determining the NRV of an ecosystem quality, then we use the plant diversity of Great Plains grasslands (GPG) to illustrate this process and, in so doing, provide specific values of these qualities and their NRV for GPG managers—information not available elsewhere. Finally, we discuss potential decision thresholds for specific GPG plant communities based on these values, and the implications of defining NRV in different ways.

A General Process for Quantifying Natural Range of Variation

Assuming that the community or ecosystem of interest is already well defined, the first step in describing NRV is to determine which attributes of that ecosystem will be used to describe that NRV. The attributes can be any of a wide variety of processes and properties, but they of course must be relevant to the management issue at hand and sufficient information about them must be available. A large part of the literature on NRV focuses on fire return intervals in forested systems and the resulting distribution of forest ages and types across the landscape (Bergeron et al. 2004; Carlson and Kurz 2007; Doyon et al. 2008; Mori and Lertzman 2011), but other attributes for which NRV has been quantified include spruce beetle irruption frequency and extent in Alaskan boreal forests (Sherriff et al. 2011), net ecosystem production in tropical forest (Sierra et al. 2007), and magnitude, frequency, and duration of river flows, as well as the spatial distribution and diversity of specific geomorphological forms, in the Colorado Front Range and Florida Everglades (Harwell 1997; Wohl 2011).

Second, appropriate spatial and temporal scales must be identified for the attributes of interest. Both must be broad enough that they allow for variation or heterogeneity in the ecosystem qualities of interest, but narrow enough that they encompass an ecosystem that is relatively consistent in terms of climatic, edaphic, topographic, and biogeographic conditions that are relevant to the management issue (Morgan et al. 1994). For example, differences in valley geometry, as well as variations in vegetation and hydrological flow regimes associated with elevation, translate into different NRVs among reaches within streams and among streams in mountainous regions (Wohl 2011). Bergeron et al. (2004) used dendrochronological techniques to estimate the mean fire intervals for mixed and coniferous boreal forests in eastern Canada prior to 1850, when European settlers began to impact the fire behavior of the region. The authors later decided that the length of time covered using these techniques (300–400 years) was too short to adequately capture the NRV given the long life span of the tree species and communities that they were investigating, as well as the somewhat anomalous climate conditions of the 1770–1850 period (Cyr et al. 2009). Other issues to consider when choosing specific locations from which to gather NRV information include the location's history of management and other human influences, and the presence and abundance of exotic species (Landres et al. 1999).

Next, information must be compiled and translated into actual values of the attributes of interest. Methods used to do this depend on the choices made above and fall into five basic categories: study of historic records, analysis and interpretation of proxy records, modeling, space-for-time substitutions, and analysis of long-term monitoring data (Morgan et al. 1994; Egan and Howell 2001a). The last of these methods might be considered ideal, but consistently collected data over time periods long enough to address many NRV questions are rare. This is because the data-collection period must not only adequately cover the ecosystem of interest spatially, but also must be long relative to the return interval of external forces driving variability (disturbances, climatic fluctuations, etc.) and to the life span of the organisms of interest. Space-for-time substitutions can be used when a sufficiently large unaltered area, such as a large wilderness area, contains the range of conditions encompassed by the target ecosystem (Morgan et al. 1994). For example, current vegetation could be sampled in areas that have experienced various levels of grazing by native herbivores in order to describe the NRV of composition and productivity in a grassland ecosystem that evolved with these grazers. Dynamic simulation models that incorporate the effects of disturbances and stochastic fluctuations (as in weather) have been used to estimate the NRV of net ecosystem production, fire, and landscape dynamics, for example (Baker 1992; Sierra et al. 2007; Doyon et al. 2008). Models have the advantage of being able to cover a wide range of possible conditions that other methods may not, but they must be adequately calibrated to the location of interest to provide reasonable estimates. All three of these methods are subject to the same primary difficulty of finding an ecosystem minimally impacted by fire suppression, pollution, predator control, and other ecological disruptions caused by modern humans, to monitor, measure, or use to calibrate a model. Proxy methods that construct a chronology of past events based on pollen, microfossils, seeds, tree rings, fire scars, or lake sediments can avoid this problem, but they may not provide the temporal, spatial, or biological resolution desired (Swanson et al. 1994). Early land survey or forest reserve data may provide quantitative historical information (e.g., Graves 1899), but their use may be limited by lack of or unknown rigor, lack of detail, and their scarcity through time and space (Morgan et al. 1994). Finally, historical records such as photographs and explorers' journals can provide qualitative information where quantitative data cannot be obtained. For example, Higgins (1986) compiled and interpreted historical fire accounts from early European explorers' journals of their travels in the northern Great Plains to estimate fire frequency and seasonal distribution, but he pointed out the many limitations of this approach, including potential exaggeration by the journalists. Ideally, values of the attributes of interest will be derived using a combination of approaches so that the weaknesses of each approach are compensated for by the strengths of the other.

After values for the attributes of interest have been derived, the NRV of those attributes must be quantified. This can be done with a variety of metrics, the use

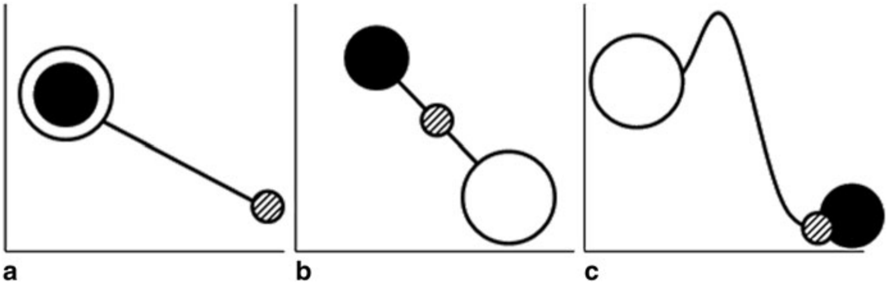


Fig. 8.1 Three potential relationships between natural range of variation (NRV; *open circle*), desired conditions (*filled circle*), and current conditions (*hatched circle*) and ease of path among them (*line*). **a** Desired conditions are within NRV; moving current conditions to either will require effort. **b** Desired conditions differ substantially from NRV, and moving current conditions to desired conditions will be difficult as the tendency is to move towards NRV. **c** A substantial ecological barrier (threshold) makes returning current conditions to within NRV extremely difficult, so desired conditions are in an alternate state

of which depends on the attributes of interest. Mean, median, standard deviation, percentiles, skewness, confidence intervals, and range describe an expected value and the magnitude and shape of the variation around that value. The coefficient of variation (CV; standard deviation divided by mean) is useful for describing the relative magnitude of fluctuations in an attribute through time, and frequency quantifies the rate of recurrence of an event type or ecosystem state. Mean and 95 % confidence intervals appear to be popular metrics for describing NRV (Bergeron et al. 2004; Sierra et al. 2007; Doyon et al. 2008; Cyr et al. 2009), but the full range of measured values is also sometimes used (Carlson and Kurz 2007; Sherriff et al. 2011). Although the central limit theorem ensures that the distribution of the mean of a large number of samples from any population will generally approximate a normal distribution, this does not mean that a given ecological variable will have a normal distribution. Thus, caution should be exercised when using parameters describing a normal distribution (mean, standard deviation) to describe the NRV of an ecological variable. Consequently, as with any quantitative data, it is always wise to perform a variety of exploratory data analyses to understand the shape of the data's distribution (Ellison 2001); this shape may itself be a useful means for describing the NRV (Landres et al. 1999).

Once the NRV is quantified, it is used for its intended purpose, which is usually to evaluate current conditions and determine desired conditions (Fig. 8.1). Desired conditions may be a subset of the NRV if parts of the NRV are not socially acceptable (e.g., intense, stand-replacing fires near urban areas) or if the full NRV is no longer possible due to land development, climate change, extinction, etc. (Fig. 8.1a; Swanson et al. 1994; National Park Service 2009; Thompson et al. 2009; Unnasch et al. 2009; Duncan et al. 2010). Desired conditions that differ substantially from the NRV for the former reason but not the latter may be difficult to attain or maintain without substantial, direct management (Fig. 8.1b). On the other hand, if an ecological threshold has been crossed, current conditions may be substantially outside the system's NRV. This situation may warrant setting the desired conditions outside the

NRV because conditions within the NRV are not feasibly attainable (Fig. 8.1c). For the rest of our discussion, we will focus on the scenario in Fig. 8.1a, where desired conditions lie wholly within the NRV.

Process for Great Plains Grasslands Plant Diversity Natural Range of Variation

Focal Ecosystem

GPG cover an area of approximately 2 million km^2 in mid-continental North America. Vegetation biomass is dominated by grasses, but forbs generally provide much of the diversity. Temperature and precipitation gradients across the region result in a general west-to-east increase in productivity and plant species richness (Teeri and Stowe 1976; Risser et al. 1981) and a gradient of broad grassland types (Lauenroth et al. 1999; Fig. 8.2). Local variations in soils and topography yield a variety of plant assemblages at finer spatial scales. Temporal climate variability is high throughout the region, more so than in the remainder of North America east of the Rocky Mountains (Borchert 1950). These climatic forces, as well as periodic fire and herbivory by large ungulates (bison, elk, and pronghorn), insects, and prairie dogs, shaped the evolution of GPG (Axelrod 1985; Anderson 2006). Today, major factors affecting this highly endangered ecosystem include land use patterns, exotic and invasive species, atmospheric nitrogen deposition, altered fire and grazing regimes, and climate change (Samson and Knopf 1994), and most of these have been shown to affect plant diversity in the region (Symstad and Jonas 2011).

The term “Great Plains grasslands” encompasses far too broad an area for meaningful characterization of NRV of any ecosystem characteristic, but this area hosts at least 19 National Park Service (NPS) units where plant diversity has been identified as an important measure of ecosystem health (DeBacker et al. 2004; Manier et al. 2011; Symstad et al. 2011), as well as a wide variety of other federal, state, and private lands where native grasslands are being managed and restored. In order to serve this broad audience but also provide meaningful values, we present NRV information for one or two sites from four of the five major grassland types in the Great Plains (Fig. 8.2), separated by topoedaphic class within these sites when appropriate.

Attributes of Interest

We focus on plant diversity for three reasons. First, the NPS mission is to preserve and protect the landscapes and organisms within its holdings for the enjoyment of future generations. Consequently, maintaining or restoring diversity is increasingly becoming an explicit management goal of many NPS units. Second, a large body of research investigating the relationship between biodiversity and ecosystem functioning over the past 15 years has shown that greater plant diversity on average results not

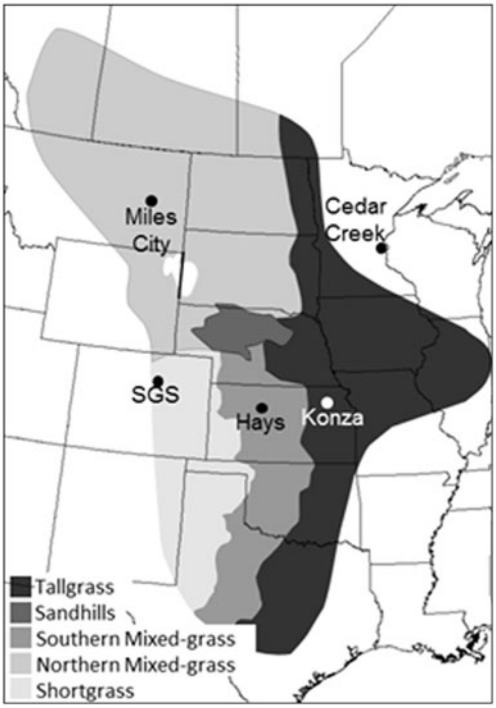


Fig. 8.2 Major grassland types of North America's Great Plains (after Lauenroth et al. 1999), and locations of datasets described in Table 8.1

only in higher production, but also in more stable production (Tilman 2001; Hooper et al. 2005; Balvanera et al. 2006; Cardinale et al. 2006, 2007; Fargione et al. 2007; Schmid et al. 2009; Isbell and Wilsey 2011), and that maintaining high levels of multiple ecosystem functions (e.g., nutrient retention and belowground carbon storage in addition to aboveground production) requires more species than maintaining a high level of just one ecosystem function (Hector and Bagchi 2007; Zavaleta et al. 2010). Consequently, Briske et al. (2006) proposed species loss as one category of threshold that rangelands cross when progressing from a desirable state to an undesirable state. Little information to evaluate this suggestion is available, however. Thus, our third reason for focusing on plant diversity is to begin filling that information gap. The Natural Resource Conservation Service's nationwide effort to describe the dynamics of rangeland (including GPG) vegetation in response to various management practices provides a wealth of information on the variability of dominant plant species (Bestelmeyer et al. 2003, 2009; see Bowker et al. (Chap. 7) for a thorough description of this process), but provides no information on plant diversity.

We use two metrics of plant diversity—species richness and species evenness (hereafter richness and evenness). Richness is the number of species in a given area (i.e., those counted in a sample of fixed area), whereas evenness is a metric quantifying the relative abundance of species in that area. We use the Shannon evenness index, calculated as $(-\sum p_i \ln p_i)/\ln S$, where p_i is the proportional abundance of species i, and S is the total number of species (i.e., richness; Magurran 1988). Evenness ranges from 0 to 1, with values near 0 indicating greater dominance by a single species and values near 1 indicating nearly equal abundance of all species present. We use both of these metrics because they are expected to respond to ecosystem stressors differently. Richness will be more sensitive to stressors that cause the loss of only relatively rare species, whereas a stressor that causes significant shifts in the way resources are partitioned among species could affect evenness without substantially affecting richness. Because of the different information contained in these two metrics, we do *not* use any of the indices that combine richness and diversity into one value and are usually referred to as diversity indices.

Spatial and Temporal Scale

Although plant diversity is an important descriptor of a plant community, any metric describing it is complicated by its sensitivity to the area over which it is sampled. Consequently, we focus on richness and evenness at the −quadrat or transect scale (see Table 8.1 for sizes and explanation of scale chosen for each site) because it is the scale most comparable among management units and between current and reference conditions. In addition, because annual fluctuations in climate are significant drivers in GPG vegetation, we felt it was important to characterize variability with a high temporal resolution—an annual time step.

Information Sources and Approach

Given these attributes, metrics, and goals, the best method for describing NRV in GPG vegetation was analysis of long-term monitoring data. We found six datasets that have sufficient temporal length (> 10 years) and resolution (annual), have enough detail (abundance of individual species recorded in a fixed location) to calculate richness and evenness, and are from relatively unimpacted ecosystems (e.g., fire not excluded, low exotic species abundance). Data come from five sites: the Shortgrass Steppe (SGS), Konza Prairie (Konza), and Cedar Creek (Cedar Creek) long-term ecological research stations, Fort Hays State University (Hays), and Fort Keogh Livestock and Range Research Laboratory (Miles City). Two datasets (Konza FRI and Konza Grazed), from two separate experiments at Konza, were kept separate in our analyses. Table 8.1 describes these datasets, two of which are historic and four of which are modern. We refined the datasets to include only quadrats (transects for Konza datasets; Table 8.1) for which data were reported for at least 80 % of the time series. For all datasets except Cedar Creek, exotic species occur in up to 68 % of sample units, but, averaged over all sample units in each dataset, they comprise < 6 % of total richness and < 5 % of plant cover/density. Exotics are more abundant in the Cedar Creek datasets. Consequently, for this dataset, we only included plots that never had > 25 % exotic species cover. Each dataset was accompanied by weather data from a nearby (<12 km) meteorological station for the period during which the vegetation data were collected.

Table 8.1 Datasets used to assess natural range of variability (NRV) in Great Plains grasslands (GPG)

Name	Location	Subset ($n =$ sample size)	Period of record	Data	Source, reference
SGS	Shortgrass Steppe LTER North-central Colorado Loamy and sandy plains Shortgrass prairie Annual precip: 348 mm Mean annual temp: 9.0 °C	Grazing exclosure $n = 12$[a]	1998–2006	Number of individuals for each species counted in 1-m^2 quadrat	Control treatment of *Bouteloua gracilis* removal experiment http://sgs.cnr.colostate.edu/dataset_view.aspx?id=BOGRRmvlDnsty (accessed 27 Sept 2011) (Munson and Lauenroth 2009)
		Moderate summer cattle grazing $n = 12$[a]	1998–2006		
Hays	Fort Hays State Univ. Central Kansas Loamy and limy rolling hills Southern mixed-grass prairie Annual precip: 577 mm Mean annual temp: 12.1 °C	Big bluestem: swales $n = 3$ or 4	1932–1968, 1970–1972	Basal cover of each species in 1-m^2 quadrat mapped using pantograph	Historic rangeland monitoring data digitized by Adler et al. (2007)
		Little bluestem: shallow limestone on hill brows and slopes $n = 8$–12	1932–1972		
		Ecotone between little bluestem and shortgrass $n = 6$ (1933–1942) $n = 14$ (1942–1972)	1933–1972		
		Shortgrass: deep soils on level uplands $n = 4$–6	1932–1972		
		Grazed shortgrass: deep soils on level uplands $n = 8$–15	1935–1972		

Table 8.1 (continued)

Name	Location	Subset (n = sample size)	Period of record	Data	Source, reference
Miles City	Fort Keogh Livestock and Range Research Lab Eastern Montana Silty and clayey ecological sites Northern mixed-grass prairie Annual precip: 305 mm Mean annual temp: 8.2 °C	*Light grazing (7.9 ha • AUM*$^{-1}$*)* $n = 7–13$	*1933–1945*	Basal cover or stem count of each perennial species mapped in 1-m^2 quadrat using pantograph; number of individuals in same 1-m^2 quadrat recorded for annual species[b]	Historic rangeland monitoring data digitized by Anderson et al. (2011)
		Moderate grazing (6.2 ha • AUM^{-1}) $n = 11–18$	1933–1945		
		Heavy grazing (4.7 ha • AUM^{-1}*)* $n = 8–11$	*1933–1945*		
Cedar Creek	Cedar Creek LTER East-central Minnesota Well-drained sandy plains Sand tallgrass prairie/oak savanna Annual precip: 804 mm Mean annual temp: 6.7 °C	Nitrogen addition (N) = 0 g/m^2 ($n = 10$)[c]	1982–2004	Aboveground biomass: above-ground biomass clipped, dried and weighed by species at peak biomass in 0.1 × 3 m strips alternating location within each 4 m × 4 m plot	Field D of Experiment 001 (long-term nitrogen deposition) http://www.lter.umn.edu/research/data/ (accessed 20 Sept 2011) (Tilman 1987)
		$N = 1.02, 2.04, 3.40, 5.44, 9.52, 17.0$, *or* $27.2\ g/m^2$ *(n = 5 each)*	*1982–2004*		
Konza FRI	Konza Prairie LTER Eastern Kansas (Flint Hills) Cherty silt loam (upper slope), silty clay loam (lower slope) Tallgrass prairie	*Upper (n = 4), mid (n = 4), or lower (n = 4) slope burned every year*	*1983–2007 (upper and lower); 1991–1992, 1997–2001, 2006 (mid)*	Percent canopy cover (modified Daubenmire cover classes) recorded by species in five 10-m^2 quadrats along 50-m transect[d]	Ungrazed watersheds of dataset PVC02 http://www.konza.ksu.edu/KNZ/pages/data/Knzdsdetail.aspx?datasetCode=PVC02 (accessed 14 July 2008) (Collins et al. 1995; Collins 2000)

Table 8.1 (continued)

Name	Location	Subset ($n=$ sample size)	Period of record	Data	Source, reference
	Annual precip: 824 mm Mean annual temp: 13.2 °C	Upper ($n=4$ or 8), mid ($n=4$ or 8), or lower ($n=4$) slope burned every 4 years	1983–2007 (upper and lower); 1991–2011, 2006 (mid)	Evenness calculated only after 1988 because cover estimation methods changed in 1989	
		Upper ($n=4$), mid ($n=4$) or lower ($n=4$) slope burned every 20 years	*1983–2007 (upper and lower); 1991–2011, 2006 (mid)*		
Konza Grazed	Konza Prairie LTER Eastern Kansas (Flint Hills) Cherty silt loam (upper slope), silty clay loam (lower slope) Tallgrass prairie Annual precip: 816 mm Mean annual temp: 13.0 °C	*Moderate summer bison grazing on upper ($n=8$), mid ($n=4$), or lower ($n=8$) slope burned every year*	*1994–2007 (upper and lower); 1997–2007 (mid)*	Percent canopy cover (modified Daubenmire cover classes) recorded by species in five 10-m^2 quadrats along 50-m transect[d]	Bison-grazed watersheds of dataset PVC02 http://www.konza.ksu.edu/KNZ/pages/data/Knzdsdetail.aspx?datasetCode=PVC02 (accessed 14 July 2008) (Hartnett et al. 1997; Towne et al. 2005)
		Moderate summer bison grazing on upper ($n=8$), mid ($n=8$), or lower ($n=8$) slope burned every 4 years	1994–2007 (all)		
		Moderate summer bison grazing on upper ($n=8$), mid ($n=8$), or lower ($n=8$) slope burned every 20 years	*1994–2007 (all)*		

In the subset column, treatments considered to most closely approximate reference conditions are in normal font; management practices that deviate from approximate reference are given in *italics*

[a] Species richness and evenness did not differ among subsets through time (repeated measures ANOVA subset and subset × year effects $p > 0.32$)

[b] Because methods used to measure abundance varied among species (and within species among years), evenness could not be calculated

[c] Combination of treatments A and I (Tilman 1987), which did not differ in species richness and evenness through time ($p > 0.10$)

[d] Data analyzed at transect level because quadrats were not labeled consistently across years

Quantifying NRV

For each dataset, we calculated richness and evenness for each quadrat (or transect for Konza datasets) in each year (including all species), then tested for significant differences in richness and evenness among community types (Hays), treatments (SGS, Miles City, Cedar Creek), or soil type × treatment combinations (Konza FRI and Konza Grazed) with repeated measures ANOVAs. These effects or their interaction with each year were significant ($p < 0.05$) for all but one dataset (SGS). Thus, each community type, treatment, or soil type × treatment was treated separately for all datasets except SGS, in which all subsequent calculations combined the two grazing treatments into a single time series. We calculated annual richness and evenness means and 95 % confidence interval of those means for each community type/treatment/soil × treatment, as well as the median and 10th, 25th, 75th, and 90th percentiles and full range of each response variable over the entire time series for a single reference treatment (where applicable; see Table 8.1). As a measure of temporal variability, we calculated the CV in richness and evenness through time for each quadrat/transect.

Since species richness or evenness fluctuations may be related to fluctuations in precipitation via individual species' response to moisture, we also quantified temporal variability of precipitation as the CV of mean annual precipitation over the period in which vegetation data were collected. Interannual fluctuations in GPG aboveground net primary production are also considerable and have been characterized much more fully than fluctuations in species richness or evenness (Webb et al. 1978, 1983; Sala et al. 1988; Smart et al. 2007). To put the variability of species richness and evenness in the context of variability of this other important GPG attribute, we calculated the CV of total density, cover, or biomass (measure varies among datasets; see Table 8.1) through time for each quadrat or transect.

Variation of Plant Species Richness and Evenness in Great Plains Grasslands

Median species richness in 1-m^2 quadrats varies from 4 to 12, and evenness varies from 0.34 to 0.68, in western GPG (SGS, Hays, and Miles City; Figs. 8.3–8.5). In the more productive and diverse eastern GPG, median species richness is 14 in just 0.3 m^2 in the sand savanna (Cedar Creek; Fig. 8.6a), and 37–54 species in five 10-m^2 quadrats in the Flint Hills of Kansas (Konza; Fig. 8.7a–c, g–i); median evenness was similar between the sites, ranging from 0.623 to 0.675 (Figs. 8.6b, 8.7d–f, j–l). The low evenness values at Hays (Fig. 8.4, *right*) compared to the other sites may be due to the way that species abundance was measured (basal cover vs. density or foliar cover) rather than an inherent difference in structure of the plant community.

Temporal variability of species richness in these datasets is fairly consistent across the wide variety of conditions occurring in shortgrass prairie in Colorado (SGS) to tallgrass sand savanna in Minnesota (Cedar Creek), with richness CVs in the

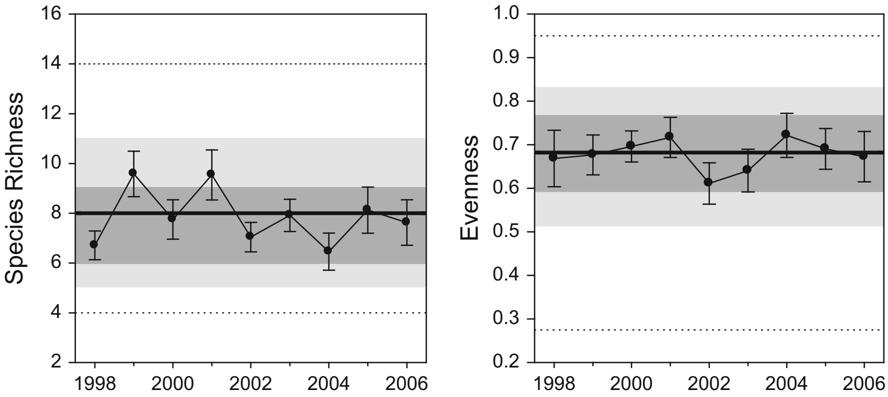


Fig. 8.3 Mean (± 95 % CI) annual species richness (*left*) and evenness (*right*) in 1-m^2 quadrats in northeastern Colorado shortgrass prairie (SGS dataset). *Bands* indicate 25–75th (*darker gray*) and 10–90th (*lighter gray*) percentile range of all values measured in indicated time period, and *dotted lines* indicate the full range of these values. The *thick horizontal line* is the long-term median of these values

0.20–0.45 range (Table 8.2). Flint Hills tallgrass prairie (Konza) shows much lower richness CVs (< 0.20), but this is likely due to the much larger area covered by each sample. The chances of one or a few species winking in or out from one year to the next are presumably much higher in a small quadrat than in the 50-m^2 area sampled for the Konza datasets. Higher species richness CVs seem to be related more to whether a time series includes the 1930s Dust Bowl period of extreme drought in this region rather than to the length of the time series. Comparing richness CVs from SGS and Cedar Creek (Dust Bowl not included) to those of Hays and Miles City (Dust Bowl included), versus comparing SGS and Miles City (< 15 years) to Cedar Creek and Hays (≥ 25 years), illustrates this point (Table 8.2). Temporal variability of evenness spans a larger range among the time series than does species richness variability, but evenness CV is not consistently greater than that of richness at a given site.

Temporal variability of species richness in these datasets is consistently lower than that of production (Table 8.2) suggesting that this may be true across a wider range of locations than those investigated here. In contrast, temporal variability of evenness relative to that of production is inconsistent within and across sites, as is the variability of richness or evenness relative to precipitation. The latter indicates that, although GPG production is generally related to temporal variations in precipitation (Webb et al. 1978, 1983; Sala et al. 1988; Smart et al. 2007), the same is not true for GPG diversity. Indeed, analyses investigating the relationship between plant species richness and a variety of weather variables in GPG suggests that there is little consistency in these relationships across sites or among management treatments within sites (Jonas et al. in revision). It is important to note, however, that temporal variability in richness and evenness is frequently of similar magnitude to differences in the metrics among management treatments (e.g., fire; Fig. 8.7d–f vs. Fig. 8.8b) or

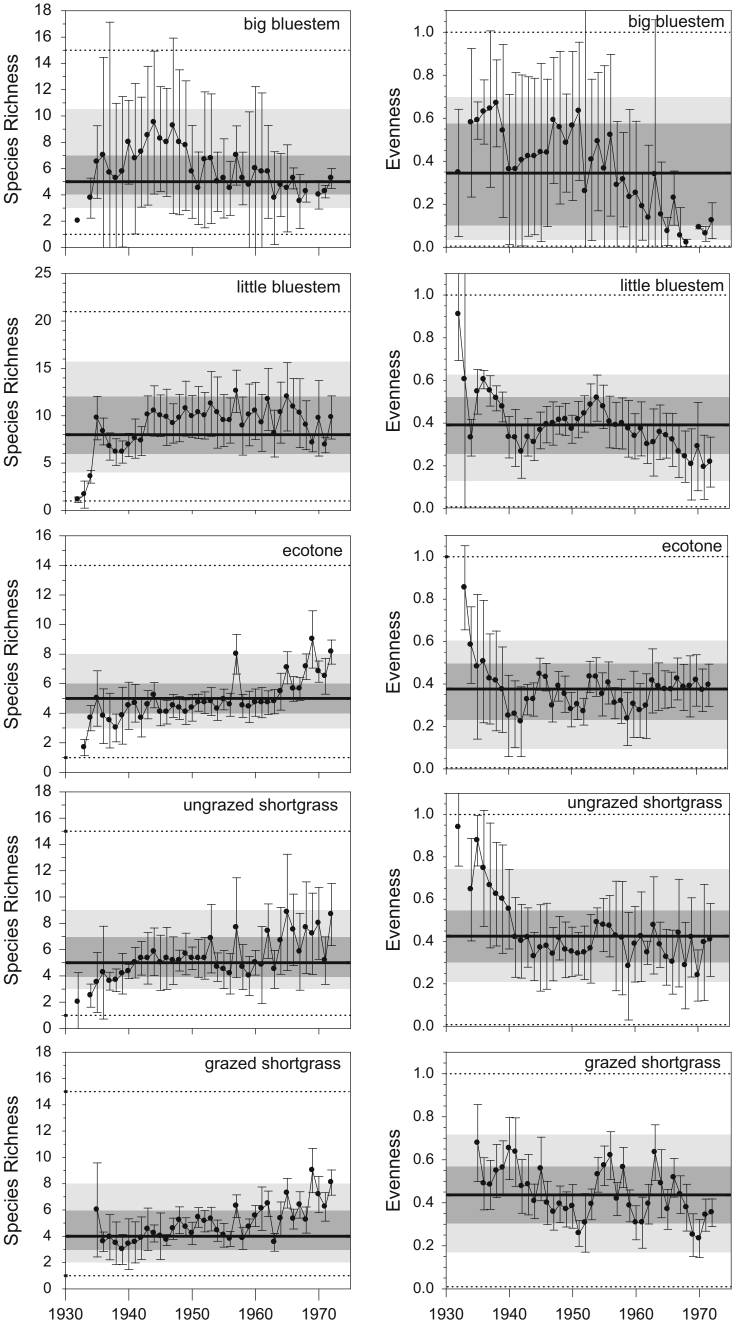


Fig. 8.4 Mean (± 95 % CI) annual species richness (*left*) and evenness (*right*) in 1-m^2 quadrats in five vegetation types in central Kansas mixed-grass prairie (Hays dataset). Shading and lines as in Fig. 8.3

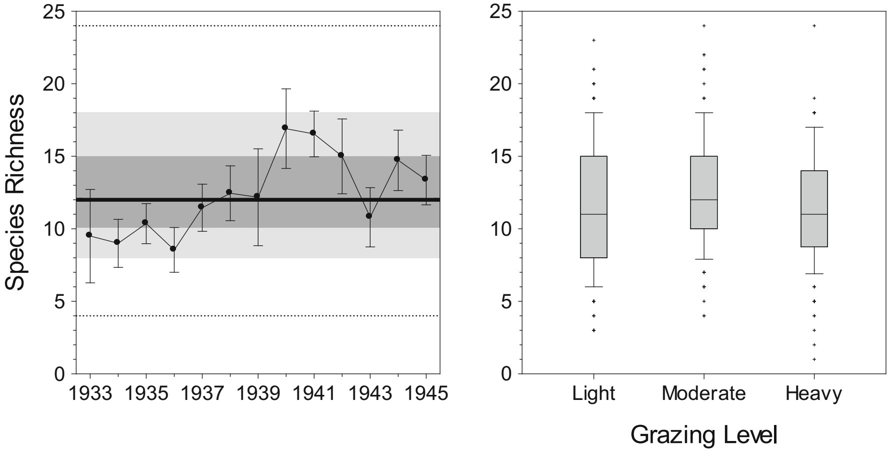


Fig. 8.5 (*Left*) Mean (± 95 % CI) annual species richness in 1-m^2 quadrats in moderately grazed mixed-grass prairie in eastern Montana (Miles City dataset). Shading and lines as in Fig. 8.3. (*Right*) Box-whisker diagrams for light, moderate, and heavy grazing treatments at the same location and over the same time period depicted at *left*. For each treatment, the median (*central line*), middle quartiles (*box*), 10th and 90th percentiles (*whiskers*), and outlying values (+ *'s*) are shown

experimental treatments simulating anthropogenic stresses (e.g., nitrogen deposition; Fig. 8.6b vs. Fig. 8.6d). This "noise" that makes detecting effects of these factors on GPG difficult has been noted previously by other authors (Gibson and Hulbert 1987; Biondini et al. 1989; Coppedge et al. 1998; Wienk et al. 2009).

Where Should the Decision Threshold Be?

Possible ranges of plant species richness and evenness that might be considered NRVs for a variety of GPG communities are illustrated in Figs. 8.3–8.7. By definition, the range spanning the 25–75th percentiles ("interquartile" range; dark gray band in Figs. 8.3–8.7) includes half of the values measured, whereas the 10–90th percentile range (light gray band) includes 80 % of them. In most cases, the latter range is substantially larger than the former, indicating substantially different deviations from the central tendency that would be needed to spur a change in management activity if the decision threshold were simply the outside bound of the NRV. For example, the declining values of species richness in fertilized plots at Cedar Creek (Fig. 8.6a) could prompt action in 1985 if the decision threshold were determined by the interquartile range, but they would not garner attention until 1988, or possibly even 1996, if the decision threshold were based on the wider 10–90th percentile range. And, of course, there would be no action at all if the decision threshold was based on the full range of values. Basing NRVs and decision thresholds on this full range is particularly problematic for Hays, the longest dataset and one that began during the Dust Bowl,

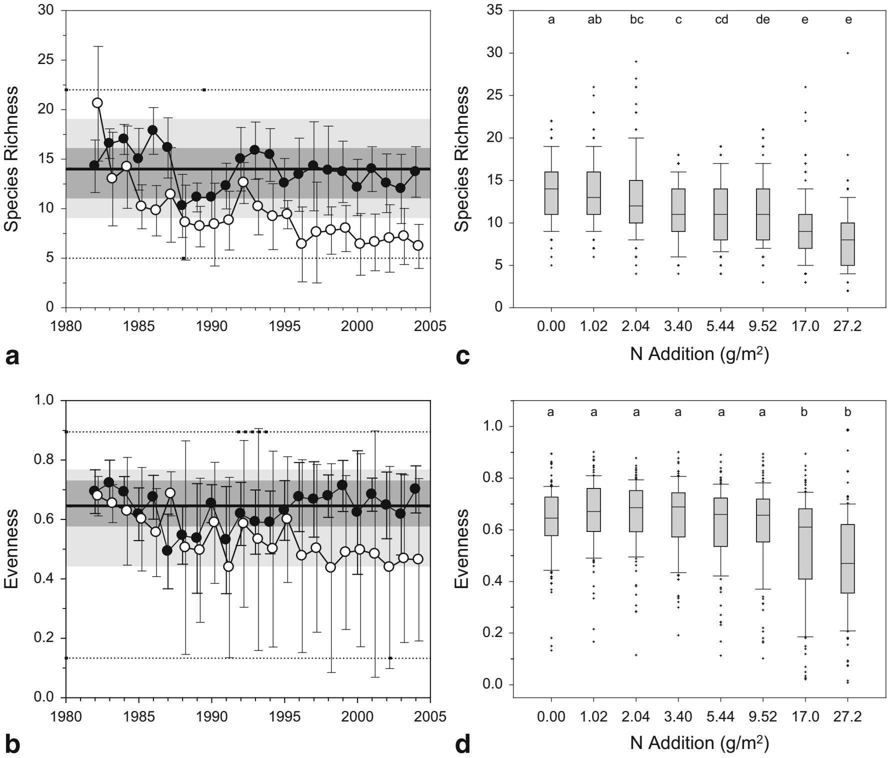


Fig. 8.6 **a**, **b** Mean (± 95 % CI) annual species richness (**a**) and evenness (**b**) in 0.3-m^2 quadrats in unfertilized (reference condition; *black circles*) and fertilized (17.0 g nitrogen/m^2, *white circles*) sand savanna in east-central Minnesota (Cedar Creek dataset). Fertilized time series are offset slightly for clarity. Shading and lines as in Fig. 8.3. **c**, **d** Box-whisker diagrams for nitrogen (N) addition treatments at the same location and over the same time period depicted in (**a**) and (**b**). For each treatment, the median (*central line*), middle quartiles (*box*), 10th and 90th percentiles (*whiskers*), and outlying values (+ *'s*) are shown. Lower-case letters above boxes indicate significant differences among treatments ($p < 0.05$)

because the range of evenness values experienced by the system essentially equals the full range of values mathematically possible for this index.

A single location's species richness value falling outside of the range chosen as the NRV under current conditions would generally not warrant the management action prescribed by the decision threshold. Decisions would be based on the mean or median of the sample frame and the confidence in that estimate. A single year's crossing out of the NRV also might not warrant immediate action. For example, if we say that the NRVs in Fig. 8.6a are independent of the time series shown therein, the drop in mean species richness of the unfertilized plant community at Cedar Creek below the interquartile range in 1988 might not spur action because richness returns to within this range the next year and stays there for the duration of the monitoring. Thus, a decision threshold might require that the metric of interest show a consistent trend of

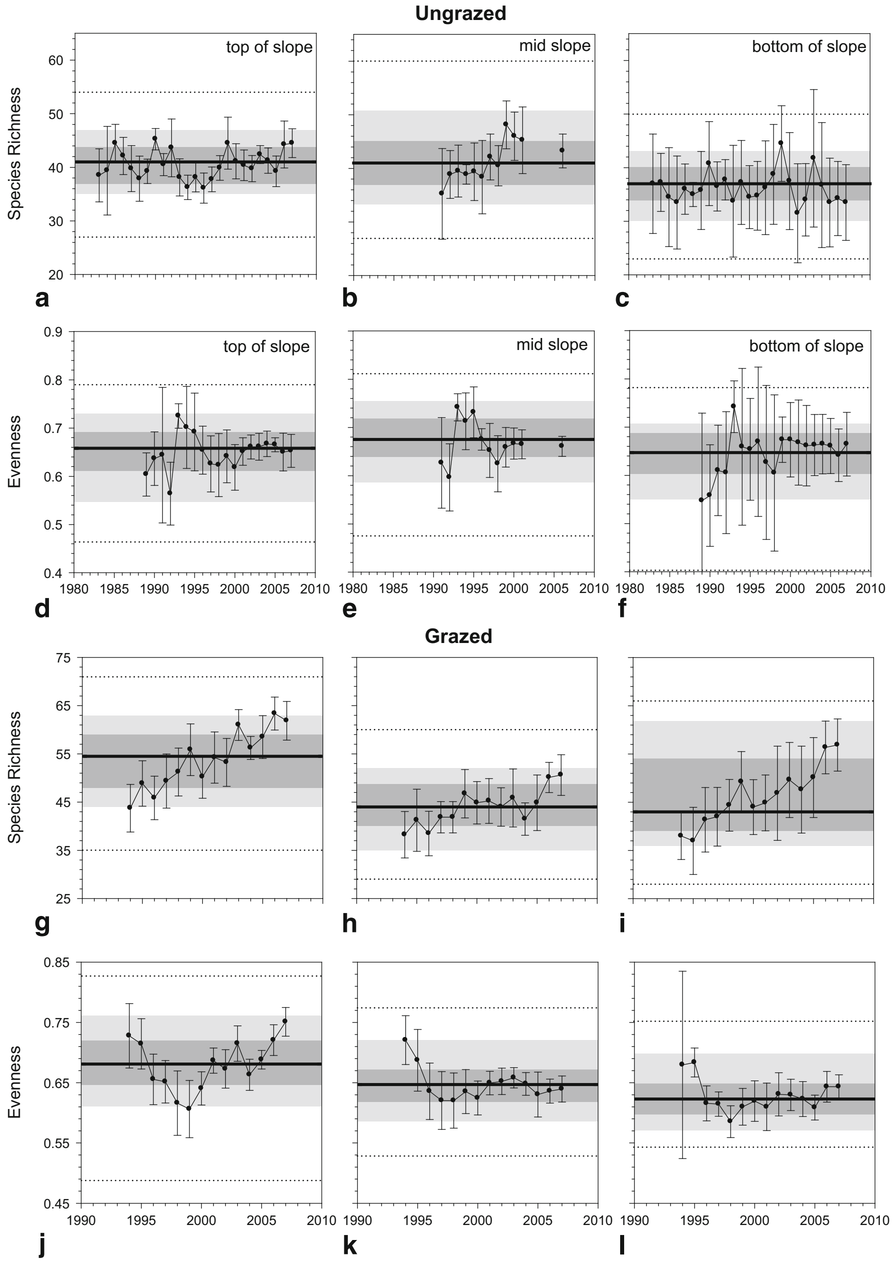


Fig. 8.7 Mean (± 95 % CI) annual species richness and evenness for five 10-m^2 quadrats in quadrennially burned, ungrazed (**a**–**f**) and grazed (**g**–**l**) tallgrass prairie in three different topographic positions/soil types in eastern Kansas (**a**–**f**, Konza FRI dataset; **g**–**l**, Konza Grazed dataset). *Shading and lines* as in Fig. 8.3

Table 8.2 Temporal variability of plant species richness, evenness, and production in reference (*plain font*) or alternative reference (*italics*) conditions for datasets described in Table 8.1

Dataset	Prairie type	Subset	CV of species richness	CV of evenness	CV of "production" (total density, cover, or biomass)	CV of precipitation
SGS	Shortgrass	Grazed and ungrazed combined	0.231 (0.112–0.395)	0.677 (0.275–0.950)	0.502 (0.237–0.900)	0.333
Hays	Mixed-grass	Big bluestem	0.358 (0.326–0.380)	0.696 (0.486–0.838)	0.477 (0.369–0.563)	0.258
		Little bluestem	0.319 (0.228–0.423)	0.493 (0.235–0.693)	0.404 (0.309–0.532)	
		Little bluestem-shortgrass ecotone	0.382 (0.219–0.536)	0.482 (0.227–0.846)	0.444 (0.327–0.580)	
		Shortgrass	0.408 (0.292–0.490)	0.469 (0.374–0.755)	0.571 (0.488–0.650)	
		Shortgrass (grazed)	0.429 (0.291–0.545)	0.448 (0.346–0.649)	0.489 (0.405–0.612)	
Miles City	Mixed-grass	Low grazing (7.9 ha • AUM^{-1})	0.360 (0.207–0.614)	Could not be calculated		0.331
		Medium grazing (6.2 ha • AUM^{-1})	0.282 (0.186–0.436)			
		High grazing (4.7 ha • AUM^{-1})	0.330 (0.239–0.528)			
Cedar Creek	Sand tallgrass		0.233 (0.195–0.278)	0.183 (0.102–0.326)	0.619 (0.405–0.882)	0.210
Konza FRI	Tallgrass	Annually burned upland	0.088 (0.074–0.104)	0.105 (0.084–0.120)	0.195 (0.139–0.235)	0.234
		Annually burned slope	0.060 (0.039–0.102)	0.181 (0.132–0.218)	0.129 (0.097–0.196)	
		Annually burned lowland	0.077 (0.069–0.088)	0.107 (0.101–0.115)	0.175 (0.144–0.255)	
		Quadrennially burned upland	0.106 (0.068–0.131)	0.080 (0.060–0.105)	0.191 (0.155–0.233)	
		Quadrennially burned slope	0.131 (0.078–0.178)	0.087 (0.036–0.113)	0.195 (0.120–0.275)	
		Quadrennially burned lowland	0.112 (0.094–0.140)	0.099 (0.063–0.150)	0.270 (0.212–0.351)	
Konza Grazed	Tallgrass	Grazed annually burned upland	0.117 (0.086–0.143)	0.073 (0.057–0.092)	0.240 (0.187–0.328)	0.219
		Grazed annually burned slope	0.084 (0.062–0.103)	0.054 (0.037–0.076)	0.232 (0.182–0.300)	
		Grazed annually burned lowland	0.199 (0.163–0.250)	0.081 (0.049–0.097)	0.223 (0.167–0.299)	
		Grazed quadrennially burned upland	0.127 (0.108–0.156)	0.076 (0.051–0.140)	0.187 (0.115–0.288)	
		Grazed quadrennially burned slope	0.110 (0.070–0.147)	0.060 (0.041–0.090)	0.214 (0.134–0.337)	
		Grazed quadrennially burned lowland	0.160 (0.071–0.221)	0.064 (0.048–0.085)	0.233 (0.155–0.291)	

Values shown are the mean (min–max) coefficient of variation (CV) for transects (Konza) or quadrats over the time described for the dataset in Table 8.1. For comparison, CV of precipitation over that time period is also given deviating from the NRV. Taking action also requires understanding the reason for the departure. In the Cedar Creek example, the precipitous drop in richness was caused by a severe drought (Tilman and El Haddi 1992), a natural environmental driver usually requiring no management intervention. On the other hand, if the departure from NRV were linked to a fertilizer spill, remedial action should be taken.

How long to wait, though, will depend on the relative costs and benefits of acting when the system can actually recover on its own versus being too slow in taking action so that the system crosses into an undesirable state from which it cannot recover. These costs and benefits are formally codified in a utility function (Martin et al. 2009; Nichols et al. Chap. 2), which also incorporates knowledge about the ecosystem's response to specific management actions with respect to management objectives. Since utility functions are unique to each management situation and are relatively rare for vegetation management in natural areas, we do not address this issue. We emphasize, however, that using management thresholds and assessment points based simply on NRV is a reasonable first step towards constructing more rigorous decision-making models in GPG.

A judicious approach in this situation would be to designate the interquartile NRV as the bounds for instigating more attention, such as more detailed analyses of data (e.g., closer look at species composition, exotic species richness and abundance, or woody plant cover) or increasing the intensity of sampling (surveillance assessment points; Mitchell et al., Chap. 10), and to designate the broader 10–90th percentile NRV as the bounds to instigate a specific-management action (action assessment point). The rationale for having these two different assessment points is to avoid the cost (monetary and ecological) of initiating an action when unnecessary, such as when the system returns to the narrower NRV prior to reaching the action assessment point.

Some Notes About the Values We Present

Designating what constitutes the "reference" condition is far from straightforward in an ecosystem that does not lend itself to most historical reconstruction techniques (e.g., dendrochronology). Although there is uncertainty about the intensity, frequency, and spatial patterning of fire and grazing before European settlement in all of these systems, we are certain that the conditions maintained at the sites from which we obtained our data substantially diverged from presettlement conditions. For example, grazing at these sites, if it occurred at all, was much more regulated (timing and intensity) than what occurred when large herds of bison, elk, and pronghorn freely roamed the Great Plains, and there are known differences in behavior and dietary preferences between these native grazers and domestic cattle (Plumb and Dodd 1993; Hartnett et al. 1997; Towne et al. 2005). Prescribed fire, if it occurred at all, was also applied quite regularly and uniformly and during conditions that do not necessarily mimic those of presettlement times (Higgins 1986). Because it is impossible to return to presettlement fire and grazing regimes and undo landscape fragmentation, atmospheric pollution (including greenhouse gas emissions), etc., the species richness and evenness values we present are reasonable alternatives.

Although we used only one "reference" condition to illustrate fluctuations in species richness and evenness through time (Figs. 8.3–8.7), we acknowledge that others may disagree with those choices by presenting summary information on temporal variability (Table 8.2) and overall variation (Figs. 8.5, 8.8) for conditions that could also be considered reference. For example, although the assumed fire return interval in presettlement tallgrass prairie is 3–5 years (Knapp et al. 1998; hence our using the 4-year fire return interval as the reference condition for Konza datasets), shrub cover has steadily increased under these conditions at Konza, whether grazed or not (Heisler et al. 2003; Briggs et al. 2005; Ratajczak et al. 2011), indicating the vegetation is not in reference condition despite the maintenance of the historical fire frequency. Indeed, shrub cover at Konza has increased slightly even under annual burning, indicating that management more aggressive than the historical fire regime may not be able to attain or maintain vegetation within the NRV in the face of changing atmospheric and herbivory conditions (Briggs et al. 2005).

We also acknowledge that some datasets probably do not adequately describe NRV for certain conditions. In particular, species richness has continuously increased since bison grazing was initiated in 1994 in the Konza datasets (Fig. 8.7g–i), indicating that the system has not yet reached equilibrium with the new conditions. Similarly, the Miles City dataset, begun in the Dust Bowl, may not be long enough to capture the full recovery of the system from this severe drought (Albertson and Weaver 1944; Adler and Levine 2007; i.e., compare Figs. 8.4 and 8.7).

The datasets we used cover only a small subset of the many grassland-vegetation assemblages that occur across the Great Plains. However, the values we derived for NRVs (interquartile and 10–90th percentile ranges) provide a starting point for managers at other sites. For example, species richness in a native prairie unit at Fort Union Trading Post National Historical Site in western North Dakota is well within the center of the interquartile range for Miles City. This position is consistent with other aspects of the vegetation at Fort Union that suggest that it is in good condition (Symstad 2011). On the other hand, the mixed-grass prairie at Scotts Bluff National Monument in western Nebraska has a substantial amount of invasive, annual brome (*Bromus* spp.) grasses, suggesting poor condition. Species richness there is well below the 10th percentile of Miles City mixed-grass prairie and near the 25th percentile for the mixed-grass communities at Hays (Symstad 2005). When determining whether the values presented here are relevant for other sites, managers would benefit from a short term but spatially extensive sampling of one or more reference sites relevant to their location. Given the paucity of long-term monitoring in GPG vegetation, however, we believe the temporal variability information we present (CVs in Table 8.2) is likely to be the best available.

Unfortunately, the natural variability of these metrics through time may make detecting trends towards undesirable changes in GPG vegetation difficult. For example, Collins et al. demonstrated that annual spring burning of tallgrass prairie in the Flint Hills of Kansas is known to lead to statistically significantly lower species richness compared to areas burned every 3–4 years (Collins et al. 1995; Collins 2000;

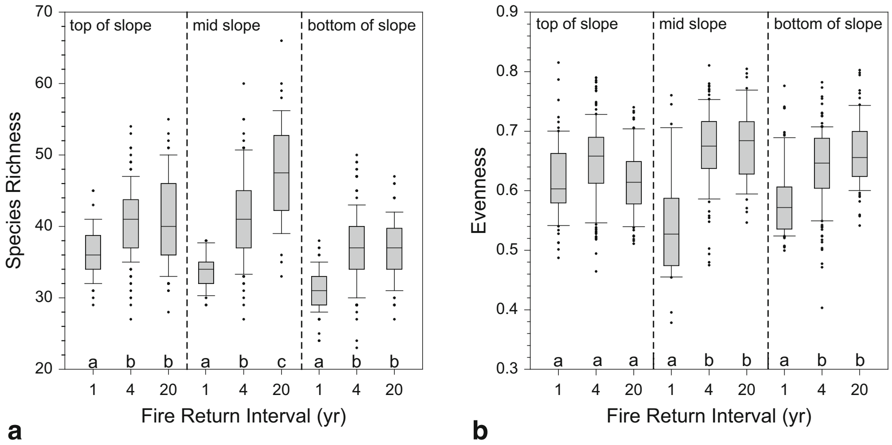


Fig. 8.8 Box-whisker plots of species richness (**a**) and evenness (**b**) values from five 10-m^2 quadrats for three fire return interval treatments in three topographic positions/soil types in ungrazed, eastern Kansas tallgrass prairie, 1983–2007 (richness) or 1989–2007 (evenness) (Konza FRI dataset). For each treatment × topographic position combination, the median (*central line*), middle quartiles (*box*), 10th and 90th percentiles (*whiskers*), and outlying values (+ *'s*) are shown. *Lower-case letters* below boxes indicate significant differences among treatments within topographic position ($p < 0.05$)

Fig. 8.8a). In the Konza FRI dataset, the difference between annual and quadrennial burning is reflected by the fact that the median richness of annually burned prairie is below the 25th percentile for the quadrennially burned (reference condition) in each of the three slope positions prairie (Fig. 8.8a). On the other hand, even fairly low levels of nitrogen fertilization have been shown to significantly reduce species richness at Cedar Creek (Tilman 1987; Clark and Tilman 2008; Fig. 8.6c), but the median value of richness in significantly impacted plots (2.04 and 3.40 g N/m^2 added per year) is well within the interquartile range of richness for the unfertilized plots (Fig. 8.6c). In this case, the experimental evidence supports setting surveillance-assessment points within a narrower range than the interquartile based on temporal variability, but action assessment points/decision thresholds would need to be tempered by the fact that a severe drought (in 1987–1988) drove species richness and evenness well outside the interquartile range (Fig. 8.6a, b).

Conclusions

In ecosystems exhibiting a wide range of natural variability, identifying ecological or decision thresholds can be challenging. Employing the NRV concept to describe attributes of interest based on long-term information for an ecosystem is a promising avenue for determining surveillance and action-assessment points or decision thresholds. This case study presents actual values of plant species richness and evenness and their NRVthat managers of GPG vegetation can use as starting points for establishing thresholds for heightened attention or taking action in their own locations. In addition, it illustrates the limitations of using just NRV information to establish these thresholds. Continued attention to the importance of plant diversity in GPG and other ecosystems, long-term monitoring of plant diversity metrics being established by the NPS and other agencies and organizations, research to better understand the response of these metrics to various stressors and management actions, and concerted efforts by managers to determine the costs and benefits of acting at various levels of diversity will help refine decision thresholds for this vast but highly threatened ecosystem.

Acknowledgments At the time of the research for this manuscript Jayne Jonas was with IAP World Service, Inc. This project would not have been possible without access to the datasets provided by the Cedar Creek, Konza Prairie, and Shortgrass Steppe LTER Programs, the Kansas Agricultural Experiment Station—College Pasture at Hays, Kansas, and the Fort Keogh Livestock and Range Research Lab, Miles City, Montana. In particular, we thank Troy Mielke (Cedar Creek), Gene Towne (Konza), Seth Munson (SGS), and Peter Adler (Hays and Miles City) for assistance with information on the datasets. We also thank Greg Eckert, Diane Larson, and Glenn Guntenspergen for their helpful comments on an earlier version of this manuscript. Funding for this project was provided by the U.S. Geological Survey's Park Monitoring Project. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Adler, P. B., and J. M. Levine. 2007. Contrasting relationships between precipitation and species richness in space and time. *Oikos* 116:221–232.

Adler, P. B., W. R. Tyburczy, and W. K. Lauenroth. 2007. Long-term mapped quadrats from Kansas prairie: Demographic information for herbaceous plants. *Ecology* 88:2673.

Albertson, F. W., and J. E. Weaver. 1944. Nature and recovery of grassland from the Great Drought of 1933 to 1940. *Ecological Monographs* 14:393–479.

Anderson, R. C. 2006. Evolution and origin of the Central Grassland of North America: Climate, fire, and mammalian grazers. *Journal of the Torrey Botanical Society* 133:626–657.

Anderson, J., L. Vermiere, and P. B. Adler. 2011. Fourteen years of mapped, permanent quadrats in a northern mixed prairie, USA. *Ecology* 92:1703.

Axelrod, D. I. 1985. Rise of the grassland biome, central North America. *Botanical Review* 51: 163–201.

Bagchi, S., D. D. Briske, X. B. Wu, M. P. McClaran, B. T. Bestelmeyer, and M. E. Fernandez-Gimenez. 2012. Empirical assessment of state-and-transition models with a long-term vegetation record from the Sonoran Desert. *Ecological Applications 22*:400-411.

Baker, W. L. 1992. Effects of settlement and fire suppression on landscape structure. *Ecology* 73:1879–1887.

Balvanera, P., A. B. Pfisterer, N. Buchmann, J.S. He, T. Nakashizuka, D. Raffaelli, and B. Schmid. 2006. Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecology Letters* 9:1146–1156.

Bennetts, R. E., J. E. Gross, K. Cahill, C. McIntyre, B. B. Bingham, A. Hubbard, L. Cameron, and S. L. Carter. 2007. Linking monitoring to management and planning: Assessment points as a generalized approach. *The George Wright Forum* 24:59–77.

Bergeron, Y., S. Gauthier, M. Flannigan, and V. Kafka. 2004. Fire regimes at the transition between mixedwood and coniferous boreal forest in northwestern Quebec. *Ecology* 85:1916–1932.

Bestelmeyer, B. T., J. R. Brown, K. M. Havstad, R. Alexander, G. Chavez, and J. E. Herrick. 2003. Development and use of state-and-transition models for rangelands. *Journal of Range Management* 56:114–126.

Bestelmeyer, B. T., A. J. Tugel, G. L. Peacock, Jr., D. G. Robinett, P. L. Shaver, J. R. Brown, J. E. Herrick, H. Sanchez, and K. M. Havstad. 2009. State-and-transition models for heterogeneous landscapes: A strategy for development and application. *Rangeland Ecology & Management* 62:1–15.

Biondini, M. E., A. A. Steuter, and C. E. Grygiel. 1989. Seasonal fire effects on the diversity patterns, spatial distribution and community structure of forbs in the northern Mixed Prairie, USA. *Vegetatio* 85:21–32.

Borchert, J. R. 1950. The climate of the central North American grassland. *Annals of the Association of American Geographers* 40:1–39.

Briggs, J. M., A. A. K. Knapp, J. M. Blair, J. L. Heisler, G. A. Hoch, M. S. Lett, and J. K. McCarron. 2005. An ecosystem in transition: Causes and consequences of the conversion of mesic grassland to shrubland. *Bioscience* 55:243–254.

Briske, D. D., S. D. Fuhlendorf, and F. E. Smeins. 2006. A unified framework for assessment and application of ecological thresholds. *Rangeland Ecology and Management* 59:225–236.

Cardinale, B. J., D. S. Srivastava, J. E. Duffy, J. P. Wright, A. Downing, M. Sankaran, and C. Jouseau. 2006. Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* 443:989–992.

Cardinale, B. J., J. P. Wright, M. W. Cadotte, I. T. Carroll, A. Hector, D. S. Srivastava, M. Loreau, and J. J. Weis. 2007. Impacts of plant diversity on biomass production increase through time because of species complementarity. *Proceedings of the National Academy of Sciences U S A* 104:18123–18128.

Carlson, M., and W. A. Kurz. 2007. Approximating natural landscape pattern using aggregated harvest. *Canadian Journal of Forest Research* 37:1846–1853.

Clark, C. M., and D. Tilman. 2008. Loss of plant species after chronic low-level nitrogen deposition to prairie grasslands. *Nature* 451:712–715.

Collins, S. L. 2000. Disturbance frequency and community stability in native tallgrass prairie. *The American Naturalist* 155:311–325.

Collins, S. L., S. M. Glenn, and D. J. Gibson. 1995. Experimental analysis of intermediate disturbance and initial floristic composition: Decoupling cause and effect. *Ecology* 76:486–492.

Coppedge, B. R., D. M. Engle, C. S. Toepfer, and J. H. Shaw. 1998. Effects of seasonal fire, bison grazing and climatic variation on tallgrass prairie vegetation. *Plant Ecology* 139:235–246.

Cyr, D., S. Gauthier, Y. Bergeron, and C. Carcaillet. 2009. Forest management is driving the eastern North American boreal forest outside its natural range of variability. *Frontiers in Ecology and the Environment* 7:519–524.

DeBacker, M. D., A. N. Sasseen, C. Becker, G. A. Rowell, L. P. Thomas, J. R. Boetsch, and G. D. Willson. 2004. *Vegetation community monitoring protocol for the Heartland I & M Network and Prairie Cluster Prototype Monitoring Program*. Republic: National Park Service.

Doyon, F., S. Yamasaki, and R. Duchesneau. 2008. The use of the natural range of variability for identifying biodiversity values at risk when implementing a forest management strategy. *Forestry Chronicle* 84:316–329.

Duncan, S. L., B. C. McComb, and K. N. Johnson. 2010. Integrating ecological and social ranges of variability in conservation of biodiversity: Past, present, and future. *Ecology and Society* 15(1):5.

Egan, D., and E. A. Howell, editors. 2001a. *The historical ecology handbook: A restorationist's guide to reference ecosystems*. Washington, DC: Island Press.

Egan, D., and E. A. Howell. 2001b. Introduction. In *The historical ecology handbook: A restorationist's guide to reference ecosystems,* ed. D. Egan, and E. A. Howell, 1–23. Washington, DC: Island Press.

Ellison, A. M. 2001. Exploratory data analysis and graphic display. In *Design and analysis of ecological experiments,* ed. S. M. Scheiner, and J. Gurevitch, 37–62. New York: Oxford University Press.

Fargione, J., D. Tilman, R. Dybzinski, J. H. R. Lambers, C. Clark, W. S. Harpole, J. M. H. Knops, P. B. Reich, and M. Loreau. 2007. From selection to complementarity: Shifts in the causes of biodiversity-productivity relationships in a long-term biodiversity experiment. *Proceedings of the Royal Society B-Biological Sciences* 274:871–876.

Ford, E. S., A. E. Kelly, S. M. Teutsch, S. B. Thacker, and P. L. Garbe. 1999. Radon and lung cancer: A cost-effectiveness analysis. *American Journal of Public Health* 89:351–357.

Gibson, D. J., and L. C. Hulbert. 1987. Effects of fire, topography and year-to-year climatic variation on species composition in tallgrass prairie. *Vegetatio* 72:175–185.

Graves, H. S. 1899. The Black Hills Forest Reserve. In *The nineteenth annual report of the survey, 1897-1898. Part V. Forest Reserves,* 67-164. Washington, DC: U.S. Geological Survey.

Groffman, P. M., J. S. Baron, T. Blett, A. J. Gold, I. Goodman, L. H. Gunderson, B. M. Levinson, M. A. Palmer, H. W. Paerl, G. D. Peterson, N. L. Poff, D. W. Rejeski, J. F. Reynolds, M. G. Turner, K. C. Weathers, and J. Wiens. 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? *Ecosystems* 9:1–13.

Hartnett, D. C., K. R. Hickman, and L. E. Fischer Walter. 1996. Effects of bison grazing, fire, and topography on floristic diversity in tallgrass prairie. *Journal of Range Management* 49:412–420.

Hartnett, D. C., A. A. Steuter, and K. R. Hickman. 1997. Comparative ecology of native and introduced ungulates. *Ecological Studies* 125:72–101.

Harwell, M. A. 1997. Ecosystem management of south Florida. *Bioscience* 47:499–512.

Hastings, A., and D. B. Wysham. 2010. Regime shifts in ecological systems can occur with no warning. *Ecology Letters* 13:464–472.

Hector, A., and R. Bagchi. 2007. Biodiversity and ecosystem multifunctionality. *Nature* 448: 188–191.

Heisler, J. L., J. M. Briggs, and A. K. Knapp. 2003. Long-term patterns of shrub expansion in a C4-dominated grassland: Fire frequency and the dynamics of shrub cover and abundance. *American Journal of Botany* 90:423–428.

Higgins, K. F. 1986. *Interpretation and compendium of historical fire accounts in the northern Great Plains. Resource Publication, 161*. Washington, DC: U. S. Fish and Wildlife Service.

Hooper, D. U., F. S. Chapin, III, J. J. Ewel, A. Hector, P. Inchausti, S. Lavorel, J. H. Lawton, D. M. Lodge, M. Loreau, S. Naeem, B. Schmid, H. Setälä, A. J. Symstad, J. Vandermeer, and D. A. Wardle. 2005. Effects of biodiversity on ecosystem functioning: A consensus of current knowledge. *Ecological Monographs* 75:3–35.

Isbell, F. I., and B. J. Wilsey. 2011. Increasing native, but not exotic, biodiversity increases above ground productivity in ungrazed and intensely grazed grasslands. *Oecologia* 165:771–781.

Knapp, A. K., J. M. Briggs, D. C. Hartnett, and S. L. Collins, eds. 1998. *Grassland dynamics: Long-term ecological research in tallgrass prairie*. New York: Oxford University Press.

Landres, P. B., P. Morgan, and F. J. Swanson. 1999. Overview of the use of natural variability concepts in managing ecological systems. *Ecological Applications* 9:1179–1188.

Lauenroth, W. K., I. C. Burke, and M. P. Gutmann. 1999. The structure and function of ecosystems in the central North American grassland region. *Great Plains Research* 9:223–259.

Magurran, A. E. 1988. *Ecological diversity and its measurement*. Princeton: Princeton University Press.

Manier, D., D. Shorrock, E. W. Schwieger, I. W. Ashton, B. Frakes, M. Britten, D. Pillmore, and J. Burke. 2011. *Rocky Mountain Network vegetation composition structure and soils monitoring protocol: Small park grasslands, shrublands, and woodlands*, Version 1.0. Natural Resource Report NPS/ROMN/NRR-2011/383. Fort Collins, CO: National Park Service.

Martin, J., M. C. Runge, J. D. Nichols, B. C. Lubow, and W. L. Kendall. 2009. Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications* 19:1079–1090.

Morgan, P., G. H. Aplet, J. B. Haufler, H. C. Humphries, M. M. Moore, and W. D. Wilson. 1994. Historic range of variability: A useful tool for evaluating ecosystem change. *Journal of Sustainable Forestry* 2:87–111.

Mori, A. S., and K. P. Lertzman. 2011. Historic variability in fire-generated landscape heterogeneity of subalpine forests in the Canadian Rockies. *Journal of Vegetation Science* 22:45–58.

Munson, S. M., and W. K. Lauenroth. 2009. Plant population and community responses to removal of dominant species in the shortgrass steppe. *Journal of Vegetation Science* 20:224–232.

National Park Service. 2009. *Interim technical guidance on defining meaningful desired conditions for natural resources*. Fort Collins: National Park Service Biological Resource Management Division.

Plumb, G. E., and J. L. Dodd. 1993. Foraging ecology of bison and cattle on a mixed prairie: Implications for natural area management. *Ecological Applications* 3:631–643.

Ratajczak, Z., J. B. Nippert, J. C. Hartman, and T. W. Ocheltree. 2011. Positive feedbacks amplify rates of woody encroachment in mesic tallgrass prairie. *Ecosphere* 2:121.

Risser, P. G., E. C. Birney, H. D. Blocker, S. W. May, W. J. Parton, and J. A. Wiens. 1981. *The true prairie ecosystem*. Stroudsburg: Hutchinson Ross Publishing Company.

Sala, O. E., W. J. Parton, L. A. Joyce, and W. K. Lauenroth. 1988. Primary production of the central grassland region of the United States. *Ecology* 69:40–45.

Samson, F. B., and F. L. Knopf. 1994. Prairie conservation in North America. *Bioscience* 44: 418–421.

Scheffer, M., and S. R. Carpenter. 2003. Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends in Ecology & Evolution* 18:648–656.

Schmid, B., P. Balvanera, B. J. Cardinale, J. Godbold, A. B. Pfisterer, D. Raffaelli, M. Solan, and D. S. Srivastava. 2009. Consequences of species loss for ecosystem functioning: meta-analyses of data from biodiversity experiments. In *Biodiversity, ecosystem functioning, and human wellbeing,* ed. S. Naeem, D. E. Bunker, A. Hector, M. Loreau, and C. Perrings, 14–29. Oxford: Oxford University Press.

Sherriff, R. L., E. E. Berg, and A. E. Miller. 2011. Climate variability and spruce beetle (*Dendroctonus rufipennis*) outbreaks in south-central and southwest Alaska. *Ecology* 92:1459–1470.

Sierra, C. A., M. E. Harmon, F. H. Moreno, S. A. Orrego, and I. J. Del Valle. 2007. Spatial and temporal variability of net ecosystem production in a tropical forest: Testing the hypothesis of a significant carbon sink. *Global Change Biology* 13:838–853.

Smart, A. J., B. H. Dunn, P. S. Johnson, L. Xu, and R. N. Gates. 2007. Using weather data to explain herbage yield on three Great Plains plant communities. *Rangeland Ecology & Management* 60:146–153.

Stafford Smith, M. 1996. Management of rangelands: Paradigms at their limits. In *The ecology and management of grazing systems,* ed. J. Hodgson and A. W. Illius, 325-357. Wallingford: CAB International.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. *Ecological Applications* 16:1267–1276.

Swanson, F. J., J. A. Jones, D. O. Wallin, and J. H. Cissel. 1994. Natural variability—implications for ecosystem management. In *Volume II: Ecosystem management: principles and applications. Gen. Tech. Rep. PNW-GTR-318*. ed. M. E. Jensen and P. S. Bourgeron, 80–94. Portland: U.S. Department of Agriculture, Forest Service.

Symstad, A. J. 2005. A pilot study for assessing the potential plant community impacts of reintroducing grazing to Scotts Bluff National Monument: Results and implications for further investigation. Keystone: USGS Northern Prairie Wildlife Research Center, Black Hills Station.

Symstad, A. J. 2011. A vegetation management plan for Fort Union Trading Post National Historic Site: Final report for interagency agreement number F154910005, Natural Resource Report NPS/FOUS/NRR–2011/456. Fort Collins: National Park Service.

Symstad, A. J., and J. L. Jonas. 2011. Incorporating biodiversity into rangeland health: Plant species richness and diversity in Great Plains grasslands. *Rangeland Ecology & Management* 64:555–572.

Symstad, A. J., R. A. Gitzen, C. L. Wienk, M. R. Bynum, D. J. Swanson, A. D. Thorstenson, and K. J. Paintner. 2011. Plant community composition and structure monitoring protocol for the Northern Great Plains I & M Network: Version 1.00, Natural Resource Report NPS/NRPC/NRR–2011/291. Fort Collins: National Park Service.

Teeri, J. A., and L. G. Stowe. 1976. Climatic patterns and the distribution of C^4 grasses in North America. *Oecologia* 23:1–12.

Thompson, J. R., S. L. Duncan, and K. N. Johnson. 2009. Is there potential for the historical range of variability to guide conservation given the social range of variability? *Ecology and Society* 14(1):18.

Thrush, S. F., J. E. Hewitt, P. K. Dayton, G. Coco, A. M. Lohrer, A. Norkko, J. Norkko, and M. Chiantore. 2009. Forecasting the limits of resilience: Integrating empirical research with theory. *Proceedings of the Royal Society B-Biological Sciences* 276:3209–3217.

Tilman, D. 1987. Secondary succession and the pattern of plant dominance along experimental nitrogen gradients. *Ecological Monographs* 57:189–214.

Tilman, D. 2001. Effects of diversity and composition on grassland stability and productivity. In *Ecology: Achievement and Challenge*, ed. M. C. Press, N. J. Huntly, and S. Levin, 183–207. Cambridge: Cambridge University Press.

Tilman, D., and A. El Haddi. 1992. Drought and biodiversity in grasslands. *Oecologia* 89:257–264.

Towne, E. G., D. C. Hartnett, and R. C. Cochran. 2005. Vegetation trends in tallgrass prairie from bison and cattle grazing. *Ecological Applications* 15:1550–1559.

Unnasch, R. S., D. P. Braun, P. J. Comer, and G. E. Eckert. 2009. The ecological integrity assessment framework: A framework for assessing the ecological integrity of biological and ecological resources of the National Park System. Fort Collins: National Park Service.

Webb, W. L., S. R. Szarek, W. K. Lauenroth, R. S. Kinerson, and M. Smith. 1978. Primary productivity and water use on native forest, grassland, and desert ecosystems. *Ecology* 59: 1239–1247.

Webb, W. L., W. K. Lauenroth, S. R. Szarek, and R. S. Kinerson. 1983. Primary production and abiotic controls in forests, grasslands, and desert ecosystems in the United States. *Ecology* 64:134–151.

Wienk, C., A. Thorstenson, J. Freeman, and D. Swanson. 2009. Northern Great Plains Fire Ecology Program Review 1997–2007. Internal report, U.S. Department of Interior, National Park Service, Midwest Regional Office, Fire & Aviation Management, Omaha, Nebraska.

Wohl, E. 2011. What should these rivers look like? Historical range of variability and human impacts in the Colorado Front Range, USA. *Earth Surface Processes and Landforms* 36:1378–1390.

Zavaleta, E. S., J. R. Pasari, K. B. Hulvey, and G. D. Tilman. 2010. Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. *Proceedings of the National Academy of Sciences U S A* 107:1443–1446.

Chapter 9
Evaluating Bioassessment Designs and Decision Thresholds Using Simulation Techniques

Craig D. Snyder, Nathaniel P. Hitt, David R. Smith and Jonathan P. Daily

Abstract Natural resource managers face numerous choices when developing bioassessment programs but seldom have the opportunity to compare the performance of alternative designs. As a result, managers often lack a basis for establishing decision thresholds based on their objectives for evaluating resource condition, accounting for uncertainty, and controlling costs. In this chapter, we illustrate how simulation techniques may be used to optimize bioassessment decision thresholds and sampling designs with a case study of benthic macroinvertebrate communities in Shenandoah National Park, USA. We evaluated the effects of sampling effort (6 levels) and taxonomic resolution (family vs. genus) on the sensitivity of a commonly used index of stream condition (Macroinvertebrate Biotic Integrity Index, MBII) to classify resource condition as affected by ecological change. We computed expected utility values to compare decision thresholds, which integrated statistical power and differential risk tolerance for misclassification (i.e., type I and II error rates). Our analysis revealed important differences among bioassessment designs. MBII sensitivity increased with sampling effort, but improvements were modest across the highest sampling levels. Genus-level assessments were generally most sensitive to ecological change, even though precision increased at the family level due to decreased variation in reference communities. However, the sensitivity-cost relationship revealed no single, optimal combination of taxonomic resolution and sampling effort. Rather, we found that for a given cost, equivalent sensitivities could be obtained from larger samples at the family-level or smaller samples at the genus level. An analysis of expected utility demonstrated that the optimal decision threshold depends on prior probability of resource condition, i.e., reference, early warning, or impaired. We conclude that simulation methods provide a flexible approach to evaluate and optimize bioassessment designs and decision thresholds based on objective-specific utility values.

C. D. Snyder (✉) · N. P. Hitt · D. R. Smith · J. P. Daily
Leetown Science Center, Aquatic Ecology Branch, U.S. Geological Survey,
11649 Leetown Road, Kearneysville, WV 25430, USA
e-mail: csnyder@usgs.gov

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_9,

Keywords Bioassessment design · Metric sensitivity-cost tradeoffs · Simulation · Benthic macroinvertebrates · Sampling effort · Taxonomic resolution · Decision thresholds

Introduction

Over the last 30 years, biological assessments of environmental quality (i.e., bioassessments) have been used widely to assess ecological status and trends in freshwater ecosystems. Although worldwide adoption of bioassessments indicates their value for evaluating resource condition, their application varies considerably within and among regions. Natural resource managers require a means to compare alternative bioassessment designs given their objective of evaluating resource condition while incorporating uncertainty and controlling cost. However, such comparisons are rare, in part due to available datasets coming from different regions at different times using different sampling methods. In this chapter, we demonstrate how simulation methods may solve this problem by enabling quantitative comparisons and optimization of bioassessment designs.

Typically, bioassessment programs sample organismal diversity and abundance to calculate indices across a gradient of reference and impaired conditions (Fig. 9.1). Such dose-response patterns (i.e., change in resource across stress gradient) define the ecological change that is of primary concern by managers. The ecological change may exhibit ecological thresholds as zones of abrupt nonlinear change (Andersen et al. 2009) or wedge-shaped patterns due to multiple, unmeasured limiting factors (Cade and Noon 2003). In either scenario, bioassessment sensitivity will be influenced by the magnitude of ecological change (i.e., responsiveness) and the variability within reference conditions (i.e., precision; Fig. 9.1).

When resource condition is discrete, such as for reference versus impaired conditions, then changes in condition can define *decision thresholds* (Smith et al., this volume). Decision thresholds provide boundaries for managers to recognize ecological change with well-understood probabilities of falsely detecting change (type I errors) or failing to detect change (type II errors; Fig. 9.1). Moreover, managers may have an objective to maintain a certain resource condition (i.e., reference condition) and have different tolerance for misclassifying condition depending on stress level (Fig. 9.1). *Utility values* reflect these different objectives and tolerances (Keeney 1992) and provide a common currency to compare alternative bioassessment designs (i.e., expected utility; Gregory and Keeney 2002).

Bioassessment could be incorporated into adaptive management in cases where there are specified management actions and models for predicting effects of actions on resource condition (see Martin et al. 2009). However, most managers use bioassessment as a form of surveillance monitoring without management actions specified a priori that could affect resource condition directly (Nichols and Williams 2006; Smith et al., Chap. 6). Decision thresholds, ecological change, and utility values are the unifying concepts discussed by Martin et al. (2009) and reviewed in

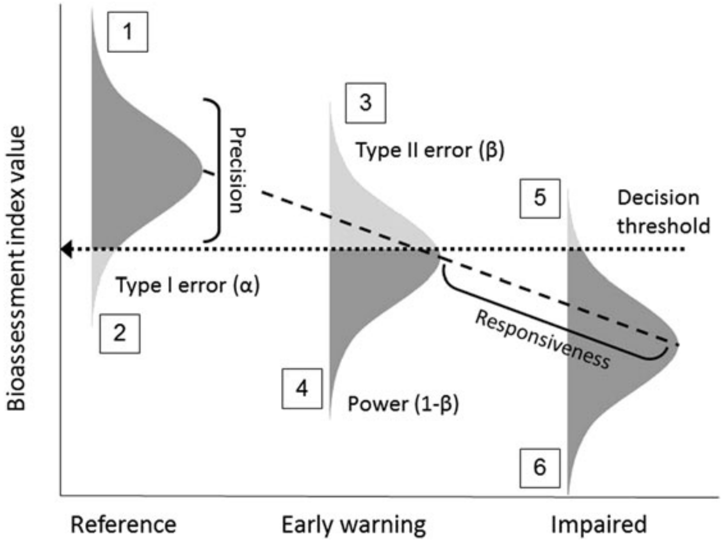


Fig. 9.1 Conceptual model of decision thresholds and error probabilities in biological monitoring. Precision indicates the variation within reference conditions and responsiveness indicates the change in central tendency from reference to impaired states. Numbers correspond to utility value categories in Fig. 9.7

Smith et al. (this volume). Utility values reflect the management objectives. Models of ecological change link resource condition to stressors and management actions. Decision thresholds can be determined by optimizing expected utility value. In surveillance monitoring, a sensitivity or power analysis is a special case of maximizing expected utility value where full value is assigned to detecting a departure from desired resource condition and no value is assigned to correctly classifying reference condition or to committing type I or II errors.

Benthic macroinvertebrates, the benthos-dwelling freshwater invertebrates visible to the naked eye, are commonly used in bioassessment due to their well-known sensitivities to environmental stressors (Lenat 1993; Klemm et al. 2002; Yuan 2004), established sampling protocols (Hauer and Resh 2006), and ubiquity in lotic and lentic habitats (Merritt and Cummins 1996). Moreover, benthic macroinvertebrates have revealed environmental gradients not detected by physiochemical data, presumably because biota may respond to interactive effects of multiple physical and chemical stressors simultaneously (Yoder and Rankin 1998). As such, benthic macroinvertebrates are commonly used to assess environmental quality worldwide (Capitulo et al. 2001; Hering et al. 2004; Mazor et al. 2006; Ollis et al. 2006; Morse et al. 2007).

Despite the widespread and growing use of benthic macroinvertebrates in bioassessment, different methods are often employed without an understanding of how methodological decisions affect bioassessment performance and utility values. In this chapter, we focus on two important methodological issues: sampling effort and taxonomic resolution. In this context, "sampling effort" refers to the number of individual organisms used to characterize a site and "taxonomic resolution" is determined by laboratory method, usually between classifying individuals at the family or genus level. Both of these factors may have important consequences for bioassessment designs, but their interactive effects and trade-offs remain poorly understood.

Considerations of sampling effort often involve decisions about how to subsample collected organisms from a site. Such methods are referred to as "fixed-count" approaches because organisms are included in the final sample for a site based on a target number of individuals, usually between 100 and 300. In contrast, "fixed-area" approaches do not employ a subsampling protocol, but instead quantify all individuals sampled from a given area of stream substrate. Because fixed-count approaches will typically require taxonomic identifications for fewer individuals, these methods have been recommended for rapid assessments of biological condition (Plafkin et al. 1989; Growns et al. 1997; Burton and Gerritsen 2003).

Fixed-count methods have been criticized for their sensitivity to species richness which may bias bioassessment inferences (Courtemanch 1996; Vinson and Hawkins 1996; Nichols et al. 2006). Some research suggests a fixed-count threshold of 300 individuals, above which relatively little additional information is gained (Ostermiller and Hawkins 2004; Cao and Hawkins 2005; Van Sickle et al. 2007). In practice, stream bioassessment programs employ many different levels of subsampling: 100 individuals (e.g., Burton and Gerritsen 2003), 200 individuals (e.g., Clarke et al. 2002; Mazor et al. 2006), 300 individuals (e.g., Van Sickle et al. 2005), 400 individuals (e.g., Marshall et al. 2006), and 700 individuals (e.g., Clarke et al. 2006). Other bioassessment designs use a fixed-area approach to avoid subsampling altogether (e.g., USGS National Water Quality Assessment Program, Moulton et al. 2002; USEPA Environmental Monitoring and Assessment Program, Lazorchak et al. 1998).

Natural resource managers must also determine the appropriate level of taxonomic resolution for community samples. This is typically a choice between the "lowest practical" level of taxonomic resolution (usually genus or species) and the family level, representing an important trade-off between statistical power and cost (Sovell and Vondracek 1999). Some genera and species within the same family clearly show different responses to environmental degradation (Lenat and Resh 2001), and some bioassessment studies are more sensitive at the genus level than family level (Hawkins et al. 2000; King and Richardson 2002; Pond et al. 2008).

In contrast, other studies found that relatively little information is lost using family-level taxonomy (Bailey et al. 2001; Waite et al. 2004; Marshall et al. 2006) and thus lowest-level taxonomy would probably not be worth the increased analytical costs. This may be due to an increased misidentification rate for genus- and species-level taxonomy (New 1996; Marchant 2007) or the reduction of variation in reference conditions yielding increased precision to detect non-reference sites (see Hawkins et al. 2000). Nonetheless, the increased costs of lowest-level taxonomy may be justified if inferences about environmental quality are obscured at the family level (Lenat and Resh 2001).

Simple comparisons of bioassessment studies are often confounded by differences in sampling time or location (e.g., seasons and ecoregions, respectively). Moreover, in observational field studies, ecological changes across stress levels can be confounded by unknown structural variation and sampling error. As a result, few tools are currently available for managers to evaluate alternative bioassessment design strategies.

Simulation methods provide a potential solution to this problem. By simulating community change along a stress gradient, the "true" impairment is known so the relative performance of alternative bioassessment strategies may be quantified and separated from effects due to sampling error. Prior studies have demonstrated the utility of simulations in this regard (Field et al. 2004; Cao and Hawkins 2005), but did not apply these tools to optimize power-cost relations of alternative monitoring designs.

In this chapter, we used modeling and simulation techniques to evaluate bioassessment designs based on objectives for statistical power, risk tolerance, and financial costs. We demonstrated how such utility values affect optimal decision thresholds for defining "impaired" sites and provided recommendations for maximizing expected utility. Although our analysis focused on benthic macroinvertebrates in streams, this approach may be applied to terrestrial or marine ecosystems.

Methods

We simulated ecological change in benthic macroinvertebrate communities using taxa-specific dose response and capture probability models (Fig. 9.2). First, we simulated reference communities for two sites within the Shenandoah National Park (SNP), USA, contrasting high- and low-richness species pools. Second, we modeled community change from taxa-specific responses to a generalized stressor gradient in the mid-Atlantic highlands, USA. Third, we sampled the simulated-stressed communities with different combinations of sample effort and taxonomic resolution and calculated the sensitivity of a commonly used bioassessment index (Macroinvertebrate Biotic Integrity Index, MBII; Klemm et al. 2003). Finally, we evaluated expected utilities for model sensitivity, sensitivity-cost efficiencies, and risk tolerance and concluded with a demonstration of how prior knowledge and uncertainty about site condition (i.e., reference, early warning, or impaired sites) may be used to optimize decision thresholds.

Simulating Reference Communities

The SNP study area encompasses approximately 800 km^2 within the Blue Ridge ecoregion and consists primarily of deciduous and coniferous forests (Young et al. 2009). This study area was useful for several reasons. First, it supports some of the

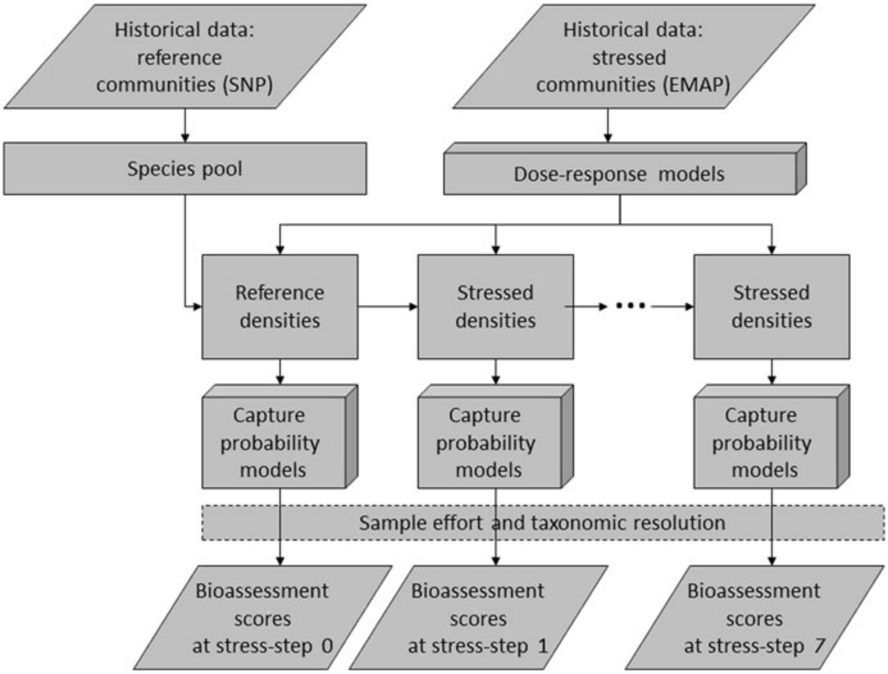


Fig. 9.2 Conceptual model of simulation methods. *Three-dimensional boxes* indicate taxa-specific models. Historical data from the Shenandoah National Park (*SNP*) were used to model the species pool for each study site. Data from the mid-Atlantic highlands Environmental Monitoring and Assessment Program (*EMAP*) were used to develop taxa-specific dose-response models

most high-quality streams in the eastern USA (Lynch 1987; Summers and Tonnessen 1998) and therefore provides meaningful reference communities. Second, several streams within the Park support long-term benthic macroinvertebrate community datasets as part of the National Park Service's Vital Signs Program (Marshall and Piekielek 2007). Third, the study area enabled the use of empirical data to model dose-response patterns within the mid-Atlantic highlands, a region encompassing approximately 200,000 km^2 in the Appalachian Mountains of eastern North America (Fig. 9.3).

We used data from annual benthic macroinvertebrate surveys conducted from 1990 to 2004 at two SNP sites, Hazel River and Paine Run. We selected these focal streams because they both contained $>$ 10 years of annual sampling data but supported different species richness and environmental conditions (Table 9.1). At each sampling site, National Park Service personnel used 0.1 m^2 portable invertebrate box samplers (PIBS) to collect benthic macroinvertebrates from three randomly selected riffle locations (Moeykens and Voshell 2002). Subsamples were then pooled into a composite fixed-area sample for each site. We evaluated data from collections taken during spring months (April–June).

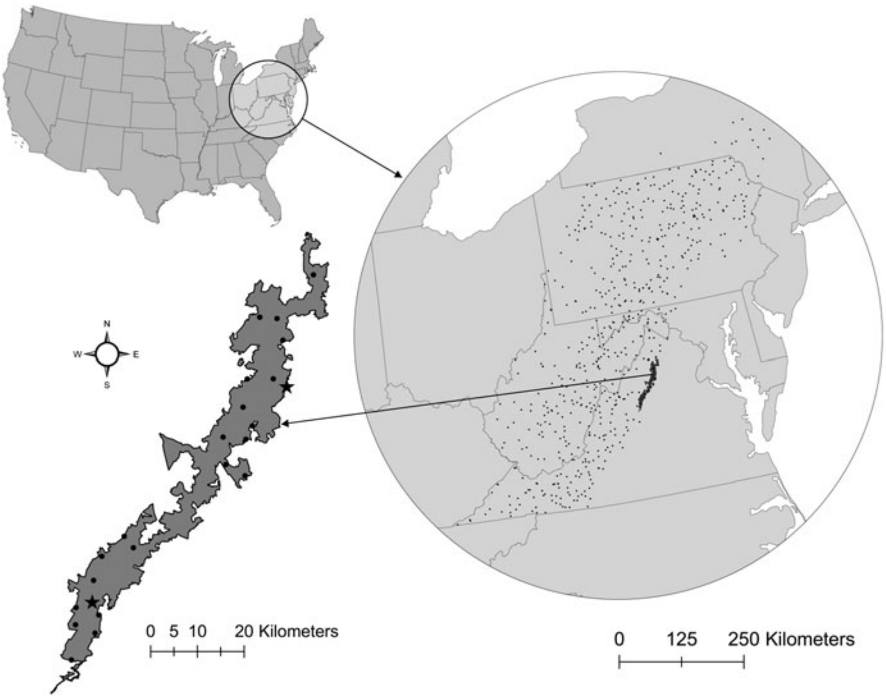


Fig. 9.3 Study sites within the mid-Atlantic highlands and Shenandoah National Park, USA. Regional inset shows Environmental Monitoring and Assessment Program (EMAP) sites used to simulate stress responses; and Shenandoah National Park inset shows the location of all Vital Signs monitoring sites (circles), as well as the Hazel River and Paine Run sites (stars) evaluated in this study.

Table 9.1 Biological and physical characteristics of two Shenandoah National Park sites evaluated in this study

Variable	Hazel River	Paine Run
Biological		
Genus richness	92	67
Family richness	51	40
Mean generic richness	43	28
Mean probability of capture	0.50	0.36
Physical		
Basin area (ha)	1,335	179
Site elevation (m)	323	557

The Hazel River site included records of 21,130 individuals from 51 families and 92 genera; Paine Run supported 25,746 individuals from 43 families and 77 genera. Cumulatively, 106 unique macroinvertebrate taxa were represented in the SNP dataset and the focal sites shared 67 genera and 40 families. The Hazel River site was substantially larger than Paine Run (Table 9.1), which probably explains why it supported greater taxonomic richness. We compared simulation models for Hazel River and Paine Run to evaluate how natural variation in community composition may affect bioassessment performance.

Simulating Ecological Change

We modeled taxa-specific dose-response patterns using data from the US Environmental Protection Agency's (USEPA) Environmental Monitoring and Assessment Program (EMAP) in the mid-Atlantic highlands, USA ($n = 439$ sites; Fig. 9.3). EMAP sites were selected using a stratified-random methodology to represent the availability of stream sizes in the study area (Herlihy et al. 2000). Macroinvertebrate communities in these sites were sampled using equivalent effort across a stressor gradient that included reference and impaired sites (Waite et al. 2004).

USEPA personnel sampled sites during baseflow conditions between April and September from 1993 to 1998. At each site, they sampled macroinvertebrates from nine transects positioned perpendicularly to the stream channel. Within each transect, they collected organisms from a $0.5\,\mathrm{m}^2$ quadrat by kick-net sampling into a 595-μm mesh for 20 s (Lazorchak et al. 1998). Quadrat samples were then combined into "riffle" and "pool" composites based on quadrat locations, constituting fixed-area samples. In the laboratory, specimens were identified to the lowest taxonomic resolution practical (usually genus or species) and enumerated (Lazorchak et al. 1998). Raw data are available at http://www.epa.gov/emap.

We limited our analysis of EMAP data using several criteria. First, we used only riffle data for analysis because pool composites were not collected at all EMAP sites and because riffle data discriminate between reference and non-reference sites more powerfully than pool data (Klemm et al. 2003). Moreover, other protocols for benthic macroinvertebrate sampling focus on riffle habitats (e.g., Moulton et al. 2002). Second, we limited data to sites sampled during the spring (April to June) to coincide with the sampling periods for the SNP long-term monitoring program. We also excluded sites with upstream catchment areas $> 500\,\mathrm{km}^2$ to limit the analysis to wadeable streams. Where sites were sampled more than once, we used the sample with the greater species richness, or in the case of ties, the more recent sample.

We used an index of land use intensity to indicate environmental quality at each site. Catchment land use is strongly correlated to water quality (e.g., nutrient concentrations) and physical habitat quality (e.g., substrate size, woody debris abundance) within the study area (Hitt and Angermeier 2008), and consequently may be used as a generalized stressor gradient. Specifically, we used the landscape development intensity index (LDI) (Brown and Vivas 2005) calibrated for streams in Maryland (D. White, USEPA, unpublished data). The LDI is an index of potential human disturbance that weights individual land uses (e.g., agriculture, urbanization, mining) based on the amount of anthropogenic energy required to maintain a given land use type (Brown and Vivas 2005). The index is based on the hypothesis that land uses which require more energy to maintain are more detrimental to biological integrity than land uses which require less energy (Odum 1988, 1996). LDI strongly correlates with biotic integrity in a variety of ecosystems (McCarrondand and Frydenburg 2005; Lane and Brown 2006; Mack 2006).

We used LDI coefficients based on 19 National Land Cover Data classifications (6 classes of agriculture, 12 of urban, and 1 of mining). To match EMAP land use data,

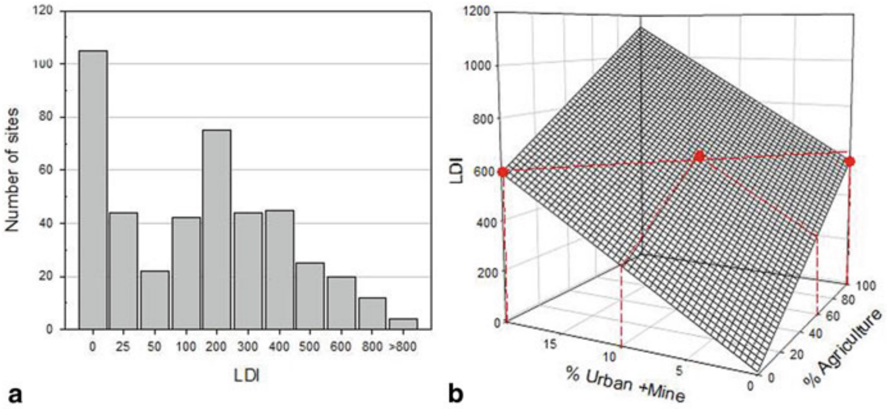


Fig. 9.4 Description of landscape development intensity (*LDI*) gradient for EMAP sites. **a** Observed distribution of LDI among 439 Environmental Monitoring and Assessment Program sites in the mid-Atlantic highlands, USA. **b** Relationship between LDI and upstream agriculture and urban/mining land uses. An example of how three different land use combinations can yield similar LDI values is highlighted at LDI = 600

we averaged LDI coefficients for agriculture (5.6), urbanization (29.3), and mining (35.5). Relative differences in these LDI coefficients suggest that urban and mining land uses are on average about 5.3 and 6.4 times more disruptive to biotic integrity than agriculture land use. This ratio roughly corresponds to ecological thresholds based on land use derived from field research. Several studies have reported significant declines in fish and benthic macroinvertebrate bioassessment indices when agricultural development exceeds 30–50 % (Quinn and Hickey 1990; Wang et al. 1997), and urban land exceeds 6–10 % of watershed area (Roy et al. 2003; Snyder et al. 2003).

LDI scores for the 439 EMAP sites ranged between 0 in completely forested watersheds to more than 3,000 in a watershed dominated by mining. Nearly 25 % of the sites (*n* = 105) were in reference condition (LDI = 0) and all but four sites had LDI scores less than 800 (Fig. 9.4a). Comparable LDI values can be observed from different combinations of land uses. For example, LDI values near 600 can be obtained for watersheds comprised of 100 % agriculture; watersheds with 20 % urban and 80 % forest; or watersheds with 60 % agriculture, 10 % urban, and 30 % forest (Fig. 9.4b).

We modeled abundance of SNP taxa (*N* = 106) as a function of LDI, upstream basin area, and site elevation. Scatter plots revealed wedge-shaped relations between macroinvertebrate abundances and anthropogenic stress levels (Fig. 9.5), which suggested that dose-response relationships may be better characterized by changes in the variance of the response than in central tendency (Scharf et al. 1998). We therefore used quantile regression techniques to model this heteroscedastic dose-response relationship (Koenker and Hallock 2001; Cade and Noon 2003). Other studies have successfully used quantile regression to model macroinvertebrate responses to land use-stressor gradients (e.g., Purcell et al. 2009).

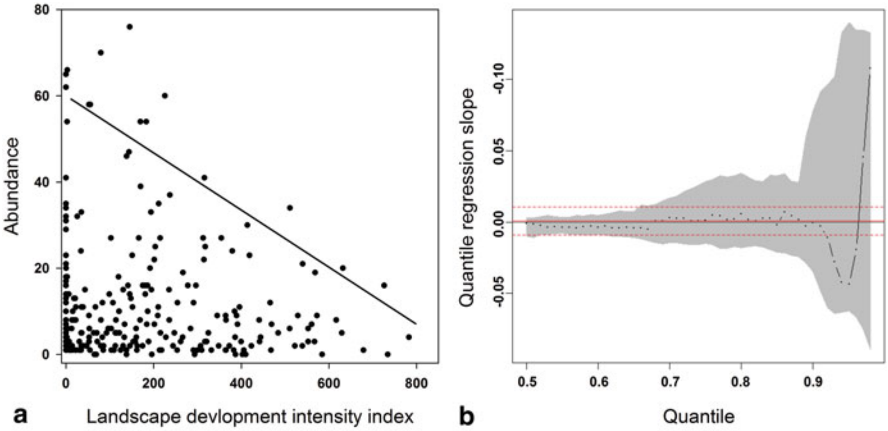


Fig. 9.5 Example of quantile regression for dose-response models. **a** Abundance of *Baetis sp.* (mayfly) along the landscape development intensity index with 95 % quantile regression line shown. **b** Analysis of tau level by 0.01 increments to assess nonzero slope for quantile regression. *Straight lines* indicate ordinary least-squares regression slopes for comparison. The *shaded region* shows range of slopes indicated by permutation procedures

For each taxon, we determined the appropriate quantile level (i.e., tau) for regression analyses by evaluating quantile model slopes across 1-unit tau increments from the median (50th quantile) to the 99th quantile in each dose-response model. We determined the significance of these dose-response models as whether the 95 % confidence intervals departed from least squares regressions. Bootstrapped 95 % confidence intervals for each tau increment were evaluated for their departure from zero to indicate significant dose-response patterns. Tau levels yielding nonzero slopes were used for subsequent quantile regressions. If taxa had too few observations to model tau levels, we combined taxa to evaluate dose-response patterns (see Appendix A). We included upstream basin area and site elevation as covariates in all quantile models because macroinvertebrate community structure varies by stream size in the study area (Waite et al. 2000). We performed quantile analyses in R using the package "quantreg" version 4.30 (R Development Core Team 2009; Koenker 2010).

Quantile regression results indicated a wide range of responses to stress among the 106 taxa in the species pool. The mean proportional change in abundance between reference conditions (i.e., LDI = 0) and non-reference conditions (i.e., LDI = 600) ranged from – 1.0 (i.e., extirpation) to 9.9 (i.e., nearly 10-fold increase in abundance). However, the majority of taxa in both sites exhibited decreasing abundances with increasing stress (Fig. 9.6a, b; Appendix A).

Taxa responses to LDI were highly correlated between the two SNP sites (Pearson's $r = 0.95$) indicating that basin area, elevation, and initial densities had only minor effects on taxonomic vulnerability to stress. However, the high-richness Hazel River community was on average more sensitive to stress than the low-richness Paine Run community (mean change = −0.18 and – 0.08, respectively). Community-level

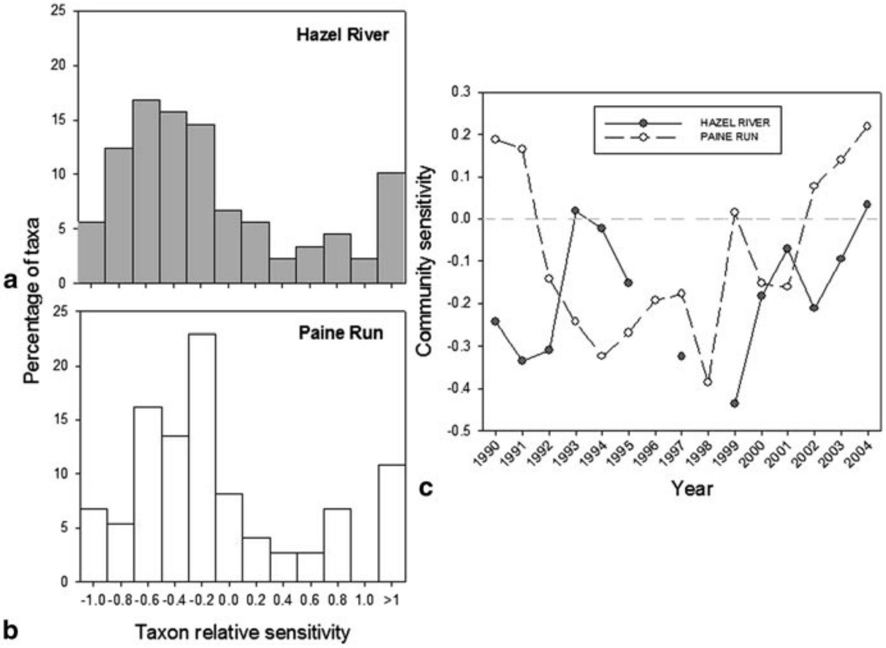


Fig. 9.6 Taxa-specific sensitivity values (**a** and **b**) and yearly community sensitivity values (**c**) for Shenandoah National Park sites. Taxon-specific sensitivity is defined as the average proportional change in taxon abundance between simulations at reference (LDI = 0) and simulations at the end of the "early-warning" portion of the stress gradient assessed (LDI = 600). A value of "− 1" indicates taxon extirpation, and a value of "+ 1" indicates taxon doubled in density. Community sensitivity is the relative sensitivity of the overall assemblage and is the sum of the products of the relative sensitivity and corresponding abundance of each taxon divided by the total abundance of the assemblage

vulnerability to stress was highly variable for both sites across the 15 years of historical monitoring data (Fig. 9.6c) indicating substantial natural variation in vulnerability to stress.

We modeled changing benthic macroinvertebrates communities at 7 stress-steps, corresponding to LDI values of 50, 100, 200, 300, 400, 500, and 600. LDI values are interpreted as a generalized stress gradient corresponding to land use intensity. At each stress-step, community composition was represented by a vector of taxa-specific densities.

Sampling Simulated Communities

Surveillance monitoring often requires estimating the abundance of taxa with imperfect capture probabilities. We used annual sampling data to estimate capture probabilities for taxa included in the SNP species pool ($N = 106$). We assumed that taxa observed within the historical records were present during all time periods, but subject to imperfect detection (Kery et al. 2009). For example, if a taxon was observed in 2 of 10 years (i.e., occurrence probability = 0.20), we assumed that 80 % of the observations failed to detect the taxon when it was present. We also assumed that capture probability would increase with abundance (MacKenzie and Kendall 2002).

We used logistic regression to model the relationship between taxa occurrence probability and abundance. Taxa-specific model coefficients were used to adjust community densities to account for imperfect detection across stress levels (Fig. 9.2) unless the fit was poor ($p > 0.10$) or an insufficient number of observations were available for logistic models. In those cases we used observed occurrence probabilities as an estimate of a constant capture probability which was invariant to abundance.

The average capture probability was less than 0.50 for both sites (Hazel River = 0.48; Paine Run = 0.36) and ranged between 0.07 (i.e., a taxon was captured 1 out of 15 years) and 1.0 (i.e., a taxon was captured every year). The average taxonomic richness was substantially less than the cumulative total number of taxa captured at both sites (Hazel River = 42.2; Paine Run = 28.1), indicating considerable turnover for many taxa. Taxa-specific capture probability coefficients are presented in Appendix A.

We then sampled the simulated communities assuming a negative binomial distribution for taxa abundances where variance > mean and a Poisson distribution otherwise. We drew 1,000 bootstrap samples with replacement from capture probability-adjusted densities under constraints of sample size and taxonomic resolution. Six levels of sampling effort were evaluated: 100, 300, and 600 fixed-count samples and 3, 6, and 9 fixed-area samples. At each stress-step, we calculated an index of stream quality, the Macroinvertebrate Index of Biotic Integrity (MBII), to evaluate bioassessment performance (Fig. 9.2). The MBII was developed to assess ecological integrity of streams in the mid-Atlantic highlands region, USA (Klemm et al. 2002, 2003). Seven metrics are included in the calculation for MBII that represent the taxonomic and functional composition of the macroinvertebrate community (Table 9.2). Unlike traditional means of computing MBII scores, no adjustments were required for basin area, and individual metric scores were standardized by the maximum observed at the site. Increasing MBII scores indicate increasing ecological quality (Klemm et al. 2002, 2003). In our model, MBII scores were calculated from each stress-step and bioassessment design scenario applied to our bootstrapped reference communities. We programmed simulations, sampling, and index calculations using MATLAB version 7.7.0.

Model Validation

To be useful, simulated communities need to reflect both natural variation in assemblage structure and changes in assemblage structure due to stress (Dale and Beyeler 2001). We evaluated these factors using nonmetric multidimensional scaling (NMS) ordinations. This method is generally considered to be the most effective ordination method for ecological community data (McCune and Grace 2002) and is widely used in bioassessment (e.g., Reynoldson et al. 1997; Cao and Hawkins 2005). First, we evaluated simulated reference communities by comparing historical community data

Table 9.2 Component metrics of the Macroinvertebrate Biotic Integrity Index (MBII) and their hypothesized response to stressors. (Klemm et al. 2002)

Metric	Response to stress
Number of Ephemeroptera (mayfly) taxa	Decline
Number of Plecoptera (stonefly) taxa	Decline
Number of Trichoptera (caddis fly) taxa	Decline
Number of collector-filterer taxa	Decline
Percent of noninsect individuals	Increase
Mean Tolerance Value	Increase
Percent of individuals in 5 dominant taxa	Increase

against 15 randomly selected bootstrap samples for the Hazel River and Paine Run sites. We reasoned that if the ordination revealed similar patterns for simulated and historical samples then our simulation models could realistically indicate patterns of natural variation in reference community composition.

Second, we ordinated simulated-stressed communities and EMAP field samples collected across the LDI stressor gradient to evaluate patterns of simulated ecological change. For this ordination, we categorized EMAP samples as reference sites ($LDI = 0$, $n = 78$) and impaired sites ($LDI > 300$, $n = 79$). We used the average densities from the 1,000 model runs to represent simulated-stressed sites in the ordination and only included EMAP taxa which were present in the SNP species pool. If our simulation models were meaningful, changes in NMS site scores among stress steps should resemble changes between reference and impaired sites from regional field data. For both ordinations, data were log-transformed taxon-specific abundances. We used PC-ORD (version 5.32) for NMS ordinations (McCune and Grace 2002).

Evaluating Utility

We evaluated bioassessment performance using three expressions of utility. The first expression was a standard sensitivity or power analysis. In a sensitivity analysis, all the utility value is assigned to correctly classify impaired conditions, which occurs with a frequency of 1—type II error rate. Value for correctly classifying reference condition is implicit in the a priori setting of the type I error rate, and it is well understood how the type I error rate affects sensitivity or statistical power. In surveillance monitoring, the type I error rate typically determines the decision threshold. Sensitivity was calculated from the distributions of MBII scores in simulated reference and simulated impaired communities. Specifically, we expressed sensitivity as the percent of 1,000 bootstrapped community samples that fell below a decision threshold for MBII corresponding to a type I error rate of 0.10 in reference conditions (see Fig. 9.1). Using this threshold, we interpolated the minimum detectable change (MDC) in LDI based on statistical power of 0.80 for each alternative monitoring design (combination of sample effort and taxonomic resolution).

In our second expression of utility, we explicitly accounted for monitoring costs. We estimated financial costs for benthic macroinvertebrate analysis (Table 9.3) and incorporated cost into a sensitivity-cost efficiency model. We then compared

Table 9.3 Description of cost estimates used in this study for lowest (LOW) and family (FAM) taxonomic resolution. Costs for alternative sampling designs were based on estimates for processing time and personnel wages provided by managers from the Mid-Atlantic region, USA (including Shenandoah National Park) and Eastern Rivers and Mountains Vital Signs networks (US National Park Service)

Sample effort	Processing time (man-hours)				Costs per man-hour ($)				Total costs per site ($)	
	Sorting		Identification		Sorting		Identification			
	LOW	FAM	LOW	FAM	LOW	FAM	LOW	FAM	LOW	FAM
100	1.0	1.0	0.8	0.5	10.00	10.00	32.00	20.00	42.00	30.00
300	2.0	2.0	1.8	1.0	20.00	20.00	72.00	40.00	92.00	60.00
600	3.0	3.0	3.5	1.6	30.00	30.00	140.00	64.00	170.00	94.00
PIBS-3	7.5	7.5	9.0	4.0	75.00	75.00	360.00	160.00	435.00	235.00
PIBS-6	15.0	15.0	18.0	8.0	150.00	150.00	720.00	320.00	870.00	470.00
PIBS-9	22.5	22.5	27.0	12.0	225.00	225.00	1,080.00	480.00	1,305.00	705.00

bioassessment designs in terms of their sensitivity-cost relationship. Examination of the sensitivity-cost plots enabled an analysis of possible trade-offs between statistical power and costs which are fundamental to the efforts to optimize monitoring program designs.

In our third expression of utility, we incorporated value for classification of both reference and impaired conditions and allowed risk tolerance to vary among resource conditions (Figs. 9.1 and 9.7). As a heuristic analysis, we evaluated a hypothetical scenario of risk tolerance by assuming that managers' tolerance for misclassifications would vary among resource conditions. We specified three states of resource condition (reference, early warning, and impaired communities) and set different risk tolerance values for classifications in each state (Fig. 9.7). The three resource states along with correct or incorrect classification resulted in six classification scenarios determined by observing a site above or below the decision threshold in reference and impaired communities (see Fig. 9.1). In our heuristic example, tolerance for misclassification was higher for reference condition (i.e., type I error) than for impaired conditions (i.e., type II error). In other words, we assigned some value to misclassifying reference as impaired knowing that would increase the sensitivity of the decision threshold to correctly detect impaired condition. We then calculated expected utility as the product of event probabilities and utility values to compare decision thresholds based on type I error rates of 0.10, 0.15, 0.20, and 0.30.

Results

Simulation Model Validation

Simulated reference communities were similar to historical samples at both sites as indicated by their close proximity in two-dimensional NMS ordination space

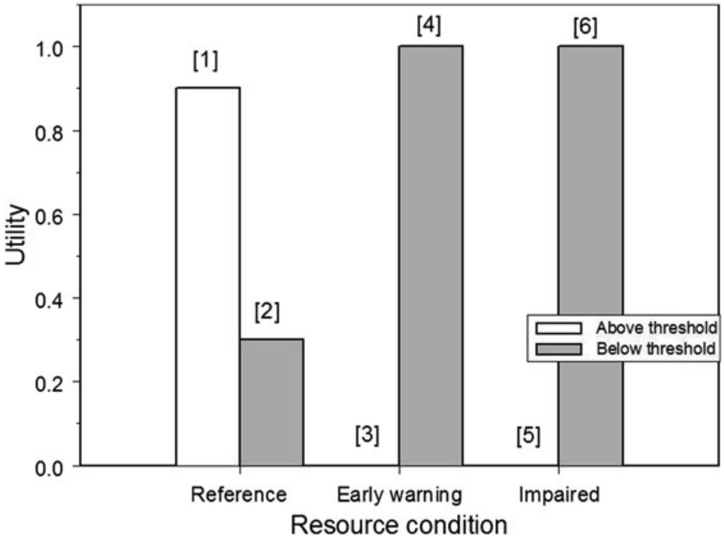


Fig. 9.7 Utility values assigned for risk analysis. Numbers for reference, early-warning, and impaired sites correspond to regions defined in Fig. 9.1. In this heuristic example, managers express greater utility for observing change in early-warning and impaired sites than in reference sites

(stress = 0.09; Fig. 9.8a). However, historical samples were more heterogeneous than modeled reference communities, as indicated by the larger space occupied by historical samples in the ordination. Post-hoc analysis revealed that the extreme values for historical samples (i.e., two Hazel River samples with high NMS axis II scores, and two Paine Run samples with highest NMS axis I scores; Fig. 9.8a) were from years with very low total macroinvertebrate abundances. This pattern suggests that our simulation procedures provided a realistic model for reference communities overall but may have underrepresented low abundance sample years.

Our ecological change model simulated realistic community responses to a generalized stressor gradient. Reference and impaired EMAP sites were highly differentiated in a two-dimensional NMS ordination, with site quality primarily associated with NMS axis I (Fig. 9.8b). Reference sites tended to have lower NMS axis I scores than impaired sites. Likewise, NMS axis I scores increased with stress for the artificially impaired assemblages in both SNP sites. NMS axis I scores were highly correlated with LDI for EMAP sites and simulated-stressed sites (Spearman $r = 0.67$ and 0.97, respectively), indicating that empirical and simulated assemblages responded to the stress gradient in similar ways. However, simulated communities at the maximum stress-step did not reach the non-reference community zone defined by EMAP sites (Fig. 9.8b). Our simulation model therefore produced realistic changes in community composition but did not represent fully impaired sites elsewhere in the mid-Atlantic highlands region.

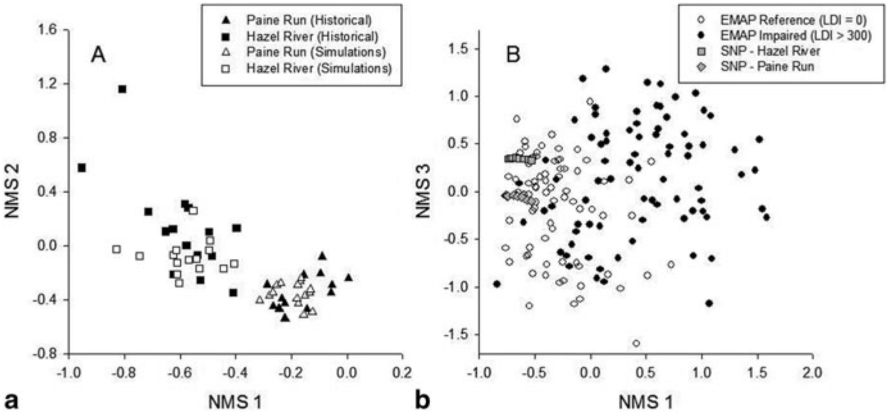


Fig. 9.8 Two-dimensional nonmetric multidimensional scaling ordinations of macroinvertebrate community structure. **a** Comparison of reference assemblages between 15 randomly selected simulated samples and historical samples collected at the two Shenandoah National Park sites over the 15-year monitoring period. **b** Comparison of assemblages at reference sites (LDI = 0, $N = 78$) and stressed sites (LDI > 300, $N = 79$) observed from EMAP data, and simulated-stressed assemblages at the two SNP sites. The simulated assemblages are defined by LDI values of 0, 50, 100, 200, 300, 400, 500, and 600 (plotted *from left to right*)

Sensitivity

Simulation-derived MBII scores ranged from 20.1 to 99.6 (Fig. 9.9). The lowest MBII values occurred in the low-richness site (Paine Run) at the lowest sample effort (100 fixed count) and highest values occurred at the high-richness site (Hazel River) at the highest sample effort (9 fixed-area samples; Fig. 9.9). Mean MBII scores increased with sample effort and were generally greater at the family level than genus level, but taxonomic resolution had a negligible effect on MBII scores at the highest sampling efforts in Hazel River (Fig. 9.9).

Sensitivity increased throughout the simulated stress gradient for all combinations of sample effort and taxonomic resolution (Fig. 9.10). Hazel River MBII scores generally exhibited greater sensitivity than in Paine Run, and Hazel River exhibited asymptotic sensitivity at approximately LDI of 400, whereas this pattern was not evident in Paine Run (Fig. 9.10). Genus-level MBII scores revealed greater sensitivities than family-level scores for all sampling efforts in both sites, but the effect of taxonomic resolution increased with sampling effort, especially in the Hazel River site (Fig. 9.10). All else being equal, sensitivity increased with sample effort and was higher for the high-diversity site and for the genus level of taxonomic resolution (Fig. 9.10).

Comparisons of mean detectable change (MDC) at 80 % sensitivity revealed important distinctions across sites, sampling efforts, and taxonomic resolutions (Fig. 9.11). Hazel River site showed lower MDC levels than Paine Run (i.e., greater sensitivity), and genus-level MBII scores improved MDC over family-level scores

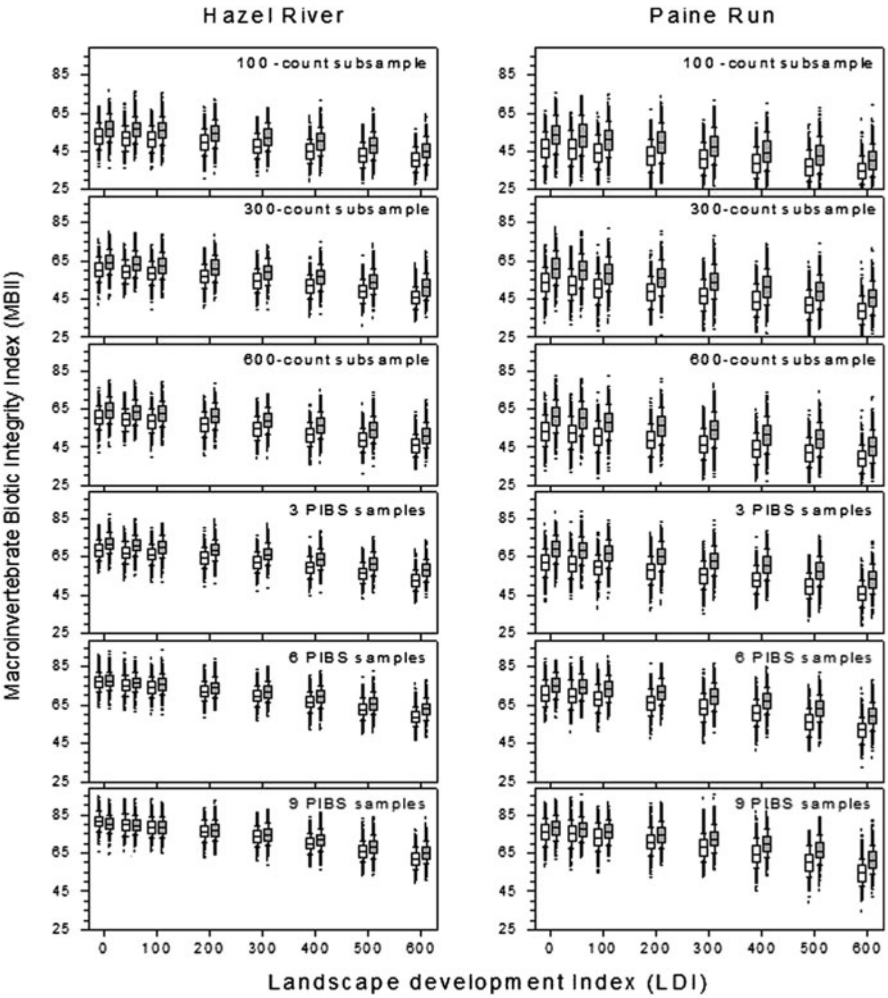


Fig. 9.9 Effects of sample effort and taxonomic resolution on Macroinvertebrate Biotic Integrity Index (MBII) scores across modeled stress-steps (Landscape Disturbance Index, LDI) in Hazel River and Paine Run. *Boxplots* show median scores (*center line*) and quartile distributions (*box range*). *Open boxes* indicate genus-level taxonomy; *closed boxes* indicate family-level taxonomy

at both sites. However, the genus-level scores in Paine Run were not as sensitive as the family-level scores from Hazel River (Fig. 9.11), demonstrating the importance of site-level conditions in evaluating sensitivity. Both sites exhibited nonlinear relationships between sample size and MDC : increasing sample effort in fixed-counts resulted in stronger improvements in MDC than increasing sample effort from fixed-area samples (i.e., asymptotic MDC for fixed-area samples). Genus-level fixed-area samples showed more improvement in MDC than family-level fixed-area samples (Fig. 9.11). However, taxonomic resolution in the low-diversity site (Paine Run) was relatively unimportant for MDC at the lowest sampling efforts.

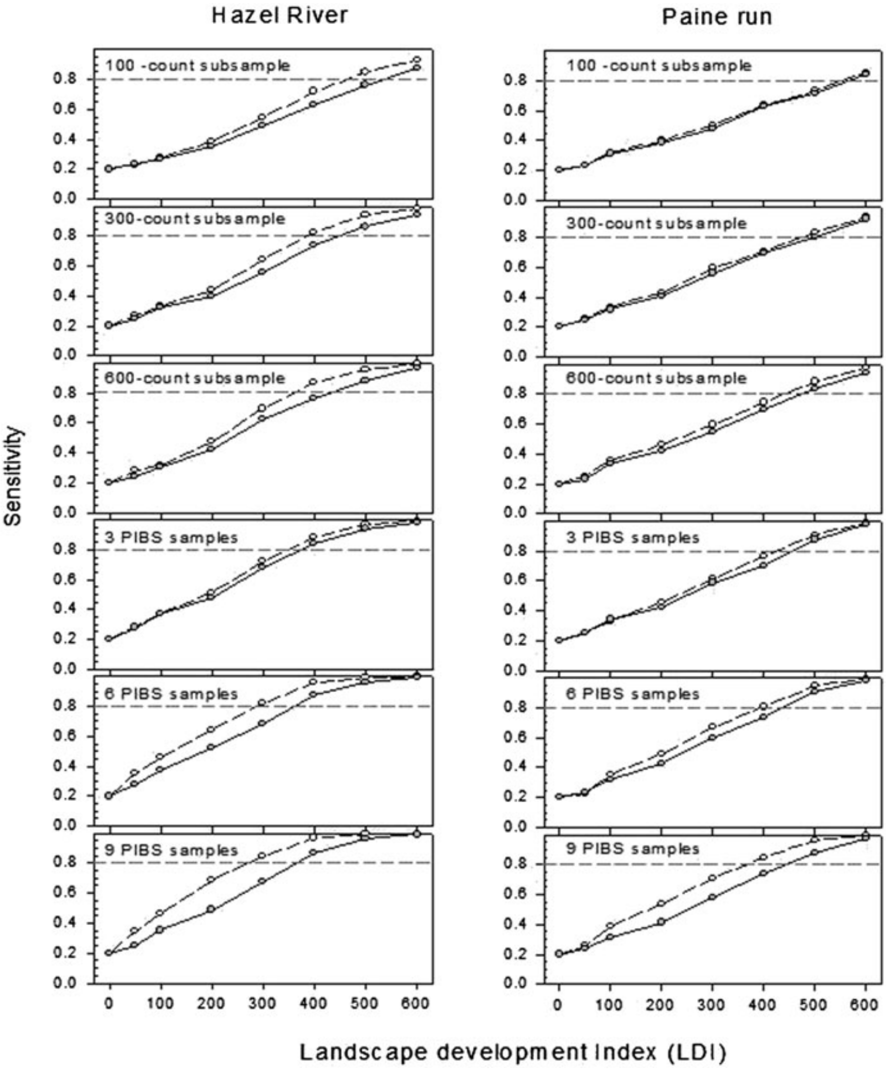


Fig. 9.10 Effects of sample effort and taxonomic resolution on Macroinvertebrate Biotic Integrity Index (MBII) sensitivity across modeled stress-steps (Landscape Disturbance Index, LDI) in Hazel River and Paine Run. Sensitivity values are based on type I error rates of 0.20. A *dashed horizontal line* at sensitivity of 0.80 (type II error = 0.20) is included to facilitate comparisons. *Open circles* indicate genus-level taxonomy; *closed circles* indicate family-level taxonomy

The increase in sensitivity with increasing sample effort was due to higher precision (i.e., lower variance) of reference MBII scores and not to greater responsiveness (i.e., change in mean MBII scores between reference and impaired assemblages). For example, MBII coefficients of variation (CV) for simulated reference assemblages at the highest sample effort (9 fixed-area samples) were about half of that observed for the lowest-sample effort (100 fixed-counts samples), and this was true for both levels

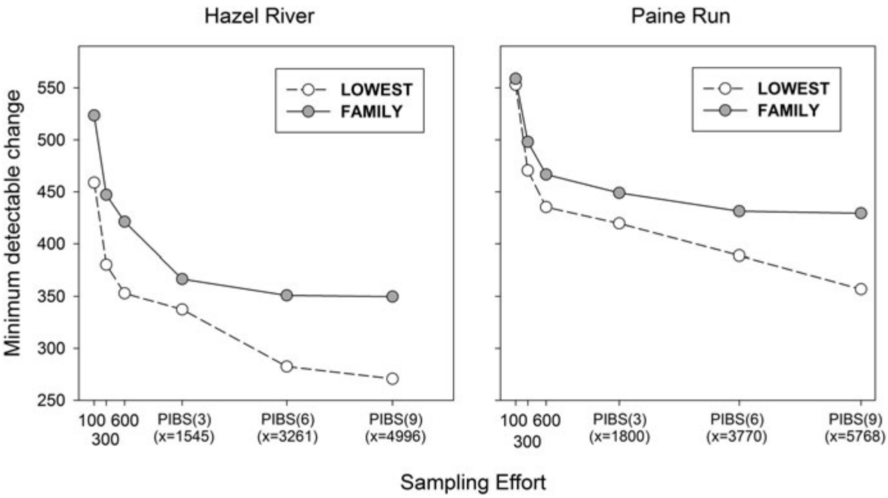


Fig. 9.11 Minimum detectable change (MDC) for alternative monitoring program designs based on combinations of sample effort and taxonomic resolution for Hazel River and Paine Run. MDC values are based on type I and II error rates of 0.20. Numbers in parentheses below fixed area samples (PIBS-3, PIBS-6, and PIBS-9) represent the mean number of individuals contained within composite samples.

of taxonomic resolution and for both sites. At the same time, the percent change in average MBII scores between reference (LDI = 0) and impaired (LDI = 400) assemblages (i.e., responsiveness) actually declined slightly with increasing sample effort.

Similarly, differences in sensitivity between sites were also due to precision and not responsiveness. For example, precision (1/CV) of MBII values at reference was between 25 and 35 % higher at Hazel River site compared to Paine Run for both genus and family levels of taxonomic resolution, and patterns were consistent across sample efforts (Fig. 9.12a). At the same time, responsiveness was actually higher at the Paine Run site for both genus and family though large differences were mainly noted for smaller sample efforts (Fig. 9.12a).

In contrast to sample effort and sites, differences in index sensitivity between levels of taxonomic resolution were due to differences in responsiveness and not to differences in precision at reference, and this was true for both sites (Fig. 9.12b). Responsiveness increased with sample effort and was always higher for genus level resolution though differences were marginal at low sample effort for the Paine Run site (Fig. 9.12b). At the same time, precision was generally better at the family level at both sites and across levels of sample effort (Fig. 9.12b).

Sensitivity-Cost Efficiency

Laboratory costs for taxonomic identification were nonlinearly related to sensitivity or statistical power (i.e., MDC at 0.20 type II error rate), revealing asymptotes for cost efficiency and trade-offs between sample size and taxonomic resolution (Fig. 9.13).

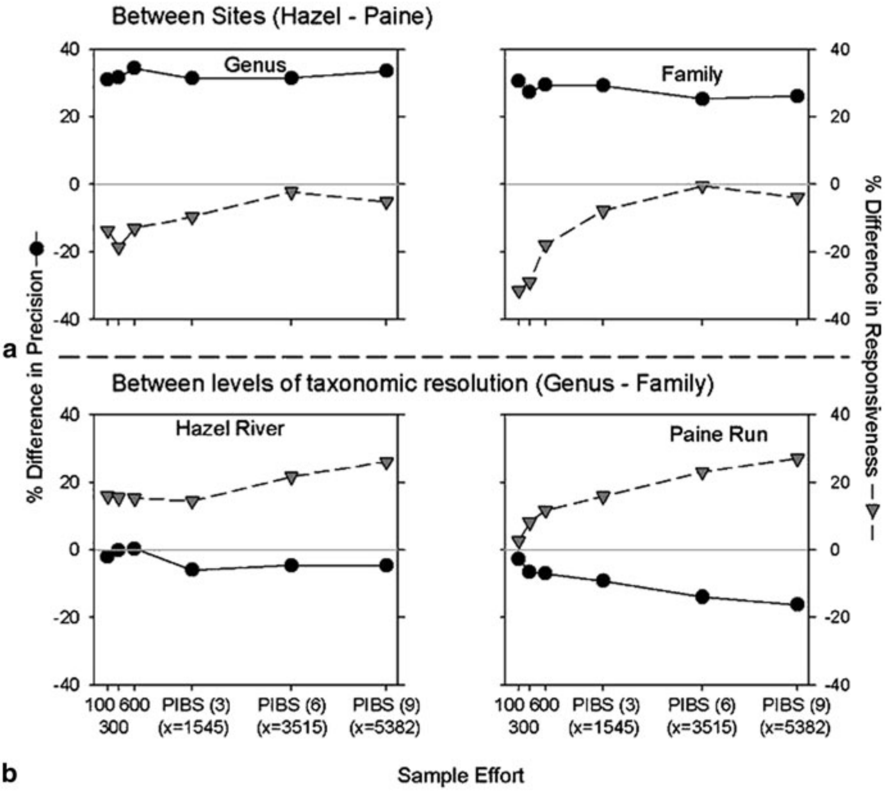


Fig. 9.12 Differences in precision and responsiveness (mean change between reference and impaired) of Macroinvertebrate Biotic Integrity Index (MBII) scores. **a** Comparisons between sites and **b** comparisons between taxonomic resolutions. Precision is expressed as the inverse of the coefficient of variation in MBII scores within reference sites. Responsiveness is expressed as the mean change in MBII scores between LDI = 0 (reference) and LDI = 400 (see methods). Site differences are reported relative to the Hazel River so that *positive values* indicate Hazel River was better (i.e., precision or responsiveness higher) and *negative values* indicate Paine Run was better. Likewise, differences between genus and family resolutions are reported relative to the genus level so that *positive values* indicate genus was better

In both sites, additional fixed-count samples dramatically increased power but had relatively small cost increases. However, additional fixed-area samples contributed relatively little power with considerable additional costs (Fig. 9.13). In the Hazel River site, genus-level taxonomy maximized power (i.e., lowest MDC) but incurred the greatest costs. In contrast, the Paine Run site revealed that increased sample size at the family level could achieve the same power as lower sample sizes at the genus level (Fig. 9.13). Designs involving higher sample efforts yielded only slight improvements in sensitivity for much greater cost.

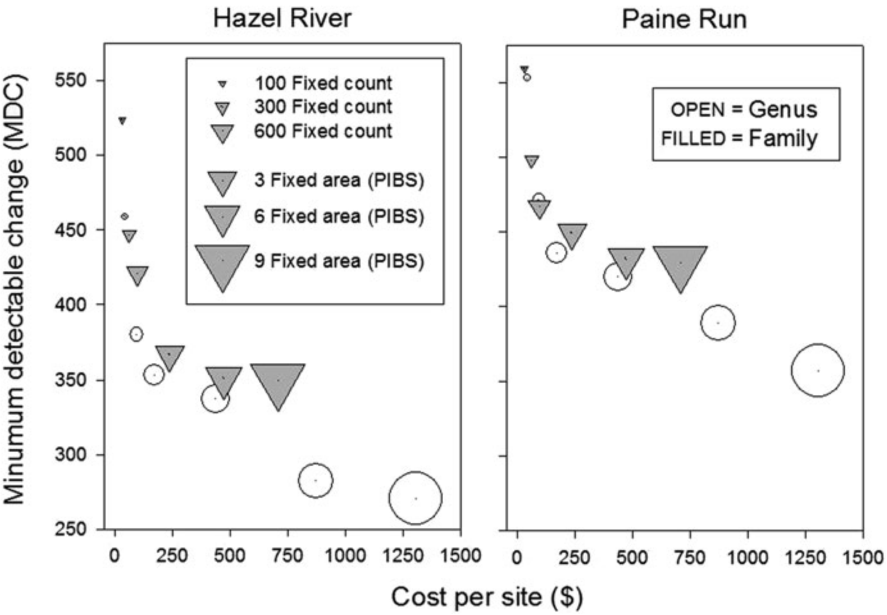


Fig. 9.13 Sensitivity and costs of alternative sampling designs defined by sample effort and taxonomic resolution in Hazel River and Paine Run. Sensitivity was expressed as minimum detectable change (MDC) at 80 % power. Lower values of MDC indicate greater sensitivity

Risk Tolerance

Decision thresholds associated with maximum expected utility varied with prior belief regarding ecological change (Fig. 9.14). In our heuristic example, we assumed that managers would be more risk-averse for misclassifications in impaired conditions than in reference conditions (Fig. 9.7). We calculated expected utility for three models of ecological change: no change from reference, gradual change, and abrupt change. If a manager believes the no-change model is most likely then a decision threshold determined by the lowest type I error rate (0.10) would maximize expected utility. However, in either change model the type I error rate that maximized the expected utility shifted as resource condition changed from reference through levels of impairment (Fig. 9.14). At the highest levels of impairment, expected utility was approximately equal for all type I error rates. If a manager believes change is most likely then a decision threshold based on high type I error rates would be warranted.

Discussion

Simulation Model Development and Validation

To be useful for bioassessment, stress-simulation models must start from realistic representations of reference community composition. In our model, we used annual

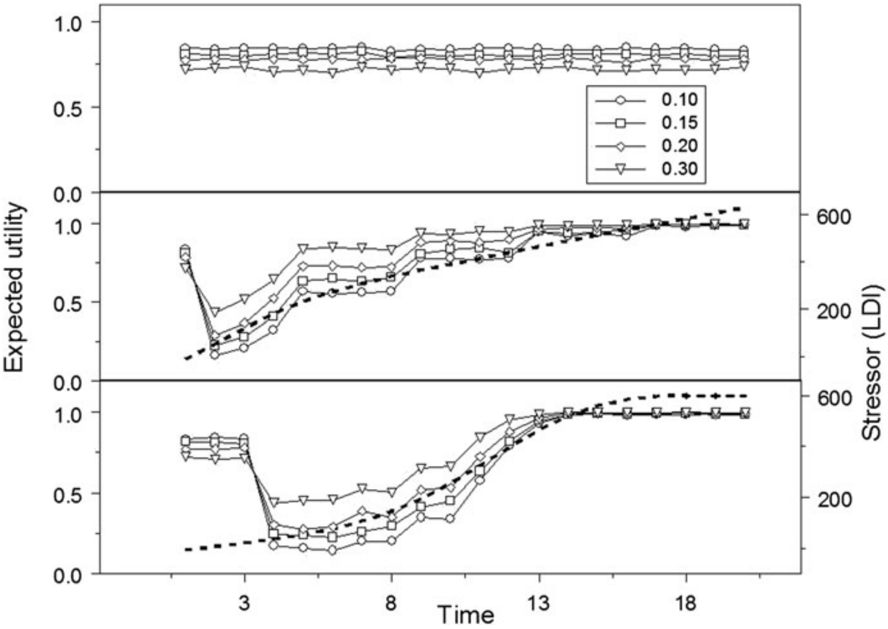


Fig. 9.14 Variation in expected utility among type I error rates over a 20-year period for three models of ecological change. Utility values for six events determined by correct or incorrect classification of resource condition are presented in Fig. 9.7. The *top panel* is a no-change model, i.e., the system remains in reference. The *middle panel* is a gradual change model, and the *bottom panel* is an abrupt change model. The underlying change in stressor is shown as *dashed lines*. Type I error rates were 0.1, 0.15, 0.2, and 0.3

monitoring data collected over 15 years to characterize natural variation in reference communities. Previous simulation models have relied on space-for-time substitutions to infer natural variability (Cao and Hawkins 2005; Hawkins et al. 2010), but spatial data may not fully portray natural variation from temporally structured natural events (i.e., floods and droughts; McElravy et al. 1989; Wagner et al. 2000; Kennen et al. 2010). We recognize that long-term datasets are rare, and that it may not be feasible to use temporal data for simulations in all circumstances. Nonetheless, in our application, temporal data were useful for quantifying natural variation and for modeling taxa-specific capture probabilities.

Comparisons of simulated and observed reference communities suggested that our model provided a reasonable representation of natural variation. However, our model did not capture years of very low macroinvertebrate densities. Based on historical data, low-density samples occurred approximately two times in 15 years and were probably the result of particularly stressful conditions (e.g., years with droughts or floods in the months preceding sampling). Consequently, our model may have overestimated precision and sensitivity to some degree because our model assumed independent taxon-specific densities. However, these effects were stochastic with respect to bioassessment design, and therefore should not bias inferences in this regard. Future research could improve upon our approach by incorporating taxa-abundance covariance structure into the simulation model.

To simulate stressed communities, we developed taxon-specific dose-response models from a regional dataset (Klemm et al. 2003). This approach is similar to that of Cao and Hawkins (2005) except that they modeled occurrence between reference and impaired sites to infer relative sensitivities of taxa, and then applied a constant to obtain changes in relative abundance. In contrast, we modeled changes in abundance directly. However, in both cases, relative sensitivities of taxa were empirically derived and not assumed, and the effects of site elevation and drainage area were included as covariates. Our model shared two other important attributes with that of Cao and Hawkins (2005): both models prohibited colonization of taxa that were not previously encountered and both assumed linear responses of individual taxa to stress.

Our model simulated realistic community responses to stress, based on the concordance between simulated stressed assemblages within the SNP and observed stressed assemblages across the mid-Atlantic highlands region. We found that increasing stress-steps in simulated communities moved community structure consistently towards the composition of known stressed communities. However, changes in actual abundances of individual taxa, and consequently the magnitude of overall changes in community composition, were less than expected based on empirical data. For example, simulated communities at the greatest stress-step (LDI = 600) did not "reach" the impaired community zone in a community ordination (Fig. 9.8b). Cao and Hawkins (2005) observed a similar pattern and attributed this to the inability of their model to account for colonization of novel species. However, in contrast to Cao and Hawkins (2005), we only evaluated taxa from the simulated sites (i.e., SNP taxa) in our NMS ordination and therefore colonization dynamics cannot explain our results. Instead, we believe that the assumption of linear responses to stress may have underestimated actual community dose-response patterns. Moreover, our model probably reflected early stages of ecological impairment and therefore may contribute to the "early-warning" detection objective of the NPS Vital Signs Program.

Sensitivity of Alternative Monitoring Designs

Our analyses revealed important effects of sampling effort and taxonomic resolution on bioassessment performance. First, we found that MBII sensitivity increased with sampling effort in both sites and levels of taxonomic resolution. This result is consistent with the findings of others (e.g., King and Richardson 2002; Ostermiller and Hawkins 2004; Cao and Hawkins 2005; Pond et al. 2008) and should not be surprising given the positive relationship typically observed between sample size and precision (Manly 2008). However, increasing sample effort beyond three fixed-area samples resulted in only modest improvements in sensitivity. This suggests that monitoring programs which rely on small subsamples may not be able to detect early-warning responses to stressors. Moreover, our cost-benefit analysis revealed that improvements in sensitivity from 100 to 600 individuals could be achieved at relatively modest increases in cost. We believe these findings are important in light of bioassessment programs that commonly rely on small fixed-count subsamples (i.e., < 300 individuals; Gerritsen et al. 2000; Burton and Gerritsen 2003).

Second, we found that genus-level assessments were more sensitive to ecological change than family-level assessments, even though precision increased at the family level (i.e., decreased variation in reference communities). This pattern was robust to sample size and environmental differences between sites. This supports the contention that benthic macroinvertebrate monitoring programs should use the lowest practical levels of taxonomic resolution (i.e., genus or species), especially for early-warning objectives (Lenat and Resh 2001; Pond et al. 2008). However, although the genus-level MBII was more sensitive at both sites, the high-diversity Hazel River site showed greater improvements with genus-level taxonomy than the low-diversity Paine Run site. Moreover, when we considered costs, trade-offs between taxonomic resolution and sample size became evident. For instance, we could achieve comparable levels of sensitivity at similar costs from 600 fixed-count subsamples identified to the genus level, or three fixed-area samples identified to the family level. Significant improvements in sensitivity are possible beyond these two alternative designs but only with substantial increases in costs, which, assuming a fixed budget, would result in the trade-off of fewer sites assessed (not considered in our study).

Other factors may affect how resource managers value bioassessment design considerations. For example, managers may use a genus-level design to enable comparisons to regional datasets that use this taxonomic resolution (e.g., Klemm et al. 2003). Alternatively, quality control and assurance issues (e.g., ambiguous and misidentified taxa) are a more serious concern for taxonomic identification at the genus and species level identifications than at the family level (New 1996; Marchant 2007). This concern may be particularly important for long-term monitoring programs where managers contract the identification of specimens to external laboratories that may change over time. A third consideration might be to employ a hybrid design wherein family-level taxonomy is used for annual status and trends assessments but genus-level taxonomy is used every 3–5 years for additional resolution.

Our simulations revealed differences in MBII sensitivity between the two SNP study sites. The MBII was more sensitive at the high-diversity Hazel River site for all alternative designs evaluated, and the effect of taxonomic resolution was greater at this site. In contrast, the effect of sample effort was fairly consistent between sites. Hawkins et al. (2000) also concluded that family-level taxonomy was sufficient to detect impairment in low-diversity streams but not in high-diversity streams (i.e., Great Britain and California, respectively). This is probably because low-diversity sites tend to support a relatively small number of genera and species per family and thus community composition patterns across taxonomic resolutions tend to be highly correlated. In contrast, high-diversity sites may contain many genera and species per family so within-family variation intolerance would be greater than in low-diversity sites. Thus, genus-level taxonomy is expected to be most responsive to ecological change at species-rich sites.

Although we found that differences in sensitivity between taxonomic resolutions *within a site* were due to responsiveness, differences *between sites* were due to higher precision at the Hazel River site and not to better responsiveness. This observation suggests that differences in sensitivity between SNP sites were not due to differences in diversity per se. Rather, reduced precision at the Paine River site was likely due to lower individual taxon capture probabilities (mean = 0.48 and 0.36 for Hazel River and Paine Run, respectively). We hypothesize that because Paine Run is much smaller than Hazel River it may be more vulnerable to natural disturbances such as floods and droughts that would likely reduce capture probabilities based on annual sampling and therefore reduce precision. In any event, these findings demonstrate the importance of site-level characteristics in evaluating sensitivity and suggest that optimum monitoring designs may be site- or region-specific.

Incorporating Risk Tolerance to Evaluate Type I Error Rate

In standard sensitivity or power analysis, the decision threshold is not evaluated because the type I error rate is fixed somewhat arbitrarily and subsequently ignored. This is an incomplete analysis because resource managers place value on correct classification of all resource conditions, costs, and risk tolerance associated with misclassification vary among resource conditions, and setting type I error rate has become a routine and uncritical process (Field et al. 2004). Utility values can be used to incorporate risk tolerance associated with classification of all resource conditions. We provided a heuristic example by assigning utility values. (In practice, these values should be elicited from the resource managers and decision makers. Thus, we present this as a heuristic example.) We found that setting of the type I error rate needs to account for prior belief in future changes to the resource. In general, if no change is likely, then type I error rate should be set low. However, if change is likely, then type I error rate should be set high. By integrating over the prior beliefs, bioassessment performance can be compared including the decision threshold/type I error rate, sample size, and other factors. Smith et al. (this volume) work through an example based on the simulation in this case study and found that a type I error rate of 0.20 was optimal for the range of sample sizes presented here using the utility values from Fig. 9.7 and the change models in Fig. 9.14.

Summary

Natural resource managers face numerous choices when developing bioassessment programs but seldom have the opportunity to compare the performance of alternative designs. As a result, managers usually fail to establish bioassessment programs based on their objectives for evaluating resource condition while accounting for uncertainty and controlling costs. In this chapter, we illustrated how simulation techniques can be used to evaluate sensitivity (statistical power) of alternative designs, and how the concept of utility values can be used to link sensitivity assessments with management values related to risk tolerance of misclassification (i.e., Type I and II error rates) and financial costs. The scope of our assessment was limited to 12 alternative designs based on 6 levels of sample effort and 2 levels of taxonomic resolution. However, simulations could also be used to evaluate the relative performance of bioassessment indices, the effects of rare taxa (i.e., include or exclude), the effects of seasonality (i.e., spring or fall), and the interactions among all variables. Our use of utility values in this chapter was heuristic, but we believe it illustrates the potential for this approach to provide quantitative links between management values and decision thresholds, something that has been lacking in bioassessment.

Appendix A

Ecological characteristics of macroinvertebrate taxa collected in Shenandoah National Park. We report information required to compute the MBII, results of taxon-specific dose-response models, and the results of capture probability models. Filter-feeders (1 if filter-feeder, blank if other) and pollution tolerance values (PTV) are required to calculate two of the seven MBII metrics. Feeding habitats were determined from Merritt and Cummins (1996) and PTV values were taken from Klemm et al. (2002). For the dose-response models, we report the proportional change in taxon abundance between LDI values of 0 (reference) and 600 (*P* Change), a measure of the relative sensitivity of each taxon to the LDI stress gradient simulated. For example, a proportional change value of "– 1" indicates increasing LDI from 0 to 600 resulted in local extirpation, and a value of "2" indicates a doubling of density. For capture probability, we report (1) the results of logistic regression models of capture probability on total community density ("Model"), and (2) the observed capture probability ("Observed") determined from 15 years of historical monitoring data. For taxa whose probability of capture was significantly related to total community density, we incorporated the logistic model parameters into simulations; otherwise, we used the observed capture probability. For the "model" capture probability, we report the log-odds of an increase in density of 500 individuals. For example, log-odds value of "2" indicates that the probability of capture doubled with an increase in total density of 500 individuals. For "observed" we simply report the ratio of the number of historical samples where the taxon was captured by the total number of samples. Both modeled and observed capture probabilities were site-specific.

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Class Insecta							
Order Plecoptera							
Family Pteronarcyidae		4					
Genus *Pteronarcys*		4	– 1.00		0.92		
Family Peltoperlidae		2					
Genus *Peltoperla*		3	– 0.40				0.13
Genus *Tallaperla*		1	– 0.73		0.85	3.03	0.93
Family Nemouridae		4					
Genus *Amphinemura*		5	– 0.09		0.23	5.80	0.87
Genus *Prostoia*		4					
Family Perlidae		3					
Genus *Paragnetina*		3	– 0.21		0.38		
Genus *Agnetina*		2	0.09		0.08		
Genus *Acroneuria*		3	– 0.27		1.00		0.73
Genus *Eccoptura*		3	– 0.18		0.15		0.53
Genus *Perlesta*		4	– 0.74		0.23		0.47
Genus *Hansonoperla*		2					
Family Perlodidae		2					
Genus *Yugus*		3	– 0.50			1.90	0.13
Genus *Remenus*		2	– 0.40				0.13
Genus *Isoperla*		3	– 0.66		0.85		0.07
Genus *Malirekus*		3					
Family Chloroperlidae		1					
Genus *Alloperla*		1	0.07	2.8	0.08		
Genus *Haploperla*		2	– 0.63		0.54		0.07
Genus *Sweltsa*		2	– 0.96		0.31	3.17	0.40
Genus *Suwallia*		1	– 1.00		0.15		0.13
Family Taeniopterygidae		3					
Genus *Oemopteryx*		3					
Genus *Taeniopteryx*		3					
Family Leuctridae		2					
Genus *Leuctra*		2	– 0.52		1.00	2E+20	1.00
Order Ephemeroptera							
Family Ephemeridae		2					
Genus *Ephemera*		2	– 0.56		0.38		
Family Ephemerellidae		3					
Genus *Serratella*		3	– 0.69		0.62		
Genus *Timpanoga*		3	– 0.84	2.8	0.08		
Genus *Drunella*		3	– 0.80		1.00	1.24	0.07
Genus *Ephemerella*		2	– 0.38	1.44	0.92	3.84	0.87
Genus *Eurylophella*		3	– 0.19		0.15		0.20
Family Ameletidae		4					
Genus *Ameletus*		4	– 0.77		0.38	1.80	0.67
Family Leptophlebiidae		3					
Genus *Paraleptophlebia*		3	– 0.75		0.62	2.79	0.47
Genus *Habrophlebia*		2	– 0.15		0.08		0.53
Genus *Habrophlebiodes*		4	– 0.64		0.69		0.33
Genus *Leptophlebia*		4					

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Baetidae							
Genus *Baetis*		3	−0.55	3E+23	1.00		0.80
Genus *Callibaetis*		4					
Genus *Centroptilum*		2	−0.67		0.08		
Family Heptageneidae		4					
Genus *Stenonema*		4	−0.53		0.85	1.26	0.53
Genus *Stenacron*		4	−0.40				0.20
Genus *Epeorus*		4	−0.84		1.00	9.19	0.93
Genus *Cinygmula*		2	−0.58	2.09	0.62		0.13
Genus *Leucrocuta*		3	−0.49		0.38		0.40
Genus *Heptagena*		4	−0.50		0.15		
Genus *Rhithrogena*		4					
Family Isonychiidae	1	2					
Genus *Isonychia*	1	2	0.42		0.08		
Order Odonata		5					
Family Gomphidae		5					
Genus *Stylogomphus*		3	−0.67		0.08		
Genus *Lanthus*		4	−0.87	2.72	0.77	1.69	0.33
Genus *Progomphus*		4					
Genus *Gomphus*		5					
Family Aeshnidae		6					
Genus *Boyeria*		6	−0.67		0.08		
Family Calopterygidae		5					
Genus *Calopteryx*		5					
Family Coenagrionidae		5	−0.16	1.95	0.08		
Genus *Argia*		5					
Order Hemiptera							
Family Veliidae		7					
Genus *Microvelia*		7	8.67				0.07
Genus *Rhagovelia*		7					
Order Megaloptera							
Family Corydalidae		5					
Genus *Corydalus*		6	−0.07	1E+46	0.08		
Genus *Nigronia*		3	−0.09		1.00	4.06	0.73
Order Trichoptera							
Family Hydroptilidae		4					
Genus *Hydroptila*		5	1.27	2.80	0.38	1.24	0.07
Family Helicopsychidae		3					
Genus *Helicopsyche*		3					
Family Hydropsychidae	1	4					
Genus *Hydropsyche*	1	4	−0.39		1.00		0.13
Genus *Cheumatopsyche*	1	6	−0.29		0.46	1.24	0.07
Genus *Diplectrona*	1	4	−0.41	2.99	1.00	4.42	0.87
Family Rhyacophilidae		3					
Genus *Rhyacophila*		3	−0.60	2.99	1.00	2.50	0.67

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Philopotamidae		3					
Genus *Chimarra*		4					
Genus *Wormaldia*		1	– 0.55	69.20	0.08		0.20
Genus *Dolophilodes*		3	– 0.85		0.92	1.76	0.60
Family Psychomyiidae		2					
Genus *Lype*		3					
Genus *Psychomyia*		2					
Family Leptoceridae		4					
Genus *Triaenodes*		4	0.61			1.22	0.07
Family Odontoceridae							
Genus *Psilotreta*		1	1.18	2.13	0.46		
Family Brachycentridae	1						
Genus *Micrasema*	1	4	– 0.30	2.73	0.23	1.24	0.07
Genus *Brachycentrus*	1	4	– 0.67	12.68	0.23		
Genus *Adicrophleps*							
Family Lepidostomatidae							
Genus *Lepidostoma*		3	– 0.38		0.54		0.60
Family Glossosomatidae							
Genus *Glossosoma*		3	– 0.97		0.62		0.33
Genus *Agapetus*		3	1.51		0.69		
Family Limnephilidae		3					
Genus *Pycnopsyche*		5	0.67		0.08	38.55	0.07
Family Goeridae							
Genus *Goera*		1	– 1.00		0.08		
Family Ueniodae	1						
Genus *Neophylax*	1	3	– 0.31	31.98	1.00		0.40
Family Polycentropodidae	1						
Genus *Neureclipsis*	1	5	– 0.68	2.80	0.08		
Genus *Nyctiophylax*	1	4	– 0.72		0.08		0.07
Genus *Polycentropus*	1	5	– 0.68	1.68	0.69	2.45	0.87
Family Molannidae							
Genus *Molanna*							
Order Coleoptera							
Family Psephenidae							
Genus *Psephenus*		5	– 0.26		0.92		0.40
Genus *Ectopria*		3	– 0.08		0.23	38.55	0.07
Family Dryopidae							
Genus *Helichus*		6	0.50		0.23	1.24	0.07
Family Elmidae							
Genus *Stenelmis*		6	2.56	5.83	0.46		0.13
Genus *Optioservus*		4	0.72	2.06	0.77		0.20
Genus *Promoresia*		3	– 0.75		0.62	1.11	0.07
Genus *Oulimnius*		3	– 0.32	31.98	1.00	2.85	0.87
Genus *Gonielmas*		4	0.67			1.24	0.07
Family Ptilodactylidae							
Genus *Anchytarsus*		5	0.60	2.80	0.08		
Order Diptera							

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Blephariceridae							
Genus *Blepharicera*		4	– 0.50	3.36	0.77		0.07
Family Tipulidae							
Genus *Tipula*		6	– 0.50		0.15		0.13
Genus *Antocha*		4	– 0.81	2.99	1.00		0.04
Genus *Dicranota*		5	0.92		0.77		0.20
Genus *Hexatoma*		5	– 0.29	3.03	0.85	10.18	0.87
Genus *Pilaria*		4	– 0.33				0.07
Genus *Ormosia*		5	– 0.43		0.08		
Genus *Erioptera*		3	– 0.50		0.08		
Family Psychodidae							
Family Dixidae							
Genus *Dixa*		6	– 0.51		0.15		0.13
Family Simulidae	1						
Genus *Prosimulium*	1	5	– 0.93	4.76	0.38	13.31	0.40
Genus *Simulium*	1	5	1.66	68.2	0.92	5.24	0.80
Family Chironomidae		6	0.43		1.00		1.00
Family Ceratopogoniidae		6	1.45	7E+170	0.92	3.85	0.53
Family Tabanidae		6	0.00				0.07
Family Athericidae							
Genus *Atherix*		4	0.28	2.80	0.08		
Family Empididae							
Genus *Hemerodromia*		6	0.50	4.07	0.38		0.53
Genus *Chelifera*		6	2.37		0.54	38.55	0.07
Genus *Wiedemannia*		6	1.00				0.07
Genus *Clinocera*		6	– 0.83	3.63	0.23		0.13
Genus *Oreogeton*		6	0.71			1.24	0.07
Non-Insect Taxa							
Class Arachnida							
Order Hydracarina							
Class Gastropoda							
Order Mesogastropoda							
Family Pleuroceridae		5	9.88	2.72	0.46		
Class Bivalvia	1						
Order Veneroida							
Famlly Sphaeriidae	1	8	0.00		0.15		
Class Turbellaria							
Order Tricladida							
Family Planariidae		1	0.00	4E+23	0.15		0.20
Class Oligochaeta		8	0.42	6E+6	0.85	1E+33	0.60
Class Crustacea							
Order Amphipoda							
Family Gammaridae		6	1.50				0.47
Order Decapoda							
Family Cambaridae							
Genus *Cambarus*					0.15		0.60

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Aeshnidae		6					
Genus *Boyeria*		6	– 0.67		0.08		
Family Calopterygidae		5					
Genus *Calopteryx*		5					
Family Coenagrionidae		5	– 0.16	1.95	0.08		
Genus *Argia*		5					
Order Hemiptera							
Family Veliidae		7					
Genus *Microvelia*		7	8.67				0.07
Genus *Rhagovelia*		7					
Order Megaloptera							
Family Corydalidae		5					
Genus *Corydalus*		6	– 0.07	1E+46	0.08		
Genus *Nigronia*		3	– 0.09		1.00	4.06	0.73
Order Trichoptera							
Family Hydroptilidae		4					
Genus *Hydroptila*		5	1.27	2.80	0.38	1.24	0.07
Family Helicopsychidae		3					
Genus *Helicopsyche*		3					
Family Hydropsychidae	1	4					
Genus *Hydropsyche*	1	4	– 0.39		1.00		0.13
Genus *Cheumatopsyche*	1	6	– 0.29		0.46	1.24	0.07
Genus *Diplectrona*	1	4	– 0.41	2.99	1.00	4.42	0.87
Family Rhyacophilidae		3					
Genus *Rhyacophila*		3	– 0.60	2.99	1.00	2.50	0.67
Family Philopotamidae		3					
Genus *Chimarra*		4					
Genus *Wormaldia*		1	– 0.55	69.20	0.08		0.20
Genus *Dolophilodes*		3	– 0.85		0.92	1.76	0.60
Family Psychomyiidae		2					
Genus *Lype*		3					
Genus *Psychomyia*		2					
Family Leptoceridae		4					
Genus *Triaenodes*		4	0.61			1.22	0.07
Family Odontoceridae							
Genus *Psilotreta*		1	1.18	2.13	0.46		
Family Brachycentridae	1						
Genus *Micrasema*	1	4	– 0.30	2.73	0.23	1.24	0.07
Genus *Brachycentrus*	1	4	– 0.67	12.68	0.23		
Genus *Adicrophleps*							
Family Lepidostomatidae							
Genus *Lepidostoma*		3	– 0.38		0.54		0.60
Family Glossosomatidae							
Genus *Glossosoma*		3	– 0.97		0.62		0.33
Genus *Agapetus*		3	1.51		0.69		
Family Limnephilidae		3					
Genus *Pycnopsyche*		5	0.67		0.08	38.55	0.07

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Goeridae							
Genus *Goera*		1	– 1.00		0.08		
Family Ueniodae	1						
Genus *Neophylax*	1	3	– 0.31	31.98	1.00		0.40
Family Polycentropodidae	1						
Genus *Neureclipsis*	1	5	– 0.68	2.80	0.08		
Genus *Nyctiophylax*	1	4	– 0.72		0.08		0.07
Genus *Polycentropus*	1	5	– 0.68	1.68	0.69	2.45	0.87
Family Molannidae							
Genus *Molanna*							
Order Coleoptera							
Family Psephenidae							
Genus *Psephenus*		5	– 0.26		0.92		0.40
Genus *Ectopria*		3	– 0.08		0.23	38.55	0.07
Family Dryopidae							
Genus *Helichus*		6	0.50		0.23	1.24	0.07
Family Elmidae							
Genus *Stenelmis*		6	2.56	5.83	0.46		0.13
Genus *Optioservus*		4	0.72	2.06	0.77		0.20
Genus *Promoresia*		3	– 0.75		0.62	1.11	0.07
Genus *Oulimnius*		3	– 0.32	31.98	1.00	2.85	0.87
Genus *Gonielmas*		4	0.67			1.24	0.07
Family Ptilodactylidae							
Genus *Anchytarsus*		5	0.60	2.80	0.08		
Order Diptera							
Family Blephariceridae							
Genus *Blephariceria*		4	– 0.50	3.36	0.77		0.07
Family Tipulidae							
Genus *Tipula*		6	– 0.50		0.15		0.13
Genus *Antocha*		4	– 0.81	2.99	1.00		0.04
Genus *Dicranota*		5	0.92		0.77		0.20
Genus *Hexatoma*		5	– 0.29	3.03	0.85	10.18	0.87
Genus *Pilaria*		4	– 0.33				0.07
Genus *Ormosia*		5	– 0.43		0.08		
Genus *Erioptera*		3	– 0.50		0.08		
Family Psychodidae							
Family Dixidae							
Genus *Dixa*		6	– 0.51		0.15		0.13
Family Simulidae	1						
Genus *Prosimulium*	1	5	– 0.93	4.76	0.38	13.31	0.40
Genus *Simulium*	1	5	1.66	68.2	0.92	5.24	0.80
Family Chironomidae		6	0.43		1.00		1.00
Family Ceratopogoniidae		6	1.45	7E+170	0.92	3.85	0.53
Family Tabanidae		6	0.00				0.07
Family Athericidae							
Genus *Atherix*		4	0.28	2.80	0.08		

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Empididae							
Genus *Hemerodromia*		6	0.50	4.07	0.38		0.53
Genus *Chelifera*		6	2.37		0.54	38.55	0.07
Genus *Wiedemannia*		6	1.00				0.07
Genus *Clinocera*		6	– 0.83	3.63	0.23		0.13
Genus *Oreogeton*		6	0.71			1.24	0.07
Non-Insect Taxa							
Class Arachnida							
Order Hydracarina							
Class Gastropoda							
Order Mesogastropoda							
Family Pleuroceridae		5	9.88	2.72	0.46		
Class Bivalvia	1						
Order Veneroida							
Famlly Sphaeriidae	1	8	0.00		0.15		
Class Turbellaria							
Order Tricladida							
Family Planariidae		1	0.00	4E+23	0.15		0.20
Class Oligochaeta		8	0.42	6E+6	0.85	1E+33	0.60
Class Crustacea							
Order Amphipoda							
Family Gammaridae		6	1.50				0.47
Order Decapoda							
Family Cambaridae							
Genus *Cambarus*					0.15		0.60
Family Limnephilidae		3					
Genus *Pycnopsyche*		5	0.67		0.08	38.55	0.07
Family Goeridae							
Genus *Goera*		1	– 1.00		0.08		
Family Ueniodae	1						
Genus *Neophylax*	1	3	– 0.31	31.98	1.00		0.40
Family Polycentropodidae	1						
Genus *Neureclipsis*	1	5	– 0.68	2.80	0.08		
Genus *Nyctiophylax*	1	4	– 0.72		0.08		0.07
Genus *Polycentropus*	1	5	– 0.68	1.68	0.69	2.45	0.87
Family Molannidae							
Genus *Molanna*							
Order Coleoptera							
Family Psephenidae							
Genus *Psephenus*		5	– 0.26		0.92		0.40
Genus *Ectopria*		3	– 0.08		0.23	38.55	0.07
Family Dryopidae							
Genus *Helichus*		6	0.50		0.23	1.24	0.07
Family Elmidae							
Genus *Stenelmis*		6	2.56	5.83	0.46		0.13
Genus *Optioservus*		4	0.72	2.06	0.77		0.20
Genus *Promoresia*		3	– 0.75		0.62	1.11	0.07
Genus *Oulimnius*		3	– 0.32	31.98	1.00	2.85	0.87
Genus *Gonielmas*		4	0.67			1.24	0.07

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Ptilodactylidae							
Genus *Anchytarsus*		5	0.60	2.80	0.08		
Order Diptera							
Family Blephariceridae							
Genus *Blepharicera*		4	– 0.50	3.36	0.77		0.07
Family Tipulidae							
Genus *Tipula*		6	– 0.50		0.15		0.13
Genus *Antocha*		4	– 0.81	2.99	1.00		0.04
Genus *Dicranota*		5	0.92		0.77		0.20
Genus *Hexatoma*		5	– 0.29	3.03	0.85	10.18	0.87
Genus *Pilaria*		4	– 0.33				0.07
Genus *Ormosia*		5	– 0.43		0.08		
Genus *Erioptera*		3	– 0.50		0.08		
Family Psychodidae							
Family Dixidae							
Genus *Dixa*		6	– 0.51		0.15		0.13
Family Simulidae	1						
Genus *Prosimulium*	1	5	– 0.93	4.76	0.38	13.31	0.40
Genus *Simulium*	1	5	1.66	68.2	0.92	5.24	0.80
Family Chironomidae		6	0.43		1.00		1.00
Family Ceratopogoniidae		6	1.45	7E+170	0.92	3.85	0.53
Family Tabanidae		6	0.00				0.07
Family Athericidae							
Genus *Atherix*		4	0.28	2.80	0.08		
Family Empididae							
Genus *Hemerodromia*		6	0.50	4.07	0.38		0.53
Genus *Chelifera*		6	2.37		0.54	38.55	0.07
Genus *Wiedemannia*		6	1.00				0.07
Genus *Clinocera*		6	– 0.83	3.63	0.23		0.13
Genus *Oreogeton*		6	0.71			1.24	0.07
Non-Insect Taxa							
Class Arachnida							
Order Hydracarina							
Class Gastropoda							
Order Mesogastropoda							
Family Pleuroceridae		5	9.88	2.72	0.46		
Class Bivalvia	1						
Order Veneroida							
Famlly Sphaeriidae	1	8	0.00		0.15		
Class Turbellaria							
Order Tricladida							
Family Planariidae		1	0.00	4E+23	0.15		0.20
Class Oligochaeta		8	0.42	6E+6	0.85	1E+33	0.60
Class Crustacea							
Order Amphipoda							
Family Gammaridae		6	1.50				0.47
Order Decapoda							

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Cambaridae							
Genus *Cambarus*					0.15		0.60
Family Aeshnidae		6					
Genus *Boyeria*		6	− 0.67		0.08		
Family Calopterygidae		5					
Genus *Calopteryx*		5					
Family Coenagrionidae		5	− 0.16	1.95	0.08		
Genus *Argia*		5					
Order Hemiptera							
Family Veliidae		7					
Genus *Microvelia*		7	8.67				0.07
Genus *Rhagovelia*		7					
Order Megaloptera							
Family Corydalidae		5					
Genus *Corydalus*		6	− 0.07	1E+46	0.08		
Genus *Nigronia*		3	− 0.09		1.00	4.06	0.73
Order Trichoptera							
Family Hydroptilidae		4					
Genus *Hydroptila*		5	1.27	2.80	0.38	1.24	0.07
Family Helicopsychidae		3					
Genus *Helicopsyche*		3					
Family Hydropsychidae	1	4					
Genus *Hydropsyche*	1	4	− 0.39		1.00		0.13
Genus *Cheumatopsyche*	1	6	− 0.29		0.46	1.24	0.07
Genus *Diplectrona*	1	4	− 0.41	2.99	1.00	4.42	0.87
Family Rhyacophilidae		3					
Genus *Rhyacophila*		3	− 0.60	2.99	1.00	2.50	0.67
Family Philopotamidae		3					
Genus *Chimarra*		4					
Genus *Wormaldia*		1	− 0.55	69.20	0.08		0.20
Genus *Dolophilodes*		3	− 0.85		0.92	1.76	0.60
Family Psychomyiidae		2					
Genus *Lype*		3					
Genus *Psychomyia*		2					
Family Leptoceridae		4					
Genus *Triaenodes*		4	0.61			1.22	0.07
Family Odontoceridae							
Genus *Psilotreta*		1	1.18	2.13	0.46		
Family Brachycentridae	1						
Genus *Micrasema*	1	4	− 0.30	2.73	0.23	1.24	0.07
Genus *Brachycentrus*	1	4	− 0.67	12.68	0.23		
Genus *Adicrophleps*							
Family Lepidostomatidae							
Genus *Lepidostoma*		3	− 0.38		0.54		0.60
Family Glossosomatidae							
Genus *Glossosoma*		3	− 0.97		0.62		0.33
Genus *Agapetus*		3	1.51		0.69		

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Limnephilidae		3					
Genus *Pycnopsyche*		5	0.67		0.08	38.55	0.07
Family Goeridae							
Genus *Goera*		1	− 1.00		0.08		
Family Ueniodae	1						
Genus *Neophylax*	1	3	− 0.31	31.98	1.00		0.40
Family Polycentropodidae	1						
Genus *Neureclipsis*	1	5	− 0.68	2.80	0.08		
Genus *Nyctiophylax*	1	4	− 0.72		0.08		0.07
Genus *Polycentropus*	1	5	− 0.68	1.68	0.69	2.45	0.87
Family Molannidae							
Genus *Molanna*							
Order Coleoptera							
Family Psephenidae							
Genus *Psephenus*		5	− 0.26		0.92		0.40
Genus *Ectopria*		3	− 0.08		0.23	38.55	0.07
Family Dryopidae							
Genus *Helichus*		6	0.50		0.23	1.24	0.07
Family Elmidae							
Genus *Stenelmis*		6	2.56	5.83	0.46		0.13
Genus *Optioservus*		4	0.72	2.06	0.77		0.20
Genus *Promoresia*		3	− 0.75		0.62	1.11	0.07
Genus *Oulimnius*		3	− 0.32	31.98	1.00	2.85	0.87
Genus *Gonielmas*		4	0.67			1.24	0.07
Family Ptilodactylidae							
Genus *Anchytarsus*		5	0.60	2.80	0.08		
Order Diptera							
Family Blephariceridae							
Genus *Blepharicera*		4	− 0.50	3.36	0.77		0.07
Family Tipulidae							
Genus *Tipula*		6	− 0.50		0.15		0.13
Genus *Antocha*		4	− 0.81	2.99	1.00		0.04
Genus *Dicranota*		5	0.92		0.77		0.20
Genus *Hexatoma*		5	− 0.29	3.03	0.85	10.18	0.87
Genus *Pilaria*		4	− 0.33				0.07
Genus *Ormosia*		5	− 0.43		0.08		
Genus *Erioptera*		3	− 0.50		0.08		
Family Psychodidae							
Family Dixidae							
Genus *Dixa*		6	− 0.51		0.15		0.13
Family Simulidae	1						
Genus *Prosimulium*	1	5	− 0.93	4.76	0.38	13.31	0.40
Genus *Simulium*	1	5	1.66	68.2	0.92	5.24	0.80
Family Chironomidae		6	0.43		1.00		1.00
Family Ceratopogoniidae		6	1.45	7E+170	0.92	3.85	0.53
Family Tabanidae		6	0.00				0.07

Taxon	MBII metric characteristics		Dose-response models	Capture probabillity models			
				Hazel River		Paine Rub	
	Filter feeders	PTV	*P* change	Model	Obs.	Model	Obs.
Family Athericidae							
Genus *Atherix*		4	0.28	2.80	0.08		
Family Empididae							
Genus *Hemerodromia*		6	0.50	4.07	0.38		0.53
Genus *Chelifera*		6	2.37		0.54	38.55	0.07
Genus *Wiedemannia*		6	1.00				0.07
Genus *Clinocera*		6	– 0.83	3.63	0.23		0.13
Genus *Oreogeton*		6	0.71			1.24	0.07
Non-Insect Taxa							
Class Arachnida							
Order Hydracarina							
Class Gastropoda							
Order Mesogastropoda							
Family Pleuroceridae		5	9.88	2.72	0.46		
Class Bivalvia	1						
Order Veneroida							
Famlly Sphaeriidae	1	8	0.00		0.15		
Class Turbellaria							
Order Tricladida							
Family Planariidae		1	0.00	4E+23	0.15		0.20
Class Oligochaeta		8	0.42	6E+6	0.85	1E+33	0.60
Class Crustacea							
Order Amphipoda							
Family Gammaridae		6	1.50				0.47
Order Decapoda							
Family Cambaridae							
Genus *Cambarus*					0.15		0.60

References

Andersen, T., J. Carstensen, E. Hernández-García, and C. M. Duarte. 2009. Ecological thresholds and regime shifts: Approaches to identification. *Trends in Ecology and Evolution* 24:49–57.

Bailey, R. C., R. H. Norris, and T. B. Reynoldson. 2001. Taxonomic resolution of benthic macroinvertebrate communities in bioassessments. *Journal of the North American Benthological Society* 20:280–286.

Brown, M. T., and M. B. Vivas. 2005. Landscape development intensity index. *Environmental Monitoring and Assessment* 101:289–309.

Burton, J., and J. Gerritsen. 2003. *A stream condition index for Virginia non-coastal streams*. Washington, D.C.: U. S. Environmental Protection Agency.

Cade, B. S., and B. R. Noon. 2003. A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and Evolution* 1:412–420.

Cao, Y., and C. P. Hawkins. 2005. Simulating biological impairment to evaluate the accuracy of ecological indicators. *Journal of Applied Ecology* 42:954–965.

Capitulo, A. R., M. Tangorra, and C. Ocon. 2001. Use of benthic macroinvertebrates to assess the biological status of Pampean streams in Argentina. *Aquatic Ecology* 35:109–119.

Clarke, R. T., M. T. Furse, R. J. M. Gunn, J. M. Winder, and J. F. Wright. 2002. Sampling variation in macroinvertebrate data and implications for river quality indices. *Freshwater Biology* 47:1735–1751.

Clarke, R. T., A. Lorenz, L. Sandin, A. Schmidt-Kloiber, J. Strackbein, N. T. Kneebone, and P. Haase. 2006. Effects of sampling and sub-sampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. *Hydrobiologia* 566:441–459.

Courtemanch, D. L. 1996. Commentary on the subsampling procedures used for rapid bioassessments. *Journal of the North American Benthological Society* 15:381–385.

Dale, V.H., and S.C. Beyeler 2001. Challenges in the development and use of ecological indicators. *Ecological Indicators* 1:3–10.

Field, S. A., A. J. Tyre, N. Jonzen, J. R. Rhodes, and H. P. Possingham. 2004. Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters* 7:669–675.

Gerritsen, J., J. Burton, and M.T. Barbour 2000. A stream condition index for West Virginia wadeable streams. Tetra Tech, Inc., Owings Mills, Maryland. Prepared for West Virgiia Department of Environmental Protection, Charleston, N.C. http://www.dep.state.WV.US/Docs/536WV-Index.pdf.

Gregory, R. S., and R. L. Keeney. 2002. Making smarter environmental management decisions. *Journal of the American Water Resources Association* 38:1601–1612.

Growns, J. E., B. C. Chessman, J. E. Jackson, and D. G. Ross. 1997. Rapid assessment of Australian rivers using macroinvertebrates: Cost and efficiency of six methods of sample processing. *Journal of the North American Benthological Society* 16:682–693.

Hauer, F. R., and V. H. Resh. 2006. Macroinvertebrates. In *Methods in stream ecology*, eds F. R. Hauer and G. A. Lambert, 435–463. San Diego: Elsevier.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10:1456–1477.

Hawkins, C. P., Y. Cao, and B. Roper. 2010. Methods of predicting reference condition biota affects the performance and interpretation of ecological indices. *Freshwater Biology* 55:1066–1085.

Hering, D., O. Moog, L. Sandin, and P. F. Verdonschot. 2004. Overview and application of the AQEM assessment system. *Hydrobiologia* 516:1–20.

Herlihy, A. T., D. P. Larsen, S. G. Paulsen, N. S. Urquhart, and B. J. Rosenbaum. 2000. Designing a spatially bananced, randomized site selection process for regional stream surveys: The EMAP mid-Atlantic pilot study. *Environmental Monitoring and Assessment* 63:95–113.

Hitt, N. P., and P. L. Angermeier. 2008. River-stream connectivity affects fish bioassessment performance. *Environmental Management* 42:132–150.

Keeney, R. L. 1992. *Value-focused thinking, a path to creative decision-making*. Cambridge: Harvard University Press.

Kennen, J. G., K. Rvia-Murray, and K. M. Beaulieu. 2010. Determining hydrologic factors that influence stream macroinvertebrate assemblages in the northeastern US. *Ecohydrology* 3:88–106.

Kery, M., J. A. Royle, M. Plattner, and R. M. Dorazio. 2009. Species richness and occupancy estimation in communities subject to temporal emigration. *Ecology* 90:1279–1290.

King, R. S., and C. J. Richardson. 2002. Evaluating subsampling approaches and macroinvertebrate taxonomic resolution for wetland bioassessment. *Journal of the North American Benthological Society* 21:150–171.

Klemm, D. J., K. A. Blocksom, W. T. Thoeny, F. A. Fulk, A. T. Herlihy, P. R. Kaufmann, and S. M. Cormier. 2002. Methods development and use of macroinvertebrates as indicators of ecological conditions for streams in the mid-Atlantic Highlands region. *Environmental Monitoring and Assessment* 78:169–212.

Klemm, D. J., K. A. Blocksom, F. A. Fulk, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, W. T. Thoeny, M. B. Griffith, and W. S. Davis. 2003. Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing mid-Atlantic highlands streams. *Environmental Management* 31:656–669.

Koenker, R. 2010. quantreg: Quantile regression. R package version 4.50. http://CRAN.R-project.org/package=quantreg. Last accessed on July 11, 2011.

Koenker, R., and K. F. Hallock. 2001. Quantile regression. *Journal of Economic Perspectives* 15:143–156.

Lane, C. R., and M. T. Brown. 2006. Energy-based land use predictors of proximal factors and benthic diatom composition in Florida freshwater marshes. *Environmental Monitoring and Assessment* 117:433–450.

Lazorchak, J. M., D. J. Klemm, and D. V. Peck, eds. 1998. *Environmental monitoring and assessment program—surface waters: Field operations and methods for measuring ecological condition of wadeable streams. EPA 620/R-94/004F*. Cincinnati: U.S. Environmental Protection Agency.

Lenat, D. R. 1993. A biotic index for the southeastern United States: Derivation and list of tolerance values, with criteria for assigning water quality ratings. *Journal of the North American Benthological Society* 12:279–290.

Lenat, D. R., and V. H. Resh. 2001. Taxonomy and stream ecology—the benefits of genus- and species-level identifications. *Journal of the North American Benthological Society* 20:287–298.

Lynch, D. D. 1987. *Hydrologic conditions and trends in Shenandoah National Park, Virginia, 1983–1984. Water Resources Investigation Report 87–4131*. Richmond: US Geological Survey.

Mack, J. J. 2006. Landscape as a predictor of wetland condition: An evaluation of the landscape development intensity index (LDI) with a large reference wetland dataset from Ohio. *Environmental Monitoring and Assessment* 120:221–241.

MacKenzie, D. I., and W. L. Kendall. 2002. How should detection probability be incorporated into estimates of relative abundance? *Ecology* 83:2387–2397.

Manly, B. F. 2008. *Statistics for environmental science and management*. 2nd ed. Boca Raton: Chapman and Hall/CRC.

Marchant, R. 2007. The use of taxonomic distinctness to assess environmental disturbance of insect communities from running water. *Freshwater Biology* 52:1634–1645.

Marshall, M., and B. Piekielek. 2007. *Eastern Rivers and Mountains Network ecological monitoring plan. Natural Resource Report NPS/ERMN/NRR-2007/017*.

Marshall, J. C., A. L. Steward, and B. D. Harch. 2006. Taxonomic resolution and quantification of freshwater macroinvertebrate samples from an Australian dryland river: The benefits and costs of using abundance data. *Hydrobiologia* 572:171–194.

Martin, J., M. C. Runge, J. D. Nichols, B. C. Lubow, and W. L. Kendall. 2009. Structured decision making as a conceptual framework to identify thresholds for conservation management. *Ecological Applications* 19:1079–1090.

Mazor, R. D., T. B. Reynoldson, D. M. Rosenberg, and V. H. Resh. 2006. Effects of biotic assemblage, classification, and assessment method on bioassessment performance. *Canadian Journal of Fisheries and Aquatic Sciences* 63:394–411.

McCarrondand, E., and R. Frydenburg. 2005. Using the human disturbance gradient to develop bioassessment procedures in estuaries. Pages 481–492 In *Estuarine indicators*, ed S. A. Bortone, 481–492. Boca Raton: CRC Press.

McCune, B., and J. B. Grace. 2002. *Analysis of ecological communities*. Geneden Beach: MJM Software Design.

McElravy, E. P., G. A. Lamberti, and V. H. Resh. 1989. Year-to-year variation in the aquatic macroinvertebrate fauna of a northern California stream. *Journal of the North American Benthological Society* 8:51–63.

Merritt, R. W., and K. W. Cummins. 1996. An introduction to the aquatic insects of North America. 3rd ed. Dubuque: Kendall/Hunt Publishing Co.

Moeykens, M. D., and J. R. Voshell, Jr. 2002. *Studies of benthic macroinvertebrates for the Shenandoah National Park Long-term monitoring systems: Statistical analysis of LTEMS Aquatic*

Dataset from 1986 to 2000 on water chemistry, habitat, and macroinvertebrates. Virginia Technical Resport No. 208–11-110A-007–374-1. Blacksburg: Virginia Tech University.

Moulton, S. R., II, J. G. Kennen, R. M. Goldstein, and J. A. Hambrook. 2002. Revised protocols for sampling algae, invertebrate, and fish communities in the National Water-Quality Assessment Program. U.S. Geological Survey Open-File Report 02–150.

Morse, J. C., Y. J. Bae, G. Munkhjargal, N. Sangpradub, K. Tanida, T. S. Vshivhova, B. Wang, L. Yang, and C. M. Yule. 2007. Freshwater biomonitoring with macroinvertebrates in east Asia. *Frontiers in Ecology and the Environment* 5:33–42.

New, T. R. 1996. Taxonomic focus and quality control in insect surveys for biodiversity conservation. *Australian Journal of Entomology* 35:97–106.

Nichols, J. D., and B. K. Williams. 2006. Monitoring for conservation. *Trends in Ecology and Evolution* 21:668–673.

Nichols, S. J., W. A. Robinson, and R. H. Norris. 2006. Sample variability influences on the precision of predictive bioassessment. *Hydrobiologia* 572: 215–233.

Odum, H. T. 1988. Self-organization, transformity, and information. *Science* 242:1132–1139.

Odum, H. T. 1996. *Environmental accounting: Energy and decision making*. New York: Wiley.

Ollis, D. J., H. F. Dallas, K. J. Esler, and C. Boucher. 2006. Bioassessment of the ecological integrity of river ecosystems using aquatic macroinvertebrates: An overview with a focus on South Africa. *African Journal of Aquatic Sciences* 31:205–227.

Ostermiller, J. D., and C. P. Hawkins. 2004. Effects of sampling error on bioassessments of stream ecosystems: Application to RIVPACS-type models. *Journal of the North American Benthological Society* 23:363–382.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. *Rapid bioassessment protocols for use in streams and rivers, benthic macroinvertebrates and fish*. Washington, D.C.: U. S. Environmental Protection Agency.

Pond, G.J., M.E. Passmore, F.A. Borsuk, L. Reynolds, and C.J. Rose. 2008. Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. Journal of the North American Benthological Society 27:717–737.

Purcell, A. H., D. W. Bressler, M. J. Paul, M. T. Barbour, E. T. Rankin, J. L. Carter, and V. H. Resh. 2009. Assessment tools for urban catchments: Developing biological indicators based on benthic macroinvertebrates. *Journal of the American Water Resources Association* 45:306–319.

Qian, S. S., R. S. King, and C. J. Richardson. 2003. Two statistical methods for the detection of environmental thresholds. *Ecological Modelling* 166:87–97.

Quinn, J. M., and C. W. Hickey. 1990. *The magnitude of the effects of substrate particle size, recent flooding, and catchment development on benthic invertebrates in 88 New Zealand rivers. New Zealand Journal of Marine and Freshwaters* 24:411–427.

Reynoldson, T. B., R. H., Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. 1997. The reference condition: A comparison of multimetric and multivariate approaches to assess water quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society* 16:833–852.

Roy, A. H., A. D. Rosemond, M. J. Paul, D. S. Leigh, and J. B. Wallace. 2003. Stream macroinvertebrate response to catchment urbanization (Georgia, USA). *Freshwater Biology* 48:329–346.

Scharf, F. S., F. Juanes, and M. Sutherland. 1998. Inferring ecological relationships from the edges of scatter diagrams: Comparison of regression techniques. *Ecology* 79: 448–460.

Snyder, C. D., J. A. Young, R. Villella, and D. P. Lemarie. 2003. Influences of upland and riparian land use patterns on stream biotic integrity. *Landscape Ecology* 18:647–664.

Sonderegger, D. L., H. Wang, W. H. Clements, and B. R. Noon. 2009. Using SiZer to detect thresholds in ecological data. *Frontiers in Ecology and the Environment* 7:190–195.

Sovell, L. A., and B. Vondracek. 1999. Evaluation of the fixed-count method for Rapid Bioassessment Protocol III with benthic macroinvertebrate metrics. *Journal of the North American Benthological Society* 18:420–426.

Summers, J. K., and K. E. Tonnessen. 1998. Linking monitoring and effects research: EMAP's intensive site network program. *Environmental Monitoring and Assessment* 51:369–380.

Van Sickle, J., C. P. Hawkins, D. P. Larsen, and A. T. Herlihy. 2005. A null model for the expected macroinvertebrate assemblage in streams. *Journal of the North American Benthological Society* 24:178–191.

Van Sickle, J., D. P. Larsen, and C. P. Hawkins. 2007. Exclusion of rare taxa affects performance of the O/E index in bioassessments. *Journal of the North American Benthological Society* 26:319–331.

Vinson, M. R., and C. P. Hawkins. 1996. Effects of sampling area and subsampling procedures on comparisons of taxa richness among streams. *Journal of the North American Benthological Society* 15:392–399.

Wagner, R., T. Dapper, and H. H. Schmidt. 2000. The influence of environmental variables on the abundance of aquatic insects: A comparison of ordination and artificial neural networks. *Hydrobiologia* 422/423:143–152

Waite, I. R., A. t. Herlihy, D. P. Larsen, and D. J. Klemm. 2000. Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the mid-Atlantic Highlands, USA. *Journal of the North American Benthological Society* 19:429–441.

Waite, I. R., A. T. Herlihy, D. P. Larsen, N. S. Urquhart, and D. J. Klemm. 2004. The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: An example from the mid-Atlantic Highlands, U.S.A. *Freshwater Biology* 49:474–489.

Wang, L., J. Lyons, P. Kanehl, and R. Gatti. 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. *Fisheries* 22:6–12.

Yoder, C. O., and E. T. Rankin. 1998. The role of biological indicators in a state water quality management process. *Environmental Monitoring and Assessment* 51:61–88.

Young, J. A., J., G. Fleming, W. Cass, and C. Lea. 2009. *Vegetation of Shenandoah National Park in Relation to Environmental Gradients, Version 2.0. Technical Report NPS/NER/NRTR—2009/142*. Philadelphia: National Park Service.

Yuan, L. L. 2004. Assigning macroinvertebrate tolerance classifications using generalized additive models. *Freshwater Biology* 49:662–677.

Chapter 10
Getting the Message Across: Using Ecological Integrity to Communicate with Resource Managers

Brian R. Mitchell, Geraldine L. Tierney, E. William Schweiger, Kathryn M. Miller, Don Faber-Langendoen and James B. Grace

Abstract This chapter describes and illustrates how concepts of ecological integrity, thresholds, and reference conditions can be integrated into a research and monitoring framework for natural resource management. Ecological integrity has been defined as a measure of the composition, structure, and function of an ecosystem in relation to the system's natural or historical range of variation, as well as perturbations caused by natural or anthropogenic agents of change. Using ecological integrity to communicate with managers requires five steps, often implemented iteratively: (1) document the scale of the project and the current conceptual understanding and reference conditions of the ecosystem, (2) select appropriate metrics representing integrity, (3) define externally verified assessment points (metric values that signify an ecological change or need for management action) for the metrics, (4) collect data and calculate metric scores, and (5) summarize the status of the ecosystem using a variety of reporting methods. While we present the steps linearly for conceptual clarity, actual implementation of this approach may require addressing the steps in a

---

B. R. Mitchell (✉)
Northeast Temperate Network, National Park Service, 54 Elm Street, Woodstock, VT, USA
e-mail: brian_mitchell@nps.gov

G. L. Tierney
Department of Environmental and Forest Biology, SUNY College of Environmental Science and Forestry, 1 Forestry Drive Syracuse, NY, USA

E. W. Schweiger
Rocky Mountain Network, National Park Service, 1201 Oakridge Drive,
Fort Collins, CO, USA

K. M. Miller
Northeast Temperate Network, National Park Service, Acadia National Park,
20 McFarland Hill Drive, Bar Harbor, ME, USA

D. Faber-Langendoen
Conservation Science Division, NatureServe, 4600 N. Fairfax Drive,
7th Floor, Arlington, VA, USA

J. B. Grace
National Wetlands Research Center, U.S. Geological Survey,
700 Cajundome Blvd., Lafayette, LA, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_10,

different order or revisiting steps (such as metric selection) multiple times as data are collected. Knowledge of relevant ecological thresholds is important when metrics are selected, because thresholds identify where small changes in an environmental driver produce large responses in the ecosystem. Metrics with thresholds at or just beyond the limits of a system's range of natural variability can be excellent, since moving beyond the normal range produces a marked change in their values. Alternatively, metrics with thresholds within but near the edge of the range of natural variability can serve as harbingers of potential change. Identifying thresholds also contributes to decisions about selection of assessment points. In particular, if there is a significant resistance to perturbation in an ecosystem, with threshold behavior not occurring until well beyond the historical range of variation, this may provide a scientific basis for shifting an ecological assessment point beyond the historical range. We present two case studies using ongoing monitoring by the US National Park Service Vital Signs program that illustrate the use of an ecological integrity approach to communicate ecosystem status to resource managers. The Wetland Ecological Integrity in Rocky Mountain National Park case study uses an analytical approach that specifically incorporates threshold detection into the process of establishing assessment points. The Forest Ecological Integrity of Northeastern National Parks case study describes a method for reporting ecological integrity to resource managers and other decision makers. We believe our approach has the potential for wide applicability for natural resource management.

Keywords Assessment point · Communication tool · Conceptual diagram · Condition metric · Ecological integrity · Ecological threshold · Forest · Index of biological integrity · Natural variability · Wetland

Introduction

Ecological thresholds have been defined in many ways, including the commonly used definition from Groffman et al. (2006): "an ecosystem quality, property or phenomenon … where small changes in an environmental driver produce large responses in the ecosystem." As scientists tasked with monitoring long-term trends in natural resource conditions, we are keenly interested in using multiple methods to detect important thresholds, be they strict ecological thresholds as defined by Groffman et al. (2006), or simply a point along a continuum that reflects a shift to an undesirable state. As communicators who are required to convey complex results to decision makers and the public, we need a simple, flexible framework suitable for reporting data analyses in a way that can be easily understood and applied. The goal of this chapter is to provide you with an approach, based on the concept of ecological integrity, that incorporates threshold ideas and reference conditions and is broadly applicable for presenting research and monitoring results to decision makers. We present two case studies using ongoing monitoring by the US National Park Service Vital Signs program (Fancy et al. 2009) that illustrate the use of an ecological integrity approach to communicate ecosystem status to resource managers. Each example has different objectives and a different emphasis in order to demonstrate some of the range of applications of the general approach.

"Ecological integrity" builds on the related concepts of biological integrity and ecological health, and is a useful endpoint for ecological assessment and reporting (Czech 2004). "Integrity" is defined as the quality of being unimpaired, sound, or complete. To have integrity, an ecosystem should be relatively unimpaired across a range of characteristics, and across spatial and temporal scales (De Leo and Levin 1997). Ecological integrity has been defined as a measure of the composition, structure, and function of an ecosystem in relation to the system's natural or historical range of variation, as well as perturbations caused by natural or anthropogenic agents of change (Karr and Dudley 1981). An ecological system has integrity "when its dominant ecological characteristics (e.g., elements of composition, structure, function, and ecological processes) occur within their natural ranges of variation and can withstand and recover from most perturbations imposed by natural environmental dynamics or human disruptions" (Parrish et al. 2003).

As Tierney et al. (2009) describe, ecological integrity can be difficult to assess. One approach builds on the Index of Biological Integrity (IBI), which was originally used to interpret stream integrity based on 12 metrics that reflected the condition, reproduction, composition, and abundance of fish species (Karr 1981). Each metric was rated by comparing measured values with the values expected under relatively unimpaired conditions, and the ratings were aggregated into a total score. Related biotic indices have sought to assess the integrity of other aquatic and wetland ecosystems, primarily via faunal (and more recently, floral) assemblages. Building upon this foundation, others have suggested measuring the integrity of ecosystems by developing suites of indicators or metrics comprising the key biological, physical, and functional attributes of those ecosystems (Andreasen et al. 2001; Parrish et al. 2003; Mack and Kentula 2010).

For the purpose of communicating information about ecosystem condition to managers, ecological integrity can be summarized as one or more metrics of ecosystem composition, structure, and function. The acceptable ranges of these metrics are established through knowledge of their natural variability at defined spatial and temporal scales and their resistance to perturbation (Tierney et al. 2009). In some cases, extensive data sets and prior research are available to determine the natural range of variation; in other cases, an initial period of baseline data collection or expert judgment can be used to establish the acceptable ranges. Regardless of the specifics of how these ranges are developed, attention to potential ecological thresholds is important. Managers are particularly concerned about nonlinear effects near thresholds that produce outsized impacts on resource condition or shift ecosystems into new and unnatural stable states (Groffman et al. 2006). An example of such a dramatic shift in an ecosystem's state (cited in Groffman et al. 2006) is Florida Bay, which in the 1990s abruptly shifted from an oligotrophic clear water system dominated by seagrasses to a turbid system dominated by phytoplankton blooms. Knowledge of the strength and location of thresholds like the one that led to the ecological shift in Florida Bay allows scientists and managers to develop precautionary "assessment points" for metrics that will trigger action before the threshold is reached (Bennetts et al. 2007). In this chapter, we use the term "ecological threshold" in the sense implied by Groffman et al. (2006), to refer to a nonlinear response by a system to a stressor. We follow Bennetts et al. (2007) in their use of "assessment point" to refer to a value along the continuum of a metric that has relevance to managers, including an ecological threshold.

Using ecological integrity to communicate with managers requires five steps, often implemented iteratively: (1) document the scale of the project and the current conceptual understanding of the ecosystem, (2) select appropriate metrics, (3) define assessment points, (4) collect data and calculate metrics, and (5) produce a report or other communication tool. In particular, steps 2 and 3 may be revisited multiple times as a monitoring program develops and data is collected and analyzed, causing scientists to rethink the metric choices and assessment points. This five-step process shares a number of characteristics with other frameworks for developing research and monitoring programs (e.g., Fancy and Bennetts 2012). The next sections cover these steps in depth, and highlight places where knowledge of ecological thresholds fits into the framework. We then present two case studies illustrating the use of an ecological integrity approach to communicate ecosystem status to resource managers.

The Ecological Integrity Framework

Define Scale and Develop Conceptual Diagram

Begin by defining the scale, specifically the spatial and temporal scale of the ecological system being evaluated. This includes documenting the geographic boundary of the system and the specific features of the system within that boundary. The spatial scale is equivalent to a statistical population, and can be general (e.g., forests of the USA) or specific (e.g., Pitch Pine Woodlands in Acadia National Park greater than 0.5 ha in areal extent and correctly identified on the 2003 vegetation map). The temporal scale is also important, and includes consideration of the timing of data collection (e.g., summer only or year-round) and the planned duration (e.g., one time or repeated). Clear spatial and temporal scales are essential for data collection, and will help guide the development of the conceptual diagram and metrics.

Conceptual ecological diagrams or models that describe major ecosystem functions and delineate linkages between key ecosystem attributes and known stressors or agents of change are an essential tool for identifying and interpreting metrics with high ecological and management relevance (Fig. 10.1) (Noon 2003). The specific features of conceptual diagrams can vary, and the approach can include models that organize the linkages among on-site condition and patch size with surrounding landscape attributes (Unnasch et al. 2009, Faber-Langendoen et al. 2012). Here we focus on the primary components of integrity: composition, structure, and function. Composition refers to the species making up the ecosystem, including overall species richness and evenness. Structure means the physical characteristics of the

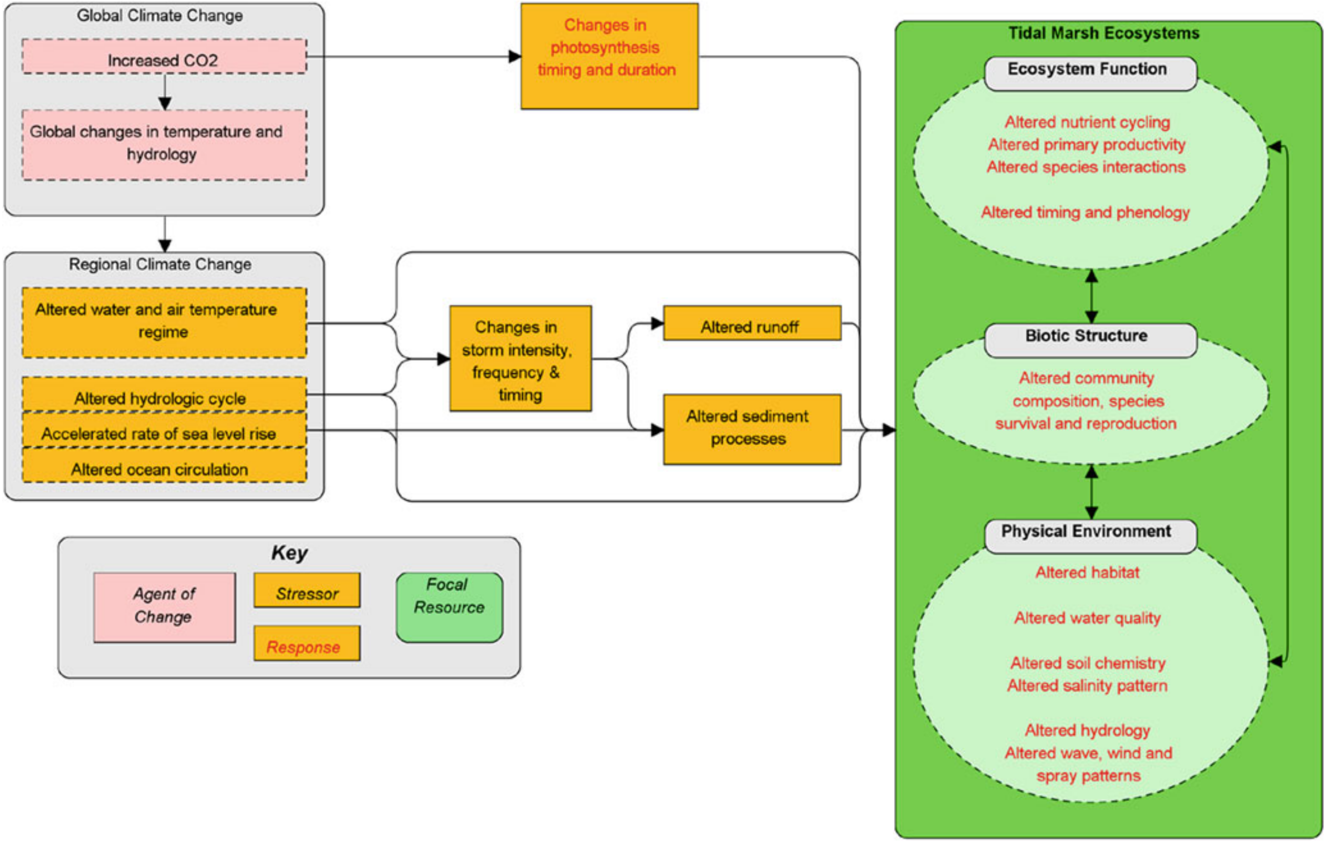


Fig. 10.1 Example conceptual diagram documenting the expected effects of climate change on tidal marsh systems. (From Stevens et al. 2010)

system at multiple scales, including vertical stratification, physical substrates and microhabitats, and landscape level features like patchiness and connectivity. Function covers dynamic characteristics like species demography and interactions as well as ecological processes like carbon, nitrogen, and water cycling. For each of these components, it is essential to document the important ecological features, and how they relate to one another, including aspects of the ecosystem that are important for resource managers.

Next, consider the ecosystem drivers, or the factors that work to maintain the system in its current state, and stressors that can disrupt the system. Formally, drivers are external forces like climate, fire, and natural disturbance that have large scale influences on natural systems (National Park Service 2012). In contrast, stressors are perturbations to a system that are either foreign to the system or are applied at an excessive (or deficient) level (Barrett et al. 1976). Stressors therefore can cause a shift in the status of a driver, with potentially cascading effects on the ecosystem. While considering the ecosystem drivers, think about and document the different pathways through which the drivers and stressors can affect ecosystem composition, structure, and function. It may help to distinguish between two different types of drivers: external drivers (like climate) create an effect, while internal drivers (like nutrient levels) convey the effect to the biota. Understanding the linkages whereby internal drivers mediate or transfer the effects of human disturbance to the biotic communities can be important for devising interventions to restore the system. The conceptual diagram is also the first place to consider the potential impacts of ecological thresholds. Are some stressors more likely to produce nonlinear or threshold effects on ecosystems than others? For example, will an increase in atmospheric deposition of nitrogen cause a sudden shift in trophic status of an aquatic system, or a gradual change?

Your conceptual diagram may be a simple figure with supplemental text that describes ecological components and potential effects of stressors (e.g., Mitchell et al. 2006), or it may be a highly structured set of models and submodels that makes specific hypotheses about the mechanistic relationships between model elements (e.g., Miller et al. 2010a). Whatever the level of detail chosen, the goal is to formally document the current understanding of the ecosystem— including known and potential threshold effects—in a way that supports the selection of a suite of metrics suitable for representing ecological integrity. Looking to the future, statistical methods now exist that can permit conceptual diagrams to be translated into formal causal network hypotheses, which can be evaluated using empirical data (Grace et al. 2010). As knowledge of the ecosystem improves, the conceptual diagram or model should be periodically updated, and the changes should be reviewed to determine whether changes in metrics or assessment points are warranted.

Select Metrics

The second step in determining ecological integrity is identifying a limited number of metrics that best distinguish condition classes or gradients from a highly impacted, degraded, or depauperate state to a relatively unimpaired, complete, and functioning state. These metrics can be a single response measure (field measurement) but more commonly they are calculated values based on field data. They may be properties that typify a particular ecosystem or attributes that change predictably in response to anthropogenic stress. The suite of metrics selected should be comprehensive enough to incorporate composition, structure, and function of an ecosystem across the spatial and temporal scales defined at the beginning of the previous step. Ideally, indicators of the magnitude of key stressors acting upon the system will be included to increase understanding of the relationships between stressors and effects (Tierney et al. 2009). Developing effective metrics requires access to existing studies or pilot data so that a variety of metrics can be calculated and assessed; this process may be iterative, as initial data collection efforts demonstrate the need for revised metrics and potentially different data.

When choosing metrics, consider the following four fundamental questions (Kurtz et al. 2001): (1) Is the metric conceptually relevant? Conceptually relevant metrics are related to the characteristics of the ecosystem or to the stressors that affect its integrity, and can provide information that is meaningful to resource managers. (2) Can the metric be feasibly implemented? The most feasible metrics can be sampled and measured using methods that are technically sound, appropriate, efficient, and inexpensive. (3) Is the response variability understood? Every metric has an associated measurement error, temporal variability, and spatial variability, and the best metrics will have low error and variability compared to the variability in the ecological component or stressor it is designed to measure. In other words, good metrics have high discriminatory ability, and the signal from the metric is not lost in measurement error or environmental noise. Ideally the metric will be measured across a range of sites that span the gradient of stressor levels (DeKeyser et al. 2003), and verified to show a clear response to the stressor. (4) Is the metric interpretable and useful? The best metrics provide information on ecological integrity that is meaningful to resource managers.

Part of the process of selecting metrics should include exploring the relationship between each metric and ecological condition, with explicit consideration of threshold behavior. Indicators with thresholds at or just beyond the limits of a system's range of natural variability (Fig. 10.2a) can be excellent ecological integrity metrics, since in this case, moving beyond the normal range produces a marked change in the value of the indicator that should be easier to detect. Indicators with thresholds within but near the edge of the range of natural variability (Fig. 10.2b) can also be suitable ecological integrity metrics, because they can serve as an early warning of potential change. However, indicators with thresholds far inside or outside the range of natural variability (Fig. 10.2c, d) are usually poor ecological integrity metrics, since they can lead to false alarms or not show a change until after the ecosystem has fundamentally changed (although see below for a situation where Fig. 10.2d may be a good metric).

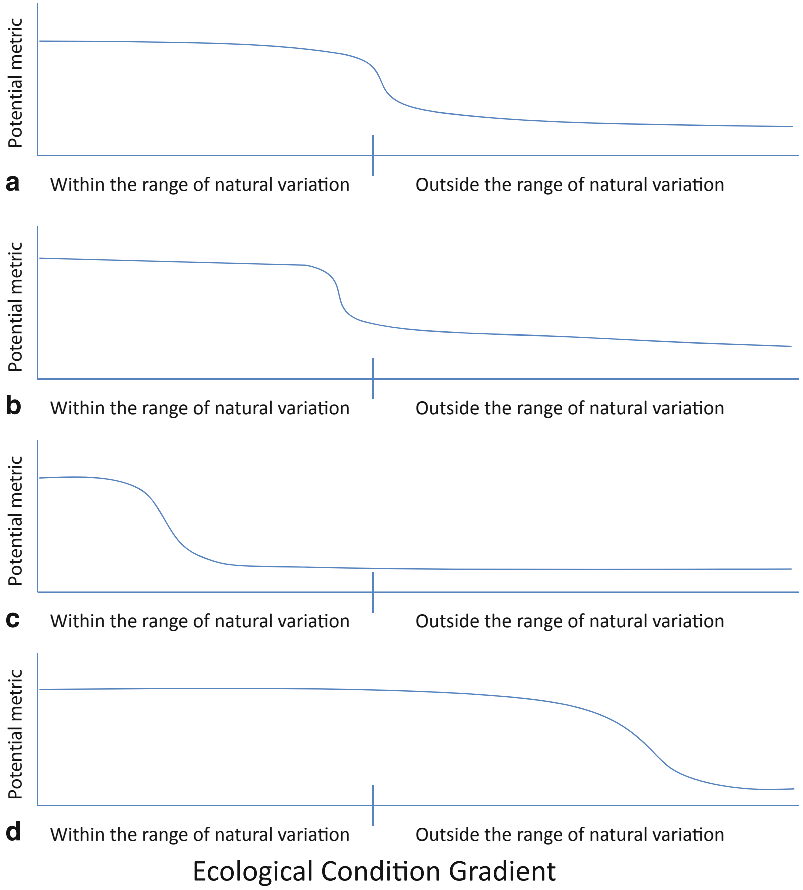


Fig. 10.2 The location of a threshold relative to the range of natural variation for a system affects the value of potential metrics. **a** and **b** are suitable ecological integrity metrics, but **c** and **d** are usually not

Determine Assessment Points

Once you have selected metrics, review the list and think about how you plan to report ecological integrity to decision makers. Is it important for describing the overall condition of systems to be able to arrive at a single number representing ecological integrity derived from the suite of metrics, such as through an Index of Biotic Integrity or other modeling approach (Karr 1991)? Or, will it be more valuable to provide a set of metrics that reflects different components of the system's overall integrity? A single value is often attractive because of its simplicity, but you risk oversimplification and difficulty interpreting its meaning. If you decide to combine a set of metrics into a single summary metric, any data analyses and modeling should be clearly documented, with assessment points usually developed for the summary metric rather than the component metrics. On the other hand, a suite of metrics can provide more nuanced insight into particular aspects of ecological integrity that may be at risk. In most situations, it will help to present a combination metric like an IBI as an overall summary, while also including some or all of its component metrics, which may have more direct management relevance and be easier to interpret.

For each final metric, establish assessment points that distinguish expected or acceptable conditions from undesired ones that warrant concern, further evaluation, or management action (Bennetts et al. 2007). Assessment points are "preselected points along a continuum of resource-indicator values where scientists and managers have together agreed that they want to stop and assess the status or trend of a resource relative to program goals, natural variation, or potential concerns" (Bennetts et al. 2007). Based on Bennetts et al., we define two categories of assessment points that are useful for ecological integrity reporting: ecological assessment points related to ecosystem condition, and management assessment points derived from the goals of resource managers. Types of management assessment points include surveillance assessment points that indicate when extra attention, research, and planning are needed; and action assessment points that define when management action should be taken. Two or more assessment points can share the same metric value, such as when action and ecological assessment points are identical. Alternatively, one category of assessment point may have multiple values, such as when one ecological assessment point represents the point where a system exceeds its range of natural variation and additional ecological assessment points indicate different levels of degradation.

Ecological assessment points are derived from some characterization of either natural or historical variability. Estimates of historical or natural variation in ecosystem attributes provide a reference for gauging the effects of current anthropogenic stressors, while at the same time recognizing the inherent natural variation in ecosystems across space, time, and stages of ecological succession (Landres et al. 1999). This may be empirically derived from the extant distribution of a metric across a defined spatial and temporal scale (especially of relatively pristine ecosystems like large wilderness areas or national parks), inferred from the best available information prior to meaningful anthropogenic disturbance (e.g., paleoecological reconstructions) or via models of ecosystem dynamics. In some cases, there is no relevant existing data for one or more metrics, and in these cases, initial assessment points should be established based on expert judgment or baseline data collection (e.g., the first 5 years of data, assuming the sample design is appropriate for this purpose). Although all of these provide useful insight, our understanding of historical and natural conditions in many ecosystems relies on a limited number of key studies, and care must be taken when extrapolating these data to other areas (Tierney et al. 2009). Whatever the source of the data, our understanding of the range of natural variation and any ecological assessment points based on this knowledge need to be periodically reviewed and updated to ensure that we are using the best available information for decisions.

If you are confident that a nonlinearity in a metric's functional form corresponds to a true ecological threshold, it may make sense to use the threshold as an ecological assessment point rather than strictly relying on the range of natural variability. This aligns with the idea of ecological integrity including resistance to perturbation in

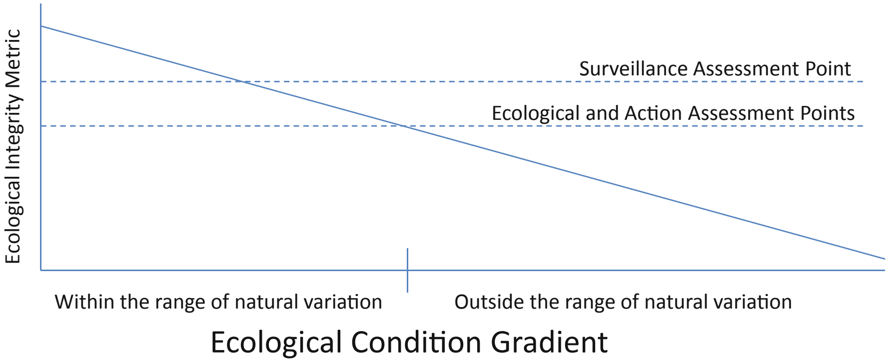


Fig. 10.3 Assessment points for an ecological integrity metric that does not exhibit threshold behavior

addition to the historical or natural variation in the system (Parrish et al. 2003). Even though an ecosystem component may function within a certain range, it may be that integrity (as measured by the particular metric) does not change noticeably unless the range of natural variation is exceeded by a large amount (e.g., Fig. 10.2d). In this situation, ecological integrity may not be threatened by exceeding the range of natural variation, but it would be altered by exceeding the ecological threshold, so the latter point may be more suitable for an ecological assessment point.

In many cases, the point where a metric's value indicates that an ecosystem has exceeded its range of natural variation—a critical ecological assessment point—can also be used as an action assessment point. This is the point where active steps need to be taken to bring the ecosystem back within the natural range. A separate surveillance assessment point can be established near this point but within the natural range of variation, indicating a need for vigilance and planning for potential corrective measures (Fig. 10.3). In other cases, particularly when metrics exhibit thresholds near an ecological assessment point, the placement of action and surveillance assessment points may need to be adjusted (Fig. 10.4). Setting an action assessment point near the ecological threshold will maximize the discriminatory ability of the assessment effort and help ensure that action is not taken in the absence of a real change in ecological integrity. A surveillance assessment point is best set where the metric's value begins to enter the zone where small changes in the ecological state begin to produce a large effect on the metric (Fig. 10.4). In making these decisions, it is important to consider the lag times associated with system response, however, because lag responses can increase the need to anticipate a system's approach to a threshold so that actions can prevent further degradation (Contamin and Ellison 2009).

Action and surveillance assessment points may be shifted from an ecological assessment point in some additional situations. One of these is when there is high uncertainty in measurements of a metric. In this case, one must balance the risk of delaying discovery of an ecological problem with that of falsely identifying a problem. If you determine the risk of delaying discovery to be more important,

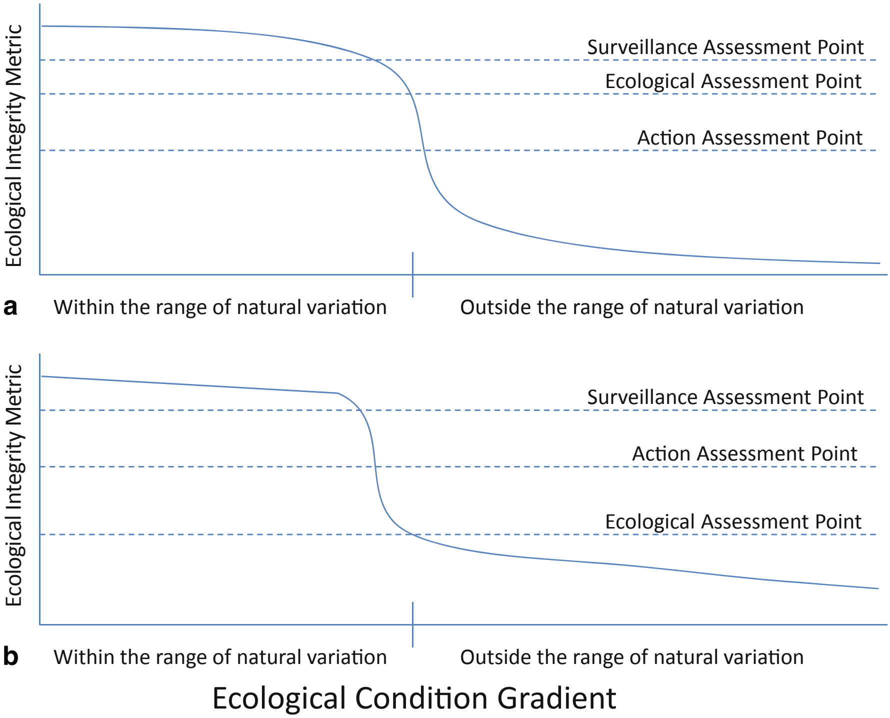


Fig. 10.4 Assessment points for ecological integrity metrics that exhibit threshold behavior. Panel **a** illustrates a case where the threshold is close to the boundary between natural and unnatural variation, while panel **b** illustrates a case where the threshold falls within the range of natural variation

assessment points should be shifted inside the natural range of variation. If the risk of falsely reporting a problem is more important, then shift the assessment point outside the natural range of variation. You may also want to consider dropping this metric, improving the precision of the measurement, or quantifying measurement error through quality assurance and quality control procedures.

Another type of situation occurs when resource managers have a goal other than ecological integrity, or when an ecosystem is already well outside of ecological integrity and interim recovery goals are needed. In these cases, ecological assessment points still serve as valuable, science-based benchmarks, but the action and surveillance points will likely be set relative to management targets or “utility thresholds” (Nichols et al. 2012) chosen for their relevance to decision makers. For example, managers of a historic site or a military base may be willing to accept some deviations from ecological integrity in order to preserve the historic scene or military readiness, and can benefit from working with scientists to set reasonable action and surveillance assessment points that protect ecological integrity as much as possible.

After establishing assessment points for each metric, you should thoroughly document the relevant spatial and temporal scales, information used in determining the natural range of variation and resistance to perturbation, implications of sampling uncertainty, and all decisions regarding where to place assessment points. These decisions may be based on statistical analyses, professional judgment (e.g., through discussions with resource managers), or they may be somewhat arbitrary. It is important to make the basis of all decisions clear and easily accessible in order to facilitate periodic reviews and revisions (Fancy et al. 2009).

Collect Data and Calculate Metrics

Some amount of data collection probably happened before metrics were chosen and assessment points defined, and this information is important to the previous steps and for future iterations of the ecological integrity framework. Existing data can be particularly valuable in determining whether metrics are feasible, with appropriately understood response variability (Kurtz et al. 2001). In many cases, though, the conceptual diagramming and metric selection process identifies new or different metrics that have not previously been collected or calculated, so new data and analyses are needed before indicators of ecological integrity can be estimated. Data collection should be matched to the desired spatial and temporal scale defined in the first step of the framework. This typically entails a sampling design focused on the statistical population, but it is also possible to use a well-chosen set of index sites to document site-specific trends, although this prohibits rigorous extrapolation to the full population.

Regularly scheduled new data collection and metric calculation, typically integrated into long-term ecological monitoring using detailed protocols (see Oakley et al. 2003 for guidelines), is essential for providing up-to-date ecological integrity data to decision makers. An extensive longitudinal data set for a population of sites provides a foundation for testing hypotheses about relationships among ecological components and stressors that are based on the conceptual diagram, and facilitates updating the diagram. Longitudinal data also help to clarify temporal variability of metrics and can uncover metrics that are highly correlated and thus duplicative and unnecessary for continued use in ecological integrity reporting.

Report Results

The final step in the ecological integrity framework is to ensure that results reach the hands of decision makers in a timely manner and in a format that is accessible and useful. This can be a one or two page "brief" that presents the highlights for upper-level administrators or a longer report with more detail for resource managers. Regardless of format, the information should describe the spatial and temporal scale and refrain from extrapolating beyond the data. It should include a simple summary that illustrates metric values for sites or management units in relation to the established assessment points, plus audience-appropriate explanations of each metric and key findings and recommendations. Ecological integrity reports also need to highlight measurement or other uncertainties, including uncertainties about ecological assessment points.

Although a variety of reporting approaches are possible, some of the authors (B. Mitchell, G. Tierney, and K. Miller) have had success using a "stoplight" system, where "Good" (green) represents an acceptable condition, "Caution" (yellow) indicates that the surveillance assessment point has been passed and a problem may exist, and "Significant Concern" (red) means that the action assessment point has been passed and that an undesirable condition exists that may require management correction (Tierney et al. 2009). A similar approach would categorize the condition of sites as "Good," "Moderate," and "Poor" (e.g., James-Pirri et al. 2012). It is important not to raise a false alarm when historical information or the data have high levels of uncertainty. One way to avoid this pitfall is by avoiding use of the "Significant Concern" category for metrics where there is uncertainty about the location of an ecological assessment point; in these cases, it may help to define a "Caution" or surveillance assessment point and defer decisions on other assessment points until additional data are available.

In our experience, the most effective reporting approach has been a tiered system, with short summaries pointing the way to a more detailed report that contains links or references to the most detailed raw data, descriptions of the conceptual diagram, metrics, assessment points, and data collection methods. Tiered reports allow decision makers to start with the simplest summaries, and drill down to the level of detail that is most appropriate for them. This approach also ensures maximum transparency, by making it easy to find the raw data, rationale for the choice of metrics, justifications for the assessment points, and data collection methods.

Wetland Ecological Integrity in Rocky Mountain National Park

The National Park Service (NPS) Rocky Mountain Inventory and Monitoring Network (ROMN) is using the ecological integrity framework to monitor and report the condition of wetlands in several park units. Here we focus on the process followed in Rocky Mountain National Park (ROMO). ROMO is a large park in the North Central Rockies of Colorado. Most of the park is designated wilderness, and it has important wetland resources that support iconic wildlife such as elk and beaver.

Define Scale and Develop Conceptual Diagram

The primary spatial scales of interest for wetland monitoring in ROMO included specific individual wetlands as well as the complete population of wetlands across the park. Most sites were selected using a spatially balanced survey design (Stevens and Olsen 2004) that allows unbiased estimation (Olsen et al. 1999) for the population of wetlands in the park. Because implementing a survey in a park like ROMO is expensive, these sites are sampled across time using a paneled structure (Urquhart

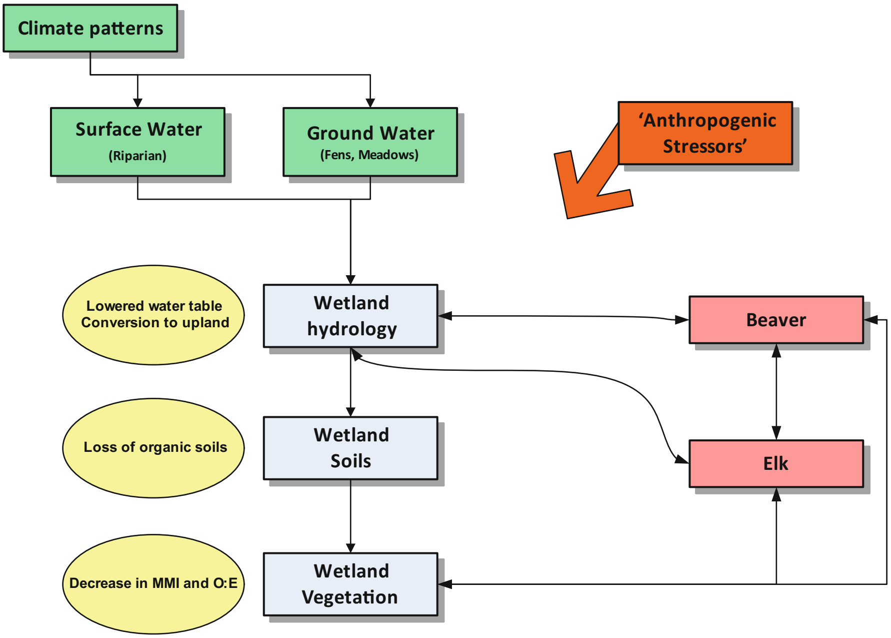


Fig. 10.5 Simple conceptual diagram of the key drivers and responses of wetlands in Rocky Mountain National Park

and Kincaid 1999). Additional annual monitoring is conducted at a subset of hand-picked "sentinel" wetlands that are either representative of key wetland types in the park or have management significance. Sentinel sites allow more detailed treatments of the ecology of place (Billick and Pierce 2010) and are more efficient to monitor, but do not statistically represent wetland resources throughout the park. The temporal scale of interest is long term. Shorter term variation is important and the sample design and analyses attempt to accommodate it, but the true power and utility of the approach may not be realized for several years.

As with all NPS monitoring networks, ROMN developed conceptual diagrams as part of its general monitoring plan (Britten et al. 2007). These were revisited during the development of the wetland monitoring protocol (Fig. 10.5) (Schweiger et al. 2010a, 2010b) to ensure that park-specific drivers and stressors were included (in ROMO, beaver and ungulate herbivory), as well as more global threats like anthropogenic hydrological alterations (Gage and Cooper 2009), climate change (Field et al. 2007), and aerial nitrogen deposition (Baron et al. 2009).

Select Metrics

Metrics were selected by the ROMN using two strategies. First, several were defined a priori based on conceptual diagramming. Second, and perhaps more importantly, a set of metrics was developed using a large pilot effort in the park and a series of analyses and models (summarized below). This ensured that the metrics would be scale-appropriate, ecologically responsive, efficient, and logistically feasible given budget constraints. Using pilot data also allowed the ROMN to work with resource managers to evaluate the management relevance of candidate metrics. All metrics were related to ecosystem composition, structure, function, and key stressors—the core elements of ecological integrity.

Compositional metrics were based on the wetland vegetation assemblage, which was particularly important at ROMO because wetlands are biodiversity hotspots, containing approximately 37 % of the park's plant taxa within less than 4 % of its area (Schweiger et al. 2010b). An a priori decision was made to focus on wetland vegetation as the primary biological response measure given cost considerations, the integrative and likely sensitive response of vegetation to wetland disturbance (Mack 2001), and its central role in nearly all wetland functions. Vascular and nonvascular vegetation was sampled using a suite of nested plots at each site (Peet et al. 1998), and the data were developed into both individual metrics and Indices of Biotic Integrity (IBI) for each wetland type.

Wetland extent was the primary structural metric. This is because larger wetlands are likely better buffered from disturbance; their vegetation typically remains more intact and diverse (Risvold and Fonda 2001); and hydrologic services like water storage and purification function more naturally (Cooper et al. 2006; Mitsch and Gosselink 2007). Extent was quantified using the survey design and analysis of field assays of individual wetland complex area and type.

Hydrology can serve a structural and functional role in wetlands, and was also selected as a core metric. The hydrology of a wetland is likely one of the most important drivers of its extent, type, and condition (Gage and Cooper 2009), but because hydrology primarily affects wetlands via patterns in hydrologic variability, it is a difficult and expensive metric to monitor. ROMN measured instantaneous ground water depths at the peak of vegetation growth and development, when deviations from the range of natural variation should be most meaningful. ROMN also continuously recorded water table depths at sentinel sites and will integrate these more meaningful data in the future.

Other functional metrics were related to water chemistry and wetland soils. For water chemistry, ROMN focused on data that could be collected using a hand-held probe: pH, specific conductance, and temperature. Nutrients and other analytes were considered, but ROMN decided that the laboratory costs would be too high. The network addressed wetland soils by determining percent organic matter, depth of peat, and a suite of structural aspects like texture and horizon depths at their monitoring sites. A more detailed set of parameters including minerals, soil pH, carbon and nitrogen content, and cation exchange capacity were collected at sentinel sites, and these more complete characterizations of the soil resource will be integrated in the future.

The ROMN wetland conceptual diagram included stressors with strong hypothesized or known effects on wetland ecological integrity. Anthropogenic disturbance was estimated at the site, meso scale (wetland buffers), and landscape scale (the catchment of each wetland) through a series of measures of land use and cover, hydrologic alterations, and physical/chemical disturbances. Example response measures included estimates of intensive land use such as roads, trails, structures, dams, and ditches that have been shown to strongly influence wetland condition (Mack 2007; Lemly and Rocchio 2009). The individual disturbance indicators were combined into a metric of Human Disturbance Index (HDI) following an approach similar to Faber-Langendoen et al. (2006) and Lemly and Rocchio (2009). The HDI provides an independent measure of wetland condition against which vegetation attributes can be assessed to determine their relationship with human disturbance.

To incorporate the important role of natural disturbances in the park, several measures of stress not directly due to anthropogenic factors were developed. For example, beavers play an important role in shaping and maintaining wetlands in the park (Baker et al. 2005), and the network included a metric of the extent of beaver presence in ROMO wetlands. Similarly, the large elk herd at the park is a stressor to woody species like *Salix* spp. (Baker et al. 2005), and ROMN defined three browse metrics, including percent of dead stems, percent of crown dieback, and percent of browsed live stems.

Determine Assessment Points

Several of the ROMN's ecological assessment points were developed based on existing literature (especially Faber-Langendoen et al. (2006) and Lemly and Rocchio (2009)), collective experience with wetlands in the park, and discussions with park resource managers. This was the case for wetland extent, fen hydrology, some water chemistry parameters, most soil metrics, and elk browse. These responses had some existing science to support their assessment points, but they were not always specific to ROMO and therefore may not necessarily reflect wetland ecology in the park. These points represent a starting point for assessing ecological integrity, and will be reviewed as more data are collected and additional research is conducted.

A key element of the ROMN approach is the empirical development of park-specific reference conditions and ecological assessment points for wetland vegetation. The ROMN protocol adopted and modified methods for quantifying reference distributions and ecological assessment points created over the last two decades (Stoddard et al. 2006). The ROMN felt this was necessary because of the paucity of established assessment points or relevant thresholds for wetland vegetation, as well as the possible inappropriateness of applying existing regional research results to ROMO. National parks like ROMO are often unique landscapes with largely intact habitats and few of the anthropogenic stressors that structure wetland condition in more developed landscapes. There are gradients in human disturbance across the park, but they encompass a different range than broader landscapes and likely reflect different stressors.

The ROMN approach required several steps and was based on pilot data from over 300 sample events at 140 sites collected between 2007 and 2009. First, data

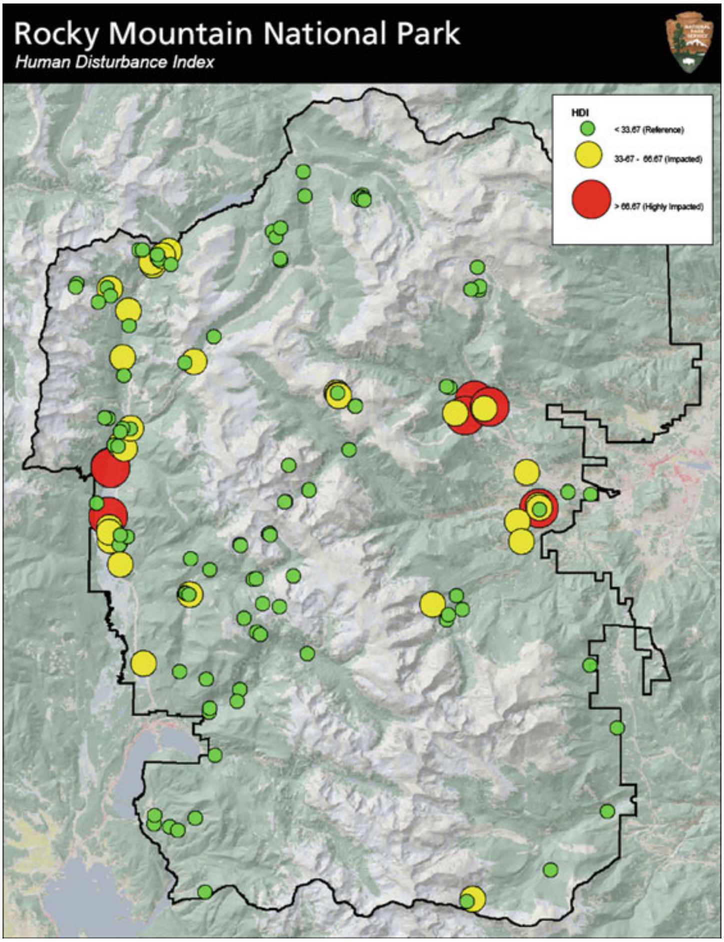


Fig. 10.6 Map of Human Disturbance Index (HDI) values in ROMO wetlands (all types). HDI ranges from ~ 0 to ~ 100. Reference, impacted, and highly impacted sites as defined by Colorado Natural Heritage Program (CNHP) arbitrary breakpoints (< 33.67 reference, 33.67–66.67 impacted, and > 66.67 highly impacted) are defined by the size and color of each point (*larger circles* and *redder colors* indicate sites with more human disturbance). A clear gradient exists from high-to-low HDI scores with higher disturbance in low elevation front country wetlands on both the east and west sides of the park

were classified into three wetland types (fens, wet meadows, and riparian) based on extensive prior wetland classification work in the region (Cooper 1998). Then the HDI was generated for each site, and assigned to a priori disturbance classes based on Colorado Natural Heritage Program break points established using professional judgment (Fig. 10.6). Third, metrics that described distinct responses of vegetation to anthropogenic disturbance were generated. Examples of metrics include percent invasive species (which might be expected to increase with disturbance), mean conservatism score (a measure of the fidelity of plant species to intact or degraded habitat that decreases with disturbance; Wilhelm and Masters 1995), and percent moss cover (which tends to decrease with disturbance). In total, ROMN created over 130 candidate metrics. The best metrics were selected from the full list by choosing the ones that were most strongly predictive of the anthropogenic disturbance gradient and that passed various statistical tests (including information content, reproducibility, independence from other metrics, and interpretability; Stoddard et al. 2008). The final metrics had meaningful responses to disturbance, were ecologically interpretable, were not redundant, and had favorable precision. ROMN also looked at relationships with environmental gradients like elevation and precipitation. If a metric responded to a natural environmental driver in the reference set of wetlands, ROMN statistically adjusted the data to remove the influence (Stoddard et al. 2008). This step was important in ROMO because several metrics did covary with environmental features, and these relationships can confound our ability to detect a response to the HDI. Finally, the best metrics were summed and scaled to range from 0 to 10, with ecological integrity increasing with the score. This final combined metric is an Index of Biological Integrity (IBI; Karr 1991; Mack 2001; Miller et al. 2006; Mack 2007) that the ROMN interprets as a synthetic estimate of wetland ecological integrity in the park.

The final IBI for each of the three wetland types contained between four and six component metrics that described a broad spectrum of wetland vegetation response to disturbance. The component metrics were distinct for each wetland type, but metrics based on either species invasiveness or species conservatism occur in each IBI. While the component metrics all had significant relationships with disturbance, each final index had stronger relationships (R^2 between 0.30 and 0.61) with the HDI than the individual component metrics. This result suggests that the indices were integrating ecological response and were likely meaningful indictors of the ecological integrity of wetlands in the park (Karr and Chu 1997).

To incorporate relevant ecological thresholds and establish ecological assessment points, the ROMN conducted a series of analyses to define condition classes specific for each wetland type. Regression tree models were used to determine change points in a predictor variable that best distinguished groups of values of a response variable. Reciprocal regression tree analyses (De'ath and Fabricius 2000) were conducted, one using IBI values as the predictor and HDI as the response and one vice versa to estimate these thresholds for each variable. Figure 10.7 is a riparian wetland example of the relationship between IBI and HDI, including the calculated threshold values. All three IBIs were best split into only two classes: “reference” or “non-reference.” Importantly, these ratings are specific to ROMO. In other landscapes, only two classes might suggest the final model was not very precise, but the ROMN believes this accurately describes the distribution in the park. Wetlands near visitor facilities, roads, and some park boundaries were often disturbed while wetlands in the wilderness backcountry were largely intact. Finally, all IBI models strongly discriminated among HDI classes, suggesting that there were ecological thresholds in the park’s wetland vegetation communities that were suitable for use as ecological

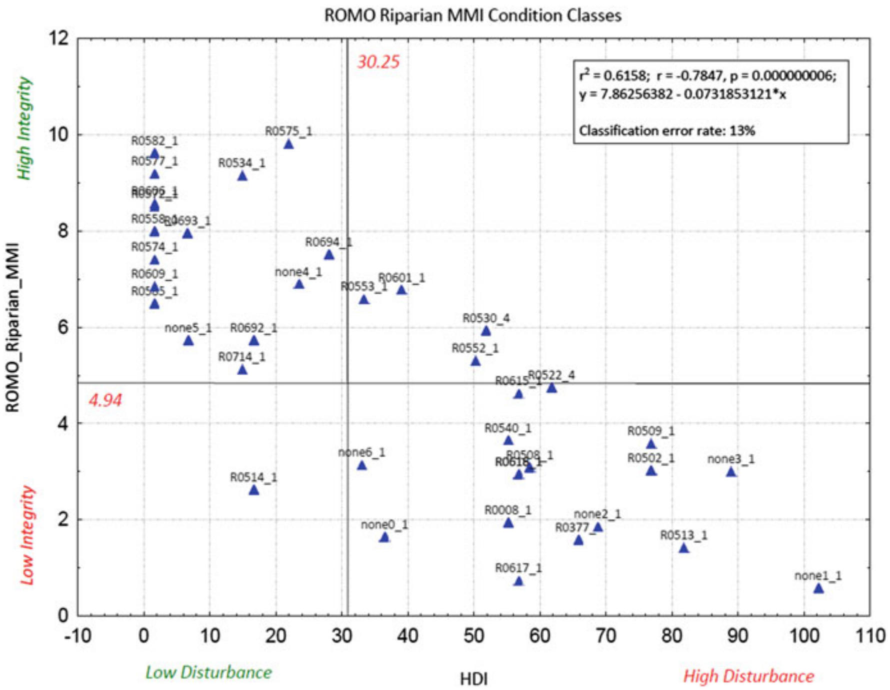


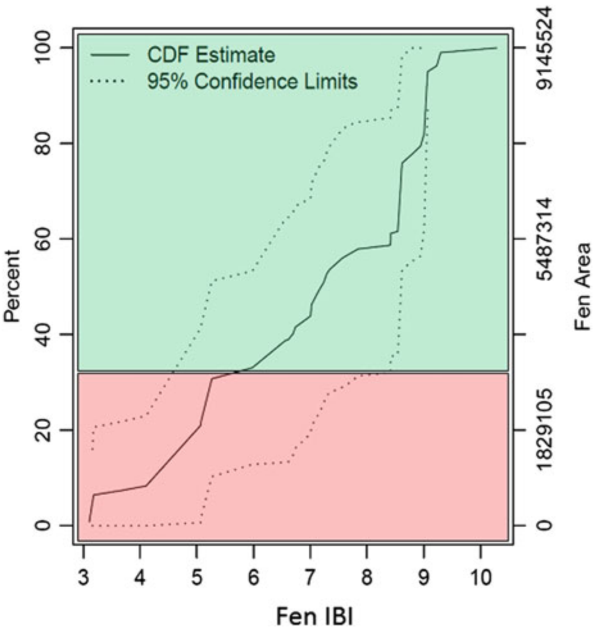
Fig. 10.7 Scatter plot of ROMO riparian wetland IBI vs. HDI. Linear model and Pearson correlations in inset show strong relationship with HDI. Classification error rate was derived using cross-validation models. Ecological assessment points based on threshold values (vertical and horizontal lines on the figure with corresponding scores of HDI = 30.25 and IBI = 4.94) were generated using regression tree models. (See Schweiger (2010b) for details)

assessment points, and that it will be possible to place novel sites into the correct ecological integrity category most of the time (Hawkins et al. 2000).

Both the reference and nonreference classes were characterized by a range of values at the appropriate scale (park rather than state or ecoregion). For reference sites, this reflects the natural range of variability, and for disturbed sites, it reflects variation due to human impacts plus the underlying natural variability. One of the primary reasons for the modeling effort conducted by the ROMN was to define the park-specific condition gradient from reference to impacted. This gradient in ROMO may be quite different from the larger landscape; a non-reference designation in ROMO may be a relatively intact wetland if the scale of the assessment were broader.

Because most of the sample sites used to generate the IBI models were from a survey design, the ROMN used design-based analysis (Olsen et al. 1999) with the IBI-based ecological assessment points (and many other metrics, see Schweiger et al. 2010b) to estimate wetland condition at the population scale. Fig. 10.8 shows an example of one of the key outputs from these analyses using the ROMO fen IBI—a cumulative distribution function (CDF). Generally speaking, a CDF is an “area so far” function of the probability distribution for a response or a metric.

Fig. 10.8 Cumulative distribution function (CDF) of ROMO fen IBI. *Green* and *red* regions on the figure are based on where the ecological assessment point of 5.64 intersects the curve. Using the properties of a CDF, this defines the percentage (or area) of the ROMO fen resource that is in impacted (30.8 ± 20 %) and reference (69.2 ± 20 %) condition classes. (See Schweiger (2010b) for details)



Graphical presentations of CDFs aid visualization of the probability distribution and readily facilitate interpreting thresholds placed within the distribution by locating the percentage of the response that is above or below the threshold. Using the threshold of 5.64 in the IBI generated from the regression tree analysis, the proportion of all fens in the park in an impacted and a reference state are shown graphically in Fig. 10.8. Sixty-nine percent (± 20 %) of ROMO fen habitat is in a reference state and 31 % (± 20 %) is in a nonreference condition.

These analyses and results define the baseline of wetland condition for long-term monitoring in ROMN. Additionally, the ROMN approach to developing metrics and assessment points facilitates the distillation of large volumes of ecological data into concise results that decision makers can use for resource management.

Collect Data and Calculate Metrics

Select elements of the pilot summarized above also served as the initial monitoring effort for wetlands in the park. The ROMN is currently reviewing what worked and what did not within the pilot and finalizing long-term plans for continued wetland monitoring in ROMO and other NPS units. Current plans include statistical sampling of the park's wetland population every 5–10 years, plus annual monitoring at four sentinel wetland complexes (Schweiger et al. 2010a). This frequency of data collection and recalculation of ecological integrity metrics will ensure that current information is available for resource managers, and that they have time to plan and react to changing conditions. The ROMN is also working directly with the park to develop related management assessment points based on the ecological assessment points developed during the pilot, the management needs of the park, and the precision of the wetland protocol.

Report Results

Now that ROMN has completed the pilot and its initial assessment of wetland ecological integrity in the park, the network is developing a suite of products to convey the results to managers, the public, and others. ROMN is developing concise “resource briefs” suitable for nonscientists, plus website content and other summaries. These documents will link to more detailed products that will include metric and assessment point justifications, field sampling methods, and analysis details (Schweiger et al. 2010a). This means that park managers and other stakeholders will have access to relevant summary information on wetland ecological integrity at ROMO, and will have the option of digging deeper to investigate the science behind each assessment.

Forest Ecological Integrity in Northeastern National Parks

Another example of the ecological integrity framework in action is forest monitoring in northeastern national parks. The NPS Northeast Temperate Inventory and Monitoring Network (NETN) monitors forests in ten national parks, including Acadia National Park (ACAD), Marsh-Billings-Rockefeller National Historical Park (MABI), and Morristown National Historical Park (MORR). Covering over 14,000 ha, ACAD is situated on the coast of Maine, and is dominated by second-growth spruce–fir forests that have had minimal management for nearly 100 years. MABI is a small (225 ha) park in rural Vermont with an ongoing forestry operation. The park is dominated by forest land, which consists of a patchwork of northern hardwoods and monoculture conifer plantations. MORR, a 691 ha park in suburban New Jersey, is predominantly northern hardwoods and is heavily impacted by invasive species and browsing by white-tailed deer. These three parks are the focus of this example, although NETN monitors and reports forest ecological integrity for the larger group.

Define Scale and Develop Conceptual Diagram

NETN defined its scale for forest ecological integrity as the long-term monitoring of forest condition (and more open woodland communities at ACAD) within park boundaries during the summer season. Within this population, permanent plots were stratified by park and selected using a spatially balanced random sample (Stevens and Olsen 2004). Like the Rocky Mountain Network, NETN developed a conceptual diagram for forests during the development of their monitoring plan (Mitchell et al. 2006; Fancy et al. 2009) and used this diagram during the metric selection process. NETN evaluates ecological integrity of forested systems using the plot data and 13 metrics of ecological composition, structure, and function that are broadly applicable across northeastern temperate forests (Tierney et al. 2009, 2010).

Select Metrics

NETN uses five composition metrics. Tree regeneration indicates the quantity and composition of established tree seedlings and therefore of potential future canopy composition, and is substantially impacted by a historically large eastern US population of white-tailed deer (*Odocoileus virginianus*) (Cote et al. 2004). Tree condition, based on qualitative observations of disease, pests, pathogens, and canopy foliage problems, provides an early warning indicator of infestation, disease, or decline of one or more species. Biotic homogenization is the process by which regional biodiversity declines over time, due to the addition of widespread exotic species and the loss of native species (Olden and Rooney 2006); this metric can be calculated between site pairs as a simple ratio of species present at two sites over the total species present at either site (Jaccard's Similarity Index; Olden and Poff 2003). Invasive exotic plant species exploit and alter habitat, and are monitored by recording the frequency of 22 exotic species that are highly invasive in northeastern forest, woodland, and successional habitats. Deer browse can affect understory plant composition in addition to tree regeneration (Augustine and DeCalesta 2003), and NETN monitors the change between monitoring events in the abundance of common preferred browse species and unpalatable species.

The network's structural forest ecological integrity metrics include two landscape metrics and three stand-level metrics. Forest patch size strongly impacts habitat suitability for a variety of taxa (Fahrig 2003), with larger forest patches supporting larger populations of fauna and more native, specialist, and forest interior-dwelling species. Human land use, based on the percentage of land area containing human land use versus "natural" land use within a 50 ha (400 m radius) circle around each forest plot, is used to estimate the impacts of habitat loss within a local neighborhood. The stand-level metrics are stand structural class, snag abundance, and coarse woody debris (CWD) volume. Using the method of Goodell and Faber-Langendoen (2007), NETN calculates stand structural stage from tree size and canopy position measurements; this metric helps the network assess altered disturbance regimes coincident with global change and exotic pest and pathogen outbreaks (Dale et al. 2001). Dead wood, in the form of snags (standing dead trees) and fallen CWD, is an important structural component that provides necessary habitat for many forest taxa. Silviculture and land management often reduce the quantity and quality of dead wood, but ecologically based land management can retain or enhance these features (Keeton 2006).

The NETN also selected three metrics of ecosystem function: canopy tree growth and mortality, acid stress, and nitrogen saturation. Decreased growth or elevated mortality rates may indicate a particular health problem, such as sugar maple decline (Duchesne et al. 2003), or may indicate a regional environmental stress (Dobbertin 2005). Acid stress (primarily from atmospheric deposition) is measured in forest soil based on the molar ratio of calcium to aluminum (Cronan and Grigal 1995). Nitrogen saturation (also from atmospheric deposition) may exacerbate the effects of acidification (Aber et al. 1998) and is measured in forest soil based on the ratio of carbon to nitrogen.

Determine Assessment Points

Once the metrics were identified, the NETN developed assessment points. NETN established action and surveillance assessment points based on ecological assessment points, using existing research whenever possible. If the available data suggested a range of values for the limits of natural variation, the network typically created a surveillance assessment point at the lower (more “natural”) value and an action assessment point at the higher (less “natural”) value (Tierney et al. 2010). Because NETN’s field methods and many of their metrics were closely related to methods used by the well-established US Forest Service’s Forest Inventory and Analysis (FIA) program (http://www.fia.fs.fed.us/), in some cases there were scientifically-based ecological assessment points or existing baseline data to facilitate the process. For example, assessment points for the tree regeneration metric were partly based on FIA research (McWilliams et al. 2005), using a stocking ratio metric and associated assessment point that varies by forest type and is partly based on a proposed metric for detecting ungulate impacts in forests (Sweetapple and Nugent 2004). In other cases, the FIA data were used as baseline data. This was the case for the tree growth rate assessment points, which were based on FIA regional and species-specific patterns (Tierney et al. 2010). A few NETN assessment points rely on comparisons to baseline data collected by the NETN. For the biotic homogenization and indicator browse metrics, assessment points were based on the changes from the baseline condition rather than a comparison to predetermined values, given the challenges of establishing historical baselines for these metrics (Tierney et al. 2010).

Most NETN assessment points were established by reviewing and applying existing research, and this is the primary place where the existence of ecological thresholds played an integral role in the process. For example, Aber et al. (2003) compiled data from sites across the northeastern USA and discovered that nitrification increased sharply below a C:N ratio of 20–25. Additionally, the Indicators of Forest Ecosystem Functioning (IFEF) database compiled data from sites across Europe and found that, below a C:N ratio of 25, overall nitrate leaching was significantly higher and more strongly correlated to nitrogen deposition (MacDonald et al. 2002). NETN used this research to establish a surveillance assessment point (“Caution” rating) at a C:N ratio of 25, and an action assessment point (“Significant Concern” rating) at a C:N ratio of 20. Ecological thresholds are also present in the minimum habitat patch sizes needed to support species. Kennedy et al. (2003) reviewed the available research, and found that minimum patch areas ranged up to 1 ha for invertebrates, up to 10 ha for small mammals, and up to 50 ha for the majority (75 %) of bird species, with much bigger patch sizes needed to support large mammals. The relatively small parks for which the NETN metric was designed could not independently support large mammal populations. Therefore, the network chose ecological assessment points based on the threshold patch sizes needed to support birds, small mammals, and invertebrates.

Collect Data and Calculate Metrics

NETN has been collecting data annually since 2006 at 350 fixed plots. Plot numbers vary across the ten parks, with as few as ten plots (one plot for each two forested hectares) at Weir Farm National Historic Site and up to 176 plots (1 plot for each 73 forested hectares) at ACAD. Half the plots at each of the network’s small parks are sampled every other year, and a quarter of ACAD’s plots are sampled each year (Tierney et al. 2010). Metrics are automatically calculated by the network’s monitoring database at a minimum of once every complete sampling cycle (every 4 years). Often interim calculations are produced using the current year’s data or a rolling window of the most recent 4 years. This frequency of data collection and metric calculation ensures that there are always current data available to address the needs of park managers.

Report Results

An integral part of the Northeast Temperate Network’s forest monitoring is producing a variety of reports that ensure park managers are aware of current forest conditions, that they have information explaining these conditions, and that they have access to the raw data if they need to explore the summary information more fully. The foundation of the NETN approach is simple summary tables (e.g., the comparison of three parks in Table 10.1) that provide an intuitive scorecard for managers and help them see the status of their park and how it compares to other network parks. Most of the metrics in the table are assessed park-wide or by management unit (for some of the larger parks). A few metrics, such as tree condition and tree regeneration, can be assessed accurately at the plot scale, allowing for the use of pie charts to convey the proportion of the park’s forest in each assessment category (Table 10.1).

The scorecard table suggests that ACAD is doing well, with all park-wide metrics within the ecological assessment points, and a small percentage of plots indicating poor integrity for the tree condition and regeneration metrics. Conditions at MABI

Table 10.1 Forest ecological integrity at three Northeast Temperate Network parks, based on a subset of ecological integrity metrics and data collected in 2007–2010. *Green* indicates that the park (or a percentage of the park for multicolored pie charts) is within the range of natural variation; *yellow* indicates that the surveillance (and first ecological) assessment point has been passed; *red* indicates that the action (and second ecological) assessment point has been exceeded

Metric	**Acadia NP**	**Marsh-Billings-Rockefeller NHP**	**Morristown NHP**
Composition: Indicator Invasive Species	🟢	🟡	🔴
Composition: Tree Condition			
Composition: Tree Regeneration			
Structure: Stand Structure	🟡	🟡	🟢
Structure: Snag Abundance	🟢	🟡	🟢
Structure: Coarse Woody Debris Volume	🟢	🟡	🟡
Function: Tree Mortality	🟢	🟡	TBD
Function: Soil Acid Stress	🟡	🟡	🟢
Function: Soil Nitrogen Saturation	🟢	🔴	🔴

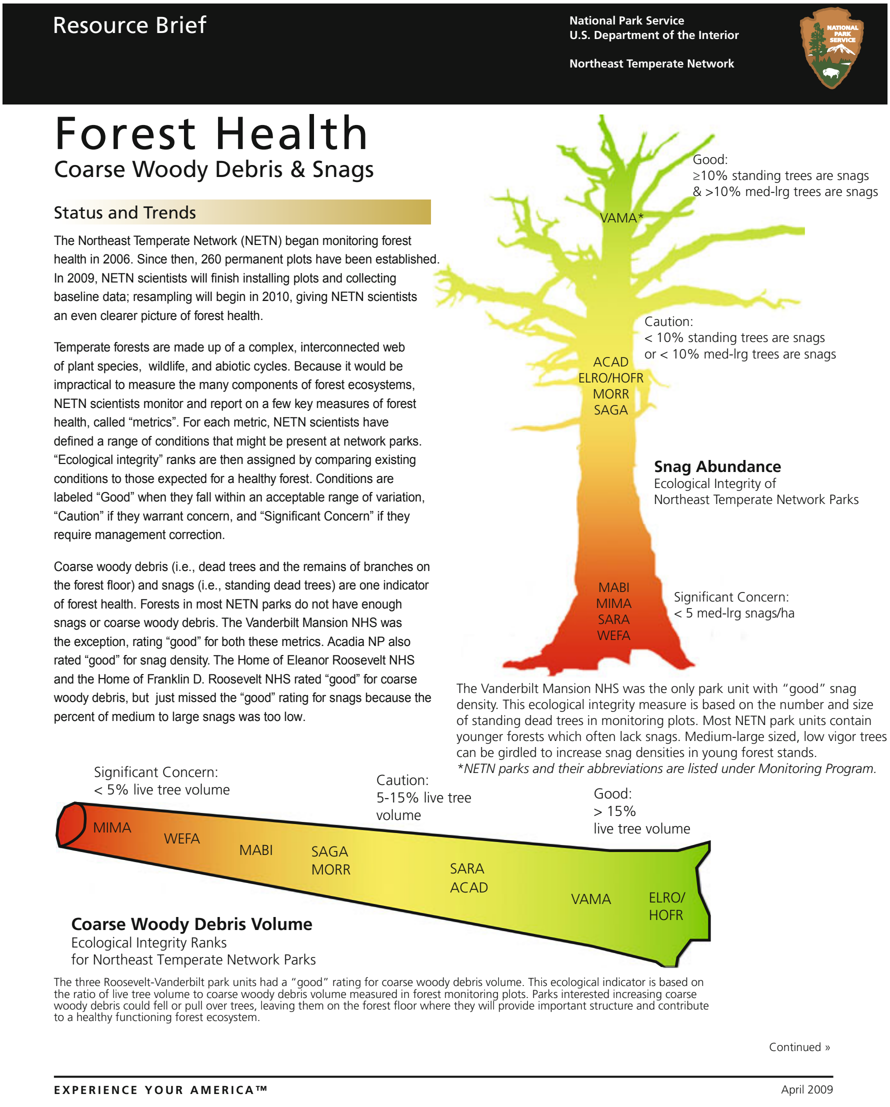
warrant close ongoing observation, since many of the metrics had scores between the surveillance and action assessment points. In addition, tree regeneration is likely inhibited at the park, and nitrogen saturation may have reached problematic levels. This park has chosen to be proactive about forest condition, and projects ranging from extensive invasive plant removal to silvicultural actions that increase snag and coarse woody debris abundance will likely improve the park's scores in the future. MORR's ratings are more checkered than the other parks. Although the structural and functional metric ratings indicate a reasonably good condition (albeit with excess nitrogen deposition), the compositional metrics indicate some problems. In particular, invasive plants are having a significant effect on plant diversity, and overabundant deer have effectively eliminated tree regeneration in much of the park. Both of these issues are high priorities for park managers.

While the ecological integrity scorecard is great for an at-a-glance summary, NETN makes sure to supplement the scores with additional details that include the actual values and assessment points for each metric, as well as discussion about the implications of the scores and possible corrective actions (e.g., Miller et al. 2010b). This information is reported at the park level, and when possible (for parks with more plots), the scores and interpretation are provided for management units within each park. Network staff also produce a series of resource briefs that highlight key information from the more technical report (e.g., Fig. 10.9); these publications are often popular with higher level managers as well as park education staff and interested members of the public. The scorecard report and resource briefs always provide citations or links to additional information, including the monitoring protocol that documents the ecological integrity metrics and assessment points (Tierney et al. 2010). All of these reports and communication tools are intended to support and supplement (rather than replace) regular in-depth data analyses and scientific reports that will explore trends and patterns in the long-term data set. The whole range of publications is organized and made available in digital format to resource managers through the NETN web site, so that they can quickly locate information when the need arises.

Conclusions

The ecological integrity framework is a powerful tool for organizing complex data sets and conveying important information to resource managers. Even when managers do not have the time or background to fully explore the statistics or threshold dynamics that led to the choices of different ecological assessment points, they intuitively grasp the idea of an ecological system being inside or outside its historical range of variability. This framework has a number of important features. It can accommodate application of ecological thresholds—where they exist—at multiple points, particularly in the choice of suitable assessment points. In many cases, thresholds facilitate the process, particularly when they occur near limits of the range of historical variation of a system. The presence of threshold behavior in a metric can help guide the development of assessment points, because thresholds indicate places where the metric value changes rapidly (and can be detected more easily) in response to changes in the system. Lack of clear thresholds make the identification of ecological assessment points somewhat more arbitrary, but also allow for greater flexibility in the choice of surveillance and action points.

Thresholds have an additional value to contribute to the framework, in that they may highlight an important exception to the usual practice of setting assessment points around the historical range of variation. If there is significant resistance to perturbation in an ecosystem, with threshold behavior not occurring until well beyond the historical range of variation, this may provide a scientific basis for shifting an



Resource Brief

National Park Service
U.S. Department of the Interior

Northeast Temperate Network

Forest Health

Coarse Woody Debris & Snags

Status and Trends

The Northeast Temperate Network (NETN) began monitoring forest health in 2006. Since then, 260 permanent plots have been established. In 2009, NETN scientists will finish installing plots and collecting baseline data; resampling will begin in 2010, giving NETN scientists an even clearer picture of forest health.

Temperate forests are made up of a complex, interconnected web of plant species, wildlife, and abiotic cycles. Because it would be impractical to measure the many components of forest ecosystems, NETN scientists monitor and report on a few key measures of forest health, called "metrics". For each metric, NETN scientists have defined a range of conditions that might be present at network parks. "Ecological integrity" ranks are then assigned by comparing existing conditions to those expected for a healthy forest. Conditions are labeled "Good" when they fall within an acceptable range of variation, "Caution" if they warrant concern, and "Significant Concern" if they require management correction.

Coarse woody debris (i.e., dead trees and the remains of branches on the forest floor) and snags (i.e., standing dead trees) are one indicator of forest health. Forests in most NETN parks do not have enough snags or coarse woody debris. The Vanderbilt Mansion NHS was the exception, rating "good" for both these metrics. Acadia NP also rated "good" for snag density. The Home of Eleanor Roosevelt NHS and the Home of Franklin D. Roosevelt NHS rated "good" for coarse woody debris, but just missed the "good" rating for snags because the percent of medium to large snags was too low.

Snag Abundance
Ecological Integrity of
Northeast Temperate Network Parks

The Vanderbilt Mansion NHS was the only park unit with "good" snag density. This ecological integrity measure is based on the number and size of standing dead trees in monitoring plots. Most NETN park units contain younger forests which often lack snags. Medium-large sized, low vigor trees can be girdled to increase snag densities in young forest stands.
**NETN parks and their abbreviations are listed under Monitoring Program.*

Coarse Woody Debris Volume
Ecological Integrity Ranks
for Northeast Temperate Network Parks

The three Roosevelt-Vanderbilt park units had a "good" rating for coarse woody debris volume. This ecological indicator is based on the ratio of live tree volume to coarse woody debris volume measured in forest monitoring plots. Parks interested increasing coarse woody debris could fell or pull over trees, leaving them on the forest floor where they will provide important structure and contribute to a healthy functioning forest ecosystem.

Continued »

EXPERIENCE YOUR AMERICA™ April 2009

Fig. 10.9 Example northeast temperate network resource brief describing status of the coarse woody debris and snag metrics in network parks

ecological assessment point beyond the historical range. For example, suppose an ecosystem has a natural range of variation in its carbon to nitrogen ratio of 40–80, but research shows no effects of nitrogen saturation until the ratio drops below 25. This situation suggests placement of an ecological assessment point at a ratio of 25, even though 40 represents the lower bound of the range of natural variation.

Other valuable features of the framework stem from its focus on providing useful and timely information to resource managers. The framework is transparent, since decisions and analyses are documented and easily available for review. It is appropriate for a variety of audiences, particularly when a hierarchy of publications is produced that vary in technical detail and allow readers to find their level of comfort, while keeping the full details within easy reach. The framework is also iterative and easily integrated into the adaptive management cycle (Lancia et al. 1996). The iterative nature of the framework is particularly apparent early in the process, when one step in the framework often requires revisiting other steps. For example, initial data collection may reveal that a metric has more variability or is more expensive than originally expected, triggering a re-evaluation of the metric selection and assessment points. Alternatively, a new publication and ongoing data collection may reveal that a hypothesized relationship in the conceptual diagram was incorrect; this may suggest new hypothesis, new metrics, and even the discontinuation of current metrics. The ecological integrity framework can play a central role in the adaptive management cycle by regularly reporting current results to managers, incorporating new information (including the results of management actions and data analyses) into the conceptual foundation of the framework, and making modifications to metrics and reporting that reflect new knowledge.

Acknowledgments We acknowledge the assistance of Donald Schoolmaster, who contributed to the IBI analyses at ROMO, and Greg Shriver, for his early contributions to the NETN monitoring approach. We also gratefully acknowledge the NETN and ROMN field crews; their dedication and long hours in the field have made a lasting contribution to our understanding of forests and wetlands managed by the NPS. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Aber, J., W. McDowell, K. Nadelhoffer, A. Magill, G. Berntson, M. Kamakea, S. McNulty, W. Currie, L. Rustad, and I. Fernandez. 1998. Nitrogen saturation in temperate forest ecosystems: Hypotheses revisited. *Bioscience* 48:921–934.

Aber, J. D., C. L. Goodale, S. V. Ollinger, M. L. Smith, A. H. Magill, M. E. Martin, R. A. Hallett, and J. L. Stoddard. 2003. Is nitrogen deposition altering the nitrogen status of northeastern forests? *Bioscience* 53:375–389.

Andreasen, J. K., R. V. O'Neill, R. Noss, and N. C. Slosser. 2001. Considerations for the development of a terrestrial index of ecological integrity. *Ecological Indicators* 1:21–35.

Augustine, D. J., and D. DeCalesta. 2003. Defining deer overabundance and threats to forest communities: From individual plants to landscape structure. *Ecoscience* 10:472–486.

Baker, B. W., H. C. Ducharme, D. C. S. Mitchell, T. R. Stanley, and H. R. Peinetti. 2005. Interaction of beaver and elk herbivory reduces standing crop of willow. *Ecological Applications* 15: 110–118.

Baron, J. S., T. M. Schmidt, and M. D. Hartman. 2009. Climate-induced changes in high elevation stream nitrate dynamics. *Global Change Biology* 15:1777–1789.

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

Bennetts, R. E., J. E. Gross, K. Cahill, C. McIntyre, B. B. Bingham, A. Hubbard, L. Cameron, and S. L. Carter. 2007. Linking monitoring to management and planning: Assessment points as a generalized approach. *The George Wright Forum* 24:59–77.

Billick, I., and M. Pierce, eds. 2010. *The ecology of place: Contributions of place-based research to ecological understanding*. Chicago: The University of Chicago Press.

Britten, M., E. W. Schweiger, B. Frakes, D. Manier, and D. Pillmore. 2007. *Rocky Mountain Network vital signs monitoring plan*. (Natural resource report NPS/ROMN/NRR-2007/010.) Fort Collins: National Park Service

Contamin, R., and A. M. Ellison. 2009. Indicators of regime shifts in ecological systems: What do we need to know and when do we need to know it? *Ecological Applications* 19:799–816.

Cooper, D. J. 1998. Classification of Colorado's wetlands for use in HGM functional assessment: A first approximation. In *A characterization and functional assessment of reference Wetlands in Colorado*. Denver: Department of Natural Resources, Colorado Geological Survey—Division of Minerals and Geology.

Cooper, D. J., J. Dickens, N. T. Hobbs, L. Christensen, and L. Landrum. 2006. Hydrologic, geomorphic and climatic processes controlling willow establishment in a montane ecosystem. *Hydrological Processes* 20:1845–1864.

Cote, S. D., T. P. Rooney, J. P. Tremblay, C. Dussault, and D. M. Waller. 2004. Ecological impacts of deer overabundance. *Annual Review of Ecology Evolution and Systematics* 35:113–147.

Cronan, C. S., and D. F. Grigal. 1995. Use of calcium/aluminum ratios as indicators of stress in forest ecosystems. *Journal of Environmental Quality* 24:209–226.

Czech, B. 2004. A chronological frame of reference for ecological integrity and natural conditions. *Natural Resources Journal* 44:1113–1136.

Dale, V. H., L. A. Joyce, S. McNulty, R. P. Neilson, M. P. Ayres, M. D. Flannigan, P. J. Hanson, L. C. Irland, A. E. Lugo, C. J. Peterson, D. Simberloff, F. J. Swanson, B. J. Stocks, and B. M. Wotton. 2001. Climate change and forest disturbances. *Bioscience* 51:723–734.

De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* 81:3178–3192.

De Leo, G. A., and S. Levin. 1997. The multifaceted aspects of ecosystem integrity. *Conservation Ecology* 1(1): 3.

DeKeyser, E. S., D. R. Kirby, and M. J. Ell. 2003. An index of plant community integrity: Development of the methodology for assessing prairie wetland plant communities. *Ecological Indicators* 3:119–133.

Dobbertin, M. 2005. Tree growth as indicator of tree vitality and of tree reaction to environmental stress: a review. *European Journal of Forest Research* 124:319–333.

Duchesne, L., R. Ouimet, and C. Morneau. 2003. Assessment of sugar maple health based on basal area growth pattern. *Canadian Journal of Forest Research* 33:2074–2080.

Faber-Langendoen, D., J. Rocchio, M. Schafale, C. Nordman, M. Pyne, J. Teague, T. Foti, and P. J. Comer. 2006. *Ecological integrity assessment and performance measures for wetland mitigation*. Arlington: NatureServe.

Faber-Langendoen, D., C. Hedge, M. Kost, S. Thomas, L. Smart, R. Smyth, J. Drake, and S. Menard. 2012. *Assessment of wetland ecosystem condition across landscape regions: A multi-metric approach*. EPA/600/R-12/021a. U.S. Environmental Protection Agency Office of Research and Development, Washington, DC.

Fahrig, L. 2003. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology Evolution and Systematics* 34:487–515.

Fancy, S. G., and R. E. Bennetts. 2012. Institutionalizing an effective long-term monitoring program in the US national park service. In *Design and analysis of long-term ecological monitoring studies*, ed. R. A. Gitzen, J. J. Millspaugh, A. B. Cooper, and D. S. Licht, 481–497. Cambridge: Cambridge University Press.

Fancy, S. G., J. E. Gross, and S. L. Carter. 2009. Monitoring the condition of natural resources in US national parks. *Environmental Monitoring and Assessment* 151:161–174.

Field, C. B., L. D. Mortsch, M. Brklacich, D. L. Forbes, P. Kovacs, J. A. Patz, S. W. Running, and M. J. Scott. 2007. North America. In *Climate change 2007: Impacts, adaptation and vulnerability. Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change*, ed. M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden, and C. E. Hanson, 617–652. Cambridge: Cambridge University Press.

Gage, E., and D. Cooper. 2009. *Historical range of variation assessment for wetland and riparian ecosystems, U.S. Forest service, Rocky Mountain region*. Fort Collins: Colorado State University.

Goodell, L., and D. Faber-Langendoen. 2007. Development of stand structural stage indices to characterize forest condition in Upstate New York. *Forest Ecology and Management* 249: 158–170.

Grace, J. B., T. M. Anderson, H. Olff, and S. M. Scheiner. 2010. On the specification of structural equation models for ecological systems. *Ecological Monographs* 80:67–87.

Groffman, P. M., J. S. Baron, T. Blett, A. J. Gold, I. Goodman, L. H. Gunderson, B. M. Levinson, M. A. Palmer, H. W. Paerl, G. D. Peterson, N. L. Poff, D. W. Rejeski, J. F. Reynolds, M. G. Turner, K. C. Weathers, and J. Wiens. 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? *Ecosystems* 9:1–13.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10:1456–1477.

James-Pirri, M.-J., J. L. Swanson, C. T. Roman, H. S. Ginsberg, and J. F. Heltshe. 2012. Ecological thresholds for salt marsh nekton and vegetation communities. In *Application of threshold concepts in natural resource decision making,* ed. G. Guntenspergen. New York: Springer.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6:21–27.

Karr, J. R. 1991. Biological integrity—a long-neglected aspect of water-resource management. *Ecological Applications* 1:66–84.

Karr, J. R., and E. W. Chu. 1997. Biological monitoring: Essential foundation for ecological risk assessment. *Human and Ecological Risk Assessment* 3:993–1004.

Karr, J. R., and D. R. Dudley. 1981. Ecological perspectives on water quality goals. *Environmental Management* 5:55–68.

Keeton, W. S. 2006. Managing for late-successional/old-growth characteristics in northern hardwood-conifer forests. *Forest Ecology and Management* 235:129–142.

Kennedy, C., J. Wilkinson, and J. Balch. 2003. *Conservation thresholds for land use planners*. Washington, D.C.: Environmental Law Institute.

Kurtz, J. C., L. E. Jackson, and W. S. Fisher. 2001. Strategies for evaluating indicators based on guidelines from the Environmental Protection Agency's Office of Research and Development. *Ecological Indicators* 1:49–60.

Lancia, R. A., C. E. Braun, M. W. Collopy, R. D. Dueser, J. G. Kie, C. J. Martinka, J. D. Nichols, T. D. Nudds, W. R. Porath, and N. G. Tilghman. 1996. ARM! For the future: Adaptive resource management in the wildlife profession. *Wildlife Society Bulletin* 24:436–442.

Landres, P. B., P. Morgan, and F. J. Swanson. 1999. Overview of the use of natural variability concepts in managing ecological systems. *Ecological Applications* 9:1179–1188.

Lemly, J., and J. Rocchio. 2009. *Vegetation index of biotic integrity (VIBI) for headwater wetlands in the southern Rocky Mountains: Version 2.0: Calibration of selected VIBI models*. Fort Collins: Colorado Natural Heritage Program.

MacDonald, J. A., N. B. Dise, E. Matzner, M. Armbruster, P. Gundersen, and M. Forsius. 2002. Nitrogen input together with ecosystem nitrogen enrichment predict nitrate leaching from European forests. *Global Change Biology* 8:1028–1033.

Mack, J. J. 2001. Vegetation index of biotic integrity (VIBI) for wetlands: Ecoregional, hydrogeomorphic, and plant community comparisons with preliminary wetland aquatic life use designations. In *D. o. S. W. Ohio Environmental Protection Agency*, ed. Wetland Ecology Group, 99. Columbus: State of Ohio Environmental Protection Agency.

Mack, J. J. 2007. Developing a wetland IBI with statewide application after multiple testing iterations. *Ecological Indicators* 7:864–881.

Mack, J. J., and M. E. Kentula. 2010. *Metric similarity in vegetation-based wetland assessment methods*. (EPA/600/R-10/140.) Washington, D.C.: U.S. Environmental Protection Agency, Office of Research and Development.

McWilliams, W. H., T. W. Bowersox, P. H. Brose, D. A. Devlin, J. C. Finley, K. W. Gottschalk, S. Horsley, S. L. King, B. M. LaPoint, T. W. Lister, L. H. McCormick, G. W. Miller, C. T. Scott, H. Steele, K. C. Steiner, S. L. Stout, J. A. Westfall, and R. L. White. 2005. Measuring tree seedlings and associated understory vegetation in Pennsylvania's forests. In *Proceedings of the Fourth Annual Forest Inventory and Analysis Symposium. General Technical Report NC-252*, ed. R. E. McRoberts, G. A. Reams, P. C. Van Deusen, W. H. McWilliams, and C. J. Cieszewski, 21–26. St. Paul: U.S. Department of Agriculture, Forest Service, North Central Research Station.

Miller, S. J., D. H. Wardrop, W. M. Mahaney, and R. R. Brooks. 2006. A plant-based index of biological integrity (IBI) for headwater wetlands in central Pennsylvania. *Ecological Indicators* 6:290–312.

Miller, D. M., S. P. Finn, A. Woodward, and A. Torregrosa. 2010a. Conceptual models for landscape monitoring. In *Conceptual ecological models to guide integrated landscape monitoring of the Great Basin*, ed. D. M. Miller, S. P. Finn, A. Woodward, A. Torregrosa, M. E. Miller, D. R. Bedford, and A. M. Brasher, 1–12. Reston: United States Geological Survey.

Miller, K. M., G. L. Tierney, and B. R. Mitchell. 2010b. *Northeast temperate network forest health monitoring report: 2006–2009*. (Natural Resource Report NPS/NETN/NRR-2010/206.) Fort Collins: National Park Service.

Mitchell, B. R., W. G. Shriver, F. Dieffenbach, T. Moore, D. Faber-Langendoen, G. Tierney, P. Lombard, and J. P. Gibbs. 2006. *Northeast temperate network vital signs monitoring plan*. Woodstock: National Park Service, Northeast Temperate Network.

Mitsch, W. J., and J. G. Gosselink. 2007. *Wetlands*. 4th ed. Hoboken: Wiley.

National Park Service. 2012. Guidance for designing an integrated monitoring program. Natural Resource Report NPS/NRSS/NRR–545. National Park Service, Fort Collins, Colorado.

Nichols, J. D., M. J. Eaton, and J. Martin. 2012. Thresholds for conservation and management: structured decision making as a conceptual framework. In *Application of threshold concepts in natural resource decision making,* ed. G. Guntenspergen.. New York: Springer.

Noon, B. R. 2003. Conceptual issues in monitoring ecological resources. In *Monitoring ecosystems: Interdisciplinary approaches for evaluating ecoregional initiatives,* ed. D. E. Busch, and J. E. Trexler, 27–72. Washington, D.C.: Island Press.

Oakley, K. L., L. P. Thomas, and S. G. Fancy. 2003. Guidelines for long-term monitoring protocols. *Wildlife Society Bulletin* 31:1000–1003.

Olden, J. D., and N. L. Poff. 2003. Toward a mechanistic understanding and prediction of biotic homogenization. *American Naturalist* 162:442–460.

Olden, J. D., and T. P. Rooney. 2006. On defining and quantifying biotic homogenization. *Global Ecology And Biogeography* 15 (2): 113–120.

Olsen, A. R., J. Sedransk, D. Edwards, C. A. Gotway, W. Liggett, S. Rathbun, K. H. Reckhow, and L. J. Young. 1999. Statistical issues for monitoring ecological and natural resources in the United States. *Environmental Monitoring and Assessment* 54:1–45.

Parrish, J. D., D. P. Braun, and R. S. Unnasch. 2003. Are we conserving what we say we are? Measuring ecological integrity within protected areas. *Bioscience* 53:851–860.

Peet, R. K., T. R. Wentworth, and P. S. White. 1998. A flexible, multipurpose method for recording vegetation composition and structure. *Castanea* 63:262–274.

Risvold, A. M., and R. W. Fonda. 2001. Community composition and floristic relationships in montane wetlands in the North Cascades, Washington. *Northwest Science* 75:157–167.

Schweiger, E. W., E. Gage, D. Shorrock, I. W. Ashton, J. Burke, D. Pillmore, and M. Britten. 2010a. Draft. Wetland ecological integrity. Main narrative version 2.1. Fort Collins: National Park Service.

Schweiger, E. W., J. Grace, K. Driver, D. Schoolmaster Jr., D. Cooper, G. Guntenspergen, D. Shorrock, I. Ashton, J. Burke, L. O'Gan, and M. Britten. 2010b. Draft. Wetland ecological integrity protocol: Rocky Mountain Network report on the 2007–2009 protocol development pilot in Rocky Mountain National Park (NPS/XXXX/NRTR—20XX/XXX). Fort Collins: National Park Service.

Stevens, D. L. Jr., and A. R. Olsen. 2004. Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* 99:262–278.

Stevens, S., B. R. Mitchell, M. Brown, and P. Campbell. 2010. *Strategy for enhanced monitoring of natural resource condition in North Atlantic coastal parks to address the effects of rapid climate change*. (Natural Resource Report NPS/NCBN/NRR-2010/272.) Fort Collins: National Park Service.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. *Ecological Applications* 16:1267–1276.

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the North American Benthological Society* 27:878–891.

Sweetapple, P. J., and G. Nugent. 2004. Seedling ratios: A simple method for assessing ungulate impacts on forest understories. *Wildlife Society Bulletin* 32:137–147.

Tierney, G. L., D. Faber-Langendoen, B. R. Mitchell, W. G. Shriver, and J. P. Gibbs. 2009. Monitoring and evaluating the ecological integrity of forest ecosystems. *Frontiers in Ecology and the Environment* 7:308–316.

Tierney, G. L., B. R. Mitchell, K. M. Miller, J. Comiskey, A. Kozlowski, and D. Faber-Langendoen. 2010. *Northeast Temperate Network long-term forest monitoring protocol: 2010 revision*. Fort Collins: National Park Service.

Unnasch, R. S., D. P. Braun, P. J. Comer, and G. E. Eckert. 2009. *The ecological integrity assessment framework: A framework for assessing the ecological integrity of biological and ecological resources of the National Park System*. Arlington: NatureServe.

Urquhart, N. S., and T. M. Kincaid. 1999. Designs for detecting trend from repeated surveys of ecological resources. *Journal of Agricultural Biological and Environmental Statistics* 4: 404–414.

Wilhelm, G. S., and L. A. Masters. 1995. *Floristic quality assessment in the Chicago region and application computer programs*. Lisle: The Morton Arboretum.

Chapter 11
Use, Misuse, and Limitations of Threshold Indicator Taxa Analysis (TITAN) for Natural Resource Management

Ryan S. King and Matthew E. Baker

Abstract Detection of ecological thresholds has broad relevance to management of ecosystems. However, ecological community data present a distinct problem because current statistical methods used for identifying thresholds were not developed for analysis of multiple, individual species abundances. We developed a new method, Threshold Indicator Taxa ANalysis (TITAN), specifically to deal with some of the limitations of existing methods for estimating community thresholds. Our objectives in this chapter are to (1) summarize the theoretical basis for the method and related methods, (2) provide a brief overview of how it works, (3) use a real data set to illustrate an application of the method, and (4) conclude the chapter by addressing several issues related to the appropriate use of the method, misconceptions about how it works or what the results mean, and limitations that could lead to erroneous conclusions. We explain that step-function conceptualizations of community thresholds are not sufficiently inclusive of all the response forms that satisfy threshold criteria, how gradual responses of univariate community metrics do not rule out community thresholds, and that linear regression techniques do not provide an adequate test for the absence of thresholds, especially in the presence of long environmental gradients. We note substantial misunderstanding in the recent literature regarding appropriate use and interpretation of statistical change points identified by taxon-specific analysis in TITAN, that univariate community metrics are inappropriate response variables for such analyses, and that extreme variation in the density of the sample distribution can affect results of any method, including TITAN. We end by reminding users that despite the additional insight it brings to community analysis, TITAN is neither a causal analysis nor a black box for developing regulatory criteria. Instead, we intend TITAN to complement current analytical approaches, while highlighting assumptions and flaws in the broader paradigms in which they are often applied.

R. S. King (✉)
Center for Reservoir and Aquatic Systems Research, Department of Biology,
Baylor University, Waco, TX 76798, USA
e-mail: Ryan_S_King@baylor.edu

M. E. Baker
Department of Geography and Environmental Systems, University of Maryland-Baltimore
County, Baltimore, MD 21250, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_11,

Keywords Biodiversity conservation · Community analysis · Indicator species · Ecological thresholds · Conservation · Environmental criteria

Introduction

There is a growing interest in the application of ecological thresholds for natural-resource management (e.g., Huggett 2005; Suding and Hobbs 2009; Dodds et al. 2010). Indeed, threshold detection has broad relevance to management of aquatic ecosystems, such as conservation (DeLuca et al. 2008; Hilderbrand et al. 2010), biological invasions (King et al. 2007), ecosystem restoration (Walsh et al. 2005a; Martin et al. 2009; Clements et al. 2010), development of numerical water-quality criteria (King and Richardson 2003; Soranno et al. 2008), ecosystem management (Richardson et al. 2007), and forecasting effects of climate change (Smol et al. 2005).

Despite the recent interest in ecological thresholds, application of the threshold concept to aquatic-resource management remains tentative, if not contentious (e.g., Gaiser et al. 2008; Richardson et al. 2008). Threshold estimation depends upon the selection of a response variable, assumed shape of the response, and appropriateness of the corresponding statistical model, any of which may contribute to different interpretations regarding the location of a threshold or whether a threshold even exists (e.g., Walsh et al. 2005b; Moore and Palmer 2005; Dodds et al. 2010).

Ecological community data present a distinct problem because current statistical methods used for identifying thresholds were not developed for simultaneous analysis of multiple, individual species abundances (Brenden et al. 2008; Anderson et al. 2009). The vast majority of taxa in community data sets have low occurrence frequencies (i.e., do not occur in a large proportion of the sample units) and have highly variable abundances which make their individual response difficult to fit with various forms of regression analysis typically used for threshold detection (e.g., piecewise regression, Toms and Lesperance 2003, significant zero crossings, Sonderegger et al. 2009). Consequently, most investigators aggregate community data into univariate responses, selecting a priori attributes that presumably represent an important facet of community structure, such as the number of taxa, or deriving synthetic variables from multivariate analysis of taxa composition among sites (e.g., dissimilarity metrics, ordination axes; King and Richardson 2005; Walsh et al. 2005a). While aggregating taxa into one or more response variables may, in some instances, increase the community signal in response to anthropogenic gradients, it also likely obscures nonlinear changes in one or more taxa, potentially underestimating or misrepresenting the effect of an anthropogenic gradient on ecological communities. Thus, evaluating ecological community thresholds with existing approaches often involves undesirable generalities, loss of information, or assumptions regarding taxon-specific responses.

We developed a new method, Threshold Indicator Taxa ANalysis (TITAN), specifically to deal with some of the limitations of existing methods for estimating community thresholds (Baker and King 2010). Since the publication of the method, we have published four additional papers describing applications of the method (King and Baker 2010; King and Baker 2011a; King et al. 2011; Bernhardt et al. 2012) and a detailed explanation of when to use it, how to interpret it, and what it does and does not do, largely in response to misunderstanding and misrepresentation of the approach by a few investigators (Baker and King 2013). Our objectives in this chapter are to (1) summarize the theoretical basis for the method and related methods, (2) provide a brief overview of how it works, (3) use a real data set to illustrate an application of the method, and (4) conclude the chapter by addressing several issues related to the appropriate use of the method, misconceptions about how it works or what the results mean, and limitations that could lead to erroneous conclusions.

Community Thresholds and Novel Gradients

We define an ecological community threshold to mean that the frequency and/or abundance of taxa will increase or decrease sharply at some level along an environmental gradient, such that an incremental change in a driver such as urban intensity, toxic compounds, or any number of anthropogenic variables results in a disproportionately large change in community structure relative to elsewhere along the gradient (Baker and King 2010). This definition does not necessarily imply a catastrophic, vertical increase or decrease in the response, preceded and followed by zones of minimal change (i.e., a step function), which is unrealistic for many ecological responses and corresponds more closely to a regime shift or alternative stable state (see reviews by Sheffer and Carpenter 2003 and Andersen et al. 2009). However, this interpretation is one that we have frequently encountered in discussions about ecological thresholds with other investigators. Moreover, our definition does not preclude the possibility that some taxa may decline prior to or following a synchronous decline in multiple taxa, but it is grounded in the fact that many interacting species may be influenced in similar ways by an environmental driver, either physiologically or through disruption of interspecific interactions (Økland et al. 2009).

Our definition of an ecological community threshold is particularly linked to and relevant in the context of anthropogenic changes to natural environments (Fig. 11.1). The physical and chemical conditions of many modern ecosystems increasingly diverge from environments known to have existed at any time in the history of Earth (Fox 2007). These "no-analog" or novel environments can lead to wholesale changes in community structure caused by a cascade of intra- and interspecific mechanisms ranging from extirpation of species due to physiological stress, decoupling of positive interactions such as facilitation, relaxation of resource limitations on some while imposing new ones on others, and altering competition or predation (Hobbs et al. 2006; Williams and Jackson 2007). Novel environmental gradients likely represent a strong selective pressure favoring native taxa that are less specialized, have greater physiological plasticity, or facilitate invasion of adaptive nonnative taxa (Stralberg et al. 2009). Species replacement results in novel biotic communities that may be difficult to manage, afford fewer ecosystem services, and may not respond to habitat

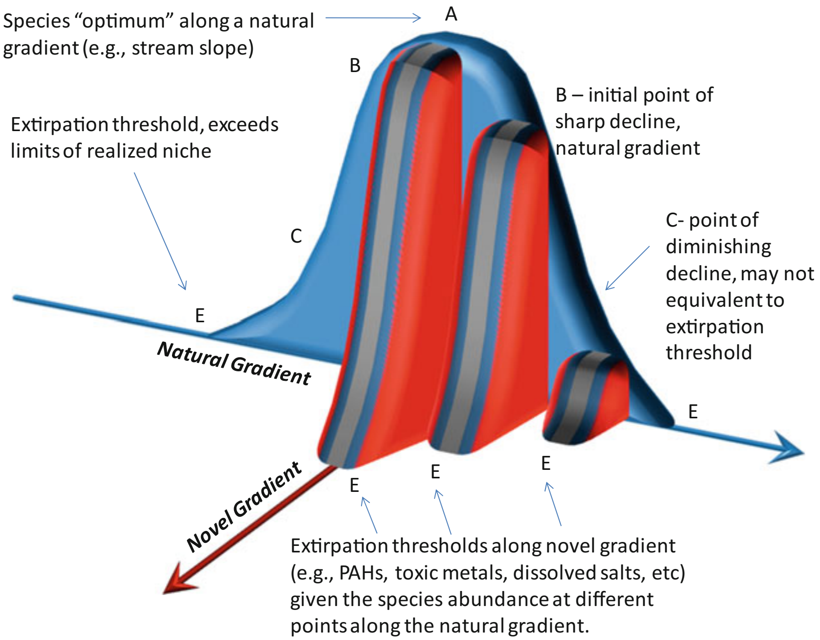


Fig. 11.1 Conceptual diagram illustrating theoretical responses of a single taxon to a natural and novel environmental gradient. The *natural gradient* represents a long gradient such as latitude. The *solid curve* along this gradient represents the range of frequency and/or abundance expected with increasing values along the axis. The *novel gradient* represents a "new" dimension and is theoretically independent (or orthogonal) to the natural gradient. In this example, the *novel gradient* has a negative effect on the taxon, but depending upon the location along the *natural gradient*, the taxon may respond differently to the novel gradient

restoration efforts (Palmer et al. 2009; Clements et al. 2010). Thus, characterizing taxon-specific responses to novel anthropogenic gradients is important for detecting critical levels of alteration, understanding mechanisms of biodiversity loss, identifying adaptive traits that confer success, assessing changes to ecosystem function, and shaping restoration strategies.

A good illustration of a community response to novel gradients is the coincident decline of stream macroinvertebrate communities in response to anthropogenic changes. Although it is widely debated whether stream communities truly exhibit "threshold" responses, we suggest there is strong theoretical and empirical evidence that they do. Many streams, particularly in old landscapes such as the Appalachian Mountains, USA, exhibit high biodiversity as a result of subtle, yet critical differences in stream flow velocities and material transport through time and space (e.g., Poff 1997). Diverse microhabitats have resulted in extensive adaptive radiation of many stream-dwelling taxa (Vinson and Hawkins 1998), whereas moderate frequency and magnitude of hydrological disturbances have maintained high levels of species richness at a local scale (Connell 1978). Facilitation among taxa is also documented in streams (Cardinale et al. 2002). Consequently, lotic species have coevolved to possess unique morphological, behavioral, and physiological adaptations that correspond to an often narrow range of environmental conditions. Small functional niches undoubtedly render many species intolerant of conditions that fall outside those experienced in evolutionary time (sensu Shelford 1913). Thus, taxa sensitive to the novel environment are selected against, sharply decline, and eventually disappear (Fig. 11.2a).

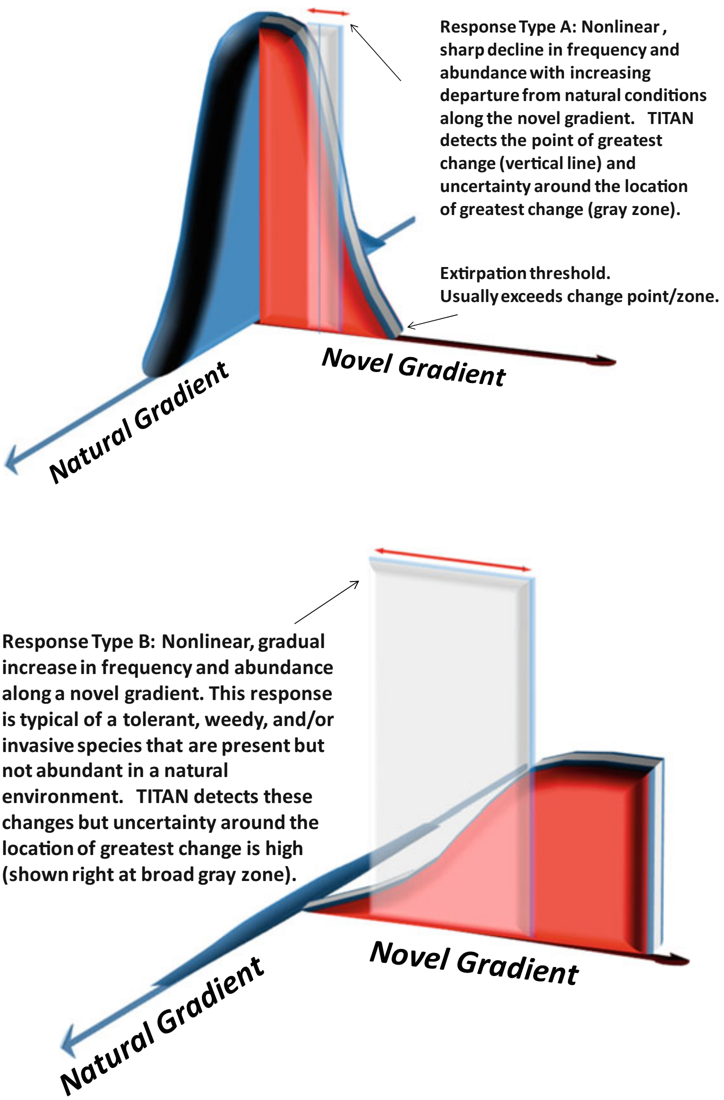
The theoretical sensitivity posed above appears borne out in some empirical responses to anthropogenic stressor gradients where marked synchrony occurs in the decline of sensitive taxa (e.g., King et al. 2011; Bernhardt et al. 2012). However, synchrony does not mean that all taxa exhibit exactly the same response function, but that their greatest declines (change in frequency and abundance) all occur within a narrow range of the environmental gradient. Many of the responding taxa may occupy distinct trophic positions and thus exhibit different responses, which is why their coincident declines are strong evidence of community organization.

In contrast to the synchronous declines of sensitive taxa, positive-responding taxa may or may not increase synchronously, and in our studies to date appear to increase gradually in frequency and abundance at various levels of increasing novel conditions (e.g., Baker and King 2010; King et al. 2011). The lack of synchronous change points, that is, locations along the novel gradient that result in the greatest amount of change in the response, and greater uncertainty in the location of individual taxa change points implies that positive responding taxa probably do not represent well-organized communities, but rather are composed of historically native taxa that either directly (resource subsidy) or indirectly (e.g., realized niche expansion, reduced competition or predation) benefited from it (Fig. 11.2b).

These theoretical and empirical responses represent the underlying basis for the development of TITAN. The following section details in brief how the method works and can be used to identify change points in individual taxa responses as well as provide an assessment of the degree of synchrony in multiple taxa responses as evidence for an ecological community threshold.

Threshold Indicator Taxa ANalysis (TITAN): What Is It and How Does It Work?

TITAN is an analytical approach for identifying and distinguishing threshold-type responses in ecological community data sets. Its basic premise is that community response to environmental gradients, particularly novel environmental gradients, is best detected empirically by aggregating the responses of individual taxa rather than seeking change in community summary metrics (King and Baker 2010). This is consistent with what Ferrier and Guisan (2006) identify as a “Predict first, assemble later” approach to modeling communities. Analyzing individual taxa in this way requires confronting the considerable variability in numerical abundance data and uncertainty typically associated with sparse community data matrices, where many



Response Type A: Nonlinear, sharp decline in frequency and abundance with increasing departure from natural conditions along the novel gradient. TITAN detects the point of greatest change (vertical line) and uncertainty around the location of greatest change (gray zone).

Extirpation threshold. Usually exceeds change point/zone.

Response Type B: Nonlinear, gradual increase in frequency and abundance along a novel gradient. This response is typical of a tolerant, weedy, and/or invasive species that are present but not abundant in a natural environment. TITAN detects these changes but uncertainty around the location of greatest change is high (shown right at broad gray zone).

Fig. 11.2 Conceptual diagram illustrating theoretical responses of different taxa to a novel environmental gradient. The x–y intercept represents absence of the anthropogenic stressor or a level that falls within the normal range of conditions experienced during evolutionary time. The response of Taxon A represents no change in its distribution along the gradient until a critical change point or zone is reached (*shaded region*), which leads to a nonlinear decline and eventual extirpation at a level beyond the initial change point. Taxon B represents a native taxon that is tolerant of the novel gradient and either directly (resource subsidy) or indirectly (e.g., realized niche expansion, reduced competition or predation) benefits, resulting in an indeterminate increase in its frequency and abundance among sites with increasing values of the gradient. Taxon B could also be an invasive taxon that is able to cross ecosystem boundaries and proliferate because of the altered and more favorable novel conditions that previously limited its distribution taxa occur infrequently and irregularly in samples measured across time and space. Confounding detection efforts, taxon abundances are often observed to vary as a result of stochastic processes, sampling error, and strong correlation among unmeasured environmental factors (e.g., King et al. 2005, 2011). Fortunately, this autocorrelation means that the majority of variation in community structure can be explained with relatively few dimensions. It also means that some taxa (though which ones are not always clear) are likely to respond similarly to strong environmental drivers. This redundancy is the basis for multivariate ordination methods that presuppose a unimodal response model, but such methods are designed to detect species optima and do not necessarily describe change. At its core, TITAN is a filtering process that seeks to separate true and reliable response patterns of change from the high levels of unexplained variability, or noise, in community data.

TITAN works by integrating a relatively simple and elegant measure of association in taxon abundance with a nonparametric technique for detecting change. Indicator species analysis (Dufrene and Legendre 1997) uses abundance-weighted occurrence frequency to describe association between a particular taxon and groups of samples defined by their order along an environmental gradient. Baker and King (2010) provide explicit detail; however, for an intuitive understanding, it is perhaps useful to consider the patterns of abundance in Fig. 11.2. A taxon's indicator value (IndVal) at any position along the gradient is a function of the relative abundance on either side of a partition, weighted by its occurrence within each partition (i.e., the product of both). In TITAN, the larger IndVal on each side of a partition is retained and compared across partitions to find the value of the environmental gradient that results in the greatest change in taxon abundance and frequency within the observed sample. This value is a change point. However, a change point is not necessarily a "statistical threshold" per se. It is simply the value of x that best partitions the data so that difference in frequency and abundance is maximized. This analysis is repeated for each taxon to provide a set of observed change points and the direction of that change.

To facilitate comparison across taxa, TITAN compares each taxon's maximum IndVal score to those expected if the same sampled abundances were randomly distributed across the environmental gradient. A good indicator species is one that occurs frequently, so that changes in its abundance are easy to detect, but that is not the only kind of response worth noting. IndVal scores will always be small for rare, variable, or sensitive taxa, even though they can nonetheless represent important changes within a community. By comparison to the average IndVal scores derived by random permutation, TITAN standardizes measures of change for any given taxon to units of standard deviation (z scores; Baker and King 2010). Standardization emphasizes observed changes for each taxon relative to their own patterns of variability in abundance and occurrence.

To better understand uncertainty surrounding the observed change points, TITAN employs a bootstrap resampling technique (resampling with replacement; Manly 1997). However representative or large a sample may be, it remains only one estimate of true underlying population patterns, and given a taxon-specific pattern of abundance, another sample may yield an altogether different change point. Thus, for every taxon the entire analysis is repeated many times (we recommend a minimum of 500), each with a slightly different collection of replicates drawn from the original sample set to obtain a distribution of potential change points.

Information provided by the bootstrap is critical for interpreting results in TITAN. In addition to the location and dispersion of the change-point distribution, TITAN evaluates consistency in the response direction as purity, and the frequency of a strong response magnitude as reliability (Baker and King 2010). Combined with a minimum occurrence frequency, these diagnostic indices are used as filters to help distinguish the signal produced by indicator taxa responses from stochastic noise along the gradient. This filtering is part of what distinguishes TITAN from many other multivariate techniques based on weighted averaging or dissimilarity.

Once indicator taxa have been identified, TITAN provides information that can be used to identify a potential community-level threshold. A plot of filtered indicator taxa showing change-point quantiles from bootstrap replicates provides evidence regarding the existence of synchronous changes in the community structure. Because the magnitude of all responses is standardized across taxa as z scores, their sum reflects the magnitude of community change at any point along the gradient. Distinct peaks in the sum(z) curve (maxima) plotted across the environmental gradient are another indication of coincident change in community structure. When bootstrap replicates used to compare the location of the sum(z) maxima across many sample replicates show a narrow band, this constitutes evidence for a threshold response (Baker and King 2010; King et al. 2011).

Case Study: Macroinvertebrate Community Response to a Phosphorus Gradient in the Everglades

The Everglades (Florida, USA) is a large subtropical wetland that has experienced significant anthropogenic changes in the past several decades. Modifications to hydrology, fire frequency and intensity, and other environmental factors all have played a role in the alteration structure and functioning of the Everglades ecosystem, but phosphorus (P)-enriched runoff from the Everglades Agricultural Area (EAA) is widely viewed as the primary stressor (SFWMD 1992). An extensive canal-and-levee system has compartmentalized most of the remaining Everglades ecosystem, a system that also serves as a conduit for P from the EAA. Water-control structures along the canals function as point sources of P to downstream portions of the wetland ecosystem. In areas near water-control structures, P has been found to be largely responsible for facilitating invasion of cattail, vines, willows, and other plants that are strongly limited by P (King et al. 2004; Richardson et al. 2007). Periphyton (floating and attached mats of algae and bacteria), macroinvertebrate, and fish communities have also changed dramatically in areas with even modest levels of P enrichment. Thus, P enrichment in the Everglades serves as an excellent example of a novel environmental gradient.

The data to be used for this example are from a previous study designed to identify a concentration of surface-water total P (TP) that corresponded to abrupt changes in

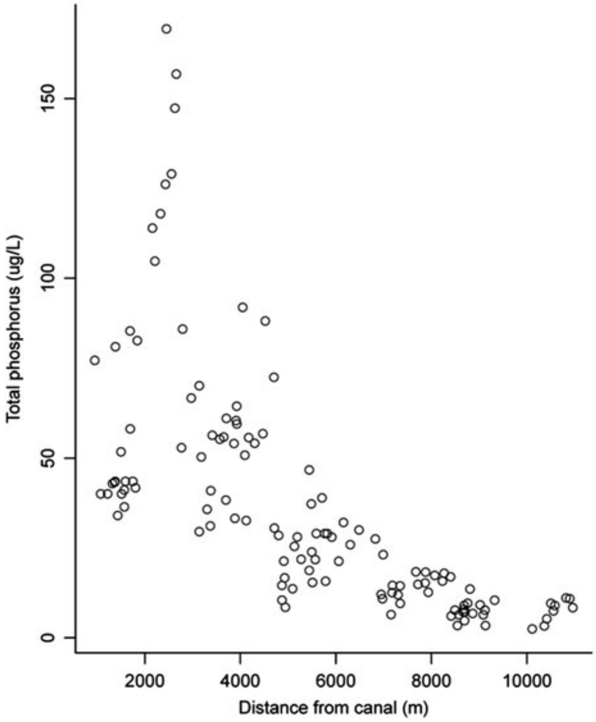


Fig. 11.3 Relationship between surface-water total phosphorus (TP) and distance from the canal in the northern Everglades

macroinvertebrate species composition in the Florida Everglades, USA (King and Richardson 2003). Macroinvertebrate species composition (no/m^2, 164 taxa, species or morphospecies-level taxonomy) was measured from 126 marsh sampling stations along a 10-km TP gradient. This gradient corresponds closely to proximity canal inflow structures, the point-sources of P to the interior marsh. Concentrations of TP in the data set ranged from < 10 to > 100 µg/L. The authors used several community variables and estimated TP change points using a univariate method called nonparametric change-point analysis (nCPA), a binary partitioning method that is computationally similar to regression tree analysis but incorporates bootstrapping for confidence interval (CI) estimation and allows specification of distribution families for response data (Qian et al. 2003). The resulting change points from their analysis ranged from approximately 10–25 µg/L TP, and authors concluded that TP $>$ 12–15 µg/L was likely to correspond to ecologically significant changes in taxonomic composition.

In our example, we reanalyzed macroinvertebrate community response to TP as well as a second variable, distance from the canal inflow structures (canal, m), that was not analyzed by King and Richardson (2003). Because canal represents the source of TP, and TP responds in a nonlinear, negative manner to increasing distance from the canal (Fig. 11.3), taxa that respond in a strongly negative direction to increasing TP presumably should respond *positively* to increasing distance from the canal. Thus, taxa that are classified as negative indicators (z−) for TP should be positive indicators (z+) for canal, assuming there is good correspondence in the explanatory power between the variables. To be clear, this is by no means a causal analysis but in this case substantial experimental work has been done to validate observational results implying TP as a driver of community thresholds (King and Richardson 2003). We will discuss the issue of confounded, intercorrelated variables in the concluding section of this chapter.

Prior to TITAN analysis, we $\log_{10}(x+1)$ transformed taxa abundances to reduce the influence of highly variable taxa on indicator score calculations, although it is certainly acceptable to use untransformed abundances in this nonparametric analysis. Taxa with < 5 occurrences were deleted (following previous analyses of these data). We ran TITAN using each variable separately to compare the relative strength of community response (sum(z)), number of threshold indicator taxa, and the correspondence between individual taxa responses to both variables. The minimum split size (minimum number of observations required on each side of a candidate change point) was set to 5, number of permutations was set to 250, and the number of bootstrap replicates was set to 100 (although we suggest using the default number of 500 in most cases for more precise confidence limits and purity/reliability estimates). Step-by-step instructions on how to load TITAN, import data, run the analysis, and graph results are provided online (http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2009.00007.x/suppinfo).

The first sets of results to examine are the individual threshold indicator taxa change points, confidence limits, and other diagnostic statistics. Table 11.1 reports a list of taxa that are deemed highly probable indicators of decline in response to a novel gradient, in this case, positive change with increasing distance from the canal (taxa classified as positive indicators, or z+) and negative change with increasing levels of surface-water TP. Recall that these predictors are negatively related so it is important to recognize that taxa that are negatively affected by a novel gradient can be either classified as negative (z−) or positive (z+) indicators depending upon the direction of the values of the gradient. The output from TITAN will not make value judgments about which direction is "good" or "bad" so it is imperative that the investigator be aware of this fact when sorting through output.

There are several columns of results included in Table 11.1. The first, Freq (Frequency of occurrence) merely summarizes how many times a taxon occurred in the data set. Note that a few taxa only occurred five times, the minimum requirement for inclusion in this analysis. The next column (Obs.cp) is the observed value of the predictor that resulted in the maximum indicator z score in the data set. The next two columns report the raw IndVal and its standardized z score (z). Note that some taxa have relatively low IndVal scores (recall that IndVal is scaled from 0–100) but relatively high z scores. The reason for this is that IndVal does not reflect the magnitude of difference in frequency and abundance between the groups of samples on either side of the change point, whereas the z score does. The z does this by subtracting the average IndVal score obtained by randomly reshuffling the data (250 permutations) from the observed IndVal and dividing this difference by the standard deviation of

Table 11.1 Indicator taxa identified by TITAN in response to distance from the canal and total phosphorus

Distance from canal (m)								Total phosphorus (ug/L)					
Taxon code	FrEq	Obs. cp	IndVal	z	CI-90 %	Purity	Rel05	Obs. cp	IndVal	z	CI-90 %	Purity	Rel05
Ablarham	15	6,844	28.25	8.64	6,202	1.00	1.00	15.81	31.08	10.46	20.08	1.00	1.00
Acentria	7	6,327	11.82	4.29	5,094	1.00	0.95	14.56	14.65	5.00	28.80	1.00	0.95
Aphaopac	104	4,425	64.15	9.08	3,467	1.00	1.00	36.91	64.16	8.54	42.40	1.00	1.00
Beartrun	66	4,300	72.11	14.49	3,510	1.00	1.00	29.33	66.55	13.54	38.79	1.00	1.00
Bezzsp2	41	5,302	53.84	11.70	5,169	1.00	1.00	16.27	50.56	10.38	30.90	1.00	1.00
Bratunid	77	3,453	63.67	10.05	3,335	1.00	1.00	32.43	53.61	7.54	41.14	1.00	1.00
Caendimi	90	4,175	72.73	12.62	2,418	1.00	1.00	30.33	63.82	10.19	40.17	1.00	1.00
Calanoid	26	3,510	30.95	6.42	3,467	1.00	1.00	29.07	25.63	4.94	35.07	0.98	0.95
Callflor	**89**	**2,075**	**70.08**	**6.91**	**1,733**	**1.00**	**1.00**						
Celiepon	14	6,844	20.86	5.40	6,202	0.99	0.95	13.65	25.51	7.12	16.27	1.00	0.98
Ceriodap	**39**	**8,081**	**48.38**	**7.62**	**6,994**	**0.99**	**0.99**						
Cernotin	15	6,994	40.54	14.03	7,789	1.00	1.00	9.04	50.07	11.77	15.11	1.00	1.00
Cladopel	37	3,884	36.29	5.68	3,508	1.00	1.00	32.43	31.12	4.67	43.21	1.00	0.99
Cladotan	14	7,956	42.71	13.38	7,831	1.00	1.00	10.92	35.70	9.94	14.56	1.00	1.00
Coryspd	5	8,832	14.59	6.35	6,994	1.00	0.99	13.16	14.29	6.24	15.14	1.00	0.95
Cyprbrev	13	8,807	45.29	12.76	7,789	1.00	1.00	13.65	36.11	12.26	15.37	1.00	1.00
Dasyhele	**75**							**14.62**	**44.44**	**3.49**	**68.52**	**0.99**	**0.96**
Dicromod	21	5,277	23.06	4.76	3,506	0.98	0.96	33.70	22.63	5.04	40.17	0.99	0.99
Enalcivi	70	7,956	72.81	13.60	5,225	1.00	1.00	18.61	70.31	13.21	30.98	1.00	1.00
Ferrissi	59	4,175	48.31	6.29	3,403	1.00	1.00	30.33	44.05	5.44	37.01	0.99	0.99
Fittsert	18	4,300	23.68	6.15	3,861	1.00	1.00	9.64	23.02	3.98	35.24	0.99	0.95
Hyalazte	125	3,585	57.29	7.04	3,467	1.00	1.00	30.33	55.56	5.93	60.07	1.00	1.00
Hydbisp2	42	5,094	47.86	10.26	4,300	1.00	1.00	33.00	47.10	9.51	43.16	1.00	1.00
Ischnura	8	8,382	24.34	8.39	7,831	1.00	1.00	7.42	33.31	9.51	12.36	1.00	0.99
Labrneop	14	6,202	22.37	6.20	5,169	1.00	1.00	13.16	25.17	6.66	28.08	1.00	0.98
Laevpeni	**94**	**2,418**	**66.42**	**7.52**	**1,568**	**0.99**	**0.99**						
Micrdila	**58**	**4,175**	**54.94**	**9.07**	**3,467**	**1.00**	**1.00**						
Nanoalte	6	8,682	23.08	8.98	7,931	1.00	0.99	8.48	31.58	11.19	9.62	1.00	0.99
Nilothau	10	8,807	36.60	12.19	7,956	1.00	1.00	10.92	34.48	11.91	12.06	1.00	1.00

Table 11.1 (continued)

Distance from canal (m)								Total phosphorus (ug/L)					
Taxon code	FrEq	Obs. cp	IndVal	z	CI-90 %	Purity	Rel05	Obs. cp	IndVal	z	CI-90 %	Purity	Rel05
Oecetspe	22	6,535	39.48	10.73	5,302	1.00	1.00	17.21	34.22	9.72	29.07	1.00	1.00
Oxyethir	22	7,956	62.88	18.84	7,413	1.00	1.00	14.62	45.63	13.34	14.77	1.00	1.00
Palapalu	83	4,175	84.18	16.04	3,585	1.00	1.00	28.08	71.97	13.34	36.91	1.00	1.00
Paracala	12	8,081	34.15	10.09	6,224	1.00	1.00	10.52	31.12	8.27	14.56	1.00	1.00
Parakspc	35	6,744	79.55	21.67	6,317	1.00	1.00	14.62	76.21	20.94	17.76	1.00	1.00
Parapony	11	7,956	35.48	12.69	7,931	1.00	1.00	12.06	34.38	11.92	12.66	1.00	1.00
Paraspb	21	8,382	73.54	18.72	7,953	1.00	1.00	10.92	59.88	17.32	14.62	1.00	1.00
Physcube	67	5,739	61.40	10.19	5,035	1.00	1.00	31.68	56.51	9.29	39.55	1.00	1.00
Plansp1	69	6,327	51.08	6.43	6,202	1.00	0.99	14.62	52.28	6.03	17.76	1.00	0.99
Polyhalt	14	6,844	32.56	11.86	6,327	1.00	1.00	16.87	30.43	12.11	16.87	1.00	1.00
Polypspa	21	6,844	46.10	14.86	6,202	1.00	1.00	14.62	43.71	13.67	21.61	1.00	1.00
Proclad	8	9,719	52.38	11.94	7,981	1.00	1.00	6.48	48.05	11.24	12.71	1.00	1.00
Spongill	13	6,077	24.53	8.09	5,277	1.00	1.00	8.22	34.80	7.93	21.92	1.00	1.00
Tanylimn	18	5,244	23.74	6.16	5,060	0.99	0.99	29.83	21.29	4.76	33.70	1.00	0.99
Tanyspg	**85**							**40.11**	**50.79**	**4.68**	**48.60**	**1.00**	**1.00**
Tanyspr	46	6,744	96.19	25.14	6,177	1.00	1.00	14.62	86.23	22.61	25.73	1.00	1.00
Tanyspt	5	7,413	13.89	5.87	6,869	0.99	0.95	7.64	25.87	7.74	9.43	1.00	0.95
Trichorx	11	6,844	22.20	8.02	5,706	1.00	1.00	21.38	20.75	7.45	24.71	1.00	1.00
Unionico	5	8,807	19.01	7.85	7,956	1.00	0.97	9.64	20.00	8.55	10.72	1.00	0.99

Taxa are listed in alphabetic order by code name. Frequency specifies the number of times each taxon occurred in the data set (out of 126 sample units). The observed change point (Obs.cp) is the value of the predictor that resulted in the maximum observed indicator score (IndVal) among all possible change point values. The IndVal score is scaled from 0–100, with 100 indicating a taxon that occurred in all of the samples above or below a change point value and in none of the samples on the other side of the change point, whereas the z score is scaled to reflect how different the IndVal was from values that were obtained by random reshuffling of the data (250 permutations). The CI-90 % is the value where 90 % of the change points estimated among all of the bootstrap replicates fell either below (for a declining, z− taxon) or above (for an increasing, z+ taxon). Purity is proportion of the bootstrap replicates that correctly assigned the taxon as an increaser or decreaser, whereas reliability (rel05) is the proportion of the bootstrap replicates that resulted in a p-value of < 0.05. Only indicator taxa that passed screening criteria ($p < 0.05$, purity > 0.95, reliability (rel05) > 0.95) for at least one variable are shown; taxa that passed criteria for only one variable are shown in bold the random IndVals. The advantage of this standardization is that a taxon with relatively few occurrences but very strong fidelity to one end of the gradient can achieve a large z score despite having a relatively small IndVal. Nevertheless, both statistics are useful measures of taxa responses to the gradient.

The next column, CI-90 %, reports metrics of uncertainty about the change point location of individual taxa. The 90 % CI is the 90th quantile of the distribution of change points computed from 100 bootstrap replicates. In the case of an increasing taxon (z+, distance from the canal), this is the 90th quantile on the left end of the distribution, or the lowest level of the predictor where change points begin to be detected using bootstrapping. In the case of a decreasing taxon, the 90th quantile is near the highest level of the predictor where change points are detected using bootstrapping. These CIs serve as conservative estimates of change-point locations. Note that in some cases the 90 % CI is substantially higher (z−) or lower (z+) than the observed change point. In these instances, this suggests a broader range of change much like the taxon illustrated in Fig. 11.2b. We also acknowledge that CIs can be inaccurate for taxa with relatively few occurrences, which would be the case for any analysis, so strict interpretation of CIs for individual taxa is discouraged (Baker and King 2010). However, the CIs do provide an informative reflection of variability in IndVal scores among different samples of the data.

The final two columns report metrics of uncertainty about the repeatability of a taxon as a potential threshold indicator. Purity is the proportion of times that a taxon is given the same classification in each bootstrap replicate as in the observed data set. So, taxa that receive a purity score of 1.00 were assigned as a z+ taxon (distance from canal) or z− (TP) in every bootstrap replicate. The second metric is reliability. It counts the number of times out of the n number of bootstrap replicates that an individual taxon achieved a p–value < 0.05. The closer this value is to 1.00, the more likely the taxon is indeed responding in a predictable manner to the novel gradient. Note that we did not include a column for observed p-values; all of the p-values for these taxa were less than or equal to the lowest possible value for 250 permutations (< 0.004). The observed p-value is neither informative, nor is it an appropriate metric of statistical significance (Baker and King 2013). Taxa with purity > 0.95 and reliability > 0.95 always achieve $p < 0.05$ because the former two metrics are based on resampling of the data and thus much more robust indicators of taxa response. Thus, this list of taxa is of those that passed these two filtering criteria (purity > 0.95, reliability > 0.95). Note that for the sake of space, we did not include in Table 11.1 the taxa that increased along the variables representing the novel gradient (z− for distance from canal, z+ for TP), but did in the figures to follow.

Before we move on to a discussion of the graphical results, note in Table 11.1 the remarkable degree of overlap in the list of taxa that increase with distance from the canal and decrease with increasing TP. We expected substantial overlap given the correlation between variables but also some disagreement because of the moderate amount of variability in TP concentration as a function of distance from the canal (Fig. 11.3). The concordance between predictors provides at least some support for the idea that the effect of the canal is a function of TP, given that we already know

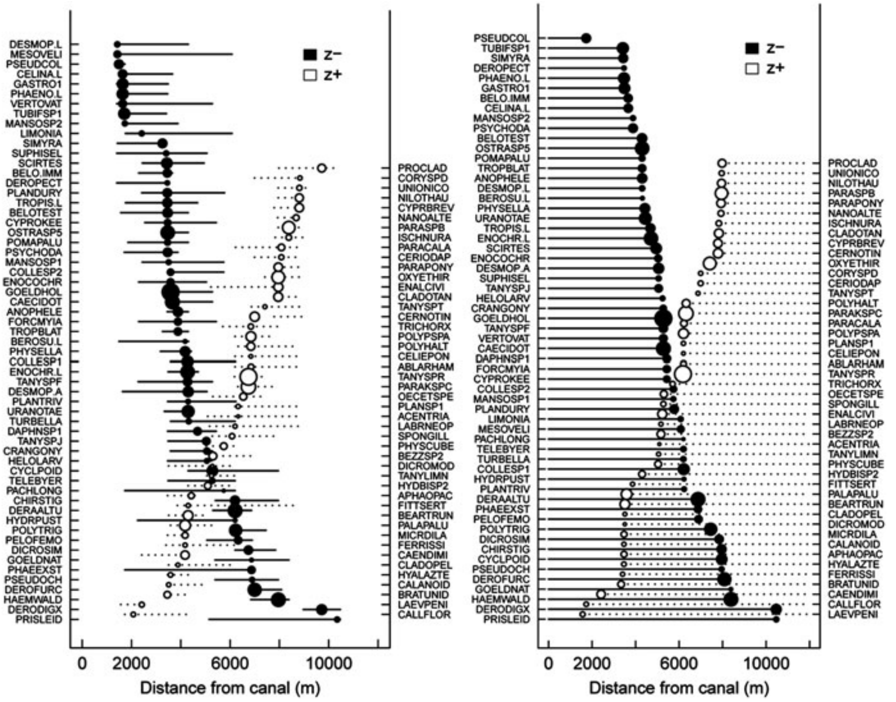


Fig. 11.4 Robust indicator taxa identified by TITAN in response to distance from the canal. In panel **a**, *horizontal lines* represent 90 % CIs of observed change points (*open or closed circles*) for each taxon. In panel **b**, taxa are ordered based on the 95th quantile of the bootstraps. *Horizontal lines* extend from the lowest (declining taxa) or the highest predictor value (increasing taxa) to the 95th quantile of the distribution of change points for each taxa ("diving board plot")

that the Everglades is very P limited and that experiments have validated strong community responses to TP in isolation.

Graphical evaluation of individual taxa results from TITAN provides a much cleaner depiction of most of the tabular output presented in Table 11.1. Figure 11.4 illustrates two types of graphs we developed for displaying indicator taxa results. In Fig. 11.4a, the open and filled circles are the observed change points for each indicator taxon that passed the screening criteria (as shown in Table 11.1). The horizontal lines intersecting each point are the 5th and 95th quantiles of the bootstrap distribution of change points for each taxon, i.e., confidence or variability bands such as those illustrated in Fig. 11.2a, b. The *y*-axes show the code names of the indicator taxon in rank order of the observed change point, starting with the lowest change point on the top left axis (negative indicator taxa, z−) and the highest change point on the top right (positive indicator taxa, z+). The size of the open or filled circles is proportional to the indicator z score, so larger circles are taxa with stronger relative responses to the gradient. Note that there actually were more taxa that favored conditions near the canal than the natural condition at the opposite end of the gradient, as shown by the longer list of taxa on the left *y*-axis.

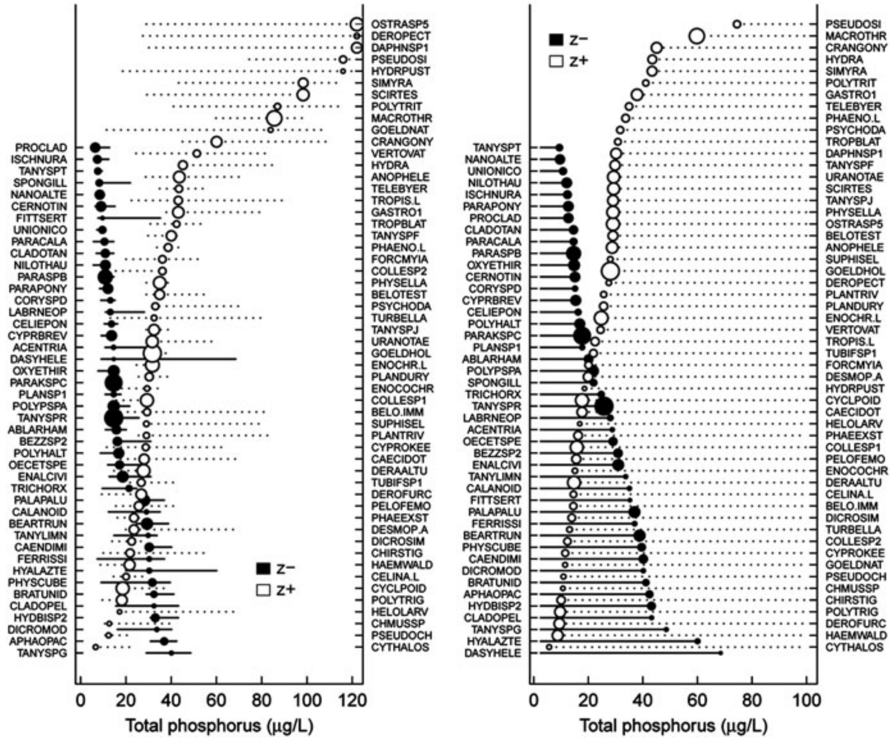


Fig. 11.5 Robust indicator taxa identified by TITAN in response to surface-water total phosphorus (TP). In panel **a**, *horizontal lines* represent 90 % CIs of observed change points (*open* or *closed circles*) for each taxon. In panel **b**, taxa are ordered based on the 95th quantile of the bootstraps. Horizontal lines extend from the lowest (declining taxa) or the highest predictor value (increasing taxa) to the 95th quantile of the distribution of change points for each taxa ("diving board plot")

The second figure (Fig. 11.4b) illustrates the same information but in a different way. In this figure, taxa are plotted in rank order of the CI-90 %. The open and filled symbols are placed at CI-90 % and sized in proportion to z scores. The horizontal lines are drawn from the 90 % CI to the *y*-axis to facilitate visualization of overlap of the increasing and decreasing indicators as well as the degree of synchrony in change point locations. We have termed this a "diving board plot" because the open or filled symbol at the end of the horizontal line represents the point where a taxon is likely to "dive."

In both subfigures of Fig. 11.4, the pattern of many taxa increasing and decreasing in a relatively narrow range of the novel gradient should be evident, particularly in the Fig. 11.4b. The zone of overlap nicely illustrates a region of substantial turnover in taxonomic composition. Despite relatively strong synchrony, there is evidence for some degree of a continuum of change in the middle of the gradient rather than a catastrophic "step function" type threshold.

However, Fig. 11.5 suggests that part of the reason for continuum of change with distance from the canal is the variability in TP at different distances from the canal,

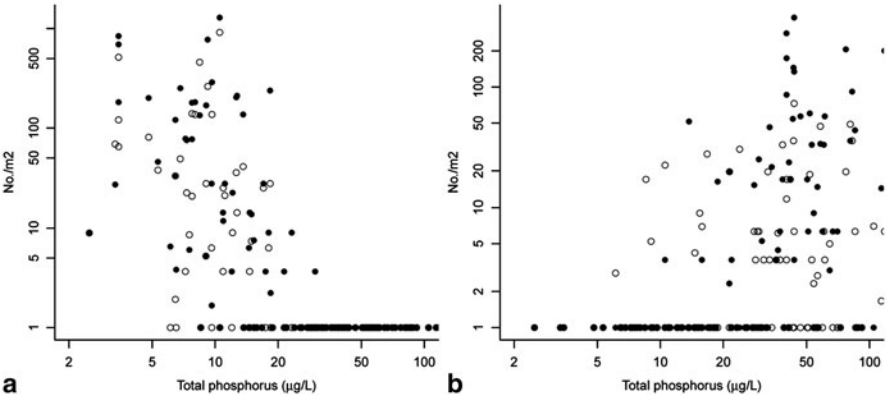


Fig. 11.6 Scatterplots of response of **a** two representative negative indicators (*Tanytarsus* sp. R, *filled symbols*, and *Parakiefferiella* sp. C, *open symbols*) and **b** two positive indicators (*Goeldichironomus holoprasinus*, *filled symbol*, and *Caecidotea*, *open symbol*) to the phosphorus gradient. Note that TP is expressed on a logarithmic scale

as shown in Fig. 11.3. The response of negative indicators to TP is much more abrupt and synchronous, ranging from about 10–40 ug/L TP. The strongest threshold indicators responded quite synchronously between 10 and 20 ug/L TP, evidenced by the clustering of larger filled symbols at low levels of TP. In some cases, taxa occurred in nearly every sample unit to the left of the threshold and none of the sample units to the right (Fig. 11.6a, b).

Recall that TITAN also provides a second set of results that attempts to synthesize individual taxa responses into an index of community-level change, sum(z). Table 11.2 reports several community-level results based on the aggregate response of negative and positive responding taxa, respectively. The raw sum(z) value is the sum of all z scores (not just ones that met screening criteria) at the value of the predictor where sum(z) is maximized. This value is computed for negative and positive responders separately. The sum(z) value is quite useful by itself because it provides a metric of the aggregate magnitude of change among negative and positive indicator taxa in the community. It is reasonable to compare the sum(z) value among multiple predictors as a metric of explanatory power, but it is not reasonable to compare sum(z) among different data sets because the absolute value is dependent upon the number of taxa in the data set. In this case, the sum(z) for positive responses to the canal gradient is slightly lower than sum(z) for the negative responses to the TP gradient, suggesting that TP did a slightly better job of capturing the effect of enrichment than the proximal "source" variable, canal. However, the opposite was true for the taxa that favored conditions near the canal, where sum(z) negative exceeded the TP sum(z) positive by a moderate value. This was also reflected in the larger number of taxa that were deemed negative indicators to canal than positive indicators of TP (59 vs. 54, respectively).

Table 11.2 Community-level results from TITAN using distance from canal and total phosphorus as predictors and macroinvertebrate taxa abundances as the response. (see Fig. 11.7)

	Distance from canal (m)		Total phosphorus (ug/L)	
	Negative	Positive	Negative	Positive
Sum(z)	304	311	327	283
No. of indicator taxa	59	46	44	54
Sum(z) change pt.	5,739	7,956	14.62	31.68
CI 5 %	3,485	5,169	12.36	20.08
CI 10 %	3,578	5,277	12.66	22.55
CI 50 %	5,048	6,794	15.11	29.83
CI 90 %	6,077	7,956	21.28	37.86
CI 95 %	6,202	8,382	28.31	40.11

Sum(z) is the sum of all taxa z scores at the level of the predictor that resulted in the greatest change in the aggregate response of negative (z−) or positive (z+) taxa. Sum(z) is a relative measure of response magnitude. The number of indicator taxa reflects only those taxa that passed all screening criteria ($p < 0.05$, purity > 0.95, reliability > 0.95; see Table 11.1). The sum(z) change point is the value of the predictor that resulted in the greatest aggregate change among negative and positive responding taxa, respectively. The CIs correspond to change point quantiles computed from the bootstrap replicates and are displayed visually as a cumulative frequency curve in Fig. 11.7

The next series of results are the observed and bootstrap quantiles of change points for the community-level response of negative and positive indicator taxa. The observed change points for both sum(z−) and sum(z+) are relatively similar for both predictors, reflecting the synchronous turnover in taxa in the zone of 6–8 km from the canal and 14–31 ug/L TP. The location of synchronous decline in response to TP was the tightest community-level response among the sum(z) change points, spanning a 90 % CI of 11–28 ug/L TP in comparison to the full range of values spanning ∼ 5 to 150 ug/L TP.

The last form of output from TITAN is the plot of all of the values of sum(z) along the novel gradient (Fig. 11.7). These plots are arguably less intuitive but quite informative once fully understood. The left *y*-axis is the sum of the z scores. As the value of the gradient increases, the sum of the z scores will climb as the community-level response increases in magnitude. The peak in the sum of the z scores represents the observed change point. Beyond the peak, values will decline but may show secondary peaks along the gradient where other groups of taxa change synchronously. If the peak is very sharp such that the sum of the z scores increases and decreases rapidly on either side of the peak, this is strong evidence for a sharp, synchronous change in the community. If the peak is broad, such that there are many values along the gradient that produce similar sum(z) scores, this is more indicative of a zone of change rather than an abrupt threshold. Contrasting the response of the sum(z) scores between canal and TP shows these two different responses. The response of both increasing and decreasing taxa to the canal gradient is more gradual, such that the peak in sum(z) bounces around in the middle of the gradient before declining sharply on either end. Conversely, the negative response to TP is quite sharp, with a clear peak evident at ∼ 15 ug/L. The positive response is less sharp but is still consistent with a rapid, synchronous increase in multiple taxa around 30–40 ug/L

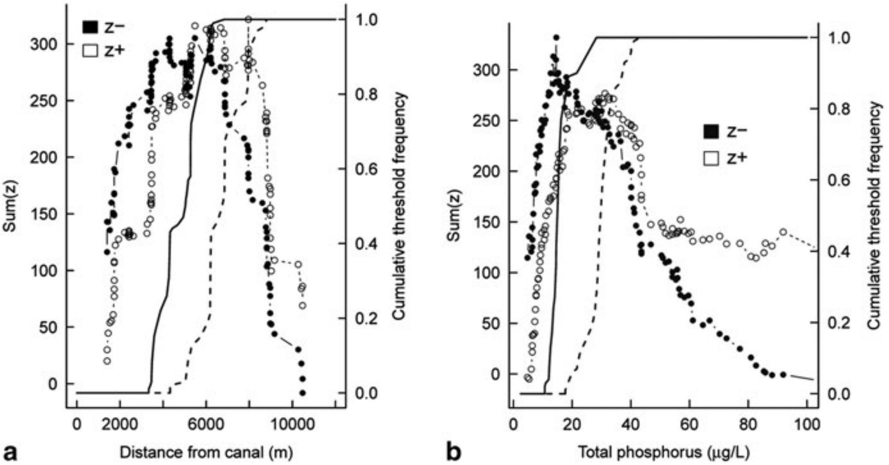


Fig. 11.7 TITAN sum(z−) and sum(z+) values for all possible change points in response to distance from canal (**a**) and total phosphorus (**b**). Peaks in sum(z−) correspond to locations along the *gradient* where synchronous declines of taxa occur. *Solid* and *dashed lines* represent the cumulative frequency distribution of change points among 100 bootstrap replicates for sum(z−) and sum(z+), respectively

TP. Finally, the right *y*-axis shows the cumulative frequency of sum(z) "peaks" among the bootstrap replicates. Steep cumulative frequency curves will span only a narrow range of *x* values and are further support for a community-level threshold. Broader curves imply more gradual change. Collectively, these results support the conclusion that macroinvertebrate communities respond strongly to TP, that the effect of TP on community structure is nonlinear, and that the source of TP, the canal, explains most of the same variability in community structure.

Misconceptions, Misuse, and Myths About Community Thresholds and TITAN

In this concluding section, we will address several ideas or issues that we consider important points of clarification for users of TITAN and anyone interested in analyzing ecological data in search of thresholds. We have attempted to group them but admit that these cover a wide range of topics and may not follow a logical progression.

Community thresholds are not necessarily "step functions." We have encountered numerous investigators who react to the idea of thresholds with much doubt if not disdain. Although reasons vary, one apparent reason for this reaction is the preconceived notion that a threshold necessarily implies a single point along a gradient where everything falls apart, and prior to and following that point essentially nothing happens. Perhaps the root of this confusion is the related but distinct theoretical ideas of "regime shifts" or "alternative stable states" where an entire ecosystem undergoes a shift from one structural and functional identity to a very different one once the system is pushed beyond some theoretical limit (e.g., Sheffer and Carpenter 2003). The best (and one of the few) example of such a shift is lake eutrophication where a clear-water system with rooted macrophytes suddenly shifts to a turbid, phytoplankton system once nutrient loading exceeds its assimilative capacity (Sheffer et al. 2001). While such examples probably exist for ecological communities in response to novel gradients (and is actually approximated fairly well by the synchronous declines of taxa to TP in our Everglades example), this definition is by no means sufficiently inclusive of other responses that also may be deemed thresholds. Per our own definition, a community threshold need only demonstrate that a certain subset of taxa change in a relatively synchronous manner at a particular level of a novel environmental gradient, and that additional change prior to or beyond that point is entirely acceptable. Once investigators are willing to acknowledge this definition, we submit that the threshold concept will be less offensive to their sensibilities.

Linear responses of univariate community metrics to environmental gradients do not rule out community thresholds. It is of no surprise that different investigators can come to different conclusions about the response of communities to novel gradients using essentially identical data sets. One of the best examples of this has been the ongoing debate about stream community response to watershed urbanization (e.g., Walsh et al. 2005; Cuffney et al. 2010; King and Baker 2010, 2011). Some investigators claim no evidence in support of thresholds based on the result that variance in univariate community metrics is well explained by a linear regression. Others have claimed that a piecewise model with an immediate linear decline across a substantial fraction of the gradient followed by a second zone of essentially little or no change is the most probable response, whereas we have consistently detected a narrow zone of urban intensity where multiple taxa begin to decline, indicative of what we consider to be a community threshold (King and Baker 2010, 2011; King et al. 2011; Baker and King 2013). How can one reconcile these disparate results?

It is our opinion, one we base on multiple empirical lines of evidence, that the use of univariate community metrics as a response variable coupled with the very casual application of linear models has obscured nonlinear changes in community data. We have demonstrated this phenomenon using a simulation where we programmed taxa responses to sharply decline at a particular level of a novel gradient and combined these responses with more gradual increases (as in Fig. 11.2) and other taxa with no response. Once these responses were combined into a single value per sample unit, the response appeared roughly linear for most of the gradient whereas TITAN revealed very sharp synchronous declines in the taxa that we had programmed (King and Baker 2010). In sum, we caution the use of univariate metrics for community threshold analysis without careful consideration of the location, magnitude, and direction of individual taxa responses.

Linear regression does not provide a "test" for the absence of thresholds. We also caution the use of linear regression for "testing" for the presence or absence of thresholds (see King and Baker 2011). A significant *p*-value for a regression slope does not mean that response is necessarily linear. Only a graphical examination of the *x*–*y* relationship and the residuals from that relationship can yield the necessary information about the appropriateness of the linear model. If investigators are set on analyzing univariate metrics, we suggest they read Zuur et al. (2010) and consider using loess regression (e.g., King and Baker 2011), or even better, generalized additive models (GAMs) in accordance with the assumptions of the analysis to determine whether the response has evidence of nonlinearity (e.g., Bernhardt et al. 2012). GAMs fit smoothing functions to the response but will only smooth the response if the addition of greater model complexity is deemed worthwhile based on cross validation. If the estimated degrees of freedom exceed 1 and the smoother p-values are < 0.001, it is highly likely that the response is nonlinear. GAMs also permit specification of appropriate distribution families to match the distribution of response data (e.g., negative binomial, Gaussian, etc).

Beware of long environmental gradients when dismissing thresholds. Another issue related to the use of univariate response variables for threshold identification is gradient length. The problem is that predictor variables that span a very wide range of conditions (long gradients; for example, urban intensity from none to downtown Chicago) can obscure sharp, nonlinear patterns at low levels of the gradient. If a variable does not respond immediately to a novel gradient, (generally, they do not) it may appear to do so if the response location is extreme relative to the complete gradient length. This was a major point of King and Baker (2011), who critiqued a different study that concluded responses to urbanization were linear but missed the lack of response at low levels of the gradient because the gradient was so long and difficult to visualize without looking more carefully at a narrower range of values. Once viewed at levels of urban intensity between 0 and 20, the responses were essentially flat until a critical level of urbanization was reached, which happened to be similar to the levels of urbanization identified as community thresholds by TITAN. Thus, we strongly recommend that users graph their data in such a way (log transformed or truncated axis) to reveal low-level responses to novel gradients (Zuur et al. 2010).

Do not use community metrics as response variables in TITAN. TITAN is designed for taxa abundance data sets (matrix of abundances of multiple taxa by sample units). Patterns of frequency of occurrence among sample units are the key component of the IndVal calculation. Data sets dominated by species that occur in all samples are poorly suited for TITAN because presence/absence of taxa no longer contributes any information to the analysis. This issue becomes particularly problematic when investigators attempt to use a matrix of community metrics (e.g., number of Ephemeroptera, Plecoptera, and Trichoptera taxa (EPT), percentage filterers, ratio of weevils to platypus, etc). TITAN is not intended for use in this way because most contain few if any nonzero values and do not approximate negative binomial distributed abundance data for which TITAN was designed. Such responses are better modeled with other approaches, such as GAM (Zuur et al. 2010), nonparametric multiplicative regression (NPMR; McCune 2006), or, in the rare case when data (residual variance) are normally distributed, piecewise regression.

Statistical change points are not necessarily "thresholds." Large data sets with numerous taxa will almost certainly yield at least a few taxa that are identified as having change points. In fact, even linear responses will yield a change point in

TITAN because the method is designed to find taxa that have greater frequency and abundance at one end of the gradient than another. A linear response will yield such a pattern. Thus, one or a few change points alone does not imply a community threshold using TITAN, although these responses are still potentially useful to managers. First, we strongly encourage users to evaluate the uncertainty around observed change point locations of individual taxa. If it is very broad, there is weak evidence for a sharp change in its distribution with respect to the novel gradient. Narrow confidence or variability bands provide greater support for a threshold-type response for a single taxon. However, more importantly, we recommend that users focus on the distribution of multiple taxa change points. Are the confidence limits narrow, and do they overlap (i.e., are they relatively synchronous)? Does the sum(z) peak sharply or is it poorly defined? If taxa change points are relatively widely distributed with broad confidence limits and poorly defined sum(z) peaks, the response is probably better characterized as a zone of gradual change. The main point here is that TITAN provides a lot of different types of information but it is up to the investigator or manager to interpret the output. See Baker and King (2013) for a thorough treatment of this issue.

Density of the distribution of sample units can affect results. The distribution of sample units along environmental gradients can be an important factor for any threshold analysis. If most of the data points are located near the low end of the gradient, it can lead to misleading change points because there may not be sufficient distribution of sample units at other levels of the gradient, particularly if there are large gaps. Using TITAN or most any other method under these circumstances will likely yield biased results and should be interpreted with caution, if at all (Daily et al. 2012).

TITAN is not a causal analysis. TITAN was designed primarily for use with observational data, particularly large biomonitoring data sets that span a wide range of novel environmental gradients. Such gradients are almost always confounded by multiple, correlated variables which make it very difficult to make strong inference about the cause of the observed response (e.g., King et al. 2005). We strongly caution users to think carefully about their data prior to using TITAN. All of the criteria used to define reference conditions and classify sites into comparable physiographic groupings should be applied to TITAN (e.g., Stoddard et al. 2006).

TITAN is not intended to be black box for developing regulatory criteria. TITAN has great potential to inform managers about critical levels of anthropogenic changes that are associated with rapid changes in ecological communities. However, the potential for confounded variables or study designs that lead to misleading results is certainly a distinct possibility. We strongly discourage using output from TITAN as the sole basis for supporting management decisions. We further suggest that multiple lines of evidence be used to further support or refute TITAN results. Manipulative field experiments and lab studies are certainly recommended when applicable. At a minimum, we recommend that investigators carefully examine the list of taxa identified as threshold indicators and apply knowledge of species sensitivities, evolutionary relationships, and life-history characteristics to support statistical conclusions.

Conclusion

TITAN is a tool. All tools have limitations and can be misused. However, TITAN has distinct advantages for detecting change in taxa distributions that may help identify levels of environmental change associated with disproportionate declines or increases in species abundances, which in turn may be indicative of community thresholds.

References

Andersen, T., J. Carstensen, E. Hernandez-Garcia, and C. M. Duarte. 2009. Ecological regime shifts: Approaches to identification. *Trends in Ecology and Evolution* 24:49–57.

Baker, M. E., and R. S. King. 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. *Methods in Ecology and Evolution* 1:25–37.

Baker, M. E., and R. S. King. 2013. Of TITAN and Straw Men: An appeal for greater understanding of community data. *Freshwater Science* 32:(1) pages forthcoming.

Bernhardt, E. S., B. D. Lutz, R. S. King, J. P. Fay, C. E. Carter, A. M. Helton, and D. Campagna. 2012. How many mountains can we mine? Assessing regional degradation of Central Appalachian rivers by surface coal mining. *Environmental Science and Technology* 46:8115–8122.

Brenden, T. O., L. Wang, and Z. Su. 2008. Quantitative identification of disturbance thresholds in support of aquatic-resource management. *Environmental Management* 42:821–832.

Clements, W. H., N. K. M. Vieira, and D. L. Sonderegger. 2010. The use of ecological thresholds to assess recovery in lotic ecosystems. *Journal of the North American Benthological Society* 29:1017–1023.

Connell, J. H. 1978. Diversity in tropical rainforests and coral reefs. *Science* 199:1302–1310.

Cuffney, T. F., R. B. Brightbill, J. T. May, and I. R. Waite. 2010. Responses of benthic macroinvertebrates to environmental changes associated with urbanization in nine metropolitan areas. *Ecological Applications* 20:1384–1401.

Daily, J. P., N. P. Hitt, D. R. Smith, and C. D. Snyder. 2012. Experimental and environmental factors affect spurious detection of ecological thresholds. *Ecology* 93:17–23.

DeLuca, W. V., C. Studds, R. S. King, and P. P. Marra. 2008. Coastal development and the integrity of estuarine waterbird communities: Threshold responses and the importance of scale. *Biological Conservation* 141:2669–2678.

Dodds, W. K., W. H. Clements, K. Gido, R. H. Hilderbrand, and R. S. King. 2010. Thresholds, breakpoints, and nonlinearity in aquatic ecosystems as related to management. *Journal of the North American Benthological Society* 29:988–997.

Dufrêne, M., and P. Legendre. 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs* 67:345–366.

Ferrier, S, and A. Guisan. 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology* 43:393–404.

Fox, D. 2007. Back to the no-analog future? *Science* 316:823–825.

Gaiser, E. E., J. H. Richards, J. C. Trexler, and R. F. Doren. 2008. Comment on "Estimating ecological thresholds for phosphorus in the Everglades.". *Environmental Science and Technology* 42:6770–6771.

Groffman, P. M., J. S. Baron, and T. Blett. 2006. Ecological thresholds: The key to successful environmental management or an important concept with no practical application? *Ecosystems* 9:1–13.

Hobbs, R. J., S. Arico, J. Aronson, J. Baron, P. Bridgewater, V. A. Cramer, P. Epstein, J. Ewel, C. A. Klink, A. E. Lugo, D. Norton, D. Ojima, D. M. Richardson, E. W. Sanderson, V. Fernando, Z. Montserrat, Z. Regino, and M. Zobel. 2006. Novel ecosystems: Theoretical and management aspects of the new ecological world order. *Global Ecology and Biogeography* 15:1–7.

Huggett, A. J. 2005. The concept and utility of 'ecological thresholds' in biodiversity conservation. *Biological Conservation* 124:301–310.

King, R. S., and M. E. Baker. 2010. Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. *Journal of the North American Benthological Society* 29:998–1008.

King, R. S., and M. E. Baker. 2011. An alternative view of ecological community thresholds and appropriate analyses for their detection. *Ecological Applications* 21:2833–2839.

King, R. S., and C. J. Richardson. 2003. Integrating bioassessment and ecological risk assessment: an approach to developing numerical water-quality criteria. *Environmental Management* 31:795–809.

King, R. S., M. E. Baker, D. F. Whigham, D. E. Weller, P. F. Kazyak, and M. K. Hurd. 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecological Applications* 15:137–153.

King, R. S., W. V. DeLuca, D. F. Whigham, and P. P. Marra. 2007. Threshold effects of coastal urbanization on Phragmites australis (common reed) abundance and foliar nitrogen in Chesapeake Bay. *Estuaries and Coasts* 30:469–481.

King, R. S., M. E. Baker, P. F. Kazyak, and D. E. Weller. 2011. How novel is too novel? Stream community thresholds at exceptionally low levels of catchment urbanization. *Ecological Applications* 21:1659–1678.

Manly, B. F. J. 1997. *Randomization, bootstrap, and Monte Carlo methods in biology*. 2nd ed. London: Chapman & Hall.

McCune, B. 2006. Nonparametric habitat models with automatic interactions. *Journal of Vegetation Science* 17:819–830.

Moore, A. A., and M. A. Palmer. 2005. Invertebrate biodiversity in agricultural and urban headwater streams: Implications for conservation and management. *Ecological Applications* 15:1169–1177.

Økland, B., O. Skarpaas, and K. Kausrud. 2009. Threshold facilitations of interacting species. *Population Ecology* 51:513–523.

Palmer, M. A., H. Menninger, and E. S. Benhardt. 2009. River restoration, habitat heterogeneity, and biodiversity: A failure of theory or practice? *Freshwater Biology* 55:205–222.

Poff, N. L. 1997. Landscape filters and species traits: Towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society* 16:391–409.

Qian, S. S., R. S. King, and C. J. Richardson. 2003. Two methods for the detection of environmental thresholds. *Ecological Modelling* 166:87–97.

Richardson, C. J., R. S. King, S. S. Qian, P. Vaithiyanathan, R. G. Qualls, and C. A. Stow. 2007. Estimating ecological thresholds for phosphorus in the Everglades. *Environmental Science and Technology* 41:8084–8091.

Richardson, C. J., R. S. King, S. S. Qian, P. Vaithiyanathan, R. G. Qualls, and C. A. Stow. 2008. Response to comment on "Estimating ecological thresholds for phosphorus in the Everglades". *Environmental Science and Technology* 42:6772–6773.

Scheffer, M., and S. R. Carpenter. 2003. Catastrophic regime shifts in ecosystems: Linking theory to observation. *Trends in Ecology & Evolution* 18:648–656.

Scheffer, M., S. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413 (6856): 591–596.

SFWMD (South Florida Water Management District). 1992. Surface water improvement plan for the Everglades. West Palm Beach (FL): Supporting Information Document, South Florida Water Management District.

Shelford, V. E. 1913. *Animal communities in temperate North America*. Chicago: University of Chicago Press.

Smol, J. P., A. P. Wolfe, and H. J. B. Birks. 2005. Climate-driven regime shifts in the biological communities of arctic lakes. *Proceedings of the National Academy of Science of the United States of America* 102:4397–4402.

Sonderegger, D. L., H. Wang, W. H. Clements, and B. R. Noon. 2009. Using SiZer to detect thresholds in ecological data. *Frontiers in Ecology and the Environment* 7:190–195.

Soranno, P. A., K. S. Cheruvelil, R. J. Stevenson, S. L. Rollins, S. W. Holden, S. Heaton, and E. K. Torng. 2008. A framework for developing ecosystem-specific nutrient criteria: Integrating biological thresholds with predictive modeling. *Limnology and Oceanography* 53:773–787.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16 (4): 1267–1276.

Stralberg, D., D. Jongsomjit, C. A. Howell, M. A. Snyder, J. D. Alexander, J. A. Wiens, and T. L. Root. 2009. Re-shuffling of species with climate disruption: A no-analog future for California birds? *PLoS One* 4 (9): e6825.

Suding, K. N., and R. J. Hobbs. 2009. Threshold models in restoration and conservation: A developing framework. *Trends in Ecology and Evolution* 24:271–279.

Toms, J., and M. L. Lesperance. 2003. Piecewise regression: A tool for identifying ecological thresholds. *Ecology* 84:2034–2041.

Vinson, M. R., and C. P. Hawkins. 1998. Biodiversity of stream insects: Variation at local, basin, and regional scales. *Annual Review of Entomology* 43:271–293.

Walsh, C. J., T. D. Fletcher, and A. R. Ladson. 2005a. Stream restoration in urban catchments through redesigning stormwater systems: Looking to the catchment to save the stream. *Journal of the North American Benthological Society* 24:690–705.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005b. The urban stream syndrome: Current knowledge and the search for a cure. *Journal of the North American Benthological Society* 24:706–723.

Williams, J. W., and S. T. Jackson. 2007. Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment* 5:475–482.

Zuur, A. F., E. N. Ieno, C. S. Elphick. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1:3–14.

Chapter 12
Ecosystem Trajectories: A Statistical Approach to Analyze Changing Pressure-Response Relationships Over Time

Jacob Carstensen

Abstract There is increasing empirical evidence that ecosystem responses to changing pressures follow different pathways during the degradation and recovery phases. I present a statistical inferential approach based on generalized additive models (GAM) to substantiate such conclusions. The approach analyzes the time trajectories of departures from a proposed functional relationship between pressure and response. The trajectory analysis provides a general exploratory tool to uncover changes in pressure-response relationships that may not be apparent from plotting the data as well as a model diagnosis tool. Simulations revealed that the approach can separate the time trajectory from the functional relationship, when the observed pressure variable is well determined. Four coastal ecosystems from Duarte et al. (Estuaries and Coasts 32:29–36, 2009) were reanalyzed to exemplify the approach, providing statistical evidence of separate pathways during eutrophication and oligotrophication. For the many empirical studies on ecological regime shifts and shifting baselines I recommend that the trajectory analysis, in combination with other analytical procedures, is employed to document the existence of such effects with sufficient statistical confidence.

Keywords Ecosystem restoration · Eutrophication · Generalized additive model · Multi-stressors · Regime shift · Shifting baseline · Statistical identification · Time series analysis

Introduction

On a geological time scale, our planet has experienced a relatively stable environment for the last 10,000 years (Petit et al. 1999), known to geologists as the Holocene, but for the last couple of centuries, also termed the Anthropocene (Crutzen 2002), human activities have become the main driver of environmental change at the global

---

J. Carstensen (✉)
Department of BioScience, Aarhus University,
Frederiksborgvej 399, DK-4000 Roskide, Denmark
e-mail: jac@dmu.dk

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_12,

scale and also at the local scale throughout most regions of the earth (Rockström et al. 2009). These anthropogenic drivers have exerted increased pressure on world's ecosystems (Vitousek et al. 1997), and the acknowledgment and scientific documentation of the associated deleterious effects have prompted political responses to alleviate these pressures in an attempt to restore ecosystem functioning. Action plans addressing various sources of emissions were established on the pervasive belief that ecosystem responses to increasing pressures could be reversed and that a previously observed and desired ecosystem state could be restored. The fundamental tenet was that ecosystem decline and recovery would follow the same pathway, linear or nonlinear, and that such relationships between driver and response were time invariant.

There is growing observational evidence that this tenet is essentially flawed. The ozone layer has not been restored to past levels after the implementation of the Montreal Protocol in 1987 for reducing emissions of chlorofluorocarbon (CFC) gases and this lack of recovery is believed to be caused by climate warming and release of new chemicals with a yet unknown effect to the ozone layer (Weatherhead and Andersen 2006). Commercial fish stocks have not recovered following reduced fishing pressure, and climate change and complex food-web interactions have been suggested as plausible explanations (Botsford et al. 1997). Nutrient reductions in coastal areas have not reduced phytoplankton biomass (Duarte et al. 2009) nor the extent of hypoxia (Conley et al. 2007), and this is mainly attributed to climate change and altering of the food web. Thus, many of the world's ecosystems fail to return to previously observed states after pressure reduction, because changes in other drivers have shifted the baseline. This evidently leads to nonuniform time trajectories in the relationship between pressure and response.

The concept that ecosystems respond nonlinearly to changes in the drivers, displaying hysteresis-like behavior with alternative stable states, is not new (Holling 1973; May 1977). Lakes exhibit shifts between a clear state with dominance of submerged aquatic vegetation and a turbid state dominated by phytoplankton (Scheffer et al. 1997; Carpenter et al. 1999). The shift is typically driven by enhanced nutrient input (mostly phosphorus) from human activity leading to increased phytoplankton growth and subsequently shading of the submerged aquatic vegetation (Jeppesen et al. 1999). Tropical reefs alternate between corals states and states where macroalgae overgrow the corals and prevent the settlement of coral larvae for continued recruitment (Knowlton 1992). The resilience of the coral reefs has been eroded through nutrient enrichment and overfishing (Scheffer et al. 2001), whereas the shift between states appears to be triggered by events such as hurricanes and outbreak of diseases affecting sea urchins (Mumby et al. 2007). The complex interactions in food webs may similarly lead to regime shifts through cascading effects driven by eutrophication, overfishing, and invasive species (Daskalov et al. 2007; Casini et al. 2008). Outbreaks of hypoxia can also lead to a sudden change in the biogeochemical processes causing a positive feedback of nutrients to the water column through reduced nitrification-denitrification, releases of ironbound phosphate, and reduced transfer of energy to higher trophic levels (Conley et al. 2009b). Loss of benthic macrofauna with hypoxia and thresholds associated with recolonization suggests a hysteresis-like behavior (Diaz and Rosenberg 2008). Thus, there are several examples from the literature of ecosystems displaying hysteresis behavior and the conceptual understanding of the positive feedback mechanisms required for the existence of alternative stable states has largely been established.

However, there is a gap in the literature between apparent regime shifts and the application of a rigorous mathematical-statistical framework for the actual demonstration of such threshold effects and hysteresis responses (Andersen et al. 2009). Moreover, most quantitative analyses of regime shifts are theoretical studies that examine the behavior of a simple nonlinear model that is believed to capture the essential mechanisms of the ecosystem (e.g., Carpenter et al. 1999; Ludwig et al. 2003; Guttal and Jayaprakash 2008). Although such models may mimic ecosystem observations to a reasonable degree and hence provide support for the existence of regime shifts, they do not offer statistical confidence in the existence of bistability, i.e., in terms of quantifying the probability of a hysteresis response relative to a simpler and uniform relationship.

Statistical tests can, in principle, be employed by comparing the likelihood of two such competing models, but in practice this is more complicated as it requires relatively simple mathematical representation of the ecosystem in question (i.e., few parameters) and sufficient data to estimate these models and calculate their likelihood. Consequently, scientists have resorted to simpler statistical procedures, typically identification of change-points in time (e.g., Zeileis et al. 2003; Rodionov 2004), as an exploratory data analysis indicating if abrupt changes may have occurred. Such statistical methods have started to populate the ecological literature recently and are the natural first step towards identifying potential drivers and mechanisms but do not describe any driver-response relationship as time can never be the underlying driver (Andersen et al. 2009). Moreover, change-point detection methods can identify abrupt changes, both with respect to time and potential drivers although the latter is rarely seen in the literature, but they do not provide inference for alternative stable states. Therefore, the objective here is to supplement the growing set of statistical methods used to analyze for potential regime shifts in ecosystems with a method indicative of alternative stable states. The idea is to examine the time trajectory of an ecosystem response variable relative to a hypothesized driver and test if this trajectory is time invariant, i.e., the relationship is uniform across time.

Conceptualizing Ecosystem Responses

The literature is populated with conceptual figures displaying different categories of driver-response relationships. May (1977) formulated a nonlinear differential equation and showed graphically that this model would exhibit two alternative stable states for a specific parameter setting and one unstable state constituting a divide between the two stable attractors. A perhaps better illustration of this concept was the marble rolling in a rugged landscape that could have several attractors (Scheffer 1990; Scheffer et al. 1993; Scheffer et al. 2001). The ridge between the two basins of attraction constituted the unstable state where the ball would roll either direction. Carpenter et al. (1999) presented a simple lake model with a sigmoid phosphorus influx for the sources and a linear efflux for the sinks, and graphically demonstrated how this could lead to alternative stable states and hysteresis responses. The main conclusion from these studies was that the bistability figures could actually be derived mathematically from the simple models exhibiting hysteresis.

On the other hand, the statistical approach to conceptualizing driver-response relationship has been based on experiences from data exploration. De Young et al. (2004) proposed three different types of responses: (1) linear, (2) abrupt change and reversible, and (3) abrupt change and not directly reversible, the latter representing a hysteresis-type behavior. Andersen et al. (2009) extended these to also consider the time dimension, showing that abrupt changes in time series can occur even if the driver-response relationship is strictly linear, because an abrupt change in the driver is directly mediated to the response. They cautioned about over-interpreting abrupt changes in biological time series, if the cause of the change was not within the biological system itself.

A broad range of possible responses to increasing followed by decreasing pressures on the ecosystem, derived from theory and observations, has been proposed and synthesized into a few generic classes of responses (e.g., Duarte et al. 2009; Kemp et al. 2009). Here, I will consider the four different response types presented in Duarte et al. (2009) (Fig. 12.1). The uniform relationship between response and pressure variable is an idealized situation (Fig. 12.1a), where nothing else changes over time (termed "Return to Neverland" in Duarte et al. 2009). This is the fundamental type of relationship that managerial frameworks are built around, despite increasing observational evidence that pressure-response relationships are not static. The hysteresis relationship (Fig. 12.1b) resembles those obtained from theoretical studies (e.g., May 1977; Scheffer et al. 2001) with alternative stable states within a range of the pressure variable. It involves a resistance to return to the original state when the pressure is alleviated. A shift in the ecosystem baseline (Fig. 12.1c) typically occurs in a multi-pressure system, which essentially includes all ecosystems, and illustrates that the outcome, after reducing the main pressure on the system, is different from the starting point, because other pressures have induced a shift (for most ecosystems a shift to a less desirable state). Finally, ecosystems can display combinations of hysteresis and shifting baselines (Fig. 12.1d).

One major problem in analyzing observations of pressure versus response and identifying the most appropriate relationship is that Fig. 12.1 displays a steady-state relationship, whereas observations do not necessarily represent a steady-state situation. The steady state can be assessed when all pressures and other perturbations remain at a constant level for over a sufficiently long time for the ecosystem variables to stabilize. However, ecosystem dynamics are often associated with lags and memory effects having a time scale exceeding that of the sampling. Essentially, this implies that it can be difficult to distinguish hysteresis and shifting baselines from the dynamic output of a linear system. To exemplify this, responses of four

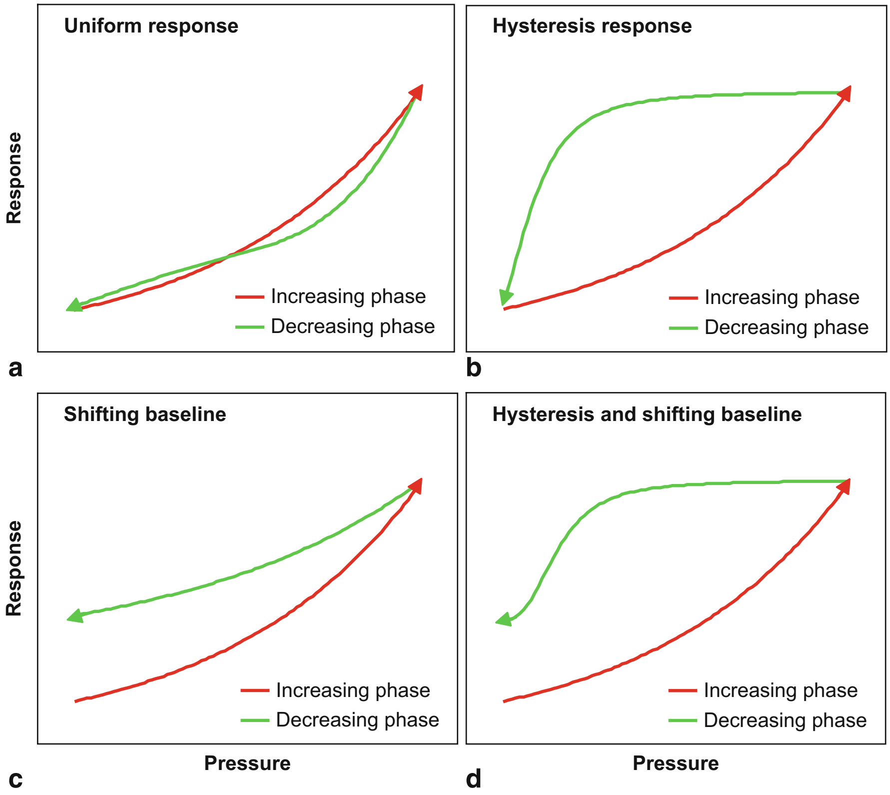


Fig. 12.1 Different types of ecosystem responses to increasing and decreasing pressure. (Redrawn from Duarte et al. 2009)

linear dynamical systems are shown for an increasing followed by decreasing input (Fig. 12.2). If there are no accumulating effects, the dynamical response to an increasing/decreasing input equals the steady-state solution (Fig. 12.2a), but there can be a delay (delayed exponential response) if the ecosystem variable linearly depends on the input as well as previous states of the response (Fig. 12.2b). Lagged responses or simple delays (i.e., the ecosystem variable depends on past and not present values of the input variable) also give rise to delays, albeit less smooth, in the response (Fig. 12.2c). Finally, a linear dynamical system combining lag and memory effects can display almost hysteresis-like behavior (Fig. 12.2d). Thus, the example illustrates that separating nonlinear dynamics leading to hysteresis and alternative stable states from linear dynamical systems can indeed be difficult.

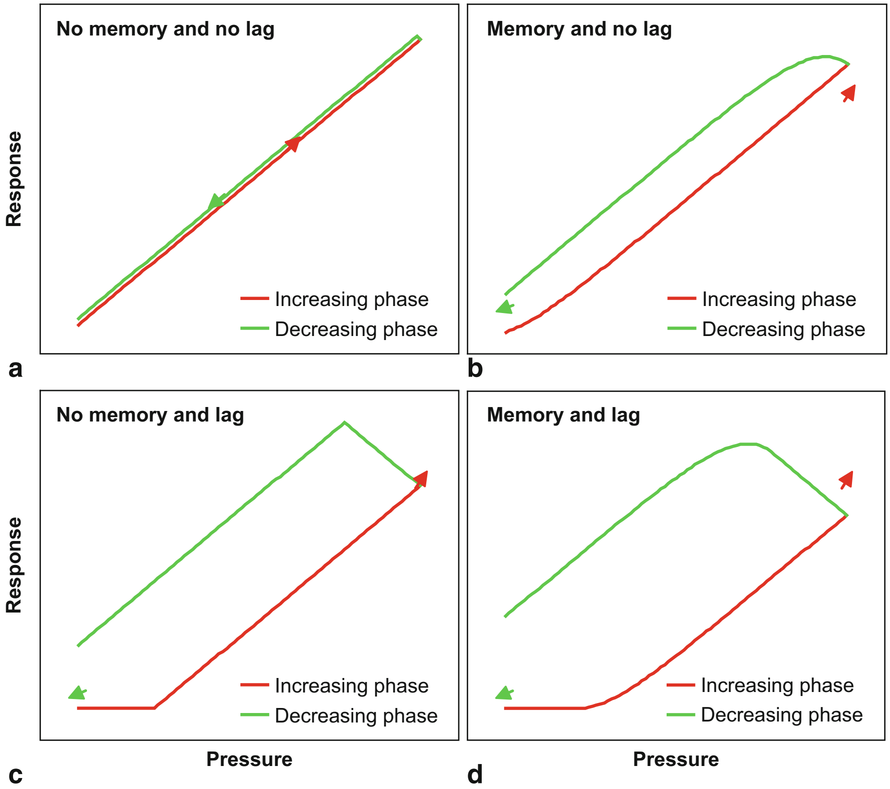


Fig. 12.2 Linear system responses to increasing and decreasing driver. **a** Direct response or simple gain, **b** memory effect (response equals 90 % weight on previous state and 10 % on driver), **c** lagged response to the driver, and **d** combination of memory effect and lagged response

Methods for Identifying Pressure-Response Relationships

The theory for identification of linear dynamical systems is well described and involves estimating the impulse response function from which the structure of the transfer function can be inferred and subsequently estimated (Box and Jenkins 1976). These standard-identification procedures are applicable only to input-output relationships (open loop) and should not be employed when there is feedback in the system (closed loop) (see e.g., Chatfield 1984 for discussion). However, I will refer to the literature for more details on identification of linear systems.

The framework for nonlinear model selection is less rigorous than for linear models, and essentially boils down to formulating a number of candidate models that are subsequently compared by various goodness-of-fit criteria. Two of the most common are Akaike's Information Criterion (AIC) (Akaike 1974) and Bayes Information Criterion (BIC) (Schwarz 1978) that combine the maximum likelihood with a penalty for the number of parameters in the model. Application of these criteria may lead to different optimal models and they should be used only as a guideline in the model selection process. Hence, nonlinear modeling involves a large degree of subjectivity in the formulation of alternative models, and for ecosystem modeling this implies careful consideration of the mechanisms underlying the observations. It should also be stressed that the use of information criteria selects for the best-fitting model, but it does not provide a formal testing to determine if one model is significantly better than another.

The above-mentioned methods for identification of linear systems and selection of the most appropriate nonlinear model underlie the assumption that the observations can be described by a parametric distribution and that the residuals are independent. This also implies that the residuals should be uncorrelated with time. Many ecosystem studies have suggested that pressure-response relationships are changing with time (see above), but few have provided statistical inference to support these conclusions (e.g., Hagy et al. 2004; Conley et al. 2009a; Carstensen and Weydmann 2012). There are many model diagnosis tools available to test the assumptions of proposed regression models (e.g., cross-validation, autocorrelation, correlation with time, Portmanteau lack-of-fit), but I will focus on examining the time trajectory of the ecosystem response to the pressure and provide a test for the significance of departures from a proposed uniform relationship using the statistical framework of Generalized Additive Models (GAM) (Hastie and Tibshirani 1990).

Let us assume that there is a uniform relationship between the pressure (x) and the response given by the parametric function denoted $f(x)$. Let us also assume that we can describe potential departures from this relationship with a smooth nonparametric function of time, denoted $s(t)$. Consequently, we are interested in testing if the combination of $f(x)$ and $s(t)$ gives a significantly better description of the response variable than just $f(x)$ alone. This can be formalized such as: Given there are n pairs of observations for the pressure and response variables ($x_i, y_i; i = 1 \ldots n$), where y_i belongs to the exponential family of distributions (e.g., Normal, Binomial, Poisson, Gamma) with location parameter (μ) that is linked (through a link function $g(\mu)$) to $f(x)$ and $s(t)$, we can test the null hypothesis

$$\mathrm{H0} : g(\mu) = f(x)$$

versus the alternative

$$\mathrm{H1} : g(\mu) = f(x) + s(t).$$

Both the parametric ($f(x)$) and the nonparametric ($s(t)$) functions are estimated by means of the backfitting algorithm (see Hastie and Tibshirani 1990) that iteratively finds an optimal fit for both functions. The significance of the alternative hypothesis is tested by calculating the log ratio of the two models' maximum likelihood values (likelihood ratio test), which is approximately $\chi^2(df)$-distributed with df equal to the approximate degrees of freedom of the smoothing function $s(t)$. The likelihood ratio test applies only because the model under H0 is a submodel of the full model (H1). A simple example of the test above is a normal distributed response with f(x) being

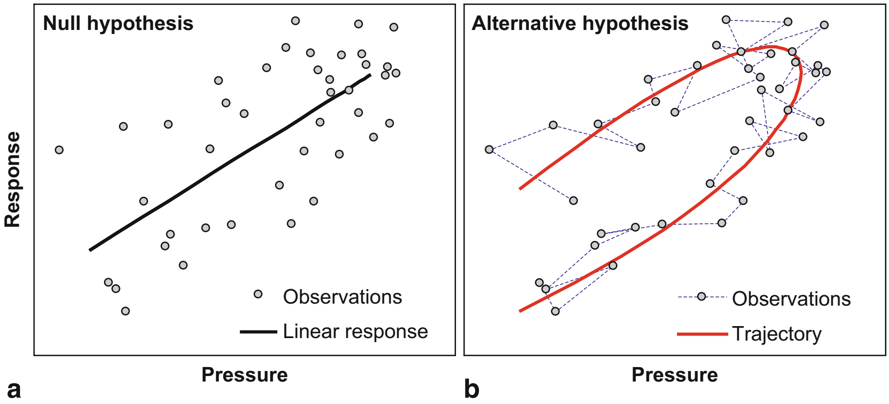


Fig. 12.3 Illustration of the linear model hypothesis (**a**) versus the time trajectory hypothesis (**b**). Observations were simulated from the trajectory model in (**b**) with a relatively small noise added. To illustrate the time dependency of the alternative hypothesis observations were connected in time (**b**)

linear and the identity function used as link function, i.e., $g(\mu) = \mu$ (Fig. 12.3). The pressure-response relationship appears linear when the time dependency of the observations is disregarded (Fig. 12.3a), but if observations are connected to constitute a time trajectory a slightly more complicated relationship emerges (Fig. 12.3b). Thus, the test formulated above examines if the likelihood of the alternative hypothesis is larger than the likelihood of the null hypothesis with sufficient confidence.

The smoothness of $s(t)$ is governed by df with lower degrees of freedom leading to rather smooth fit, whereas higher degrees of freedom result in wiggly curves. GAM normally offers to estimate the optimal degrees of freedom by cross-validation, and this is the recommended setting. However, occasionally GAM does overfit the data using cross-validation and this is reflected in high degrees of freedom in the smoothing function. Thus, if the degrees of freedom gets high (> 4 as a rule of thumb) it is recommended to constrain the degrees of freedom to a maximum of four.

The trajectory of the pressure-response relationship can be graphically shown by predicting the response variable as function of f(x) and s(t). This will produce a smooth trajectory, provided that the pressure is a smooth function of time (continuously increasing/decreasing). If the trend in the pressure variable is noisy, in the sense that there are temporal fluctuations in addition to the overall trend, the resulting trajectory based on predictions from a fluctuating input will result in less smooth trajectory. Consequently, for displaying the trajectory it can be recommended to smooth the pressure-input variable first and subsequently use the smoothed pressure time series for predicting (scoring) responses. The trajectory analysis will be exemplified in the following with a simulation example and by means of observations from four coastal ecosystems.

Noise Contamination Simulation

The noise added to the trajectory model for illustrating the difference between the two hypotheses in the previous section was small and the trajectory of the pressure-response relationship was still apparent from the observations (Fig. 12.3b). In such cases the observations themselves convincingly demonstrate a departure from the simple linear model. This is not necessarily the case if more noise is added to the relationship. To investigate the behavior of the method with more noisy data I have considered the following alternative situations: (1) the underlying relationship between pressure and response is linear versus time trajectory, (2) the pressure is increasing and decreasing linearly without versus with random variation (e.g., interannual variation), and (3) observations of the response variable are noisy versus observations of both pressure and response variables are noisy. These eight combinations were analyzed for different magnitudes of random variation. For this purpose I used the normal distribution for simulating random variates and defined a noise ratio as the standard error of the random variation divided by the range of variation in the pressure and response variables. Moreover, as many simulations and estimations were carried out without user intervention the risk of overfitting GAM was tackled by fixing the degrees of freedom of the GAM to four. However, it is important to stress that the idea was not to perform a complete power analysis to decipher when the method successfully identifies an existing trajectory for various combinations of random variation and number of observations.

One might expect that the GAM would be significant only for the four cases based on an underlying trajectory model, but two out of the four simulated examples with an underlying linear model also had a significant trajectory (Fig. 12.4c, g). At first glance this might seem surprising; however, both examples had observational noise on the observed values of the pressure variable, whereas the two other examples, with an underlying linear model and observation noise in the response variable only, did not result in significant departures from linearity (Fig. 12.4a, e). An explanation is that ordinary regression methods do not account for uncertainty in the independent (or explanatory) variable, so observation noise may lead to significant departures from linearity by sheer coincidence. Secondly, it should be noticed that the linear model in Fig. 12.4c is not significant (regressions slope not different from zero) and the linear slope in Fig. 12.4g is significant, albeit with less confidence than Fig. 12.4a, e. Moreover, the method also failed to identify the linear part of the trajectory with observation noise on both pressure and response (Fig. 12.4d) and the estimated linear component in Fig. 12.4h had a slope substantially lower than the linear part of the simulated trajectory (slope $= 1$). In fact, all the simulations with observational noise on the pressure resulted in slope estimates significantly lower than 1 ($P < 0.0001$ in Fig. 12.4c, d, g, and h, assessed by t-tests of the parameter estimates), whereas the simulations without observational noise all had slopes not significantly different from 1 ($P > 0.1$ in Fig.4. 4a, c, e, g). These results show that the nonparametric smooth curve is actually capable of explaining the linear relationship with the pressure variable as part of the smooth trend, and that the GAM to some extent render the linear model insignificant, when observation noise is added to the pressure variable, whereas this does not seem to be the case when there is no observation noise on the pressure.

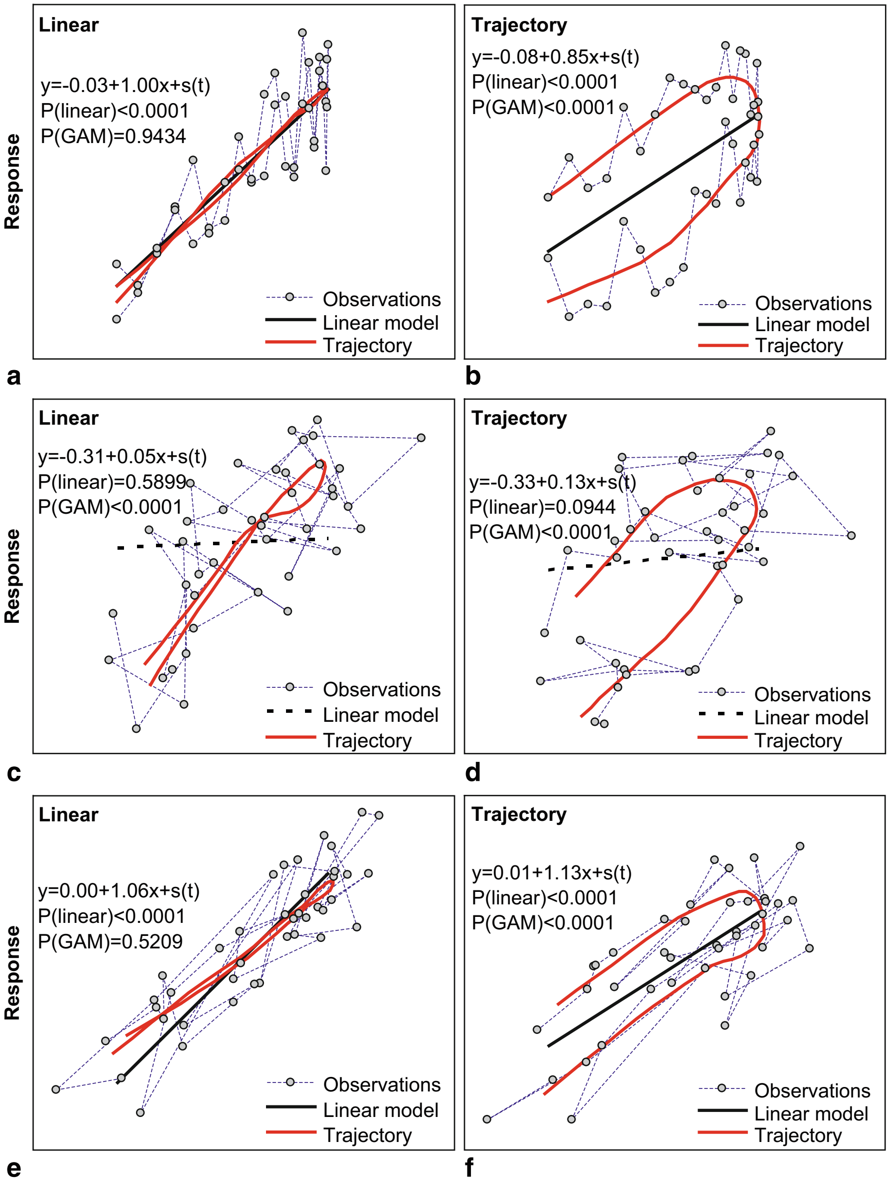


Fig. 12.4 Simulated observations from a linear (*left panel*) and trajectory (*right panel*) pressure-response relationship. **a** and **b** have observation noise on the response variable only. **c** and **d** have observation noise on both pressure and response variables. **e** and **f** have observation noise on the response variable and random variation added to the increasing and decreasing pressure trend. **g** and **h** have observation noise on the response variable and both random variation and observation noise on the pressure variable. The underlying linear model and the linear component of the trajectory had slopes equal to 1 and no intercept. The smooth component was simulated with a sine function of time. Noise ratio was set to 10 %

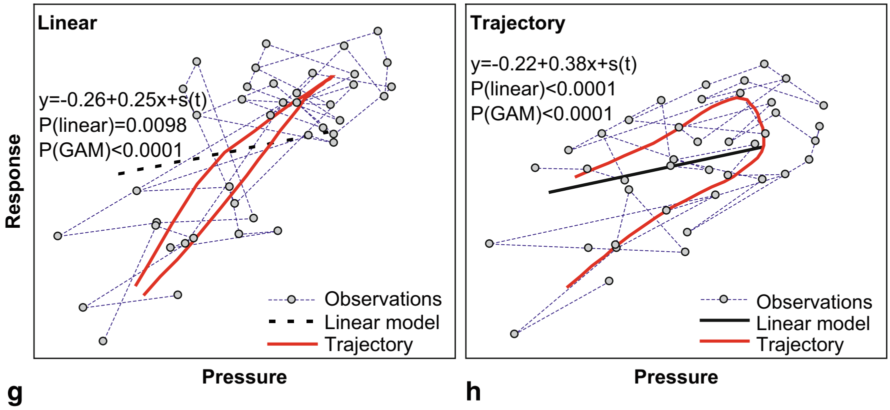


Fig. 12.4 (continued)

The results exemplified in Fig. 12.4 did not represent a single isolated case but were confirmed by numerous simulations. For each of the eight different combinations in Fig. 12.4, the probability of finding a significant time trajectory was estimated as the proportion of 10,000 replications having a significant time component (*s*(*t*)) in the GAM. The linear model with observation noise only (Fig. 12.4a) had about 10–11 % probability for a significant time trajectory, whereas the linear model that also included random variation in the pressure (Fig. 12.4e) had about 17–19 % probability for a significant time trajectory (Table 12.1). These probabilities did not decrease with increasing noise ratio, as was the case for all the other models. Both the linear model and the trajectory model with observation noise on both pressure and response (Fig. 12.4c, d) generally gave higher probabilities for a significant time trajectory than the other linear and trajectory models (Table 12.1). The probabilities for identifying a time trajectory decreased the most with the noise ratio for the models that included the most uncertainty components, i.e., observation noise on both pressure and response and random variation in pressure (Fig. 12.4g, h). Overall, there was a high probability for finding a significant time trajectory, when present, for noise ratios up to 60 %, yielding a power of approximately 80 % (Table 12.1). Even when the noise approached the range of variation in the data (noise ratio ~1) there was still a considerable probability (> 40 %) for identifying a significant time trajectory, when present (Table 12.1).

Coastal Ecosystem Recovery Example

Duarte et al. (2009) brought the concept of regime shifts and shifting baselines from theory to practice by showing that these phenomenas actually take place and should be considered in ecosystem management. Four coastal ecosystems with long-term monitoring data, all having experienced increasing nutrient inputs in the 1970s and 1980s followed by decreasing nutrient inputs during the last two decades, demonstrated idiosyncratic trajectories of phytoplankton biomass versus nutrient inputs. In

Table 12.1 Probability of identifying a significant time trajectory (GAM) for the eight simulation models exemplified in Fig. 12.4 (cf. letters in first column)

Model	Model type	Observation noise		Pressure variation	Noise ratio (%)									
		Response	Pressure		10	20	30	40	50	60	70	80	90	100
(A)	L	X			10.2	11.1	11.2	10.7	10.9	11.2	10.9	11.0	10.7	10.5
(B)	T	X			100.0	100.0	100.0	99.0	92.1	80.7	67.5	56.5	48.1	40.9
(C)	L	X	X		99.5	99.2	97.4	92.8	84.4	75.3	67.3	59.6	52.9	47.2
(D)	T	X	X		100.0	100.0	100.0	100.0	99.4	96.9	92.5	85.7	79.1	71.6
(E)	L	X		X	16.7	17.6	18.6	18.3	17.8	17.7	17.5	18.1	18.5	18.1
(F)	T	X		X	100.0	100.0	100.0	99.1	94.5	84.6	73.6	63.4	55.5	49.4
(G)	L	X	X	X	81.6	70.0	54.5	44.1	36.5	31.0	27.6	25.9	24.7	23.1
(H)	T	X	X	X	100.0	100.0	99.9	97.3	88.6	77.1	66.1	57.3	51.2	43.8

The probabilities were estimated as the number of significant ($P < 0.05$) nonparametric time trajectories ($s(t)$) out of 10,000 simulations for each combination of models and noise ratio. The same noise ratio was applied to all three sources of random variation

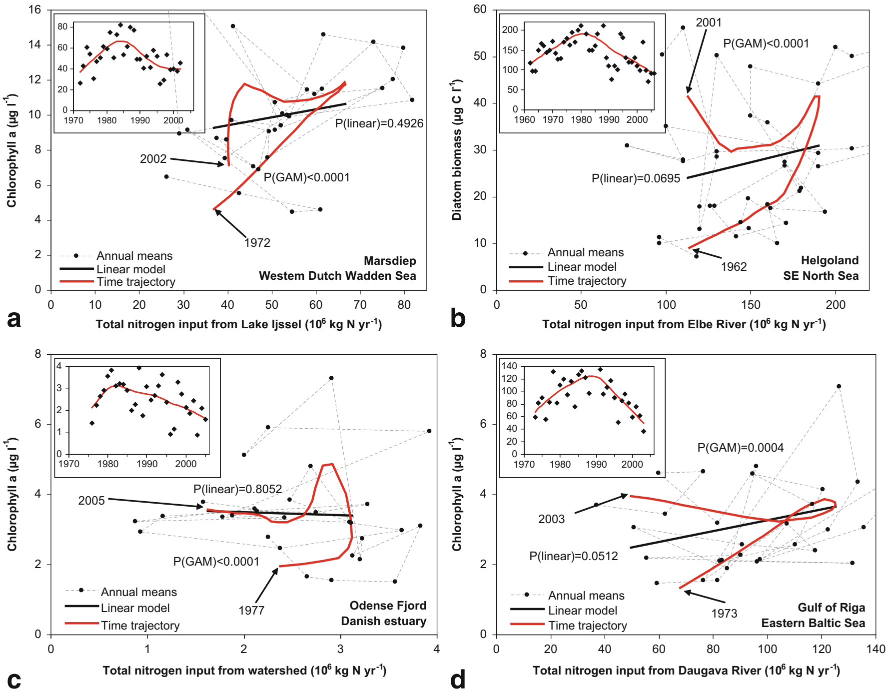


Fig. 12.5 Reanalysis of trajectories from Duarte et al. (2009) using the GAM method presented here. For comparison the linear model under the null hypothesis is also given. The sample trajectories represent four intensively studied Northern European coastal ecosystems that experienced significant eutrophication followed by oligotrophication. The full *black* symbols show the annual average values and the *red* line shows the smooth trajectory developed here. Initial and final years of the time series are indicated. *Inserts* show the time series and smooth GAM trend of total nitrogen inputs to the ecosystems. Note the difference in scaling across ecosystems

all these systems, nutrient inputs approximately doubled from the 1970s to the 1980s and then returned to the level of the 1970s. It was anticipated that the management measures to reduce nutrient inputs would return the ecosystems to their original status, i.e., phytoplankton biomass levels similar to that observed in the 1970s. However, in all four systems, recent phytoplankton biomass concentrations were almost double that of the 1970s despite similar levels of nutrient inputs. The trajectories in Duarte et al. (2009) were computed as 5-year moving averages on both nutrient inputs and phytoplankton annual means to reduce the variation in the data. Thus, although the trajectories in Duarte et al. (2009) graphically displayed departures from an anticipated linear relationship (based on the eutrophication concept originally developed for lakes, see Vollenweider 1968; Dillon and Rigler 1974), there was no statistical evidence of this. Therefore, I reanalyzed these data to examine if the time trajectories were significantly different from a linear pressure-response relationship.

All four coastal ecosystems had considerable variation in both pressure and response variables, without any visually discernible pressure-response relationship from the annual means (Fig. 12.5). None of the four systems actually had a

Table 12.2 Statistics for the null hypothesis (H0: linear pressure-response model) versus the alternative hypothesis. H1: linear pressure-response model and smooth time departure for the four coastal ecosystem exemplifying the method. For the linear models (both H0 and H1) the parametric regression line (*f(x)*) and the probability for zero slope are given. For the nonparametric time smoother (*s(t)*) the degrees of freedom (d.f.) for the smoother and its significance are given. Coefficients of determination were calculated from the model deviance

Coastal ecosystem	No. of years	Null hypothesis				Alternative hypothesis			
		f(x)	P(*f(x)*)	R^2	*f(x)*	P(*f(x)*)	d.f. *s(t)*	P(*s(t)*)	R^2
(A) Marsdiep	31	$7.57+0.046x$	0.2310	0.0643	$11.0-0.020x$	0.4926	3.36	<0.0001	0.5048
(B) Helgoland	40	$14.0+0.090x$	0.0695	0.0841	$21.4+0.042x$	0.1966	3.89	<0.0001	0.6450
(C) Odense Fjord	29	$3.65-0.077x$	0.8052	0.0023	$3.35+0.043x$	0.8485	3.88	<0.0001	0.5555
(D) Gulf of Riga	31	$1.72+0.016x$	0.0512	0.2164	$0.91+0.024x$	0.0007	2.25	0.0004	0.5144

distinctive linear response to changing nutrient inputs (null hypothesis), although the Helgoland and Gulf of Riga data (Fig. 12.5b, d) were borderline significant with *P* values close to the standard significance level of 5 % (Table 12.2). For the Marsdiep data the chlorophyll yield to increasing nutrient input was still positive under H0, albeit nonsignificant (Fig. 12.5a), and the Odense Fjord data actually gave rise to a weak negative linear relationship (Fig. 12.5c). In fact, testing for a linear model only in these four systems would suggest that there is no relationship between phytoplankton biomass and nutrient input, a result that is in contrast to our general conceptual understanding of coastal ecosystem behavior. Such analyses, based on the assumption of a time-invariant relationship between nutrient input and phytoplankton biomass, could potentially lead to erroneous conclusions for nutrient management in the coastal watersheds. The lack of explanatory power under the null hypothesis was also seen in low R^2-values ($< 22\,\%$, Table 12.2).

The alternative hypothesis, including both a linear pressure-response model and a smooth time trend, explained considerably more variation in data ($R^2 \sim 50$–$65\,\%$) but the linear component did not change much from that of the null hypothesis (Table 12.2). Thus, the smooth time trend accounted for most of the explained variation, and the *P* values associated with *s(t)* indicated a high significance. The smooth trend was selected by general cross-validation for the Marsdiep and Gulf of Riga data, whereas the degrees of freedom for the smoother were constrained to be less than four for Helgoland and Odense Fjord data. The general cross-validation method resulted in degrees of freedom equal to 9.11 and 6.54 for these two ecosystems, respectively, and therefore, the wiggliness of the smoother had to be constrained. The estimated time trajectories (Fig. 12.5) generally showed the same behavior as those found by moving averages in Duarte et al. (2009), although considerably smoother, and the statistics confirmed that there was indeed a significant departure from the simple linear pressure-response relationship across time (Table 12.2). Thus, the method delivered statistical inference to further support the theory of shifting baselines and regime shifts in coastal-ecosystem responses to nutrient input.

Discussion

Observations from ecosystem monitoring can be quite variable, often spanning several orders of magnitude, resulting in a cloud of scattered observations as the basis for identifying relationships between drivers and responses (e.g., Guildford and Hecky 2000; Ptacnik et al. 2008). The implication of the large data scatter is that many observations are required to identify potential relationships and that the true nature of the relationship is not visible. Today, many ecosystem monitoring programs have been in operation for several decades, thereby alleviating the data requirements for identifying relationships in the presence of noisy data. Despite the substantial source of information that large data sets typically offer, most studies analyze for simple and static relationships only, despite the availability of a large toolbox of statistical methods to gain further insight into the data (Andersen et al. 2009). The trajectory analysis in this study presents a specific application of the wide class of GAM, specifically designed to identify significant time departures from a proposed static relationship in data. As such, the approach does not present a novel statistical development but it documents the usefulness of analyzing time series by means of GAM to test the implicit assumption of time invariance underlying most pressure-response relationship in the literature. Therefore, this study fulfills the intended goal of providing scientists, that are less experienced with the wide variety of statistical methods, a standard approach for exploring structures in their data that may potentially lead to further model development beyond the most common and simple relationships.

The trajectory analysis provides both a general exploratory tool to uncover changes in pressure-response relationships that may not be visible from plotting the data, and a model diagnosis tool. If there are significant time departures from a proposed parametric relationship then clearly the assumption of independence across the residuals is violated and the estimated parametric relationship will be biased. Secondly, systematic deviations may give hints to refining the parametric relationship or extending the parametric component by including additional explanatory variables. For instance, plotting the smooth trend component ($s(t)$) against various explanatory factors may identify other pressures potentially affecting the ecosystem, and subsequently include these as part of the functional relationship ($f(x)$) and reassess if significant time departures are still present. Hence, the trajectory analysis becomes part of a model identification framework. Essentially, such an iterative process can continue until there are no more suggestions for model improvements and/or there are no more systematic time departures from the relationship. This identification framework may also include process-based models, although there are limitations to the number of parameters that can be identified based on statistical principles. For example, the smooth trend component for the four coastal ecosystems (Fig. 12.5) could be plotted against temperature or grazing pressure to develop an improved functional description of phytoplankton biomass responses to multiple pressures. In fact, Jurgensone et al. (2011) showed that the increasing phytoplankton biomass in the Gulf of Riga could be attributed to declines in zooplankton biomass.

The potential confounding of combined lag and memory effects with the smooth trend component was not considered for the four coastal ecosystems above, although these effects could mimic, to some extent, the observed trajectories (cf. Figs. 12.3 and 12.5). Here, model intuition should also play an important role, because phytoplankton regeneration times and the residence times are substantially shorter than the time resolution of the observations (annual values) for all systems. Although internal inputs of nutrient regenerated from sedimenting organic material could have responses on the interannual scale, the processes involved are more subtle and functional relationships to describe these would go beyond the scope of introducing the trajectory-analysis approach. However, it will be important to consider lag and memory effects for other types of ecosystem responses to recovery efforts, particularly those involving long-lived organisms (Jones and Schmitz 2009).

Another issue of confounding effects was revealed in the simulation study, where the smooth trend was also capable of explaining the underlying simulated linear relationship when observational noise was added to the pressure variable (Fig. 12.4c, d, g, h), whereas both the functional relationship and the smooth time trajectory were nicely separated when the pressure variable had no observational noise (Fig. 12.4a, b, e, f). These tendencies were further confirmed from the multiple simulations with different noise ratios (Table 12.1). Thus, the GAM is sufficiently flexible to overrule an existing functional relationship when the exact value of the pressure variable is not known. The simulations indicated that this phenomenon is pronounced only for noise ratios above 10 % on the pressure variable. Most pressure variables are relatively well determined compared to the ecosystem response. Emission estimates for various substances may have noise ratios below 10 % and climate effects, such as temperature increases, can be measured with high precision and consequently, the noise on pressure variables associated with climate change is likely considerably less than 10 %. Thus, for most pressure-response relationships the noise on the pressure variable is such that the underlying relationship can be separated from the smooth trend.

The trajectory analysis assumes separability of the functional relationship and the smooth time trajectory (additive factors under H1), but it could be argued that time interacts with the functional relationship such that the functional shape changes with time. Such models can also be analyzed within the GAM framework using thin plate splines. However, for the principle of parsimony such an avenue of analysis should be pursued only, if the separability assumption is first invalidated. The good thing about separability is that the significance of the functional relationship and smooth time trend can be tested separately, which is not the case with a thin plate spline. Furthermore, there can be an increased risk of overfitting data with a thin plate spline, which requires a less heuristic constraining of the degrees of freedom, compared to an additive form of the functional relationship and smooth time trend.

In summary, the trajectory analysis is a general exploratory tool that identifies time departures in a proposed functional relationship. It can be used in an iterative manner for model diagnosis and development. However, it should be stressed that there are many other tools that have similar objectives, and that all these tools should be used for guidance rather than providing a rigorous modeling framework.

Acknowledgments This research is a contribution to the ATP (contract # FP7-226248) and WISER (contract #FP7-226273) projects funded by Framework Program 7 of the European Commission.

References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 21:716–723.

Andersen, T., J. Carstensen, E. Hernández-García, and C. M. Duarte. 2009. Ecological thresholds and regime shifts: Approaches to identification. *Trends in Ecology and Evolution* 24:49–57.

Botsford, L. W., J. C. Castilla, and C. H. Peterson. 1997. The management of fisheries and marine ecosystems. *Science* 277:509–515.

Box, G. E. P, and G. M. Jenkins. 1976. *Time series analysis, forecasting and control*. 2nd ed. San Francisco: Holden-Day.

Carpenter, S. R., D. Ludwig, and W. A. Brock. 1999. Management of eutrophication for lakes subject to potentially irreversible change. *Ecological Applications* 9:751–771.

Carstensen, J., and A. Weydmann. 2012. Tipping points in the Arctic: Eyeballing or statistical significance? *Ambio* 41:34–43.

Casini, M., J. Lövgren, J. Hjelm, M. Cardinale, J.-C. Molinero, and G. Kornilovs. 2008. Multi-level trophic cascades in a heavily exploited open marine ecosystem. *Proceeding of the Royal Society B* 275:1793–1801.

Chatfield, C. (1984) *The analysis of time series—An introduction*. Chapmann & Hall, London.

Conley, D. J., J. Carstensen, G. Ærtebjerg, P. B. Christensen, T. Dalsgaard, J. L. S. Hansen, and A. Josefson. 2007. Long-term changes and impacts of hypoxia in Danish coastal waters. *Ecological Applications* 17(5):S165–S184.

Conley, D. J., J. Carstensen, R. Vaquer-Sunyer, and C. M. Duarte. 2009a. Ecosystem thresholds with hypoxia. *Hydrobiologia* 629:21–29.

Conley, D. J., S. Björck, E. Bonsdorff, J. Carstensen, G. Destouni, B. G. Gustafsson, S. Hietanen, M. Kortekaas, H. Kuosa, H. E. M. Meier, B. Müller-Karulis, K. Nordberg, A. Norkko, G. Nürnberg, H. Pitkänen, N. N. Rabalais, R. Rosenberg, O. P. Savchuk, C. P. Slomp, M. Voss, F. Wulff, and L. Zillén. 2009b. Hypoxia-related processes in the Baltic Sea. *Environmental Science & Technology* 43:3412–3420.

Crutzen, P. J. 2002. Geology of mankind. *Nature* 415:23.

Daskalov, G. M., A. N. Grishin, S. Rodionov, and V. Mihneva. 2007. Trophic cascades triggered by overfishing reveal possible mechanisms of ecosystem regime shifts. *Proceedings of the National Academy of Sciences of the USA* 104:10518–10523.

De Young, B., R. Harris, J. Alheit, G. Beaugrand, N. Mantua, and L. Shannon. 2004. Detecting regime shifts in the ocean: Data considerations. *Progress in Oceanography* 60:143–164.

Diaz, R. J., and R. Rosenberg (2008). Spreading dead zones and consequences for marine ecosystems. *Science* 321:926–929.

Dillon, P. J., and F. H. Rigler. 1974. The phosphorus–chlorophyll relationship in lakes. *Limnology and Oceanography* 19:767–773.

Duarte, C. M., D. J. Conley, J. Carstensen, and M. Sánchez-Camacho. 2009. Return to Neverland: Shifting baselines affect ecosystem restoration targets. *Estuaries and Coasts* 32:29–36.

Guildford, S. J., and R. E. Hecky. 2000. Total nitrogen, total phosphorus, and nutrient limitation in lakes and oceans: Is there a common relationship? *Limnology and Oceanography* 45:1213–1223.

Guttal, V., and C. Jayaprakash. 2008. Changing skewness: An early warning signal of regime shifts in ecosystems. *Ecology Letters* 11:450–460.

Hagy, J. D., W. R. Boynton, C. W. Keefe, and K. V. Wood. 2004. Hypoxia in Chesapeake Bay, 1951–2001: Long-term change in relation to nutrient loading and river flow. *Estuaries* 27: 634–658.

Hastie, T. J., and R. J. Tibshirani. 1990. *Generalized additive models*. New York: Chapman & Hall.

Holling, C. S. 1973. Resilience and stability of ecological systems. *Annual Review of Ecological Systems* 4:1–24.

Jeppesen, E., M. Søndergaard, B. Kronvang, J. P. Jensen, L. M. Svendsen, and T. L. Lauridsen. 1999. Lake and catchment management in Denmark. *Hydrobiologia* 396:419–432.

Jones, H. P., and O. J. Schmitz. 2009. Rapid recovery of damaged ecosystems. *Public Library of Science ONE* 4:e5653.

Jurgensone, I., J. Carstensen, A. Ikauniece, and B. Kalveka. 2011. Long-term changes and controlling factors of phytoplankton community in the Gulf of Riga (Baltic Sea). *Estuaries and Coasts* 34:1205–1219.

Kemp, W. M., J. M. Testa, D. J. Conley, D. Gilbert, and J. D. Hagy. 2009. Temporal responses of coastal hypoxia to nutrient loading and physical controls. *Biogeosciences* 6:2985–3008.

Knowlton, N. 1992. Thresholds and multiple stable states in coral reef community dynamics. *American Zoologist* 32:674–682.

Ludwig, D., S. Carpenter, and W. Brock. 2003. Optimal phosphorus loading for a potentially eutrophic lake. *Ecological Applications* 13:1135–1152.

May, R. M. (1977). Thresholds and breakpoints in ecosystems with a multiplicity of states. *Nature* 269:471–477.

Mumby, P. J., A. Hastings, and H. J. Edwards. 2007. Thresholds and the resilience of Caribbean coral reefs. *Nature* 450:98–101.

Petit, J. R., J. Jouzel, D. Raynaud, N. I. Barkov, J.-M. Barnola, I. Basile, M. Bender, J. Chappellaz, M. Davisk, G. Delaygue, M. Delmotte, V. M. Kotlyakov, M. Legrand, V. Y. Lipenkov, C. Lorius, L. Pépin, C. Ritz, E. Saltzmank, and M. Stievenard. 1999. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* 399:429–436.

Ptacnik, R., A. G. Solimini, T. Andersen, T. Tamminen, P. Brettum, L. Lepistö, E. Willén, and S. Rekolainen. 2008. Diversity predicts stability and resource use efficiency in natural phytoplankton communities. *Proceedings of the National Academy of Science of the USA* 105:5134–5138.

Rockström, J., W. Steffen, K. Noone, Å. Persson, F. S. Chapin III, E. F. Lambin, T. M. Lenton, M. Scheffer, C. Folke, H. J. Schellnhuber, B. Nykvist, C. A. de Wit, T. Hughes, S. van der Leeuw, H. Rodhe, S. Sörlin, P. K. Snyder, R. Costanza, U. Svedin, M. Falkenmark, L. Karlberg, R. W. Corell, V. J. Fabry, J. Hansen, B. Walker, D. Liverman, K. Richardson, P. Crutzen, and J. A. Foley. 2009. A safe operating space for humanity. *Nature* 461:472–475.

Rodionov, S. N. 2004. A sequential algorithm for testing climate regime shifts. *Geophysical Research Letters* 31:L09204.

Scheffer, M. 1990. Multiplicity of stable states in freshwater systems. *Hydrobiologia* 200/201: 475–486.

Scheffer, M., S. H. Hosper, M.-L. Meijer, B. Moss, and E. Jeppesen. 1993. Alternative equilibria in shallow lakes. *Trends in Ecology and Evolution* 8:275–279.

Scheffer, M., S. Rinaldi, A. Gragnani, L. R. Mur, and E. H. van Nes. 1997. On the dominance of filamentous cyanobacteria in shallow, turbid lakes. *Ecology* 78:272–282.

Scheffer, M., S. Carpenter, J. A. Foley, C. Folke, B. Walker (2001) Catastrophic shifts in ecosystems. *Nature* 413:491–496.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6:461–464.

Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo. 1997. Human domination of Earth's ecosystems. *Science* 277:494–499.

Vollenweider, R. A. 1968. *Scientific fundamentals of the eutrophication of lakes and flowing waters with particular reference to nitrogen and phosphorus as factors of eutrophication* (Organisation for Economic Cooperation and Developmen (OECD), Tech Rep DA S/SCI/68.27.250). Paris: OECD.

Weatherhead, E. C., and S. B. Andersen. 2006. The search for signs of recovery of the ozone layer. *Nature* 441:39–45.

Zeileis, A., C. Kleiber, W. Krämer, and K. Hornik. 2003. Testing and dating of structural changes in practice. *Computational Statistics and Data Analysis* 44:109–123.

Chapter 13
Detection of Harbingers of Catastrophic Regime Shifts in Drylands

Robert A. Washington-Allen, R. Douglas Ramsey, Thomas G. Van Niel and Neil E. West

Abstract Harbingers are early warnings of imminent ecosystem collapse and thus are aids to preventing land degradation. Dynamic systems are hypothesized to exhibit dampening or inflation of critical attributes at or near a threshold, which is a decreasing or increasing spatial or temporal trend, respectively. This behavior is diagnostic of a state change and can be operationalized as an early detection system. Consequently, we used a time series from 1972 to 1997 of seasonal soil-adjusted vegetation index (SAVI) data, a proxy for canopy cover that was derived from Landsat imagery of the Marine Corps Air Ground Combat Center. We used dynamical, trend, and autocorrelation function (ACF) time series analysis to find that the time series had an increasing linear trend that correlated with wet periods of the Pacific Decadal Oscillation (PDO) and the El Niño-Southern Oscillation (ENSO). High and low SAVI values were in wet and dry basins of attraction with a rapid shift from dry to wet period from 1981 to 1982. Mean SAVI dampening appears to have occurred 2–3 years prior to the shift. Consequently, this study suggests that this dampening trend of the mean SAVI can be used as a harbinger of land degradation.

Keywords Washington Allen—abrupt transition · Early warning · Imminent change · Remote sensing · Threshold · Time series

Introduction

The major dust storms that occurred during the US Dust Bowl era of the 1930s were hypothesized to be the result of the loss of vegetation cover at a critical point that led to bare-ground patches that were fragmented at local scales abruptly increasing in

---

R. A. Washington-Allen (✉)
Department of Geography, University of Tennessee,
Burchfiel Geography Building, Room 304, Knoxville, TN 37996–0925, USA
e-mail: washingtonra@utk.edu

R. D. Ramsey · N. E. West
Department of Wildland Resources, Utah State University, Logan, UT 84322–5230, USA

T. G. Van Niel
Division of Land and Water, Commonwealth Scientific and
Industrial Research Organization (CSIRO), Wembley, WA 6913, Australia

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_13,

number, area, and connectivity to the regional scale of the western USA (Peters et al. 2004). This loss of vegetation cover and the resulting dust storms were attributed to multiyear severe droughts and land mismanagement that led to loss of agricultural livelihoods, the economic collapse of the agricultural sector, and the creation of the Soil Conservation Service, now the Natural Resources Conservation Service (NRCS) by the US Department of Agriculture. The NRCS was created to prevent this type of catastrophe from occurring again in rangelands and it would be advantageous to them and other land management agencies if changes in rangeland ecosystem attributes, e.g., vegetation cover, that lead to land degradation could be detected prior to a threshold being exceeded. This idea would require the early detection of harbingers or indicators of the imminent approach to a threshold (Brock and Carpenter 2006; Scheffer et al. 2009).

One such harbinger is the observed increase or decrease, also called negative and positive dampening, of the variability of measured community or ecosystem-level characteristics, e.g., vegetation cover or bare-ground patch dynamics, in time and space. This behavior harkens the change from one state to another or the approach to a threshold (Allen et al. 2005; Brock and Carpenter 2006; Wardwell and Allen 2009; Biggs et al. 2009; Scheffer et al. 2009). Time series analysis of autocorrelation in individual ecosystem characteristics can be used to detect harbingers that are represented as changes in key drivers and positive feedback interactions that are expressed as near shifts or breaks in scale (Briske et al. 2010). Ludwig et al. (2000) have shown that these large fluctuations in resource variability increase the success of random events to affect system reorganization. Consequently, the purpose of the study presented in this chapter is to determine whether the temporal behavior of a remotely sensed indicator of Dryland vegetation response, i.e., a vegetation index, under conditions of a known climatic regime shift can be used to retrospectively detect local-site behavior that are harbingers or early warning indicators of the pending approach of a threshold. A secondary purpose of this study is to detect the more and less ecologically resilient portions of Drylands that are subject to military training and testing activities. The US military installations occupy over 12 million ha of land, much of which is in Drylands, and hosts the largest population density of endangered species on federal lands (Benton et al. 2008). Consequently, this information would be helpful for directing military training exercises to the more productive areas of a landscape as well as directing ecological restoration practices to the most vulnerable. Consequently, this study will result in tools that will allow responsive land management from managers, stakeholders, and other decision makers (Scheffer et al. 2009).

Definitions

Drylands are areas where the ratio of the mean annual precipitation to the mean annual potential evapotranspiration ranges between 0 and 0.65. Drylands cover some 41 % of the terrestrial surface and provide ecosystem goods and services for 36 % of the world's population (MEA 2005; Reynolds et al. 2007). However, the ecological condition and trend of Drylands at regional and global scales is unknown, primarily because they have demonstrated characteristics that are diagnostic of complex adaptive systems including emergent spatial patterns at the landscape scale (e.g., Rietkerk et al. 2004), multiple dynamic regimes or states in space and time (Archer 1989; Westoby et al. 1989; Lockwood and Lockwood 1993; Mayer and Rietkerk 2004), and nonlinear discontinuous or abrupt transitions at thresholds in space (e.g., ecotones) and time (Archer 1989; Scheffer et al. 2001).

There are a number of synonyms for thresholds including transition zones and tipping points (Washington-Allen et al. 2010). We define thresholds in the manner of Archer (1989), Westoby et al. (1989), Friedel (1991), and Scheffer et al. (2001) where a threshold is the abrupt transition from one state to another. Beisner et al. (2003) demonstrated that a state can be viewed from the community and/or the ecosystem levels of organization. For example, the US land management agencies used the concept of vegetation succession, specifically the compositional change in either plant functional types or physiognomic structure, e.g., the change of grassland to woodland, with respect to a reference state (usually the climax), as a way to monitor the ecological condition and trend of rangelands (Dyksterhuis 1949; Westoby et al. 1989; West 2003). This approach assumed linear and predictable vegetation dynamics, but actual observations demonstrated both discontinuous and continuous behavior in Dryland composition, spatial pattern, and biogeochemistry (Westoby et al. 1989; Stringham et al. 2003; Washington-Allen et al. 2008, 2009). Consequently, this community-level perspective has been expanded to include ecosystem science where changes in hydrological or biogeochemical parameters such as soil infiltration or nitrogen cycling are monitored and assessed in conjunction with vegetation compositional changes (Schlesinger et al. 1990; Stringham et al. 2003; Betlemeyer et al. 2004; Peterson et al. 2009).

Dryland degradation can be defined as a decrease in plant cover, density, productivity, or some other plant or vegetation parameter or measurement of attributes (Washington-Allen et al. 2004a, b; Washington-Allen et al. 2006). It is difficult to separate the relative contributions of climate, fire, and land management practices, particularly livestock grazing to changes in ecological indicators. Time series datasets of ecological indicators and drivers for 10 or more years, are required in order to replicate climatic drivers, such as the El Niño–Southern Oscillation (ENSO) that has a return interval of 3 to 7 years (Glantz 2001), in order to be able to separate climatic impacts from that of land management practices (Washington-Allen et al. 2006). The 40-year Landsat satellite image archive (1972 to the present), excluding Landsat 8, has pixel resolutions of 15 m in the panchromatic (black and white) and 30 m resolution for six of its multiple bands, and both 60 m and 120 m resolution for thermal bands. Ecological indicators that are diagnostic of vegetation parameters can be derived from the spectral and textural characteristics of historical imagery like Landsat (Tueller 1989; Quattrochi and Pelletier 1991; Washington-Allen et al. 2006). For example, the spectral characteristics of Landsat from blue to shortwave-infrared in detecting the reflectance of vegetation allows the development of biophysical models such as the Normalized Difference Vegetation Index (NDVI) (Rouse et al. 1973) that

Fig. 13.1 A true-color Landsat image in the southern Mojave Desert of the study area: the Marine Corps Air Ground Combat Center (MCAGCC) in Twentynine Palms, CA

relies on the difference in percent reflectance of incident near infrared (NIR) and red (R) energy from leaf tissue. NDVI is calculated as:

$$\mathrm{NDVI} = (\mathrm{NIR} - \mathrm{R})/(\mathrm{NIR} + \mathrm{R}) \tag{13.1}$$

NDVI has been significantly correlated with vegetation attributes that are commonly collected in the field including leaf area index (LAI), plant cover, phytomass, and net primary productivity (Sellers 1985). A nearly 40-year time series of Landsat NDVI can be characterized to look at possible state changes.

Methods

Study Area

The 238,645 ha Marine Corps Air Ground Combat Center (MCAGCC) was established in 1952 and is a training facility that is located in the Mojave Desert, near Twentynine Palms, California (Fig. 13.1; NRED 1999). MCAGCC is in the Great Basin section of the Basin and Range physiographic province in mountain ranges, hills, alluvial fans, drainages, playas, and lava flows. Elevations on the facility range from 213 m to 1250 m above sea level. Soil textures range from gravelly fine sand to gravelly very fine sand with indurated calcic horizons. The vegetation is sparse and is comprised mainly of small shrubs and grasses of which the predominant plant species are creosotebush (*Larrea tridentata*), burrobrush (*Coleogyne ramosissima*), cheesebush (*Hymenoclea salsola*), and catclaw acacia (*Acacia greggi*). In the more alkaline areas of the Mojave-dominant species are allscale (*Atriplex polycarpa*), wingscale (*Atriplex canescens)*, and alkali blite (*Chenopodiium rubrum*). The most common grass species is Indian ricegrass (*Oryzopsis hymenoides*). MCAGCC is habitat for the federally listed endangered species, desert tortoise *(Gopherus agassizii).* Land use at MCAGCC consists of live fire, combined arms, and maneuver training that simulates combat situations using ground infantry, tracked vehicles including tanks, and light-armored vehicles. There are also air-to-ground ordinance activities that result in localized impact zones (Natural Resources and Environmental Division (NRED) 1999).

Climatic Events

Temperature in the Mojave Desert ranges from −13.3 °C to 48.3 °C. The average precipitation from 1893 to 2001 was 137 mm year^{-1} with a range from 34 to 310 mm year^{-1} (Hereford et al. 2004). The dry season is May through September and the wet season is October through April (Hereford et al. 2004). Hereford et al. (2004, 2006) have shown that between 1893 and 2001, the Mojave experienced four major drought periods in 1893–1904, 1942–1975, 1988–1991, and 1999 (Breshears et al. 2005; Overpeck and Udall 2010) and two major wet (flood) periods in 1905, 1941, and 1976–1998. Hereford et al. (2004, 2006) using time series spectral analysis showed that these drought and flood events were reflections of both 35-year and 5-year periodicities that relate to the Pacific Decadal Oscillation (PDO), an index of the relative sea surface temperatures (SST) of the northern Pacific Ocean, and the ENSO, an atmosphere–ocean interaction that involves the warming of SST in the tropical Pacific Ocean from the equator to the southern coast of South America (El Niño) and the back-and-forth exchange of air masses between the eastern and western hemisphere at sub- and tropical latitudes (Southern Oscillation), respectively. Hereford et al. (2004, 2006) show that major climatic regime shifts in the Mojave were a function of the PDO cool phase (decreasing SST), that was manifest as the 1942–1975 drought, and PDO warm phase (increasing SST), that was manifest as the 1976–1998 wet period. The PDO warm phase also included the 1982–1985 very strong El Niño (Glantz 2001) and the 1988 La Niña drought called the "Great North American Drought" (Trenberth et al. 1988; Riebsame et al. 1991; Washington-Allen et al. 2009). La Niña is the cooling of SST in the tropical Pacific Ocean from the equator to the southern coast of South America, field studies of vegetation dynamics in the Mojave indicate that plant community composition changes with high mortality of plants during droughts and recruitment during wet periods of both annual and perennial grasses and woody plants (Beatley 1980; Webb et al. 2003; Hereford et al. 2006). The 1989–1991 drought period was particularly noteworthy for the high mortality of perennials (Webb et al. 2003), a phenomena which is currently being observed in the large die-offs of tree species in the drought–stricken southwestern USA (Breshears et al. 2005; Overpeck and Udall 2010). Additionally, Rundel and Gibson (1996) have shown that primary productivity in the Mojave increased under increased precipitation and of course decreased during droughts.

Data Acquisition

A time series of historical dry and wet season Landsat images from 1976 to 1996 (21 scenes) and 1972 to 1997 (20 scenes) were acquired of MCAGCC, respectively. The wet season dataset has a hiatus between the period 1972–1979 and is missing portion of scenes for the period 1973–1978. However, the existing dataset overlaps with the PDO cool phase from 1942 to 1975 and the warm phase from 1976 to 1998, as well as the previously mentioned very strong ENSO event of 1982–1985 (Glantz 2001). The dry season scenes were selected between May and June and the wet season scenes between October and November. For this study, six Landsat platforms with two different radiometers including Landsat Multispectral Scanner (MSS) and Thematic Mapper$^{\text{TM}}$ have been deployed. Because of the differing formats of 5-(Landsat MSS) and 8-bit$^{\text{TM}}$ and 30 m and 79 m pixel resolutions, images were: (1) rectified to a common map projection and resampled to a common resolution (60 m); (2) standardized by conversion to exoatmospheric reflectance values using Landsat postlaunch calibration gains and biases (e.g., Markham and Barker 1986); and (3) atmospherically corrected using a relative atmospheric correction procedure based on the use of pseudoinvariant features for multitemporal imagery (Jensen 2005; Washington-Allen et al. 2004a, b). The criteria for data acquisition were: low-cost acquisition of scenes that were representative of the dry and wet season, $\leq 10\,\%$ cloud cover, and anniversary dates between scenes. Path/Row location of scenes differed between platforms and in some years full scenes of the study area could not be acquired because the midline of the path/row grid passed through the study area, e.g., 1972–1981, 1983, 1985, 1986, 1989–1995 for the wet season imagery. However, at least 50 % of the study area was available for time series analysis. Table 13.1 lists the characteristics of the image data set.

Indicator Development

The effectiveness of NDVI for discriminating vegetation response is substantially reduced in drylands because of the sparser vegetation cover leading to increased sensitivity to soil background moisture and reflectance effects on the vegetation signal. Consequently the soil-adjusted vegetation index (SAVI) was developed from the NDVI to increase the vegetation signal relative to the soil noise (Huete 1988). SAVI is calculated as:

$$\text{SAVI}=[(\text{NIR}-\text{R})/(\text{NIR}+\text{R}+\text{L})]*(1+\text{L}) \qquad (13.2)$$

The L is an adjustment factor which varies from 0–1 in accordance with soil background conditions (Huete 1988). The recommended L factor of 0.5 was used for all images (Huete 1988).

Table 13.1 The Landsat satellite imagery dataset from 1972 to 1997 of the Marine Corps Air Ground Combat Center (MCAGCC) that was used in this study

Acquisition Date	Landsat scanner and number	Season
10/02/1972	MSS 1	Wet
05/16/1976	MSS 2	Dry
06/16/1977	MSS 2	Dry
06/03/1978	MSS 2	Dry
10/28/1979	MSS 2	Wet
06/25/1979	MSS 2	Dry
10/23/1980	MSS 2	Wet
06/01/1980	MSS 2	Dry
11/05/1981	MSS 2	Wet
06/14/1981	MSS 2	Dry
12/12/1982	MSS 4	Wet
05/30/1982	MSS 3	Dry
11/13/1983	MSS 4	Wet
06/29/1983	MSS 4	Dry
11/30/1984	MSS 5	Wet
06/07/1984	MSS 5	Dry
11/1/1985	MSS 5	Wet
06/19/1985	MSS 5	Dry
11/13/1986	MSS 5	Wet
06/06/1986	MSS 5	Dry
11/16/1987	MSS 5	Wet
06/09/1987	MSS 5	Dry
10/24/1988	TM 5	Wet
06/02/1988	TM 5	Dry
10/20/1989	TM 5	Wet
06/05/1989	TM 5	Dry
10/30/1990	TM 5	Wet
06/08/1990	TM 5	Dry
11/11/1991	TM 5	Wet
06/04/1991	TM 5	Dry
10/28/1992	TM 5	Wet
06/14/1992	TM 5	Dry
11/07/1993	TM 5	Wet
06/16/1993	TM 5	Dry
11/19/1994	TM 5	Wet
06/19/1994	TM 5	Dry
10/05/1995	TM 5	Wet
05/30/1995	TM 5	Dry
11/08/1996	TM 5	Wet
06/08/1996	TM 5	Dry
10/26/1997	TM 5	Wet

Mean-Variance Plots

Graphical analysis of a dynamical system uses phase diagrams or portraits that describe the motion or trajectory of states through time (Morse et al. 2000). Pickup and Foran (1987) developed a special case of this analysis called mean–variance analysis (MVA) to characterize the spatiotemporal behavior of a remotely sensed vegetation index (VI). Washington-Allen et al. (2008) provide a conceptual model of a VI in

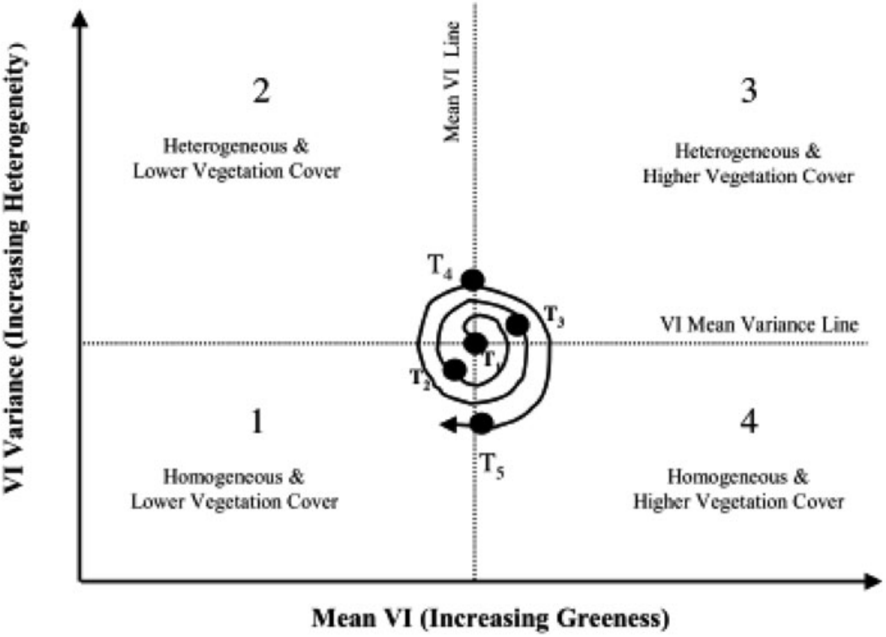


Fig. 13.2 A hypothetical statistical phase portrait of the interannual mean-variance dynamics of an agricultural landscape's vegetation index (VI). T = time. (Used with permission from *Ecology & Society*)

mean-variance space where the hypothetical trajectory of a VI of an agricultural crop that has been bred to minimize variability is described (Fig. 13.2). As previously discussed, the VI is a proxy for vegetation response, particularly percent canopy cover or biomass. Consequently, an NDVI value can be treated as percent canopy cover. The value of the VI variance represents the degree of landscape heterogeneity or balance between bare soil and vegetated patches (Pickup and Foran 1987). Thus, the variance and the mean together provide an indirect measure of a sites' susceptibility to soil erosion, where if a normal distribution is assumed, large VI variances indicate the likelihood that pixels with low VI values at the distribution's tails consist of either reduced vegetation or bare soil cover (Fig. 13.2).

The phase portrait is divided into four sectors or states that are delineated by the grand mean of the VI mean and the VI variance of a landscape's VI time series (Fig. 13.2). Each sector describes the state of a landscape's trajectory. For example, sector 1 (low mean and low variance) can be considered the most degraded state of this landscape, sector 2 (low mean and high variance) indicates that a higher proportion of the landscape tends towards bare ground and thus high susceptibility to erosion, sector 3 (high mean and high variance) means a higher proportion of the landscape has vegetation cover, but depending on skewness, a small proportion of the landscape is susceptible to erosion, and sector 4 (high mean and low variance) would be the most ideal and stable conditions for this VI landscape (Fig. 13.2). Mean-variance plots were developed for whole landscape response as well as for each individual vegetation cover class. If the time series is long enough a mean-variance portrait can be used to delineate both seasonal and interannual dynamics of a landscape and discriminate regime shifts (Washington-Allen et al. 2008 and 2009).

Time Series Analysis

Characterization of the direction and strength of a trend can be accomplished with regression analysis (Yafee and McGhee 2000). A significant slope (ß) is a measure of the direction of trend, i.e., stable (0), increasing (+ ß) and decreasing (− ß), and the magnitude of the coefficient of determination (r^2) from a linear or polynomial regression measures the strength of the trend (Yafee and McGhee 2000). The autocorrelation function (ACF) can be used to detect thresholds in a vegetation index time series (Turchin and Ellner 2000; Washington-Allen et al. 2009). An abrupt and consistent sign change, i.e., (+) to (−) or vice versa and a significant ACF value usually indicate a threshold and allows delineation of the time series into different period states.

Results

SAVI Time Series

Figure 13.3 is the wet (1976–1996, a) and dry (1972–1997, b) season SAVI images of MCAGCC. During the wet season, the vegetation response (SAVI) tends to be from 0.05 to 0.44 throughout the period 1972–1997 (Fig. 13.3). However, during the dry season the SAVI value from 1976 to 1982 ranged from 0.06 to −0.35 and from 1983 to 1996, they ranged between 0.06 and −0.64. Trend analysis of the dry and wet season time series from 1972 to 1997 indicated a significant linear fit ($r^2 = 0.52$, $p < 0.05$) of increasing greenness from 1972 to 1997 (Fig. 13.4a). Examination of the ACF for this SAVI time series indicated an abrupt dampening of ACF values from lag 10 (0.10, 1981 compared to 1982) to lag 11 (0.28) to 15 (1981–1983 compared to 1982–1984) and a switch in ACF sign from (+) to (−) correlation at lag 16 (the dry season of 1985) (Fig. 13.4b). The trend of the SAVI variance from 1972 to 1997 was slightly increasing ($r^2 = 0.032$), but not significant.

Mean-Variance Plots

The wet season phase diagram indicated oscillating temporal dynamics that suggested one domain of attraction about the mean of the variance and the mean from 1972 to

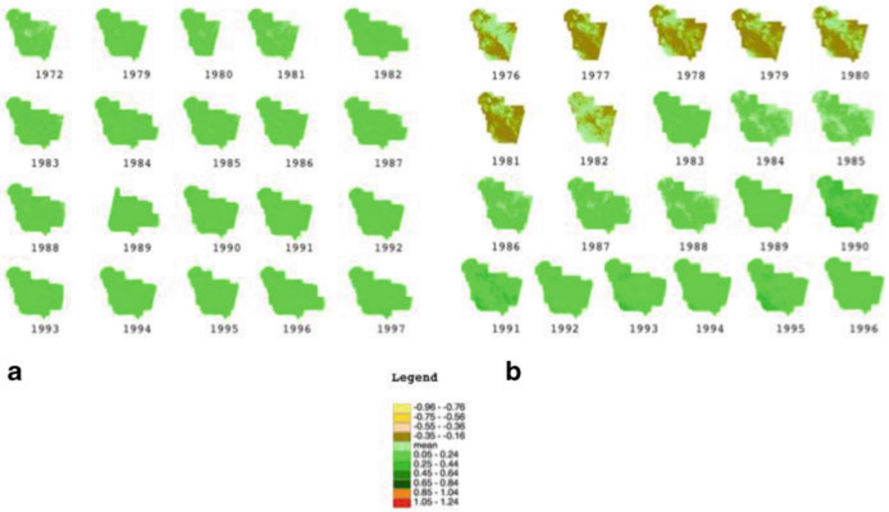


Fig. 13.3 **a** The wet (1972–1997) and **b** dry (1976–1996) season soil-adjusted vegetation index (SAVI) image time series of the Marine Corps Air Ground Combat Center (MCAGCC) in the southern Mojave Desert near Twentynine Palms, CA. The legend value range is ± 1 standard deviations between SAVI values from the mean

1997 (Fig. 13.5a). However, the dry season phase diagram indicated three domains of attraction: (1) from 1976 to 1982 of low SAVI and relatively higher variance, particularly 1980, than the other domains; (2) from 1983 to 1989 the orbit near the SAVI grand mean and variance; and (3) with relatively lower variance and higher mean SAVI from 1990 to 1997 (Fig. 13.5b). When the ACF was calculated on the dry season SAVI mean from 1976 to 1997 a switch in ACF sign from (+) to (−) correlation at lag 6 (1981 compared to 1982) to 7 (1982 compared to 1983) was detected (Fig. 13.6a). The trend in dry season SAVI mean ACF dampening was from lag 1 (1976 compared to 1977) to lag 6 (1981 compared to 1982) (Fig. 13.6a). When ACF was computed on the dry season SAVI variance from 1976 to 1996, a switch in ACF sign from (+) to (−) correlation at lag 6 (1981 compared to 1982) to 7 (1982 compared to 1983) was detected (Fig. 13.6b). The dry season SAVI ACF variance values from lag 3 to 6 decreased or dampened before the switch (Fig. 13.6b).

Discussion

The SAVI Landsat satellite time series was acquired at a time coincident with the major climatic regime shifts of the PDO and ENSO, specifically the 30-year (1942–1975) PDO cool phase drought and the 20-year (1976–1998) PDO warm phase wet period that included the very strong 1982–1985 El Niño wet period and the severe

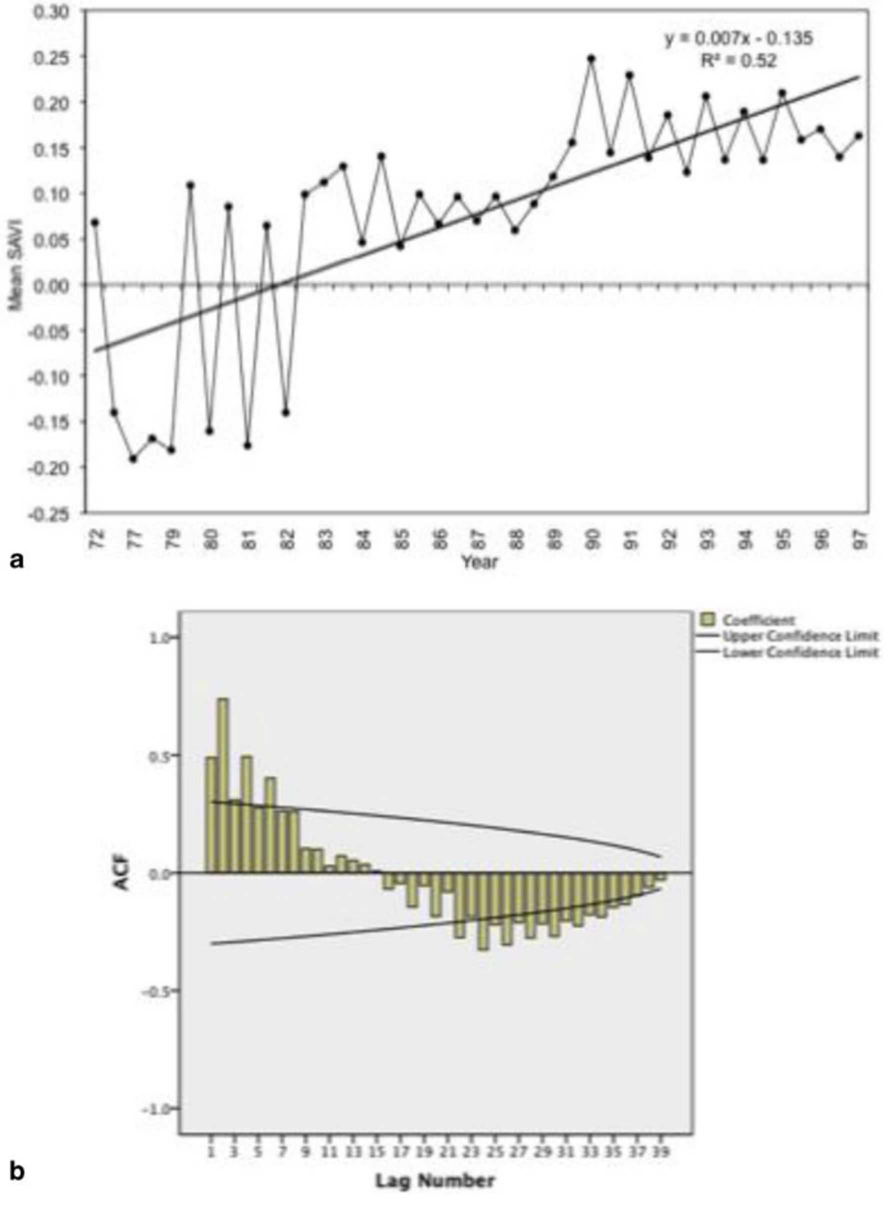


Fig. 13.4 Times series of the combined wet and dry season mean soil-adjusted vegetation index (SAVI) response for the Marine Corps Air Ground Combat Center (MCAGCC) from 1972 to 1997 (**a**). The linear fit suggests an increasing trend in greenness from 1972 to 1997. The autocorrelation function (ACF) of the combined time series (**b**) indicates significant changes during the dry season relative to the wet and that a regime shift had occurred between 1981 and 1982

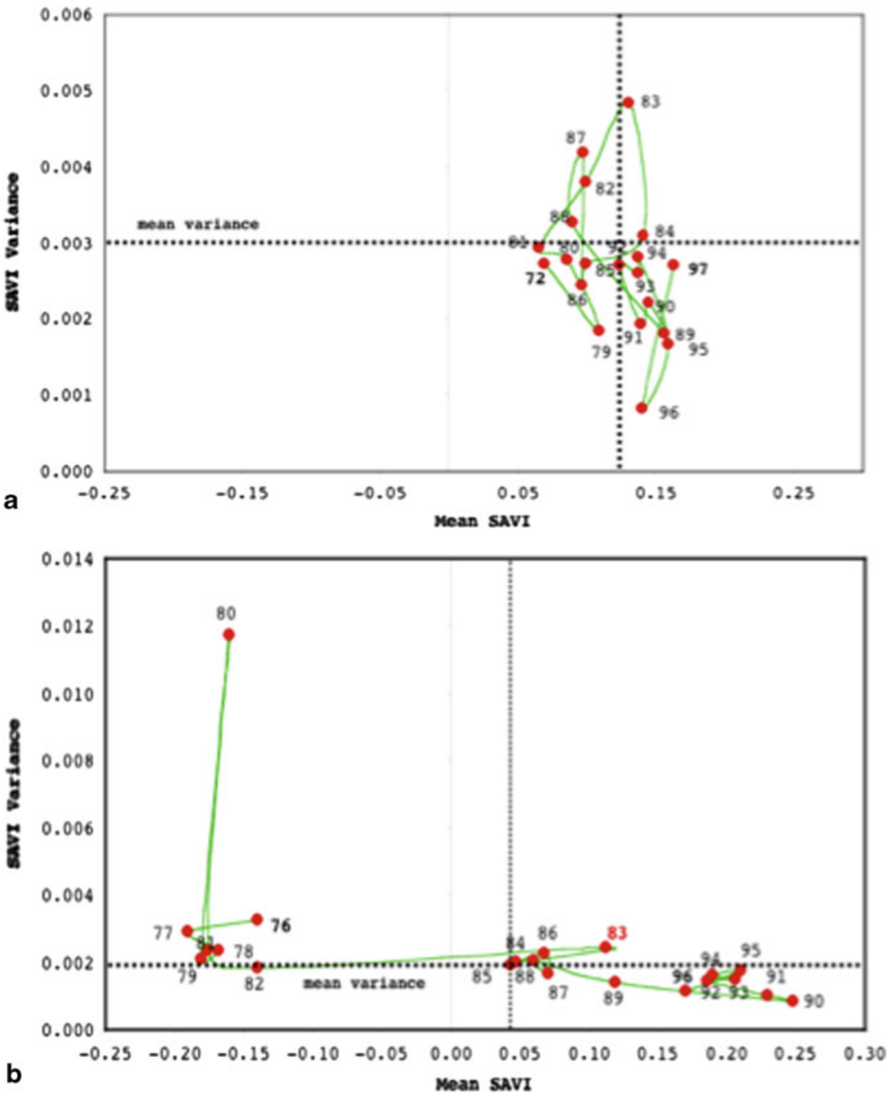


Fig. 13.5 Mean-Variance analysis of the dry (**a**) and wet (**b**) season soil-adjusted vegetation index (SAVI) scenes of the Marine Corps Air Ground Combat Center (MCAGCC) from 1972 to 1997. Three domains of attraction are evident during the dry season time series, whereas only one basin is evident during the wet season

droughts of the 1988–1991 La Niña. However, due to the data hiatus in the wet season imagery from 1973 to 1978, the PDO cool phase impact could not be detected.

The overall response of the Combat Center's SAVI response from 1972 to 1997 was an increasing linear trend that was buoyed by the PDO warm phase wet period and the increased wetness from the ENSO from 1982 to 1985 (Fig. 13.4). The

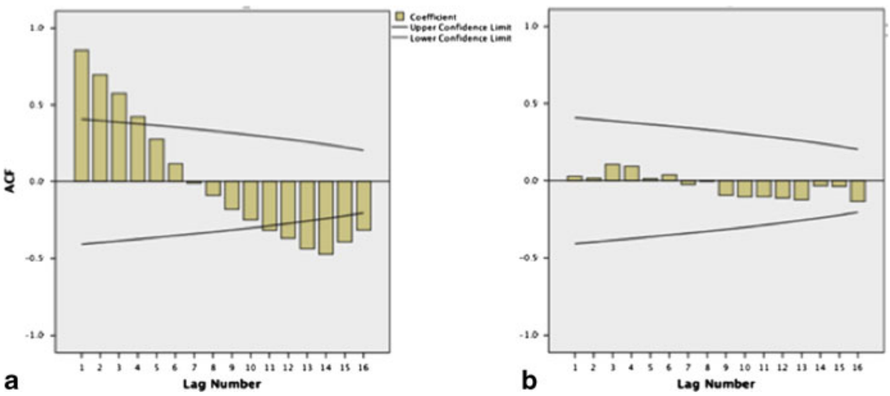


Fig. 13.6 The autocorrelation function (ACF) of the dry season mean soil-adjusted vegetation index (SAVI) (**a**) and the variance of SAVI (**b**) time series of the Marine Corps Air Ground Combat Center (MCAGCC) from 1976 to 1997

temporal response from 1972 to 1997 was bimodal with an abrupt shift occurring around 1981–1982 towards greater vegetation production. This regime shift was prominently detected by the dry season mean-variance and ACF analyses (Figs. 13.5 and 13.6) and coincided with the increased wetness of the southern Mojave with the inception of the El Niño wet period. The impacts of the droughts on vegetation were prominently detected in the SAVI image dry season time series from 1976 to 1981. It is likely that areas of high mortality were detected in this time sequence and could be further studied for persistence of drought impacts (Breshears et al. 2005; Washington-Allen et al. 2004a). This also applied to productive portions of the landscape and would be helpful for directing military training exercises to the more ecologically resilient portions of the Combat Center. This assessment of the response of the vegetation as measured by SAVI from 1972 to 1997, was consistent with the field vegetation and climate studies by Beatley (1980) and Hereford et al. (2006). In particular, the dry season response was more clearly indicative of vegetation response to climate than the wet season as evidenced by significant dry season ACF values in the overall time series (Fig. 13.4b). The dry season phase portrait indicated three domains of attraction for the southern Mojave with the wettest period of production during the years 1990–1996 (Fig. 13.5b). Consequently, Fig. 13.7 delineates the two major vegetation dynamic regimes: 1972–1981 and 1982–1997, of the vegetation response (SAVI) for the Combat Center.

Conclusions

A number of researchers have argued that either the dampening or increase of variance as a threshold approach may provide an early warning indicator or harbingers of change (Scheffer et al. 2009; Briske et al. 2010). However, though this behavior

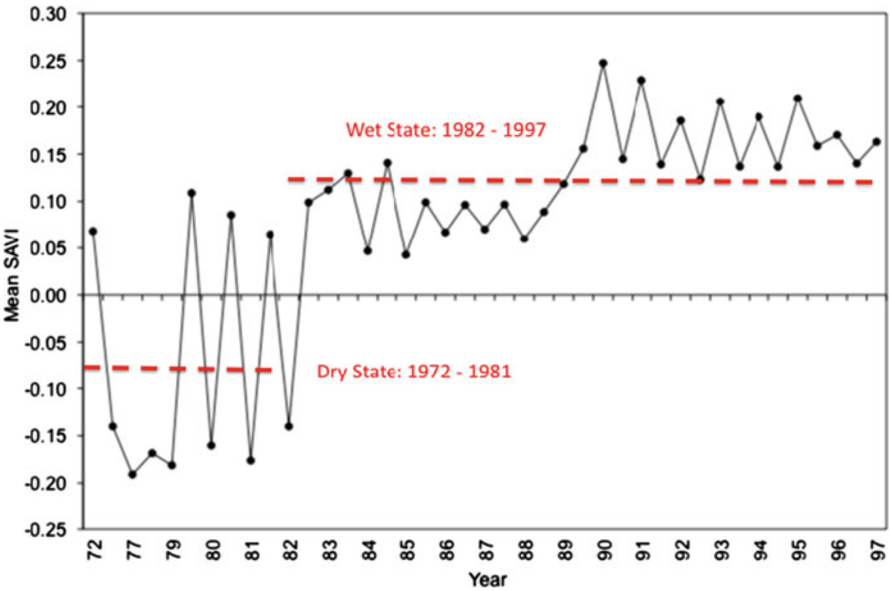


Fig. 13.7 The dry (1972–1981) and wet (1982–1997) dynamic regimes of the Marine Corps Air Ground Combat Center (MCAGCC) from 1972 to 1997

was observed for the SAVI variance, it was not significant in the overall time series nor for the dry season SAVI variance time series (Fig. 13.6b). We did find, however, that the ACF behavior of the SAVI mean dampened at a threshold, i.e., the ACF decreased from 1972 to 1981 and then changed sign from (+) to (−) ACF, and then increased incrementally after the threshold was crossed, peaking at lag 26 (1990) (Fig. 13.6a). In fact, the same dampening phenomenon is evident from 1990 in the entire time series and at lag 14 (1990) to 1997 of the dry season ACF (Fig. 13.6a). This dampening behavior suggests that a regime shift from the wet period of 1976–1998 to (observed) drought conditions may occur. Albeit a postmortem, the dampening of the ACF function from 1990 was a harbinger of the 1999–2003 droughts in the Mojave (Hereford et al. 2006) and the present droughts from 1999 to 2009 in the southwestern USA (Overpeck and Udall 2010). Consequently, the dampening of the ACF after a peak may be a harbinger of a dynamic regime shift induced by ENSO and the ACF sign change from (−) to (+) ACF or vice versa is indicative of a threshold. For MCAGCC this provided an early warning for drought mitigation techniques, as they go from a wet cycle to a dry, and implementation of ecological restoration techniques, as they take advantage of the water subsidies provided by an El Niño wet period (Holmgren and Scheffer 2001; Holmgren et al. 2001).

Acknowledgments We thank Brian Beach, Eli Rodemaker, Karin Callahan, and Todd Sajwaj for their work on data processing standardization and automation. Research was funded in part by a NASA Earth Science Enterprise Grant NNS04AB23 to R.D.R. and the Strategic Environmental

Research and Development Program (SERDP) under contract DE-AC05-00OR22725 with Oak Ridge National Laboratory, managed by UT-Battelle, LLC to R.A.W. SERDP is funded by the U.S. Department of Defense in partnership with DOE and EPA.

References

Allen, C. R., L. Gunderson, and A. R. Johnson. 2005. The use of discontinuities and functional groups to assess relative resilience in complex systems. *Ecosystems* 8 (8): 958–966.

Archer, S. 1989. Have southern Texas savannas been converted to woodlands in recent history? *American Naturalist* 134:545–561.

Beatley, J. C. 1980. Fluctuations and stability in climax shrub and woodland vegetation on the Mojave, Great Basin, and transition deserts of southern Nevada. *Israel Journal of Botany* 28:149–168.

Beisner, B. E., D. T. Haydon, and K. Cuddington. 2003. Alternative stable states in ecology. *Frontiers in Ecology and the Environment* 1:376–382

Benton, N., J. D. Ripley, and F. Powledge. 2008. Conserving biodiversity on military lands: A guide for natural resources managers. 2008 edition, NatureServe Arlington, Virginia. http://www.dodbiodiversity.org.

Betelmeyer, B. T., J. E. Herrick, J. R. Brown, D. A. Trujillo, and K. M. Havstad. 2004. Land management in the American southwest: A state-and-transition approach to ecosystem complexity. *Environmental Management* 34 (1): 38–51.

Biggs, R., S. R. Carpenter, and W. A. Brock. 2009. Turning back from the brink: Detecting an impending regime shift in time to avert it. *Proceedings of the National Academy of Science USA* 106 (3): 826–831.

Breshears, D. D., N. S. Cobb, P. M. Rich, K. P. Price, C. D. Allen, R. G. Balice, W. H. Romme, J. H. Kastens, M. Lisa-Floyd, J. Belnap, J. J. Anderson, O. B. Myers, and C. W. Meyer. 2005. Regional vegetation die-off in response to global-change-type drought. *Proceedings of the National Academy of Science* 102 (42): 15144–15148

Briske, D., R. A. Washington-Allen, C. Johnson, T. Stringham, D. Lockwood, J. Lockwood and H. H. Shugart. 2010. Catastrophic thresholds, perspectives, definitions, and applications: A synthesis. *Ecology and Society* 15:37. http://www.ecologyandsociety.org/vol15/iss3/art37/.

Brock, W. A., and S. R. Carpenter. 2006. Variance as a leading indicator of regime shift in ecosystem services. *Ecology and Society* 11 (2): 9. http://www.ecologyandsociety.org/vol11/iss2/art9/.

Dyksterhuis, E. J. 1949. Condition and management of rangeland based on quantitative ecology. *Journal of Range Management* 2:104–115.

Friedel, M. H. 1991. Range condition assessment and the concept of thresholds: A viewpoint. *Journal of Range Management* 44:422–426.

Glantz, M. H. 2001. *Currents of change: El Niño and La Niña impacts on climate and society*. 2nd ed. Cambridge: Cambridge Univ. Press.

Hereford, R., R. H. Webb, and C. I. Longpre. 2004. Precipitation history of the Mojave Desert Region, 1893–2001. USGS Fact Sheet 117–03, http://geopubs.wr.usgs.gov/fact-sheet/fs117–03/.

Hereford, R., R. H. Webb, and C. I. Longpre. 2006. Precipitation history and ecosystem response to multidecadal precipitation variability in the Mojave Desert and vicinity, 1893–2001. *Journal of Arid Environments* 67:13–34.

Herrick, J. E., V. C. Lessard, K. E Spaeth, P. L Shaver, R. S. Dayton, D. A. Pyke, L. Jolley, and J. J. Goebel. 2010. National ecosystem assessments supported by scientific and local knowledge. *Frontiers in Ecology and the Environment* 8:403–408. doi:10.1890/100017.

Holmgren, M. and M. Scheffer. 2001. El Niño as a window of opportunity for the restoration of degraded arid ecosystems. *Ecosystems* 4:151–159.

Holmgren, M., M. Scheffer, E. Ezcurra, J. R. Gutiérrez, and G. M. J. Mohren. 2001. El Niño effects on the dynamics of terrestrial ecosystems. *Trends in Ecology and Evolution* 16:89–94.

Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* 25:295–309.

Jensen, J. R. 2005. *Introductory digital image processing: A remote sensing perspective*. 3rd ed, 316. Upper Saddle River, NJ: Prentice Hall.

Lockwood, J. A., and D. R. Lockwood. 1993. A unified paradigm of rangeland ecosystem dynamics through the application of catastrophe theory. *Journal of Range Management* 46:282–288.

Ludwig, J. A., J. A. Wiens, and D. J. Tongway. 2000. A scaling rule for landscape patches and how it applies to conserving soil resources in savannas. *Ecosystems* 3:84–97.

Markham, B. L. and J. L. Barker. 1986. Landsat MSS and TM-post calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. *Earth Observation Satellite Co. (EOSAT) Landsat Technical Notes* 1:3–8.

Mayer, A. L., and M. Rietkerk. 2004. The dynamic regime concept for ecosystem management and restoration. *BioScience* 54:1013–1020.

MEA (Millennium Ecosystem Assessment). 2005. *Ecosystems and Human Well-being: Desertification Synthesis*. Washington: World Resources Institute.

Morse, D. R., J. N. Perry, and R. H. Smith. 2000. A glossary of terms used in nonlinear dynamics. In *Chaos in Real Data: The Analysis of Non-linear Dynamics from Short Ecological Time Series*, eds J. N. Perry, R. H. Smith, I. P. Woiwod, and D. R. Morse, 191–218. Dordrecht: Kluwer Academic Publishers.

Natural Resources and Environmental Division (NRED). 1999. Biological Assessment: Effects of Training and Land Use at Marine Corps Air Ground Combat Center, Twentynine Palms, on the Desert Tortoise (Gopherus agassizii). Marine Corps Air Ground Combat Center, Twentynine Palms, California.

Overpeck, J., and B. Udall. 2010. Dry times ahead. *Science* 328:1642–1643.

Peters, D. P. C., R. A. Pielke, Sr., B. T. Bestelmeyer, C. D. Allen, S. Munson-McGee, and K. M. Havstad. 2004. Cross-scale interactions, nonlinearities, and forecasting catastrophic events. *Proceedings National Academy of Science* 101:15130–15135.

Pickup, G., and B. D. Foran. 1987. The use of spectral and spatial variability to monitor cover change on inert landscapes. *Remote Sensing of Environment* 23:351–363.

Quattrochi, D. A., and R. E. Pelletier. 1991. Remote sensing for analysis of landscapes: An introduction. In *Quantitative Methods in Landscape Ecology*, ed M. G. Turner and R. H. Gardner, 51–76. New York: Springer-Verlag.

Reynolds J. F., D. M. Stafford Smith, E. F. Lambin, et al. 2007. Global desertification: Building a science for dryland development. *Science* 316:847–51.

Riebsame, W. E., S. A. Changnon, and T. R. Karl. 1991. *Drought and natural resources management in the United States: impacts and implications of the 1987–89 drought*. Boulder: Westview Press.

Rietkerk, M., S. C. Dekker, P. C. Rutler, and J. van de Koppel. 2004. Self-organized patchiness and catastrophic shifts in ecosystems. *Science* 305: 1926–1929

Rouse, J. W., Jr., R. H. Haas, J. A. Schell, and D. W. Deering. 1973. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Progress Report RSC 1978–1, Remote Sensing Center, Texas A & M University, College Station, TX 93p. (NTIS No. E73–106393).

Rundel, P. W. and A. C. Gibson. 1996. *Ecological communities and processes in a Mojave Desert ecosystem*. Rock Valley Nevada. Cambridge: Cambridge University Press.

Scheffer, M., S. R. Carpenter, J. A. Foley, C. Folke, and B. Walker. 2001. Catastrophic shifts in ecosystems. *Nature* 413(6856):591–596.

Scheffer, M., J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. van Nes, M. Rietkerk and G. Sugihara. 2009. Early-warning signals for critical transitions. *Nature* 461:53–59.

Schlesinger, W. H., J. H. Reynolds, G. L. Cunningham, L. F. Huenneke, W. M. Jarrell, R. A. Virginia, and G. W. Whitford. 1990. Biological feedbacks in global desertification. *Science* 247(4946):1043–1048.

Sellers, P. J., 1985. Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing* 6:1335–1372.

Stringham, T. K., W. C. Krueger, and P. L. Shaver. 2003. State and transition modeling: An ecological process approach. *Journal of Range Management* 56:106–113

Trenberth, K. E., G. Branstator and P. A. Arkin. 1988. Origins of the 1988 North American drought. *Science* 242:1640–1645.

Tueller, P. T. 1989. Remote sensing technology for rangeland management. *Journal of Range Management* 42:442–453.

Turchin, P. and S. P. Ellner. 2000. Modelling time-series data. In *Chaos in real data*, eds J. N. Perry, R. H. Smith, I. P. Woiwod, and D. R. Morse, 33–48. Dordrecht: Kluwer Academic Publishers.

Wardwell, D. A., and C. R. Allen. 2009. Variability in population abundance is associated with thresholds between scaling regimes. *Ecology and Society* 14 (2): 42. http://www.ecologyandsociety.org/vol14/iss2/art42/.

Washington-Allen, R. A., D. Briske, H. H. Shugart, and L. F. Salo. 2010. Catastrophic thresholds, perspectives, definitions, and applications: An introduction. *Ecology and Society* 15: 38. http://www.ecologyandsociety.org/vol15/iss3/art38/.

Washington-Allen, R. A., R. D. Ramsey, and N. E. West. 2004a. Spatiotemporal mapping of the dry season vegetation response of sagebrush steppe. *Community Ecology* 5:69–79.

Washington-Allen, R. A., T. G. Van Niel, R. D. Ramsey, and N. E. West. 2004b. Assessment of localized impacts of grazing using piospheric analysis. *GI Science and Remote Sensing* 41:95–113.

Washington-Allen, R. A., R. D. Ramsey, N. E. West, and R. A. Efroymson. 2006. A remote sensing-based protocol for assessing rangeland condition and trend. *Rangeland Ecology and Management* 59:19–29.

Washington-Allen, R. A., R. D. Ramsey, N. E. West, and B. E. Norton. 2008. Quantification of the ecological resilience of drylands using digital remote sensing. *Ecology and Society* 13:33. http://www.ecologyandsociety.org/vol13/iss1/art33/.

Washington-Allen, R. A., N. E. West, R. Douglas Ramsey, D. K. Phillips and H. H. Shugart. 2009. Retrospective assessment of soil stability on a landscape subject to commercial grazing. *Environmental Monitoring and Assessment*: doi:10.1007/s10661–008-0661–3.

Webb, R. H., M. B. Murov, T. C. Esque, D. E. Boyer, L. A. DeFalco, D. F. Haines, D. Oldershaw, S. J. Scoles, K. A. Thomas, J. B. Blainey, and P. A. Medica. 2003. Perennial vegetation data from permanent plots on the Nevada Test Site, Nye County, Nevada, US Geological Survey Open File Report 03–336.

West, N. E. 2003. Theoretical underpinnings of rangeland monitoring. *Arid Land Research and Management* 17:333–346

Westoby, M., B. H. Walker, and I. Noy-Meir. 1989. Opportunistic management for rangelands not at equilibrium. *Journal of Range Management* 42:266–274.

Yafee, R. A., and M. McGhee. 2000. *Introduction to time series analysis and forecasting with applications of SAS and SPSS*. San Diego: Academic Press.

Chapter 14
Ecological Thresholds for Salt Marsh Nekton and Vegetation Communities

Mary-Jane James-Pirri, Jeffrey L. Swanson, Charles T. Roman,
Howard S. Ginsberg and James F. Heltshe

Abstract Salt marsh vegetation and nekton respond to stressors in estuarine systems, providing ideal indicators of change. Here we characterize the structure of these communities along a gradient of anthropogenic stress (e.g., human population size of watersheds, degree of hydrological alteration of marshes), identify ecological thresholds that are linked to community condition, and provide potential decision thresholds for land managers based on an evaluation of community condition. Salt marsh nekton and vegetation community data (species composition and abundance) were compiled from over 180 discrete data sets from marshes along the Atlantic coast from Maine to Virginia. Using multivariate techniques (e.g., Principal Component Analysis, Canonical Correspondence Analysis, Analysis of Similarities), patterns of community change along a gradient of hydrologic impact and degree of watershed development were elucidated. Several levels of community complexity, individual species' abundances, and relative abundances of life history-based groups were used to identify potential metrics for ecological thresholds. The nekton community displayed shifts in community structure along a gradient of human population size (e.g., anthropogenic stress gradient) in surrounding watersheds, from resident fish-dominated communities at marshes in watersheds with low human populations to shrimp-dominated communities (Palaemonidae species) in watersheds with high human populations. Vegetation communities from reference (relatively hydrologically undisturbed) marshes were dominated by obligate halophytes (e.g., salt meadow

---

M.-J. James-Pirri (✉)
Graduate School of Oceanography, University of Rhode Island,
Narragansett, RI 02882, USA
e-mail: mjjamespirri@mail.uri.edu

J. L. Swanson · H. S. Ginsberg
Department of Computer Science and Statistics, University of Rhode Island,
Kingston, RI 02881, USA

C. T. Roman
National Park Service, Graduate School of Oceanography,
University of Rhode Island, Narragansett, RI 02882, USA

J. F. Heltshe
USGS Patuxent Wildlife Research Center, Coastal Field Station, Woodward Hall-PLS,
University of Rhode Island, Kingston, RI 02881, USA

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0_14,

grasses) with low proportions of invasive plants. Impacted (e.g., tidally restricted) marshes had fewer halophytic obligate wetland plants, more facultative wetland plant species, and higher proportions of invasive species. Shifts in characteristics of nekton and vegetation communities toward the impacted state can be used as ecological thresholds upon which decision thresholds for land managers can be based. For example, detection of declining proportions of killifish and transient fish and increasing proportions of Palaemonidae shrimp (the ecological threshold metrics), would indicate a community changing from moderate to poor condition (the decision threshold), thereby triggering management actions. Ecological thresholds are presented for both nekton and vegetation communities and are presented in the context of a desired resource goal (good, moderate, poor condition) that can be used by resource managers to evaluate responses to restoration activities or the overall condition of the marsh community. Findings from this study are most directly relevant to northeastern US salt marshes, but the methods to derive the ecological threshold metrics can be applied to other regions.

Keywords Grass shrimp · Palaemonidae shrimp · Killifish · Fundulidae species · Salt marsh community condition · Salt marsh nekton · Salt marsh vegetation · Anthropogenic stress

Introduction

Salt marsh vegetation and nekton species composition and abundance clearly respond to stressors in estuarine systems, thus providing ideal indicators of condition. For example, vegetation changes have been documented in response to salt marsh hydrologic alterations that are related to ditching activity (e.g., Bourn and Cottam 1950; Niering and Warren 1980), restriction of tidal action (e.g., Roman et al. 1984, 1995), and submergence related to sea-level rise (e.g., Donnelly and Bertness 2001; Hartig et al. 2002). With nutrient enrichment of estuaries, changes in vegetation species composition and primary productivity have been noted (e.g., Nixon and Oviatt 1973; Wigand et al. 2001). Characteristics of nekton (free-swimming fishes and crustaceans), such as fish abundance, species richness, and growth rates of the common mummichog (*Fundulus heteroclitus*), have been shown to increase with nutrient loading (LaBrecque et al. 1996; Tober et al. 1996) and respond quite rapidly to hydrologic alterations (e.g., Able et al. 2000; Roman et al. 2002). Further, nekton represent a valuable monitoring indicator because of their essential role as forage for piscivorous birds, economically valuable fishes, and marine mammals (e.g., Friedland et al. 1988; Sekiguchi 1995; Smith 1997).

The purpose of this study was to identify ecological thresholds that resource managers can use to evaluate salt marsh condition. A quantitative knowledge of condition can inform resource management decisions. We define ecological thresholds after Martin et al. (2009) as changes in metric values (in this case nekton and vegetation community composition) that produce measureable shifts in community structure; and decision thresholds as changes in metric values that should prompt specific management actions. We analyzed extensive nekton and vegetation data sets from marshes throughout the Maine to Virginia region, including marshes that varied along a gradient from relatively undisturbed to highly impacted based on watershed land use, watershed human population, and hydrologic alteration.

Methods

General Data Description

Existing data collected during monitoring efforts of salt marsh nekton and vegetation communities at numerous sites were used for this project (Swanson 2009). The majority of data were collected between 2001 and 2006 using protocols developed by the National Park Service for their Inventory and Monitoring Program (Roman et al. 2001; Raposa and Roman 2000; James-Pirri et al. 2008; James-Pirri et al. 2011). Both nekton and vegetation data were multivariate data of community composition (species) and abundance (density for nekton or percent cover for vegetation). Sites and data spanned a wide latitudinal range (Fig. 14.1; Maine to Virginia) and included varying degrees of disturbance as defined by watershed development (rural, agricultural, suburban, and urban), human population size in the watershed, and hydrologic impacts (e.g., tidal restrictions and hydrological alterations for mosquito control). Sites were characterized as either reference or impacted sites. Reference sites were grid ditched or occasionally unditched, but contained no other hydrological alterations. Impacted sites were also grid ditched and had other hydrological alterations including tidal restrictions, open marsh water management for mosquito control, or were recently tidally restored. The majority of sites had data from multiple years and some had data from both before and after recent human-made alterations or restoration activities. The term "site" is used herein to identify data from a specific marsh-year combination (e.g, data from Sachuest reference marsh sampled in 2000, or Sachuest reference marsh sampled in 2003).

Nekton and Vegetation Collection Methods and Data Description

Nekton data were collected using two habitat-dependent enclosure gear types: a throw trap in salt marsh pools and tidal creeks (Rozas and Minello 1997; Raposa et al. 2003) and a ditch net in salt marsh grid ditches or small tidal creeks (James-Pirri et al. 2010, James-Pirri et al. 2011). All sampling was conducted after the marsh surface drained of water. Sampling stations were randomly located within each habitat. All sampling was in summer (June–September), usually with 10–20 stations sampled

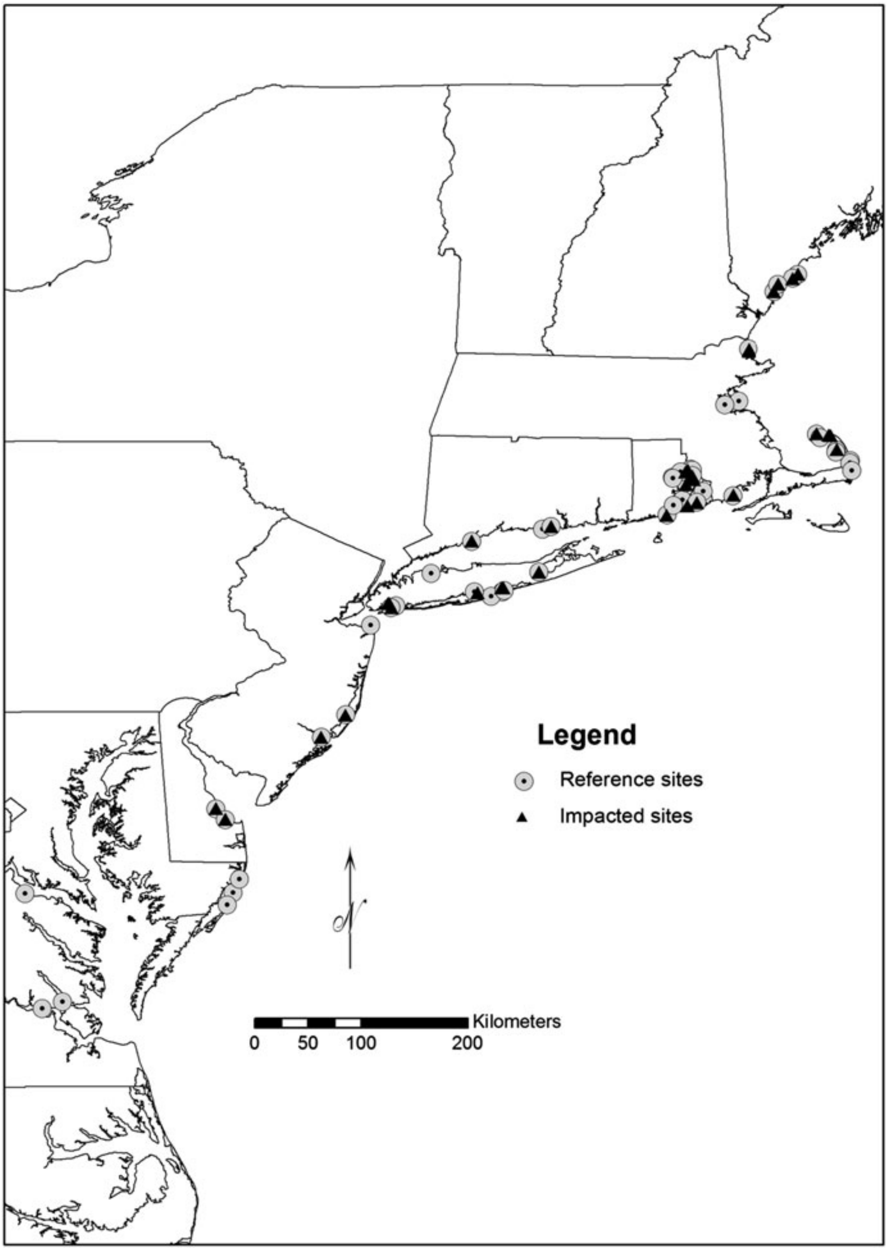


Fig. 14.1 Location of salt marsh sites where data were collected. Impacted sites include hydrologically altered, tide restricted, and recently tidally restored sites

with each gear type during each summer sampling period, resulting in over 7,850 individual samples in the database. Nekton were identified to the level of species, if possible. Average density (number individuals m^{-2}) was calculated for each species at each site and the data were standardized as relative species abundance, calculated as the proportion of total nekton density for each site. Nekton data consisted of 81 individual salt marshes (43 reference and 38 hydrologically altered or impacted marshes) resulting in a total of 185 sites (96 reference and 89 altered or impacted sites), since several marshes were sampled over multiple years. A total of 50 species (35 fish species and 15 crustacean species), with an additional 12 categories that were only identified to family, were present in the data (Table 14.1).

At each site salt marsh vegetation was sampled in a minimum of 20 randomly located 1 m^2 vegetation plots oriented along transects that traversed the elevation gradient (Elzinga et al. 2001; Roman et al. 2001; James-Pirri et al. 2007; James-Pirri et al. 2011). Vegetation was sampled annually, from late August to early September, using the point intercept method (50 point grid) to estimate percent cover for each cover type (Kent and Coker 1992; Elzinga et al. 2001). Cover type categories included all live vegetation (including vascular plants and algae), open water, and wrack/litter. Vegetation data were from 70 individual salt marshes (40 reference and 30 altered or impacted marshes), resulting in a total of 156 sites (80 reference and 76 altered or impacted sites). There were 181 cover types represented (179 plant or macro-algae species and two non-vegetative covers: standing water and litter/wrack) (Table 14.2). Percent cover for each cover type was estimated for each vegetation plot (4,220 vegetation plots) and average percent cover was estimated for each site.

Watershed Landuse and Population Estimation

The marsh sites were located in 48 individual subwatersheds. Since there was no national GIS watershed data set, subwatersheds were delineated from individual state GIS coverages using the smallest available watershed. When state subwatershed data were not available (e.g., Long Island, New York, and Massachusetts) subwatersheds were delineated using Digital Elevation Model data in ArcGIS 9.1 software (ESRI 2005).

Landuse/land cover data were estimated for each subwatershed from the 2001 National Land Cover Data Set using ArcGIS 9.1. To standardize landuse among the subwatersheds, watershed landuse data were converted to proportional data or a percent that each land use type occupied within the subwatershed. In addition, a smaller area (5 km radius from the middle of each site) around each marsh was delineated in ArcGIS 9.1 and estimates of landuse in this smaller area were calculated.

Population data for all subwatersheds were derived from the 2000 U.S. Census, mapped by census block. The human population of an individual census block was used only if more than 50 % of the block was within the watershed boundary.

Table 14.1 Distribution of nekton species data and life history groups used in analyses

Grouping and Latin name	Common name	Presence (percent of sites) (%)	Relative abundance (percent of total nekton density) (%)
Resident fish			
Cyprinodon variegatus	Sheepshead minnow	44	3
Dormitator maculatus[a]	Fat sleeper	1	< 1
Fundulus species	Killifish species	5	< 1
Fundulus diaphanus	Banded killifish	4	< 1
Fundulus heteroclitus	Mummichog	100	30
Fundulus luciae	Spotfin killifish	12	< 1
Fundulus majalis	Striped killifish	43	1
Gambusia holbrooki	Eastern mosquitofish	2	< 1
Gambusia species	Mosquitofish species	5	< 1
Lucania parva	Rainwater killifish	32	1
Resident shrimp			
Palaemonetes pugio	Daggerblade grass shrimp	49	22
Palaemonetes species	Grass shrimp species	39	35
Palaemonetes vulgaris	Marsh grass shrimp	1	1
Resident crustacean			
Carcinus maenas[b]	Green crab	56	1
Pagurus longicarpus	Longwrist hermit crab	4	< 1
Pagurus species	Hermit crab species	6	1
Uca pugnax	Atlantic marsh fiddler crab	7	< 1
Uca species	Fiddler crab species	6	< 1
Transient fish			
Alosa aestivalis[a]	Blueback herring	1	< 1
Alosa pseudoharengus	Alewife	2	< 1
Alosa sapidissima	American shad	1	< 1
Ammodytes americanus[a]	American sand lance	1	< 1
Anchoa mitchilli[a]	Bay anchovy	1	< 1
Anguilla rostrata	American eel	44	< 1
Apeltes quadracus	Fourspine stickleback	29	< 1
Brevoortia tyrannus	Atlantic menhaden	9	1
Centropristis striata	Black sea bass	1	< 1
Clupea harengus	Atlantic herring	2	< 1
Gasterosteus aculeatus	Threespine stickleback	15	< 1
Gobiosoma bosc	Naked goby	1	< 1
Gobiosoma ginsburgi	Seaboard goby	2	< 1
Goby species	Goby species	1	< 1
Lepomis species[a]	Sunfish species	1	< 1
Menidia beryillina	Inland silverside	30	< 1
Menidia menidia	Atlantic silverside	62	2
Menidia species	Silverside species	8	< 1
Morone americana	White perch	4	< 1
Mugil cephalus	Striped mullet	3	< 1
Mugil curema	White mullet	3	< 1
Notropis species	Eastern shiner species	2	< 1
Opsanus tau	Oyster toadfish	1	< 1

Table 14.1 (continued)

Grouping and Latin name	Common name	Presence (percent of sites) (%)	Relative abundance (percent of total nekton density) (%)
Paralichthys dentatus[a]	Summer flounder	1	< 1
Pomatomus saltatrix[a]	Bluefish	1	< 1
Pseudopleuronectes americanus	Winter flounder	9	< 1
Pungitius pungitius	Ninespine stickleback	29	< 1
Syngnathus fuscus	Northern piepfish	12	< 1
Tautoga onitis	Tautog	1	< 1
Tautogolabrus adspersus	Cunner	1	< 1
Trinectes maculatus	Hogchoker	1	< 1
Transient Crustacean			
Callinectes sapidus	Blue crab	35	< 1
Cancer irroratus	Rock crab	1	< 1
Crangon septemspinosa	Sevenspine bay shrimp	37	1
Dyspanopeus sayi	Say mud crab	1	< 1
Hemigrapsus sanguineus[b]	Asian shore crab	4	< 1
Libinia dubia[a]	Longnose spider crab	1	< 1
Libinia species[a]	Spider crab species	1	< 1
Limulus polyphemus[c]	American horseshoe crab	8	< 1
Mud crab species	Mud crab	6	< 1
Neopanope species[a]	Grassflat crab spp	1	< 1
Ovalipes ocellatus	Lady crab	3	< 1
Panopeus herbstii	Atlantic mud crab	1	< 1
Rhithropanopeus harrisii	Harris mud Crab	2	< 1

[a] Indicates species was found at only one site (a site was defined as a specific salt marsh and year of data)

[b] Indicates exotic species. Species that were present at 10 % or more of the sites were considered common species in preliminary analyses

[c] The American horseshoe crab was grouped with the transient crustaceans

Analyses

Several multivariate statistical analyses (e.g., Principal Component Analyses, Canonical Correspondence Analysis, Analysis of Similarities (ANOSIM), Similarity Percentages (SIMPER) (Clarke and Warwick 2001; Clarke and Gorley 2006) were performed to determine relationships among species composition and abundance, degree of alteration (reference sites versus impacted sites), landuse pattern, and human population density.

For nekton, the analyses included several iterations using all species, common species (those present at $> 10\,\%$ of the marshes), and groups based on life history association with the salt marsh environment (Table 14.1). The life history groups (after Bigelow and Schroeder 1953; Ayvazian et al. 1992; Able et al. 1996; Deegan et al. 1997; Collette and Klein-MacPhee 2002; Elliot et al. 2007) were salt marsh resident fish (e.g., killifish, Fundulidae species), resident shrimp (e.g., marsh

Table 14.2 Wetland status and salinity tolerance for vegetation cover types. Wetland status and salinity tolerance was based on data from USDA Plants database (USDA 2009)

Latin name	Common name	Presence (percent of sites) (%)	Relative cover (percent of total cover) (%)
Wetland, high salinity tolerance			
Ascophyllum species	Brown algae species	3	< 1
Atriplex patula	Spear saltbush	32	< 1
Chaetomorpha linum	Green algae species	1	< 1
Cladophora species	Green algae species	3	< 1
Distichlis spicata	Saltgrass	91	11
Fucus and Ascophyllum	Brown algae species	1	< 1
Fucus vesiculosus	Bladderwrack	4	< 1
Iva frutescens ME[a]	Jesuit's bark	33	1
Juncus balticus MD[a]	Baltic rush	8	< 1
Juncus gerardii	Saltmeadow rush	41	5
Limonium carolinianum[b] NY[a]	Carolina sea lavender	22	< 1
Limonium nashi[b] NY[a]	Sea lavender	21	< 1
Limonium species[b]	Sea lavender species	2	< 1
Macro or filamentous green algae	Macroalgae	13	< 1
Plantago maritima	Goose tongue	36	1
Puccinellia distans and *P. maritima*	Weeping/seaside alkaligrass	26	1
Ruppia maritima	Widgeongrass	3	< 1
Salicornia bigelovii[b]	Dwarf saltwort	2	< 1
Salicornia maritima[b]	Slender glasswort	51	1
Salicornia species[b]	Glasswort species	29	1
Salicornia virginica[b]	Virginia glasswort	6	1
Schoenoplectus maritimus CT[a], NJ, NY, RI	Cosmopolitan bulrush	3	< 1
Schoenoplectus robustus	Sturdy bulrush	6	< 1
Solidago sempervirens	Seaside golden rod	39	< 1
Spartina alterniflora	Smooth cordgrass	97	25
Spartina patens	Saltmeadow cordgrass	97	24
Suaeda linearis[b]	Annual seepweed	3	< 1
Suaeda maritima[a]	Herbaceous seepweed	22	< 1
Symphyotrichum tenuifolium[b]	Perennial saltmarsh aster	11	< 1
Wetland, Medium Salinity Tolerance			
Agalinis maritima ME[a]	Saltmarsh false foxglove	23	< 1
Agrostis stolonifera	Creeping bentgrass	10	< 1
Amaranthus cannabinus	Tidalmarsh amaranth	4	< 1
Baccharis halimifolia	Eastern baccharis	4	< 1
Eleocharis parvula	Dwarf spikerush	2	< 1
Eleocharis rostellata ME[a], *RI*	Beaked spikerush	1	< 1
Glaux maritima MD[a], NJ, RI	Sea milkwort	26	2 %
Hierochloe odorata MD[a]	Sweetgrass	1	< 1
Iris versicolor	Harlequin blueflag	1	< 1

Table 14.2 (continued)

Latin name	Common name	Presence (percent of sites) (%)	Relative cover (percent of total cover) (%)
Juncus arcticus MD[b]	Arctic rush	1	< 1
Juncus canadensis	Canadian rush	2	< 1
Juncus effusus var. conglomeratus	Common rush	6	< 1
Kosteletzkya virginica	Virginia saltmarsh mallow	2	< 1
Lythrum lineare	Wand lythrum	1	< 1
Lythrum salicaria[c]	Purple loosestrife	3	< 1
Polygonum sagittatum	Arrowleaf tearthumb	2	< 1
Potentilla anserina anserina	Silverweed cinquefoil	5	< 1
Schoenoplectus americanus	American or Olney bulrush	18	< 1
Schoenoplectus pungens	Common threesquare	4	< 1
Spergularia salina	Salt sandspurry	1	< 1
Symphyotrichum novi-belgii[b]	New York aster	3	< 1
Symphyotrichum subulatum ME[a], NY	Eastern annual saltmarsh aster	3	< 1
Triglochin maritimum[b] NJ[a]	Seaside arrowgrass	31	1
Typha angustifolia	Narrowleaf cattail	12	< 1
Wetland, Low Salinity Tolerance			
Argentina anserina	Silverweed cinquefoil	8	< 1
Carex comosa	Longhair sedge	1	< 1
Carex scoparia	Broom sedge	3	< 1
Carex species	Sedge species	4	< 1
Cyperus esculentus	Yellow nutsedge	2	< 1
Eleocharis quadrangulata CT[a], NY	Squarestem spikerush	1	< 1
Hibiscus moscheutos	Crimsoneyed rosemallow	1	< 1
Lycopus americanus	American water horehound	4	< 1
Panicum rigidulum var. pubescens	Redtop panicgrass	1	< 1
Photinia pyrifolia	Red chokeberry	1	< 1
Phragmites australis[b,c]	Common reed	29	3
Pluchea odorata var. ordorata	Sweetscent	17	< 1
Poa palustris	Fowl bluegrass	1	< 1
Scirpus cyperinus	Woolgrass	3	< 1
Typha latifolia	Broadleaf cattail	1	< 1
Verbena hastata	Swamp verbena	1	< 1
Wetland, No Salinity Tolerance			
Boehmeria cylindrica	Smallspike false nettle	1	< 1
Carex straminea NY[a]	Eastern straw sedge	2	< 1
Cuscuta gronovii[c]	Scaldweed	1	< 1
Cyperus filicinus	Fern flatsedge	2	< 1
Cyperus strigosus	Strawcolored flatsedge	1	< 1
Dichanthelium scabriusculum CT[a], MA, NY, MD	Wooly rosette grass	3	< 1

Table 14.2 (continued)

Latin name	Common name	Presence (percent of sites) (%)	Relative cover (percent of total cover) (%)
Doellingeria umbellata var. umbellata	Parasol whitetop	1	< 1
Epilobium coloratum	Purpleleaf willowherb	1	< 1
Epilobium leptophyllum	Bog willowherb	3	< 1
Eupatorium dubium ME[a]	Coastal plain joe pye weed	3	< 1
Galium palustre	Common marsh bedstraw	1	< 1
Galium tinctorium	Stiff marsh bedstraw	3	< 1
Galium trifidum	Threepetal bedstraw	1	< 1
Glyceria canadensis	Rattlesnake mannagrass	1	< 1
Ilex verticillata	Common winterberry	3	< 1
Impatiens capensis NY[a]	Jewelweed	1	< 1
Lycopus uniflorus	Northern bugleweed	1	< 1
Mentha arvensis	Wild mint	1	< 1
Myrica gale	Sweetgale	2	< 1
Onoclea sensibilis	Sensitive fern	3	< 1
Petasites hybridus	Pestilence wort	1	< 1
Polygonum hydropiper	Marshpepper knotweed	1	< 1
Ribes lacustre CT[a], MA	Prickly currant	1	< 1
Rosa palustris	Swamp rose	1	< 1
Salix bebbiana MD[a]	Bebb willow	1	< 1
Sparganium eurycarpum	Broadfruit bur-reed	1	< 1
Spartina pectinata	Prairie cordgrass	7	< 1
Sphagnum species	Sphagnum species	1	< 1
Spiraea alba	White meadowsweet	4	< 1
Spiraea tomentosa	Steeplebush	4	< 1
Teucrium canadense	Canada germander	3	< 1
Triadenum virginicum	Virginia marsh St. Johnswort	4	< 1
Vaccinium corymbosum	Highbush blueberry	1	< 1
Vaccinium macrocarpon	Cranberry	3	< 1
Wetland, unknown salinity tolerance			
Oclemena nemoralis CT[a]	Bog aster	1	< 1
Platanthera clavellata NY[a]	Small green wood orchid	1	< 1
Polygonum arifolium	Halberdleaf tearthumb	1	< 1
Ptilimnium capillaceum RI[a]	Herbwilliam	1	< 1
Schoenoplectus torreyi	Torrey's bulrush	1	< 1
Scutellaria galericulata	Marsh skullcap	3	< 1
Transitional, high salinity tolerance			
Ammophila breviligulata	American beachgrass	5	< 1
Atriplex prostrata	triangle orache	1	< 1
Atriplex species	Saltbush species	2	< 1
Plantago species	Plantain species	1	< 1
Transitional, Medium Salinity Tolerance			
Agrostis hyemalis	Winter bentgrass	1	< 1

Table 14.2 (continued)

Latin name	Common name	Presence (percent of sites) (%)	Relative cover (percent of total cover) (%)
Calystegia sepium	Hedge false bindweed	8	< 1
Lathyrus japonicus	Beach pea	1	< 1
Morella pensylvanica	Northern bayberry	7	< 1
Panicum virgatum	Switchgrass	22	< 1
Rosa rugosa	Rugosa rose	3	< 1
Toxicodendron radicans	Eastern poison ivy	7	< 1
Transitional, low salinity tolerance			
Agrostis scabra	Rough bentgrass	1	< 1
Ambrosia trifida	Great ragweed	1	< 1
Amphicarpaea bracteata	American hogpeanut	1	< 1
Artemisia stelleriana	Oldwoman	1	< 1
Chenopodium album	Lambsquarters	1	< 1
Festuca rubra	Red fescue	21	< 1
Juncus tenuis	Poverty rush	1	< 1
Lonicera japonica[c]	Japanese honeysuckle	1	< 1
Polygonum ramosissimum NY[b]	Bushy knotweed	1	< 1
Thelypteris palustris NY[a]	Eastern marsh fern	4	< 1
Viburnum dentatum	Southern Arrowwood	1	< 1
Transitional, No Salinity Tolerance			
Acer rubrum	Red maple	2	< 1
Achillea millefolium	Common yarrow	3	< 1
Asclepias syriaca	Common milkweed	1	< 1
Betula populifolia	Gray birch	1	< 1
Elymus repens	Quackgrass	8	< 1
Euthamia graminifolia tenuifolia	Flat-top/slender goldentop	4	< 1
Fragaria virginiana	Virginia strawberry	1	< 1
Holcus lanatus	Common velvetgrass	3	< 1
Juncus greenei	Greene’s rush	1	< 1
Parthenocissus quinquefolia	Virginia creeper	2	< 1
Phytolacca americana	American pokeweed	1	< 1
Polygonum scandens	Climbing false buckwheat	1	< 1
Populus deltoides	Eastern cottonwood	1	< 1
Prunus serotina	Black cherry	3	< 1
Ribes hirtellum	Hairystem gooseberry	2	< 1
Rosa virginiana	Virginia rose	3	< 1
Schizachyrium scoparium	Little bluestem	1	< 1
Smilax rotundifolia	Roundleaf greenbrier	1	< 1
Solanum dulcamara[c]	Climbing nightshade	1	< 1
Solidago canadensis	Canada goldenrod	2	< 1
Solidago rugosa	Wrinkleleaf goldenrod	3	< 1
Solidago simplex	Mt. Albert goldenrod	1	< 1
Upland, medium salinity tolerance			
Nuttallanthus canadensis	Canada toadflax	1	< 1
Rosa carolina	Carolina rose	3	< 1
Upland, Low Salinity Tolerance			

Table 14.2 (continued)

Latin name	Common name	Presence (percent of sites) (%)	Relative cover (percent of total cover) (%)
Polygonella articulata	Coastal jointweed	1	< 1
Prunus maritima CT[a], MD, ME	Beach plum	2	< 1
Upland, no salinity tolerance			
Artemisia vulgaris	Common wormwood	1	< 1
Celastrus orbiculatus	Oriental bittersweet	2	< 1
Cirsium arvense[c]	Canada thistle	1	< 1
Daucus carota	Queen Anne's lace	1	< 1
Deschampsia flexuosa	Wavy hairgrass	3	< 1
Eragrostis spectabilis	Purple lovegrass	1	< 1
Erechtites hieracifolia	American burnweed	3	< 1
Heliopsis helianthoides	Smooth oxeye	1	< 1
Hieracium aurantiacum	Orange hawkweed	1	< 1
Hypochaeris radicata	Hairy cat's-ear	1	< 1
Lactuca canadensis	Canada lettuce	1	< 1
Lechea intermedia	Largepod pinweed	1	< 1
Lonicera morrowii[c]	Morrow's honeysuckle	1	< 1
Rumex acetosella[c]	Common sheep sorrel	3	< 1
Solanum ptychanthum	West Indian nightshade	1	< 1
Solidago odora	Anisescented goldenrod	1	< 1
Sonchus arvensis	Field sowthistle	1	< 1
Status unknown, medium salinity tolerance			
Lechea maritima	Beach pinweed	1	< 1
Lepidium species	Pepperweed species	1	< 1
Status unknown, low salinity tolerance			
Agrostis gigantea	Redtop	1	< 1
Artemisia campestris	Field sagewort	1	< 1
Status Unknown, No Salinity Tolerance			
Rhus copallinum	Winged sumac	3	< 1
Rhus species	Sumac species	1	< 1
Spiraea species	Spirea species	1	< 1
Non-vegetative covers			
Litter/wrack	Litter/wrack	Not used	
Water	Water	Not used	

Nomenclature from USDA Plants Database

"-" indicates either wetland status or salinity tolerance or could not be definitively determined and was assigned as "unknown"

[a] Indicates state-listed species and state where listed

[b] Indicates common species

[c] Indicates invasive species

grass shrimp, Palaemonidae species), resident crustaceans other than Palaemonidae shrimp (e.g., *Carcinus maenus*, green crab), transient fish (e.g., flounders, silversides (*Menidia* species)), and transient crustaceans (e.g., *Callinectes sapidus*, blue crab) (Table 14.1). Resident shrimp and resident fish composed the majority of the data, (57 and 36 %, respectively, of the total relative abundance of the entire data set). The other three groups each represented less than 5 % (transient fish, 4 %; transient crustaceans, 2 %; resident crustaceans, 1 %). The resident categories were species that primarily rely on the salt marsh for most life history stages, while the transient categories included species that use salt marshes but were not exclusively found in salt marshes (e.g., flounders, silversides, blue crab).

Due to the large size of the original vegetation data initial analyses were conducted on the 21 common vegetation cover types (Table 14.2), defined as those that were present at 25 % or more of the sites. To incorporate more than the common vegetation cover types in subsequent analyses, all vegetative cover types were categorized according to wetland status (wetland, transitional, or upland plants) and salinity tolerance (none, low, medium, or high tolerance), based on information from the USDA Plants Database (USDA 2009). Plants were also identified as state-listed (e.g., endangered, threatened, vulnerable) or invasive/exotic if appropriate, as identified in the USDA Plants Database (Table 14.2). Cover types of individual species in each category were summed to obtain a percent cover for each category. The wetland status and salinity tolerance categories resulted in a total of 17 groups (15 wetland status and salinity tolerance groups and two groups for state-listed and invasive status), some salinity tolerance and wetland status groups that had few occurrences (< 5) were combined into similar groups for analyses.

If patterns in either nekton or vegetation communities were observed in initial exploratory analyses, these were followed with additional analyses to determine if there were any significant correlations with anthropogenic stressor indicators such as the human population size of the watershed or surrounding land use.

Results

Nekton Analyses

The nekton community was dominated by the common mummichog (*Fundulus heteroclitus*) and Palaemonidae grass shrimp (*Palaemonetes pugio*, *P. vulgaris*, and *Palaemonetes* species), together comprising 88 % of the total relative nekton abundance (Table 14.2). All other species contributed 3 % or less to the total nekton density. Mummichogs were present at all sites and Palaemonidae shrimp were present at most sites (89 % of sites) (Table 14.1). Palaemonidae shrimp were the only species that exceeded an average density of 70 individuals m^{-2} at any one site.

Initial Principal Component Analyses using all nekton species indicated that little variation was explained by the first three principal components when either all sites ($n = 185$) were included or when reference ($n = 96$) and impacted ($n = 89$) sites were analyzed separately, with only 16–21 % of the variability explained by the first three principal components. In all of these iterations, the first three principal components were generally representative of species that only appeared sporadically in the data (e.g., Harris mud crab, *Rhithropanopeus harrisii*; Eastern mosquito fish, *Gambusia holbrooki*; fiddler crab species, *Uca* species; Table 14.1), and accounted for only a small percent of the total variability in the data.

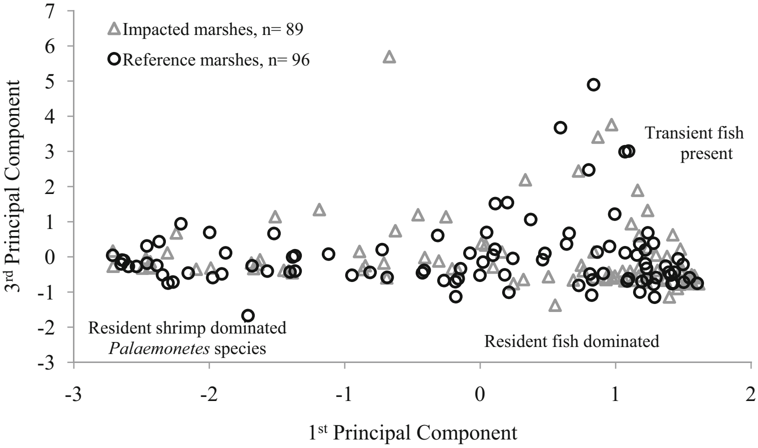


Fig. 14.2 Graph of first and third principal components scores from the analyses using nekton life history groups. Variation explained: 1st PC: 38 %; 3rd PC: 22 %

Principal Component Analyses using the 17 common species increased the amount of variability explained, with the first three principal components accounting for 34 % of the variability (all sites, $n = 185$). When analyzed separately by reference and impacted sites, there was an improvement with the first three principal components accounting for 36 and 37 % of the variability, respectively. A graphic representation of the first and second principal components (accounting for 26 % of the variability) indicated the majority of sites were closely grouped together, thus limiting the ability to distinguish differences among sites based on common nekton species.

Analysis of the five nekton life history groups resulted in a high percentage of the variability (85 %) accounted for by the first three principal components (all sites, $n = 185$). The first and third principal component separated sites according to three of the five life history groups, resident fish, resident shrimp, and transient fish (Fig. 14.2). The first principal component, accounting for 38 % of the variation, had a high negative (88 %) correlation between resident fish (positive loading) and resident shrimp (negative loading), separating sites that had high relative abundances of resident fish from those with high relative abundances of resident shrimp. The third principal component accounted for 22 % of the variability, and had high loadings for transient fish and was indicative of sites where transient fish species were present. The second principal component accounted for 26 % of the variation and had high positive loadings for both resident and transient crustaceans. Similar results were observed when just the 96 reference sites were analyzed (85 % of the variation explained by the first three principal components). A slightly higher proportion of the variation (87 %) was explained when only the 89 impacted sites were used in the analyses. Interestingly, the trends in the correlation matrix and proportion explained by the three principal components were nearly identical when reference and impacted sites were analyzed separately as when all sites were analyzed together.

Watershed Population Size and Nekton Guilds

Human population size of the watershed was used as an indicator of development. Six sites located within New York City (Jamaica Bay) were removed prior to the analyses because they were extreme outliers in regards to total watershed population size (~ 1.8 million people, the next largest watershed population was 55,000 people). There was no strong functional relationship between the dominant life history groups (first principal component (resident fish and resident shrimp)) and watershed population (all sites, Pearson Correlation Coefficient -0.03, $p = 0.67$ $n = 173$; reference sites, Pearson Correlation Coefficient 0.02, $p = 0.88$, $n = 90$; impacted sites, Pearson Correlation Coefficient -0.10, $p = 0.36$, $n = 83$). This was likely attributed to the high variability in dominant resident fish and shrimp groups at low population levels (e.g., $< 1{,}000$ people), as they spanned the entire range of the first principal component score. Based on these results sites with low populations in the surrounding watershed ($< 1{,}000$ people) were removed from further analyses to determine if a stronger correlation could be found.

The first principal component was significantly correlated with log-transformed population size (Pearson Correlation coefficient: -0.036, $p < 0.01$, $r^2 = 0.12$, impacted and reference sites together ($n = 94$)) when the low population sites ($< 1{,}000$ people) were removed, with the first three principal components accounting for 82 % of the total variation. The principal component loadings also remained similar with the first principal component distinguishing between resident fish-dominated sites (positive loading) and resident shrimp-dominated sites (negative loading). Transient fish were represented by high positive loading in the second principal component, while transient crustaceans and resident crabs were represented by positive loadings in the third principal component. When reference sites ($n = 51$) were analyzed separately, the first three principal components accounted for 84 % of the total variation (41, 23, 20 %: first, second, and third principal components, respectively) and a strong correlation was observed between the first principal component and total population (log transformed) (Pearson Correlation Coefficient: -0.64, $p < 0.01$, $r^2 = 0.40$) (Fig. 14.3). Similar to the reference sites the three principal components for the impacted sites ($n = 43$) accounted for 85 % of the variation in the nekton groups and had similar percentages accounted for by the first three first principal components, but the correlation between the first principal component and population size was not significant (Pearson Correlation Coefficient: 0.03, $p = 0.84$).

Plotting the first and third principal components for the reference sites categorized by population size showed that in high population watersheds (20,000–100,000 people) the nekton communities were always dominated by the resident Palaemonidae shrimp (Fig. 14.4). As watershed population size decreased, the nekton community became dominated by resident fish (e.g., Fundulidae species), and at the lowest watershed populations the presence of transient fish species increased.

Additional iterations using total species richness, fish species richness, and exotic species richness yielded no useful results for distinguishing among sites with varying types of hydrological impacts or with watershed population size.

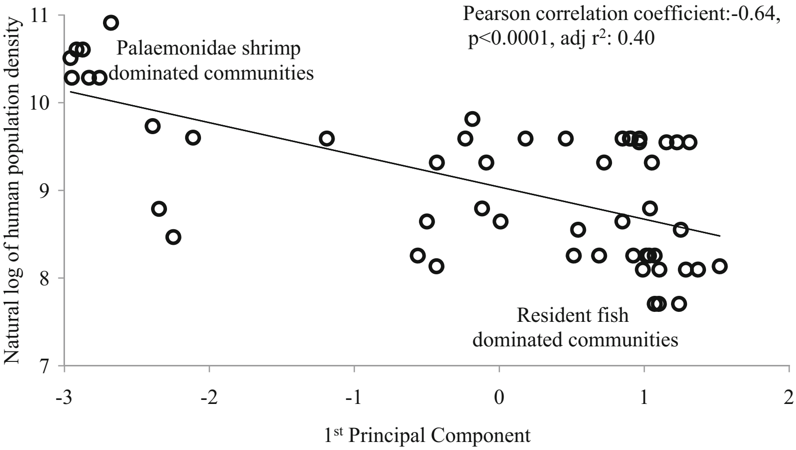


Fig. 14.3 First principal component scores versus watershed human population size for reference sites ($n = 51$) with populations > 1,000 and < 100,000 people

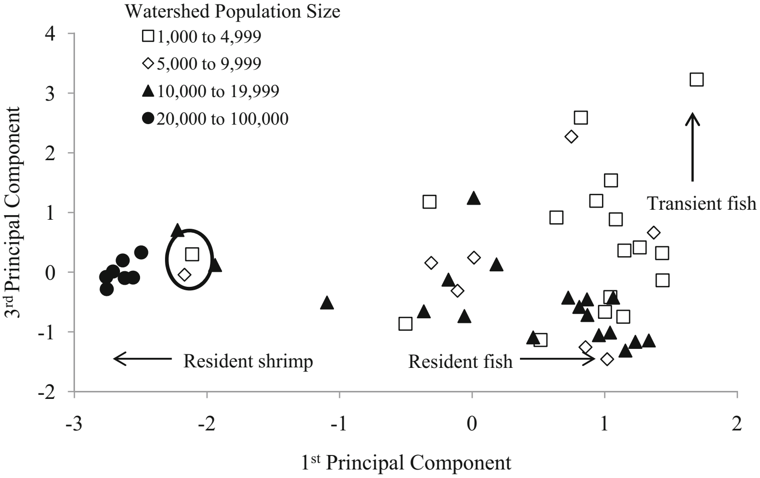


Fig. 14.4 First and third principal components for reference sites ($n = 51$) in watersheds with > 1,000 and < 100,000 people. The circle indicates two marshes in small watersheds (in Rhode Island) that are surrounded by other watersheds with dense human population

Land Use and Nekton Guilds

Principal Component Analyses using watershed land cover types (15 land cover categories) were conducted to determine if the amount of developed land cover was correlated with nekton community composition. The first three principal components accounted for 54 % of the total variation. The first principal component score, which accounted for 22 % of the variability, loaded positively for all the developed land cover categories and negatively on the natural land use types (e.g., rocks and sand, emergent wetlands, grassland, and water); however, there were also several natural land cover categories (e.g., shrub lands and forests types) that had positive loadings. This pattern of loadings would make it difficult to use these land cover categories as indicators of watershed development that might correlate with the nekton life history groups. Additionally, there were no useful correlations (less than 10 % of the variability explained) between the first principal component for the nekton groups and the developed land cover categories for all iterations (all sites, reference sites, and impacted sites).

The land cover within a 5 km buffer from the center of each site was estimated and nekton residency structure was examined using Canonical Correspondence Analysis to evaluate the relationship between the nekton life history groups and immediately adjacent land cover. A significant correlation was observed between the nekton life history groups and adjacent land cover for all reference sites ($n = 96$) with the first axis (Pearson Correlation: 0.80, $p = 0.01$) and second axis (Pearson Correlation: 0.65, $p = 0.01$), accounting for 43 % of the variation. Reference sites tended to be grouped by marshes with resident fish and transient fish (negative scores for the first and second axes) and those that had predominately resident shrimp and resident crabs (positive first and second axis scores). The land cover scores for wetland and forested lands had negative first and second axes scores while the developed land use categories had generally positive first and second axes scores. Therefore, sites with high proportions of resident and transient fish tended to be surrounded by more natural land (wetlands and forests); whereas, sites that were dominated by resident shrimp and resident crabs tended to be associated with marshes that were surrounded by developed land (Fig. 14.5). This analysis was in agreement with the previous Principal Component Analyses indicating that shrimp-dominated communities were associated with watersheds having higher human populations, corresponding with more developed land.

Ecological Thresholds for Salt Marsh Nekton Communities

We propose a series of metrics and ecological thresholds for salt marsh nekton community condition (Fig. 14.6, Table 14.3) that may be indicative of changes in community condition. We developed ecological thresholds for three life history groups (resident fish, transient fish, and resident (Palaemonidae) shrimp) based on the 25th and 75th percentiles of relative abundance (*sensu* Plafkin et al. 1989; Deegan et al. 1997; Hughes et al. 2002) along a gradient of watershed population size using data from the reference sites as they provided the best correlation among these life history groups with watershed population size (Fig. 14.4). If it is assumed that reference sites in watersheds with low populations (1,000–5,000 people) were representative of a desired condition and sites in watersheds with the higher populations

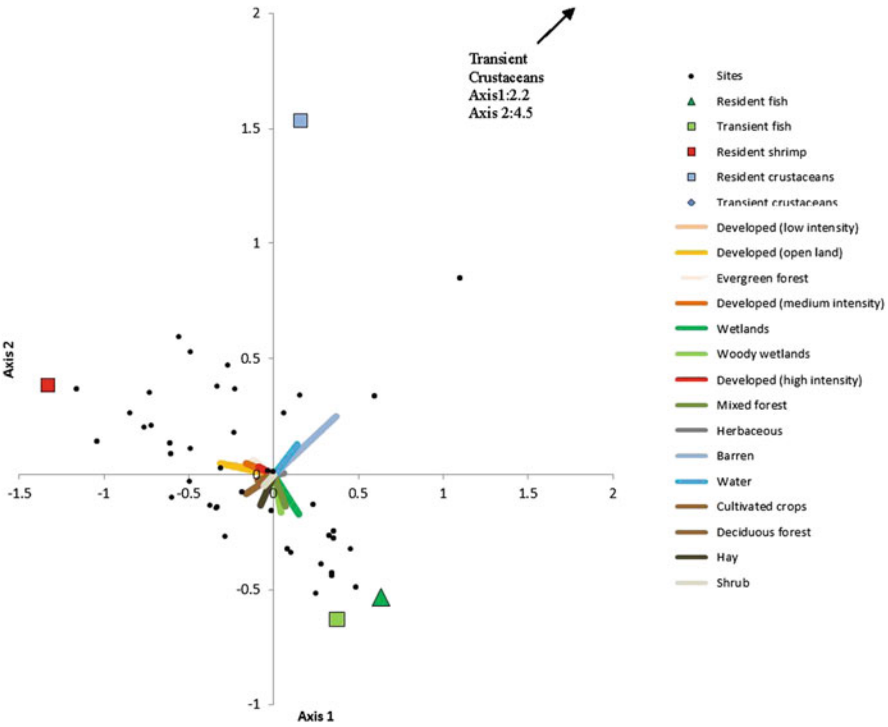


Fig. 14.5 Canonical Correspondence Analysis biplot showing relationship among nekton life history groups and immediate adjacent land cover (5 km radius) for all reference sites ($n = 96$). The five nekton groups are shown as symbols and land cover types shown as lines

were representative of a stressed condition, then the relative abundance of nekton in these life history groups can provide the basis for ecological thresholds at differing levels of anthropogenic stress (Table 14.3)

Other metrics that could be used as a basis for ecological thresholds are the presence of invasive or exotic species and species richness. Invasive and/or exotic species (e.g., Asian shore crab (*Hemigrapsus sanguineus*), green crab) are viewed as potential threats to the salt marsh nekton community (Mack et al. 2000; USEPA 2008), and while they accounted for less than 0.5 % in our data their presence in the nekton community should be noted (Table 14.3). There was no correlation with nekton species richness for any of the iterations in our analyses; however, species richness is clearly an indicator of ecosystem condition, with richness declining as condition deteriorates (Karr 1981). Ecological thresholds for species richness were estimated using the 10th, 50th, and 75th percentiles of species richness for all salt marsh sites (Table 14.3).

To evaluate condition of the nekton community at an individual marsh site, a score of 5 = good, 3 = moderate, or 1 = poor was assigned for each of the five metrics,

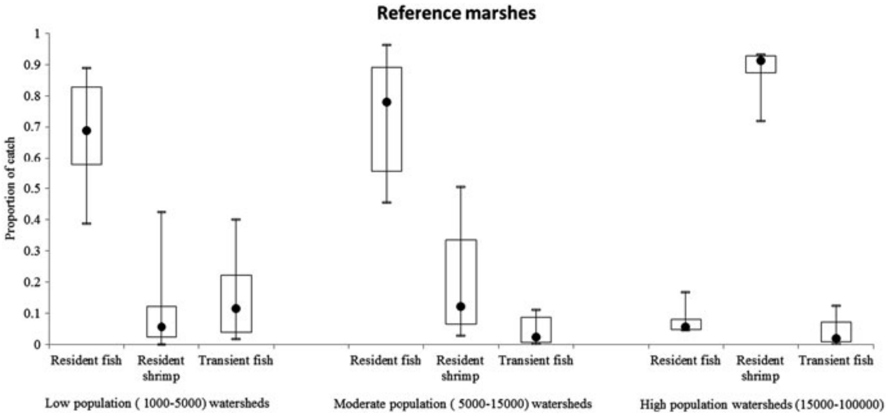


Fig. 14.6 Estimation of metric values for ecological thresholds associated with nekton life history groups. Metric values were based on proportion of catch (relative abundance) for reference sites ($n = 51$) from three classes of watershed population size. Box plots are composed of 25th and 75th percentile (*box*) with the 50th percentile shown as a solid circle, and 10th and 90th percentiles shown as vertical lines

Table 14.3 Ecological thresholds for the relative abundance total nekton across a range of community condition. Scores associated with each condition category are indicated in parentheses

Metric	Good condition (5)	Moderate condition (3)	Poor condition (1)
Resident fish (relative abundance)[a]	> 60 %	30–60 %	< 30 %
Resident shrimp (relative abundance)[a]	< 15 %	15–50 %	> 50 %
Transient fish (relative abundance)[a]	> 15 %	15–4 %	< 4 %
Exotic species	absent	–	present
Species richness[b]	≥ 10	6–9	≤ 5
Overall Condition (sum of scores)	21–25	20–12	5–11

[a] Based on quartile values for reference sites in low population (1,000–5,000), moderate population (5,000–15,000), and high population (15,000–100,000) watersheds (refer to Fig. 14.6)

[b] Based on 10th, 50th, and 75th percentile values for all sites

then summed to derive an overall condition score that ranged from 5 to 25, similar to previous indices of nekton community condition (e.g., Karr 1981; Deegan et al. 1997; Vile 2008) (Table 14.3). For example, assuming the data reveal that the proportion of resident fish, resident shrimp, and transient fish was 75 % (score of 5), 10 % (score of 5), and 10 % (score of 5), respectively, with exotic species absent (score of 5) and a richness of 9 species (score of 3), the overall condition score would be 21 or good. A declining trend in the overall score over time or a difference in score between a reference and impacted site could be used as a decision threshold to initiate management action and/or further study. An improving trend in condition score could suggest that implemented management actions (e.g., restoration efforts) were successful.

Vegetation Analyses

The initial Principal Component analyses using all 156 sites and the 21 common vegetation cover types indicated that only 40 % of the variation was explained by the first three principal components. Similar results were obtained when reference ($n = 80$) and impacted sites ($\mathrm{n} = 76$) were analyzed separately, yielding 47 % and 41 % of the total variation explained by the first three components, respectively. There were no significant correlations with the first principal component and watershed population size (log transformed).

Similar to the nekton analyses, iterations were conducted using the 21 common vegetation covers omitting sites in watersheds with extremely high human populations (~1.8 million people). This did not improve the amount of variation explained by the first three principal components when either all sites were analyzed or when reference and impacted sites were analyzed separately, with the first three principal components again explaining only about 44 % of the variation. As with the previous analyses, there were no significant correlations with watershed population size.

More of the variation in the data were described by the 21 common vegetation covers when sites in both high (~1.8 million) and low (< 1,000) population watersheds were removed. In this iteration, the amount of variation explained by the first three principal components for the reference sites ($n = 43$) was 54 and 56 % for the impacted sites ($n = 44$). None of these iterations yielded significant correlations between the first principal component and watershed population size.

Grouping all vegetation covers by wetland status and salinity tolerance (Table 14.2) yielded the following results. For impacted sites ($n = 76$), the first three principal components accounted for 68 % of the variation, while the first three principal components of the reference sites ($n = 80$) only explained 48 % of the variation. Neither of these iterations resulted in useful relationships with watershed population size.

Canonical Correspondence Analysis indicated that land use within a 5 km radius of the reference sites was correlated with the wetland status and salinity tolerance groups. The first three axes explained 47 % of the variation in marsh vegetation, with high correlations between land use cover and the first axis observed only for herbaceous (0.703), hay (0.524), and emergent wetland (0.455) cover types. These landuse categories tend to be associated with marshes having higher proportions of wetland and transitional plants with medium salinity tolerances. It is likely that this was a result of a correlation of the marshes with themselves, since most sites were contained within a larger area of emergent wetland. For impacted sites, the first three axes accounted for 67 % of the variation in wetland vegetation, with the first axis accounting for 44 % of the variation. There were high correlations with some of the developed land use categories with the first axis (medium intensity developed land, − 0.77; high intensity developed land, − 0.53, low intensity development, − 0.34). These land use covers tended to be associated with sites that had higher proportions of invasive and low salinity tolerant wetland plants (Fig. 14.7).

There were distinct differences in the percent cover of the wetland status and salinity tolerance groups among the different types of site impacts (Fig. 14.8). ANOSIM

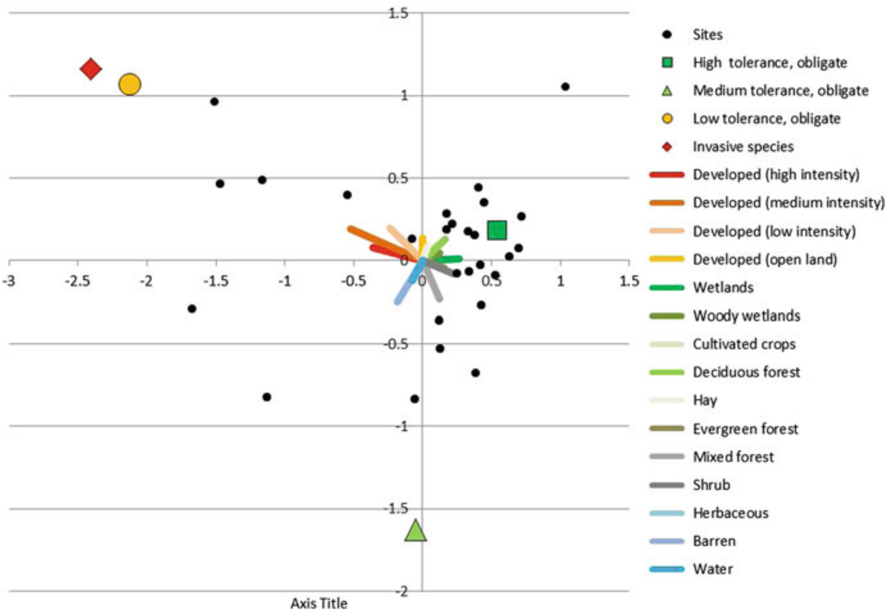


Fig. 14.7 Canonical Correspondence Analyses biplot showing relationship among wetland status and salinity tolerance groups for marsh vegetation and immediate adjacent land cover (5 km radius) for impacted marshes ($n = 76$). The four primary vegetation groups responsible for differences among marshes, as determined by SIMPER analysis, are showing as symbols and land cover types shown as lines

indicated that reference and hydrologically altered (e.g., open marsh water management for mosquito control) sites had similar percent cover for the vegetation groups; similarities were also observed between sites that were tide-restricted or had been recently tidally restored. All other comparisons were significantly different (Global $R = 0.36$, ANOSIM for all comparisons $p < 0.01$). SIMPER analyses indicated that these four significant comparisons (reference vs. tide-restricted, reference vs. tide-restoring, tide-restricted vs. hydrologically altered, tide-restoring vs. hydrologically altered) the same four vegetation groups explained from 84 to 92 % of the overall dissimilarity between the groups. High salinity tolerant wetland plants accounted for the majority of the dissimilarity (42–68 %) for all four significant comparisons. Low salinity tolerant wetland plants accounted for 9–17 % of the dissimilarity between comparisons, followed by invasive species, which explained 8–17 % of the dissimilarity, and medium salinity tolerant wetland plants, which accounted for 6–8 % of the dissimilarity.

Ecological Thresholds for Vegetation Communities

Vegetation groups based on wetland status and salinity tolerance were found to be the best descriptors across a gradient of physical marsh impacts. Four of the vegetation

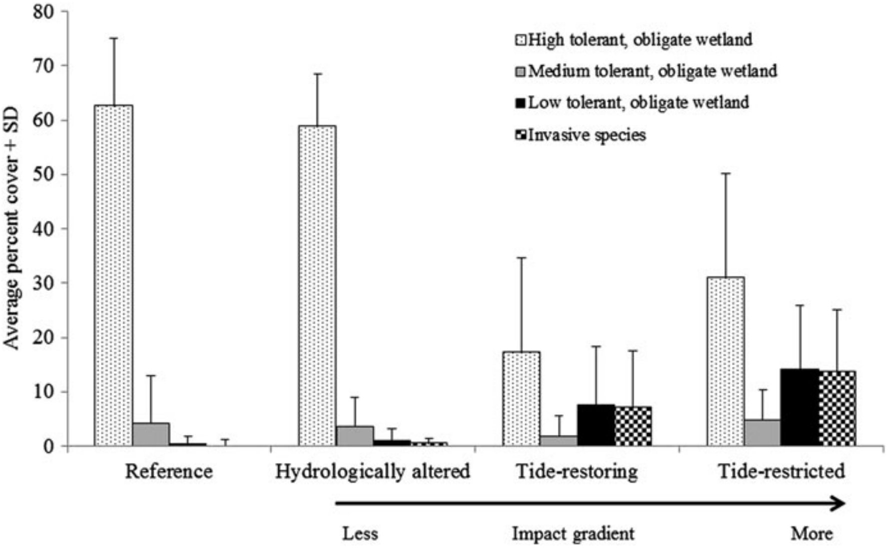


Fig. 14.8 Average percent cover (+ standard deviation) for the four primary vegetation groups by degree of impact to marsh sites

groups, high salinity tolerant wetland plants, medium salinity tolerant wetland plants, low salinity tolerant wetland plants, and invasive plants accounted for the majority of variation, and thus seem to be good metrics to assess change (Figs. 14.7 and 14.8). Previous efforts have similarly identified the abundance or extent of salinity tolerant vegetation as a metric to assess wetland status using smaller data sets (e.g., Konisky et al. 2006; Wigand et al. 2010). Konisky et al. (2006) used the abundance of halophyte and brackish vegetation, while Wigand et al. (2010) incorporated the extent of *Spartina alterniflora* and *S. patens*, as metrics in their assessment methodology.

The four vegetation wetland status and salinity tolerance groups accounted for 84–92 % of the dissimilarity between marshes with differing degrees of physical impact types. Additionally, the rank order of importance of these vegetation groups was the same for all of the significant comparisons (from reference to severely impacted marshes), indicating that these groups were important descriptors for all types of marshes. Therefore, these groups can be useful indicators of vegetation community change if the impact type is used as a proxy for anthropogenic stressor (in increasing order of stress: reference, hydrologically altered, tide-restoring, and tide-restricted). The percent cover of these groups can be used to assess the status of salt marsh communities, assuming that reference marshes represent desired conditions or a restoration target condition. Like the metrics for nekton, the 25th and 75th quartiles of the percent cover of the four vegetation groups were used as ecological threshold metrics (Fig. 14.9, Table 14.4). Scores were assigned to each condition category (good = 5, moderate = 3, poor = 1), with the sum of the scores providing a guide to the condition of the vegetation community. The scores range from 4 to 20, with scores less than 9 representative of a poor salt marsh vegetation community, scores

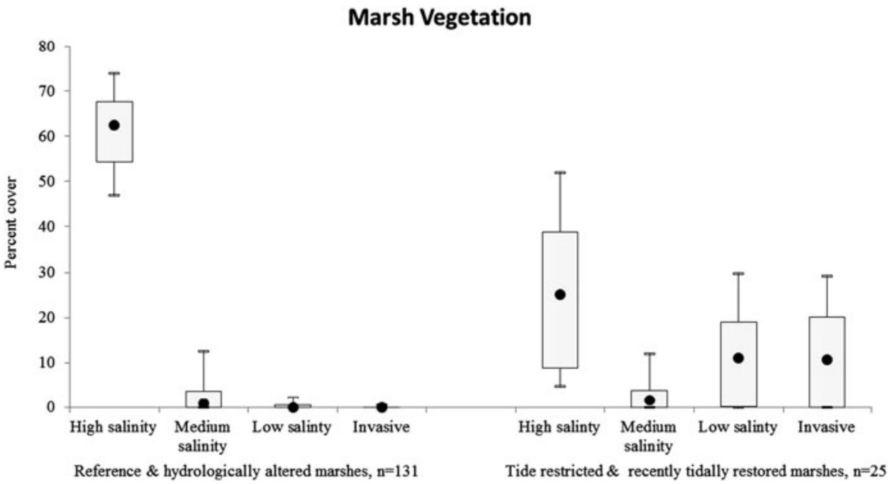


Fig. 14.9 Estimation of metric values for ecological thresholds associated with salt marsh vegetation groups. Metric values were based on the percent cover of each vegetation group for relatively undisturbed marshes (reference and hydrologically altered marshes) and disturbed marshes (tidally restricted and recently tidally restored marshes). Box plots are composed of 25th and 75th percentile (*box*) with the 50th percentile shown as a solid circle, and 10th and 90th percentiles shown as vertical lines

Table 14.4 Ecological thresholds for percent cover the four primary vegetation groups across a range of community condition

Metric	Good condition[a] (5)	Moderate condition[b] (3)	Poor condition[c] (1)
High salinity tolerant, wetland plants	> 55 %	40–55 %	< 40 %
Medium salinity tolerant, wetland plants	< 4 %	4–12 %	> 12 %
Low salinity tolerant, wetland plants	< 2 %	2–20 %	> 20 %
Invasive plant species	< 1 %	1–20 %	> 20 %
Overall Condition (sum of scores)	18–20	10–17	4–9

[a] Based on quartile values for reference and hydrologically altered marshes (refer to Fig. 14.9)

[b] Based on quartiles in between “Good” and “Poor” condition

[c] Based on quartile values from restricted and restored marshes (refer to Fig. 14.9)

between 10 and 17 a moderate vegetation community, and a score above 18 indicative of a salt marsh vegetation community in good condition.

Discussion

Life history characteristics and trophic status of fishes are commonly used as indicators of biotic integrity as fish community structure reflects the quality of habitat. These indices or metrics condense various attributes of the fish community (e.g., species richness, dominance) and often vary in relation to differences in environmental quality (Hughes et al. 2002). The majority of research on biotic indices for fishes has been focused on estuarine and freshwater assemblages (e.g., Karr 1981; Deegan et al. 1997; Martinho et al. 2008; Vile 2008) with little effort specifically targeted on salt marsh nekton communities. We found that grouping nekton species by life history characteristics was the best descriptor of salt marsh nekton communities. Our analyses revealed that three nekton groups (resident shrimp, resident fish, and transient fish) explained the majority of the variation and are reasonable candidates for detecting changes in the salt marsh nekton community. Additionally, nekton community composition changed as human population increased in the watershed. The nekton community of relatively undisturbed marshes (e.g., reference marshes) in low population watersheds was characterized by a high proportion of resident marsh fish species, the presence of transient fish, and a low proportion of Palaemonidae shrimp (Fig. 14.8). As watershed population size increased there was a transition from a resident fish-based community to a Palaemonidae shrimp-based community.

When the threshold values were applied to nekton communities sampled from impacted marshes (only reference sites were used in our previous analyses for defining metrics), the majority (73 %) of the marshes were classified as "moderate" condition. Twenty-one percent were classified as "poor" condition and the remaining 6 % as "good" condition. This classification corresponded relatively well with the marsh impact type, as the majority of the nekton communities classified as moderate were from marshes that had undergone open marsh water management activity for mosquito control (which is a subtle impact when compared to a tidal restriction) or were sites that had recently been tidally restored. When applied to the reference sites, 75 % of the sites were classified as having "good" or "moderate" nekton communities. The sites that were classified as having "poor" nekton communities (26 sites) were those from areas of dense human population areas (Rhode Island and Long Island, New York); however, four sites, were from low population watersheds in Maine and Massachusetts and were classified as low primarily due to low species richness (Maine sites) or a Palaemonidae shrimp-dominated community (Massachusetts).

The utility of these metrics as decision thresholds is demonstrated by examining the East Harbor tidal restoration project in Cape Cod National Seashore, MA. After 140 years of severe tidal restriction, tidal flow was reintroduced to the system in 2002. When the ecological threshold criteria were applied to nekton community data sampled in 2003, shortly after tidal flow reintroduction and again three years after tide restoration, the scores increased from "moderate" to "good," indicating that the nekton community was responding positively to the tidal restoration. This change was primarily due to an increase in resident fish, decrease in resident Palaemonidae shrimp, and increase in species richness. Conversely, scores from the urban Jamaica Bay, New York, Gateway National Recreation Area, indicate that the nekton community of the Big Egg marsh may be deteriorating, as scores from 2003 to 2006 show a decline from "moderate" to "poor" condition. This was primarily due to an increase in resident Palaemonidae shrimp and decrease in transient fish species. Thus, the change in condition based on these metrics can be used as decision thresholds to initiate further management action.

Similar to the nekton metrics, the vegetation metrics were applied to a data base of 20 salt marshes (seven tide-restricted or tide-restoring sites and 13 reference sites) that were not used in the previous estimation of ecological thresholds (due a different method of estimating percent cover, the visual estimate method). The sum of the percent cover for the four vegetation metrics were calculated and assigned a score based on the threshold values defined in Table 14.4. Four of the seven restricted/restoring sites were classified as poor condition as expected. Three of the sites were estimated as poor condition were from a tide-restricted site in Rhode Island (Sachuest Point marsh) representing the vegetation community prior to re-introduction of tidal flow and then 2 years immediately flowing tidal restoration. Prior to restoration, this site was dominated by invasive species (e.g., common reed, *Phragmites australis*). In the years immediately following tidal restoration there was a decrease in abundance of common reed and other low salinity tolerant vegetation, such as cattail (*Typha angustifolia*; Roman et al. 2002). While the vegetation metrics still ranked the condition of the newly restored marsh as poor, the decrease in low salinity tolerant vegetation was identified by the metric, and the score increased from 3 to 5 after tidal flow was reintroduced to the restricted marsh. Data collected eight years post-restoration at this same site, yielded a condition of moderate for the restored marsh, indicating that the vegetation of the restored marsh had transitioned to the high salinity tolerant wetland vegetation commonly found in marshes with unimpaired tidal flow. The data sets from the reference site (Sachuest Point reference marsh) scored as moderate to good condition, and similarly, of the thirteen sites considered reference marshes, ten were classified in good condition and three were classified as moderate condition. The three reference marshes classified as moderate tended to have lower scores for invasive species and percent of low salinity tolerant wetland vegetation (i.e., fewer invasive species and lower percent cover of brackish wetland plants). While it is widely accepted that marshes unimpaired tidal flow tend to be dominated by salinity tolerant wetland vegetation such as *Spartina* grasses and that tidally restricted marshes tend to have higher percent cover of less saline tolerant species and more invasive species (e.g., *Phragmites*; e.g., Roman et al 2002; Buchsbaum et al. 2006; Konisky et al. 2006), our analyses provide ecological thresholds linked to community condition that characterize reference and impacted sites. Similar to the nekton metrics, the ecological thresholds for salt marsh vegetation appear to be applicable along the northeastern US coast.

Future Directions

Because we have linked these ecological thresholds to salt marsh community condition, any shift from one condition to another or a cumulative decrease in condition over all metrics can be used as decision thresholds to guide future management actions. For example, if it is found than an individual marsh deviates from the reference standard, or if monitoring data indicate that the condition scores are changing, then these results can be used as decision thresholds to trigger management activities.

Alternatively, the thresholds can be used to interpret restoration response monitoring data. This research provides resource managers with an approach to defining the condition of salt marshes, based on the species composition and abundance of nekton and vegetation communities. Supported by an extensive nekton and vegetation data base, including multiple years of monitoring from over 80 salt marshes within the Maine to Virginia region and including a broad diversity of conditions from relatively undisturbed to highly impacted, we present ecological thresholds linked to salt marsh community condition (from good to poor) for numerous metrics within a framework similar to indices of biotic integrity. Within the northeast, several programs are supporting the collection of long-term data on salt marsh nekton and vegetation that could easily support the condition metrics presented in this paper (e.g., National Park Service, US Fish and Wildlife Service, National Estuarine Research Reserves, state agencies).

The metrics and ecological thresholds presented herein are relevant to salt marshes of the northeastern United States. The methods used to define the metrics and derive the ecological thresholds can be applied to other coastal regions, but a strong foundation of region-specific data would be required to set the threshold values and perhaps establish different metrics. In addition to the need for region-specific ecological threshold metrics, it is expected that additional metrics may be included for northeastern US salt marshes or condition thresholds presented in this paper will be refined as new data become available and additional analyses are conducted.

Acknowledgments We are grateful to the many graduate and undergraduate students at the University of Rhode Island, staff from the National Park Service (NPS), U.S. Fish and Wildlife (USFWS), National Marine Fisheries Service Restoration Center, and volunteers who participated in the many projects that resulted in the data used in our analyses. Thanks are also extended to Kenny Raposa (Narragansett Bay National Estuarine Research Reserve), Susan Adamowicz (USFWS), Patti Rafferty (NPS), and Marci Cole and Wenley Ferguson (Save The Bay) and for providing additional data sets. Thanks are extended to Serena Moseman-Valtierra who graciously reviewed a draft of the manuscript and provided helpful suggestions. The authors would like to Glenn Guntenspergen for editing this volume. Reference to commercial products in this manuscript does not imply endorsement by the US Government and views and opinions of authors expressed herein do not necessarily reflect those of the US Government. This research was supported by the USGS Status and Trends Program and the National Park Service Inventory and Monitoring Program through a grant to the University of Rhode Island.

References

Able, K. W., D. M. Nemerson, P. R. Light, and R. O. Bush. 2000. Initial response of fishes to marsh restoration at a former salt hay farm bordering Delaware Bay. In *Concepts and Controversies in Tidal Marsh Ecology*, ed. M. P. Weinstein and D. A. Kreeger, 749–773. The Netherlands: Kluwer Academic Publishers.

Able, K. W., D. A. Witting, R. S. McBride, R. A. Rountree, and K. J. Smith. 1996. Fishes of polyhaline estuarine shores in Great Bay–Little Egg Harbor, New Jersey: A case study of seasonal and habitat influences. In *Estuarine Shores: Evolution, Environments, and Human Alterations*, ed. K. F. Nordstrom and C. T. Roman, 335–353. New York: John Wiley and Sons.

Ayvazian, S. G., L. A. Deegan, and J. T. Finn. 1992. Comparisons of habitat use by estuarine fish assemblages in the Acadian and Virginian Zoographic Provinces. *Estuaries* 15:368–383.

Bigelow, H. B., and W. C. Schroeder. 1953. Fishes of the Gulf of Maine. United States Fish and Wildlife Service. *Fisheries Bulletin* 53:1–575.

Bourn, W. S., and C. Cottam. 1950. Some biological effects of ditching tidewater marshes. Research Report 19. US Fish and Wildlife Service, Washington, DC.

Buchsbaum, R. N., J. Catena, E. Hutchins, and M-J. James-Pirri. 2006. Changes in salt marsh vegetation, *Phragmites australis*, and nekton in response to increased tidal flushing in a New England salt marsh. *Wetlands* 26:544–557.

Clarke, K. R., and R. N. Gorley. 2006. PRIMER (Plymouth Routines in Multivariate Ecological Research) v6: User Manual/Tutorial. Plymouth: PRIMER-E.

Clarke, K. R., and R. M. Warwick. 2001. *Change in marine communities: An approach to statistical analysis and interpretation*. 2nd ed. Plymouth: PRIMER-E

Collette, B. B., and G. Klein-MacPhee, eds. 2002. *Bigelow and Schroeder's fishes of the Gulf of Maine*. 3rd ed. Washington: Smithsonian Institute Press.

Deegan, L. A., J. A. Finn, S. G. Ayvazian, C. A. Ryder-Kieffer, and J. Buonaccorsi. 1997. Development and validation of an estuarine biotic integrity index. *Estuaries*. 20:601–617.

Donnelly, J. P., and M. D. Bertness. 2001. Rapid shoreward encroachment of salt marsh cordgrass in response to accelerated sea-level rise. *Proceedings of the National Academy of Science* 98:14218–14223.

Elliot, M., A. K. Whitfield, I. C. Potter, S. J. M. Blaber, D. P. Cyrus, F. G. Nordlie, and T. D. Harrison. 2007. The guild approach to categorizing estuarine fish assemblages: A global review. *Fish and Fisheries* 8:241–268.

Elzinga, C. L., D. W. Salzer, J. W. Willoughby, and J. P. Gibbs. 2001. *Monitoring Plant and Animal Populations*. Malden: Blackwell Scientific.

ESRI. 2005. *ArcGIS 9.1 software. Environmental Systems Research Institute (ESRI)*. Redlands.

Friedland, K. D., G. C. Garman, A. J. Bejda, and A. L. Studholme. 1988. Interannual variation in diet and condition in juvenile bluefish during estuarine residency. *Transactions of the American Fisheries Society* 117:474–479

Hartig, E. K., V. Gornitz, A. Kolker, F. Mushacke, and D. Fallon. 2002. Anthropogenic and climate-change impacts on salt marshes of Jamaica Bay, New York City. *Wetlands* 22:71–89.

Hughes, J. E., L. A. Deegan, M. J. Weaver, and J. E. Costa. 2002. Regional application of an index of estuarine biotic integrity based on fish communities. *Estuaries* 25:250–263.

James-Pirri, M. J., R. M. Erwin, D. J. Prosser, and J. Taylor. 2011. Responses of salt marsh ecosystems to mosquito control management practices along the Atlantic Coast (U.S.A.) Restoration Ecology. doi:10.1111/j.1526-100X.2010.00767.x (published online Feb 2011).

James-Pirri, M. -J., C. T. Roman, J. L. Swanson. 2010. A quantitative gear to sample nekton in salt marsh ditches and small tidal creeks. *Transactions of the American Fisheries Society* 139:413–419. doi:10.1577/T09-106.1

James-Pirri, M. J., R. M. Erwin, D. J. Prosser. 2008. US Fish and Wildlife Service (Region 5) Salt Marsh Study, 2001–2006: An assessment of hydrologic alterations on salt marsh ecosystems along the Atlantic coast. USGS Patuxent Wildlife Research Center and University of Rhode Island, Final Report to U.S. *Fish and Wildlife Service* :427.

James-Pirri, M. -J., C. T. Roman, and J. Heltshe. 2007. Power analysis to determine sample size for monitoring vegetation change in salt marsh habitats. *Wetlands Ecology and Management* 15:335–345. doi:10.1007/s11273-007-9034-x.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6:21–27.

Kent, M., and P. Coker. 1992. *Vegetation description and analysis: A practical approach*. England: Wiley.

Konisky, R. A., D. M. Burdick, M. Dionne, and H. Neckles. 2006. A regional assessment of salt marsh restoration and monitoring in the Gulf of Maine. *Restoration Ecology* 14:516–525.

LaBrecque, E., C. Fritz, J. Tober, P. J. Behr, and I. Valiela. 1996. Abundance and age-specific growth rates in relation to population densities of *Fundulus heteroclitus* in Waquoit Bay estuaries subject to different nitrogen loads. *Biological Bulletin* 191:319–320.

Mack, R. N., D. Simberloff, W. M. Lonsdale, H. Evans, M. Clout, and F. A. Bazzaz. 2000. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* 10:689–710.

Martin, J., M. C. Runge, J. D. Nichols, B. C. Lubow, and W. L. Kendall. 2009. Structured decision making as a conceptual framework to identify thresholds for conservation and management. *Ecological Applications* 19:1079–1090.

Martinho, F., I. Viegas, M. Dolbeth, R. Leitao, H. N. Cabral, and M. A. Pardal. 2008. Assessing estuarine environmental quality using fish based indices: Performance evaluation under climatic instability. *Marine Pollution Bulletin* 56:1834–1843.

Niering, W. A., and R. S. Warren. 1980. Vegetation patterns and process in New England salt marshes. *Bioscience* 30:301–307.

Nixon, S. W., and C. A. Oviatt. 1973. Analysis of local variation in the standing crop of *Spartina alterniflora*. *Botanica Marina* 16:103–109.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish. U.S. Environmental Protection Agency, Office of Water Regulations and Standards: Washington, D.C. EPA 440-4-89-001.

Raposa, K. B., and C. T. Roman. 2000. Monitoring nekton in shallow estuarine habitats. Part of a series of monitoring protocols for the Long-term Coastal Ecosystem Monitoring Program at Cape Cod National Seashore. USGS Patuxent Wildlife Research Center, Coastal Research Field Station, University of Rhode Island, Narragansett, RI 02882. http://science.nature.nps.gov/im/monitor/protocoldb.cfm.

Raposa, K. B., C. T. Roman, and J. F. Heltshe. 2003. Monitoring nekton as a bioindicator in shallow estuarine habitats. *Environmental Monitoring and Assessment* 81:239–255.

Roman, C. T., R. A. Garvine, and J. W. Portnoy. 1995. Hydrologic modeling as a predictive basis for ecological restoration of salt marshes. *Environmental Management* 19:559–566.

Roman, C. T., M. J. James-Pirri, and J. F. Heltshe. 2001. Monitoring Salt Marsh Vegetation: Part of a series of monitoring protocols for the Long-term Coastal Ecosystem Monitoring Program at Cape Cod National Seashore. USGS Patuxent Wildlife Research Center, Coastal Research Field Station, University of Rhode Island, Narragansett, RI 02882. http://science.nature.nps.gov/im/monitor/protocoldb.cfm.

Roman, C. T., W. A. Niering, and R. S., Warren. 1984. Salt marsh vegetation change in response to tidal restriction. *Environmental Management* 8:141–150.

Roman, C. T., K. B. Raposa, S. C. Adamowicz, M. J. James-Pirri, and J. G. Catena. 2002. Quantifying vegetation and nekton response to tidal restoration of a New England salt marsh. *Restoration Ecology* 110:450–460.

Rozas, L. P., and T. J. Minello. 1997. Estimating densities of small fishes and decapod crustaceans in shallow estuarine habitats: a review of sampling design with focus on gear selection. *Estuaries* 20:199–213.

Sekiguchi, K. 1995. Occurrence, behavior and feeding habits of harbor porpoises (*Phocoena phocoena*) at Pajaro Dunes, Monterey Bay, California. *Aquatic Mammals* 21:91–103.

Smith, J. P. 1997. Nesting season food habits of four species of herons and Egrets at Lake Okeechobee, Florida. *Colonial Waterbirds* 20:198–220.

Swanson, J. L. 2009. *Finding indicators of salt marsh system health using nekton and vegetation data. Master's Thesis*. Kingston: University of Rhode Island.

Tober, J., C. Fritz, E. LaBrecque, P. J. Behr, and I. Valiela. 1996. Abundance, biomass, and species richness of fish communities in relation to nitrogen-loading rates of Waquoit Bay estuaries. *Biological Bulletin* 191:321–322.

USDA. 2009. website. Natural Resources Conservation Service. Plants Database. http://plants.usda.gov/index.html. Accessed June 2009.

U.S. Environmental Protection Agency (EPA). 2008. Effects of climate change for aquatic invasive species and implications for management and research. National Center for Environmental Assessment, Washington, DC; EPA/600/R-08/014. http://www.epa.gov/ncea.

Vile, J. 2008. *Fish IBI report, 2006 sampling, round 2, year 2 of 5. Vols 1. New Jersey Department of Environmental Protection*. Trenton. http://www.state.nj.us/dep/wms/bfbm/download/ibi2006Vol1.pdf.

Wigand, C., R. McKinney, M. Chintala, S. Lussier, and J. Heltshe. 2010. Development of a reference coastal wetland set in Southern New England (USA). Environmental Monitoring and Assessment 161:583–598. doi:10.1007/s10661-009-0770-7.

Wigand, C., R. Cameleo, R. McKinney, G. Thursby, M. Chintala, and M. Charpentier. 2001. Outline of a new approach to evaluate ecological integrity of salt marshes. *Human and Ecological Risk Assessment* 7:1541–1554.

Index

G. R. Guntenspergen (ed.), *Application of Threshold Concepts in Natural Resource Decision Making,* DOI 10.1007/978-1-4899-8041-0,