Ecology, 71(6), 1990, pp. 2060–2068

LARGE-SCALE MANAGEMENT EXPERIMENTS AND LEARNING BY DOING[1]

CARL J. WALTERS
Resource Ecology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

C. S. HOLLING
Department of Zoology, University of Florida, Gainesville, Florida 32611 USA

Abstract. Even unmanaged ecosystems are characterized by combinations of stability and instability and by unexpected shifts in behavior from both internal and external causes. That is even more true of ecosystems managed for the production of food or fiber. Data are sparse, knowledge of processes limited, and the act of management changes the system being managed. Surprise and change is inevitable. Here we review methods to develop, screen, and evaluate alternatives in a process where management itself becomes partner with the science by designing probes that produce updated understanding as well as economic product.

INTRODUCTION

Most of the world's ecosystems are affected to some degree by harvesting and related activities aimed at particular "resource" types or species. But in no place can we claim to predict with certainty either the ecological effects of the activities, or the efficacy of most measures aimed at regulating or enhancing them. Every major change in harvesting rates and management policies is in fact a perturbation experiment with highly uncertain outcome, no matter how skillful the management agency is in marshalling evidence and arguments in support of the change. Practicing resource managers have long been aware of this point, and have tried to invest in adequate monitoring and evaluation programs even while maintaining a public stance of confidence in their predictions. Recently, it has become fashionable to admit at least some degree of ignorance and to label substantial management initiatives as experiments, even when a scientist would shudder at how poorly they are designed. There are major opportunities for research ecologists to become involved in the design and conduct of such experiments, to the mutual benefit of scientists, managers, and resource users.

This paper discusses some challenges for justifying and designing experimental management programs. The first challenge is to demonstrate that a substantial, deliberate change in policy should even be considered, given the alternative of pretending certainty and waiting for nature to expose any gaps in understanding. A second challenge is to expose uncertainties (in the form of alternative working hypotheses) and management decision choices in a format that will promote both intelligent choice and a search for imaginative and safe experimental options, by using tools of statistical decision analysis. A third challenge is to identify experimental designs that distinguish clearly between localized and large-scale effects, and hence, make the best possible use of opportunities for replication and comparison. A fourth challenge is to develop designs that will permit unambiguous assessment of transient responses to policy changes, in the face of uncontrolled environmental factors that may affect treated and reference experimental units differently. Finally, there is need for imaginative ways to set priorities for investments in research, management, and monitoring, and for design of institutional arrangements that will be in place for long enough to measure large-scale responses that may take several decades to unfold.

PASSIVE VS. ACTIVE ADAPTATION

There are three ways to structure management as an adaptive process (Walters 1986): (1) evolutionary or "trial and error," in which early choices are essentially haphazard, while later choices are made from a subset that gives better results; (2) passive adaptive, where historical data available at each time are used to construct a single best estimate or model for response, and the decision choice is based on assuming this model is correct; or (3) active adaptive, where data available at each time are used to structure a range of alternative

[1] For reprints of this Special Feature, see footnote 1, page 2037.

response models, and a policy choice is made that reflects some computed balance between expected short-term performance and long-term value of knowing which alternative model (if any) is correct. Most theoretical literature on resource management is aimed at providing single best predictions of policy choice, and hence, presupposes that a passive strategy is best. Perturbation experiments are most likely to arise when an active adaptive strategy is adopted; the main use of active strategies has been in agriculture (field tests, rotation policies) and fisheries (varying harvest rates, hatchery systems).

There are two fundamental objections to passive adaptive policies. First, they are likely to confound management and environmental effects. For example, in fisheries with long monitoring histories (50+ yr), we still have bitter debates about the relative importance of fishing and environmental factors in driving population declines and cycles (Walters and Collie 1988). Second, passive policies may fail to detect opportunities for improving system performance if the "right" model and the "wrong" model predict the same response pattern when the system is managed as though the wrong model were correct.

Passive adaptive approaches sometimes do result in informative experiments. For example, much of the Florida Everglades system has been lost to agricultural development, and water regimes in the remaining marsh have been altered by a system of upstream dikes and canals developed to permit storage, diversion, and flood control. These changes had many effects on aquatic communities and vegetation patterns, but the most publicized one has been a drastic decline in wading bird populations (Ogden 1978, Kushlan and Frohring 1986, Frederick and Collopy 1988). Efforts to provide more water to the Everglades National Park, through a quota delivery plan in the 1970s, did not reverse the decline. By the early 1980s, a consensus had emerged that the basic causes of the decline were overall habitat loss and reshaping of the seasonal hydropattern in ways that caused nesting failures. In 1983, Congress passed the Fascell Bill authorizing tests of alternative water delivery plans to the park for 2 yr, and this bill has been extended to the present (1989). In 1985, an "experimental" water release plan, called "The Rainfall Plan," was adopted. Under this plan, water is discharged to the Park so as to restore as "natural" a pattern of seasonal and interannual variation as possible. This plan is passively adaptive in the sense that it was justified by assuming that a single best hypothesis (animals require a natural pattern) is correct, yet is also a disturbance experiment that will cause a substantial change in the timing and distribution of water flows into the Park relative to the past two decades (Kushlan 1987). It is hoped that the plan will represent the first step in an "Iterative Testing Process" to reshape hydrological regimes over time so as to permit recovery of as much of the natural Everglades ecosystem as possible (S. S. Light, J. R. Wodraska, and S. Joe, *unpublished manuscript*).

Would the Everglades Rainfall Plan have been adopted under an active adaptive planning process involving deliberate development of a range of alternative hypotheses about wading bird dynamics and a corresponding evaluation of experimental policy options? At least two broad alternatives deserve consideration. First, at least one species (White Ibis, *Eudocimus albus*) may have shifted its nesting away from the Everglades to more northern areas in the Carolinas (Frederick and Collopy 1988); perhaps changes in such areas, rather than deterioration in the Everglades, are partly responsible for the decline. Second, Everglades nesting colonies were historically concentrated in coastal mangrove areas, from which the birds could initially move out to forage in "short hydroperiod" marshes around the system margin and later use the existing marsh core and estuary as seasonal drying progressed. Now birds nest mainly in interior water storage (conservation) areas. The margin areas are mostly outside the area impacted by the Rainfall Plan, and have been much affected by drainage for agricultural and urban development. They are not currently used consistently by the birds, but they may have been critical for nesting success in the past. If either the "other opportunities" or "marginal areas" hypothesis is correct, populations will fail to recover under the Rainfall Plan and the monitoring program in the experimental area will give no clues as to why the plan failed. An actively adaptive planning process might have resulted in an experimental and monitoring program on a substantially different spatial scale, and a qualitatively different set of manipulations, than is presently considered necessary or feasible.

Embracing Uncertainty: Assessing the Value of Experimental Decision Choices

The design and justification of experimental management plans involve a more complex assessment of risks and benefits than scientists might consider in developing plans for obtaining statistically significant results. In particular, the process of policy design needs to involve a more elaborate and productive interplay between ecologists and decision makers than has traditionally been seen (Holling 1978). This section illustrates some key issues and complexities with an example from fish and wildlife harvest policy design.

For many harvested populations, managers have been "successful" in establishing regulatory programs (restrictions, monitoring, enforcement) to provide relatively stable population sizes and harvests. Usually it

is claimed that the current population size is in some sense "optimum," such that lower or higher sizes would on average be less productive of a harvestable excess each year. This claim is often supported by estimation of population model parameters from historical data.

Now suppose independent analysis based on "habitat capacity" or alternative parameter estimation procedures suggests that the optimum population size has in fact been grossly underestimated, so the current population size is far below optimum. For example, Anderson (1975) presented production parameter estimates for the Mallard (*Anas platyrhynchos*) suggesting that the continental population should perhaps be double the level targeted by management in the 1970s. A series of analyses of the Fraser River sockeye salmon (*Oncorhychus nerka*) have suggested that spawning stocks should be far larger than at present (Collie and Walters 1987), and that the present value of moving to a higher stock level could be as much as a half-billion dollars to the fishing industry.

The usual approach to these cases has been to engage in a "battle of the models," with some scientists and managers defending the status quo and others calling for harvest reductions to allow populations to rebuild to more productive levels. In such battles, the status quo argument usually prevails, since it is a "proven" strategy (low risk) and does not involve short-term pain to the harvesters. However, there is no way to prove that a higher stock would be more productive without actually trying it; no behavioral measurements on the current population, or highly localized experiments with increased density, could be guaranteed to anticipate the full variety of responses (such as colonization of marginal habitats) that might accompany the population increase.

However, in a few cases, such as the Fraser sockeye, an actively adaptive approach to management has been adopted. The essential step in these cases has been to avoid a battle of the models, with key parties (scientists, managers) agreeing instead to "embrace uncertainty" by laying out a decision table of possible outcomes under different alternative hypotheses and policy choices. The simplest possible table might look as follows:

	Alternative hypotheses	
Policy options	Current stock is optimum	Optimum is at higher stock
Maintain status quo	1.0	1.0
Experimentally increase stock	0.5	2.0

Each number in this table is an estimate of the expected long-term value of following the policy option in the row, given that the hypothesis in the column is correct. To simplify the comparison of choices, we normalized the values so that the status quo model/policy has a value of 1.0. A rather tedious modelling process is usually involved in obtaining credible estimates. A technically difficult aspect of such simulations is in representing future learning under the experimental policy option(s); it is necessary to simulate the gathering of noisy data, updating over time of odds placed on the alternative hypotheses (i.e., Bayesian learning process), and the eventual shift to an optimum policy when the correct model becomes reasonably certain. For further details on the modelling methods involved here, see Walters (1986).

The key policy issue is with the possible outcomes of an experimental policy. One possibility is a loss in short-term yield with no compensating long-term increase (0.5), and the other is to have the short-term loss more than balanced by long-term gains (2.0). Of course, if we had calculated the values while using a high discount rate for future harvests (short planning horizon, no interest in long-term resource husbandry), the second column would also have been small. Management experimentation is often meaningless in settings where no value is placed on the long-term utility of experimental results.

How should decision makers react to these values? One possibility is to simply average them using some prior odds placed on the alternative hypotheses; in the example table, such "expected value maximization" would strongly favor the experimental policy option, unless very high prior odds were placed on the status quo hypothesis. Another possibility is to look mainly at the worst possible outcome (0.5) and to place much weight on this risk. Such risk averse behavior may be personally favored by many decision makers, but is not easily justified in the context of public policy making where the decision maker is supposed to be representing the varied interests of many actors. Decision theorists agree that it is important to identify possible choices and outcomes (the decision table) objectively, and to assess the odds of alternative outcomes (assign probabilities to alternative hypotheses), but they have not yet produced a generally accepted procedure for combining the information to arrive at optimum public policy decisions. For further discussion of issues and methods, see Raiffa (1968), DeGroot (1970), Keeney and Raiffa (1976), and Lindley (1985).

A key value of the decision table is to create frustration over a set of policy choices that all appear bad. This frustration can lead to several very useful reactions by both scientists and managers: (1) critical evaluation of the alternative hypotheses, with a more careful eye to identifying others and a serious effort to place better odds on the possibilities through more careful analysis of historical information; (2) a more precise appraisal of the value of information associated with

knowing which hypothesis is correct; (3) a search for objective ways to weigh the risks and benefits of experimentation; and, perhaps most important from a scientific point of view, (4) a search for imaginative policy options that might permit the uncertainties to be resolved through more focused and perhaps less risky experimentation. There is much for scientists and decision makers to learn from one another by working together on all of these steps.

For simple decision-making situations involving harvest management, there have been some attempts to compute optimum feedback policies while explicitly accounting for the effect of informative disturbances on future management performance (Mangel 1985, Walters 1986). These attempts examine the interplay over time of population dynamics and Bayesian learning, using the methods of stochastic dynamic programming to find an optimum policy (Walters 1981, Mangel and Clark 1988). Computational experience to date indicates that the best policy choice for any year will be either to ignore uncertainty (passive adaptive) or to make a fairly dramatic and informative experimental disturbance; minor experiments are not favored because they erode average performance without significantly improving learning rates.

Choosing the Optimum Number of Experimental Replicates

The balance of learning and risks often does not favor experimental disturbances in single, unique, managed systems. However, this conclusion changes drastically when there is a collection of similar units (lakes, distinct populations, areas) that can be managed independently. A key question in such situations is how large an experiment to conduct. Some methodological issues involved in answering this question can be illustrated with a simple hypothetical example.

Suppose there are N units that are all being managed with the same baseline policy, and someone identifies an experimental policy that might be better. Suppose the experimental policy is unlikely to produce persistent effects (over >1 yr) in any unit where it is applied, and that analysis of its possible performance relative to the baseline has resulted in the following decision table of possible outcomes:

	Hypotheses	
Policies	Conservative	Optimistic
Baseline	1.0	1.0
Experimental	0.5	1.2

Here the conservative hypothesis is that the baseline policy is already optimum, so that experimental disturbance of any unit would on average cause a considerable loss (to 0.5). The optimistic hypothesis is that the experimental policy will produce a 20% improvement per unit per year. This table would not favor experimentation if all units must be treated alike because the baseline policy has higher expected value than the experimental policy unless a probability of 2/7 or less is assigned to the conservative hypothesis.

Suppose the management agency can subject n units each year to the experimental policy ($0 < n < N$). For what n will the expected long-term value from all N units be maximum? It may appear that this question can be answered without reference to time, since there are assumed to be no persistent effects of treatment. However, there is a very important temporal dynamic that links decision making from year to year, the "information state" measure $p(t)$ that summarizes the odds placed on the conservative hypothesis by the decision maker in year t. $p(t)$ will not change over time if $n = 0$, but will move toward 0.0 or 1.0 if $n > 0$, depending on which hypothesis is correct. The effect of $n > 0$ in any year is thus to leave the decision maker at time $t + 1$ with different (on average lower) odds of making the wrong decision in year $t + 1$. The dynamics of $p(t)$ are given by Bayes theorem as $p(t + 1) = L(t)p(t)/P(t)$, where $L(t)$ is the likelihood of the responses measured in year t given the conservative hypothesis and $P(t)$ is the total probability of the responses: $P(t) = L(t)p(t) + L'(t)[1 - p(t)]$, where $L'(t)$ is the likelihood of the data given the optimistic hypothesis. The details of the $L(t)$ calculation need not concern us here. A key point is that some assumption about the likelihood of different observed outcomes is necessary for predicting the behavior of $p(t)$, and hence, assessing an optimum value of n.

Given the above decision table and a likelihood function $L(t)$, the optimum value of n can be computed as a function of $p(t)$ by the methods of stochastic dynamic programming (Walters 1981, 1986, 1990, Mangel and Clark 1988). The concept behind the computation is to work backward in time, building up an estimate of the future value of being in different information states $p(t)$ while asking at each time what the optimum n is in relation to the current state and expected future values of the possible states $p(t + 1)$ that might result from the decision at time t. A sample computation of the optimum relationship between n and $p(t)$ is shown in Fig. 1, for the case $N = 10$ and a normal likelihood function with variance 0.25 around mean performance for each unit.

Fig. 1 shows that management prescriptions can be very different from scientific ones. When no future learning is expected, the optimum is either to treat all units with the experimental policy (for $p < 2/7$), or to treat none of them. When learning is considered and the units are assumed to have independent random variation, the optimum is a graded policy that involves

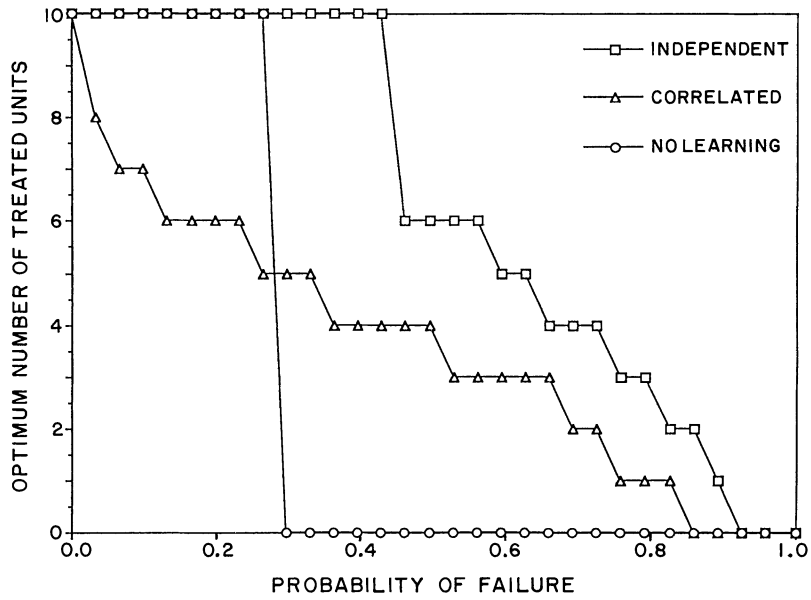

FIG. 1. Effect of the probability of failure of experimental policy on the optimum number of management units required for treatment with an experimental policy that would be best if an alternative, optimistic hypothesis is correct. Three scenarios are presented: Independent (units respond independently to environmental fluctuations); Correlated (units respond in concert with environmental fluctuations); and, No Learning (manager chooses experimentation only if current conservative policy has a low probability of being correct).

treating all the units when $p(t)$ is less than ≈ 0.45, and to treat progressively fewer units (take fewer risks, do a smaller experiment) as $p(t)$ increases. When the units are assumed to have correlated responses to environmental fluctuations, the optimum is to treat fewer units and to give up experimentation completely at a lower value of $p(t)$. The optimum scientific design in this case would be to treat $n = 5$ units every year (balanced design), to achieve a minimum variance of the treatment–control difference. Nowhere in these policies can we see patterns that can be interpreted in terms of standard scientific ideas of statistical significance.

In this example, we assumed that all costs and benefits of experimentation, including any monitoring costs, are contained in the simple decision table. A fruitful area for future statistical research would be to look more carefully at the cost structure, and particularly at how increased monitoring investment would affect the likelihood function $L(t)$, and hence, learning rates. We suspect that the result of this analysis would most often be to recommend larger (larger N, n) experiments and more crude monitoring, rather than the precise and small experiments usually favored by ecological researchers.

SCALE AND REPLICATION

Adaptive policy design must make effective use of opportunities for spatial replication and control. There is a long history of sad experience with the false premise that it is possible to "learn by doing" through sequential application of different policies to whole systems, especially in fisheries (Walters and Collie 1988). In these cases, there is little prospect of resolving the uncertainties through continued monitoring and modest policy change, and policy changes drastic enough to provide unequivocal responses would be socially or economically unacceptable.

Like ecological processes, policies act at a variety of spatial scales. For policies that can be implemented at relatively small scales on a number of experimental units, the main design problem is to control for larger scale biophysical processes and management actions that may link the units. An extreme example is in the Columbia River Basin, where the Northwest Power Planning Council is coordinating development of a massive ($100 million/yr) adaptive management program for mitigating effects of water development on anadromous salmonid populations (Lee 1989). These populations are linked through factors ranging from large-scale marine and freshwater climate variation to shared difficulties in migrating past lower river dams to mixed-stock fisheries that intercept whole collections of migrating populations. Some policy changes, such as reductions in marine harvest rates and improved fish passage facilities at lower river dams, cannot be replicated in space and will affect any localized

experiments within the Basin. Other changes, such as large hatchery developments and improved passage at upriver dams, will affect large segments of the Basin but can be replicated at least a few times. Still others, such as habitat improvements in small streams, can be replicated massively throughout the Basin (and incidentally, provide a wealth of basic scientific information on the mechanisms of population regulation in stream fishes). The challenge is to develop a nested experimental design that will permit clear separation of the effects of as many of these changes as possible, so that a sensible balance of management tools and policies can be developed.

The Problem of Transient Responses

Most actions do not simply change a managed system from one state to another; rather, they induce transient responses that may be quite complex (delays, sharp increases followed by slow decline, cycles, etc.). More important in terms of experimental design, some actions may change the sensitivity of managed systems to natural environmental factors that themselves have complex temporal patterns. In statistical terms, this change in sensitivity will result in "time–treatment interactions," where response to treatment depends on the specific time (environmental conditions) when the treatment is applied (Walters et al. 1988, 1989).

The major implication of time–treatment interaction is that treatment–reference comparisons cannot be trusted to provide a reliable estimate of the treatment transient effect, no matter how many treatment and reference replicates are used, unless treatment is initiated over a range of starting times in what we (Walters et al. 1988) have called a "staircase" experimental design. Such designs require far more experimental units than would ordinarily be considered practical in field studies.

It might be acceptable to ignore time–treatment interactions in basic research experiments. By initiating all treatment and reference comparisons at a single moment, one accepts a risk that results may be due in part to the specific time one chose to start the experiment, so that others may have difficulty in reproducing them. In management settings, time–treatment interactions have a much more damaging effect, since they can be used by proponents of any management regime to explain away any apparent failures of that regime and hence delay (often at great cost) the implementation of other options.

A costly example of the danger of ignoring time–treatment interactions in management experiments has been in hatchery rearing of Pacific salmon in the Canadian Salmonid Enhancement Program (SEP). When this $300 million program was initiated in 1974, it was billed as an experimental program to increase salmonid populations through a variety of treatments ranging from habitat improvement to hatchery rearing. The program was divided into two phases, with the first phase (10 yr) to involve mainly pilot studies (experiments) with alternative treatments, and the second to involve implementation of the best treatments found in the first phase. Due to engineering and economic cost–benefit considerations, the bulk of the first phase investment went to a set of large hatcheries with an admirable commitment to monitor post-release survival rates through coded wire tagging of juveniles. At first, the hatcheries performed well, but later there was a massive drop in marine survival rates from several facilities, especially for chinook salmon, and low survival rates have persisted to the present. SEP biologists have contended that the hatchery stocks are not failing, but rather that all of the juvenile salmon have been adversely affected by warm water conditions that have prevailed in the North Pacific since the late 1970s. When confronted with evidence that wild ("control") salmon stocks have not shown consistent survival reductions in the same marine environment over the same period (Walters and Riddell 1986), SEP proponents have countered with an argument involving time–treatment interaction: treated (hatchery) populations may be more sensitive to marine conditions than control (wild) populations, due to some mechanism such as higher prevalence among hatchery fish of diseases whose expression is temperature dependent. This apology is a "there exists" hypothesis that cannot be rejected by detailed research showing that particular mechanisms (e.g., diseases) are not responsible for the decline. The only way to reject it is to start more hatcheries later in time, then demonstrate that they undergo the same transient decline. This experimental design would obviously require far longer ($\approx$25 yr) than the decade planned for experimentation in the SEP. As an aside, a few SEP hatcheries have come on-line during the late 1970s and early 1980s, and appear to be following the same transient of high to low survival as the earlier facilities.

Where to Invest Effort

Where replication is possible in principle, we see two main challenges in the design of large-scale management experiments. First, there is a critical need for research on imaginative ways to set priorities for investing in research, monitoring, and management. We must find substitutes for many of the cumbersome, time- and worker-intensive sampling methods used in most ecological field research, even if the substitutes involve substantial loss in sampling precision. Techniques such as satellite image analysis and digital particle counting should be seen as more than labor-saving

conveniences; they are central to the future of field experimentation.

Second, we need institutional arrangements that will permit and foster experimental studies that span time scales longer than the working lives of the scientists who initiate them. In government management agencies it is possible to bureaucratize experimental programs to the point where they may even outlive their usefulness. The more difficult challenge is to develop incentive systems that will encourage persistent involvement of researchers in the design and conduct of such programs.

Where replication is impossible and the severity of disturbance experiments is limited by risks of social and economic harm, resource managers will perhaps always operate in a twilight of uncertainty about the relative importance of their actions as opposed to the effects of uncontrolled environmental and ecological factors. By careful process research and modelling we may hope to narrow the range of credible hypotheses for the patterns that are seen (Holling 1988), and we can expect better statistical tools for deciding whether significant changes are occurring over time (Carpenter et al. 1989, Goldman et al. 1989). But we must not pretend that process research and diligent data analysis alone will provide answers that resource managers can trust.

A critical antecedent to the use of adaptive management experiments and to decisions of where to invest effort is a small number of credible hypotheses to explain the patterns perceived. It is a trivial task to define testable hypotheses, but it is not easy to generate hypotheses that are relevant to changes in the external context and internal structure of managed ecosystems. These changes are the real source of the surprises and crises that pace learning. By focusing on the causes of such abrupt and unexpected changes in behavior, however, it becomes possible to use modelling to evaluate existing process knowledge so as to screen the credible hypotheses, and identify where to concentrate scarce resources.

The regional impacts of spruce budworm (*Choristoneura fumiferana*) on forests of eastern North America provide a typical example. Prior to management, the insect periodically caused extensive mortality to balsam fir over large areas in eastern Canada and the New England states. In New Brunswick, the outbreaks occurred roughly every 40 yr, causing up to 80% mortality in mature balsam fir stands. Stands of young trees rarely experienced damage. Since the early 1950s, impacts have been intensively managed by insecticide spraying. Coincident with this program was a research project that still stands as a classic example of interdisciplinary analysis of a large-scale ecological system (Morris 1963), and provided the foundation for much subsequent methodological development (Holling 1978) and process studies (Holling and Buckingham 1976). However, there are still disagreements concerning the underlying causes of the time behavior (see Clark et al. 1979 vs. Royama 1984). In this example, as well as many others (e.g., fire management, Christensen et al. 1989), sudden surprises and unexpected behaviors challenged traditional myths of causation and management. During the 1960s, for example, outbreaks began in Newfoundland where the insect was historically rare. More recently, they have become more frequent in stands of young trees. Was knowledge deficient or had the system fundamentally changed because of expansion of the geographical scale of relevant interactions as a consequence of human activities?

The expansion of budworm outbreaks to young stands has been particularly surprising, and there are at least two possible explanations for it. One is the increased incidence of warm, dry springs, perhaps as a consequence of global warming. The second concerns changes in the densities of insectivorous birds. There are about 35 species, most of which migrate to overwintering areas in the subtropics and tropics where they may be impacted by deforestation. Is it possible that the migration of the birds connects impacts of deforestation in the tropics with unexpected outbreaks of insects in coniferous forests of North America? Both explanations are defensible, but are they credible enough to justify major changes in research and management investments?

The insectivorous bird hypothesis has been explored in detail (Holling 1988). The question had to be posed in a qualitative way; i.e., how much would bird populations have to be reduced in order to change the qualitative behavior of the system? The answer, based on existing budworm population models, was by more than two-thirds, and such a reduction would likely involve even more dramatic reductions in some bird species. It is highly unlikely that such a dramatic change would go undetected by the various professional and amateur bird censuses. Hence, it is not credible that declines in populations of migratory birds are the proximate and exclusive cause of outbreaks of budworm in stands of young trees.

The identification and evaluation of the set of possibilities discussed in this example depends upon qualitative analyses of a suite of models and key processes. Not one of those models covers all the range of scales that are relevant. The methods are not available to do so. There are detailed, spatially explicit models that simulate regional budworm–forest dynamics and deal explicitly with moth dispersal over a 70 000 km^2 area (Clark et al. 1979). Other models represent local sites (stands) and were designed to simplify optimization studies (Holling et al. 1986) or qualitative stability

analysis (Ludwig et al. 1978). Each of these scale-constrained models, however, together allowed the evaluation of cross-scale interactions by focusing additional knowledge of local dispersal of larvae, fragmentation of the landscape and hemispheric movements of birds and global changes in climate.

The range of defensible hypotheses in the budworm example is typical of many current issues. What is needed are techniques to concentrate on credible possibilities and structure their evaluation. A blend of scale-constrained models, scale-unconstrained process knowledge, good old-fashioned natural history and active adaptive management provides a fruitful direction for both the science and the management of regional renewable resource systems.

Conclusions

Two kinds of science influence renewable resource policy and management. One is a science of parts, e.g., analysis of specific biophysical processes that affect survival, growth, and dispersal of target variables. It emerges from traditions of experimental science where a narrow enough focus is chosen in order to develop data and critical tests that will reject invalid hypotheses. The goal is to narrow uncertainty to the point where acceptance of an argument among scientific peers is essentially unanimous. It is appropriately conservative, unambiguous, and incomplete. The other is a science of the integration of parts. It uses the results of the first, but identifies gaps, invents alternatives, and evaluates the integrated consequence against planned and unplanned interventions in the whole system that occurs in nature. Typically, alternative hypotheses are developed concerning the integrated properties of the whole to reveal the simple causation that often underlies the time and space dynamics of complex systems. Often there is more concern that a useful hypothesis will be rejected than a false one accepted; "don't throw out the baby with the bath water." Since uncertainty is high, the analysis of uncertainty becomes a topic in itself.

Renewable resource policy and management has its own profile for achieving degrees of integration and accepting degrees of uncertainty. The degree of match with the science is a measure of the usefulness of the science for decision and action.

Policy is politics. Design and acceptance of policy occurs when simple explanations of the causes of and solutions to problems of economic or social consequence achieve sufficient credibility in scientific, government, and public communities. Scientific uncertainty can be high so long as acceptability is high. Remember the phosphate debate of the '60s, when Lake Erie was "dying," that led to the Great Lakes Water Quality Agreement. Or think of the present science and politics of the Greenhouse Gas Effect. Both the science of parts (e.g., the role of phosphates in lakes) and the science of integration play roles (e.g., global circulation models). In both cases, decisions are not made because of a well-proofed argument in the tradition of experimental science, but because of the accumulation of credible evidence supporting a simple and widely perceived explanation in a political environment that demands action. Hence, resource policy decisions can be facilitated by explicit ways to identify alternatives, their likelihood and their outcomes in an environment that engages science, government, and the public.

When policies are defined, management begins and the same process of design and analysis occurs, but now in an environment where action has to be taken, however uncertain the outcome. That is where active adaptive management can play a central role, because its premise is that knowledge of the system we deal with is always incomplete. Not only is the science incomplete, the system itself is a moving target, evolving because of the impacts of management and the progressive expansion of the scale of human influences on the planet. Hence, the actions needed by management must be ones that achieve ever-changing understanding as well as the social goals desired. That is the heart of active experimentation at the scales appropriate to the question. Otherwise the pathologies of management are inevitable—increasingly fragile systems, myopic management, and social dependencies leading to crises (Holling 1986).

Literature Cited

Anderson, D. R. 1975. Optimal exploitation strategies for an animal population in a Markovian environment: a theory and an example. Ecology **56**:1281–1297.

Carpenter, S. R., T. M. Frost, D. Heisey, and T. K. Kratz. 1989. Randomized intervention analysis and the interpretation of whole-ecosystem experiments. Ecology **70**: 1142–1152.

Christensen, N. L., J. K. Agee, P. F. Brussard, J. Hughes, D. H. Knight, G. W. Minshall, J. M. Peek, S. J. Pyne, F. J. Swanson, J. W. Thomas, S. Wells, S. E. Williams, and H. A. Wright. 1989. Interpreting the Yellowstone fires of 1988. BioScience **10**:678–685.

Clark, W. C., D. D. Jones, and C. S. Holling. 1979. Lessons for ecological policy design: a case study of ecosystem management. Ecological Modelling **7**:1–53.

Collie, J. S., and C. J. Walters. 1987. Alternative recruitment models of Adams River sockeye salmon (*Oncorhynchus nerka*). Canadian Journal of Fisheries and Aquatic Sciences **44**:1551–1561.

DeGroot, M. H. 1970. Optimal statistical decisions. McGraw-Hill, New York, New York, USA.

Frederick, P. C., and M. W. Collopy. 1988. Reproductive ecology of wading birds in relation to water conditions in the Florida Everglades. Florida Cooperative Fish and Wildlife Research Unit, School of Forestry Research and Conservation, University of Florida, Technical Report Number **30**.

Goldman, C. R., A. Jassby, and T. Powell. 1989. Interannual fluctuations in primary production: meteorological forcing at two subalpine lakes. Limnology and Oceanography **34**: 310–323.

Holling, C. S., editor. 1978. Adaptive environmental assessment and management. John Wiley and Sons, London, England.

———. 1986. The resilience of terrestrial ecosystems; local surprise and global change. Pages 292–317 *in* W. C. Clark and R. E. Munn, editors. Sustainable development of the biosphere. Cambridge University Press, Cambridge, England.

———. 1988. Temperate forest insect outbreaks, tropical deforestation, and migratory birds. Memoirs of the Entomological Society of Canada **146**:21–32.

Holling, C. S., and S. Buckingham. 1976. A behavioral model of predator–prey functional responses. Behavioral Science **3**:183–195.

Holling, C. S., G. B. Dantzig, and C. Winkler. 1986. Determining optimal policies for ecosystems. Pages 453–473 *in* M. Kallio, A. E. Andersson, R. Seppala, and A. Morgan, editors. Systems analysis in forestry and forest industries. Volume 21. TIMS studies in the management science. North-Holland, Amsterdam, The Netherlands.

Keeney, R. L., and H. Raiffa. 1976. Decisions with multiple objectives. John Wiley, New York, New York, USA.

Kushlan, J. A. 1987. External threats and internal management: the hydrologic regulation of the Everglades, Florida, USA. Environmental Management **11**:109–119.

Kushlan, J. A., and P. C. Frohring. 1986. The history of the southern Florida Wood Stork population. Wilson Bulletin **98**:368–386.

Lee, K. N. 1989. The Columbia River Basin: experimenting with sustainability. Environment **31**(6):7–11, 30–33.

Lindley, D. V. 1985. Making decisions. John Wiley, New York, New York, USA.

Ludwig, D., D. D. Jones, and C. S. Holling. 1978. Qualitative analysis of insect outbreak systems: the spruce budworm and forest. Journal of Animal Ecology **44**:315–332.

Mangel, M. 1985. Decision and control in uncertain resource systems. Academic Press, New York, New York, USA.

Mangel, M., and C. W. Clark. 1988. Dynamic modelling in behavioral ecology. Princeton University Press, Princeton, New Jersey, USA.

Morris, R. F., editor. 1963. The dynamics of epidemic spruce budworm populations. Memoirs of the Entomological Society of Canada **31**.

Ogden, J. C. 1978. Recent population trends of colonial wading birds on the Atlantic and Gulf coastal plains. *In* A. Sprunt IV, J. C. Ogden, and S. Winkler, editors. Wading birds. National Audubon Society Research Report **7**:135–153.

Raiffa, H. 1968. Decision analysis: introductory lectures on choices under uncertainty. Addison-Wesley, Reading, Pennsylvania, USA.

Royama, T. 1984. Population dynamics of the spruce budworm *Choristoneura fumiferana*. Ecological Monographs **54**:429–462.

Walters, C. J. 1981. Optimum escapements in the face of alternative recruitment hypotheses. Canadian Journal of Fisheries and Aquatic Sciences **38**:704–710.

———. 1986. Adaptive management of renewable resources. McGraw-Hill, New York, New York, USA.

———. 1990. Value of short term recruitment forecasts for harvest management. Canadian Journal of Fisheries and Aquatic Sciences **47**, *in press*.

Walters, C. J., and J. S. Collie. 1988. Is research on environmental effects on recruitment worthwhile? Canadian Journal of Fisheries and Aquatic Sciences **45**:1848–1854.

Walters, C. J., J. S. Collie, and T. Webb. 1988. Experimental designs for estimating transient responses to management disturbances. Canadian Journal of Fisheries and Aquatic Sciences **45**:530–538.

Walters, C. J., J. S. Collie, and T. Webb. 1989. Experimental designs for estimating transient responses to habitat alterations: is it practical to control for environmental changes? Canadian Special Publication of Fisheries and Aquatic Sciences, Habitat Symposium, *in press*.

Walters, C. J., and B. Riddell. 1986. Multiple objectives in salmon management: the chinook sport fishery in the Strait of Georgia, British Columbia. Northwest Environmental Journal **2**:1–15.