Economic Valuation of Aquatic Ecosystems

Anthony Fisher, Michael Hanemann,

John Harte, Alexander Horne,

Gregory Ellis, and David Von Hippel

University of California, Berkeley

October 1986

Table of Contents

CHAPTER 1: INTRODUCTION AND OVERVIEW

In this chapter we indicate the ways in which aquatic ecosystems are valuable to mankind, and make a first pass at suggesting how these values might be assessed. Our object is to give an adequate appreciation of the many and varied kinds of goods and services provided by aquatic ecosystems, while at the same time beginning the process of organizing the discussion of methods of measurement of the worth of these benefits. The chapter concludes with a detailed outline of the plan of the rest of the study.

A. Goods and Services Provided by Aquatic Ecosystems

The steps involved in determining the economic value of ecological goods and services are to identify what benefits ecosystems provide for mankind, to characterize these benefits in ecological terms, and then to assess their economic value. Even the first step should not be thought of as completed for any actual ecosystem. Indeed, it is virtually certain that as our understanding of ecosystems progresses in the future, we will discover the existence of presently unrecognized goods and services provided by healthy ecosystems. The characterization of goods and services by ecologists must include not only a description of the nature of the good or service, such as how many trout for sports fishing a particular stream maintains, but also how the continuing provision of that benefit is linked to the future state of health of the ecosystem. Generally the ability of ecologists to characterize the magnitude of the benefit under ambient circumstances far exceeds their ability to assess how continuing provision is linked to environmental quality. Finally, valuation must take into account not only the effect of a change in environmental quality on the ability of an ecosystem to provide the benefit under discussion, but also its effect on

the overall health of the ecosystem, which in turn may influence the future ability of the system to provide benefits not presently identified. This "insurance" factor is most difficult of all to include in the benefit-cost calculus because it requires having to guess the value and the ecological interconnectedness of benefits that we have not even identified as of yet.

In order to guide our thinking about methods of measuring benefits we have chosen to categorize the goods and services provided by aquatic ecosystems as being those for which the environment is an input, that is, the ecosystem provides a factor or means in the production of a good or service to be consumed, and those for which the environment itself is a final good. This distinction is,in a sense, artificial, since many goods and services provided by aquatic ecosystems fall in both categories. It will, however, be useful because, as explained in section B below and further in chapter 7, it corresponds in some ways to a distinction between approaches to economic valuation.

Goods and Services for which Aquatic Ecosystems Provide Inputs to the Production Process

The most obvious set of goods for which aquatic ecosystems provide basic inputs are "fisheries" products. These products, as indicated in Table 1, include harvested fish, shellfish, and crustaceans; aquatic plants such as kelp, which is used in the manufacture of chemicals and food products; and, to a small extent, aquatic mammals, now used mostly for garments. The rivers and reservoirs that allow hydroelectric production and its control contain aquatic ecosystems. Some types of damage to these ecosystems, e.g. siltation of reservoirs caused by soil erosion and runoff, can affect the output of the hydroelectric system. Rivers, lakes, bays, and estuaries are

TABLE 1: GOODS AND SERVICES PROVIDED BY AQUATIC ECOSYSTEMS

Goods and Services for which the Environment Provides Inputs

Fisheries Products: Fish, Shellfish, Crustacea, Kelp, Aquatic Mammals

Hydroelectric Power

Transportation

Treatment of Human Wastes

Treatment of Industrial Wastes

Water Purification

Drinking Water Storage

Information Produced via Scientific Research

Goods and Services for which the Ecosystem is a Final Good

Recreational Use of Aquatic Areas (Public Access and Commercial)

> Direct Use of Water: Boating, Rafting, Sailing, Canoeing, Scuba-diving, Swimming, Wading

> Recreational Use of Aquatic Organisms: Fishing, Waterfowl Hunting, Collection of Shellfish and Crustacea

> Waterfront Recreational Activities: Strolling, Hiking, Sunbathing, Team Sports (e.g. Volleyball), Off-Road Vehicle Use, Horseback Riding, Nature Study (e.g. Birdwatching)

Amenities

> Scenic Values

> Modulation of Local Climates by Large Bodies of Water

> Status and Enjoyment of Owning or Having Access to Aquatic Areas

> Informal Education of Children

> Psychological Benefit of Availabilty of Pristine Areas

Future Goods and Services

> Preservation of Genetic Information: Protection of endangered Species, Preservation of Gene Pool

> Preservation of Wild Areas for Use by Future Generations and for Future High-Value Development

3

also used as transportation arteries, and thus provide an input to the process of moving people and goods from place to place.

An extremely important and often overlooked set of processes in which aquatic ecosystems play roles are human and industrial waste-treatment and water purification. When human wastes are discharged into bodies of water, biological and physical processes combine to break down organic matter and release nutrients in the wastes, and to kill pathogenic organisms. In a similar manner many industrial wastes are broken down when disposed of in aquatic environments. Coupled with these waste-treatment functions, wastewaters disposed of in lakes, rivers, marshes, and other aquatic areas are purified and recycled either by evaporation and subsequent precipitation or by percolation through benthic (bottom) sediments and soil to groundwater aquifers. Wastewater added to a lake might undergo biological treatment by aerobic (oxygen-using) bacteria associated with oxygen-producing algae growing at the water's surface, chemical treatment by entrapment of metals and other substances in the anaerobic (oxygen-free) bottom waters and sediments, and physical treatment by filtering through sediments and soils before it reaches a subterranean aquifer that supplies fresh water to consumers. Properly functioning aquatic ecosystems in reservoirs also provide appropriate conditions for the storage of drinking water. Clean and/or potable water is an essential input to the production of a vast number of products and services.

Aquatic environments also provide opportunities for scientific research and development. In this case knowledge is the product for which the environment is an input. This knowledge may take the form of information about the improved cultivation of a valuable organism, for example, or data that enables prediction of the behavior of other aquatic ecosystems, and how the goods and services that they provide will vary under changing

4

conditions.   The study of one small lake, for example, might provide

information valuable in protecting a number of lakes in an area from acid

rain or some other pollutant stress.

Uses of Aquatic Ecosystems in which the Environment is the "Final Good"

Perhaps the most obvious set of goods and services in which aquatic

ecosystems are in a sense final goods are the recreational uses of watery

areas.   These recreational goods, as listed in Table 1, include direct uses

of water, the recreational pursuit and harvest of aquatic organisms, and

waterfront recreational activities.   Examples of activites involving the

direct use of water are boating, rafting, sailing, canoeing, scuba-diving,

swimming, and wading.   Fishing, hunting of waterfowl, and collection of

shellfish and crustacea are examples of the recreational use of aquatic

organisms.   Waterfront recreational activities include strolling, hiking,

sunbathing, sports such as volleyball, the use of off-road vehicles,

horseback riding, and nature study (e.g. birdwatching).   Many of the

recreational goods mentioned above are available in both public areas and

through commercial interests such as tourist hotels and lodges close to the

water, tour boats, and fishing and other guide services.   Virtually all of

these goods and services depend on good water quality for their value.

A much more amorphous class of benefits provided by aquatic ecosystems

can be loosely described as "amenities".   These include the pure scenic

value of a waterfront area or lake, the modulation of local climates by

large bodies of water, and the status and enjoyment provided by owning or

having access to areas near the water.   While the practical nature of these

amenities is clear to everyone, there is a "spiritual" side to the scenic

value of aquatic ecosystems that may represent the dominant benefit that

these ecosystems provide.   In the informal education of many children, nature plays an extremely important role.   From the autobiographies of numerous writers, artists, scientists, and others we read often of how early exposure of pristine wildlands shaped these peoples' minds beneficially. Such writings reveal the awareness of ecosystem benefits by those that are most able to express these experiences vividly, but these same benefits accrue, of course, to a far wider spectrum of people who are not necessarily as conscious of, or articulate about, their existence.

Beyond the formative years of childhood, amenity values continue to enrich peoples' lives, but in ways that can be distictly different from the ways in which children benefit.   In particular, a greater awareness of the amenities occurs as we mature and the experience of nature becomes less formative than it is restorative.   The person in an office in downtown San Francisco, for example, may take comfort in the fact that pristine areas are available for him or her to enjoy.   This thought, that escape from the "rat race" is possible, may make it easier to live and work happily in a city. If such a person were asked what this amenity was worth, he or she might quote some figure, but it is possible, since the scenic area has always been available, that the individual would undervalue this amenity relative to what would be considered his or her "share" of the value of the scenic area to society as a whole.

A final class of goods and services provided by aquatic ecosystems can be loosely described as future goods and services, and the preservation thereof.   This includes the preservation of diverse genetic information, the preservation of ecosystems for future generations of humans to enjoy, and the preservation of aquatic areas for future development.   The protection of endangered species--for their future commercial use, aesthetic

6

value, use as objects for scientific study, and existence value--is one example in which the preservation of genetic information can provide future goods and services for society. In preserving a diversity of plants and animals we are also preserving a library of genes that, with man's growing ability to manipulate genomes, may someday become tools useful in producing valuable drugs or chemicals. The preservation of scenic and wild areas for future generations to use--our National Parks are examples--provides future goods and services in the form of both recreational opportunities and aesthetic values, as described above. The knowledge that scenic areas will be available to their descendants in the future may also provide the benefit of peace of mind to a person living today. Finally, preservation of some aquatic areas may allow them to be developed for high-value uses in the future. Mining in a scenic lake area rich in some ore, for example, might have to be done today in such a way that the scenic value of the place is lost indefinitely--through poisoning of the aquatic ecosystem by acids leached from mine tailings, soil erosion from road constuction, or physical rearrangement of the area--but it might be possible to mine the same region at some future time, using an as-yet undeveloped technology, in such a way that the aesthetic value of the area remains intact. In the latter case the area continues to provide recreational and aesthetic goods and services in addition to the valuable ore. As described in chapter 5 the presence of future-worth considerations can greatly influence regulatory choices regarding the control of pollution of aquatic ecosystems.

B.   Economic Valuation

As the discussion in the first part of this chapter has suggested, the goods and services that can be provided by aquatic ecosystems are many and

varied. Yet for the purpose of characterizing evaluation, we must try to collect them into a manageable number of categories, corresponding to methods of evaluation. This we attempt in the table, Table 2, below, with the hope that no major items, at least, are lost in the process. Types of goods and services are classified into those involving the aquatic ecosystem, the environment, as input, and those involving it as a final good (or service). By environment as input, we mean that it enters into a kind of mixed biological-economic production function, along with conventional inputs such as labor and capital, to yield some desired final good--as the table suggests, a supply of fresh water for drinking, perhaps, or a shellfish harvest. The consumer of the water, or the shellfish, is assumed to care only about the good he consumes, and not the input mix used to produce it. By contrast, when the environment is valued as a final good, it enters directly into the consumer's utility function. Thus improved water quality can yield benefits both as an input to some production process, and directly to on-site recreationists, nearby property owners, and so on.

A couple of more exotic, or less tangible, goods are also indicated in the table. One is the conservation of genetic information. This can be considered as affecting <u>future</u> commercial harvesting, for example of a plant or animal species for some yet-to-be-discovered medicinal property. The other intangible good is the existence of an unspoiled environment, unrelated to any use or consumption of its resources now or in the future. Some people derive satisfaction simply from the knowledge of existence, and this has been termed "existence value" in the literature of environmental economics.

Now, why is it sensible to classify the goods and services provided by aquatic ecosystems in this fashion? Consider the first column in the table, headed "method of evaluation." It is our view that a particular method can

TABLE 2:

METHODS OF VALUATION FOR GOODS AND SERVICES

PROVIDED BY AQUATIC ECOSYSTEMS

Method of Valuation	Type of Benefit
	Environment as Input
Shifting Supply, Given Demand	Water Supply and Quality
	Commercial Harvesting (includes genetic conservation for future harvest)
	Environment as Final Good
Travel Cost	Recreation
Comparative Property Values	Amenities
Contingent Valuation	Existence Value

be identified as best suited to each of the categories. Thus, if the environment is viewed as an input to a production process, such as the commercial harvesting of shellfish, an improvement in quality due to reduced pollution loadings can be expected to lead to a shift (down and to the right, on a conventional diagram) in the cost or supply of shellfish. Given an independent estimate of the demand for the particular shellfish product, the shift in supply generates an increase in combined consumer and producer surplus, the area bounded by the demand and supply curves. Of course, establishing the nature of the connection between reduced pollution and the supply shift is a difficult empirical problem. In section A of chapter 7 below we consider the problem in some detail, and illustrate our method of solution with some computations based on estimates of relevant demand and supply parameters in the literature. The use of a change in combiend surplus to capture the welfare effect of reduced pollution is justified in chapter 6, a theoretical discussion of the economic concept of benefit.

An aquatic ecosystem can also, as we have noted, be viewed as an input to the generation of fresh water supplies in a region. Reducing pollution loadings in the system similarly results in a downward shift in the cost or supply of providing fresh water. We shall have more to say about this contribution also in chapter 7.

Turning to the environment as final good, the first item in our table is recreation. There is a large literature on methods of valuing outdoor recreation resources, discussed in some detail in section B of chapter 7. Here we just note that the preferred method, rooted in economic theory and validated in many empirical applications, is the travel cost method. The name is derived from the use of travel cost (from the point of visitor origin to the recreation destination) as the measure of price in an analysis of the demand for recreation at the site in question. Thus our focus has

shifted from supply to demand. There is however an interesting parallel to the analysis of the environment as input. Suppose an improvement in water quality makes available a site that can be assumed to perfectly substitute for another (in the provision of recreation). Then recreation at the first or unimproved site is _in effect_ available at lower cost, to those who live nearer the newly available site. Of course, this analytical device requires the assumption that the newly available site provide the same recreation services as the other, so that consumers are indifferent as to which is chosen as "input."

Reducing pollution in an aquatic ecosystem can also lead to enhanced amenities. Clean water makes nearby residential property more desirable. An extensive literature has explored the relationship between changes in environmental amenities and property values--the extent to which it exists, the circumstances under which it can be estimated, its magnitude in particular cases, and so on. This literature is reviewed in section B of chapter 7.

We come, finally, to existence value. This differs in an important way from all of the other goods, or benefits, discussed thus far in that it is not associated with _use_ of the resources of an ecosystem. In fact it is often classified, along with option value, as an "intrinsic", or non-use benefit of preserving or improving an ecosystem. We shall have more to say about option value very shortly. With respect to existence value, there is a double problem for measurement. First, one cannot measure units of consumption (to which a value might then be imputed). To some extent this is true also for amenities--as in the case of an improved view. But the value of the view may be captured by a change in property value, since the view is associated with a piece of property, and property is valued in

market transactions.

The second difficulty in measuring existence value is that it is a pure public good, and one whose consumption is not associated with consumption of some private good such as residential property. About the only approach that can be employed here--and has been, in a small number of empirical studies--is so-called contingent valuation. This is simply asking individuals what they would be willing to pay for the continued existence of an area or species. The literature has also addressed the difficulties with this approach--the hypothetical nature of the question, its unfamiliarity to respondents, their propensity for strategic behavior, and so on. We provide a review with special reference to the application to aquatic ecosystems in section C of chapter 7.

We mentioned option value as the other commonly identified non-use environmental benefit. Yet it appears nowhere in our table. The reason is that, in our judgment, it is not a separate benefit, corresponding to a separate good or service provided by an aquatic ecosystem. It is instead an adjustment, or "correction factor," to an estimate of any of the other kinds of benefits listed in the table, to take account of uncertainty about their future values. This is a complex issue, however, that has generated considerable confusion and controversy in the literature. Chapter 5 defines option value and some of its properties in an analysis of the valuation of pollution control in a dynamic, uncertain setting. Further discussion, focusing on different concepts of option value, is provided in chapter 6.

C. **Plan of the Study**

In the next chapter we discuss a kind of "quick and dirty" alternate approach to valuation, the construction of a suite of indicator variables (SIV) that might be used to characterize the response of an aquatic ecosystem to reduced pollution or other disruption. This chapter includes a review of what might be termed ecological scoring methods, such as the HEP and HES systems. It also introduces concepts which will be useful later on.

Chapter 3 is about one of these: the dynamics of ecosystem recovery. A model is developed that generates the often-observed and potentially important hysteresis phenomenon, in which a recovering ecosystem does not retrace the path of its decline. The point of the model is to enable prediction of the recovery behavior of ecosystem populations in which we are primarily interested, higher trophic levels such as fish, from that of the much more readily observed lower trophic levels such as phytoplankton. Chapter 4 is an analysis of error propagation in measuring recovery. That is, suppose we are uncertain about the degree of phytoplankton recovery. How does this translate into uncertainty about recovery of the fish population?

Chapters 2, 3 and 4 are primarily about the behavior of aquatic ecosystems, with no systematic discussion of economic valuation. In chapter 5 we begin this discussion. A model is developed to value the control of pollution, taking account of key features of the ecosystem behavior discussed in the earlier chapters: recovery lags, irreversibilities, and uncertainty. The model does not address the question of how to estimate the different categories of benefits identified in the preceding section (of the Introduction). This is the task of chapter 7, divided into three parts, also noted in the preceding section: the environment as input (water supply, commercial harvest), the environment as final good (recreation,

amenities), and non-use benefits (existence value). The discussion of methods of benefit estimation is preceded, in chapter 6, by a theoretical analysis of the economic concept of benefit. Specifically, we motivate use of combined consumer and producer surplus as the preferred measure of a welfare change following an environmental improvement.

In chapter 8 we consider appropriate directions for further work. Our present intention is to proceed in two areas: (1) comparative analysis of models for policy evaluation, and (2) development of a case study. Both are elaborated in chapter 8.

Chapter 2.   A SUITE OF INDICATOR VARIABLES (SIV) INDEX FOR AN AQUATIC
              ECOSYSTEM

I.   The Need for a SIV-Index

        Assessment of the damage to ecosystems ideally requires

an accurate and precise measurement of the harmful effects. The

results of such measurements are needed to establish a numerical

relationship between pollution and economic damage to the

ecosystem. Although not often used exactly in this way there are

several habitat evaluation procedures available to assess the

"health or state" of the ecosystem. These measures include

several separate procedures (see reviews by U.S. Water Research

Council, 1981; Putnam, Hayes, and Bartless, 1983; Canter, 1984)

and cover most types of aquatic ecosystem but focus on streams

and wetlands rather than large lakes, reservoirs, large rivers,

estuaries or the open ocean. None of these indices is ideal but

they have served well in some circumstances, especially for

evaluation of game habitat used for recreational sport, for

example, deer hunting.

        Any of these evaluation systems can be used to give a

numerical value for the ecosystem over a sustained period of

time. The resulting long-term data base is then used to show if

a decline or improvement has occurred. When compared with an unaffected or control ecosystem an ecosystem value can be expressed as a percentage of the optimum even if the evaluation procedure does not cover all the period of degradation (or improvement) of the system.

Of considerable practical interest is the need for the maintenance of a complete habitat in the kind of restoration that occurs when sewage or other wastewaters are cleaned up. For example, a relatively simple single parameter (e.g., the fish of concern) or multiple parameters (e.g., the index proposed in this paper) can be assessed routinely while habitat evaluations are extensive, expensive, and one-time measurements.

II. The Requirements for an ideal index: Selection of variables for use in a SIV index

There are three main requirements:

oData must be inexpensive to collect.

oData must already be available for some ecosystems for use in trial projects.

oThe connection between the variable and its biological effect must be known from experimental studies.

The purpose of a SIV index is to determine aquatic ecosystem health over time and/or space. The choice of variables can change depending on the ecosystem chosen. For example, dissolved oxygen fluctuations can be deadly in mid-western rivers

2.

in summer but the same quantity of waste is unlikely to trouble the temperate open ocean. Since biologically non-functional variables decrease the precision of any index they should only be used where important.

III. <u>Review and critique of ecological indexes which could be used to estimate ecosystem health</u>.

<u>Critique of existing habitat and other evaluation procedures as applied to aquatic ecosystems</u>

Existing habitat evaluation methods usually focus on

o the physical structure of the ecosystem -- e.g., stream sinuosity, mean depth, percentage of cover, size of the lake

o indigenous, rare, or sensitive species, diverse species composition, and

o maintenance of indigenous (native) sport or game species.

The habitat evaluation procedures are derived from common sense evaluations once made by wildlife managers. The purpose was usually to decide what mitigation should be given if an area was to be physically destroyed -- as for example if a housing development or a dam were to be built in the area. In many cases mitigation was the creation, donation, or restoration of a piece of land which was of comparable ecological worth to

that being destroyed. An example might be the degradation of a stream by treated sewage could be compensated for by the creation of a marshland on the treatment plant property.

The evaluation procedures have a terrestrial bent (e.g., deer, partridge) since lakes and streams cover only a small portion of the landscape. Thus physical features such as trees, browse, overhanging banks (for fish), are important, even dominant in existing habitat evaluations -- and rightly so for terrestrial and some aquatic systems.

However, most lakes, oceans, estuaries, larger streams, and rivers are structured on the basis of thermal stratification, the chemical stratification which follows, and an ever-changing biotic structure. Wetlands are intermediate in this respect depending on the degree of submersion and the life times of the plants which constitute the base of the food chain.

Pollution in aquatic systems alters the biotic structure, sometimes the overall chemical structure, but rarely the thermal or physical structure of aquatic ecosystems. In this it differs from terrestrial habital destruction. The rebuilding of a damaged landscape requires the regrowth of a complex of physical habitats, while the restoration of an aquatic one may in principle require only the cessation of pollution. In both cases it is assumed that the biotic component is readily available to migrate in from adjacent areas.

Most of the indices, especially the habitat evaluation

4.

procedure (HEP), the habitat evaluation system (HES), and the ecosystem scoping method (ESM), also incorporate an implicit (HEP, HES) or explicit (ESM) belief that diversity = stability = desirability. That is, the more different types of organisms there are (or the more links there are in the food web) the higher the ecosystem will score. Thus the most valuable ecosystems tend to be the most diverse by this rationale.

The diversity-stability argument has a 20 year history in ecology. One might sum up the conclusion as the relationship between diversity and stability depends on the definition of stability and the time scale of observation. For example, if stability is equated with constancy over time then, when using typical northern temperate human time scales of years the simple non-diverse arctic owl-lemming-grass food chain appears unstable. When viewed over decades the opposite conclusion can be drawn (i.e., a perpetually oscillating population). Other definitions of stability can lead to yet other relationships with diversity which are not discussed here. It is unfortunate that the early discoveries of high diversity in tropical forests and coral reefs were not put in a better perspective for seasonally-controlled temperate-polar systems.

The intent of this paper is to review in brief existing habitat evaluation procedures and attempt to derive a specifically aquatic index which can be used to describe the "health" of the ecosystem. Such an index will be imperfect but

is needed if one is to assess change over time, and thus see effects such as hysteresis (e.g., Edmondson and Litt, 1982, also see later in this report), improvement, degradation and ascribe some economic value to the measured changes.

The choice of variables for an aquatic health index can, in theory, be made from any or all trophic levels in the ecosystem. Unfortunately the organisms of most direct economic interest (recreational or sports fish and shellfish) do not seem to be either easy or inexpensive to sample or to use for robust indices. Because of their size, relative rarity and biological complexity fish and shellfish produce variables which vary widely from the mean value. These parameters have a high coefficient of variation and when combined into any index these large errors propagate to the point of rendering the index useless for practical purposes.

An example of this is the "scope for growth" (SFG) index which has been widely proposed for the assessment of the health of fish and shellfish. In a recent (1983-1984) and costly study of the effects of the large sewage effluents of Los Angeles, the California Dept. of Fish and Game (Monterey Office), together with the local discharger and various other regulatory agencies use the SFG method. Analysis of this data shows that changes shown near outfalls using the "scope for growth" (SFG) method are not statistically significant. Both increases and decreases in SFG relative to controls occurred at outfalls but similar changes

6.

occurred between replicated samples in the same place. The SFG
method has a poor and inexplicably variable precision relative to
other methods of growth measurement. SFG can only resolve
changes of 282% (average of all Cal-COMP data) while simple
measurement of length or weight have uniform precision and can
resolve differences of 4% length and 14% weight (Horne, 1984).

If we are to detect the biological effects of pollution
near outfalls, a more precise measurement of mussel growth must
be used to replace scope for growth tests. Such a precise method
has been developed for Region #2 (San Francisco Bay Regional
Water Quality Control Board) by scientists at the University of
California, Berkeley. It is clear that SFG is still very much at
the research stage and not a monitoring tool.

Why is the Scope for Growth Test so Imprecise?

The reasons are both physiological and statistical and
both are inevitable. The physiological reason is that it is
common for organisms moved from field to laboratory to experience
long-term stress (see Knight and Foe, report to RWQCB, 1984).
This together with individual genetic variation gives a highly
variable end result.

An implicit assumption in this method is that SFG
represents an obsolute measure of mussel health. For example, it
is assumed that "healthy mussels" are always of approximately 40
joules h^{-1}. Values measured on mussels transplanted to other

7.

sites are often much lower than this due to transplant effects alone. In addition, spawning stress will reduce growth. These stresses and other uncontrolled variables reduce the utility of SFG as a monitoring tool to almost zero. The statistical reason for the low precision of scope for growth results from SFG being a value calculated from a series of ratios and assumed values. Errors propagate through such an equation. It is usually better to measure a biologically integrated change directly -- i.e., measure growth directly rather than indirectly.

The problem of high variance is apparently inherent in these higher trophic level indices. Even relatively simple values, such as the percentage survival of animals exposed to an environmental pollutant, can be variable since animals which appear identical in size, condition, and amount of pollutant absorbed may have a very different genetic makeup (Hilvsum, 1983; Horne and Roth, in prep.)

There are two ways to overcome this dilemma. First, simpler, more abundant organisms can be used to construct a robust index. Second, functional components of one or more groups of organisms can be used instead of their abundance.

IV. Proposed Suite of Indicator Variables (SIV) index:
strengths, weaknesses.

Lacking any absolute ideal indicator(s) for ecosystem
health an index is an obvious second choice. This has a history
in economics (price index) and in ecology (diversity index,
striped bass index). Again in common with economics (consumer
price index) but not usual in biology, an index with several
components seems desirable. The problem with an index based on
any one variable in ecology is two-fold, lack of robustness and
risk of being misleading. Over the last century several single
indices have been proposed as "master variable". Acidity (pH)
has often been proposed (Schindler et al., 1985) but is
misleading for acid rain studies and alkalinity has been
substituted (Hendriksen, 1979). While alkalinity is an
appropriate guide to the susceptibility of a lake to acid
oligotrophication (acid-induced impoverishment) it is not a good
indicator of the effects of point or non-source wastewater
pollution.

What is required is a suite of independent variables
which would, if taken together, reliably show the current state
of the ecosystem. Only if the majority of variables indicate a
change in the same direction will there be good probability that
the damage is serious (ecologically important) and persistent.
It should be noted that this majority indicator approachimplies
that the "cost" of a false warning is greater or equal to the

"cost" of a false assumption that all is well. This could be
described in terms of the "crying wolf" paradox. It is not usual
to consider the damage done by false warnings of severe damage.
However, from an ecological viewpoint there is only a certain
amount of public concern for ecosystem preservation. Thus false
warnings can detract from the effort required to respond to true
warnings. An example of this is the hue and cry over DDT and its
environmental effects. The cancers and genetic damage now
ascribed to PCB are not effects of DDT. Although there are
serious effects and a ban on DDT use was appropriate the toxicant
PCB was overlooked for many years since its chromatographic
signature was confused with DDT. A decade was lost when PCB-
filled devices could have been phase out.

The plankton

Large numbers of independent (i.e., physically
unconnected) organisms can be sampled with low statistical
variance. For example, counting 100 single-celled free-floating
phytoplankton gives 95% confidence limits of being within $+20\%$ of
the true number (100). It is not always easy to be sure one has
overlooked some algae when examining lots of similar-looking
cells. If a similar number of cells were counted but were
contained in 16 filaments the 95% confidence limit would only be
$+50\%$ -- a much larger error (Land, Kipling, and Le Cren, 1959,
pg. 158). In addition to counting errors if the organisms are

10.

also physically well-mixed then genetic variation between individuals is muted. These conditions are best met in the aquatic ecosystem in that group of organisms called the plankton. The word plankton means wanderer and basically refers to those small plants and animals which are more or less at the mercy of water currents. In this paper I will use the term in its widest extent to cover small unattached organisms in ponds, lakes, streams, rivers, estuaries, oceans and coastal fringes including salt marshes. Thus true animal plankton (zoo-plankton), plant plankton, (phyto-plankton) as well as the invertebrate insect drift in rivers and streams is encompassed.

As defined widely plankton includes the young stages of almost all the commercially valuable fish and shellfish and most of the sport fish and shellfish. Those which are not included depend heavily on the plankton for food in the adult stages. For not fully understood evolutionary reasons the majority of large valuable fish and almost all shellfish need a planktonic life-stage and some such as salmon, dungeness crabs, grey mullet or eels swim or crawl thousands of miles to achieve this planktonic goal.

The functional components of aquatic ecosystems

The previously mentioned high variance (= high risk of incorrect predictions) was first recognized in the study of stream benthos (e.g., Wurtz, 1960). Here extreme patchiness

11.

(large rock adjacent to gravel, sand etc.) could only be overcome by very large numbers of replicate samples. Typically 73 replicate collections in a stream riffle might be needed for 95% confidence in the numbers of invertebrates collected relative to 3-6 replicates which are normally the limit (Needham and Usinger, 1956). This patchiness was later found to be common in most aquatic ecosystems and remains a partially solved problem (Richerson, et al., 1970; Riley, 1976; Sandusky and Horne, 1978).

In addition, particularly in streams, wetlands, and estuarine-ocean systems the identification of individual organisms is often impossible. The animals in the above-mentioned ecosystems are numerically dominated by juvenile stage of such groups as clams, oysters, polychaetes, insects, fish and crabs. The taxonomic keys for juveniles in many cases have not yet been written and even when published require expert taxonomists. This problem was again first tackled by stream ecologists who proposed to simplify their ecosystem by using functional group classification instead of taxonomic identification. Thus shredders, scrapers, filterers replaced large crayfish, caddis-flies, and may-flies even though the functional classification cut across traditional taxonomic lines.

In smaller ecosystems such as ponds and small streams it has been possible to measure whole-ecosystems variables such as net photosynthesis or respiration using whole-lake oxygen fluxes, isotope dispersion, or even carbon depletion. The process has

12.

provisionally been extended to incorporate large lakes (Talling, 1976).

Functional components have the advantage of built-in robustness since they incorporate ecosystem homeostasis as explained below (i.e., the intertia and redundancy in ecosystems which tend to reduce overall change). A typical example of this would be the replacement of the attached stream algae <u>Cladophora</u> by the attached stream algae <u>Tabellaria</u> near the inflow of a well-treated but nutrient-rich domestic sewage outflow in the Truckee River, near Lake Tahoe (Horne et al., 1978). Insect and presumably fish populations did not respond to this food chain switching presumably because either algae was equally acceptable (or unacceptable) as food.

Combined plankton-functional component index -- the SIV index

For purposes of monitoring the ecosystem effect of pollutants a combination of both the plankton and functonal components will be valuable. Large numbers of individuals (n) can be measured which will reduce type II errors and concomittant failure to detect pollution's effects until it is too late. A large n will also reduce type I errors and risk of overstating an effect. The use of juvenile stages of commercially and recreationally important fish and shellfish will assist in the economic analysis and will also include "sensitive" species (sensu EPA guidelines on NPDES permits). Both indigenous and

rare species can be accommodated in such an index. Finally, the
robustness of the index will be ensured by incorporation of
ecosystem homeostasis by the use of functional component
variables.

The drawbacks to the SIV index in principle are similar
to those of any other environmental scoping or health assessment
namely:

o   Require some measurement or knowledge of the
     ecosystem.

o   Is hard to extrapolate backwards in time to pre- or
     low-pollution eras.

o   May miss important effects if one component of the
     index was capable of indicating serious harm but the
     other components lagged behind in their responses.

The proposed SIV index has the advantage for aquatic
ecosystem pollution studies that these drawbacks can be minimized
particularly in the cost of data collection since the precision
of the index can be very high.

The main purpose of any index is to show changes over
time or space. High precision is vital if change is to be
detected in time for restorative measures to be put into effect.

The literature shows a number of multi-parameter indices
or ranking systems used to measure the "trophic state" of lakes
(i.e., their basic fertility of productivity). These include

14.

those by Lueschow et al, 1970; Shannon and Brezonik, 1972;
McColl, 1972; Michalski and Conroy, 1972; Sheldon, 1972;
Uttormark and Wall, 1975; Carlson, 1977; and the EPA's own
modified index derived from an extensive study of 757 specially
selected lakes (See Hern, Lambou, Williams, and Taylor, 1981).
The SIV index does not attempt to improve these models especially
those by Carlson 1977 and Hern (EPA) et al., 1981. Our purpose
is to extend their use to cover both toxic and biostimulatory
effects of point and non-point wastewater discharges as well as
extend coverage beyond lakes to all aquatic ecosystems.

For example, one improvement of the model suggested by
EPA (Hern et al., 1981) to use chlorophyll _a_ not nutrient levels
as a basis for trophic classification fits directly into the
functional component mechanism of the SIV index.

Multi-parameter indices also exist which attempt to
measure higher trophic level productivity including that of
fish. This is a measure of ecosystem "health". Such attempts
range from pioneering concepts such as those of Thienemann (1927)
and Rawson (1951) to complex but realistic simulation models
(e.g., Steele, 1974; Powell, in press). A "rough indicator of
edaphic (= nutrient) conditions" combined with lake bathymetry
(morphological structure) was the morphoedaphic index (MEI) of
Rawson (1955) and Ryder (1965) and Ryder et al. (1974).

The MEI uses mean depth and fish harvest statistics and
was designed for use in lakes. Since the most productive systems

15.

(lakes, streams, estuaries and coastal ocean waters) are shallow
and well-stirred this index has limited use when expanded from
typical thermally stratified lakes to all aquatic ecosystems.

Complex simulation models of the planktonic community
are not yet usable as indices even though multiple parameters are
involved. A primary reason is that such models are not normally
designed to work with the kind of pollution stress normally
imposed by toxic wastewater. Typically, the models will be
perterbed by nutrients or the introduction of a natural change
such as increased predation. Most chemical poisoning or aquatic
habitat structural alteration has few natural analogs and these
are yet little studied. The few potential analogous systems
natural springs with high acidity as toxic metals have been
little studied for metal toxicity dynamics. Almost no examples
of organic biocide accumulation are available in natural aquatic
ecosystems. However, metal or organic toxicants are a prime
cause of aquatic ecosystem degredation, second only to dissolved
oxygen reduction and diversion of water.

The construction of a numerical SIV index with some of
the properties mentioned previously cannot be easily formulated
in the abstract (see e.g., Boesch, 1977). Thus the index must be
built on a case study and then generalized if possible. The task
is formidable but an equal problem is acquiring an adequate data
base which would also be available for other ecosystems. Records
of planktonic and other biological variables are often available

16.

in the open literature. In contrast, pollution loading values are hard to find over long periods -- although they are usually available somewhere in the files of individual dischargers (Russel and Horne, 1977) or in the files of the local regulatory agency (Horne, Fischer and Roth, 1982).

V. Summary

An index which will measure the health of aquatic ecosystems would be very useful in determining the amount of damage, or recovery from damage, in aquatic ecosystems. The index should ideally be robust, precise, and multi-dimensional and reflect changes due to either toxicity or biostimulation. An index of selected indicator variables--the SIV index--is proposed which builds on the existing EPA and other indices used to estimate "trophic state". The SIV index differs from the existing habitat evaluation indices in that the bias is towards aquatic ecosystems rather than terrestrial ones. This bias is needed since the damage caused by humans to the two habitats is of a different kind. The structure of aquatic ecosystems is dynamic and is maintained by short-term biological and chemical inputs. Terrestrial ecosystems depend much more on the physical structures such as trees and hills. Water pollution usually destroys the chemical and biological structure while terrestrial disruption, such as housing developments or dams, destroys the entire physical structure.

The SIV index follows recent trends to use functional components of the ecosystem rather than only taxonomic classification. The index is comprehensive in that it uses both types of information. A major difference from other indices is an emphasis on precision so that small changes in the health of the ecosystem can be detected with statistical confidence. In this way damage can

17.

be detected before it is too late and recovery techniques modified during restoration to maximize benefits. The only way to achieve precision and avoid both type 1 and type 2 errors is to make a large number of measurements. This can be done if the variables chosen are inexpensive to measure, and this concept drives the choice of variables in the SIV index.

Common, numerous, and functionally important variables would be chosen for the SIV index. In most open-water aquatic ecosystems the plankton provide a good source of information on the health of the ecosystem. The plankton include the young stages of most commercially and recreationally important fishes, their food, and the photosynthetic base of the entire food chain. The plankton are sufficiently numerous and homogeneous to sample at a reasonable cost and are most directly exposed to water-born pollutants. For wetlands and streams the same principles apply but the collection techniques must be modified by the use of analogs to achieve the same high precision at a comparable cost.

Future research should focus on long-term data sets from already damaged test ecosystems where data are readily available and easily supplemented. This concept is opposite of the NSF long-term research program which considers only pristine ecosystems. Thus data from various less accessible "grey" literature will be the principle source of information.

References

Cantor, L.W. 1984. A comparison of habitat evaluation
methods. EPA. Environmental Impact Seminar,
Washington, D.C. Feb. 1984. 73 p.

Carlson, R.E. 1977. A trophic state index for lakes. *Limnol.*
Oceanogr. 22 361-369.

Edmondson, W.T. and A. Litt. 1982. *Daphnia* in Lake
Washington. *Limnol. Oceanogs.* 27:272-293.

Hendricksen, A. 1979. A simple approach for identifying and
measuring acidification of freshwater. *Nature.*
278:542-545.

Hern, S.C., V.W. Lambou, L.R. Williams, and W.D. Taylor, 1981.
Modifications of models predicting trophic state of
lakes: adjustment of models to account for the
biological manifestations of nutrients. (Summary) U.S.
EPA Pb. 81-144 362.

Horne, A.J., J.C. Sandusky, J.C. Roth and S.J. McCormick.
1978. Biloogical and chemical conditions in the Truckee
River, California -- Nevada during the low flow
conditions of the 1977-8 severe drought. Rept. to
McLaren Engineering Co., Sacramento, California. 87 p.

Horne, A.J., H.B. Fischer, and J.C. Roth. 1982. Proposed
Monitoring Master Plan for the San Francisco Bay-Delta

Aquatic Habitat Program. Calif. State Water Resources
Board, Sacramento. 180 p.

Lueschow, L.A., J.M. Helm, D.R. Winter, and G.W. Karl. 1970.
Trophic nature of selected Wisconsin Lakes. Wisc. Acad.
Sci. Arts Lett. 58:237-264.

Lund, J.W.G., C. Kipling, and E.D. Le Cren. 1959. The inverted
microscope method of estimating algal numbers and the
statistical basis of estimations by counting.
Hydrobiologia. 11:143-170.

Michalski, M.F. and N. Conroy. 1972. Water quality
evaluation. Lake Alert study. Ontario Min. Envir.
Rep. 23 p.

McColl, R.H.S. 1972. Chemistry and trophic state of seven New
Zealand lakes. N.Z. J. Mar. Freshwat. Res. 6:399-447.
Needham, P.R. and R.L. Usinger, 1956. Variability in
the macrofauna of a single riffle in Prosser Creek,
California, as indicated by the Surber sampler.
Hilgardia 24, 14:383-409.

Rawson, D.S. 1955. Morphometry as a dominant factor in the
productivity of large lakes. Verh. Int. Verein.
Limnol. 12:164-175.

Richerson, P., R. Armstrong, and C.R. Goldman (1970).
Comtemperaneous disequilibrium, a new hypothesis to
explain the "paradox of the plankton." Proc. Natl.
Acad. Sci. 67:873-880.

Riley, G.A. 1976. A model of phytoplankton patchiness. Limnol. Oceanogr. 16:453-466.

Russell, P.P. and A. J. Horne, 1977. The relationship of wastewater chlorination activity to Dungeness Crab landings in the San Francisco Bay Area. U.C. Berkeley-SERHL Rept. 77-1, 37 p.

Ryder, R.A., S.A. Kerr, K.H. Loflus, and H.A. Regier. 1974. The morphoedaphic index, a fresh yield estimator -- review and evaluation. J. Fish. Res. Bd. Can. 31:663-688.

Sandusky, J.C. and A. J. Horne. 1978. A pattern analysis of Clear Lake phytoplankton. Limnol. Oceanogr. 23:636-648.

Shannon, E.E. and P.L. Brezonik. 1972. Eutrophication analysis: a multivariate approach. J. Sanit. Eng. Div. ASCE. 98:37-57.

Sheldon, A.L. 1972. A quantitative approach to the classification of inland waters. pp. 205-261 in Natural Environments ed. J. V. Kratilla, Johns Hopkins Press. Baltimore.

Steele, J.H. 1974. The structure of marine ecosystems. Blackwell, 128 p.

Talling, J.F. 1976. The depletion of carbon dioxide from Lake Water by phytoplankton. J. Ecol. 64:79-121.

Wurtz, C.B. 1960. Quantitative sampling. Natilus, 73:131-5.

The Hysteresis Effect in the Recovery of Damaged Aquatic Ecosystems:

an Ecological Phenomenon with Policy Implications

Abstract

The individual species or functional components of an ecosystem can be expected to respond at different rates to the application and/or removal of pollutant stress. These rates are primarily dependent on the generation time (a function of body size and complexity) of the organism and its place in the trophic hierarchy (e.g. producer, grazer or carnivore). Even in the absence of population extinctions, a non-retraceable behavior (or hysteresis effect) is expected. Conceptually, the lower trophic levels will follow a series of nested hysteresis curves, while organisms at higher trophic levels, such as sports fish, will probably respond more erratically. To explore these issues, we develop an illustrative hysteresis trophic-link model (HTLM) that incorporates limited ecological reality but is simple enough to expand to an arbitrary number of functional groups. This model is compared to a conceptual model for biotic hysteresis for a system with three trophic levels. We show how hysteresis might influence population changes at higher trophic levels (e.g. fish) caused by pollution. These changes cannot be measured directly because large fish are difficult to sample with high precision.

Introduction

In most aquatic ecosystems damage occurs by two mechanisms. These are physical destruction (for example, lake edge filling) or chemical perturbation (notably, additions of biostimulants and toxicants). With the exception of

sediment loading, most pollutants regulated by the U.S. Environmental Protection Agency (EPA) cause damage by chemical perturbation of ecosystems.

It is often assumed that above the dose-response threshold, the change in some component of an aquatic ecosystem is linearly proportional to the amount of pollution, as for example in the Dillon-Rigler (1974), Vollenweider (1968), or Vollenweider and Kerekes (1980) phosphorus- (or nitrogen-) chlorophyll models of lake eutrophication. Studies on lake restoration have shown that non-linearity and time lags in the recovery of systems perturbed by pollution occur for at least some lakes (e.g. Shagawa Lake, Maguey et al, 1973; Lake Washington, Edmondson, 1972). The reasons for non-linearity have not been well studied, but they appear to be partially due to the varying "turnover-times" of the physical, abiotic, and simple biotic components of a complete aquatic ecosystem (Edmondson, 1982; Horne, unpublished). Further step-function-type responses and time lags may be introduced by "higher-order interactions" that occur far from the site of the pollutant action. Examples of these interactions are species displacement such as occured for lake trout in the Great Lakes, or indirect competition from changes in species dominance (Christie, 1974). Given these complications, it is not surprising that the recovery of an ecosystem's more complex biotic levels, such as that of a damaged sports fishery, does not proceed either in a simple linear or virtually instantaneous manner upon removal of a pollutant load.

It is important to distinguish between the purely physiochemical and the biotic responses to removal of a pollutant from an aquatic ecosystem. All pollutants will decrease when the source is shut off and the internal pollutant load is diluted as new clean water flushes out the system. In many cases the pollutant load will be negligible in months or years--as is the case following the onset of phosphorus removal by new sewage-treatment plants (Goldman and Horne, 1983, pp. 392-4). In any event the physiochemical

response is generally predictable from a knowledge of the pollutant, the hydraulic residence time in the system, the mean depth, and the characteristics of the bottom sediment.

In contrast, the biotic response may be delayed or may occur in spurts. In extreme cases the biota may never return to their original states. The time path of ecosystem recovery is not predictable at present since the reasons for non-linearity are unclear. The response of the biota to a decrease in pollution may also fail to mirror the response of the system to the original increase in pollution, that is, the response may be non-retraceable. This paper attempts to provide a theoretical basis for a mathematical description of the biotic restoration of damaged aquatic ecosystems. In particular the non-linear and non-retracable character of the process of recovery from pollution--defined here as the hysteresis effect--will be considered.

In the following sections we present the general methods and theoretical basis for the hysteresis trophic-link model (HTLM), describe in a theoretical way our concepts of "ideal" and "non-ideal" biotic hysteresis, show the specific form used for the HTLM and some initial results from the modelling effort, and discuss the merits and drawbacks of the HTLM approach in providing information useful in setting environmental policy.

Methods and Theoretical Basis

Time lag effects may have many sources, but it is most logical (in the sense of Occam's razor) to examine first the turnover time of the components of the ecosystem as a possible source. If a population is to recover quickly when the pollutant load is removed it must grow and breed quickly. Since the larger organisms depend on the smaller ones as food sources, populations of larger organisms cannot grow until populations of smaller organisms are in

3

place.   The turnover time for biota is usually the generation time and can range from a few hours for simple bacteria and algae to decades for very large fish such as striped bass or sturgeon.   Generation time is primarily a function of two variables: the sexuality of reproduction and the structural complexity of the adults.   Asexual reproduction (vegetative or parthenogenetic reproduction) is typical of simple animals and plants growing under favorable conditions.   Sexual reproduction is typical of more complex organisms or of simple ones growing under unfavorable conditions.   Sexual reprodiction uses more time than asexual reproduction, and confers few, if any, short-term benefits.   In addition, complex organisms must spend time in building their large complex body structures.   This involves several moults, a long adolescence, and differing environmental requirements for adult and young, depending on the species involved.   The organisms in the trophic levels usually present in aquatic ecosystems have the following typical characteristic sizes (length, 1) and generation times (gt):

phytoplankton	1 = 0.02 mm,	gt = 3 days
zooplankton	1 = 1 mm,	gt = 3 weeks
ichthyoplankton	1 = 1 cm,	gt = 1 year
juvenile piscivorous and planktivoral adult fish	1 = 5 cm,	gt = 1 year
piscivorous fish	1 = 20 cm,	gt = 3 years
large sports fish	1 = 50 cm,	gt = 10 years.

The aquatic ecosystem we use in our model is simplified in the sense that side, across, and multiple-step (omnivory) food-chain links are omitted (Figure 1).   Although this may seem like a major simplification when one considers the apparently highly cross-linked structure of some aquatic food webs (e.g. Figure 2), the dynamics of many food webs are in fact much less cross-linked, in terms of energy or food flow, than they appear to be.   This
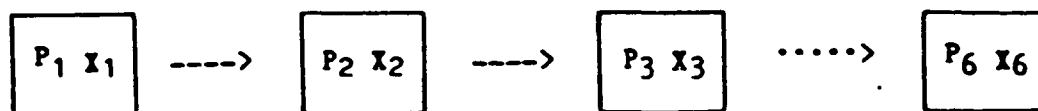
Figure 1. Schematic diagram of a trophic-link model. Here P describes the effect of the pollutant on each trophic level, and X is a measure of the biomass present for each functional class of organism (e.g. primary producers, filter-feeders, carnivores, etc.).
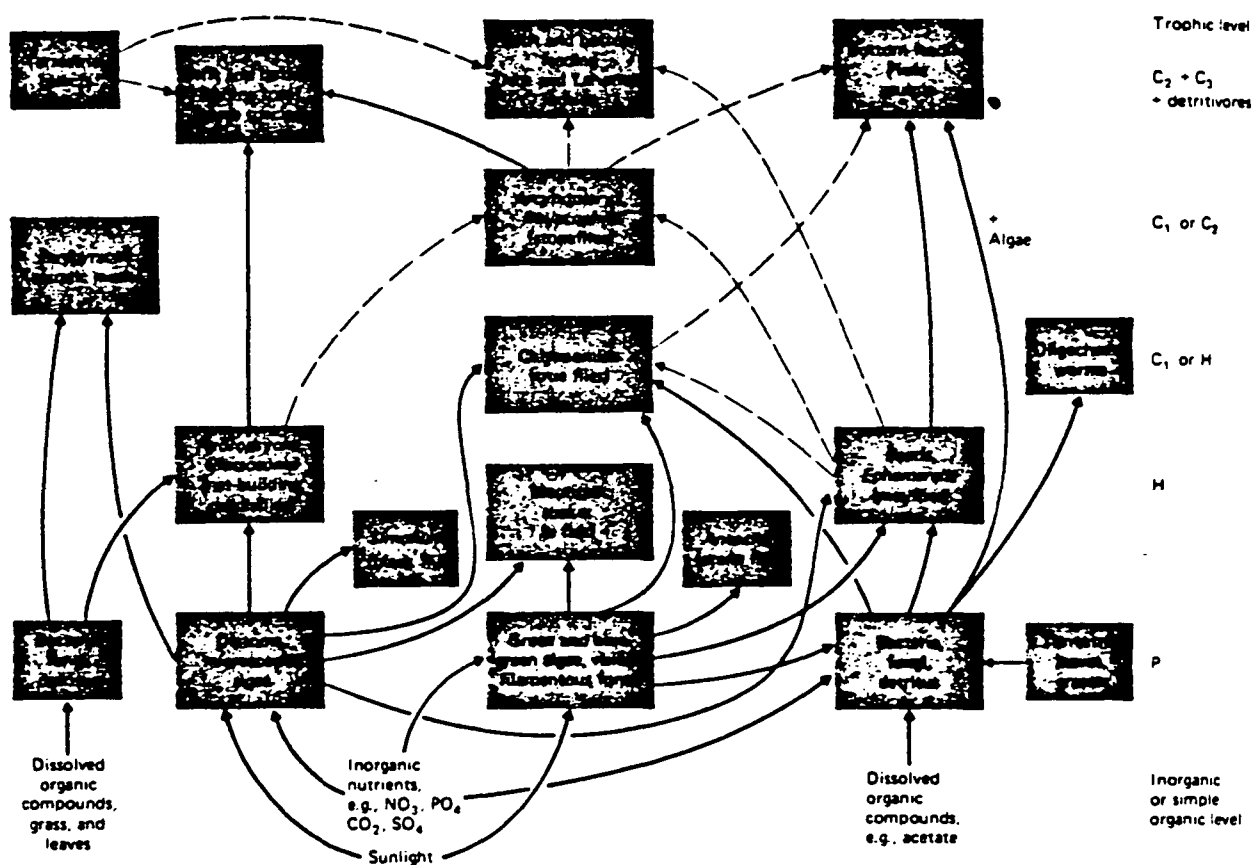

Figure 2. A qualitative food web for the Truckee River, California. Solid lines indicate measured pathways. Broken lines are assumed pathways derived from other studies of adjacent waters. Note that the omnivorous feeders (e.g. dace, trout, sculpin) use more than one trophic level. Most herbivores prefer microscopic diatoms to large filamentous green and blue-green algae. (Reproduced from Goldman and Horne, 1983)

5

is illustrated by one of the few known quantitative examples of an aquatic food chain, that of the River Thames below Kennet mouth (Goldman and Horne, 1983). Figure 3a shows the complete food web for the Thames system. As complicated as this looks, placing of the organisms in this web into functional groups results in the much more simplified structure shown in Figure 3b. Thus while the assumption of a linear food chain is certainly a simplification, it may not be a bad starting point for modelling some aquatic ecosystems.

In the linear food chain depicted in Figure 1, the rate of change of the phytoplankton population can be described by the equation

(1)
$$\frac{dX}{dt} = r_X (X)(1 - (X/K_X)) - B_{xy}XY - b_X X, \text{ where}$$

X = the population density of phytoplankton (e.g. chlorophyll a per cubic meter of water),

r_X = the maximal growth rate of the phytoplankton population,

K_X = a carrying capacity constant,

B_{xy} = a rate constant describing predation of zooplankton on phytoplankton,

Y = the population of zooplankton that feed on the phytoplankton (X), and

b_X = the rate of loss of phytoplankton due to washout and other linear, donor-controlled mechanisms.

In this system we assume that each organism eaten is killed and that no significant amount of prey is uneaten.

Analogous equations can be used to describe the rate of change of the higher trophic-level populations. For example:
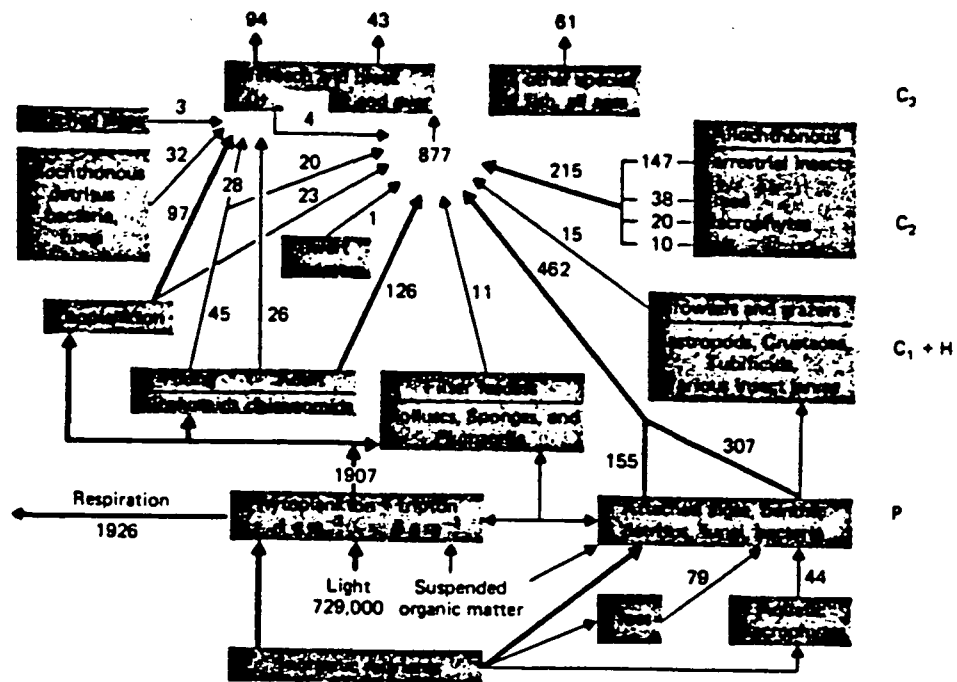
6

Figure 3a. This figure presents a dynamic food web for a natural system: an energy flowchart for the River Thames below Kennet mouth. In general, primary producers are shown at the bottom, invertebrate animals at the center, and fish at the top of the chart, but to avoid complex networks of arrows sources of attached algae, detritus, and allochthonous materials are shown in two places. Heavy arrows indicate the largest channels of energy flow. Note the twin flow of energy to fish from low-quality attached algae and high-quality animal food from terrestrial insects and adult chironomids. Energy input from dissolves organic matter was not measured directly. (Redrawn from Mann et al, 1972, reproduced from Goldman and Horne, 1983)
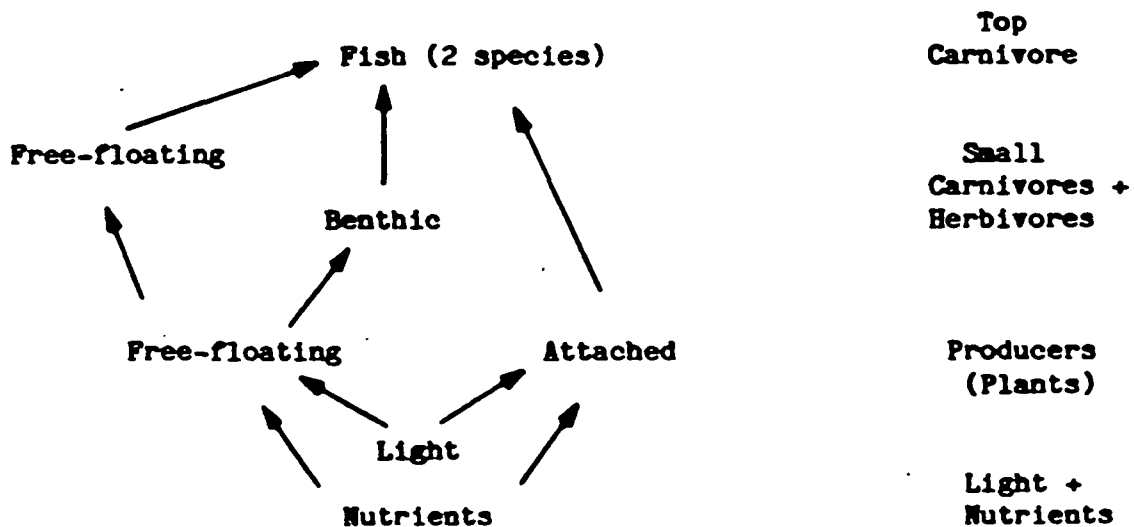
Top
Carnivore — Fish (2 species)

Free-floating

Small
Carnivores +
Herbivores — Benthic

Free-floating   Attached

Producers
(Plants)

Light

Light +
Nutrients — Nutrients

Figure 3b. The major energetic pathways from figure 3a. This diagram shows that modeling using single-link trophic models is possible if the organisms in the ecosystem are classified into functional rather than taxonomic groups.

7

(2)
$$\frac{dY}{dt} = E_{xy}B_{xy}XY - B_{yz}YZ - b_yY, \text{ where}$$

X, Y, and B_{xy} are as above,

E_{xy} = a factor describing the proportion of biomass consumed from trophic level x (phytoplankton) that is retained in trophic level y (zooplankton),

B_{yz} = a rate constant describing predation on zooplankton by icthyoplankton (small fish that feed on zooplankton),

Z = the population of icthoplankton, and

b_y = the loss rate of zooplankton due to washout, death, or other donor-related mechanisms.

This pair of equations can be expanded to an arbitrarily large set describing an arbitrarily long food chain.

Changes in pollution will affect some of the growth rates directly, but all populations will be affected as a result of trophic interactions. A straightforward example of such an interaction is the following. Suppose a pollutant acted so as to decrease the growth rate (r_x) of the phytoplankton in an aquatic ecosystem. This pollutant could be toxic to the phytoplankton or could be an inert pollutant, like silt in a lake, that affects r_x by decreasing the light available for photosynthesis. In either case, a reduction in the phytoplankton growth rate reduces the phytoplankton population, which reduces the amount of food available to the zooplankton, which reduces the zooplankton population, which reduces the amount of food available for small fish, and so on. Alternatively, a pollutant may cause an overall increase in total phytoplankton (e.g. through eutrophication) but bring about a decrease in zooplankton levels by allowing undesirable algal species to dominate at the expense of species that serve as food for the zooplankton. In this paper we have used mathematical relationships like those described above to generate a series of "hysteresis relationships"

charting the response of each trophic level in a hypothetical three-level aquatic food chain to the pollution and subsequent clean-up of the ecosystem.

We have also assumed, in making our calculations, that the onset of pollution and its clean-up are instantaneous. This is perhaps appropriate for longer-lived organisms such as fish, but has some inappropriate features for algae, which turnover rapidly and thus may respond to intermediate as well as initial and final levels of the pollutant. If it proves important to do so, a gradual change in pollution may be modeled in future work, but for our initial analysis the step-function approach is more enlightening and expedient.

Biotic Hysteresis: Theoretical Concept

The ecological hysteresis response will resemble the physical hysteresis effect observed in the magnetization of a ferromagnet. When a magnet is placed next to an unmagnetized bar of iron, the latter becomes magnetized. When the first magnet is taken away, the iron bar loses its magnetic properties much more slowly than it gained them. Similarly, as the level of pollution in an aquatic ecosystem is decreased, the biological response to the decrease does not trace out in reverse the path it followed in response to the initial pollution of the system. Nevertheless, ideally, the system, returns to its starting point. For the purposes of this paper we define "ideal" biotic hysteresis to occur when a population of organisms perturbed by pollution returns to its initial population level within a period of time short enough to be relevant to policy decisions. This time period might be 10 to 20 years. In an ecosystem with several trophic levels (phytoplankton, large zooplankton and small fish, and large fish, i.e. producers, grazers, and large carnivores) and a single type of pollutant (such as sewage) a series of response-and-recovery curves such as those shown in Figure 4 would be
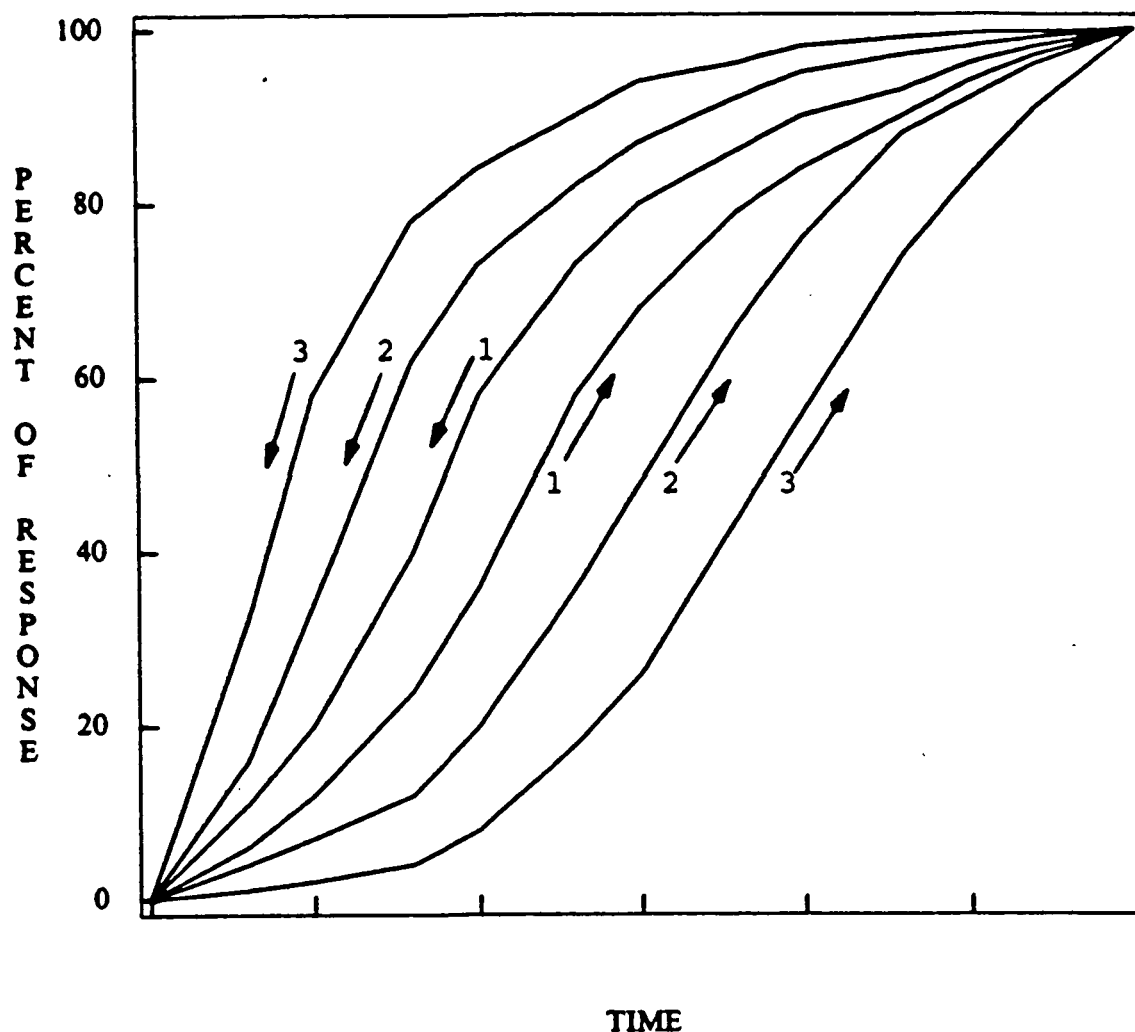
Figure 4. A theoretical distribution of ideal hysteresis curves for an aquatic ecosystem with three trophic levels. Curves marked "1" represent the time-path of the response of a population in a lower trophic level (e.g. phytoplankton) to a pollutant stress and the path of recovery once the stress has been removed. "Response" paths are marked with left-to-right arrows, while "recovery" paths are indicated by right-to-left arrows. Curves marked "2" and "3" represent time-paths for middle (e.g. zooplankton) and higher (e.g. fish) trophic levels, respectively. Note that populations in higher trophic levels exibit greater lags in both response and recovery than those in lower trophic levels.

expected. Even the rapidly growning phytoplankton (generation time 1-10 days) can exibit an ideal hysteresis response to the pollutant. For higher trophic levels (copepod zooplankton and fish), which respond to the altered phytoplankton population, there will be a delay in the initiation of the exponential section of the curve in Figure 4 in rough proportion to the generation time. A delay must occur because complex organisms are incapable of rapidly increasing their number (that is, they have a slow numerical response) on a time scale of days. It will thus take at least the adult-to-birth-to-juvenile period before copepods or small fish can show any numerical response to the perturbation, and this response period will be slightly shorter than the complete generation time. This lag in response has the interesting consequence that the last half of the change will occur more rapidly for high than for low trophic levels. Such rapid changes would be of serious concern to resource managers since the response of pollution-control agencies may be too late to protect the resource before the numbers of important organisms are seriously depleted. These rapid changes do in fact seem to happen (see Goldman and Horne, 1983). Concern about such changes is compounded by the fact that it is difficult to measure changes in biomass stocks at higher trophic levels, such as fish. The statistical resolution for fish stock estimation is usually so poor that the majority of a fish population can be lost before biologists can detect the change with any certainty.

The ideal hysteresis effect is characterized by a cyclic (on a 10-20 year time scale) non-retracable path when the response of organisms[*] to pollution

----

[*]In figures 4 and 5 the response of each trophic level is normalized so that the "percent response" at each time point is given as a percentage of the difference between the population of the organism before the system was perturbed and the population at the point where the pollutant is removed. Thus these curves show increasingly lagged responses and recoveries from pollutant stresses, but do not reflect the relative magnitudes of the responses to pollution that might be shown by the different trophic levels.
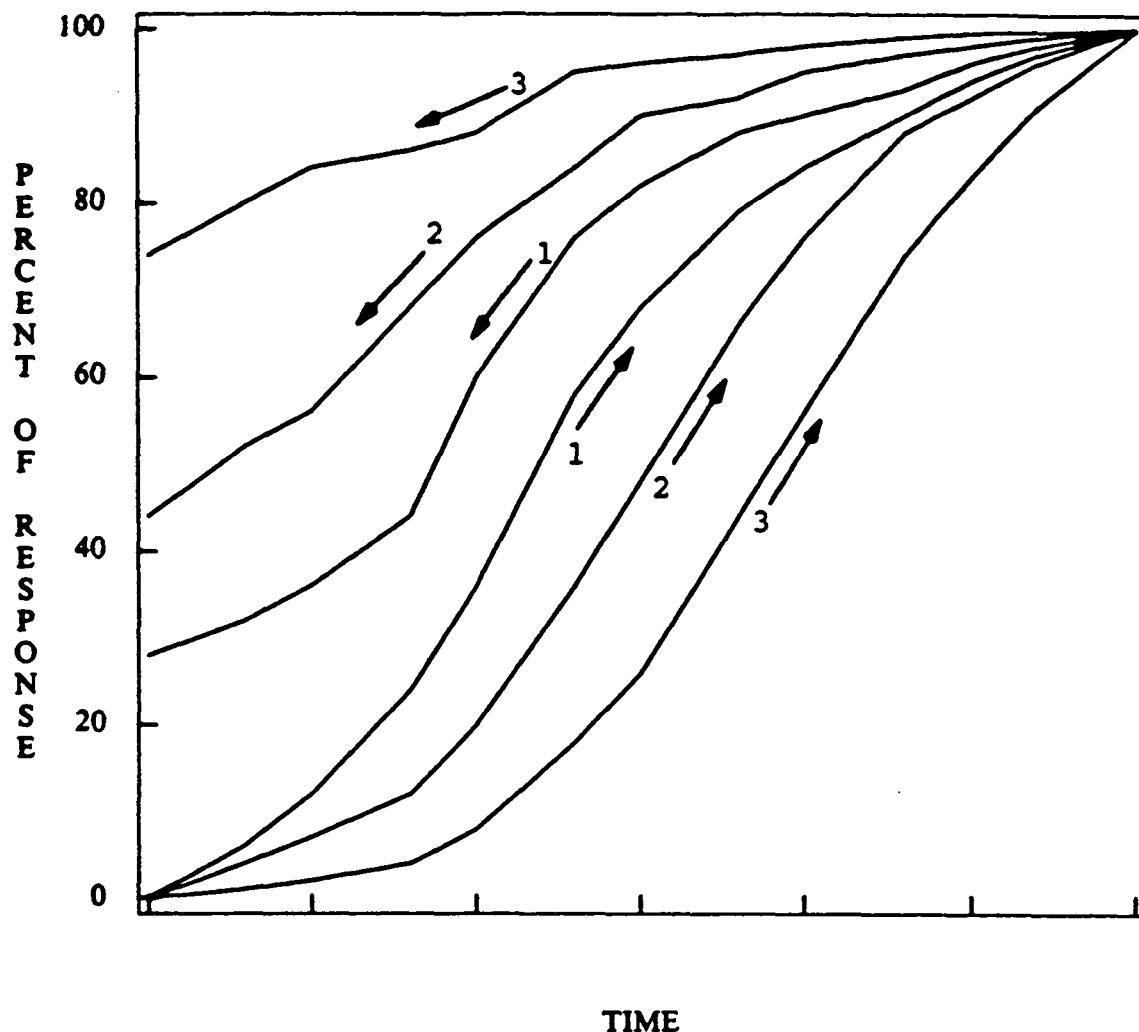
11

Figure 5. A theoretical distribution of non-ideal hysteresis curves for an aquatic ecosystem with three trophic levels. Curves marked "1" represent the time-path of the response of a population in a lower trophic level (e.g. phytoplankton) to a pollutant stress and the path of recovery once the stress has been removed. "Response" paths are marked with left-to-right arrows, while "recovery" paths are indicated by right-to-left arrows. Curves marked "2" and "3" represent time-paths for middle (e.g. zooplankton) and higher (e.g. fish) trophic levels, respectively. In this case, unlike the ideal theoretical case presented in figure 4, the populations do not recover completely within a recovery period of the same duration as the original stress.

is plotted against time for a regime in which a pollutant is added (left-to-right paths in figures 4 and 5) then removed (right-to-left paths). A damped hysteresis effect is also possible. This effect, which we have termed "non-ideal" biotic hysteresis, is characterized by non-retracable and non-cyclic behavior (as shown in figure 5), is also possible. A possible explanation of such behavior for a specific food chain (rather than a food chain of generalized trophic levels) is the following. If a species of plant or animal remains at depressed levels (e.g. as a result of a pollutant-related stress) for long periods there is in effect a vacant niche that can be occupied by a pollution-tolerant species or even another species that has no direct effect on the fish of concern (Christie, 1974). Generally the replacement species are less highly regarded by sports and/or commercial fisheries groups and are an economically inferior substitute for the original species. Thus if the return leg of the hysteresis curve is very flat after cessation of pollution, organisms at the valuable higher trophic level may be subject to "species replacement" or "competitive displacement" and never return to their original dominant position.

Methods and Initial Results from the Hysteresis Trophic-Link Model (HTLM)

Our objective in this modeling effort was to test a simple approach for describing mathematically the hysteresis phenomenon discussed above. The purpose of the model described here is solely to illustrate how a generalized ecological phenomenon of interest (hysteresis) can be demonstrated using mathematical relationships containing easily identifiable and understandable parameters. In this approach a food chain with three trophic levels--phytoplankton, zooplankton, and small fish--was assumed. The rate of change of the populations in the first two trophic levels were described by differential equations (1) and (2) above, and the rate of change of the

13

population in the third trophic level was described by

$$(3) \qquad \frac{dZ}{dt} = E_{yz}B_{yz}YZ - b_zZ, \qquad \text{where}$$

Y, Z, and B_{yz} are as previously described,

E_{yz} = that fraction of biomass in the Yth trophic level that
becomes incorporated in the Zth level, and

b_z = a rate constant describing the loss of small fish due to
old-age death and other donor-controlled mechanisms.

The constants in the three equations were obtained by assuming a value of
0.1 for E_{xy} and E_{yz}, and a value of $2 \times X^*$ for K_x. The values for X^*, Y^*,
and Z^*, the steady-state biomass populations for the three trophic levels
(that is, the relative amounts of per-unit-area biomass for which dX/dt,
dY/dt. and $dZ/dt = 0$) were taken to be 50, 10, and 1, respectively.
Generation times for the three trophic levels (T_x, T_y, and T_z) were taken to
be 3, 20, and 360 days, respectively. The following relationships were used
to derive the values of r_x, B_{xy}, and B_{yz}:

$$r_x = T_x^{-1},$$

$$E_{xy}B_{xy}X^* = T_y^{-1},$$

$$E_{yz}B_{yz}Y^* = T_z^{-1}.$$

values for b_x, b_y and b_z were derived from the steady-state forms of equations
(1) through (3).

Equations (1) through (3) were incorporated into a fortran computer
program, which was used to approximate the time path of populations X, Y, and
Z in response to a perturbation in r_x, the phytoplankton. The program calls
the NAG (Numerical Algorithms Group, 1984) subroutine D02EBF, which integrates
systems of differential equations using a variable-order, variable-step Gear
method and returns solutions to the system ($X(t)$, $Y(t)$, $Z(t)$) at specified

time points.   Details of the model and a listing of the integration program
are given in the appendix to this paper.

We should note that an analytical approximation to the solution of
equations (1) - (3) can be obtained by adding a fouth equation, namely

(4)
$$\frac{dr_x}{dt} = 0$$

to the system, deriving a 4 x 4 "community matrix" using procedures described
by May (1973) and Harte (1985), and using that matrix to explore the effects
of perturbations to the system.   A four-level food-chain model was also
developed.   This model, which adds a larger piscivorous fish to the three-
tiered food chain, uses equations (1) - (3), above, with the term $- B_{zf}ZF$
added to equation (3). A fourth equation,

(5)
$$\frac{dF}{dt} = E_{zf}B_{zf}ZF - b_f F.$$

is added to model the behavior of the population of larger fish (F).   In this
system the steady-state biomass ratios in the four trophic levels were taken
to be 500 : 100 : 10 : 1 (X^* : Y^* : Z^*: F^*), the generation time for the
larger fish (T_f) was taken to be 1080 days, E_{zf} was taken to be 0.1, and
$E_{zf}B_{zf}Z^*F^*$ was defined to equal T_f^{-1}.   This four-level system was solved as
above.   Details of the model and a listing of the computer program used to
solve it are given in the appendix.

Results

The time paths traced by the three "populations" (here taken to mean
biomass present in each trophic level per unit area of water) following a -2%
reduction in r_x are shown in figure 6.   The population of phytoplankton drops
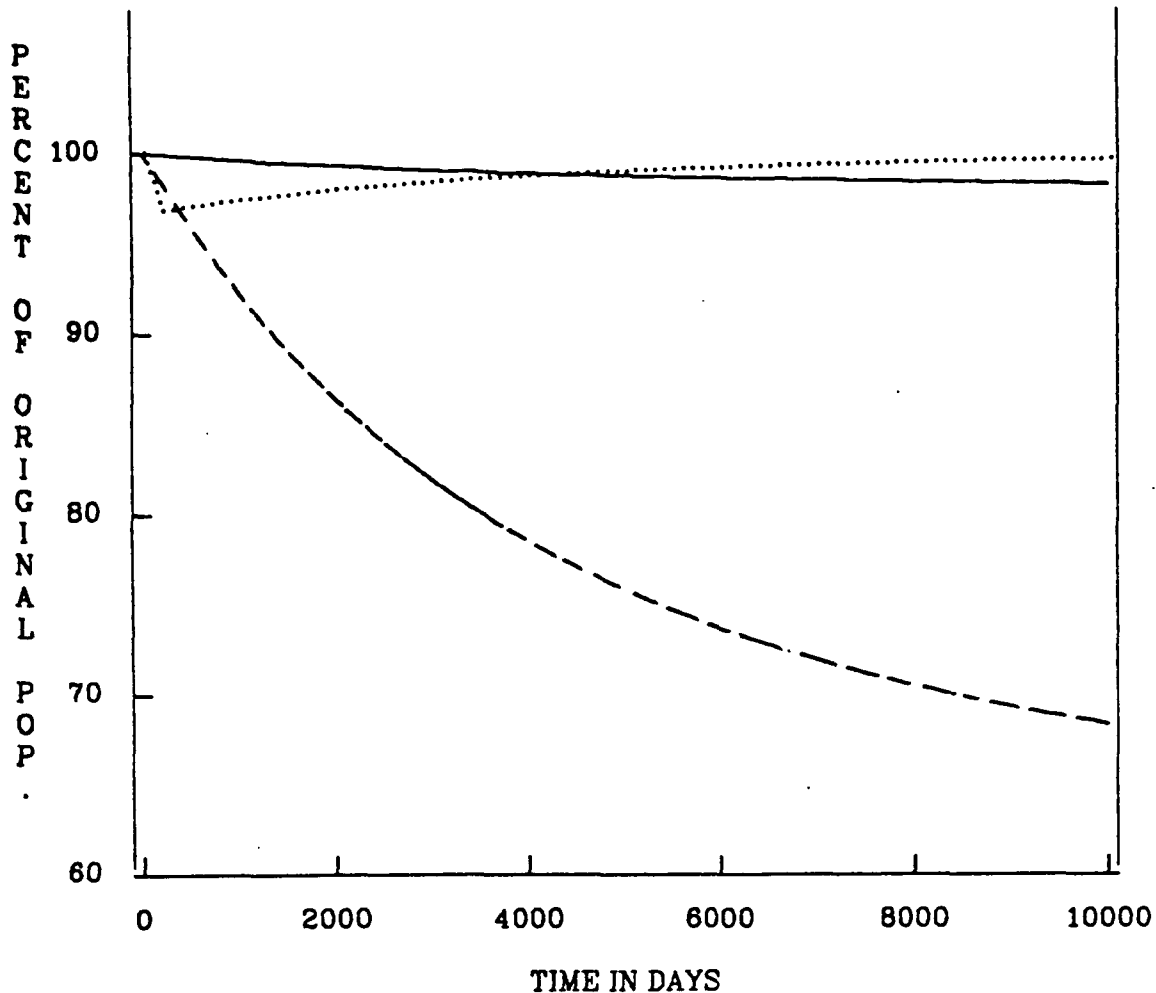rapidly in response to the reduction in its growth rate, reaching a local

Figure 6. Calculated time paths for the response of the populations (measured in biomass per unit area, initial biomass ratios: 50 phytoplankton: 10 zooplankton : 1 small fish) in a three-tiered aquatic food chain to a -2% change in the growth rate of phytoplankton. Solid, dotted, and (partially) dashed lines give the paths for phytoplankton, zooplankton , and small fish, respectively. Note that the lower trophic levels respond more quickly to the stress than higher trophic levels, but the ultimate effect on higher trophic levels is greater in magnitude.

minimum in 20 days (not visible in figure 6 due to the length of the time scale). Thereafter the population rises quickly, then falls slowly in response to the changes in the population of its predator (zooplankton). By the time 10,000 days (about 30 years) have elapsed, the phytoplankton population reaches a steady-state value equal to 99% of its original level. The population of zooplankton drops more slowly, but over a longer period. For this second trophic level the maximum deviation from the original population, -3.5%, occurs after 150 days. From there the zooplankton population rises to a level about 1% above that originally present. The population of small fish declines more slowly than those of either of the lower trophic levels, but in time exibits a greater response, reaching a new steady-state population 70% as large as the original group. Note that the deviations in the zooplankton and fish populations are out of phase with each other. This makes sense ecologically as well as mathematically: as fish populations decline, grazing pressure on zooplankton is decreased, allowing that population to expand. Perhaps the most important result shown in figure 6, however, is that a small (-1%) perturbation in the phytoplankton growth rate produces a large (-30%) change in the population at the highest trophic level.

Figures 7-10 present time paths for the three populations in which a -2% perturbation in r_x is applied at time zero, then removed at 300, 500, 2000, and 10,000 days, respectively. Paths for which arrows point left-to-right chart the response of the three populations to the original perturbation, while paths with right-left arrows chart the return paths for time periods of the same duration as the original perturbation. Thus in figure 7, for example, the solid curve labeled with a right-pointing arrow charts the response of the phytoplankton population to a perturbation applied for 300 days, while the solid path labeled with a left-pointing arrow charts the level
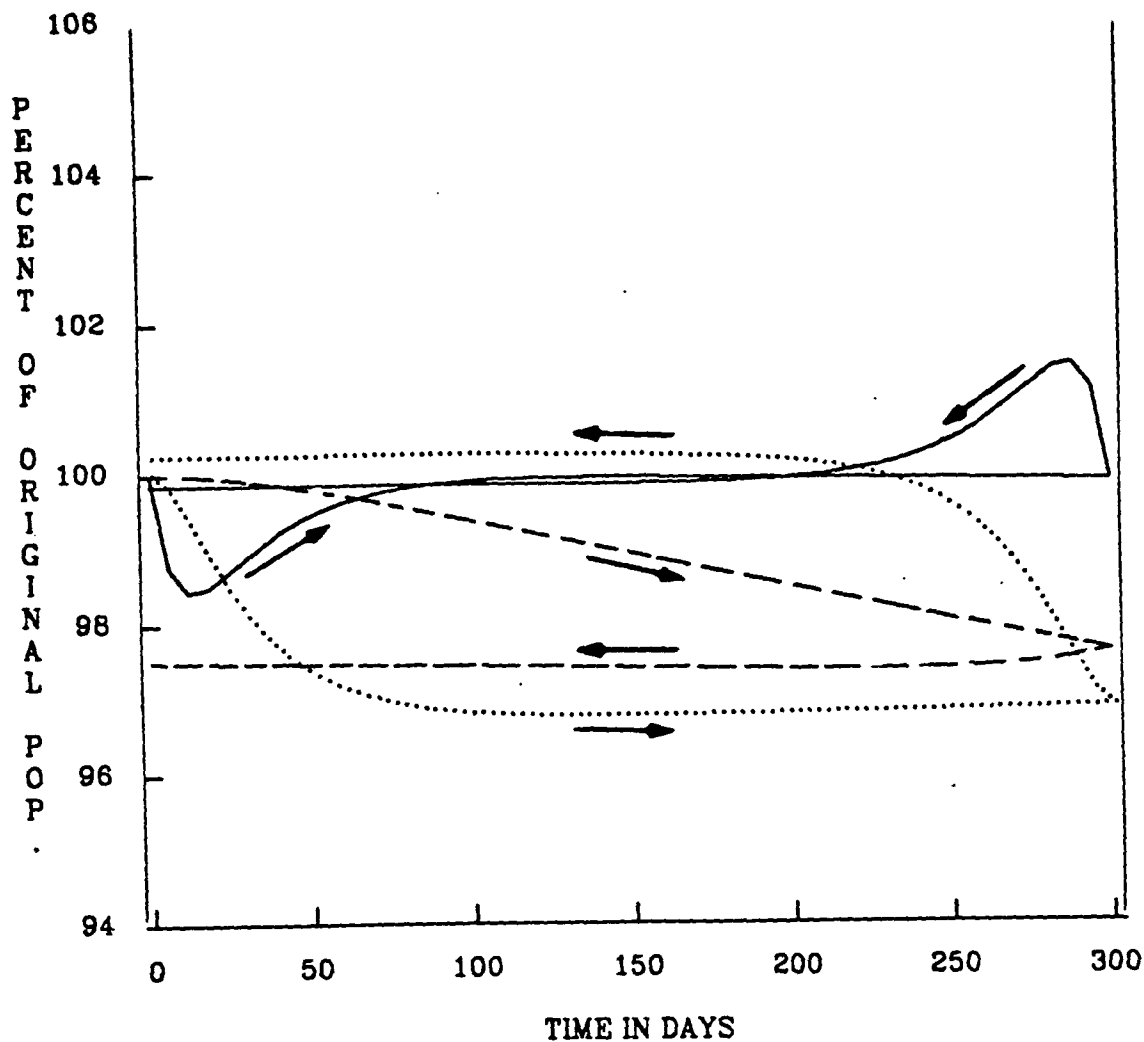
Figure 7 Calculated time paths for the response and recovery of the
populations (measured in biomass per unit area, initial
biomass ratios: 50 phytoplankton: 10 zooplankton : 1 small
fish) in a three-tiered aquatic food chain when a -2%
perturbation in the phytoplankton growth rate is applied at
time zero, then removed after 300 days. "Response" paths
are indicated by right-pointing arrows, and "recovery" paths
are marked with left-pointing arrows. Solid, dotted, and
(partially) dashed lines give the paths for phytoplankton,
zooplankton , and small fish, respectively. Note that the
population of small fish continues to decline even after the
perturbation is removed, and fails to return to its original
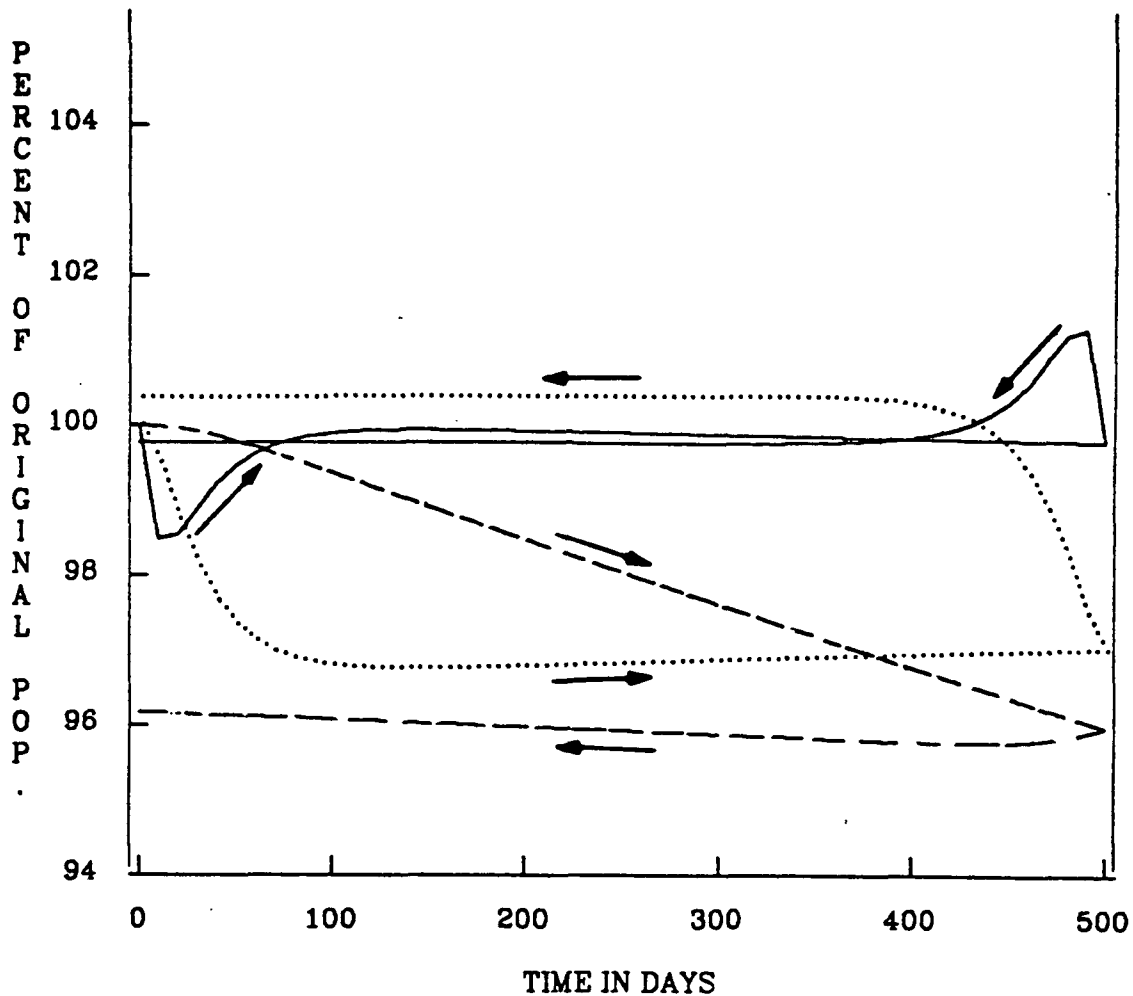position after 300 days of recovery.

Figure 8 Calculated time paths for the response and recovery of the populations (measured in biomass per unit area, initial biomass ratios: 50 phytoplankton: 10 zooplankton : 1 small fish) in a three-tiered aquatic food chain when a -2% perturbation in the phytoplankton growth rate is applied at time zero, then removed after 500 days. "Response" paths are indicated by right-pointing arrows, and "recovery" paths are marked with left-pointing arrows. Solid, dotted, and (partially) dashed lines give the paths for phytoplankton, zooplankton , and small fish, respectively. Note that the population of small fish shows a lag of approximately 50 days before beginning its recovery.
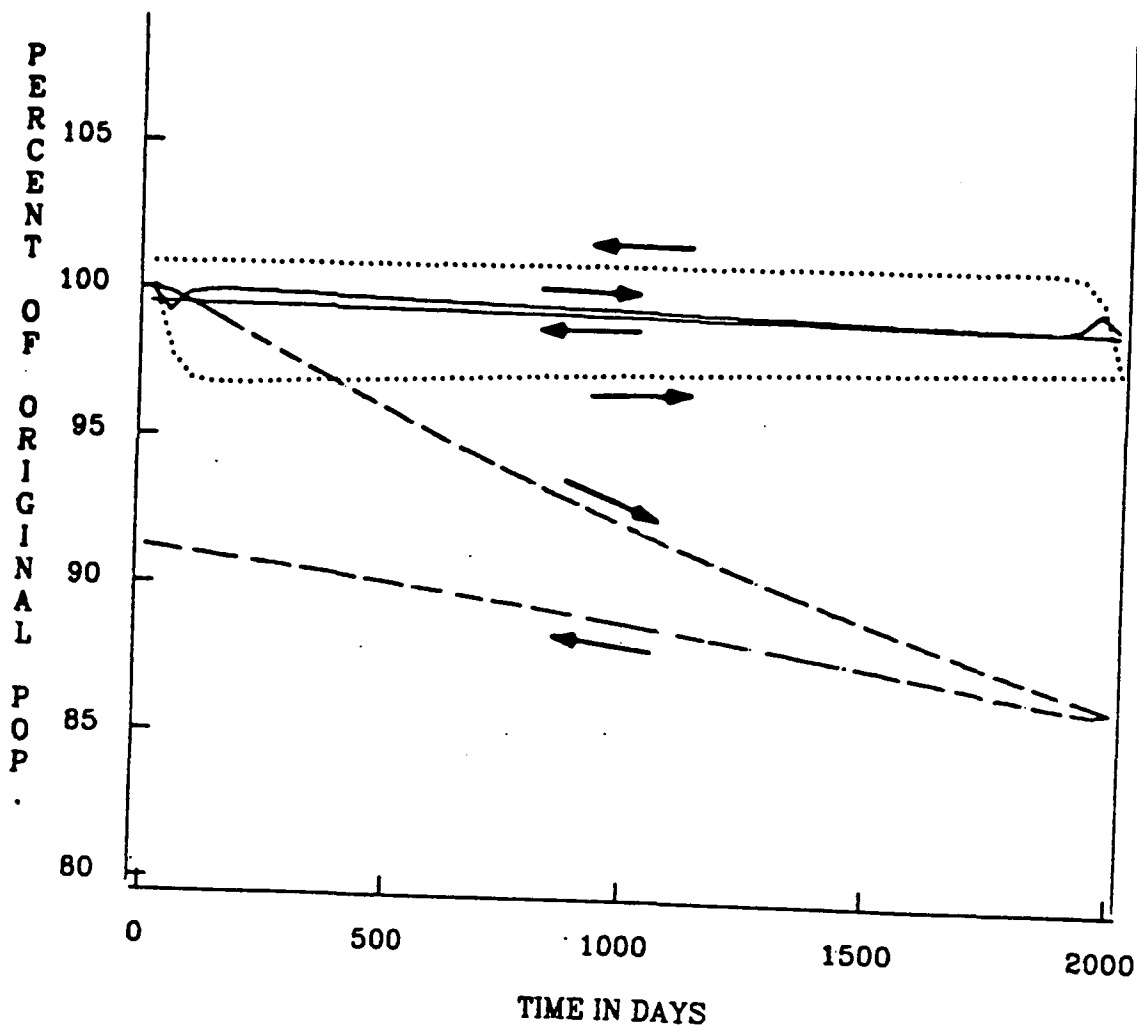
Figure 9 Calculated time paths for the response and recovery of the populations (measured in biomass per unit area, initial biomass ratios: 50 phytoplankton: 10 zooplankton : 1 small fish) in a three-tiered aquatic food chain when a -2% perturbation in the phytoplankton growth rate is applied at time zero, then removed after 2000 days. "Response" paths are indicated by right-pointing arrows, and "recovery" paths are marked with left-pointing arrows. Solid, dotted, and (partially) dashed lines give the paths for phytoplankton, zooplankton , and small fish, respectively. Note that the population of small fish fails to return to its initial level after 2000 days of recovery.

Figure 10 Calculated time paths for the response and recovery of the populations (measured in biomass per unit area, initial biomass ratios: 50 phytoplankton: 10 zooplankton : 1 small fish) in a three-tiered aquatic food chain when a -2% perturbation in the phytoplankton growth rate is applied at time zero,then removed after 10,000 days. "Response" paths are indicated by right-pointing arrows, and "recovery" paths are marked with left-pointing arrows. Solid, dotted, and (partially) dashed lines give the paths for phytoplankton, zooplankton , and small fish, respectively. Note that the population of small fish fails to return to its initial level even after 10,000 days of recovery.
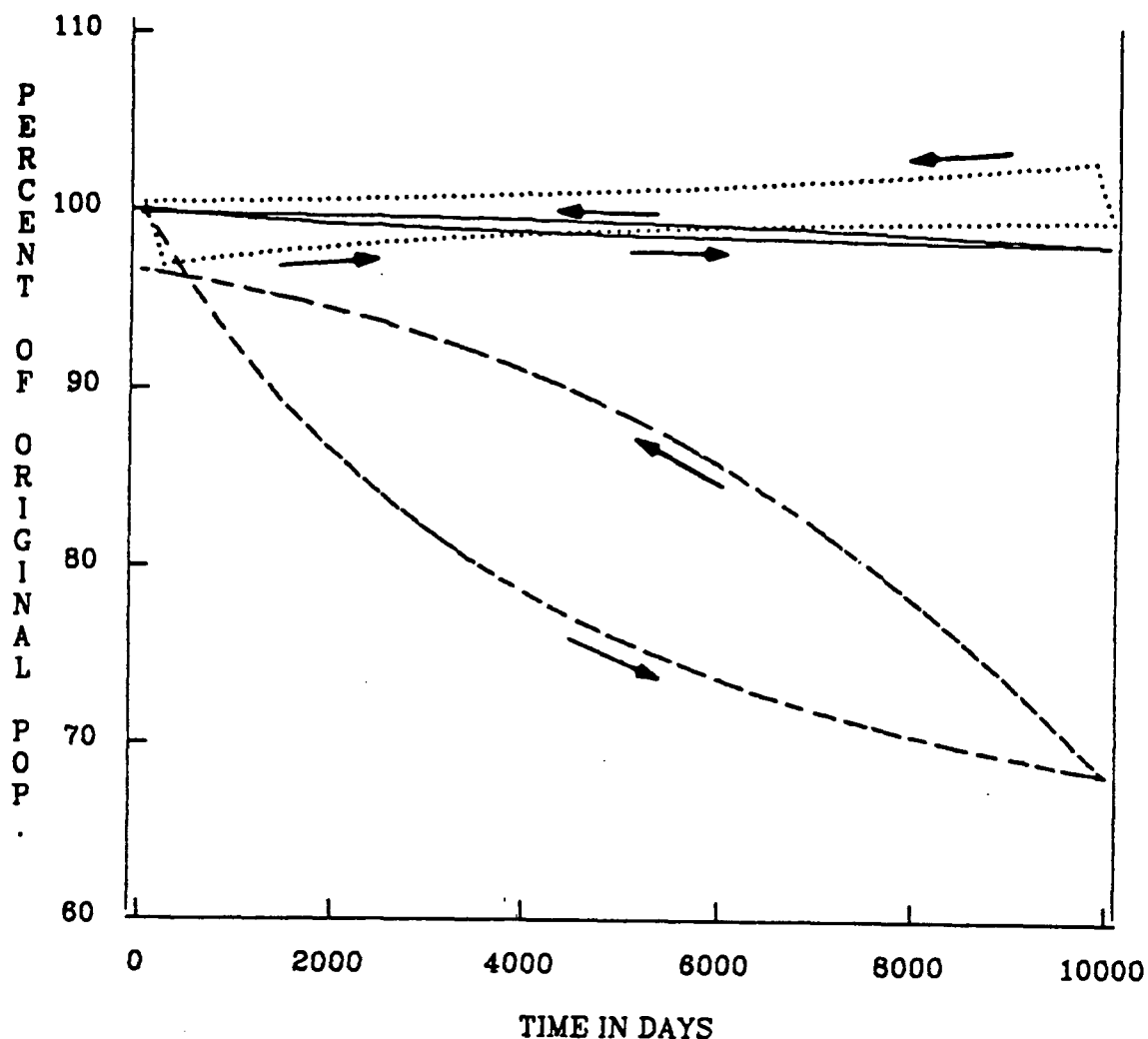
of the phytoplankton population after the perturbation is removed. For the return paths time runs right-to-left, thus the points on the return paths directly above "50" on the time axis are actually 250 days from the point where the perturbation was removed. The presentation of the hysteresis curves in figures 7-10 are different from those in figures 4 and 5 in that they are not normalized to the response of each population to the perturbation, rather they indicate the percentage change in each population. This allows the relative magnitudes of the population changes in the different trophic levels as well as the shapes of the hysteresis curves to be compared.

Figures 7-10 present a series of hysteresis curves in which time paths for the fish populations show a progression from non-ideal- toward ideal-hysteresis behavior, as those terms are defined above. For each time interval the phytoplankton population can be seen, after perturbation of the system, to decline rapidly to just above 98% of its original level, remaining near that value for the duration of the perturbation. When the stress is removed, the phytoplankton population quickly increases to 2% over its pre-perturbation level, then declines to its original level and remains relatively stable thereafter. In each of figures 7-10 the zooplankton population decreases rapidly following perturbation, then drifts slowly higher as fish populations decline. When the perturbation is removed zooplankton quickly increase, due to the increased availability of phytoplankton, then decline slowly to near their original level as fish populations increase. The population of fish shows a slow and steady decline over a 300-day perturbation. The decline continues for about 150 days after the perturbation is removed. In figure 8, the fish population again declines throughout the perturbation period and into the return period, but starts to recover approximately 50 days after the perturbation is removed. Figure 9 shows even less lag before the fish population starts to recover. Figures 7-

22

10 show ideal biotic hysteresis behavior for the populations in the two lower trophic levels, which return to roughly their original values. Note, however, that even in this case, where a recovery period of 10,000 days is allowed, the fish population does not quite return to its original level.

These results suggest the following conclusions. First, organisms at higher trophic levels show responses to perturbation of the ecosystem that are less immediate but greater in relative magnitude than the responses of lower trophic-level organisms. Second, organisms at higher trophic levels exibit a more pronounced lag in recovery from stress once the perturbation is removed. This lag has ecological importance beyond what we have been able to include in our modelling effort, as a period in which the population of an organism is low may provide an opportunity for another organism, quite often one that is economically less desirable, to come in and occupy the former's ecological niche.

Thus far the three-tiered ecosystem has been challenged with only a -2% reduction in the phytoplankton growth rate. Figure 11 shows the response of the system over the 3000 days following a -25% perturbation in r_x. In this case, the population of small fish declines to less than 10% of its original level. After the perturbation is removed, the fish population slowly increases, but remains at less than 10% of its original level even after 2000 days. If the system is allowed 10,000 days of recovery following a 3000-day -25% perturbation in r_x, the population of small fish gradually rises to 38% of its original level, still ow enough to constitute an example of non-ideal biotic hysteresis. It is probable that in a real system a sustained 90%--or even 60%--reduction in a fish species would result in another, perhaps less desirable, species occupying its ecological niche. This means that some component of an aquatic ecosystem may never recover from a stress, even if
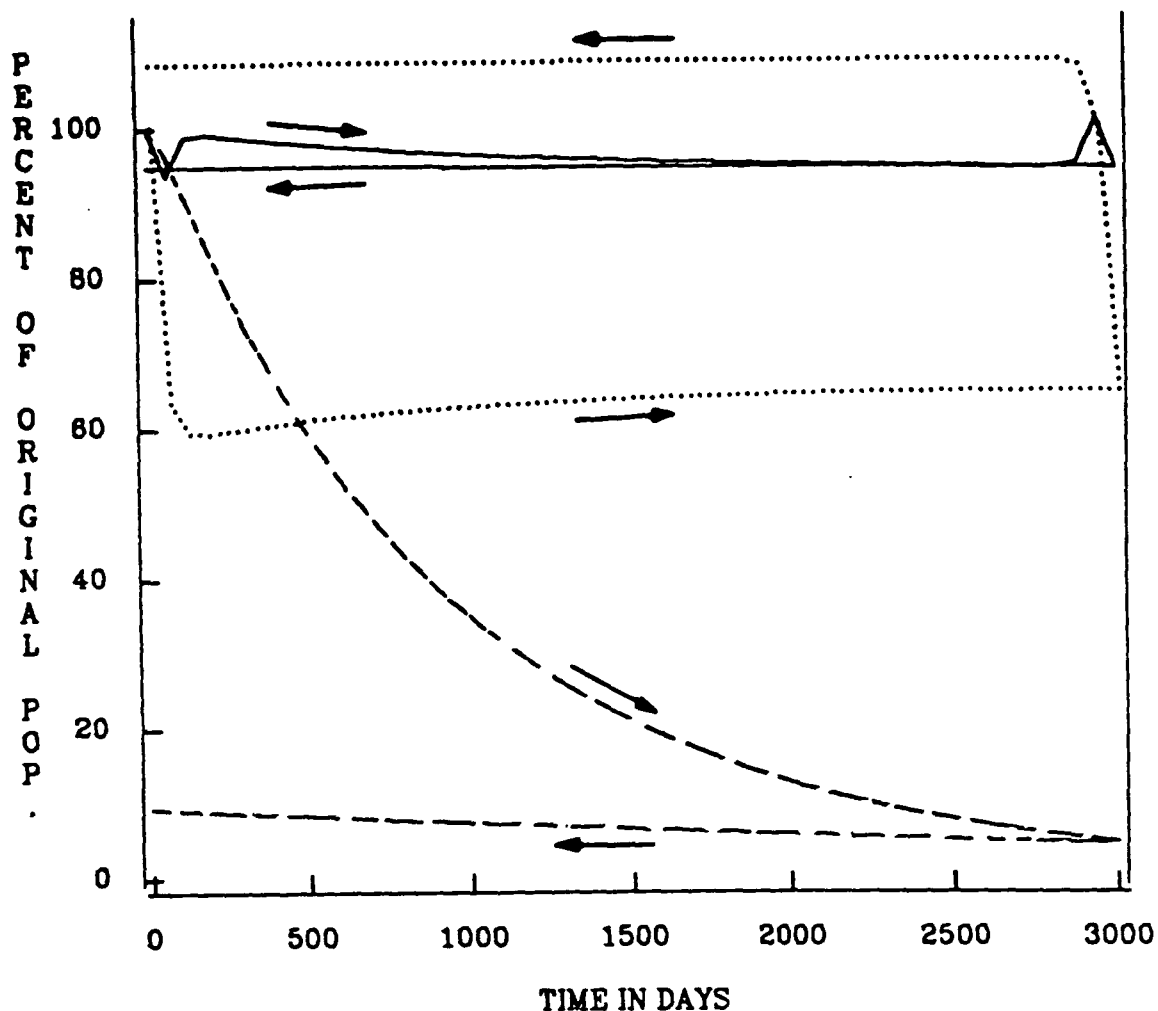
Figure 11 Calculated time paths for the response and recovery of the populations (measured in biomass per unit area, initial biomass ratios: 50 phytoplankton: 10 zooplankton : 1 small fish) in a three-tiered aquatic food chain when a -25% perturbation in the phytoplankton growth rate is applied at time zero, then removed after 3000 days. "Response" paths are indicated by right-pointing arrows, and "recovery" paths are marked with left-pointing arrows. Solid, dotted, and (partially) dashed lines give the paths for phytoplankton, zooplankton, and small fish, respectively. Note that the population of small fish falls to a critical level and fails to return to its initial level after 3000 days of recovery.

some fraction of the population remains after the stress is removed.

Figure 12 illustrates that assumptions as to the shape of the "biomass pyramid"--that is, the ratios of biomass-per-unit-area present for each trophic level--can have a profound effect on the magnitude of the magnification of perturbations down the food chain from producer to carnivore. Here we show that the effect of a -1% change in the growth rate of phytoplankton is greater on the fish population in a food chain with biomass ratios of 100 : 10 : 1 (phytoplankton : zooplankton : fish) than for food chains in which the trophic level ratios are smaller. It should be remembered that we know only that this result pertains to the simple predator-prey model we have been studying: the effect of the shape of biomass pyramids on responses to stress has yet to be investigated for other types of models.

Figure 13 presents the response of the populations in a four-tiered food-chain model to a -2% perturbation in the growth rate of the phytoplankton. Note that, as in the three-tiered case (figure 6) the relative magnitude of changes in the populations of the various trophic levels increase as the organisms get larger. Another similarity is that the lag in response to the perturbation is longer for higher trophic levels. The four-level model does, however, appear to be more stable: a -2% perturbation in r_x results in only a 10% decrease in the steady-state value of the larger fish population, while the highest trophic level in the three-tiered case is decreased 30% in population. In the four-tiered model all four populations oscillate in a damped fashion toward a steady state value. This is the sort of behavior that one might expect from a real ecosystem. It is also gratifying to note that the oscillations in the populations of each predator-prey pair are out of phase with each other. This makes ecological as well as mathematical sense. As the population of larger fish, for example, declines, grazing pressure on small fish decreases, allowing that population to expand. This increase in

25
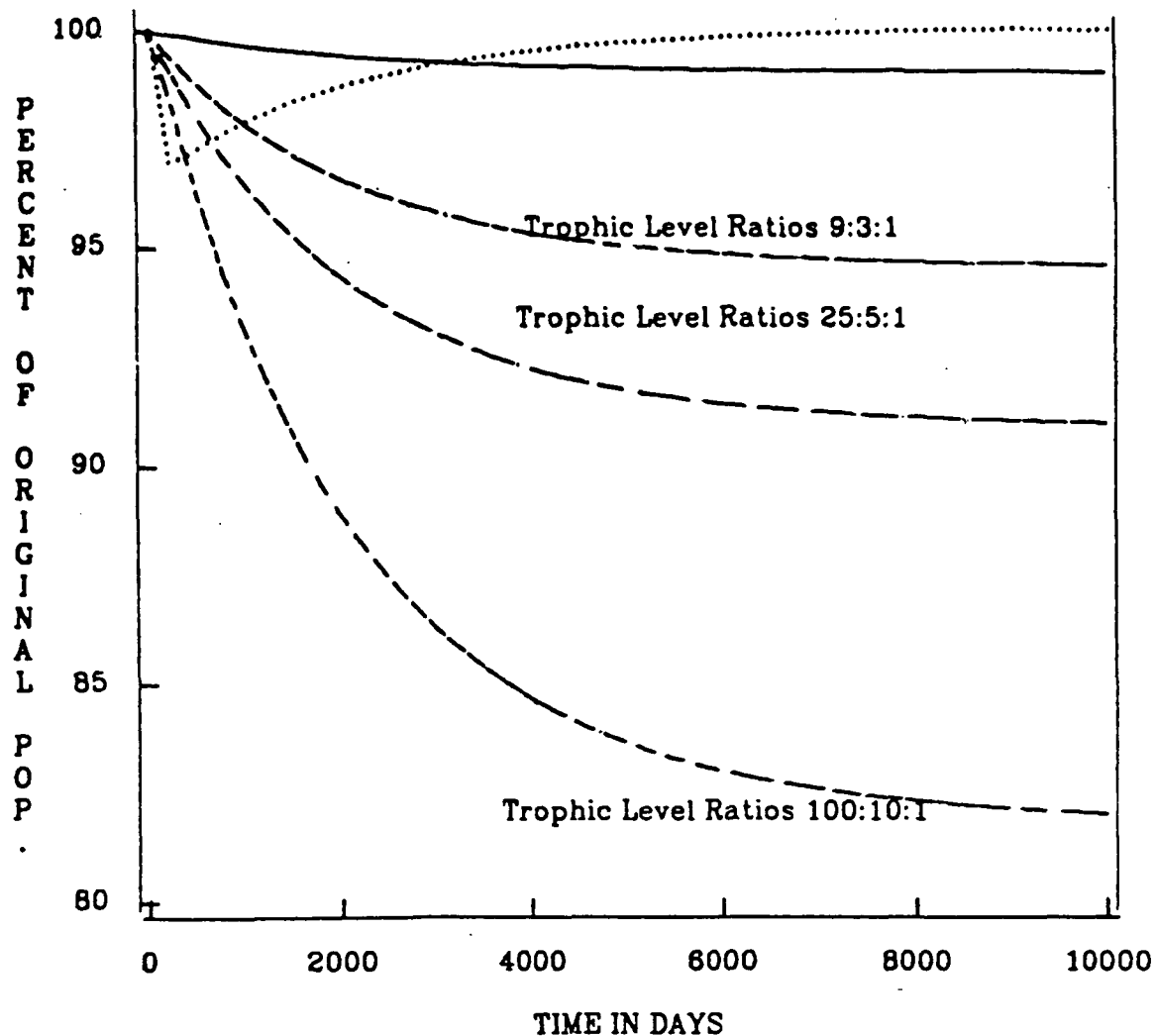
Figure 12 The response of three different three-tiered aquatic ecosystems to a -1% change in the phytoplankton growth rate. The partially dashed curves give the response of the small fish populations to the perturbation for food chains in which the initial biomass ratios (per-unit-area biomass of phytoplankton: zooplankton: small fish) are as indicated. The solid and dashed lines give the response of phytoplankton and zooplankton populations for a food chain with 100:10:1 biomass ratios.

Figure 13 The response of the populations in a four-tiered aquatic ecosystem (measured in biomass per unit area, initial biomass ratios: 500 phytoplankton: 100 zooplankton : 10 small fish: 1 larger fish) to a -2% perturbation in the phytoplankton growth rate. The paths for the responses of the phytoplankton, zooplankton, small fish, and larger fish populations are given by the upper solid curve, the dotted curve, and partially dashed curve, and the lower solid curve, respectively.

small fish creates favorable conditions for the larger fish, which multiply until the small fish have been overgrazed. At this point the population of larger fish starts to decline, the small fish start to increase, and the cycle starts again.

Figures 14 and 15 show the response of the four-tiered ecosystem to a -2% changes in r_x, and chart recovery paths for cases in which the perturbation is removed after 2000 and 4000 days, respectively. These two figures illustrate how important the timing of the removal of a stress can be. When the stress is removed after 2000 days there is a pronounced lag in the return path of the larger fish population. After 2000 days of recovery that population is still less than its pre-perturbation level. If the stress is removed after 4000 days, the population of larger fish returns to its original level after 2000 days, and is actually 10% above its original level after 4000 days of recovery. This does not imply, certainly, that it would be prudent to delay the clean-up of a polluted aquatic ecosystem in the hopes that recovery will be faster if one waits longer; it merely illustrates that the recovery of a perturbed ecosystem may not be a simple monotonic function of the length of time over which it has been polluted.

Our mathematical models tend to validate both the ideal and non-ideal theoretical hysteresis models. Lower trophic levels tend to return to their original levels after a relatively short recovery time, and thus show ideal hysteresis. For higher trophic levels (and especially with more severe stresses) the non-ideal hysteresis model dominates: larger organisms respond to a stress more slowly and recover more slowly, and frequently fail to return to their initial positions within a time-frame relevant to policy decisions. We should note, however, that by the nature of the mathematics used all of the populations we have modelled will eventually return to their original levels, given a sufficiently long recovery period.

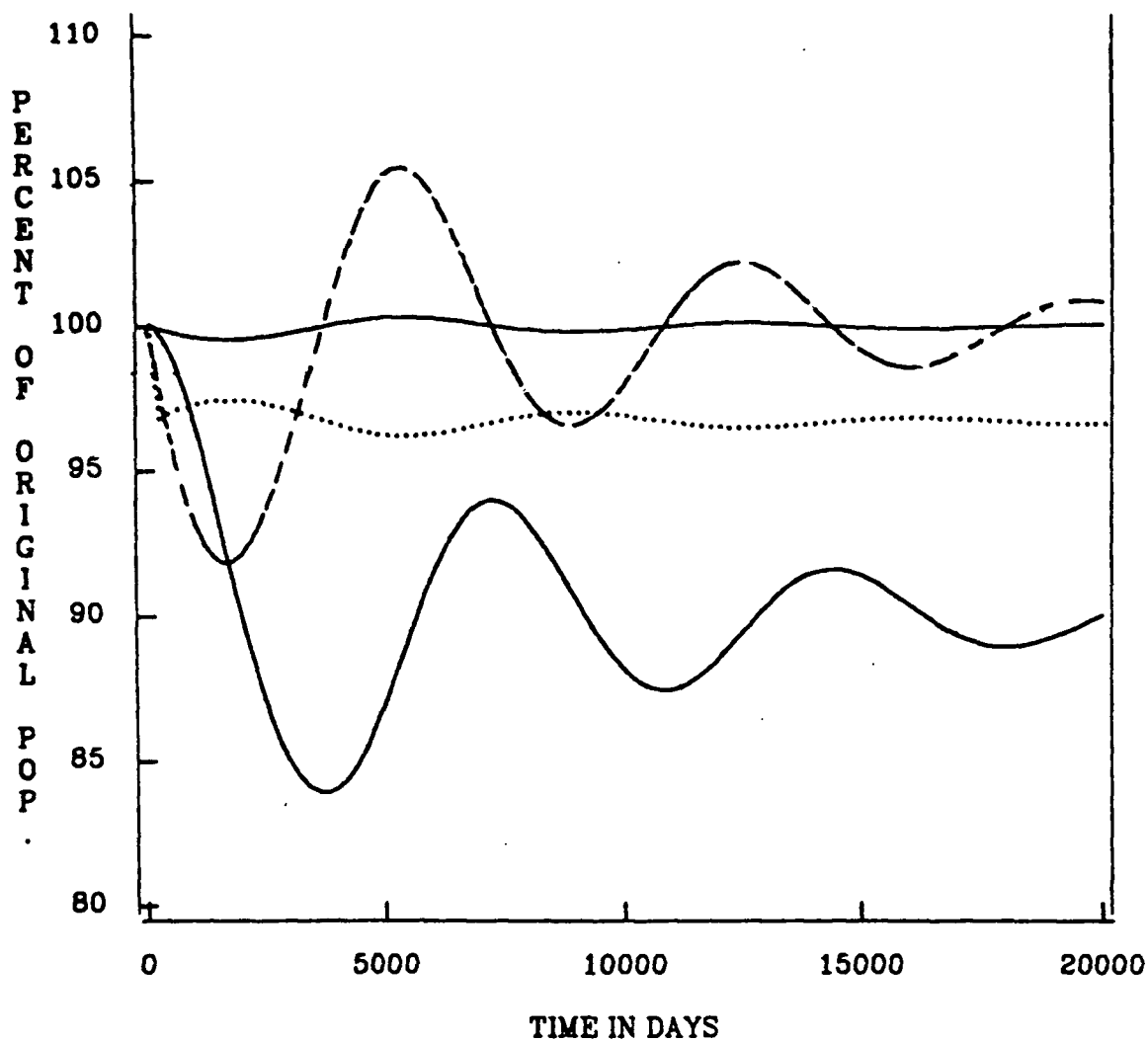**Figure 14**   Calculated time paths for the response and recovery of the populations in a four-tiered aquatic ecosystem (measured in biomass per unit area, initial biomass ratios: 500 phytoplankton: 100 zooplankton : 10 small fish: 1 larger fish) to a -2% perturbation in the phytoplankton growth rate applied at time zero and removed after 2000 days. "Response" paths are indicated by right-pointing arrows, and "recovery" paths are marked with left-pointing arrows. The paths for the responses of the phytoplankton, zooplankton, small fish, and larger fish populations are given by the upper solid curves, the dotted curves, the partially dashed curves, and the lower (more highly arched) solid curves, respectively. Note that the population of larger fish fails to return to its original position after 2000 days of recovery.
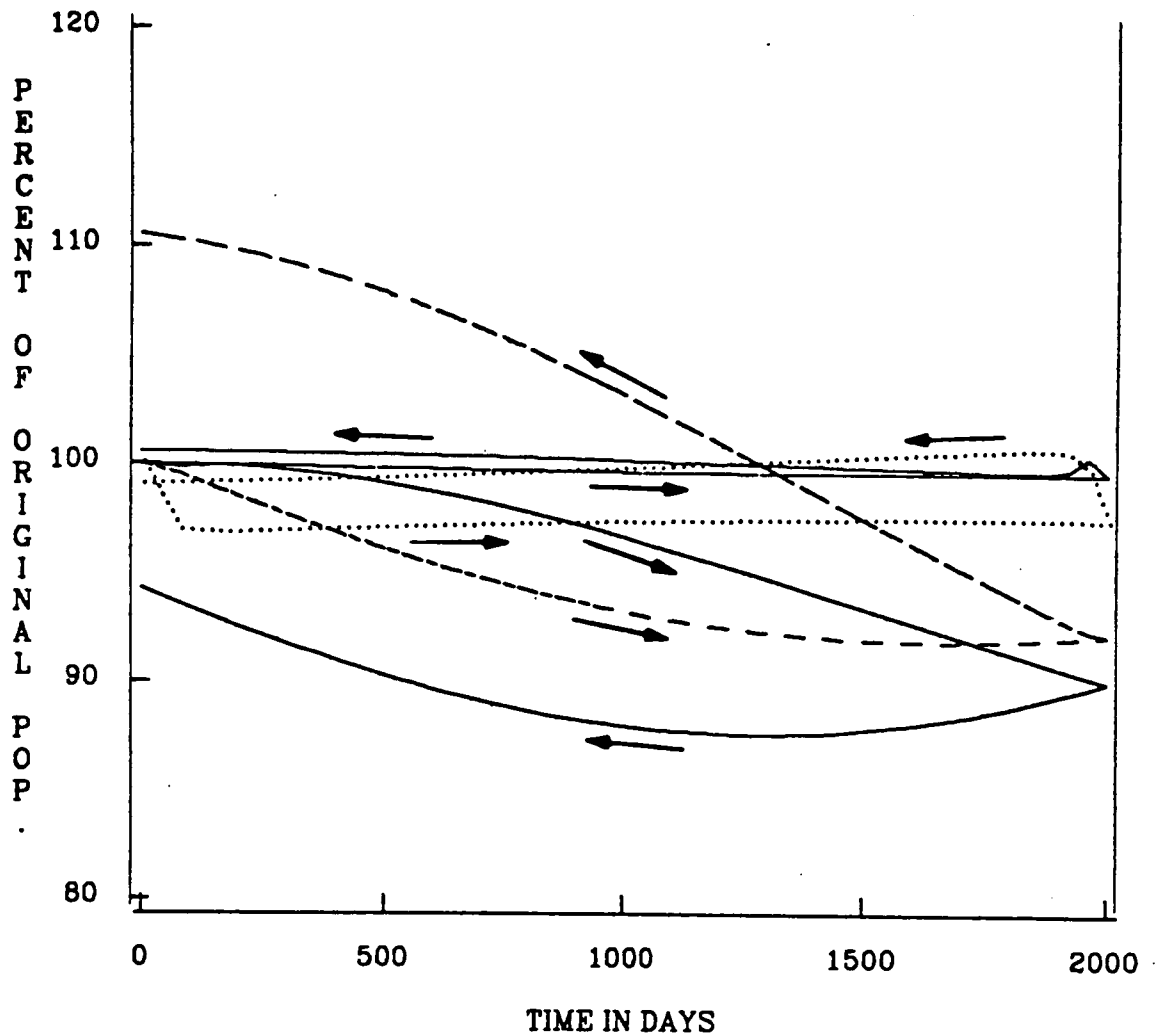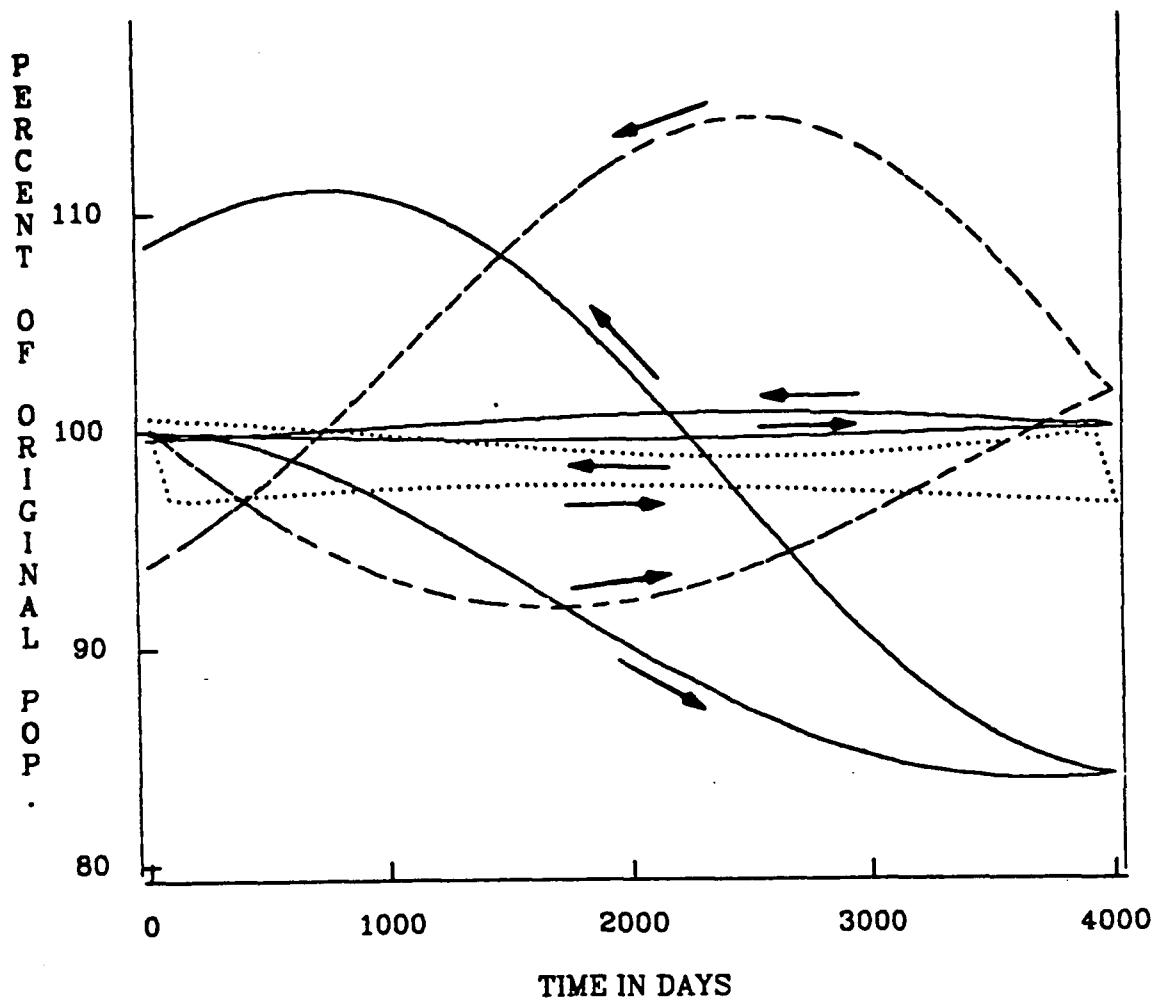
Figure 15 Calculated time paths for the response and recovery of the populations in a four-tiered aquatic ecosystem (measured in biomass per unit area, initial biomass ratios: 500 phytoplankton: 100 zooplankton : 10 small fish: 1 larger fish) to a -2% perturbation in the phytoplankton growth rate applied at time zero and removed after 4000 days. "Response" paths are indicated by right-pointing arrows, and "recovery" paths are marked with left-pointing arrows. The paths for the responses of the phytoplankton, zooplankton, small fish, and larger fish populations are given by the upper solid curves, the dotted curves, the partially dashed curves, and the lower (more highly arched) solid curves, respectively. Note that the population of larger fish returns to its original position after 2000 days of recovery and actually overshoots its level by 4000 days after the perturbation is removed.

We expect that the addition of higher trophic levels including larger, longer-lived organisms will show the non-ideal hysteresis model to be more useful for population changes occuring within a time-frame relevant to policy-making.

Discussion

Mathematical models of ecosystem perturbations are often used in ecology (Patten, 1975; O'Neill, 1976) and aquatic ecology (O'Melia, 1972; Bierman et al, 1980; Inoue et al, 1981). The drawbacks of such models are now sufficiently well understood as to allow for their restricted use.

Our mathematically-derived curves for the pollution and recovery of an aquatic ecosystem demonstrate a hysteresis effect. These curves agree closely with the ideal and non-ideal conceptual hysteresis models described above. We can use the information in our mathematically-derived curves to choose which of the conceptual models is more realistic.

The non-ideal conceptual model selected by this process is of great interest since it forecasts that the most economically valuable species, such as commercial and sports fish, will not directly and reversibly return to their original levels. This is due to the time lags that come about in part because organisms in higher trophic levels are slower to multiply and in part because increases in these levels must follow recovery of their prey populations. This type of sustained hysteresis effect is apparently inherent in ecosystems including linked trophic levels.

Our model differs from many perturbation models (e.g. O'Neill, 1976) in that we have assumed that the disturbance caused by pollution is small but continuous. This kind of small change is to be expected from "modern" pollution, where sophisticated treatment of waste is mandated and disposal of the end product of the treatment process cannot be postponed or diverted.

Sewage and industrial-waste effluents from large cities or companies are examples of such waste streams. Similarly, it is unlikely that total restoration of a grossly polluted ecosystem would be considered. Rather, a small upgrading (e.g. through control of point-sources of toxic metals, a decrease in suspended solids, or a reduction in chlorine loading) of a partially restored or partially damaged system is envisaged, as opposed to a massive ecological change. This sort of approach is typical of pollution-control strategies currently used in the U.S.

There are, however, two potential drawbacks to our simple mathematical model. First, pollution-induced changes in real aquatic ecosystems are unlikely to be quite as steady and continuous as we have modeled them. For example, many fish scarcely feed over the winter, and are thus unaffected by decreases in algae or zooplankton populations over that time period. Second, our model predicts that small fish will rather quickly be forced nearly to extinction if larger (e.g. 25%) continuous depressions of primary production are used. This is probably unrealistic due to the patchy nature of the seasonal and spatial distribution of food for higher-trophic-level organisms. We expect that some clarification of these drawbacks will result from our future comparisons of the simple Trophic-Link Model (three trophic and four levels) with a five-level version, and the comparison of both of these with real data (yet to be assembled).

Our deterministic TLM may also be insensitive to other likely ecosystem stresses that are stochastic in nature. A cool spring and summer may, for example, result in the year's juvenile fish crop being undersized at the end of the growth season, leaving them more vulnerable to cannibalism overwinter (Kipling, 1976). How would such a random event affect the hysteresis loops we have modeled, especially in the recovery phase? In progressing from a

deterministic to a stochastic modelling approach, the major difference we would anticipate would be that the position of the system would be described in probabilistic terms. For example, with respect to the -25% perturbation shown in figure 11, instead of the small fish population becoming critically low after 3000 days with a probablity of one, it might do so with a probability of 0.9, and have an additional probability of 0.1 of becoming critical at some other time. Ginzburg et al (1982) present a methodology for obtaining such extinction probabilities within the framework of a stochastic single-species population model. We intend to consider whether a similar approach is feasible for a multi-species model with realistic parameters.

We realize that the results of the HTLM are dependent on the form of the different differential equations used, the values chosen for the parameters, the method of solution of the equations, and the functional components of the ecosystem that the model describes. We intend, in fact to examine how changes in the form and parameters of HTLM's affect the results of such models. While no one trophic link model can predict the behavior of a variety of ecosystems, or even one specific ecosystem, with great certainty, we hope that advanced forms of the HTLM can be developed that can, when properly specified and calibrated with field data from a specific ecosystem, yield meaningful insights into the future behavior of that ecosystem in response to pollutant stresses. This does not mean that we believe any such model can be used to definitively predict that reducing the annual loading of compound X by 100 tons per year will result in a 5.5% increase in the number of game fish. The appropriate use for a properly calibrated model would be as an aid in making the type of yes/no choices that regulators often face. Suppose, for example, that a regulator wished to know whether or not to order the clean-up of a specific lake. If a carefully constructed and calibrated HTLM indicated that a substantial fraction of the population of an important

game fish would be likely to be lost if clean-up were delayed, the regulator might, after weighing the evidence, decide to proceed with pollution abatement. In such a case it would not matter if the model predicted a 40%, 60%, 80%, or 100% reduction in fish: the conclusion drawn by the regulator would be the same.

We feel that the simplicity of the HTLM framework will make it possible to easily calibrate models for specific situations. These models could then be run to yield qualitative information that, because of the simplicity of the models, can be traced back to allow a better understanding of the ecology behind the result.

Summary

Our initial results suggest that the hysteresis effect may be one reason why some valuable fisheries resources (e.g. the Great Lakes, where sports fisheries have failed to re-establish themselves following pollution control efforts) thave failed to respond to reduction in pollution. An understanding of hysteresis phenomena may also make it possible to predict (in an approximate way) how long it will take to see a recovery of a fish resource. An equally important application of the concept is to use it to gain a qualitative feeling for why some components of ecosystems and not others fail to show ideal hysteresis behavior and consequently become locally extinct. Further calculations using more trophic levels, different values for key parameters, and generation times derived from data on natural ecosystems, may show how useful the hysteresis concept can be for economic evaluation of pollution-control benefits that may be long delayed by ecosystem hysteresis.

References

Bierman, V.J., D.M. Dolan, E.F. Stoermer, J.E. Gannon, and V.E. Smith (1980). "The Development and Calibration of a Spatially Simplified Multi-Class Phytoplankton Model for Saginaw Bay, Lake Huron". Great Lakes Environmental Planning Study Conference No. 3. U.S. Environmental Protection Agency, Grosse Pointe, Michigan. 126 pp.

Christie, W.J. (1976). "Change in the Fish Species Composition of the Great Lakes". J. Fish. Res. Bd. Can., 31: 827-854.

Dillon, P.J., and F.H. Rigler (1974). "The Phosphorus-Chlorophyll Relationship in Lakes". Limnol. Oceanogr. 19: 767-773.

Edmondson, W.T. (1972). "The Present Condition of Lake Washington". Verh. Int. Ver. Limnol. 18: 284-291.

Edmondson, W.T. (1982). "Daphnia in Lake Washington". Limnol. Oceanogr. 27: 272-293.

Ginzburg, L.R., L.B. Slobodkin, K. Johnson, and A.G. Bindman (1982). "Quasiextinction Probabilities as a Measure of Impact on Population Growth". Risk Analysis 2: 171-181.

Goldman, C.R., and A.J. Horne (1983). Limnology. McGraw-Hill, 462 pp.

Harte J. (1985). Consider a Spherical Cow: A Course in Environmental Problem Solving. William Kaufman Inc., Los Altos, CA. 283 pp.

Hedgepath, J.W. (1977). "Models and Muddles: Some Philosophical Observations". Helgol. Wiss. Meeresunters. 30: 92-104.

Inoue, Y., S. Iwai, S. Ikeda, and T. Kunimatsu (1981). "Eutrophication of Lake Biwa--Nutrient Loadings and Ecological Model". Verh. Internat. Verein. Limnol. 21:248-255.

Kipling, C. (1976). "Year-Class Strengths of Perch and Pike in Windermere". Freshwater Biology Association Annual Report No. 44, pp. 68-75.

Malueg, K.W., R.M. Brice, D.W. Schults, and D.P. Larson (1973). The Shagawa Lake Project. U.S. EPA Report # EPA-R3-73-026, 49 pp.

May, R.M. (1973). Stability and Complexity in Model Ecosystems. Princeton University Press, 235 pp.

Numerical Algorithms Group (1984), "DO2EBF - NAG FORTRAN Library Routine Document". In NAG FORTRAN Library Manual, Mark 11, v. 1. Numerical Algorithms Group, Downers Grove, Illinois.

O'Melia, C.R. (1972). "An Approach to Modeling of Lakes". Schweiz. Z. Hydrol. 34:1-33.

O'Neill, R.V. (1976). "Ecosystem Persistence and Heterotrophic Regulation". Ecology 57:1244-1253.

Patten, B.C. (Ed.) (1975). Systems Analysis and Simulation in Ecology, Volume 3. Academic Press, N.Y., N.Y. 601 pp.

Rigler, F.H. (1976). Book Review. Limnol. Oceanogr. 21:481-483.

Vollenweider, R.A. (1969). The Scientific Basis of Lake and Stream Eutrophication, with Particular Reference to Phosphorus and Nitrogen as Eutrophication Factors. Technical Report # DAS/DSI 68.27, OECD, Paris, France.

Vollenweider, R.A., and J.J. Kerekes (1980). OECD Eutrophication Program, Synthesis Report. OECD, Paris, France.

Winberg, G.G. (1971). Methods for the Estimation of Production of Aquatic Animals. Academic Press, N.Y., N.Y., 175 pp.

APPENDIX: DETAILS OF MATHEMATICS

DETAILS OF MATHEMATICS

ASSUMPTIONS FOR THREE-LEVEL MODEL:

$$dX/dt = r_x X(1 - X/K_x) - B_{xy}XY - b_x X$$

$$dY/dt = E_{xy}B_{xy}XY - B_{yz}YZ - b_y Y$$

$$dZ/dt = E_{yz}B_{yz}YZ - b_z Z;$$

X = Phytoplankton, Y = Zooplankton, Z = Small Fish;

Steady-State Populations: $X^* = 50$, $Y^* = 10$, $Z^* = 1$;

$E_{xy} = 0.1$, $E_{yz} = 0.1$, $K_x = 100$;

Generation Times: $T_x = 3$ days, $T_y = 20$ days, $T_z = 360$ days.

$T_x = r_x^{-1}$, $T_y = (E_{xy}B_{xy}X^*)^{-1}$, $T_z = (E_{yz}B_{yz}Y^*)^{-1}$

So...

$$r_x = (T_x)^{-1} = (3)^{-1} = 1/3,$$

$$B_{xy} = 1/T_y E_{xy}X^* = (20 \times 0.1 \times 50)^{-1} = 10^{-2},$$

$$B_{yz} = 1/T_z E_{yz}Y^* = (360 \times 0.1 \times 10)^{-1} = 1/360.$$

At Steady-State:

$$r_x X^*(1 - X^*/K_x) - B_{xy}X^*Y^* - b_x X^* = 0$$

$$E_{xy}B_{xy}X^*Y^* - B_{yz}Y^*Z^* - b_y Y^* = 0$$

$$E_{yz}B_{yz}Y^*Z^* - b_z Z^* = 0.$$

So...

$$b_x = (1/3)(1 - 1/2) - (10^{-2} \times 10) = 1/6 - 0.1 = .0666667$$

$$b_y = (0.1 \times 10^{-2} \times 50) - (1/360) = 0.05 - 1/360 = 0.04722$$

$$b_z (0.1 \times 1/360 \times 10) = 1/360.$$

ASSUMPTIONS FOR FOUR-LEVEL MODEL:

$$dX/dt = r_xX(1 - X/K_x) - B_{xy}XY - b_xX$$

$$dY/dt = E_{xy}B_{xy}XY - B_{yz}YZ - b_yY$$

$$dZ/dt = E_{yz}B_{yz}YZ - B_{zf}ZF - b_zZ$$

$$dF/dt = E_{zf}B_{zf}ZF - b_fF;$$

X = Phytoplankton, Y = Zooplankton, Z = Small Fish, F = Larger Fish;

Steady-State Populations: $X^* = 500$, $Y^* = 100$, $Z^* = 10$, $F^* = 1$;

$E_{xy} = 0.1$, $E_{yz} = 0.1$, $E_{zf} = 0.1$, $K_x = 1000$;

Generation Times: $T_x = 3$ days, $T_y = 20$ days, $T_z = 360$ days, $T_f = 1080$ days.

$T_x = r_x^{-1}$, $T_y = (E_{xy}B_{xy}X^*)^{-1}$, $T_z = (E_{yz}B_{yz}Y^*)^{-1}$, and $T_f = (E_{zf}B_{zf}Z^*)^{-1}$

So...

$$r_x = (T_x)^{-1} = (3)^{-1} = 1/3,$$

$$B_{xy} = 1/T_yE_{xy}X^* = (20 \times 0.1 \times 500)^{-1} = 10^{-3},$$

$$B_{yz} = 1/T_zE_{yz}Y^* = (360 \times 0.1 \times 100)^{-1} = 1/3600, \text{ and}$$

$$B_{zf} = 1/T_fE_{zf}Z^* = (1080 \times 0.1 \times 10)^{-1} = 1/1080.$$

At Steady-State:

$$r_xX^*(1 - X^*/K_x) - B_{xy}X^*Y^* - b_xX^* = 0$$

$$E_{xy}B_{xy}X^*Y^* - B_{yz}Y^*Z^* - b_yY^* = 0$$

$$E_{yz}B_{yz}Y^*Z^* - B_{zf}Z^*F^* - b_zZ^* = 0, \text{ and}$$

$$E_{zf}B_{zf}Z^*F^* - b_fF^* = 0.$$

So...

$$b_x = (1/3)(1 - 1/2) - (10^{-3} \times 100) = 1/6 - 0.1 = .0666667$$

$$b_y = (0.1 \times 10^{-3} \times 500) - (1/3600 \times 10) = 0.05 - 1/360 = 0.04722$$

$$b_z (0.1 \times 1/3600 \times 100) - (1/1080) = 1/360 - 1/1080 = 1/540, \text{ and}$$

$$b_f (0.1 \times 1/1080 \times 10) = 1/1080.$$

```
c          This program, which incorporates the NAG subroutine do2ebf, can
c          be used to solve three coupled differential equations.
c             ..scalars in common
           implicit double precision (a-h,o-z)
           double precision H, xend
           integer I
c             ...
c             ..local scalars..
           double precision tol, x
           integer Ifail, IR, IW, mped, nout
c             ..local arrays..
           double precision W(3,21), Y(3)
c             ..subroutine references..
c           d02ebf
c             ..
           external fcn, out, pederv
           common xend, H, I
           open(8, file='output')
c          opens file, named "output", in which results are to be placed
           data nout /6/
           write (nout,99996)
           write (8,99996)
           write (nout,99994)
           write (8,99994)
           N = 3
           IW = 21
           MPED = 0
           IR = 2
           tol = 10.0d0**(-5)
           write (nout,99999) tol
           write (nout,99998)
           write (8,99999) tol
           write (8,99998)
           x = 0
           xend = 1.0d4
c          Program is now set to calculate a "response" path.  To calculate a
c          "return" path one would substitute post-perturbation values for
c          y(1-3) below
           y(1) = 50.0d0
           y(2) = 10.0d0
           y(3) = 1.0d0
           H = (xend-x)/50
c          Prints out solution at 49 evenly spaced points between x(0) and xend
           I = 49
           Ifail = 1
           call D02EBF(x, xend, N, y, tol, IR, fcn, mped, pederv,
          * out, W, IW, Ifail)
           write (nout,99997) Ifail
           write (8,99997) Ifail
```

A-3

```fortran
      subroutine pederv(x, y, PW) ·
c        ..scalar arguments..
      double precision x
c        ..array arguments..
      double precision PW(3,3), y(3)
c        ..
      PW(1,1) = -1.00d0*2.0d0*(1.0d0/(3.0d0*100.0d0))*y(1) +
     +  + 1.00d0*(1.0d0/3.0d0) - 0.06666666666666667d0
     +  - (1.0d0/100.0d0)*y(2)
      PW(1,2) = (1.0d0/100.0d0)*y(1)
      PW(1,3) = 0.0d0
      PW(2,1) = (1.0d0/100.0d1)*y(2)
      PW(2,2) = -(1.0d0/36.0d1)*y(3) - (4.72222222222d-2) +
     +     (1.0d0/100.0d1)*y(1)
      PW(2,3) = -(1.0d0/36.0d1)*y(2)
      PW(3,1) = 0.0d0
      PW(3,2) = (1.0d0/36.0d2)*y(3)
      PW(3,3) = (1.0d0/36.0d2)*y(2) - 1.0000d0*(1.0d0/3.6d2)
      return
      end
      subroutine out(x, y)
c        ..scalar arguments..
      double precision x, u
c        ..array arguments..
      double precision y(3)
      double precision z(3)
c     u allows time to be counted "backwards" (for return paths), while z(3)
c     is a set of variables that allow the populations, y(t), to be normalized
c     with respect to one another.  The equations for z(1-3) below express
c     each y(t) as a percentage of the initial population in that trophic level
c        ..
c        ..scalars in common..
      double precision H, xend
      integer I
c        ..
c        ..local scalars..
      integer J, nout
c        ..
      common xend, H, I
      data nout /6/
      z(1) = y(1)/0.5d0
      z(2) = y(2)*10.0d0
      z(3) = y(3)*1.0d2
      u = 1.0d4 - x
      write (nout,99999) x, (z(J),J=1,3)
      write (8,99999) x, (z(J),J=1,3)
      x = xend - dble(I)*H
      I = I - 1
      return
99999 format (1H , F8.2, 3E13.5)
      end
```

```fortran
      If (tol.lt.o) write (nout,99995)
      If (tol.lt.o) write (8,99995)       .
   20 continue
      mped = 1
c      mped = 1 indicates that routine is using supplied Jacobian (in PEDERV)
c      rather than calculating it internally (which happens when mped = 0)
      write (nout,99993)
      write (8,99993)
      tol = 10.0d0**(-5)
      write (nout,99999) tol
      write (8,99999) tol
      write (8,99998)
      write (nout,99998)
      x = 0
      xend = 1.0d4
      y(1) = 50.0d0
      y(2) = 10.00d0
      y(3) = 1.0d0
      H = (xend-x)/50
      I = 49
      Ifail = 1
      call D02EBF(x, xend, N, y, tol, IR, fcn, mped, pederv,
     * out, W, IW, Ifail)
      write (nout,99997) Ifail
      If (tol.lt.o) write (nout,99995)
      write (8,99997) Ifail
      If (tol.lt.o) write (8,99995)
   40 continue
      stop
99999 format (22h0CALCULATION WITH TOL=, e8.1)
99998 format (40h T AND SOLUTION AT EQUALLY SPACED POINTS )
99997 format (8h Ifail= I1)
99996 format (4(1x/), 31h D02EBF EXAMPLE PROGRAM RESULTS      x)
99995 format (24h RANGE TOO SHORT FOR TOL)
99994 format (32h0CALCULATING JACOBIAN INTERNALLY)
99993 format (31h0CALCULATING JACOBIAN BY PEDERV)
      end
      subroutine fcn(T, y, F)
c         ..scalar arguments..
      double precision T
c         ..array arguments..
      double precision F(3), y(3)
c         ..
      F(1) = 1.00d0*(1.0d0/3.0d0)*y(1)*(1.0d0-(y(1)/100.0d0)) -
     +   (1.0d0/100.0d0)*y(1)*y(2)
     + - 0.0666666666666667d0*y(1)
      F(2) = (1.0d0/100.0d1)*y(1)*y(2) - ((1.0d0/36.0d1)*y(2)         )) -
     + (4.722222222d-2)*y(2)
      F(3) = (1.0d0/36.0d2)*y(2)*y(3) - 1.00d0*(1.0d0/3.6d2)
c      Program is now set at steady state.   To model a perturbation
c      in the phytoplakton growth rate, replace "1.00d0" in the expression
c      for F(1) (and also in the expression for PW(1,1), below)
c      with, for example, "0.98d0" (for a 2% decrease)
      return
      end
```

A-5

```
c         ..scalars in common
          implicit double precision (a-h,o-z)
          double precision H, xend
          integer I
c         ...
c         ..local scalars..
          double precision tol, x
          integer Ifail, IR, IW, mped, nout
c         ..local arrays..
          double precision W(4,22), Y(4)
c         ..subroutine references..
c         d02ebf
c         ..
          external fcn, out, pederv
          common xend, H, I
          open(8, file='output')
c         Places the output of this program into a file named "output"
          data nout /6/
          write (nout,99996)
          write (8,99996)
          write (nout,99994)
          write (8,99994)
          N = 4
          IW = 22
          MPED = 0
          IR = 2
          tol = 10.0d0**(-5)
          write (nout,99999) tol
          write (nout,99998)
          write (8,99999) tol
          write (8,99998)
          x = 0
c         Program is now set to calculate time paths starting with steady-state
c         conditions.   To calculate "return" paths, replace the values of
c         y(1-3) below with post-perturbation values
          xend = 2.0d4
          y(1) = 500.0d0
          y(2) = 100.0d0
          y(3) = 10.0d0
          y(4) = 1.0d0
          H = (xend-x)/50
          I = 49
          Ifail = 1
          call D02EBF(x, xend, N, y, tol, IR, fcn, mped, pederv,
         * out, W, IW, Ifail)
          write (nout,99997) Ifail
          write (8,99997) Ifail
          If (tol.lt.o) write (nout,99995)
          If (tol.lt.o) write (8,99995)
       20 continue
c         This section, which is optional, calculates time points based on values
c         of the Jacobian matrix of the system supplied in "PEDERV", below
          mped = 1
          write (nout,99993)
          write (8,99993)
```

A-6

```fortran
      tol = 10.0d0**(-5)
      write (nout,99999) tol
      write (8,99999) tol
      write (8,99998)
      write (nout,99998)
      x = 0
      xend = 2.0d4
      y(1) = 500.0d0
      y(2) = 100.0d0
      y(3) = 10.0d0
      y(4) = 1.0d0
      H = (xend-x)/50
      I = 49
      Ifail = 1
      call D02EBF(x, xend, N, y, tol, IR, fcn, mped, pederv,
     * out, W, IW, Ifail)
      write (nout,99997) Ifail
      If (tol.lt.o) write (nout,99995)
      write (8,99997) Ifail
      If (tol.lt.o) write (8,99995)
   40 continue
      stop
99999 format (22hCALCULATION WITH TOL=, e8.1)
99998 format (40h T AND SOLUTION AT EQUALLY SPACED POINTS)
99997 format (8h  Ifail= I1)
99996 format (4(1x/), 31h D02EBF EXAMPLE PROGRAM RESULTS/1x)
99995 format (24h RANGE TOO SHORT FOR TOL)
99994 format (32hCALCULATING JACOBIAN INTERNALLY)
99993 format (31hCALCULATING JACOBIAN BY PEDERV)
      end
      subroutine fcn(T, y, F)
c        ..scalar arguments..
      double precision T
c        ..array arguments..
      double precision F(4), y(4)
c        ..
c     To calculate response to a perturbation in the phytoplankton growth rate,
c     replace "1.00d0" in F(1), and PW(1,1) below with, for example "0.98d0"
c     (for a -2% perturbation
      F(1) = 1.00d0*(1.0d0/3.0d0)*y(1)*(1.0d0-(y(1)/100.0d1)) -
     +  (1.0d0/100.0d1)*y(1)*y(2)
     + - 0.066666666666667d0*y(1)
      F(2) = (1.0d0/100.0d2)*y(1)*y(2) - ((1.0d0/36.0d2)*y(2)*y(3)) -
     + (4.72222222222d-2)*y(2)
      F(3) = (1.0d0/36.0d3)*y(2)*y(3) - ((1.0d0/1080.0d0)*y(3)*y(4)) -
     +  1.00d0*(1.0d0/5.4d2)*y(3)
      F(4) = (1.0d0/1080.0d1)*y(3)*y(4) - ((1.0d0/1080.0d0)*y(4))
      return
      end
```

```fortran
      subroutine pederv(x, y, PW)
c        ..scalar arguments..
      double precision x
c        ..array arguments..
      double precision PW(4,4), y(4)
c        ..
      PW(1,1) = -1.00d0*2.0d0*(1.0d0/(3.0d0*100.1d0))*y(1) +
     + + 1.00d0*(1.0d0/3.0d0) - 0.06666666666666667d0
     + - (1.0d0/100.0d1)*y(2)
      PW(1,2) = (1.0d0/100.0d1)*y(1)
      PW(1,3) = 0.0d0
      PW(1,4) = 0.0d0
      PW(2,1) = (1.0d0/100.0d2)*y(2)
      PW(2,2) = -(1.0d0/36.0d2)*y(3) - (4.72222222222d-2) +
     +     (1.0d0/100.0d2)*y(1)
      PW(2,3) = -(1.0d0/36.0d2)*y(2)
      PW(2,4) = 0.0d0
      PW(3,1) = 0.0d0
      PW(3,2) = (1.0d0/36.0d3)*y(3)
      PW(3,3) = (1.0d0/36.0d3)*y(2) - 1.00d0*(1.0d0/5.4d2) -
     +     (1.0d0/1080.0d0)*y(4)
      PW(3,4) = (1.0d0/1080.0d0)*y(3)
      PW(4,1) = 0.0d0
      PW(4,2) = 0.0d0
      PW(4,3) = (1.0d0/1080.0d1)*y(4)
      PW(4,4) = (1.0d0/1080.0d1)*y(3) - (1.0d0/1080.0d0)
      return
      end
      subroutine out(x, y)
c        ..scalar arguments..
      double precision x, u
c        ..array arguments..
      double precision y(4)
      double precision z(4)
c     "u" allows time to be counted "backwards" for return time paths;  z(1-4)
c     is a set of variables that allow the time points for y(1-4) to be
c     expressed as percentages of the initial populations in each trophic
c     level
c        ..
c        ..scalars in common..
      double precision H, xend
      integer I
c        ..
c        ..local scalars..
      integer J, nout
c        ..
      common xend, H, I
      data nout /6/
      z(1) = y(1)/0.5d1
      z(2) = y(2)
      z(3) = y(3)*1.0d1
      z(4) = y(4)*1.0d2
      u = 2.0d3 - x
      write (nout,99999) u, (z(J),J=1,4)
      write (8,99999) u, (z(J),J=1,4)
      x = xend - dble(I)*H
      I = I - 1
      return
99999 format (1H , F8.2, 4E13.5)
      end
```

CHAPTER 4

Ecotoxicology and Benefit-Cost Analysis:
The Role of Error Propagation

Introduction

An understandable desire exists on the part of policy makers to devise a set of procedures, an analytical approach, that can be used to guide policy. Such an approach would obviate the need for trusting to historical practice, or to the intuition of wise but inevitably fallible and probably biased individuals, or to the awkward and time-consuming process of making every decision by plebiscite. It would "rationalize" policy making and, if the procedure were appropriately chosen, optimize the well-being of the affected sector of the public. Pollution abatement policy is a prime example, for it is here that a vigorous effort is underway to promote benefit-cost analysis as the appropriate analytical approach for determining proper emission levels (see U.S. Executive Order 12291).

Despite the advantages in efficiency of decision making, and possibly in enhancement of societal welfare, that may accrue to a society that employs the benefit-cost approach to set pollution emission levels, there are major pitfalls lurking that need to be identified and discussed. These pitfalls fall into two categories: limitations in the ability of ecologists to describe precisely the ecological consequences of pollutant emission rates, and limitations in the ability of economists to describe precisely the economic consequences of ecological changes.

Quite generally, the economic and ecological analyses that are required to characterize and quantify costs and benefits of particular pollutant abatement strategies consist of a sequence of steps. Table 1 shows what a typical sequence of steps would have to look like for a believable benefit-cost

1

Change in a polluting activity
(e.g., placement of scrubbers
in power plants)

1. ↓ (combustion science)

Change in emission levels
(e.g. reduction in SO_2 output)

2. ↓ (atmospheric sciences)

Change in primary stress on
ecosystem (e.g., increase in pH
of precipitation at a particular
watershed)

3. ↓ (biogeochemistry)

Change in secondary stresses (which
act directly on biological populations
and processes) (e.g., increase in pH
of surface waters and soils)

4 ↓ (biological toxicology)

Direct biological effects of changes
in secondary stresses (e.g., increase
in populations of acid-sensitive
plankton)

5. ↓ (ecology)

Indirect ecological changes
stimulated by the direct biological
effects (e.g., improvement in
fish productivity)

(environmental sciences,
sociology, ...)

6.

Change in pattern of direct use of ecosystem
(e.g., fishermen flock to site)

Change in indirect ecological benefits to
society (e.g., hydrologic integrity of
watershed is enhanced, leading to reduction
in fluctuations of water supplies to people)

Direct market value of
changed use patterns and
of indirect benefits (e.g.,
value of user-day fees and
additional water supply);
value of other benefits
(e.g., feelings of civic
accomplishment, spiritual
satisfaction)
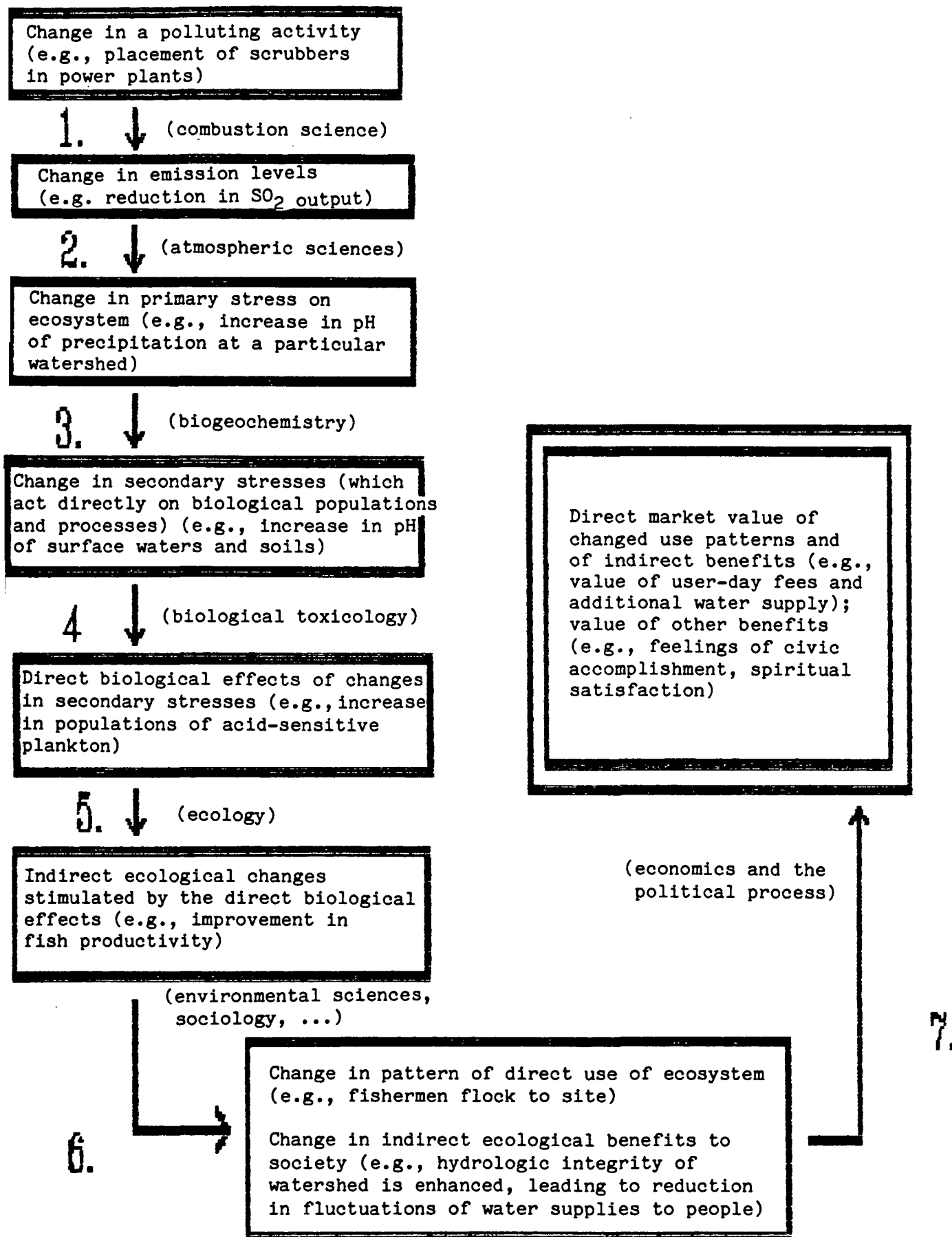
(economics and the
political process)

7.

Table 1. The stages of ecosystem impact assessment

analysis, with the example of acid rain used to provide specificity. The information that must be used to quantify any given step in the sequence must come from analysis at the preceding stage. Thus the possibility exists that error may propagate through the sequence to the point where the final output--for example, the economic benefit of a particular level of pollution abatement--is so uncertain as to be of little or no use in a benefit-cost analysis or related procedure.

Whether or not this occurs will depend in part on the degree to which ecologists and other environmental scientists can characterize the uncertainty in a manner that can be used by economists. To take a simple example, consider the statement that the decrease in fish mortality following pollution abatement in a particular lake is uncertain. This statement may mean that the decrease in mortality cannot be predicted accurately but that the odds of any specified degree of decrease in mortality are known (from some combination of measurement and modeling). Or it may mean that only the range of uncertainty is known but that the probabilities of any particular value of mortality within that range are not known. In the former case, economists may be able to estimate an expected value of benefit of any particular degree of abatement (using methods such as those described elsewhere in this report), whereas in the latter case the opportunity to characterize the benefit of any particular degree of abatement is considerably more limited.

In the remainder of this chapter we discuss in a systematic and general manner the subject of error propagation in environmental impact assessment, with an emphasis on impacts involving ecosystems. We deduce some general results about error propagation that are independant of the method of analysis. One key result is that error tends to "biomagnify" in ecological food chains, so that a small degree of uncertainty about the effect of a pollutant on the lowest trophic level is likely to translate into much more

3

substantial uncertainty about the effects on higher trophic levels, in which we are often more interested. We also explore the origin of some of the most refractory types of error in impact assessment. To relate the analysis to the specific concerns of practitioners of economic evaluation we also show how the relevant issue is not merely one of the magnitude of the range of uncertainty but also of the type of uncertainty; this is because economic analysis, which must begin where ecological analysis leaves off, can cope with some kinds of uncertainties better than others. Of particular concern in the context of benefit-cost analysis is the degree to which sources of ecological uncertainties can be characterized in ways that will be of use to economists. The overall dimensions and a few critical elements of this problem are discussed here, but it will be shown that considerable work on the part of ecologists will be necessary to bridge the gap between what is now known and what needs to be known to provide a plausible underpinning for the successful application of benefit-cost methods of decision-making.

Uncertainty in Impact Assessment: an Example

Examples of error propagation in environmental science abound. Consider the acid rain example from Table 1. Analysts have attempted to establish the existence and value of a threshold level of precipitation pH, below which lakes would become acidic and above which the natural restorative capacity of lakes and surrounding soils would afford protection. The existence of such a threshold would make the task of setting standards easier because such a threshold would provide a natural level to aim for--tightening the standard beyond the threshold would lead to diminishing returns.

However, uncertainties in impact assessment render the threshold notion a highly dubious one in this context. It is likely, in fact, that one's

perception of the location of the threshold for a particular class of lakes depends on how long one has been observing those lakes under various levels of exposire; whereas precipitation with a pH of, say, 4.5 might acidify the lakes in 10 years, precipitation with a higher pH of, say, 4.9 might acidify the lakes in 30 years, a period longer than anyone has had the opportunity to observe. Thus the threshold concept is time-dependent and intrinsic uncertainty characterizes its evaluation

The threshold value for one class of lakes might not be of much use for others. For example, in eastern North America it has been pointed out that over several decades, the period over which observations have been made, lakes receiving precipitation with a pH of less than about 4.7 have had their chemistry altered by the precipitation. Even if we accept this relatively short time-frame for that particular group of lakes, there is still uncertainty as to the value of this "threshold" in other areas. In the mountains of the western United States, for example, the susceptibility of lakes to acidification appears to be greater than in watersheds of the northeastern U.S. (Roth et al, 1985). A more complete discussion of uncertainties plaguing the use of the threshold concept in ecotoxicology is found in Cairns and Harte (1985).

Even if we had confidence in the location of a pH threshold, we would still not know exactly what the effect on precipitation pH would be for any specified emissions reduction plan. Here the uncertainty stems from the complexity of the source-receptor relation.

The uncertainty in deducing the effect of a particular level of emissions reduction on precipitation pH must be combined with the further uncertainty in deducing the effect of a reduction in precipitation pH on surface water acidity. By combining these two uncertainties, the overall uncertainty in steps 2 to 4 of Table 1 can be determined. At the other stages in the impact

5

assessment further opportunity for error arises. The combined error is almost invariably sufficiently large to make it difficult to obtain a precise characterization of the ecological benefits from a particular emissions-reduction plan.

The fact that one cannot precisely characterize the benefits of a pollution-abatement policy should not he taken to mean that the policy is unwarranted. Even though an economic analysis might not produce a reliable cost-benefit ratio, it can lead to a range of uncertatinty in that ratio, which can then be evaluated through the political process to determine what policy action is warranted. The first step, however, must be to have a systematic approach to the analysis of uncertainty; this is discussed in the following section.

A Framework for Analysis

The sequence of steps in an environmental impact assessment as shown on the left hand side of Table 1 provides a convenient framework for analysing the propagation of error in such assessments. Generally, the relation between the ith and the $i+1$st stage in the sequence is likely to look like one of the three graphs shown in figure 1. In each of the graphs, the horizontal axis represents the variable describing the ith stage and the vertical axis represents the subsequent one down the chain. The first of these three graphs illustrates a linear relation, in which the response, or output, at the subsequent stage is proportional to the input from the one before, as, for example, if the loss of organisms is proportional to the concentration of a pollutant. The second one illustrates a threshold process, in which an output is only weakly dependent on an input for small values of the input, but when the input exceeds a critical value, then the output rises sharply. The

6
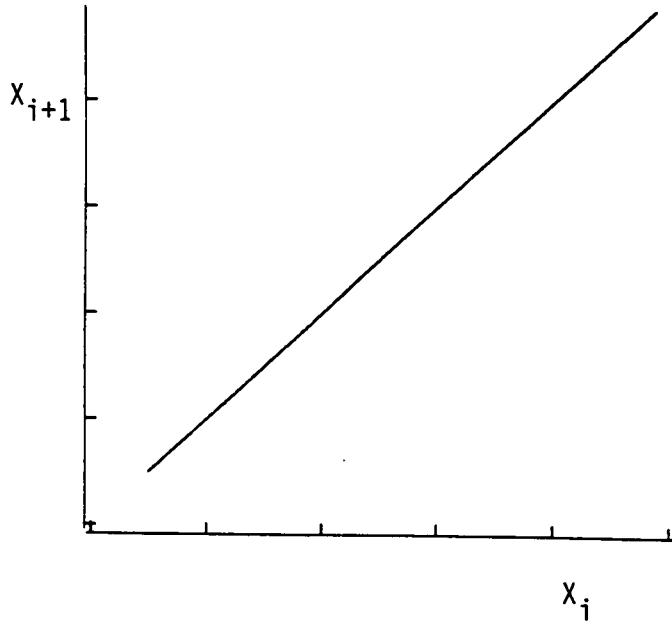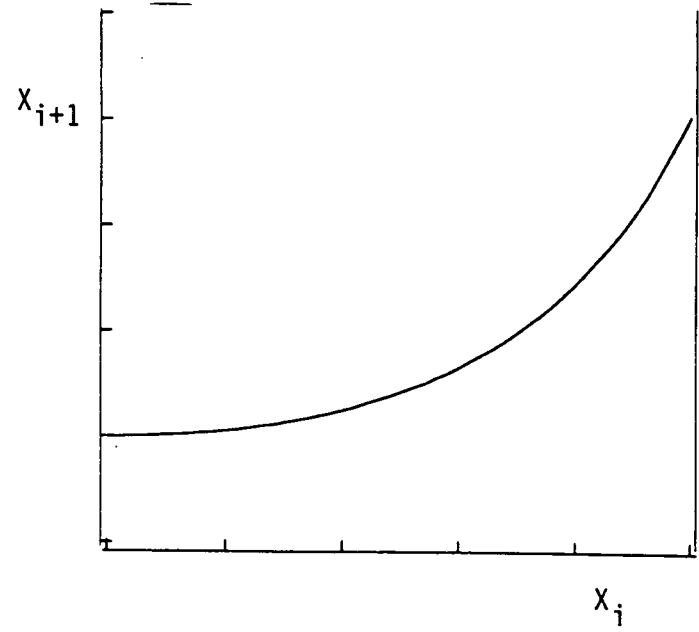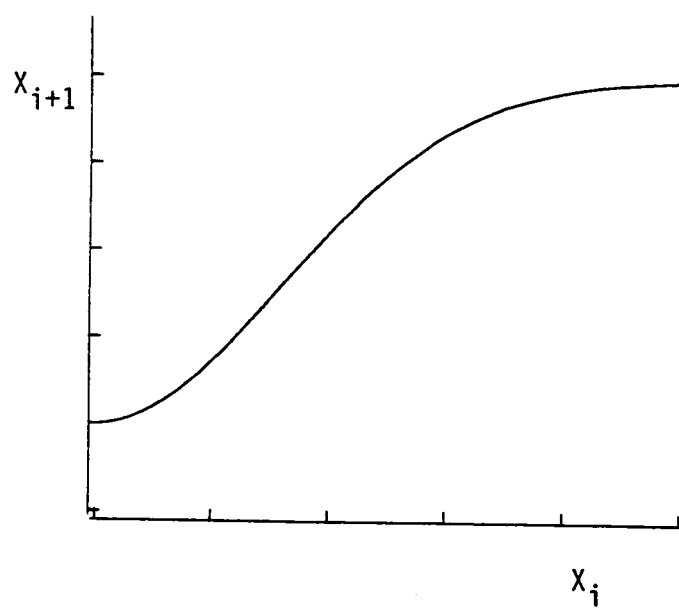
Figure 1

Illustration of a linear (a), a threshold (b), and a saturation (c) process relating variables describing successive stages in the assessment chain.

third graph in Figure 1 illustrates a saturation process, in which an output ceases to be strongly dependent on input once the input exceeds a critical value.

These three basic types of relations between sequential stages in the impact chain can be modified or combined to describe, generically, most real processes. For example, the graphs can be turned upside down to describe processes in which an output is a decreasing function of input. Or graphs 1-b and 1-c can be combined to describe a process with a threshold at a relatively low value of the input and a saturation effect at a higher one.

If knowledge of the functional relation between two sequential stages in the chain were complete, and the input data were known with perfect precision and accuracy[*] then a graph of the function describing the relation might, indeed, look something like one of the plots in Figure 1, But, in reality, there is always uncertainty in both knowledge of functional relations and in the data needed to substitute into those functions. These uncertainties will propagate down the impact chain, sometimes leading to a surprisingly high level of uncertainty at the end.

Two types of uncertainty were alluded to above. one results from poor knowledge of the dynamics of the processes--i.e. uncertainty in our understanding of the form of the relation between variables--and one results from incertain numerical values for data. For example, suppose that we are interested in estimating the uncertainty in our knowledge of the lessening of damage to plankton populations due to an expected decline in the rate of input of a pollutant to a lake. Because it is difficult to predict with high

----
[*] "Precision" refers to the detail with which a number is expressed--the number of significant figures. "Accuracy" refers to how close the number is to the true, or real, value. Thus if I state my height is 3.47258 meters, I am being precise but inaccurate. Oftentimes authors will substitute precision for accuracy, providing more significant figures than the data deserve and giving the illusion that they are highly accurate.

accuracy how the concentration of a pollutant in a lake will respond to a change in the input rate, there will be uncertainty in our knowledge of what the concentration of pollutant in the lakewater will be. On top of that we will have, at best, only partial knowledge of how the plankton population will respond to any precisely stated change in the pollutant concentration. In other words, even with perfectly accurate data describing the pollutant, our knowledge of the functional form of the relation between pollutant concentration and plankton survivability is uncertain.

Because of the uncertainty in our knowledge of functional relations, the graphs shown in Figure 1 must be modified as in Figure 2. Furthermore, because the input data (the horizontal axis variable) are likely to be uncertain, the output (the vertical axis variable) is also going to have an uncertainty that reflects the fuzziness of the input data. At each stage in the chain, the uncertainty may be amplified or damped as uncertainty in the output from one stage becomes uncertainty in the input to the next. Figure 3 provides a generic illustration of how the error will propagate down the chain. The range of uncertainty is shown to broaden in the figure, a result of the width and steepness of the functional forms assumed. If probability distributions characterizing the likelihood of the parameters taking on particular values within the range of uncertainty are known, then a more sophisticated analysis can be carried out; shown here is the simpler case in which only the propagation of the range of uncertainty is described.
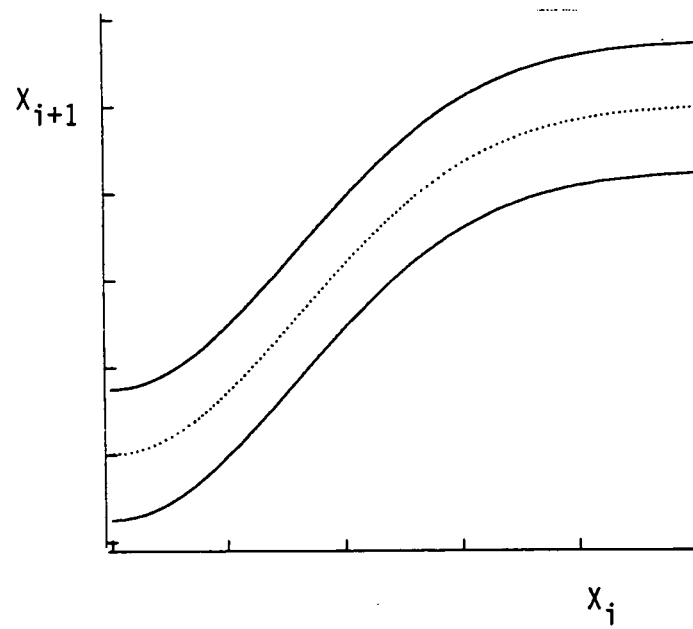
A useful analysis of the consequences for policy makers of this sort of error propagation is given in Reckhow (1984). In the following section, we discuss some general results about uncertainty that can be deduced from the above considerations.
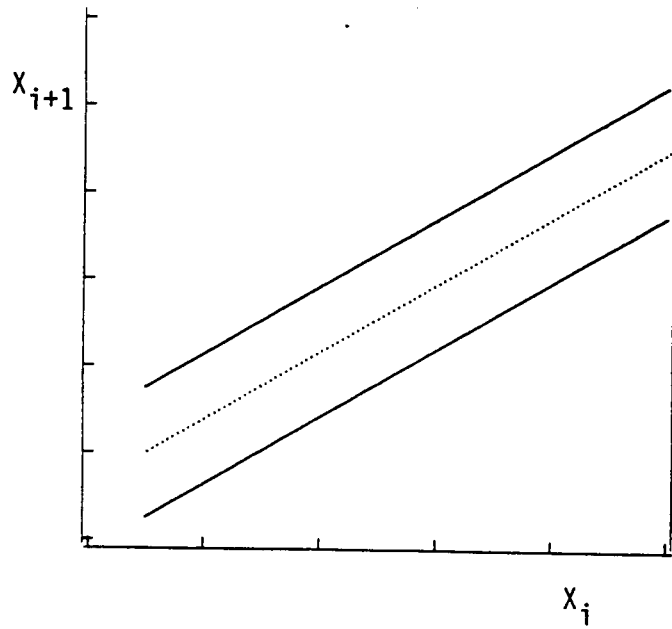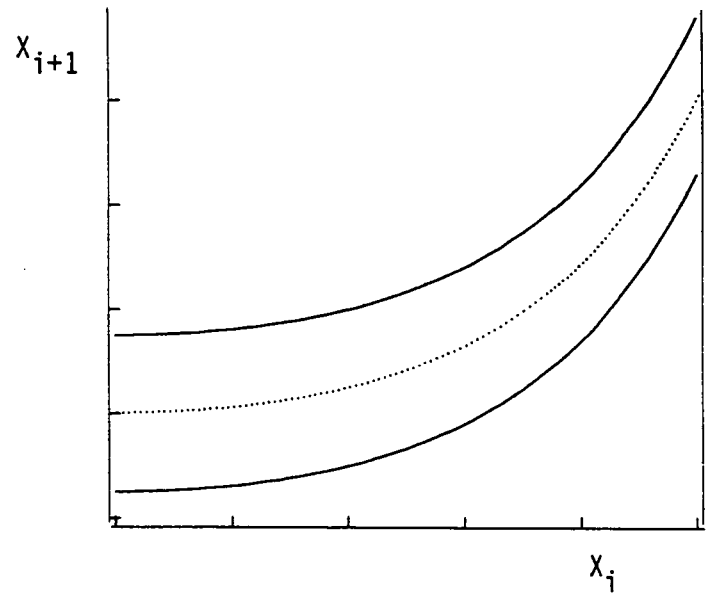
Figure 2

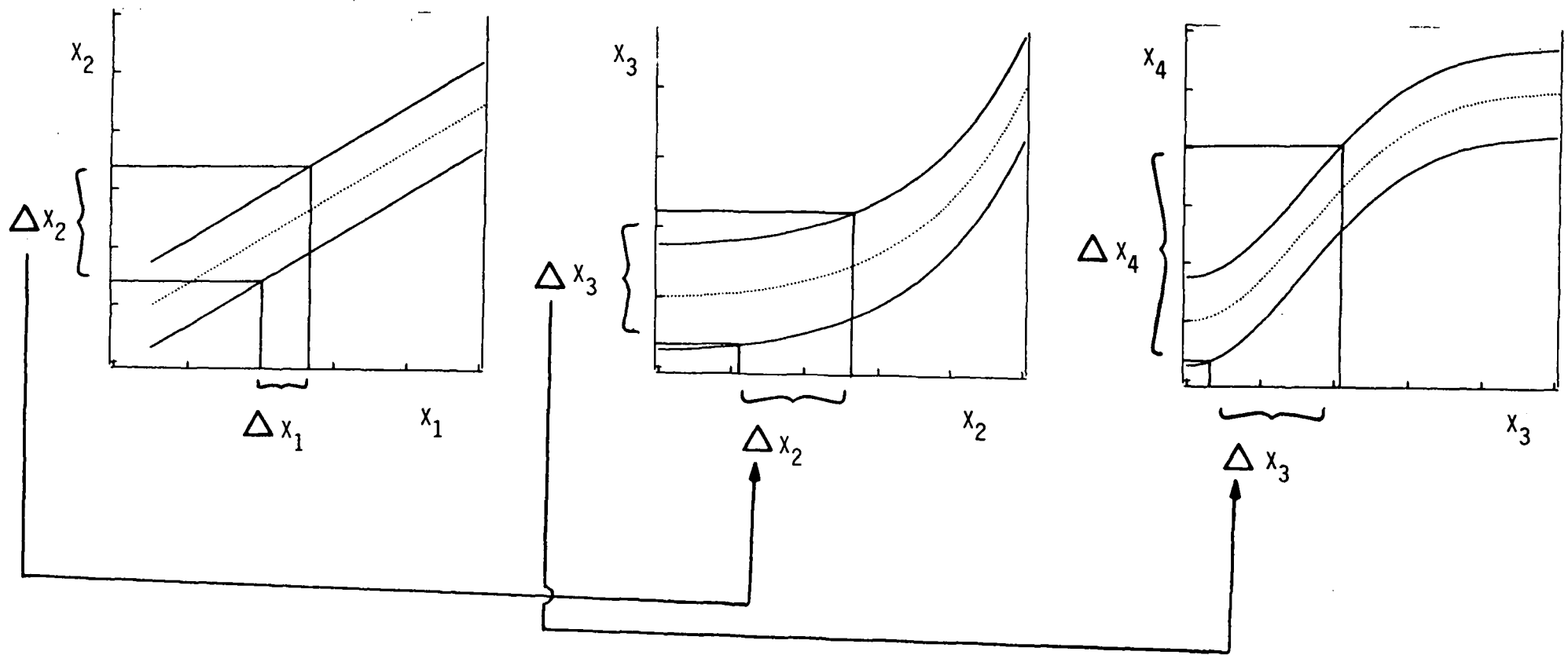Examples of error bands in the curves shown in Figure 1.

Figure 3

Illustration of the propagation of error along the assessment chain. In each graph, the uncertainty in X_i is "passed along" to X_{i+1} in the manner shown.

General Results: The "Fallacy of the Mean" and "Error Biomagnification"

Quantities such as fish productivity or water clarity, indeed any parameter
to which a numerical range can be ascribed, can be characterized by a mean
value and a range of uncertainty about that mean.   Because it is much simpler
to focus on a mean value, which is a single number, rather than on the range
of uncertainty, which is at the very least a range of numbers (often with a
complicated interpretation attached explaining what that range really refers
to) it is not uncommon for analysts to be asked questions such as "if I take
the mean value of the pollutant concentration and substitute that into the
formula relating concentration to plankton survivability, then what mean value
will I obtain for plankton survivability?"   This question reflects a
fundamental confusion: a function evaluated at the mean value of its
independent variable is generally not equal to the mean value of the function.
Indeed, as shown below, considerable error can result if mean values are
estimated by commiting this "fallacy of the mean".

How will the general shape of the graph (as is Figure 1) of the relation
between two successive stages in impact assessment influence the error
committed by assuming that a function of the mean equals the mean of the
function?   Figure 4 illustrates the answer to this question.   In this
figure, the parameter, a, has an equal probability of lying anywhere in the
range from B to C and its mean is midway between at E.   At the upper end of
this range, x(a) takes on the value D while at the lower end it takes on the
value A.   As the figure shows, if the relation between an independent
variable, a, and a dependent variable, x, is linear, then despite uncertainty
in our knowledge of a, the mean value of x, denoted by $\bar{x}$, is equal to x(a)
evaluated at $\bar{a}$, the mean value of a.   In equation form, $\bar{x} = x(\bar{a})$.   For the
case of a threshold-type relation, this figure shows why $\bar{x} > x(\bar{a})$, while for a
saturation process, $\bar{x} < x(\bar{a})$.

12
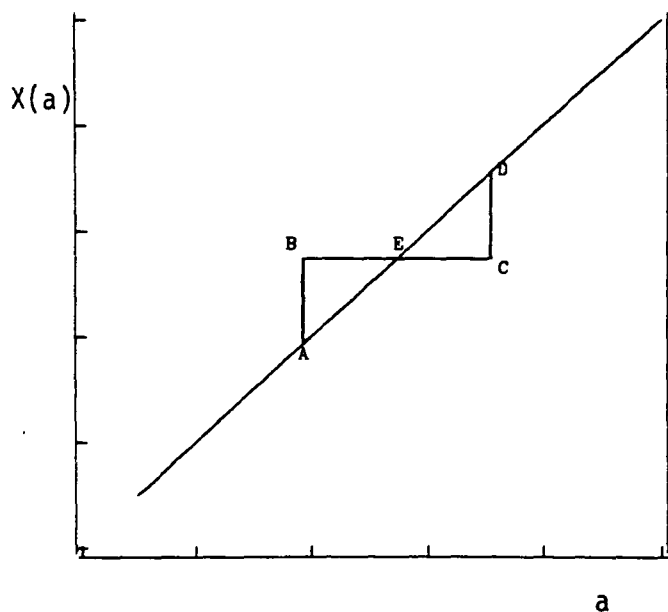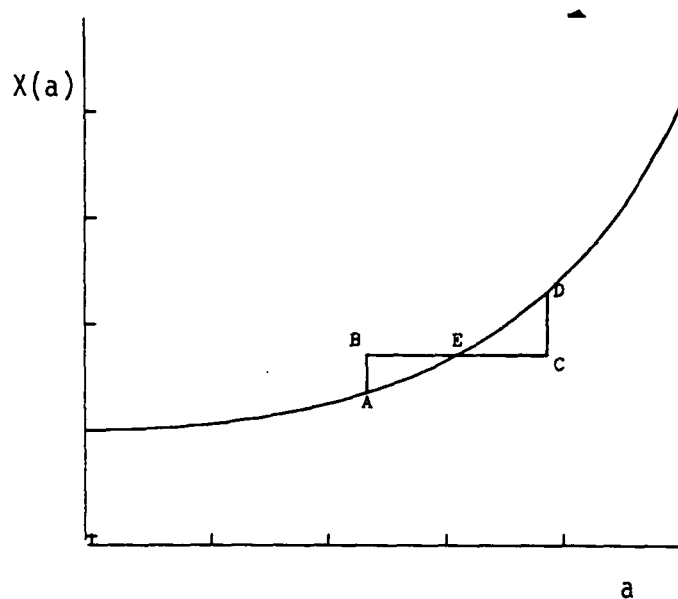
a.   IF BE=EC, THEN AB=CD

b.   IF BE=EC, THEN DC>BA
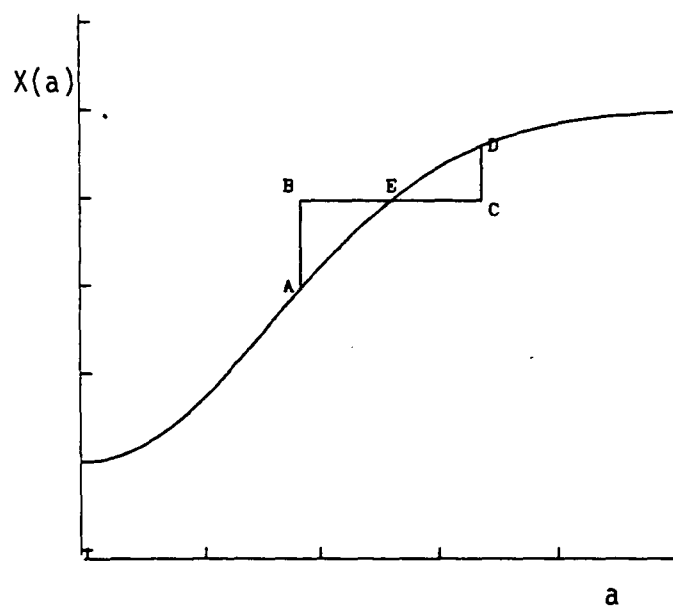
c.   IF BE=EC, THEN AB>CD

Figure 4

The relation between the mean value of X and the value of X evaluated at the
mean value of the parameter, a, upon which it depends, is shown for the
three cases of a linear (a), upward curving (b), and downward curving (c)
relation between X and a.

13

This can be very important in practice; for relations characterized by very steeply curved functions, the use of the mean value of the independent variable for evaluating the mean value of the dependent one can lead to a gross under- or over-estimation, depending on the type of curvature in the functional relation. To illustrate this, we present the following example.

The attenuation of light with depth in a relatively transparent lake obeys a simple formula: $I(d) = I_0 \exp(-\nu d)$, where $I(d)$ is the intensity at depth d, I_0 is the intensity of light at the surface, and ν is a constant characterizing the transparency of the water. The more opaque the water, the larger the value of ν. Primary productivity of aquatic plants at any particular depth will be roughly proportional to the value of I at that depth, although it also depends, of course, on concentrations of essential nutrients such as nitrate and phosphate. Suppose siltation results in a large value of ν. We will assume that the mean value of ν is 0.3/meter and that the range of uncertainty is $\pm$ 0.02/meter. We will interpret this range to mean (for the sake of simplicity) that the actual value of ν is equally likely to lie anywhere in the range from 0.28 to 0.32/meter. Suppose erosion control is expected to reduce the value of ν to 0.17 $\pm$ 0.09, with the range of uncertainty increased because it is not known how effective the control program will be. At a depth of, say, 20 meters, the mean value of I prior to the erosion control that would be calculated (incorrectly) by substituting the mean value of ν into the formula for $I(d)$ is $I_0 \exp(-6.0)$ or $0.0025\ I_0$. After the control is implemented, the similarly incorrect value is $I_0 \exp(-3.4) = 0.033 I_0$, an increase of I by a factor of about 12. However, if the actual mean value of I is calculated properly, not by substituting into $\exp(-\nu d)$ the mean value of ν but rather averaging over the range of uncertainty in ν, then we find that erosion control results, on the average, in twice as great an increase in mean light intensity at 20 meters. Leaving

14

aside subtleties such as whether plants respond to the average light intensity they receive or to some more complicated value that depends on the fluctuations, there is clearly a large potential for error in naively estimating mean values by being oblivious to the uncertainties.

We emphasize that the propagation of error by this means can result either from a situation where one knows what the uncertainties are but uses the incorrect formula relating mean values, or from a situation where one simply under- or overestimates the magnitudes of the uncertainties but uses a correct averaging procedure for estimating mean values.

In the modular approach to error propagation discussed in the previous section, there is an opportunity for errors of this type to either be reinforced or to cancel. If a sequence of relations between the variables describing the successive stages in the impact chain are all of, say, the threshold type, or more generally, of any similar curvature, then the error propagation that results from ignorance of the true range of uncertainty will be reinforcing, leading to greater and greater error as one moves along the chain. In contrast, if curves of types 1.b and 1.c from Figure 1 are equally represented in the chain, then the tendency will be for the errors of that type to cancel.

Next, we turn to the topic of "error biomagnification". Error, like many a toxic substance, will frequently increase as one probes higher up the food chain (not to be confused with the impact assessment chain in Fig. 1), although the mechanism that accounts for error biomagnification is quite different from that for toxic substance biomagnification. To see how error biomagnification arises, consider the following relatively simple model for a food chain. Figure 5 illustrates the model, showing the inflows and outflows of biomass from each link in the chain. The links can be thought of as
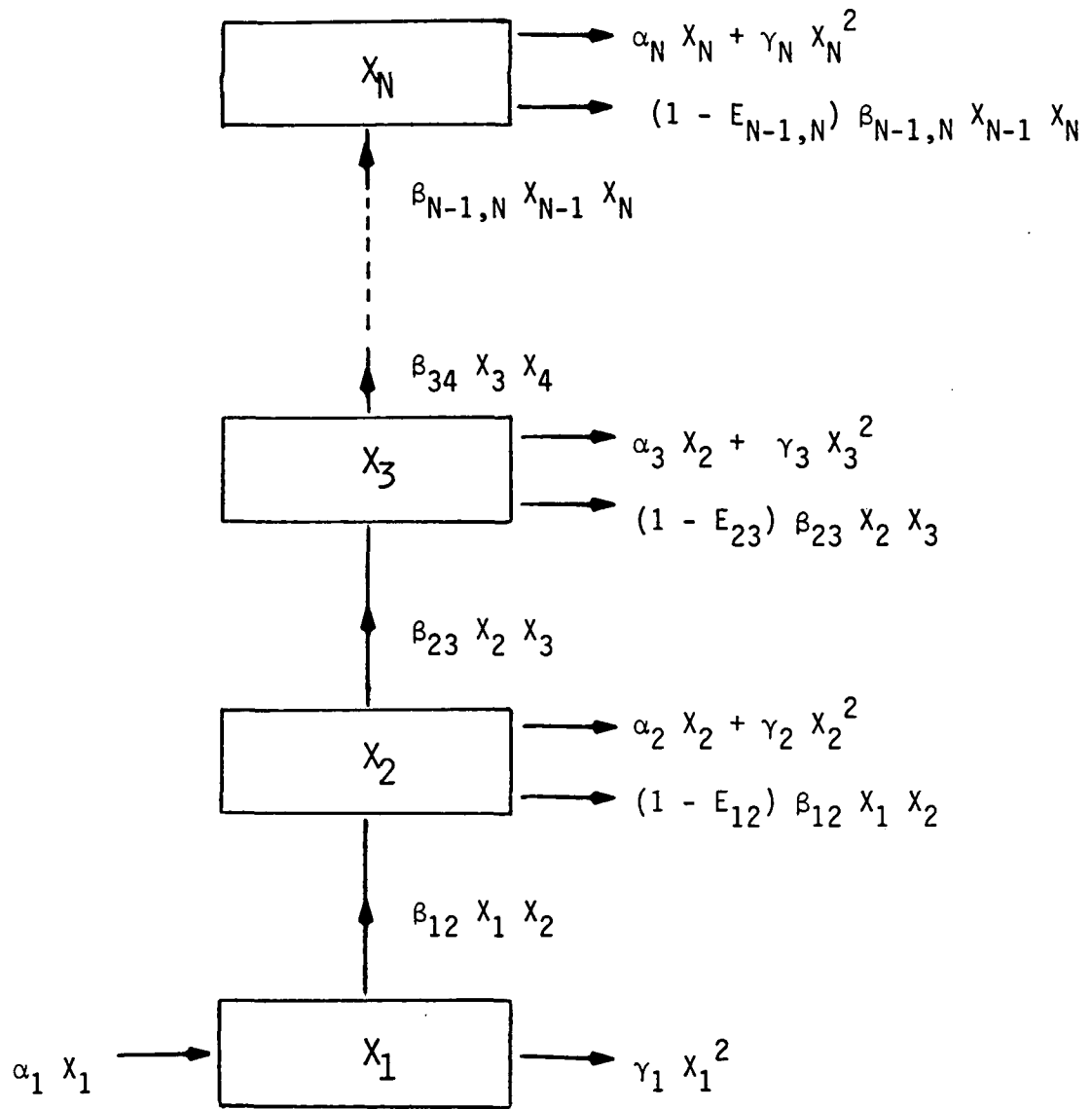
Figure 5

A trophic chain and the rates of biomass input and output from each link in the chain as described by a simple Lotka-Volterra model.

species (for example, grass, which is eaten by rabbits, which are eaten by lynx, etc.) or as functional groupings of species (for example, primary producers, herbivores, first carnivores, . . . and on up to top carnivores). In equation form, the model reads as follows:

$$\frac{dX_1}{dt} = \alpha_1 X_1 - \gamma_1 X_1^2 - \beta_{12} X_1 X_2$$

$$\frac{dX_2}{dt} = E_{12} \beta_{12} X_1 X_2 - \alpha_2 X_2 - \gamma_2 X_2^2 - \beta_{23} X_2 X_3$$

$$\frac{dX_3}{dt} = E_{23} \beta_{23} X_2 X_3 - \alpha_3 X_3 - \gamma_3 X_3^2 - \beta_{34} X_3 X_4$$

$$\vdots$$

$$\frac{dX_N}{dt} = E_{N-1,N} \beta_{N-1,N} X_{N-1} X_N - \alpha_N X_N - \gamma_N X_N^2.$$

In these equations, the X_i are the biomasses of the components; the coefficients β_{ij} are rate constants describing the predation of species j upon species i; the coefficients E_{ij} describe the efficiency of incorporation of prey biomass by the predator; and the coefficients α_i and γ_i are growth and death rates for the individual species. The presence of the γ_i terms represents a negative feedback mechanism induced by the finite carrying capacity of any realistic environment. They result in steady-state solutions that are stable against perturbations such as the removal of some percentage of the biomass of the system. Indeed, the only solution to these equations is one in which all the X_i approach time-independent values. Although real populations are not found in steady-state (that is, the numbers of individuals in real populations generally exibit both cyclic and random time dependence), models with steady-state solutions are often used to study the

time-averaged behavior of such populations. Although simple models of this sort are generally unreliable for making detailed predictions of the values of the variables, $X_i(t)$, they are useful for exploring the qualitative features of ecosystems.

Suppose that the growth rate of the primary producers is affected by a pollutant, but that there is some uncertainty about the magnitude of the effect. In other words, suppose that the value of α_1 is known only to be in the range between $\bar{\alpha}_1 + \sigma$ and $\bar{\alpha}_1 - \sigma$ where $\bar{\alpha}_1$ is the mean value and σ is a measure of the uncertainty in the mean. How will the uncertainty in affect the uncertainty in the steady-state values of the individual variables, X_i? A simple two-level model illustrates the general idea:

$$\frac{dX_1}{dt} = \alpha_1 X_1 - \gamma_1 X_1^2 - \beta_{12} X_1 X_2$$

$$\frac{dX_2}{dt} = E_{12} \beta_{12} X_1 X_2 - \alpha_2 X_2 - \gamma_2 X_2^2.$$

For this case the steady-state solutions for the X_i are:

$$X_1 = \frac{\alpha_2 \beta_{12} + \alpha_1 \gamma_2}{E_{12} \beta_{12}^2 + \gamma_1 \gamma_2} \quad \text{and}$$

$$X_2 = \frac{\alpha_1 E_{12} \beta_{12} - \alpha_2 \gamma_1}{E_{12} \beta_{12}^2 + \gamma_1 \gamma_2}.$$

A measure of the relative uncertainty in the X_i caused by the uncertainty in α_1 is $(\sigma/X_1)(\partial X_1/\partial \alpha_1)$. Thus the ratio of the relative uncertainty in X_1 to that in X_2, which we denote by R_{12}, is

$$R_{12} = \frac{(\sigma/X_1)(\partial X_1/\partial \alpha_1)}{(\sigma/X_2)(\partial X_2/\partial \alpha_1)}.$$

This can be shown to equal $(\gamma_2 X_2)/(\alpha_2 + \gamma_2 X_2)$, which is less than unity. In

other words, the relative error in X_2 induced by the uncertainty in α_1 is necessarily greater than that for X_1. For this two-level model, if the uncertainty lies in our knowledge of α_2, the parameter characterizing the death rate of the predator rather than the growth rate of the prey, then the result is ambiguous; the value of R_{12} will depend on the relative magnitude of α_1 and $\gamma_1 X_1$. In particular, if the latter term is not small compared to the former, then again R_{12} will be less than unity. Thus in this two-level model, if the uncertainty lies in our characterization of the base of the food chain, then uncertainty "biomagnifies" up the chain, whereas if it lies at the top of the chain, then it may or may not magnify down the chain.

Results from three-level and four-level models are shown in Figure 6, both for the case in which the original uncertainty lies at the base of the food chain and the error propagates up to higher trophic levels and for the case in which the original uncertainty lies at the top of the food chain and the error "bounces off" the base and propagates back up. Note how a relatively small initial error in either the phytoplankton growth rate or the fish death rate results in progressively larger uncertainty as one progresses up the food chain.

It would be of considerable interest to characterize the system properties that determine the degree to which error "biomagnification" occurs. It is likely that properties of the food chain such as the ratios of biomasses or population densities at successive levels and the ratio of predation rates to other death rates will be important factors in more complex situations than in the grossly over-simplified models treated here.

The implications of this for ecological impact assessment can be of great importance. The interest of the public is usually in the higher levels of the food chain--be it fish for recreation and food or exotic wildlife for

Figure 6a

The response of the populations in a three-tiered aquatic ecosystem (measured in biomass per unit area, initial biomass ratios: 50 phytoplankton: 10 zooplankton : 1 small fish) to -1%, -2%, and -3% changes in the phytoplankton growth rate. Solid, dotted, and (partially) dashed lines give the paths for phytoplankton, zooplankton , and small fish, respectively. This figure corresponds to a situation in which the degree of perturbation in the growth rate, caused, for example by pollution, is uncertain, but is known to lie within some range. The effect of this uncertainty on the relative magnitudes of population changes in the three trophic levels is shown.
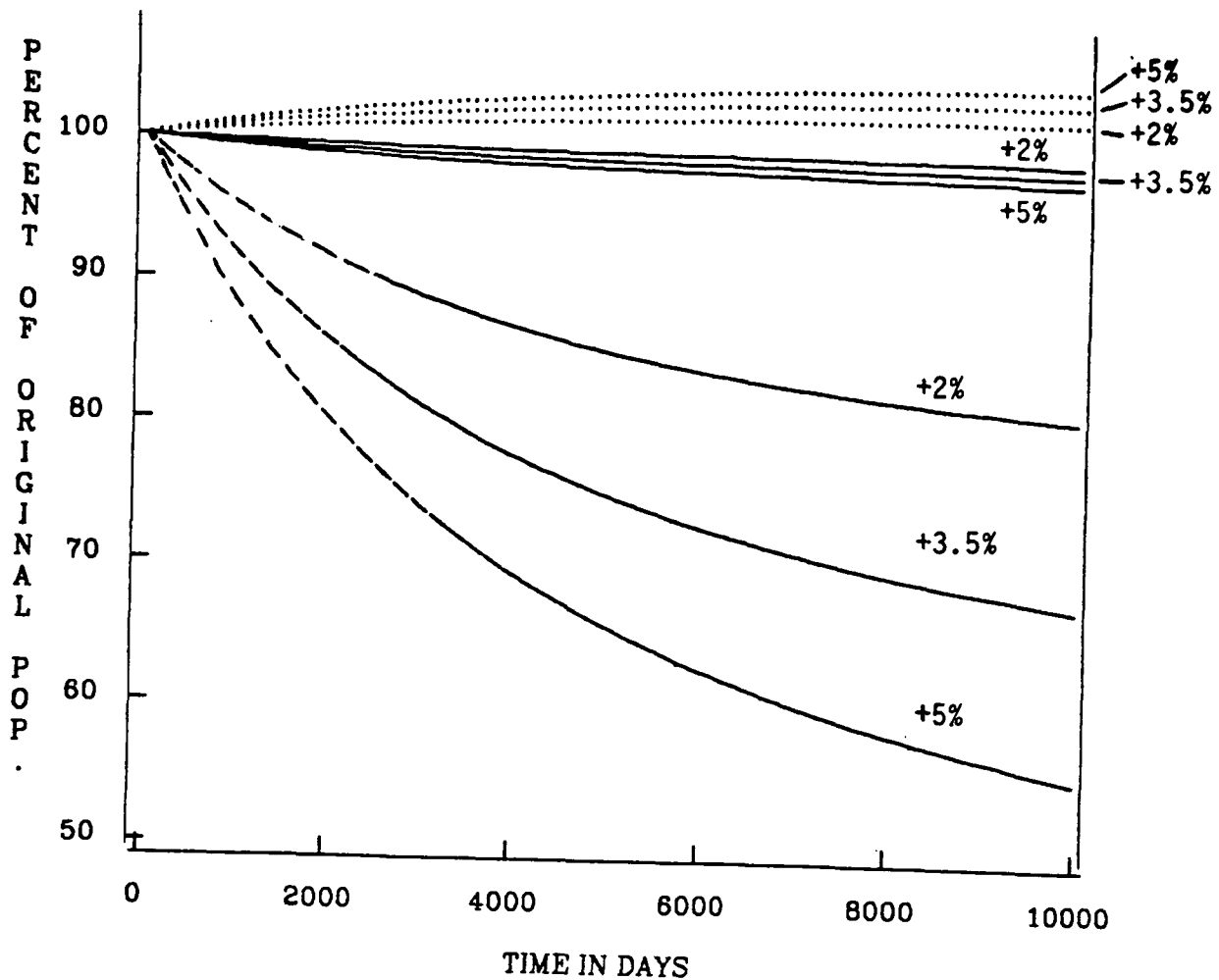
Figure 6b

The response of the populations in a three-tiered aquatic
ecosystem (measured in biomass per unit area, initial biomass
ratios: 50 phytoplankton: 10 zooplankton : 1 small fish) to +2%,
+3.5%, and +5% changes in the rate at which fish die off.   Solid,
dotted, and (partially) dashed lines give the paths for
phytoplankton, zooplankton , and small fish, respectively.   This
figure corresponds to a situation in which the degree of
perturbation in the die-off rate is uncertain, but is known to lie
within some range.   The effect of this uncertainty on the
relative magnitudes of population changes in the three trophic
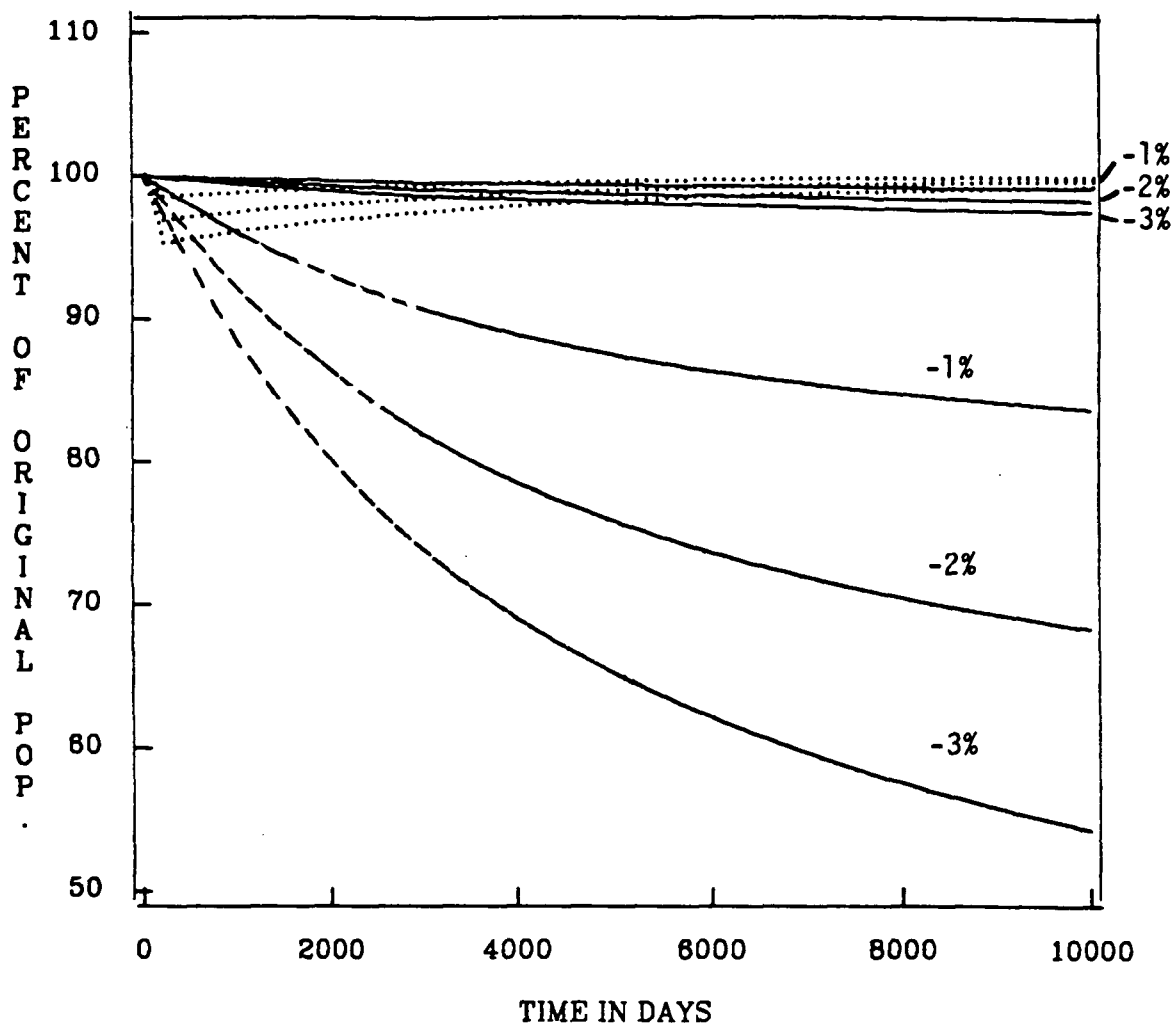levels is shown.
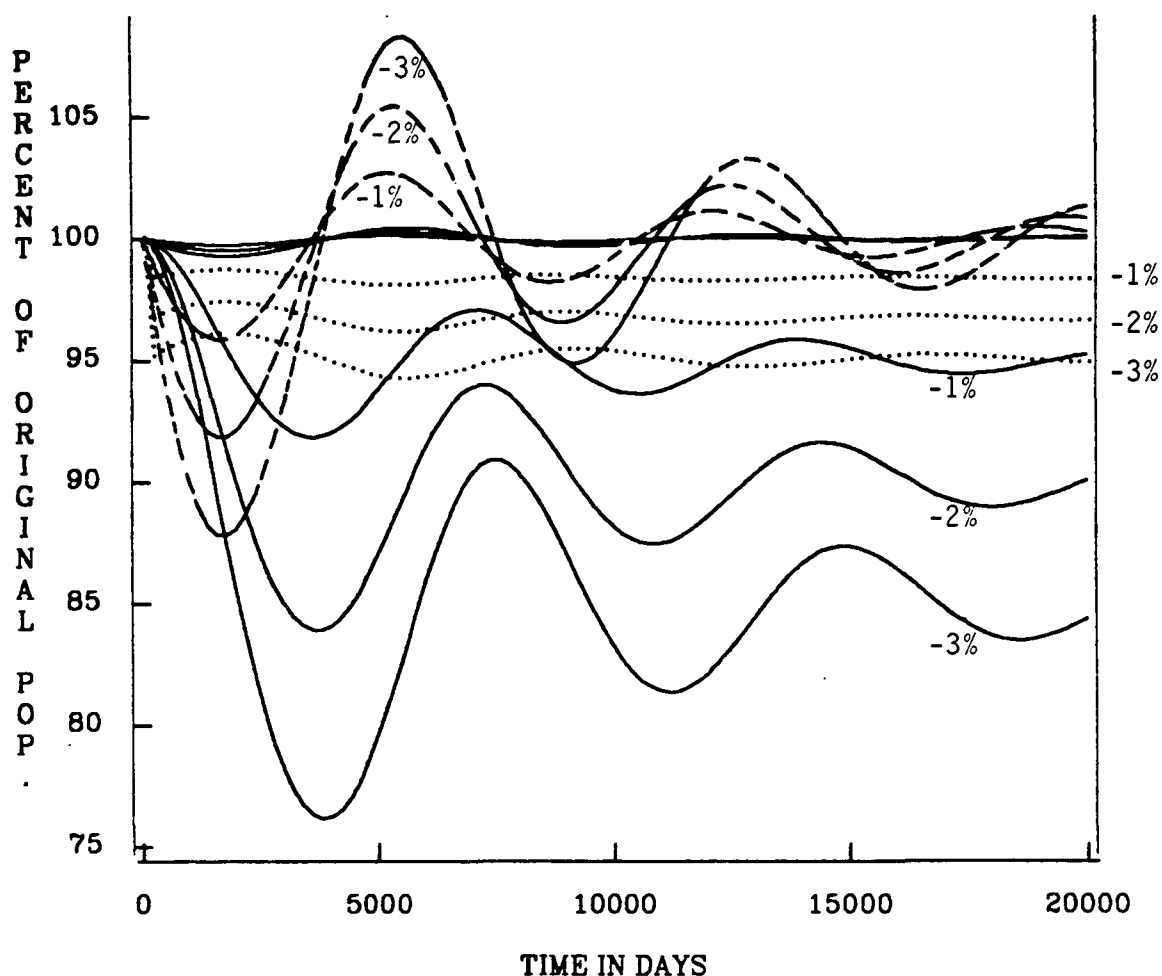
Figure 6c

The response of the populations in a four-tiered aquatic ecosystem (measured in biomass per unit area, initial biomass ratios: 500 phytoplankton: 100 zooplankton : 10 small fish: 1 larger fish) to -1%, -2%, and -3% changes in the phytoplankton growth rate. The paths for the responses of the phytoplankton, zooplankton, small fish, and larger fish populations are given by the upper solid curve, the dotted curve, and partially dashed curve, and the lower solid curve, respectively. This figure corresponds to a situation in which the degree of perturbation in the growth rate, caused, for example by pollution, is uncertain, but is known to lie within some range. The effect of this uncertainty on the relative magnitudes of population changes in the four trophic levels is shown.

nature study (or the public itself, which occupies the top carnivore spot in the global ecosystem!). The increase in error as it propagates up the chain will tend to render difficult the prediction of the magnitude of precisely those effects that the public is most concerned about. While an enormous effort is sometimes expended trying to determine precisely the environmental concentration of a pollutant, the effort may be misplaced if error propagation leads to large uncertainties higher up in the food chain where the public welfare is more directly and obviously involved.

Like toxic substance biomagnification, this magnification of error is unavoidable. It is a consequence of the fundamental ecological dynamics of a food chain and can not be circumvented. Like toxic substance biomagnification, whose effects at the higher trophic levels can be minimized by keeping the level of the toxicant in the environment to a minimum, the effect of error propagation up a food chain can be minimized by keeping to a minimum the initial error in our knowledge of the effect of the toxicant on the growth of the primary producers.

We have not discussed here the question raised in the Introduction concerning the probability distribution of the quantity of interest within its range of uncertainty. As mentioned previously, when a parameter such as a fish population is uncertain, but a probability distribution for it is calculable, then economic valuation is easier than when such a probability distribution is unknown. Consider an uncertainty in the effect of a toxicant on the growth rate of a species of phytoplankton, as in our simple food chain model, that has the characteristic that the error in our knowledge of it is gaussian-distributed. What will the distribution of biomagnified error be in the fish population? Unfortunately, no general statement that is model-independent can be made about this at present. The particular, unabashedly unrealistic, model used to motivate the existence of the phenomenon of error

23

biomagnification provides a precise answer to this question, but other models will generally provide other answers. Because we lack confidence in any particular model or class of models for the analysis of complex ecosystems, further work is clearly needed here.

Since our ability to characterize ecological uncertainty with probability distributions is presently limited, it might seem like a sensible strategy for ecologists to place more emphasis on reducing the range of uncertainty. As we show in the following Section, that approach, too, has its limits and, indeed, they are even more stubborn than are the problems discussed heretofore.

Refractory Error in Ecology

Some types of uncertainty in impact assessment are easily remedied. If a few more observers spend a little more time gathering data or improving their models, a noticeable improvement will result and these remediable types of errors will be eliminated or at least greatly reduced in magnitude. A more interesting class of errors can not be pushed to zero, however, or even significantly reduced in magnitude regardless of how much effort is expended to do so. These are the refractory or intrinsic uncertainties whose origin we now discuss. In a general sense, they stem from two sources: uniqueness and sensitivity to initial conditions. We explain these in turn.

The uniqueness of individual ecosystems and of the planetary environment in its entirety renders it impossible to achieve the sina qua non of the classical scientific experimental approach--replication of the system under investigation. Without the benefit of replicable systems, a statistically meaningful analysis of the effect of a toxin on an ecosystem is unattainable. The reason is that in any dose-response study, be it at the level of an individual organism or at the ecosystem level, one's interest is always in the

24

difference between a treatment and a control system. Inherently, this requires at least two initially identical systems. If replication of the treatment and control systems is also desired so that a measure of the statistical significance of the dose-response relarion can be derived, then even more identical systems are required. Ecosystems, unfortunately, are not so obliging. Two nearby lakes, two forests in the same region, and even two patches of meadow close by one another differ in myraid ways; ecologists will never be aware of all of them, let alone be able to quantify them.

To attempt a resolution of this dilemma, interest in ecological microcosms has recently accelerated. Microcosms are segments of natural ecosystems of a size convenient for laboratory replication and analysis. Lake microcosms, for example, consist of containers filled with lake water and possibly lake sediments taken from a real lake. If appropriate precautions are taken in the design, initiation, and operation of these systems, they can be replicated adequately for periods of up to several months and used for toxicological testing. Because they can be put together in such a way that a large fraction of the natural ecological diversity in the parent system is present in the microcosms, they offer a partial solution to the problem of uniqueness. Valuable as the microcosm approach is for ecotoxicological testing, problems of size or scale inherently limit its usefulness. Most importantly, it is not feasible to place large plants an animals in them; to do so would result in wildly unrealistic behavior, both with respect to chemical concentrations and population densities in the microcosms. Therefore, the very types of organisms of greatest interest to the public can not be studied in such systems. In addition, long-term microcosm investigations (usually of more than a few months duration) are not possible without jeopardizing the ecological realism (that is, the degree of similarity between the control microcosms and the parent ecosystem from which the

25

microcosms were derived) of the microcosms.

Which brings us to the second refractory source of uncertainty--
sensitivity to initial conditions. Ecosystems, like the global climate
system, are complex at many spatial and temporal dimensions. That is, within
such systems microscopic behavior and macroscopic behavior are present and are
strongly coupled. For example, the population dynamics of microbes can
affect the health of fish in a lake, and at a molecular level, the diffusion
of nutrients and the turbulence of the water can affect the microbe
populations. In the global climate system, atmospheric turbulence influences
climate on a macroscopic scale. In systems where such different dimensions
are coupled and chaotic or turbulent behavior is important, the ability to
predict the future consequences of the system is severly limited. In a
profound analysis of the effect of turbulence on climate prediction, Lorenz
(1969) showed that microscopic turbulence introduces an intrinsic source of
error in the prediction process. In particular, it renders the future
behavior of the climate incredibly sensitive to initial conditions. The
amount of detailed initial conditions one needs to measure in order to predict
future climate with any specified degree of accuracy increases faster than
exponentially with the period of time into the future one wants to predict the
climate. Long term prediction with the same detail and accuracy as we now
can achieve for one or two day predictions thus becomes intrinsically
impossible for a practical reason: we can not gather sufficiently detailed
measurements on today's climate.

The deep reason for this phenomenon is the extreme sensitivity of complex
systems possessing many scales of motion, such as systems with turbulence, to
small changes in initial conditions. Platt et al. (1977) investigated marine
ecosystems and found a similar sensitivity to initial conditions. It is
likely, in fact, that ecosystems, generally, are characterized by such a

26

sensitivity, although this has not been investigated yet.

Conclusion

The major advances in environmentally relevant ecological research in the past decade have not been in the direction of developing models that can predict with greater accuracy the future state of a disturbed ecosystem or the distribution of values of some uncertain parameter within its range of uncertainty. Rather the direction of progress has been in characterizing the features of ecosystems that render them either vulnerable or susceptible to change when subjected to stress and in identifying the major sources of uncertainty. Rather than making substantial progress in the development of one "correct" mathematical model for predicting the future behavior of an ecosystem, the effort has been to search for relatively model-independent truths. Valuable as this information is, it does not necessarily provide the type of information economists need if they are to apply valuation procedures to realistic situations. Error propagation and the existence of refractory sources of uncertainty in ecology must be taken into account if realistic goals for benefit-cost analysis in environmental policy are to be set. Perhaps most importantly, uncertainty about uncertainty—that is, uncertainty about the probability distribution of ecological variables within their range of uncertainty—limits progress toward more rational decision making. Perhaps error distributions can be better characterized and refractory uncertainties can be reduced by more intensive analysis of ensembles of models in conjunction with properly designed laboratory and field studies. In any event, progress toward the goal of more rational decision making will require that economists and ecologists working at the interface of these two disciplines are aware of the internal constraints of each others' field, while at the same time they sharpen their tools within their own.

References

Cairns, J., and J. Harte, 1985. "The Myth of the Threshold in Ecotoxicology", in preparation.

Lorenz, E., 1969. "The Predictability of a Flow Which Possesses Many Scales of Motion". Tellus 21, (3), pp. 289-307.

Platt, T., K. Denham, and A. Jasby, 1977. "Modelling the Productivity of Phytoplankton", in The Seas: Ideas and Observations on Progress in the Study of the Seas, E. Goldberg, ed. Vol. 6. Wiley, N.Y., N.Y..

Reckhow, K., 1984. "Decision Theory Applied to Lake Management", Duke University School of Forestry and Environmental Studies, Durham, N.C..

Roth, P., C. Blanchard, J. Harte, H. Michaels, and M. El-Ashry, 1985. "The American West's Acid Rain", World Resources Institute, Research Report #1. Washington, D.C..

U. S. Executive Order 12291, 1981.

Chapter 5

Hysteresis, Uncertainty, and Economic Valuation

I. INTRODUCTION

The purpose of this chapter is to investigate some issues that arise when one attempts to conduct a benefit evaluation for the control of pollution in an aquatic ecosystem. Obviously, the extent of the benefits depends on the nature of the ecosystem's response to control. We are concerned with two aspects of ecosystem behavior in particular. The first is the phenomenon known as "hysteresis", as discussed in chapter 3. Recall that this is the notion that a damaged ecosystem may not respond immediately to a cessation in pollution discharges and, when it does respond, may not exactly retrace the trajectory of its decline. Indeed, because of some irrecoverable losses from the system, it may _never_ return to its original state. The second aspect of ecosystem behavior we focus on is the stochasticity of natural phenomena which, as emphasized in chapter 4, implies that the ecosystem response is inherently uncertain.

Both the uncertainty and the dynamic constraints on ecosystem behavior need to be taken into account in evaluating the benefits of control and in the related decision on whether, or when, to control. Recovery dynamics, for example, may favor doing nothing, as in the case where the system is so far gone that recovery is impossible, or they may favor early action precisely to forestall more damaging, long-lasting consequences.

When uncertainty is factored into the analysis, an additional consideration arises which is sometimes overlooked. The temporal resolution of uncertainty--

the possibility of acquiring better information about the future consequences of controlling or continuing pollution--adds an extra element to the decision calculus. Regardless of whether the decisionmaker exhibits risk aversion or risk neutrality, if further information is forthcoming, there is a premium on those initial actions which preserve future flexibility and a discount on those which reduce flexibility and preclude the exploitation of the additional information at a later date. In the present context, this could be information about either the dynamics of ecosystem behavior or the social valuation of ecosystem products. If we control pollution now and, subsequently, learn that the ecosystem was not at a threshold of irreversible damage, we can always resume pollution later; but if we do not control now and then observe irreversible changes in the ecosystem, we cannot undo them by controlling later. Similarly, if we control now and then learn that future generations place a low value on ecosystem services, we can resume pollution; but if we do not control now and the ecosystem is irreversibly damaged, it is too late to act if we subsequently discover that future generations place a high value on the ecosystem. In each case there is an asymmetry in our ability to exploit future information and a premium associated with the action that preserves flexibility.

This flexibility premium has been recognized in the environmental valuation literature under the name of "quasi option value" (Arrow and Fisher [1974]) or "option value" (Henry [1974]).[1] Within the context of an irreversible land development decision where the future benefits of preservation in an undeveloped state are uncertain, these authors show that, when a decisionmaker ignores the possibility of acquiring further information about the future value of undeveloped land, he inevitably understates the net benefit of preservation over development and prejudices the decision somewhat in favor of immediate development.

The present work extends these results in several ways. First, we consider a decision framework where the irreversibility is associated with not taking action now (i.e., not controlling): In effect, we are dealing with the sin of omission rather than commission. More importantly, we consider a multiperiod decision problem, rather than the two-period problem of previous work. This change is important not merely because it is a step in the direction of greater realism--most practical policy issues involve a sequence of decision points--but also because it enables us to investigate some questions that are obscured within a two-period framework.

Suppose continued stress on a system is certain to trigger irreversible changes, beyond some critical point or period, but we do not know the period. Is there an analog to the two-period option value? Or suppose the critical period is known, but the damaging consequences are delayed as with certain kinds of health impacts. How does this affect the control decision? Still another issue we can consider in a multiperiod setting is the distinction between ordinary lags and irreversibility. Irreversible environmental degradation may be regarded as an extreme form of a lagged recovery in which the lag period is infinite (or, at any rate, longer than the effective planning horizon). What about less extreme lags where, if pollution continues beyond a certain point, the ecosystem is disabled for a certain (finite) period of time but then recovers: Do the option value arguments still apply?

Uncertainty, or more precisely the nature of learning, is necessarily treated differently in a multiperiod setting. In the two-period models, uncertainty is assumed completely resolved by the start of the second period. By contrast, we assume that the decisionmaker acquires some, but not all of the information over the first period, more over the second, more still over

the third, and so on. Partial, not perfect, information at any time is accordingly part of the structure of our model.

The chapter is organized as follows: In the next section we develop a model to evaluate pollution control, taking account of both the relevant physical constraints and the uncertainties. The model is used in sections III and IV to study the implications of various interesting combinations of recovery dynamics and uncertainties, of the sort just noted. Conclusions are offered in section V.

II. A FRAMEWORK FOR BENEFIT EVALUATION AND DECISION

We model the decision on whether or not to control pollution from the point of view of an environmental authority concerned with the net present value (benefits minus costs) of control. The optimal control is defined as the choice that maximizes this value. The important contraints are those that emerge from the discussion of the preceding section: (1) Beyond some

point in time, failure to control is not readily reversible; and (2) the benefits of control are uncertain due to a lack of knowledge about the timing and nature of ecosystem recovery and the willingness of individuals to pay for the goods and services it can produce.

Though recovery is a continuous process, evaluation and control take place in a discrete setting. Thus, we assume that a decision to control pollution can be made in each period $t = 1, 2, 3, \ldots$. The outcome of the decision can be represented by a sequence X_1, X_2, X_3, $\ldots$, where $X_t = 1$ corresponds to building a treatment plant, say, and $X_t = 0$ corresponds to not building. Note that we are considering a binary choice, neglecting intermediate levels of control. The results we obtain can be extended to the case of continuous control, but this is somewhat beside the point and comes at a substantial cost in complexity.

Associated with the choice of X_t is a set of benefits and costs. The capital and operating costs of the control facility in period t are denoted by C_t, and the benefits are denoted B_t; the net benefits are $NB_t \equiv B_t - C_t$. In the most general model, the benefits and costs accruing during any time period depend not only on the current pollution control decision, X_t, but also on all previous decisions, X_1, $\ldots$, X_{t-1}.

An essential feature mentioned above is that the benefits and costs of ecosystem recovery are uncertain. Thus, we write the overall net benefit function as

$$NB(X_1, X_2, X_3, \ldots; \theta) = NB_1(X_1; \theta) + \beta \, NB_2(X_1, X_2; \theta) + \beta^2 \, NB_3(X_1, X_2, X_3; \theta) + \ldots$$

where

$$NB_t(X_1, \ldots, X_t; \theta) \equiv B_t(X_1, \ldots, X_t; \theta) - C_t(X_1, \ldots, X_t; \theta).$$

Here β is a one-period discount factor, and θ is a random variable (or vector of random variables) representing the present uncertainty concerning the future consequences of pollution control.

With regard to the cost functions, it seems reasonable to assume that, with probability 1,

$$C_t(0, \ldots, 0; \theta) = 0$$

and

$$C_t(X_1, \ldots, X_{t-1}, 1; \theta) \geq C_t(X_1, \ldots, X_{t-1}, 0; \theta).$$

That is to say, pollution control is costly. Finally, in order to keep the decision problem simple while still making it interesting, we focus on a three-period model. This is significantly more general than the two-period models which have been used in irreversibility literature so far (for example, Arrow and Fisher [1974], Henry [1974], Epstein [1980]). With minimal notational clutter, it permits us to consider scenarios involving a variety of types of irreversibility, which is our primary objective in this paper.

Given this structure, the social decision problem is to maximize the discounted present value of expected net benefits:

$$(1) \qquad \max_{X_1, X_2, X_3} E\{NB(X_1, X_2, X_3; \theta)\}.$$

Two aspects of this problem need to be addressed, both pertaining to the treatment of uncertainty. First, what about attitudes toward risk? Should one assume risk aversion on the part of the social decisionmaker and, therefore, include a risk-premium term when taking the expectation in (1), or should one assume risk neutrality following the arguments, for example, of Samuelson [1964] or Arrow and Lind [1970]? Although it clearly makes a difference in practice, the question of risk aversion is not fundamental to the results that we will obtain: They are qualitatively independent of any assumption about risk preferences. The second aspect of modeling uncertainty in a dynamic setting is its behavior over time. Uncertainty means a lack of information; yet, it is likely that this situation changes--that information is acquired over time. Our analysis is largely concerned with the consequences of a failure on the part of the decisionmaker to take this prospect into account. We will show how this affects the social decision and how conventional benefit-cost analysis must be adjusted to incorporate this consideration.

Suppose, first, that the decisionmaker does not have to commit himself in the first period to an entire intertemporal control strategy; he can postpone the choice of X_2 to t = 2 and the choice of X_3 to t = 3. Suppose, moreover, that in each time period (except t = 3), he recognizes that further information about the future consequences of control will become available which he can exploit in making these future decisions. Define

(2a) $$\hat{V}_3(X_3|X_1, X_2) \equiv E_3\{NB_3(X_1, X_2, X_3; \theta)\}$$

(2b) $$\hat{V}_2(X_2|X_1) \equiv E_2\{NB_2(X_1, X_2; \theta) + \max_{X_3} \beta \hat{V}_3(X_3|X_1, X_2)\}$$

(2c) $$\hat{V}_1(X_1) \equiv E_1\{NB_1(X_1; \theta) + \max_{X_2} \beta \hat{V}_2(X_2|X_1)\}.$$

where $E_t\{\cdot\}$ denotes an expectation with respect to the information set available at time t--i.e., E_1 is the expectation with respect to the decision-maker's prior distribution for θ, E_2 is the expectation with respect to his posterior distribution in t = 2 which is updated in a Bayesian manner on the basis of the information obtained by the beginning of the second period, etc. One point must be emphasized: We assume that the acquisition of information does not depend on the choice of X_t; it emerges either with the passage of time (e.g., as period 2 approaches, one can make a more accurate assessment about the social value of environmental quality in the second period) or as the result of a separate research program on ecosystem dynamics.[2]

Following the Backwards Induction Principle of dynamic programming, in the third period the decisionmaker selects

(3a) $$\hat{X}_3 \equiv \arg \max \hat{V}_3(X_3|X_1, X_2),$$

in the second he selects

(3b) $$\hat{X}_2 \equiv \arg \max \hat{V}_2(X_2|X_1),$$

and in the first he selects

(3c) $$\hat{X}_1 \equiv \arg \max \hat{V}_1(X_1).$$

In each case we are assuming that, however X_1, ..., X_{t-1} are chosen, X_t is chosen optimally in the light of these previous decisions. Where it is necessary to emphasize this dependence, we shall write $\hat{X}_t$ as an explicit function of

the previous choice variables--e.g., $\hat{X}_2 = \hat{X}_2(X_1)$. In the terminology of stochastic control theory, $(\hat{X}_1, \hat{X}_2, \hat{X}_3)$ represents a closed-loop policy: At each decision point, both current information and all future anticipated information are considered in choosing a control.

We wish to contrast this with a policy in which the prospect of future information is disregarded. There are two ways to model this. One is to assume that, although the decisionmaker is still free to postpone his choice of X_2 and X_3 until the second and third periods, respectively, in each period he ignores the possibility of future learning and deals with uncertainty about future consequences by replacing random variables with his current estimate of their mean. Define

(4a) $$V_3^*(X_3|X_1, X_2) \equiv E_3\{NB_3(X_1, X_2, X_3; \theta)\}$$

(4b) $$V_2^*(X_2|X_1) \equiv \max_{X_3} E_2\{NB_2(X_1, X_2; \theta) + \beta\, NB_3(X_1, X_2, X_3; \theta)\}$$

(4c) $$V_1^*(X_1) \equiv \max_{X_2, X_3} E_1\{NB_1(X_1; \theta) + \beta\, NB_2(X_1, X_2; \theta) + \beta^2\, NB_3(X_1, X_2, X_3; \theta)\}.$$

In the third period, the decisionmaker selects

(5a) $$X_3^* \equiv \arg\max V_3^*(X_3|X_1, X_2),$$

in the second he selects

(5b) $$X_2^* \equiv \arg\max V_2^*(X_2|X_1),$$

and in the first he selects

(5c) $$X_1^* \equiv \arg\max V_1^*(X_1).$$

In the terminology of stochastic control theory, this is an open-loop feedback policy: As new information becomes available, the decisionmaker incorporates it in his choice of a control; but he assumes that no further information will become available.

The other approach to modeling the disregard of future information is to assume that the decisionmaker does not wait (or cannot wait) until the second and third periods to choose X_2 and X_3 but, instead, chooses them in the first period along with X_1. This decision, denoted $(X_1^{**}, X_2^{**}, X_3^{**})$, is the solution to

$$(6) \quad \max_{X_1, X_2, X_3} E_1 \{NB_1(X_1; \theta) + \beta\, NB_2(X_1, X_2; \theta) + \beta^2\, NB_3(X_1, X_2, X_3; \theta)\}.$$

This is known as an open-loop control where all decisions are made simultaneously on the basis of the information available at the beginning of the initial period. Comparing (5) and (6), it is clear that $X_1^* = X_1^{**}$, but in general, $X_2^{**} \neq X_2^*$ and $X_3^{**} \neq X_3^*$--there is no difference between the open-loop and open-loop feedack controls in the first period but in subsequent periods they differ. Thus our discussion below of the relation between $\hat{X}_1$ and X_1^* also applies to X_1^{**}, but it does not apply to relations in $t = 2$ and $t = 3$. Since, in a three-period model, unlike a two-period model, the choice of X_2 is of substantive interest, the sharp distinction between open-loop and open-loop feedback policies is one of the benefits that we gain by switching to a multiperiod setting. It will become clear below that, for our purposes, useful results can be obtained by comparing the closed-loop policy with the open-loop feedback policy.

We can pursue this comparison in two ways. We can ask a policy question: How do $\hat{X}_t$ and X_t^* differ? In particular, under what circumstances is it true that $\hat{X}_t \geq X_t^*$ (i.e., the case for intervening to control pollution is

strengthened when the prospect of further information is recognized)? Or we can ask a __benefit evaluation__ question: How do $\hat{V}_t(\cdot)$ and $V_t^*(\cdot)$ differ? What correction is required when expected benefits are estimated by replacing uncertain future quantities with a current estimate of their expected value?

Given the constraint that $X_t = 0$ or 1, these questions can be answered by observing that, from (2)-(4),

$$(7a) \qquad \hat{X}_1 \geq (\leq) X_1^* \qquad \text{as} \qquad OV_1 \geq (\leq) 0$$

and, for any given X_1,

$$(7b) \qquad \hat{X}_2(X_1) \geq (\leq) X_2^*(X_1) \qquad \text{as} \qquad OV_2(X_1) \geq (\leq) 0$$

where

$$(8a) \qquad OV_1 \equiv [\hat{V}_1(1) - \hat{V}_1(0)] - [V_1^*(1) - V_1^*(0)]$$

$$(8b) \qquad = [\hat{V}_1(1) - V_1^*(1)] - [\hat{V}_1(0) - V_1^*(0)];$$

and, given X_1,

$$(9a) \qquad OV_2(X_1) \equiv [\hat{V}_2(1|X_1) - \hat{V}_2(0|X_1)] - [V_2^*(1|X_1) - V_2^*(0|X_1)]$$

$$(9b) \qquad = [\hat{V}_2(1|X_1) - V_2^*(1|X_1)] - [\hat{V}_2(0|X_1) - V_2^*(0|X_1)].$$

The quantities OV_1 and $OV_2(X_1)$ are the correction factors required when the prospect of future information is disregarded and benefits are measured in terms of $V_t^*(\cdot)$ instead of $\hat{V}_t(\cdot)$; they are multiperiod generalizations of the Arrow-Fisher-Henry concept of option value.

To interpret them, consider (8b) and (9b) and observe that the term $[\hat{V}_t(X_t) - V_t^*(X_t)]$ can be cast in the form of

$$(10) \qquad \hat{V}_t(\cdot) - V_t^*(\cdot) = E_t\{\max_{X_{t+1}\cdots} F_t(\cdot; \theta)\} - \max_{X_{t+1}\cdots} E_t\{F_t(\cdot; \theta)\}.$$

This is a measure of the value of information acquired after the beginning of period t that can be exploited in the subsequent choice of X_{t+1}, X_{t+2}, ..., conditional on the choice of X_t in period t. Thus, in (8b), $[\hat{V}_1(1) - V_1^*(1)]$ is the expected value of the information that might be acquired in time to influence the second- and third-period choices conditional on controlling pollution in the first period, while $[\hat{V}_1(0) - V_1^*(0)]$ is the expected value of subsequent information conditional on not controlling pollution in the first period. The correction factor OV_1 is simply the difference between these two conditional values of information; similarly, for OV_2. Thus, if $OV_t \geq 0$, the value of information associated with setting $X_t = 1$ exceeds that associated with a decision to set $X_t = 0$ and the case for controlling pollution in period t is strengthened when the prospect of future information is considered. Conversely, if $OV_t \leq 0$, the case for pollution control is weakened.

However, without placing further structure on the model, it is impossible to determine which outcome is the more likely. From the convexity of the maximum operator and Jensen's Inequality applied to (10), it follows that $\hat{V}_t(\cdot) - V_t^*(\cdot) \geq 0$. Thus, each component of OV_t is nonnegative; but this tells us nothing about the sign of their difference. In the following sections we consider some alternative model structures embodying features of ecosystem dynamics discussed in section II and explore their effect on OV_t and their implications for pollution control policy.

III. CRITICAL PERIOD IRREVERSIBILITY

Suppose that, at some point in the evolution of the ecosystem, if the policymaker does not intervene and control pollution at that time, it could never be optimal for him to control pollution subsequently. We shall call a time period with this property a "critical" period. Whether such a phenomenon exists and what factors bring it about depends on the specifics of the ecosystem structure. In the context of the three-period model, suppose that, while it might pay to introduce controls after pollution has continued unchecked for one more period, it could never pay to introduce controls after pollution has continued unchecked for two more periods in a row. More formally, we assume that, with probability 1,

$$(11) \qquad E_t\{NB_3(0, 0, 1; \theta)\} \leq E_t\{NB_3(0, 0, 0; \theta)\} \qquad\qquad t = 2, 3.$$

Thus, if pollution is not controlled in the first period ($X_1 = 0$), the second period becomes critical.

From (2a,b) and (4b), when $X_1 = 0$, we have

$$(12a) \quad \hat{V}_2(0|0) = E_2\{NB_2(0, 0; \theta) + \beta \max [E_3 NB_3(0, 0, 1; \theta), E_3 NB_3(0, 0, 0; \theta)]\}.$$

$$(12b) \quad \overset{*}{V}_2(0|0) = E_2 NB_2(0, 0; \theta) + \beta \max [E_2 NB_3(0, 0, 1; \theta), E_2 NB_3(0, 0, 0; \theta)].$$

Applying (11) yields

$$(13a) \qquad \hat{V}_2(0|0) = E_2 NB_2(0, 0; \theta) + \beta E_2\{E_3 NB_3(0, 0, 0; \theta)\}.$$

$$(13b) \qquad \overset{*}{V}_2(0|0) = E_2 NB_2(0, 0; \theta) + \beta E_2 NB_3(0, 0, 0; \theta).$$

However, by the Total Probability Theorem, $E_t\{h(\theta)\} = E_t\{E_{t+1} h(\theta)\}$ for any function of a random variable, $h(\theta)$. Therefore, we obtain the key result that

$$(14) \qquad \hat{V}_2(0|0) - V_2^*(0|0) = 0.$$

Because the second period is critical when $X_1 = 0$, it follows that, if the decisionmaker does not control pollution in that period, he anticipates that he will never choose to control it subsequently. Since the anticipated future decisions are exactly the same under both the closed-loop and open-loop feedback policies, the expected future benefits are identical under both policies. In effect, any subsequent information is expected to have no economic value because it is not anticipated to have any effect on future decisions; hence, (14). Substituting this into (9) yields

$$(15) \qquad OV_2(0) = \hat{V}_2(1|0) - V_2^*(1|0) \geq 0.$$

From (7b), this implies that $\hat{X}_2(0) \geq X_2^*(0)$. That is, if pollution is not controlled in the first period, we have a situation where, once the potential for the acquisition of future information is recognized, the case for controlling pollution in the second period is strengthened, and there is a positive flexibility premium associated with setting $X_2 = 1$.

The key to this analysis is equation (11) which embodies our particular assumption that the second period is critical when $X_1 = 0$. Without imposing any additional restrictions, it is impossible to determine the signs of OV_1 or $OV_2(1)$. For example, from (11), one cannot infer that $\hat{V}_2(0|1) = V_2^*(0|1)$. Therefore, the indeterminacy concerning the relation between $\hat{X}_1$ and X_1^*, or $\hat{X}_2(1)$ and $X_2^*(1)$, remains.

Generalizing from this particular example, a period is critical whenever an equation analogous to (11) holds, i.e., whenever the situation is such that, if the decisionmaker does not control in that period, with probability 1 he anticipates that it would never pay to control in future periods regardless of the information subsequently acquired. By construction, when a period t is critical, we have $\hat{V}_t(0|\cdot) = V_t^*(0|\cdot)$ which implies that $OV_t(\cdot) \geq 0$ and $\hat{X}_t(\cdot) \geq X_t^*(\cdot)$.

It may be useful to compare our notion of a critical period with the concept of irreversibility employed by Arrow and Fisher [1974] and by Henry [1974] which, in the present context, would be represented by a constraint of the form

$$(16) \qquad\qquad X_1 = 0 \;\rightarrow\; X_2 \geq X_3.$$

Our assumption (11) implies (16) but is somewhat broader and illuminates the two crucial ingredients required to extend their results to more general settings. First, what is irreversible is the policy, not the fate of any particular biotic components. The ecosystem dynamics may be such that, if $X_2 = 0$, the lake trout become extinct without this necessarily implying (11) as long as the trout are sufficiently unimportant relative to the decision-maker's other objectives. The truth or falsity of (11) depends on values as well as biology. Second, what is at issue is economic rather than technical irreversibility. The technology may be such that the decision on X_2 is physically reversible in later periods (e.g., setting $X_2 = 0$ corresponds to permitting the construction of a steel mill on the edge of a lake which could subsequently be converted to a nonpolluting bowling alley); the question is whether it could ever pay to reverse the current decision. Moreover, what matters is the present anticipation of whether it could ever pay to reverse

that decision. Our assumption (11) does not preclude the possibility that, ex post, at the end of period 3, it might actually turn out that it would have been optimal to choose $X_3 = 1$ even with $X_2 = 0$. What is required is that, ex ante, this choice is always deemed implausible. Thus, we can admit the possibility that

$$NB_3(0, 0, 1; \theta) > NB_3(0, 0, 0; \theta)$$

for some realizations of θ as long as the prior density on θ and the subsequent updated posterior densities are sufficiently bounded to ensure that the expected benefits satisfy the inequality in (11).

IV. DELAYED AND TEMPORARY IRREVERSIBILITY

In this section we consider two forms of irreversibility which are weaker than the critical-period concept introduced above and yield somewhat different results. First, we consider what might be called "delayed" irreversibility: If pollution is not controlled, the consequences are (economically) irreversible, but the irreversibility sets in only after a lag. Thus, if pollution is permitted to continue now, there is an intermediate period during which it may or may not be optimal to impose controls; but, after this intermediate period, it can never pay to control. Within the framework of our three-period model, we identify "now" with period 1, the intermediate period during which it may or may not be optimal to control with period 2, and the subsequent future with period 3. The assumption of delayed irreversibility is captured by combining (11) together with the assumption that

(17) $\qquad E_t\{NB_3(0, 1, 1; \theta)\} \leq E_t\{NB_3(0, 1, 0; \theta)\} \qquad t = 2, 3$

with probability 1. The question to be addressed is how this type of irreversibility affects the pollution-control decision in period 1.

Substituting (11) and (17) into (2c) and (4c) yields the following expressions for $\hat{V}_1(0)$ and $V_1^*(0)$:

$$\hat{V}_1(0) = E_1 \ NB_1(0; \ \theta) + \beta \ E_1 \ \{\max \ [E_2 \ NB_2(0, \ 0; \ \theta) + \beta \ E_2 \ NB_3(0, \ 0, \ 0; \ \theta),$$
(18a)
$$E_2 \ NB_2(0, \ 1; \ \theta) + \beta \ E_2 \ NB_3(0, \ 1, \ 0; \ \theta)]\}$$

$$V_1^*(0) = E_1 \ NB_1(0; \ \theta) + \beta \ \max \ [E_1 \ NB_2(0, \ 0; \ \theta) + \beta \ E_1 \ NB_3(0, \ 0, \ 0; \ \theta),$$
(18b)
$$E_1 \ NB_2(0, \ 1; \ \theta) + \beta \ E_1 \ NB_3(0, \ 1, \ 0; \ \theta)].$$

By inspection, it can be seen that, while $\hat{V}_1(0) - V_1^*(0) \geq 0$, it is <u>not</u> true in general that $\hat{V}_1(0) = V_1^*(0)$. Since it can also be shown that $\hat{V}_1(1) - V_1^*(1) \geq 0$, from (8a,b), this is a situation where the sign of OV_1 and the relation between $\hat{X}_1$ and X_1^* are indeterminate.

Observe that the formula for $\hat{V}_1(0)$ in (18a) involves information acquired between the first and second periods but not that acquired between the second and third periods--the expectation $E_3\{\cdot\}$ does not appear. The latter information has no economic value when $X_1 = 0$ because the irreversibility has set in by then, but the former does have some value because it can be exploited during the intermediate period ($t = 2$) where there is still some flexibility. Of course, if $X_1 = 1$, there is sufficient flexibility to exploit <u>both</u> sets of information. But this fact, by itself, does not guarantee that the overall value of information associated with setting $X_1 = 1$ necessarily exceeds that associated with setting $X_1 = 0$. The point is that, with delayed irreversibility, the first

period is not critical because, if one does not control, then it is <u>not</u> true that it can never be optimal to control subsequently; it may still be optimal to control during the intervening period before the irreversibility sets in. Thus, with delayed irreversibility, the introduction of future learning into the decision calculus need not tilt the balance in favor of immediate control.

We now examine what might be called "temporary" irreversibility as opposed to the "permanent" irreversibility considered so far. We consider two scenarios. In the first we suppose that, if pollution is not controlled in any period, the consequences are temporarily irreversible and are felt in the following period but not necessarily thereafter. In effect, the system has a one-period memory with

$$(19) \qquad E_t\{NB_2(0,\ 1;\ \theta)\} \leq E_t\{NB_2(0,\ 0;\ \theta)\}$$

$$(20a) \qquad NB_3(X_2,\ X_3;\ \theta) \equiv NB_3(0,\ X_2,\ X_3;\ \theta) = NB_3(1,\ X_2,\ X_3;\ \theta)$$

$$(20b) \qquad E_t\{NB_3(0,\ 1;\ \theta)\} \leq E_t\ \{NB_3(0,\ 0;\ \theta)\}.$$

In this case $\hat{V}_1(0)$ and $V_1^*(0)$ are given by

$$\hat{V}_1(0) = E_1\ NB_1(0;\ \theta) + \beta\ E_1\left\{\max\left(E_2\ NB_2(0,\ 1;\ \theta) + \beta\ E_2\ \max\ [E_3\ NB_3(1,\ 0;\ \theta),\right.\right.$$
(21a)
$$\left.\left. E_3\ NB_3(1,\ 1;\ \theta)],\ E_2\ NB_2(0,\ 0;\ \theta) + \beta\ E_2\ NB_3(0,\ 0;\ \theta)\right)\right\},$$

$$V_1^*(0) = E_1\ NB_1(0;\ \theta) + \beta\ \max\ \{E_1\ NB_2(0,\ 1;\ \theta) + \beta\ \max\ [E_1\ NB_3(1,\ 0;\ \theta),$$
(21b)
$$E_1\ NB_3(1,\ 1;\ \theta)],\ E_1\ NB_2(0,\ 0;\ \theta) + \beta\ E_1\ NB_3(0,\ 0;\ \theta)]\}.$$

It follows that, while $\hat{V}_1(0) - V_1^*(0) \geq 0$, it is not true in general that $\hat{V}_1(0) = V_1^*(0)$. Thus, with this type of temporary irreversibility, the sign of OV_1 and the relation between $\hat{X}_1$ and X_1^* are indeterminate.

We now change the scenario by assuming that, if pollution is not controlled in the first period, the consequences are temporarily irreversible in the second period but the third period is entirely independent of what has happened previously, i.e., the system makes a fresh start and has no memory in the third period. Thus, we retain (19) while assuming that the third-period benefit functions satisfy the restrictions

$$NB_3(X_3; \theta) \equiv NB_3(1, 1, X_3; \theta) = NB_3(1, 0, X_3; \theta)$$

(22)

$$= NB_3(0, 1, X_3; \theta) = NB_3(0, 0, X_3; \theta).$$

The new formulas for $\hat{V}_1(0)$ and $V_1^*(0)$ are

$$\hat{V}_1(0) = E_1 \, NB_1(0; \theta) + \beta \, E_1 \, NB_2(0, 0; \theta)$$

(23a)

$$+ \beta^2 \, E_1 \, \{\max \, [E_3 \, NB_3(0; \theta), \, E_3 \, NB_3(1; \theta)]\}$$

$$V_1^*(0) = E_1 \, NB_1(0; \theta) + \beta \, E_1 \, NB_2(0, 0; \theta)$$

(23b)

$$+ \beta^2 \max \, [E_1 \, NB_3(0; \theta), \, E_1 \, NB_3(1; \theta)].$$

Similarly, substitution of (19) and (22) into (2c) and (4c) yields the following formulas for $\hat{V}_1(1)$ and $V_1^*(1)$:

$$\hat{V}_1(1) = E_1 \, NB_1(1; \, \theta) + \beta \, E_1 \, \{max \, [E_2 \, NB_2(1, \, 0; \, \theta), \, E_2 \, NB_2(1, \, 1; \, \theta)]\}$$

(24a)

$$+ \, \beta^2 \, E_1 \, \{max \, [E_3 \, NB_3(0; \, \theta), \, E_3 \, NB_3(1; \, \theta)]\}$$

$$V_1^*(1) = E_1 \, NB_1(1; \, \theta) + \beta \, max \, [E_1 \, NB_2(1, \, 0; \, \theta), \, E_1 \, NB_2(1, \, 1; \, \theta)]$$

(24b)

$$+ \, \beta^2 \, max \, [E_1 \, NB_3(0; \, \theta), \, E_1 \, NB_3(1; \, \theta)].$$

In this case, although it is still true that $[\hat{V}_1(1) - V_1^*(1)] \geq 0$ and $[\hat{V}_1(0) - V_1^*(0)] \geq 0$, we <u>can</u> determine the sign of OV_1 since application of (8) yields

$$OV_1 = \beta \, E_1 \, \{max \, [E_2 \, NB_2(1, \, 0; \, \theta), \, E_2 \, NB_2(1, \, 1; \, \theta]\}$$

(25)

$$- \, \beta \, max \, [E_1 \, NB_2(1, \, 0; \, \theta), \, E_1 \, NB_2(1, \, 1; \, \theta)] \geq 0.$$

It follows, therefore, that $\hat{X}_1 \geq X_1^*$.

In the first scenario, based on (19) and (20a,b), if one fails to control in the first period, it may nevertheless be optimal to control in the second, despite the irreversibility embodied in (19), because second-period decisions influence third-period outcomes. Thus, when $X_1 = 0$, information acquired between the first and second periods still has some economic value because it may shed light on third-period outcomes and can, therefore, affect the second-period decision. When $X_1 = 1$, information acquired between the first and second periods also has an economic value. Consequently, the net effect of incorporating future learning into benefit estimation is ambiguous: it may strengthen or weaken the case for initial control.

By contrast, in the second scenario, based on (19) and (22), the second-period decision cannot affect third-period outcomes at all because of the total lack of memory between these two periods. Therefore, the temporary irreversibility in (19) ensures that it is never optimal to control in the second period when one has not also controlled in the first. As a result, the information acquired between the first and second periods has some value when $X_1 = 1$ but none when $X_1 = 0$. Moreover, because the system makes a fresh start in the third period, the information acquired between the second and third periods is equally valuable regardless of whether $X_t = 0$ or 1, $t = 1, 2$. Hence, the case for initial control is unambiguously strengthened when one recognizes the possibility of future learning.

While it is clear that the first scenario of temporary irreversibility is incompatible with the concept of a critical period, the second scenario can still be related to that concept, albeit in a somewhat unusual manner. Under the second scenario, if the decisionmaker decides not to control in the first period, he anticipates that it could never be optimal for him to reverse this decision during the subsequent interval lasting until the system's memory is "reset." Once that has occurred, all future decisions are entirely independent of prior events. Thus, there is a sense in which the first period is "locally" critical.

V. CONCLUSIONS

It has long been recognized that the selection of an optimal pollution control or other environmental policy is highly dependent on the treatment of time and uncertainty in the benefit cost calculus. A delay in ecosystem recovery, for example, may reduce the present value of the benefits from

control; but if the recovery lags caused by continuing pollution are growing faster than the discount rate, this would tilt the balance in favor of early control, as shown in a somewhat different context by Krutilla and Fisher (1975). Similarly, depending upon one's view of the degree of risk aversion appropriate for public policy decisions, the presence of uncertainty may require an adjustment to the expected monetary benefits and costs of control. Since there may be uncertainty about the consequences of both control and no control, this could cut either way.

While not denying the importance of these issues for empirical policy analysis, in this chapter we have focused on a different aspect of benefit evaluation involving flexibility, the temporal resolution of uncertainty, and the value of information. In a dynamic system, information about the consequences of previous actions may arrive over time, and this prospect must be taken into consideration when one makes policy decisions. Future observations have no economic value, however, if (1) they are entirely uninformative in the sense that the prior and posterior distributions coincide or (2) they are informative but they cannot affect subsequent decisions because the policymaker lacks freedom of action. Thus, flexibility is a necessary ingredient for information to have economic value. This must be borne in mind when one contemplates an action with irreversible consequences, because the resulting lack of flexibility nullifies the value of any subsequent information.

In many pollution control issues this may be a relevant consideration because the ecological consequences of a failure to control may be irreversible. Actually, we have shown that what is crucial is economic irreversibility. That is to say, if in some time period the decisionmaker anticipates that, unless he controls then, it would never pay to control in the future,

regardless of the subsequent information, a decision not to control then would effectively eliminate future flexibility. In that case, there is a positive flexibility premium associated with a decision to control: When future learning is taken into account, the balance is tilted in favor of control. We have termed this a critical-period irreversibility. In other cases, however, the issue is less clear cut. For example, it may happen that the irreversible consequences are delayed in their onset or are only temporary in their effects. In such cases, we show that the conditional value of future information when one fails to control now is not necessarily zero; conceivably it may exceed the value of information associated with a decision to control. The prospect of future learning then has an ambiguous effect--it may strengthen or weaken the case for control. Our intuition is that the value of information conditional on control will ordinarily exceed the value of information conditional on no control but this is an empirical issue to be resolved through specific case studies. Such an application is the focus of our current research and will be reported separately.

[1]The term "option value" has also been used in connection with a different concept related to risk version in an atemporal setting. Major references include Schmalensee [1972], Bohm [1975], Graham [1981], Bishop [1982], Smith [1983], and Freeman [1984].

[2]Obviously, if the control decision itself generates information, this may alter the balance of the argument. If, by not controlling now, one generates potentially useful information which can be exploited in future decisions (for example, because the major uncertainty concerns the consequences of not controlling), this would weaken the case for control. If, on the other hand, one generates useful information by controlling now (because the major uncertainty concerns the consequences of control), this would strengthen the case for control. In the absence of a specific case study, it is difficult to say a priori whether or not there is dependent learning and, if there is, which form it takes. For this reason we have focused on the case of independent learning. For a further discussion of this issue see Fisher and Hanemann [1985].

REFERENCES

Kenneth J. Arrow and Anthony C. Fisher, "Environmental preservation, uncertainty, and irreversibility," Quart. J. Econ. 88, 312-319 (May 1974).

Kenneth J. Arrow and R. C. Lind, "Uncertainty and the evaluation of public investment decisions," Amer. Econ. Rev. 60, 364-378 (1970).

R. C. Bishop, "Option value: An exposition and extension," Land Econ. 58, 1-15 (February 1982).

Peter Bohm, "Option demand and consumer's surplus: Comment," Amer. Econ. Rev. 65, 733-736 (September 1975).

Larry G. Epstein, "Decision making and the temporal resolution of uncertainty," Inter. Econ. Rev. 21, 269-283 (1980).

Anthony C. Fisher and W. Michael Hanemann, "Quasi-option value: Some misconceptions dispelled," J. Environ. Econ. Manag. (forthcoming).

A. Myrick Freeman, "The size and sign of option value," Land Econ. 60, 1-13 (February 1984).

D. A. Graham, "Cost-benefit analysis under uncertainty," Amer. Econ. Rev. 71, 715-725 (September 1981).

Claude Henry, "Investment decisions under uncertainty: The irreversibility effect," Amer. Econ. Rev. 64, 1006-1012 (December 1974).

A. J. Horne, "A Suite of Indicator Variables (SIV) Index for an Aquatic Ecosystem," Energy and Resources Group, University of California, Berkeley (August 1985).

A. J. Horne, J. Harte, and D. F. Von Hippel, "Predicting the Recovery of Damaged Aquatic Ecosystems: A Hysteresis Trophic Link Model (HTLM)," Energy and Resources Group, University of California, Berkeley (August 1985).

J. V. Krutilla and A. C. Fisher, The Economics of National Environments: Studies in the Valuation of Commodity and Amenity Resources, Johns Hopkins Press, Baltimore (1975).

P. A. Samuelson, "Discussion," Amer. Econ. Rev. 54, 93-96 (1964).

Richard Schmalensee, "Option demand and consumer's surplus: Valuing price changes under uncertainty," Amer. Econ. Rev. 62, 813-824 (December 1972).

V. Kerry Smith, "Option value: A conceptual overview," Southern Econ. J. 654-668 (January 1983).

CHAPTER 6   THE ECONOMIC CONCEPT OF BENEFIT

INTRODUCTION

Consider the question: "What is the value of the wetlands of San Francisco Bay?". Why might such a question be addressed to an economist rather than a philosopher or a poet? To explain this it is vital to distinguish between two different meanings that might be attached to the original question: (i) How much value do people place on the wetlands (assuming an adequate base of information)? (ii) How much value ought they to place on them? The latter question is certainly the province of the philosopher or the poet; the economist too may have some thoughts about the question, but these arise from his private sentiments, not from his professional discipline. The former question – the positive question – is the one that the discipline of economics addresses. When we talk of benefits and benefit measurement in this report, we have this interpretation in mind – the values that people actually place on ecosystems.

This itself raises a host of questions: Which people? In what units should values be measured? Why do people have these values? Just how do we ascertain them? We will comment briefly on each of the first three questions. The answers to the fourth question will take up the remainder of this chapter, as well as Chapter 7. <u>Which people?</u> This is specified, in principle, by the agency commissioning the benefit assessment. A related, and more complex question, is: How do we add up different people's values? Again, this is specified, in principle, by the agency commissioning the study; however, here there is a body of economic theory which can guide the answer – see, for example, Sen (1973), Blackorby and Donaldson (1978), and Boadway and Bruce (1984, Chapter 9). To save space, we will duck this issue here. <u>What units?</u> Values can be measured in monetary units or in units of any commodity that people happen to value. For example, we could measure the value to an

individual of aquatic ecosystems in units of chocolate truffles - Lake Tahoe is
worth 100 truffles, say, while Mono Lake is worth only 82 truffles. Different
systems of units will generate the same ordinal ranking of ecosystems, but not
necessarily the same cardinal index of value. We choose to adopt money -
purchasing power - as our unit of measurement because this is the predominant
convention. It is possible to develop an analogous theory of benefit
measurement based on chocolate truffle units, but we shall not explore this
here (aggregation across individuals would presumably be more difficult).

How do we ascertain values? In principle there are two ways to proceed:
(i) Ask people directly, and (ii) Rely on revealed preference - observe their
behavior when they make choices on which the aquatic ecosystem somehow impinges,
and infer their values from this behavior. In this chapter we focus on the
latter approach exclusively. An immediate implication is an answer to the
question: Why do people have these values? The answer is that it doesn't
matter. We rely on preferences as revealed by actual behavior, without needing
to know how these prefences might be decomposed into alternative motives. Or
rather, there are two circumstances in which motives might matter. The first
is when a knowledge of motives gives us reason to believe that preferences (and
behavior) will be different in the future. Stability of preferences is
essential to extrapolation from observed behavior. If preferences are not
stable, this poses both a philosophical and a practical problem. The
philosophical problem is: Which set of preferences do we rely on? The practical
problem is: How can we predict what the new preferences will be if it is
decided to rely on them? The other circumstance in which we might care about
motives has to do with aggregation across individuals: specifically, a knowledge
of motives may help us to identify groups of individuals who have different
preferences. For empirical purposes, it might be more appropriate to analyze
the behavior of each group separately, rather than to aggregate them into a

single group.

Given the focus on revealed preference, why is the presence of markets required for the success of our endeavor? One answer common among non-economists, but erroneous, is that values are embodied in market prices and expenditures. Markets are needed because market prices establish values: if a commodity sells for $10, that is the value of the commodity. However, this is not exactly true. If I buy the commodity at a price of $10, then it certainly must be worth $10 to me - but it may be worth even more; i.e., the price is a lower bound on value. If I do not buy the commodity at this price, it is not worth $10 to me; i.e., the price is an upper bound. Let us switch from prices to expenditures and focus on the first case. Suppose I buy 5 units of the commodity at the going price of $10, so that my total expenditure is $50. This expenditure is clearly a lower bound on the value of the commodity to me. The problem, however, is that this lower bound may be inadequate for our purposes. Ultimately we are interested in net benefits - i.e. benefits minus costs. If the cost of supplying the commodity is also $10 a unit, the cost amounts to $50 and the difference between that and our lower bound estimate of benefits is zero - because we underestimate benefits when we use expenditures, we underestimate net benefits, possibly to the point of absurdity. Moreover, consider some change in the supply of the commodity (for example an improvement in its quality) which leads me to spend $70 on it. For the same reason as before, this $70 is a lower bound on the value of the improved commodity to me. But the change in expenditure conveys absolutely no information about the change in value: the difference between two lower bounds is not necessarily a lower bound on the difference in the quantities being bounded.

In short, we do not care about markets because market expenditures directly indicate values. At best they provide bounds on values, but these bounds are frequently so imprecise as to be useless, and the changes in market

expenditures provide no information about changes in values. Instead, we care about markets because they provide a forum for choice behavior - performing tradeoffs between goods and money - from which we can indirectly infer preferences. That is the essence of the revealed preference approach. Moreover, as will be shown in the next section, these market transactions, or tradeoffs, can convey information about preferences for other items of value which are not themselves traded in a market, as long as the preferences for the latter items interact (in a sense to be made specific below) with preferences for the traded items. We turn, now, to an elaboration of this argument.

THE BASIC FRAMEWORK

The revealed preference approach to benefit assessment can be explained in terms of two basic consumer choice models. Both models pertain to an individual consumer - we want to avoid the complications associated with estimation and interpretation of aggregate demand functions. In the first model, the individual has preferences for various marketed commodities, whose consumption is denoted by the vector x, and for various environmental resources, which are denoted by q: this could be a vector but, for simplicity of notation, we treat it as a scalar. These preferences are represented by a utility function $u(x,q)$ which is continuous and non-decreasing in all arguments (we assume that the x's and q are all "goods"), and strictly quasiconcave in x (we assume strict quasiconcavity rather than quasiconcavity in order to rule out demand correspondences). At this point, we do not assume that u( ) is (strictly) quasiconcave in q. The individual chooses his consumption of the marketed goods - the x's - by maximizing his utility subject to a budget constraint

$$\max_{x} \quad u(x,q) \quad \text{s.t.} \quad \Sigma p_i x_i = y, \quad x_i \geq 0 \quad i=1..N \tag{1}$$

where the p_i's are the prices of the marketed goods, and y is the
individual's income. Note that he does not determine the level
of the q variables. These are in the nature of public goods for
him, and he takes them as given.

The utility maximization generates a pattern of consumption
behavior represented by the ordinary demand functions $x_i = h^i$ (p,
q,y) $i=1,\dots,N$. For convenience we assume that these represent
an interior solution, so that problems associated with corner
solutions (discussed in Bockstael, Hanemann, and Strand [1984,
Chapter 9] can be ignored. Substitution of these demand
functions into the direct utility function yields the indirect
utility function $v(p,q,y) \equiv u[h(p,q,y),q]$. Alternatively, as a
dual to (1) there is an expenditure minimization problem

$$\min_{x} \sum p_i x_i \quad s.t. \quad u(x,q) = u, \quad x_i \geq 0 \qquad (2)$$

which yields a set of compensated demand functions, $x_i = $
$g^i(p,q,u)$, and the expenditure function $m(p,q,u) \equiv \sum p_i g^i(p,q,u)$.

These constructs can be employed to define what we mean by
the benefits to the individual from a change in q. Suppose that
q changes from q^0 to q^1, while prices and income remain constant
at (p,y). Accordingly, the individual's welfare changes from $u^0 \equiv$
$v(p,q^0,y)$ to $u^1 \equiv v(p,q^1,y)$. Two alternative measures of this

welfare change are the quantities C and E defined, respectively, by

$$v(p, q', y - C) = v(p, q^0, y) \tag{3}$$

$$v(p, q', y) = v(p, q^0, y + E). \tag{4}$$

Each of these represents an adjustment to the individual's income calculated to offset the effects of the change in q. C, the compensating variation, is the amount of money by which the individual's income must be adjusted after the change in order to render him as well off as he was before the change. If $u^1 < u^0$, so that $C < 0$, this is the minimum compensation that he would require in order to acquiesce in the change. Similarly, E, the equivalent variation, is the amount of money by which the individual's income must be adjusted before the change in order to render him as well off as he would be after it. If $u^1 > u^0$, so that $E > 0$, this is the minimum compensation that he would require in order to forego the change while, if $u^1 < u^0$ so that $E < 0$, this is the most he would be willing to pay to avoid the change.

The second model is based on the household production approach, in which the individual gains utility from "composite commodities" which he produces himself from private goods. One

version of this model is

$$\max_{x,z} \ w(x,z) \ \text{s.t.} \ f(x,z,q) = 0 \ , \ \Sigma p_i x_i = y \qquad (5)$$

where z denotes the vector of composite goods, f(.) is the

production function for these goods written in implicit form, and

w(.) is a utility function defined over the z's and, perhaps,

some of the x's. In this formulation we are assuming that the

individual derives utility from q not directly, but indirectly,

in so far as they contribute to the production of z's. The

utility maximization in (5) can be solved in two stages. In the

first stage one obtains

$$u(x,q) \equiv \max_{z} \ w(x,z) \ \text{s.t.} \ f(x,z,q) = 0, \qquad (6)$$

while in the second stage one solves (1) using the function

u(x,q) derived from (6). That is to say, a household production

model can always be "collapsed" into a model in the form given in

(1). Moreover, welfare measures for changes in q can be defined

as in (3) and (4) using the indirect utility function v(p,q,y)

associated with u(x,q) in (6). One consequence of the household

production approach, however, is that it generates demand (and

supply) curves for the z's - as well as demand curves for the x's

- which are of some empirical as well as theoretical interest.

Given this framework, our analysis will be concerned with

three sets of issues that have arisen in the literature on

environmental benefit evaluation; (i) What is the relation

between C and E - we know they must have the same sign, but how

much can they differ in magnitude? (ii) How can we measure C and E from observed demand behavior - after all, since we do not observe utility directly we cannot estimate the indirect utility function v(p,q,y) directly? (iii) Is there any relation between C or E and expenditures on some of the private goods - the x's - which might be specially related to the q's in terms of either consumer preferences or household production technology? Can we use expenditure on some goods as proxies for C or E?

To answer these questions, it is convenient to consider three possible markets. One is the market for x's, in which there are observable demand curves. The second is the market for z's, which may arise in connection with the household production model (5). The third market is entirely hypothetical. Suppose that the individual could actually buy q in a market at some given price, π . Instead of (1) he would now solve

$$\max_{x,q} \quad u(x,q) \quad s.t. \quad \sum p_i x_i + \pi q = y \tag{7}$$

(at this point we assume strict quasiconcavity of u(.) with respect to q in order to ensure an interior solution). Denote the resulting ordinary demand functions for the x's by $\hat{h}^i$ (p,π,y), and the ordinary demand function for q by $\hat{h}^q$ (p,π,y). The associated indirect utility function is denoted by $\hat{v}$ (p,π,y) $\equiv$ u[$\hat{h}$(p,π,y), $\hat{h}^q$(p,π,y)]. Similarly, we could define a dual expenditure minimization problem analogous to (2), in which both the x's and q are the choice variables. The resulting

compensated demand functions are denoted by $\hat{g}(p,\bar{\pi},u)$ and $\hat{g}^q(p,\bar{\pi},u)$, and the expenditure function is $\hat{M}(p,\pi,u) \equiv \Sigma\, p_i\, \hat{q}^i(p,\pi,u) + \pi\, \hat{q}^v(p,\pi,u)$.

These utility maximization and expenditure minimization problems are hypothetical because, in fact, environmental quality, q, is not a marketed commodity. Nevertheless, they are of theoretical interest because they shed light on the solutions to (1), (2), (5), and (6). For example, it is convenient to introduce the following:

DEFINITION: q is normal (inferior) if $\hat{h}^v_y > 0 \; (< 0)$.

(We shall now adopt the convention of using subscripts to denote derivatives.)

Maler (1974) proved:

PROPOSITION 1: Assume that, if elements of q change, they all change in the same direction. Then, if all of the q's which change are normal (inferior), $C \leq E \;\; (C \geq E)$.

Moreover,

PROPOSITION 2: If $\hat{h}^v_y = 0$ for all the q's which change, C = E.

Suppose, however, that there <u>are</u> income effects in the demand functions for q; the question remains: just how much can C and E differ? To answer this, we must investigate the q-market in more detail.

HOW MUCH CAN C AND E DIFFER?

Willig (1976) established that, unless the income elasticity

of demand for a commodity is very high, the compensating and

equivalent variations for a price change will not differ

considerably.   Some environmental economists do not believe that

the same holds true of compensating and equivalent variations for

change in q - see, for example, Maler (1985, p.39) or Knetsch and

Sinden (1984), who present empirical evidence of a considerable

disparity between C and E. However, Randall and Stoll (1980) have

shown that Willig's analysis carries over to changes in fixed

parameters such as the q's, and Brookshire, Randall and Stoll

(1980) have interpreted this result as implying that C and E

should not be very different in value.   How can these divergent

views be explained or reconciled?

In the paper reproduced in the Appendix to this chapter I

reexamine randall and Stoll's analysis and show that, while it is

indeed accurate, its implications have been misunderstood.   There

is no presumption that C and E must be close in value and, unlike

price changes, the difference between them depends not only on an

income effect but also on a substitution effect.   Specifically,

the magnitude of the difference depends on (i) the magnitude of

the change in q, (ii) the size of the income effects, and (iii)

the degree of substitutability between private consumption

activities (the x's) and the level of environmental quality q in the individual's preferences, all of which are empirical issues. Moreover, I suggest that the substitution effects are likely to exert far greater leverage, in practice, on the relation between C and E than the income effects. Thus, large empirical divergences between C and E may be indicative not of some failure in the survey methodology but of a general perception on the part of the individuals surveyed that the private market goods available in their choice set are, collectively, a rather imperfect substitute for the public good under consideration.

MEASURING C AND E FROM DEMAND CURVES

Analysis of the market for q is useful in that it gives us an idea of the factors that affect the relation between C and E, but it is of no value when it comes to measuring C or E in practice because, by definition, no such market exists - the demand curve for q can never be observed. What _can_ be observed is behavior in the x market - the market for private goods. This raises the question, therefore, of whether the values of C and E can be inferred from knowledge of the demand curves from the x's. There are two ways in which this can be accomplished. The first is to uncover the in direct utility function from the fitted demand curves for the x's, and then employ the formulas in (3) and (4). The second is based upon results developed by Maler (1971, 1974) which establish a relation between areas under demand curves for

the x's and the quantities C and E.

In the first approach one postulates a specific functional
form for either the direct utility function u(x,q) or the
indirect utility function v(p,q,y), and derives the appropriate
formula for the corresponding ordinary demand functions - by
analytically solving the direct utility maximization problem or
by differentiating the indirect utility function and applying
Roy's Identity. Alternatively, one can start out with a given
system of ordinary demand functions h^i (p,q,y) $i=1,..,N$, and then
attempt to recover the corresponding indirect utility function by
applying the integrability techniques developed by Hurwicz and
Uzawa. As a simple example, suppose that N=2 and the demand
function for the first good takes the semi-log form

$$\ln x_1 = \alpha - \beta (p_1/p_2) + \gamma (y/p_2) + \delta q ; \tag{8}$$

in Hanemann (1980a, 1981) it is shown that the indirect utility
function is

$$v(p_1, p_2, q, y) = - \frac{e^{-\gamma(y/p_2)}}{\gamma} + \frac{A e^{\delta q - \beta(p_1/p_2)}}{\beta} \tag{9}$$

where $A = e^{\alpha}$. Application of (3) and (4) yields the following
formulas for C and E

$$C = - \frac{p_2}{\gamma} \ln \left\{ 1 + \frac{\gamma}{\beta} (x_1' - x_1^0) \right\} \tag{10a}$$

$$E = \frac{p_2}{\gamma} \ln \left\{ 1 - \frac{\gamma}{\beta} (x_1' - x_1^0) \right\} \tag{10b}$$

where $x^0 = h$ (p,q^0,y) and $x' = h$ (p,q^1, y). Thus, to estimate C and

E one first fits the demand function (8) and then substitutes the estimated values of the coefficients α, β, γ and δ into the formulas in (10a,b).

The alternative approach to computing C and E· developed by Maler, is based on the following decomposition of the formula for C (a similar analysis applies to E)·

$$C = y - m(p, q', \cup^o)$$

$$= \left\{ m(p, q^o, \cup^o) - m(\tilde{p}, q^o, \cup^o) \right\} - \left\{ m(p, q', \cup^o) - m(\tilde{p}, q', \cup^o) \right\} + \left\{ m(\tilde{p}, q^o, \cup^o) - m(\tilde{p}, q', \cup^o) \right\}$$

$$= \int_p^{\tilde{p}} \sum_i \left[g^i(p, q', \cup^o) - g^i(p, q^o, \cup^o) \right] dp_i + \left\{ m(\tilde{p}, q^o, \cup^o) - m(\tilde{p}, q', \cup^o) \right\} \tag{11}$$

where $\tilde{p}$ is an arbitrary price vector. Assuming that $q^1 > q^0$, we know that C > 0. Since $m_q \leq 0$, we also know that the second term in (11) is non-negative. The first term is the sum of areas between compensated demand curves corresponding to q^1 and q^0, between the actual price p_i and the ith element of $\tilde{p}$ (this line integral is path-independent). It should be emphasized that the first item is not necessarily positive; it can be shown that the increase in q raises the compensated demand for the ith private good $(\partial g^i / \partial q > 0)$ if this good is a complement to q in the Hicks-Allen sense, and lowers the compensated demand $(\partial g^i / \partial q < 0)$ if the good is a substitute. Moreover, if q is a scalar, at least one of the private goods must be a Hicks-Allen substitute for q. Nevertheless, we know that the sum of the two terms in (8) must

be positive.

Maler's trick is to select $\tilde{p}$ in such a way that the second term in (11) vanishes. For this purpose, he introduces two assumptions. The novel assumption is that there exists a set of commodities with the property that, if these commodities are not being consumed, the marginal utility of q is zero. Let I be the index set of these commodities, and $\bar{I}$ its complement. Partition the vector x accordingly: x = (x_I , $x_{\bar{I}}$). Maler's assumption, which he calls weak complementarity, is:

(WC)   There exists a non-empty set I such that $\dfrac{\partial u(0, x_{\bar{I}}, q)}{\partial q} = 0$

$$(12)$$

His second assumption is:

(NE) The commodities in I are non-essential: there exists some price vector such that $g^i(.) = 0$ and $h^i(.) = 0$ all $i \in I$.

We can now apply these assumptions to (11) by choosing the price vector $\tilde{p}$ so that $\tilde{p}_i = p_i$ for $i \in \bar{I}$ while, for $i \in I$, $\tilde{p}_i$ is simply the cut-off price of the ith compensated demand function - i.e. $\max\left[q^i(\tilde{p}_I, p_{\bar{I}}, q^0, u^0), q^i(\tilde{p}_I, p_{\bar{I}}, q^1, u^0) \right] = 0$. Since sign ($m_q$) = - sign ($u_q$), this yields Maler's result:

PROPOSITION 3;   If u(x,q) satisfies (WC) and (NE),

$$(13)$$

$$C = \int_{p_I}^{\tilde{p}_I} \sum_{i \in I} \left[q^i(p, q^1, u^0) - q^i(p, q^0, u^0) \right] dp_i .$$

This proposition establishes a relationship between C and the areas between two sets of compensated demand functions. It is useful here to make a distinction between two sets of circumstances: (i) there is a set of goods with the property that q has no value only when _none_ of these goods is being consumed, and (ii) there is a set of goods with the property that q has no value when any _one_ of them is not being consumed. In the first case, C is measured by the area between compensated demand curves summed over all of the goods in I; in the second case it is measured by the area between compensated demand curves for any one of the goods in I, and we obtain the same answer regardless of the particular good selected. Note that, in order to make use of the proposition, one still needs to know something more than ordinary demand functions unless there are no income effects in the demand for the goods in I, in which case the compensated and ordinary demand functions coincide. If there _are_ income effects and one attempts to calculate the area in (13) using ordinary instead of compensated demand functions, i.e. one calculates the area

$$S = \int_{P_I}^{\tilde{P}_I} \sum_{i \in I} \left[h^i(p, q^1, y) - h^i(p, q^0, y) \right] dp_i,$$ (14)

this is likely to be of limited value. The issue is examined in Hanemann (1980b), where it is shown that under some circumstances S may not even have the correct sign. The requirement that one employ the compensated demand function in (13) implies that,

wherever there are income effects, Maler's method for calculating

C and E has the same information requirements as the method based

on direct application of (3) and (4). Finally, as an

illustration, it turns out that semi-log demand function (8)

satisfies the WC condition since, on differentiating the indirect

utility function (9), one finds that

$$\lim_{P_1 \to \infty} \frac{\partial v(p, q, y)}{\partial q_0} = 0, \tag{15}$$

which is equivalent to (12). The compensated demand function

corresponding to (13) is

$$x_1 = g'(p, q, u) = \frac{\beta}{\gamma} \left[1 - \frac{u\beta}{\gamma A} e^{\beta(P_1/P_2) - \delta q_0} \right]^{-1} \tag{16}$$

and it is straightforward to verify that (10a) and (10b) combine

to satisfy (13).

THE LIMITS TO REVEALED PREFERENCE

Both of the methods for measuring C and E from observed

demand functions rely on the assumption that all the relevant

components of the indirect utility function can be recovered from

demand functions. However, that assumption is not always true:

it holds when the underlying direct utility function has the form

$$u = \bar{u}(x, q) \tag{17}$$

as has implicitly assumed up to now, but not when the utility

function can be cast into the form

$$u = u(x,q) = T[\bar{u}(x,q),q] \tag{18}$$

where T(.) is increasing in its first argument and $\bar{u}(x,q)$ is a conventional direct utility function.   It can be shown that both utility models imply exactly the same ordinary demand functions for x's

$$\arg \max_{x} \bar{u}(x,q) = \arg \max_{x} T[\bar{u}(x,q),q]$$

even though they imply different things about the individual's preferences.   The crucial feature of (18) is that the marginal rates of substitution between the x's - The indifference map for the x's - is independent of the transformation function T(.), even though that function influences the way in which q affects the individual's utility.   This does not arise in the case of the utility function in (17).   Thus, with (17), all aspects of the individual's preferences for q are captured directly or indirectly in his ordinary demand functions for x's.   This is not so for (18): some aspects of the individual's preferences for q are not reflected in his ordinary demand functions, not even indirectly.

Another way of making the same point is to observe that the compensating variation for a change from (p,q^0,y) to (p,q^1,y) associated with u(x,q), C, can be decomposed into two elements

$$C = \bar{C} + C^{*}, \tag{19}$$

where $\overline{C}$ satisfies $\overline{v}(p,q^1,y-\overline{C}) = \overline{v}(p,q^0,y)$, $\overline{v}(.)$ being the indirect utility function corresponding to $\overline{u}(x,q)$, and C^* satisfies

$$T\left[\overline{v}(p,q^0,y-C^*), q'\right] = T\left[\overline{v}(p^0,q^0,y), q^0\right]. \tag{20}$$

Assuming that $q^1 > q^0$ and $T(.,q)$ is increasing in q, it can be shown that $C^* > 0$, so that

$$C > \overline{C} > 0. \tag{21}$$

A similar result can be shown to hold for equivalent variation measures:

$$E = \overline{E} + E^* > \overline{E} > 0, \tag{22}$$

where E is the true equivalent variation associated with the full utility function u(x,q) in (18), $\overline{E}$ is the equivalent variation associated with the sub-function $\overline{u}(x,q)$, and E^* is calculated from the transformation function $T(.,q)$, along the lines of (20). Since $\overline{C}$ and $\overline{E}$ are derived from the sub-function containing the interactions between the x's and q, we can regard them as the "consumption - or use - related" components of benefits. Similarly, we can regard C^* and E^* as the "non-consumption related" or "non-use related" components of benefits - they arise from that part of the individual's preferences which do not affect his choice of x.

The practical implications of (18) for the revealed preference approach - the measurement of C and E on the basis of observed demands for the x's - are highly important. If we only have data on ordinary demand functions for the x's, we can only recover $\bar{u}(x,q)$, but never $T(.,q)$ nor the full utility function $u(x,q)$ in (18). That is, we can only measure $\bar{C}$ and $\bar{E}$ - not C^* or E^* and, therefore, not the full value of C or E. This is a significant limitation to the revealed preference approach.

It is sometimes thought that Maler's Weak Complementarity (WC) assumption eliminates this problem, but I would dispute this. Differentiate (18) to obtain the marginal utility of q.

$$\frac{\partial u(x,q)}{\partial q} = \frac{\partial T}{\partial \bar{u}} \frac{\partial \bar{u}(x,q)}{\partial q} + \frac{\partial T}{\partial q} . \tag{23}$$

If we apply WC to $u(x,q)$, this requires that

$$x_I = 0 \implies \frac{\partial \bar{u}(0, x_{\bar{I}}, q)}{\partial q} = 0 \quad \text{and} \quad \frac{\partial T[u(0, x_{\bar{I}}, q), q]}{\partial q} = 0. \tag{24}$$

But, by itself, this is not enough to ensure that

$$\frac{\partial T[., q]}{\partial q} \equiv 0, \tag{25}$$

which is what one requires in order to rule out the

representation in (18). Suppose, for example, that

$$u(x,q) = \begin{cases} \bar{u}(x,q) & \text{if } x_I = 0 \\ T[\bar{u}(x,q),q] & \text{if } x_I > 0. \end{cases} \qquad (26)$$

This satisfies (24) but not (25), and therefore $C^* > 0$ and $E^* > 0$. In this case WC does not eliminate the problem.

To summarize, the only circumstance in which the revealed preference approach to the measurement of C and E is fully satisfactory is when (25) holds - i.e. the utility function is represented by (17) rather than (18). But there is no way to verify this from data on ordinary demand functions for x's. It could be verified if there were a market for q and one could observe demand functions for q as well as the x's. Indeed, in that case, T(.,q) could be recovered along with u(x,q) so that, if (25) were violated, C and E could still be calculated because one would obtain the full indirect utility function associated with (18). But, in the absence of a market for q, the problem remains.

In practice, there are two possible solutions. The first is simply to assume that the utility function takes the form of (17) and not (18) - which is what is generally done. The second is to collect additional behavioral data besides ordinary demand functions for the x's. For example, after measuring $\bar{C}$ by the

revealed preference approach one could conduct interviews to elicit the willingness to pay for an improvement in q directly; if the interviews yielded an estimate close to $\bar{C}$ in value one would conclude that $C^* = 0$ and hence, the utility model corresponds to (17) rather than (18). If they yielded an estimate much greater than $\bar{C}$ one would take the difference to be a measure of C^*. Alternatively, instead of contingent valuation exercises, one could conduct what has been called [Hanemann (1985)] "contingent behavior" exercises in which one attempts to elicit a hypothetical demand function for q. Both of these approaches remain subjects for future research.

THE SIGNIFICANCE OF EXPENDITURE DATA

In the theory of the welfare measurement of price changes it is well known that calculation of expenditure changes provide <u>bounds</u> on the compensating and equivalent variations, even if they are not exactly equal to these welfare measures. If prices change from p^0 to p^1 and the quantities demand change correspondingly from x^0 to x^1, then the compensating variation for the price change, C^P, and the equivalent variation, E^P, satisfy

$$C^P \geq \sum_i (p_i^0 - p_i^1) x_i^0$$

and

$$E^P \leq \sum_i (p_i^0 - p_i^1) x_i^1 .$$

although, in general, there is no determinate relation between C

or E and the overall change in expenditure $\sum p_i^\circ x_i^\circ - \sum p_i' x_i'$.

When dealing with changes in q, as opposed to price changes, some authors have wondered whether one can obtain a relation between the welfare measures C and E and the change in expenditures on some or all of the private market goods, $\sum p_i [h^i(p, q^\circ, y) - h^i(p, q', y)]$. In general, I do not believe that this is a useful approach; with one exception described below, there does not appear to be any determinate relation between changes in expenditure on x's and either C and E. Indeed, the effect of an increase in q on the demand function for any of the x's is by no means obvious. Given that $(\partial u/\partial q) > 0$, it is sometimes assumed that $\partial h^i/\partial q \geq 0$ all i - an increase in quality can never <u>lower</u> the demand for any of the x's. In fact, this is <u>not</u> true; in general, an increase in q <u>will</u> affect the demand for the x's, but note that the effect could be in either direction, depending on the specifics of the utility function. Even if q is a Hicks-Allen complement with some private good -say, x_i - it is not necessarily true that an increase in q will raise the demand for that good.

This pessimistic conclusion is based on the following proposition which links the demand functions $x_i = h^i(p, q, y)$ to the hypothetical demand functions $x_i = \hat{h}^i(p, \pi, y)$ associated with the utility maximization problem (7):

PROPOSITION 4:   Let $\Pi = \hat\Pi$ (p,q,y) be defined implicitly by

Appendix equation (11).   Then,

$$h^i(p,q,y) = \hat{h}^i[p, \hat\Pi(p,q,y), y + \hat\Pi(p,q,y)\cdot q] \qquad i=1,..,N. \qquad (27)$$

It follows as a corollary that

$$\frac{\partial h^i}{\partial q}(p,q,y) = \hat\Pi(p,q,y)\cdot\frac{\partial\hat{h}^i}{\partial y} + \left[\frac{\partial\hat{h}^i}{\partial\Pi} + q\frac{\partial\hat{h}^i}{\partial y}\right]\cdot\frac{\partial\hat\Pi}{\partial q}(p,q,y)$$

$$(28)$$

$$= \hat\Pi\cdot\frac{\partial\hat{h}^i}{\partial y} - \left\{\frac{\partial\hat{h}^i/\partial\Pi + q\,\partial\hat{h}^i/\partial y}{\partial\hat{h}^q/\partial\Pi + q\,\partial\hat{h}^q/\partial\Pi}\right\}\left(\hat\Pi\frac{\partial\hat{h}^q}{\partial y}-1\right).$$

Given that $u_q > 0$, $\hat\Pi > 0$.   If $u(x,q)$ is quasiconcave in q, the

denominator of the second term on the RHS is negative.   Thus, the

sign of $\partial h^i/\partial q$ depends upon a complex set of factors.   The

numerator of the term in braces on the RHS will be recognized as

the cross-price derivative of the compensated demand curve from q

$$\frac{\partial\hat{g}^q(p,\hat\Pi,u)}{\partial p_i} = \frac{\partial\hat{g}^i(p,\hat\Pi,u)}{\partial\Pi} = \frac{\partial\hat{h}^i}{\partial\Pi} + q\frac{\partial\hat{h}^i}{\partial y}$$

and this is positive or negative according as x_i and q are

substitutes or complements.   Moreover,

$$\hat\Pi\frac{\partial\hat{h}^q}{\partial y}-1 = \omega\eta-1 \gtrless 0 \quad as \quad \eta \gtrless \frac{1}{\omega}$$

where $\omega \equiv \frac{\hat\Pi q}{y}$.   Thus, if $\partial\hat{h}^o/\partial y > 0$ and

$$[\omega\eta-1]\,\frac{\partial\hat{g}^q}{\partial p_i} > 0 \qquad (29)$$

this is a sufficient condition for $\partial h^i/\partial q > 0$. Even if $\partial\hat{h}^i/\partial y < 0$, it

can still happen that $\partial h^i/\partial q > 0$ if (29) holds and that term is sufficiently large.

Without belaboring it further, the point is that an increase in q could either lower or raise the expenditure on x. This should make us cautious about expected any simple relation between the change in expenditure on some of the x's and C or E since it is quite possible that C and E are positive while the change in expenditure is negative. One case in which more definitive results can be obtained is where q is a perfect substitute for some of the x's - say x_1. In that case the direct utility function takes the form

$$v(x,q) = \bar{v}\,[\,x_1 + \psi(q),\, x_2,\ldots,\, x_N\,] \tag{30}$$

where $\psi(.)$ is some increasing function of q. Let $\bar{h}^1(p,y)$ and $\bar{v}(p,y)$ be the ordinary demand function for good 1 and the indirect utility function associated with $\bar{u}(.)$. The following may be shown:

PROPOSITION 5: If u(x,q) has the form given in (30),

$$h^1(p,q,y) = -\psi(q) + \bar{h}^1[\,p_1,\ldots,p_N,\, y + p_1\psi(q)\,] \tag{31a}$$

$$v(p,q,y) = \bar{v}[\,p_1,\ldots,p_N,\, y + p_1\psi(q)\,]\;. \tag{31b}$$

It follows from (31b) that

$$C = E = P_1 [\psi(q') - \psi(q^0)],$$

while the change in expenditure on x is

$$\Delta = P_1 [h'(p, q^0, y) - h'(p, q', y)]$$

$$= P_1 [\psi(q') - \psi(q^0)] + P_1 \{ \bar{h}'(p, y + P_1 \psi(q^0)) - \bar{h}'(p, y + P_1 \psi(q')) \}$$

$$\lessgtr C, E \quad as \quad \frac{\partial \bar{h}'}{\partial y} = \frac{\partial h'}{\partial y} \gtrless 0 . \tag{32}$$

Thus, if x_1 is a normal good and a perfect substitute for q, the change in the expenditure on x_1 understates the true benefit from an increase in q. In this case, moreover, there are no income effects in the demand curve for q, so that the compensating and equivalent variations coincide. Apart from this special case, however, there does not appear to be any determinate relation between Δ and C or E.

NON-USE VALUES

This above framework can be used to shed some light on the concept of existence value due originally to Krutilla (1967). This is based on the notion that, even if he did not consume any of the x's that are associated with q, an individual might still feel some improvement in q and be willing to pay something to secure it. How can this be explained in terms of the utility model discussed above?

Smith and Desvousges (1986) have made an important
distinction between existence values under conditions of
certainty and uncertainty. The phenomenon of consumer choices
under uncertainty -e.g. the individual does not know whether or
not he will want in the future to consume certain x's that are
associated with q - raises many important issues that transcend
the theory developed above, which is firmly rooted in the context
of decisions under certainty. Accordingly, I focus here on the
concept of existence values under the conditions of certainty -
an individual places some value on an improvement in q even
though he does not himself consume any of the x's that might be
associated with q, and has no doubt that he will never consume
these goods in the future. Under these circumstances, how can we
use the theoretical framework developed above to give some
operational meaning to this concept?

Two quantities identified above may have some bearing on
this question. The first is based on the decomposition in (11).
Suppose that Weak Complementarity does not apply so that $\partial u/\partial q > 0$
even when there is zero consumption of x's that are
conventionally associated with q. In that case one could regard
the quantity

$$m(\tilde{p}_x, p_{\bar{x}}, q^0, u^0) - m(\tilde{p}_x, p_{\bar{x}}, q^1, u^0) = C - \int_{p_x}^{\tilde{p}_x} \sum_{i \in I} \left[\dot{q}^i(p, q^1, u^0) - \dot{q}^i(p, q^0, u^0) \right] dp_i \qquad (33)$$

as a measure of the non-use benefits associated with the improvement in q - these are the benefits that would accrue to the individual even if he were consuming none of the x_I's. Operationally, one would measure them by computing C from the indirect utility function using (3), and then subtracting the area between the compensated demand curves represented by the integral on the RHS of (33). of course, if Weak Complementarity holds, this quantity is zero. As already noted, that would apply to the semi-log demand function (8). Interestingly, it does <u>not</u> apply to another common functional form, the linear ordinary demand function

$$ x_1 = h'(p_1, p_2, q, y) = \alpha - \beta(p_1/p_2) + \gamma(y/p_2) + \delta q. \tag{34} $$

It can be shown that the corresponding compensated demand function $g'(p,q,u)$ is independent of q so that the integral in (11) and (33) is zero and

$$ C = m(\tilde{p}_1, p_2, q^0, u^0) - m(\tilde{p}_1, p_2, q', u^0) \tag{35} $$

where the cut-off price is

$$ \tilde{p}_1 = \frac{p_2}{\beta}\left(\alpha + \gamma\frac{y}{p_2} + \delta q'\right). \tag{36} $$

In this case, therefore, <u>all</u> of the benefit from a change in q is

associated with term $[m(\breve{p}, q^0, u^0) - m(\hat{p}, q^1, u^0)]$.   For this reason it may appear unsatisfactory to equate that quantity with the notion of non-use or "existence" value.

The other candidate is the quantity C^* (and E^*) defined in (19) and (22) in connection with the utility representation (18). That is to say, if the utility function is represented by (18) rather than (17), one could regard the "extra" component of benefits over and above $\bar{C}$ or $\bar{E}$ as a form of non-use value. This interpretation was, indeed, suggested above. An extreme example arises when the utility function takes the form

$$u(x, q) = T[\bar{u}(x), q]$$
(37)

i.e. $\bar{u}(x, q)$ is completely independent of q. In that case the demand function for the x's are entirely independent of q – $x_i = \dot{h}^i(p, y)$ all i — but the individual still places some value on changes in q. From (19)

$$C = C^*$$
(38)

where C^* is defined by (20). In this special case the revealed preference approach provides no information (except to confirm that $\bar{C} = \bar{E} = 0$) and the only way to measure C is through some form of contingent valuation or contingent behavior experiment. If the utility function has the general form (18), but not the

extreme form in (37), a similar conclusion would apply: the only way to measure the non-use benefits C^* and E^* is by contingent valuation and/or contingent behavior procedures.

REFERENCES

W. Michael Hanemann (1980a) "Measuring the Worth of Natural Resource facilities: Comment" Land Economics Vol. 56 November 1980 pp. 482-486

W. Michael Hanemann (1980b) "Quality Changes, Consumer's Surplus and Hedonic Price Indices," University of California, Department of Agricultural and Resource Economics, Working paper No. 116, Berkeley, November 1980.

W. Michael Hanemann (1981) "Some Further Results on Exact Consumer's Surplus" University of California, Department of Agricultural and Resource Economics, Working paper No. 190, Berkeley, December, 1981.

W. Michael Hanemann (1985) "Some Issues in Discrete - and Continuous - Response Contingent Valuation Studies," Northeastern Journal of Agricultural Economics, April 1985

John V. Krutilla (1967) "Conservation Reconsidered," American Economic Review Vol. 57, September 1967, pp 777-786

Karl-Goran Maler (1971), "A Method of Estimating Social Benefits From Pollution Control," Swedish Journal of Economics Vol. 73, No. 1, pp 121-133

Karl-Goran Maler (1974), Environmental Economics: A Theoretical Inquiry (Baltimore: The Johns Hopkins University Press)

V. Kerry Smith and William H. Desvousges (1986), Measuring Water Quality Benefits (Boston: Kluwer-Nijhoff Publishing Co.)

WILLINGNESS TO PAY AND WILLINGNESS TO ACCEPT: HOW MUCH CAN THEY DIFFER?

Consider an improvement in the exogenous variables comprising an individual's choice set. Two possible monetary measures of the gain in her welfare are the compensating variation (C) and the equivalent variation (E). In the present context, these correspond, respectively, to the maximum amount the individual would be willing to pay (WTP) to secure the change and the minimum compensation that she would be willing to accept (WTA) to forego the change. How much can the two differ, and what are the factors that determine the difference? These questions were addressed by Robert Willig (1976) in his path-breaking paper on the welfare measurement of price changes. Willig argued that C and E are likely in practice to be fairly close in value, and he showed that the difference depends directly on the size of the income elasticity of demand for the commodity whose price changes.

In many empirical studies, however, analysts seek to obtain money measures of welfare changes due not to price changes but to changes in the availability of public goods or amenities, changes in the qualities of commodities, or changes in the fixed quantities of rationed goods. Karl-Göran Mäler (1974) was perhaps the first to show that the concepts of C and E can readily be extended from conventional price changes to quantity changes such as these. Subsequently, Alan Randall and John Stoll (1980) examined the duality theory associated with fixed quantities in the utility function and showed that, with appropriate modifications, Willig's formulas for bounds on C and E do, indeed, carry over to this setting. Within the environmental literature and elsewhere, Randall and Stoll's results have been widely interpreted as implying that WTP and WTA for changes in environmental amenities should not differ

greatly unless there are unusual income effects.[1] However, recent empirical
work using various types of interview procedures has produced some evidence of
large disparities between WTP and WTA measures--for example, Richard C. Bishop
and Thomas A. Hebertein (1979) and several studies described by Irene M.
Gordon and Jack L. Knetsch (1979), and by Knetsch and Sinden (1984). This has
led to something of an impasse: How can the empirical evidence of significant
differences between WTP and WTA be reconciled with the theoretical analysis
suggesting that such differences are unlikely? Can they be explained entirely
by unusual income effects or by peculiarities of the interview process?

In this note I reexamine Randall and Stoll's analysis and show that, while
it is indeed accurate, its implications have been misunderstood. For quantity
changes there is no presumption that WTP and WTA must be close in value and,
unlike price changes, the difference between WTP and WTA depends not only on
an income effect but also on a substitution effect. By the latter, I mean the
ease with which other privately marketed commodities can be substituted for
the given public good or fixed commodity, while maintaining the individual at
a constant level of utility. I show that, holding income effects constant,
the smaller the substitution effect (i.e., the fewer substitutes available for
the public good) the greater the disparity between WTP and WTA. This surely
coincides with common intuition. If there are private goods which are readily
substitutable for the public good, there ought to be little difference between
an individual's WTP and WTA for a change in the public good. But, if the pub-
lic good has almost no substitutes (e.g., Yosemite National Park or, in a dif-
ferent context, your own life), there is no reason why WTP and WTA could not
differ vastly--in the limit, WTP could equal the individual's entire (finite)
income while WTA could be infinite. My argument is developed in the following

two sections. Section I deals specifically with the two polar cases of perfect substitution and zero substitution between the public good and available private goods. Section II deals with Randall and Stoll's extension of Willig's formulas and shows that their bounds are, in fact, consistent with substantial divergences between WTP and WTA.

I. Two Polar Cases

The theoretical setup is as follows. An individual has preferences for various conventional market commodities whose consumption is denoted by the vector x as well as for another commodity whose consumption is denoted by q.[2] This could represent the supply of a public good or amenity; it could be an index of the quality of one of the private goods; or it could be a private commodity whose consumption is fixed by a public agency.[3] The key point is that the individual's consumption of q is fixed exogenously, while she can freely vary her consumption of the x's. These preferences are represented by a utility function, $u(x, q)$, which is continuous and nondecreasing in its arguments (I assume that the x's and q are all "goods") and strictly quasiconcave in x. The individual chooses her consumption by solving

$$(1) \qquad \max_{x} u(x, q) \text{ subject to } \Sigma p_i x_i = y$$

taking the level of q as given. This yields a set of ordinary demand functions, $x_i = h^i(p, q, y)$, $i = 1, \ldots, N$, and an indirect utility function, $v(p, q, y) \equiv u[h(p, q, y), q]$, which has the conventional properties with respect to the price and income arguments and also is increasing in q.[4] Now suppose that q rises from q^0 to $q^1 > q^0$ while prices and income remain constant

at (p, y). Accordingly, the individual's utility changes from $u^0 \equiv v(p, q^0, y)$ to $u^1 \equiv v(p, q^1, y) \geq u^0$. Following Mäler, the compensating and equivalent variation measures of this change are defined, respectively, by[5]

$$(2) \qquad\qquad v(p, q^1, y - C) = v(p, q^0, y)$$

$$(3) \qquad\qquad v(p, q^1, y) = v(p, q^0, y + E).$$

Dual to the utility maximization in (1) is an expenditure minimization: Minimize $\Sigma p_i x_i$ with respect to x subject to $u = u(x, q)$, which yields a set of compensated demand functions, $x_i = g^i(p, q, u)$, $i = 1, \ldots, N$, and an expenditure function, $m(p, q, u) \equiv \Sigma p_i g^i(p, q, u)$, which has the conventional properties with respect to (p, u) and is decreasing in q. In terms of this function, C and E are given by

$$(2') \qquad\qquad C = m(p, q^0, u^0) - m(p, q^1, u^0)$$

$$(3') \qquad\qquad E = m(p, q^0, u^1) - m(p, q^1, u^1).$$

It is evident from (2) and (3) that $0 \leq C \leq y$ while $E \geq 0$.[6] The questions at issue are: (1) Is it true that $E/C \approx 1$? (2) What factors affect this ratio? As a first cut at an answer, I compare two polar cases. In the first case at least one private good--say, the first--is a perfect substitute for some transformation of q. Thus, the direct utility function assumes the special form

$$(4) \qquad\qquad u(x, q) = \bar{u}[x_1 + \psi(q), x_2, \ldots, x_N]$$

where $\psi(\cdot)$ is an increasing function and $\bar{u}(\cdot)$ is a continuous, increasing, strictly quasiconcave function of N variables. As W. M. Gorman (1976) has shown, the resulting indirect utility function is

$$(5) \qquad v(p, q, y) = \bar{v}[p_1, p_2, \ldots, p_N, y + p_1 \psi(q)]$$

where $\bar{v}(\cdot)$ is the indirect utility function corresponding to $\bar{u}(\cdot)$. Substitution of (5) into (2) and (3) yields the following:[7]

PROPOSITION 1: If at least one private market good is a perfect substitute for q, then C = E.

At the opposite extreme, I assume that there is a zero elasticity of substitution not just between q and x_1 but between q and all the x's. Thus, the direct utility function becomes

$$(6) \qquad u(x, q) = \bar{u}\left[\min\left(q, \frac{x_1}{\alpha_1} \right), \ldots, \min\left(q, \frac{x_N}{\alpha_N} \right) \right].$$

where $\alpha_1, \ldots, \alpha_N$ are positive constants and $\bar{u}(\cdot)$ is a conventional direct utility function. In this case the indirect utility function $v(p, q, y)$ has a rather complex structure and changes its form in different segments of (p, q, y) space. It will be sufficient for my purposes to focus on just one of these segments. Suppose that $q \leq y/\Sigma p_i \alpha_i$; then the maximization of (6), subject to the budget constraint, yields ordinary demand functions and an indirect utility function of the form $x_i = h^i(p, q, y) = \alpha_i q$, and $u = v(p, q, y) = \bar{u}(q, \ldots, q) \equiv w(q)$. In this region of (p, q, y) space, the individual does not exhaust her

budget, and her marginal utility of income is therefore zero. Now suppose that $q^0 \leq y/\Sigma p_i \alpha_i$ and $q^1 > q^0$. Since $v(p, q^1, y) > w(q^0)$, it is evident from (2) that the individual would be willing to pay some positive but limited amount C to secure this change. However, for any positive quantity E, no matter how large, $v(p, q^0, y + E) = v(p, q^0, y) = w(q^0)$. This implies the following:

PROPOSITION 2: If there is zero substitutability between q and each of the private market goods, it can happen that, while the individual would only be willing to pay a finite amount for an increase in q, there is no finite compensation that she would accept to forego this increase.

It should be emphasized that this result obtains only in a portion of (p, q, y) space; in other regions, even with (6), E would be finite.[8] However, the result in Proposition 2 can also be established for other utility functions that permit some substitutability between q and the x's as long as the indifference curves between q and each of the x's become parallel to the q axis at some point. The lesson to be learned from these two propositions is that the degree of substitutability between q and private market goods significantly affects the relation between C and E. In the next section, I show how this observation can be reconciled with the bounds on C and E derived by Randall and Stoll.

II. Randall and Stoll's Bounds

In order to extend Willig's bounds from price to commodity space, Randall and Stoll focus on a set of demand functions different from those considered above. Suppose that the individual could purchase q in a market at some given

price, π. It must be emphasized that this market is entirely hypothetical since q is actually a public good. Instead of (1), she would now solve[9]

(7)
$$\max_{x,q} u(x, q) \text{ subject to } \Sigma p_i x_i + \pi q = y.$$

Denote the resulting ordinary demand functions by $x_i = \hat{h}^i(p, \pi, y)$, $i = 1, \ldots, N$ and $q = \hat{h}^q(p, \pi, y)$. The corresponding indirect utility function is $\hat{v}(p, \pi, y) \equiv u[\hat{h}(p, \pi, y), \hat{h}^q(p, \pi, y)]$. The dual to (7) is: Minimize $\Sigma p_i x_i + \pi q$ with respect to x and q subject to $u = u(x, q)$. This generates a set of compensated demand functions, $x_i = \hat{g}^i(p, \pi, u)$, $i = 1, \ldots, N$ and $q = \hat{g}^q(p, \pi, u)$, and an expenditure function, $\hat{m}(p, \pi, u) \equiv \Sigma p_i \hat{g}^i(p, \pi, u) + \pi \hat{g}^q(p, \pi, u)$. These functions are hypothetical since q is really exogenous to the individual, but they are of theoretical interest because they shed light on the relation between C and E.

For any given values of q, p, and u, the equation,

(8)
$$q = \hat{g}^q(p, \pi, u),$$

may be solved to obtain $\pi = \hat{\pi}(p, q, u)$, the inverse compensated demand (i.e., willingness to pay) function for q: $\hat{\pi}(\cdot)$ is the price that would induce the individual to purchase q units of the public good in order to attain a utility level of u, given that she could buy private goods at prices p. Let $\pi^0 \equiv \hat{\pi}(p, q^0, u^0)$ and $\pi^1 \equiv \hat{\pi}(p, q^1, u^1)$ denote the prices that would have supported q^0 and q^1, respectively. The two expenditure functions dual to (1) and (7) are related by:

(9)
$$m(p, q, u) \equiv \hat{m}[p, \hat{\pi}(p, q, u), u] - \hat{\pi}(p, q, u) \cdot q.$$

This implies that[10]

(10) $$m_q(p, q, u) = -\hat{\pi}(p, q, u).$$

Combining (10) with (2') and (3') yields these alternative formulas for C and E expressed in terms of the willingness-to-pay function:

(2") $$C = \int_{q^0}^{q^1} \hat{\pi}(p, q, u^0)\, dq$$

(3") $$E = \int_{q^0}^{q^1} \hat{\pi}(p, q, u^1)\, dq.$$

It can be shown that sign $(\hat{\pi}_u)$ = sign $(\hat{h}_y^q)$. Therefore, for given (π, q), the graph of $\hat{\pi}(p, q, u^1)$ lies above (below) that of $\hat{\pi}(p, q, u^0)$, and E > (<) C, accordingly as q is a normal (inferior) good. Figure 1 shows E and C for the case where q is normal: E corresponds to the area $q^0 \alpha \gamma q^1$ while C corresponds to the area $q^0 \beta \delta q^1$.

Using the technique pioneered by Willig, Randall and Stoll establish bounds on the difference between each of C and E and the area under an inverse ordinary demand function for q. From this, they derive bounds on the difference between C and E. However, the requisite inverse ordinary demand function is obtained in a rather special manner. Given any level of q, we can ask what market price π would induce the individual to purchase that amount of public good if it were available in a market, while still allowing her to purchase the quantity of the x's that she actually did buy at market prices p with income y. In conducting this thought experiment, one needs to supplement her
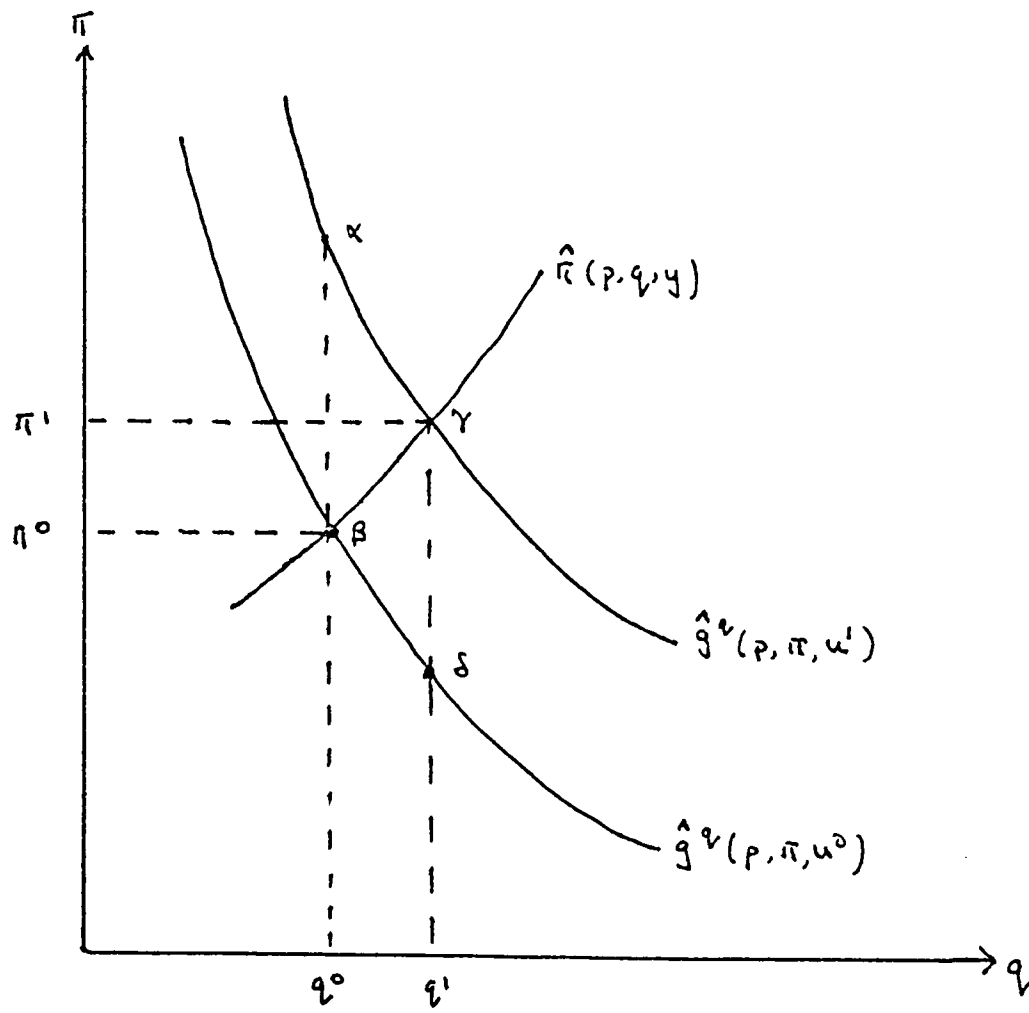
FIGURE 1. WTP and WTA for a Change in q.

income so that she can afford q as well as the x's. Thus, for given (p, q, y), we seek the price π that satisfies

$$(11) \qquad q = \hat{h}^q(p, \pi, y + \pi a).$$

The solution will be denoted by $\pi = \hat{\pi}(p, q, y)$. This inverse function is related to the inverse compensated demand function by the identities[11]

$$(12a) \qquad \hat{\pi}(p, q, y) \equiv \hat{\pi}[p, q, v(p, q, y)]$$

$$(12b) \qquad \hat{\pi}(p, q, u) \equiv \hat{\pi}[p, q, m(p, q, u)].$$

It follows from (12a) that $\pi^0 \equiv \hat{\pi}(p, q^0, u^0) = \hat{\pi}(p, q^0, y)$ and $\pi^1 \equiv \hat{\pi}(p, q^1, u^1) = \hat{\pi}(p, q^1, y)$--i.e., the graph of $\hat{\pi}(p, q, y)$ as a function of q intersects the graph of $\hat{\pi}(p, q, u^0)$ at $q = q^0$, and the graph of $\hat{\pi}(p, q, u^1)$ at $q = q^1$. This is depicted in Figure 1.[12]

Using the inverse demand function $\hat{\pi}(p, q, y)$, define the quantity

$$(13) \qquad A \equiv \int_{q^0}^{q^1} \hat{\pi}(p, q, y) \, dq$$

which corresponds to the area $q^0 \beta \gamma \delta q^1$ in Figure 1. This is a sort of Marshallian consumer's surplus, which is to be compared with C and E. Let

$$(14) \qquad \xi \equiv \frac{\partial \ln \hat{\pi}(p, q, y)}{\partial \ln y}$$

be the income elasticity of $\hat{\pi}(p, q, y)$; Randall and Stoll call this the "price flexibility of income." Assume that, over the range from (p, q^0, y) to (p, q^1, y), this elasticity is bounded from below by ξ^L and from above by

ξ^U with neither bound equal to 1. Using the mean-value theorem, as in Willig's equation (18), and the above equations (2'), (3'), (10), (12b), (13), and (14), yields Randall and Stoll's result--namely,

PROPOSITION 3: Assume $\xi^L \leq \xi \leq \xi^U$ where $\xi^L \neq 1$ and $\xi^U \neq 1$. Then,

(i) $0 \leq \left[1 + (1 - \xi^L) \frac{A}{y} \right]^{\frac{1}{1-\xi^L}} - 1 \leq \frac{E}{y}$

(ii) $0 \leq 1 - \left[1 - (1 - \xi^U) \frac{A}{y} \right]^{\frac{1}{1-\xi^U}} \leq \frac{C}{y} \leq 1$

(iii) If $\xi^U < 1$, or if $\xi^U > 1$ and $1 + (1 - \xi^U) \frac{A}{y} \geq 0$, $\frac{E}{y} \leq \left[1 + (1 - \xi^U) \frac{A}{y} \right]^{\frac{1}{1-\xi^U}} - 1$

(iv) If $\xi^L > 1$, or if $\xi^L < 1$ and $1 - (1 - \xi^L) \frac{A}{y} \geq 0$, $\frac{C}{y} \leq 1 - \left[1 - (1 - \xi^L) \frac{A}{y} \right]^{\frac{1}{1-\xi^L}}$.

Applying a Taylor approximation, as in Willig, and assuming that the conditions in (iii) and (iv) are satisfied, one obtains

(15) $\xi^L \frac{A^2}{y} \leq E - C \leq \xi^U \frac{A^2}{y}$.

This is commonly interpreted as implying that C and E are close in value, but whether or not that is correct clearly depends on the magnitudes of (A/y) and the bounds ξ^L and ξ^U. The magnitude of (A/y) depends in part on the size of the change from q^0 to q^1. But what can be said about the likely magnitude

of the income elasticity, ξ--could it happen, for example, that $\xi^L = \infty$? To answer that question, differentiate (11) implicity

$$(16) \qquad \frac{\partial \hat{\pi}(p, q, y)}{\partial y} = - \frac{\hat{h}^q_y(p, \pi, y + \pi q)}{\hat{h}^q_\pi(p, \pi, y + \pi q) + q\hat{h}^q_y(p, \pi, y + \pi q)}.$$

By the Hicks-Slutsky decomposition, the denominator is equal to the own-price derivative of the compensated demand function for q and is nonpositive

$$\hat{g}^q_\pi[p, \pi, v(p, q, y)] = \hat{h}^q_\pi(p, \pi, y + \pi q) + q \hat{h}^q_y(p, \pi, y + \pi q) \leq 0.$$

Converted to elasticity form, (16) becomes

$$(16') \qquad\qquad\qquad\qquad \xi = - \frac{\eta(1 - \alpha)}{\epsilon}$$

where $\eta \equiv (y + \pi q) \hat{h}^q_y(p, \pi, y + q\pi)/q$ is the income elasticity of the <u>direct</u> ordinary demand function for q, $\alpha \equiv q\pi/(y + q\pi)$ is the budget share of q in relation to "adjusted" income, and $\epsilon \equiv \pi\hat{g}^q_\pi[p, \pi, v(p, q, y)]/q$ is the own-price elasticity of the compensated demand function for q. The last term can be related to the overall elasticity of substitution between q and the private market goods x_1, ..., x_N. By adapting W. E. Diewert's (1974) analysis, it can be shown that, if the prices p_1, ..., p_N vary in strict proportion (i.e., $p_i = \theta\bar{p}_i$ for some fixed vector $\bar{p}$), the aggregate Allen-Uzawa elasticity of substitution between q and the Hicksian composite commodity $x_0 \equiv \Sigma\bar{p}_i x_i$, denoted σ_0, is related to the compensated own-price elasticity for q by the formula: $\epsilon = -\sigma_0(1 - \alpha)$. Hence, (16') may be written

(16'')
$$\xi = \frac{\eta}{\sigma_0}$$

where $\sigma_0 \geq 0$.

This provides an explanation of the results in the previous section. For changes in q, unlike changes in p, the extent of the difference between C and E depends not only on income effects (i.e., η) but also on substitution effects (i.e., σ_0). If, over the relevant range, either $\eta = 0$ (no income effects) or $\sigma_0 = \infty$ (perfect substitution between q and one or more of the x's), then $\xi^L = \xi^U = 0$ and, from Proposition 3, C = A = E. On the other hand, if the demand function for q is highly income elastic, or there are very few substitutes for q among the x's so that σ_0 is close to zero, this could generate very large values of ξ and substantial divergences between C and E. Suppose, for example, that, over the relevant range, a lower bound on the income elasticity of $\hat{\pi}(\cdot)$ is $\xi^L = 20$ (e.g., $\eta = 2$ and $\sigma_0 = 0.1$) and A/y = 0.05. Then, from Proposition 3 (i and iv), $C/y \leq 0.0345$ while $0.1708 \leq E/y$, so that E is at least five times larger than C.[13] Higher values of ξ^L would imply even greater differences between C and E.

III. Conclusion

A recent assessment of the state of the art of public good valuation concludes "Received theory establishes that . . . WTP . . . should approximately equal . . . WTA. . . . In contrast with theoretical axioms which predict small differences between WTP and WTA, results from contingent valuation method applications wherein such measures are derived almost always demonstrate large differences between average WTP and WTA. To date, researchers

have been unable to explain in any definitive way the persistently observed differences between WTP and WTA measures" (Cummings, Brookshire, and Schulze, p. 41). This paper offers an explanation by showing that the theoretical presumption of approximate equality between WTP and WTA is misconceived. This is because, for public goods, the relation between the two welfare measures depends on a substitution effect as well as an income effect. Given that the substitution elasticity appears in the denominator of (16") and the Engel aggregation condition places some limit on the plausible magnitude of the numerator, this suggests that the substitution effects are likely to exert far greater leverage, in practice, on the relation between WTP and WTA than the income effects. Thus, large empirical divergences between WTP and WTA may be indicative not of some failure in the survey methodology but of a general perception on the part of the individuals surveyed that the private market goods available in their choice set are, collectively, a rather imperfect substitute for the public good under consideration.

FOOTNOTES

[1]This view is expressed by, for example, Myrick Freeman (1979, p. 3);
Mark A. Thayer (1981, p. 30); Jack L. Knetsch and J. A. Sinden (1984, p. 508);
and Don L. Coursey, William D. Schulze, and John J. Hovis (1984, p. 2).

[2]I am treating q as a scalar here, but it could be a vector without
seriously affecting the analysis in this section. In the next section, how-
ever, the analysis would become significantly more complex if q were a vector
and more than one element of q changed.

[3]These alternative interpretations are offered, respectively, by Mäler,
W. Michael Hanemann (1982), and Randall and Stoll.

[4]These properties are established in my earlier paper.

[5]I have taken the liberty of defining C and E as the negative of quan-
tities appearing in Willig and in Randall and Stoll, so that sign (C) =
sign (E) = sign $(u^1 - u^0)$.

[6]I assume throughout that $q^1 > q^0$ and $u^1 \geq u^0$. The analysis could be
repeated for a case in which quality decreases and $u^1 \leq u^0$. In that case, C
and E are both nonpositive and correspond, respectively, to the compensation
that the individual would be willing to accept to consent to the change and
the amount that she would be willing to pay to avoid the change. This would
reverse the inequalities presented below, but it would not affect the sub-
stance of my argument.

[7]This result carries over, of course, if _more_ than one private good is a
perfect substitute for q. In the most general case, $u(x, q) = \bar{u}[x_1 + \psi_1(q),$
$..., x_N + \psi_N(q)]$ and $C = E = \Sigma p_i [\psi_i(q^1) - \psi_i(q^0)]$.

[8]Indeed, if $\bar{h}^i(\alpha_1 p_1, \ldots, \alpha_N p_N, y) \leq q^0$, $i = 1, \ldots, N$, it can be shown that $v(p, q^0, y) = v(p, q^1, y) = v(\bar{\alpha}_1 p_1, \ldots, \alpha_N p_N, y)$ and $C = E = 0$, where $\bar{h}^i(\cdot)$ and $\bar{v}(\cdot)$ are the ordinary demand functions and indirect utility function associated with $\bar{u}(\cdot)$.

[9]It is now necessary to assume that $u(\cdot)$ is strictly quasiconcave in both x and q.

[10]Using subscripts to denote derivatives, differentiate (9) and note that $q = \hat{g}^q(p, \pi, u) = \hat{m}_\pi(p, \pi, u)$ by Shephard's Lemma. Equations similar to (9) through (12) are presented by J. P. Neary and K. W. S. Roberts (1980).

[11]Note that $\hat{\pi}(p, q, y)$ is __not__ an inverse ordinary demand function in the sense of Ronald W. Anderson (1980) because it involves an income adjustment as well as a price effect.

[12]It is commonly supposed that $\pi^0 > \pi^1$ when $q^0 < q^1$--see, for example, Figure 7.12 in Richard E. Just, Darrell L. Hueth, and Andrew Schmitz (1982)--but this is not correct. It can be shown that $\pi^0 \gtreqless \pi^1$ according as $\eta \lesseqgtr (1/\alpha)$. Since $\Sigma \alpha_i \eta_i + \alpha \eta = 1$ by the Engel aggregation condition, where $\alpha_i \equiv p_i x_i / (y + \pi q)$ and $\eta_i \equiv (y + \pi q) h_y^i / x_i$, $\pi^0 \lessgtr \pi^1$ if and only if $\Sigma \alpha_i \eta_i \lessgtr 0$.

[13]This is actually the order of magnitude by which WTA measures exceed WTP measures in the empirical studies summarized in Table 3.2 of Ronald G. Cummings, David S. Brookshire, and William D. Schulze (forthcoming).

REFERENCES

Anderson, Ronald W., "Some Theory of Inverse Demand for Applied Demand Analysis," European Economic Review, 1980, 14, 281-90.

Bishop, Richard C. and Hebertein, Thomas A., "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased," American Journal of Agricultural Economics, December 1979, 61, 926-30.

Coursey, Don L., Schulze, William D. and Hovis, John J., "On the Supposed Disparity Between Willingness to Accept and Willingness to Pay Measures of Value: A Comment," mimeo., University of Wyoming, Department of Economics, Laramie, Wyoming, January 1984.

Cummings, Ronald G., Brookshire, David S. and Schulze, William D., Valuing Public Goods: An Assessment of the Contingent Valuation Method, Totowa, N. J.: Rowman and Allanheld, forthcoming.

Diewert, W. E., "A Note on Aggregation and Elasticities of Substitution," Canadian Journal of Economics, February 1974, 7, 12-20.

Freeman, A. Myrick, The Benefits of Environmental Improvement: Theory and Practice, Baltimore: Johns Hopkins University Press, 1979.

Gordon, Irene M. and Knetsch, Jack L., "Consumer's Surplus Measures and the Evaluation of Resources," Land Economics, February 1979, 55, 1-10.

Gorman, W. M., "Tricks With Utility Functions," in M. Artis and R. Nobay, eds., Essays in Economic Analysis, New York: Cambridge University Press, 1976.

Hanemann, W. Michael, "Quality and Demand Analysis," in Gordon C. Rausser, ed., New Directions in Econometric Modeling and Forecasting in U. S. Agriculture, Amsterdam: North Holland Publishing Co., 1982.

Just, Richard E., Hueth, Darrell L. and Schmitz, Andrew, Applied Welfare Economics and Public Policy, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1982.

Knetsch, Jack L. and Sinden, J. A., "Willingness to Pay and Compensation Demanded: Experimental Disparity in Measures of Value," Quarterly Journal of Economics, August 1984, 507-21.

Mäler, Karl-Göran, Environmental Economics: A Theoretical Inquiry, Baltimore: Johns Hopkins University Press, 1974.

Neary, J. P. and Roberts, K. W. S., "The Theory of Household Behavior Under Rationing," European Economic Review, 1980, 13, 25-42.

Randall, Alan and Stoll, John R., "Consumer's Surplus in Commodity Space," American Economic Review, June 1980, 71, 449-57.

Thayer, Mark A., "Contingent Valuation Techniques for Assessing Environmental Impacts: Further Evidence," Journal of Environmental Economics and Management, 1981, 8, 27-44.

Willig, Robert, "Consumer's Surplus Without Apology," American Economic Review, September 1976, 66, 589-97.

CHAPTER 7

METHODS OF BENEFIT MEASUREMENT

In the two preceding chapters, we have spoken of benefits in a rather general sense not specifying where they come from or how they might be measured in practice. In Chapter 5, for example, we assume the existence of a benefit function for ecosystem recovery and examine how a decision on pollution control is affected by the dynamics of recovery and the uncertainties surrounding it. In this chapter we look behind the benefit function. What kinds of benefits are provided by aquatic ecosystems, and how might they be measured? Here we take up the discussion begun in Chapter 1 drawing upon the classification of benefits and measurement approaches suggested there.

I. Aquatic Ecosystems as an Input to Production

Aquatic ecosystems function as an input to production whenever changes in an ecosystem's characteristics affect the costs of providing a good or service. For example, the number of wetland acres available as a habitat for fish may influence the cost of harvesting commercially valuable species. The quality of water withdrawn from rivers and lakes for municipal water supplies and irrigation determines the cost of subsequent water treatment and level of agricultural productivity. Finally, just as air pollution may lead to the chemical deterioration of materials, diminished water quality can lead to the corrosion of household appliances and industrial equipment. Valuing the benefits from improved environmental quality when the environment acts as an input to production is the focus of this section. We critically review a number of earlier studies in the area and go on to suggest (and illustrate) some improvements.

We focus on the examples identified in Chapter 1: supply of clean water and harvest of commercial species. Consider the former. Wetlands reduce the cost of water treatment by removing or settling pollutants. This can be represented as a shift in a marginal cost or supply curve along a given demand curve. An environmental improvement, such as provision of additional wetlands, would then involve a supply shift down and to the right, as from S to S' in Figure 1, where the shaded area between old (S) and new (S') supply curves indicates the net welfare gain, the change in consumer and producer surplus.

This is probably a typical case, but others are possible--and, it turns out, relevant to some of the existing literature. One, in particular, is worth noting. Suppose the new cost or supply curve is simply the horizontal axis. In other words, creation of the wetlands completely eliminates the need for human inputs, at least up to a point (represented by Q" on Figure 2). Then the welfare gain, illustrated in the figure, is the shaded area between old and new supply curves up to the point (Q' on the figure) where demand equals the old supply and between demand and new supply thereafter (up to Q"). Note that this is less than the area between the two supply curves. Beyond Q', consumer willingness-to-pay for water is less than the old cost of treatment so that the latter is no longer relevant.

This same point is made more dramatically in Figure 3. There the old cost of treatment or supply curve lies everywhere above the demand curve. The benefit of the environmental improvement, represented as a shift in the supply curve to coincide with the horizontal axis, is then simply the area under the demand curve (up to Q"). The area between the two supply curves, which is just the area under the old curve, or the cost of providing treatment in the absence of the wetlands, would overstate the benefit of having the wetlands for this purpose.

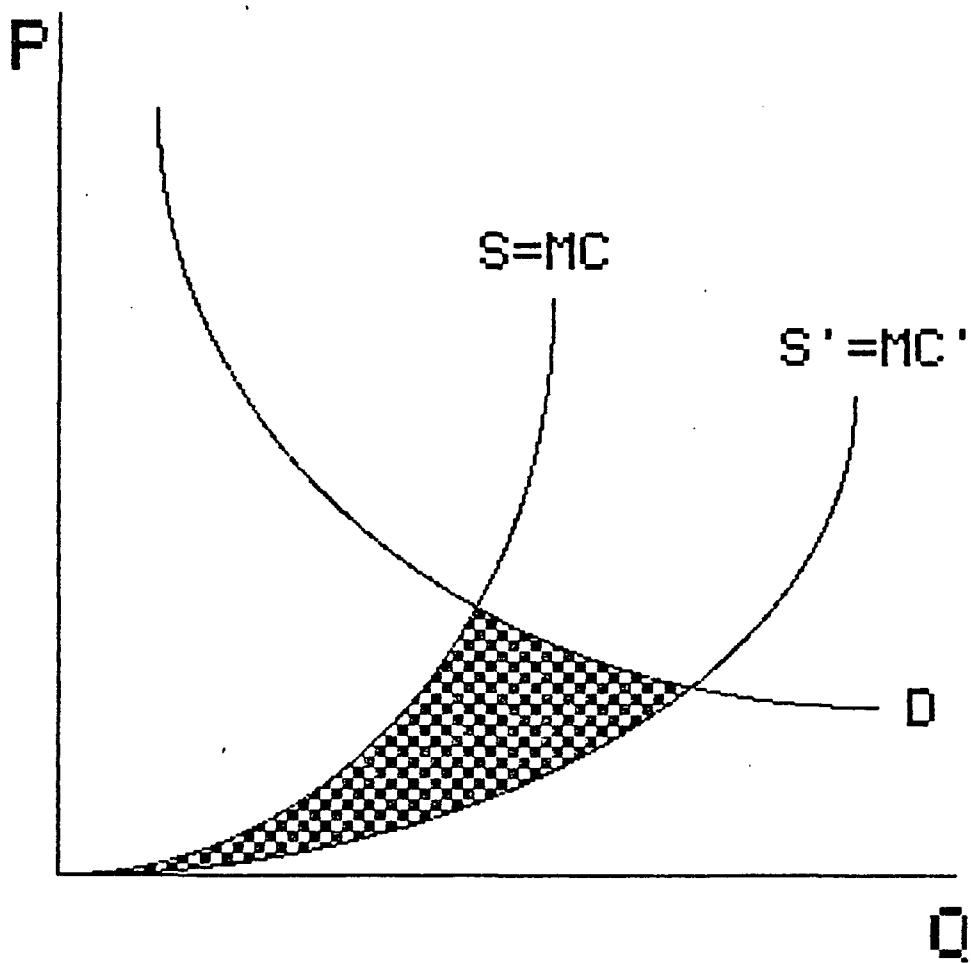# FIGURE 1

WASTE ASSIMILATION
BENEFIT PROVIDED BY
THE ECOSYSTEM

FIGURE 2

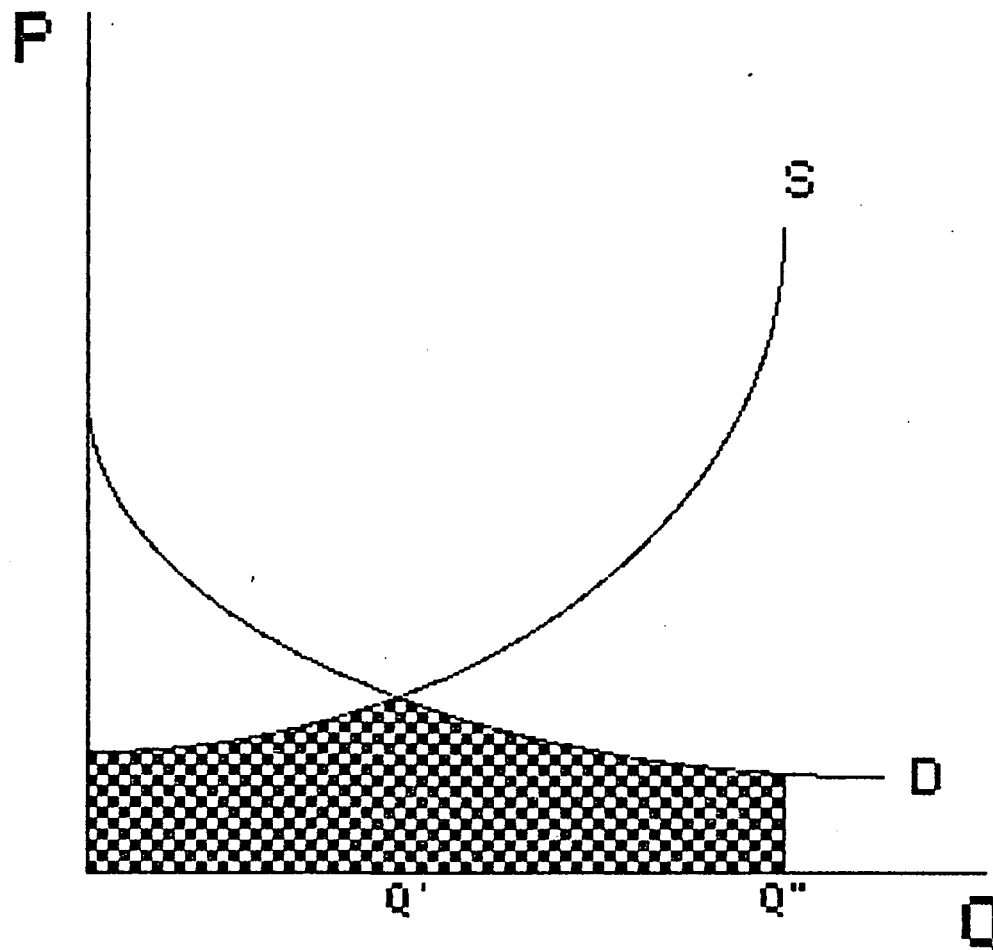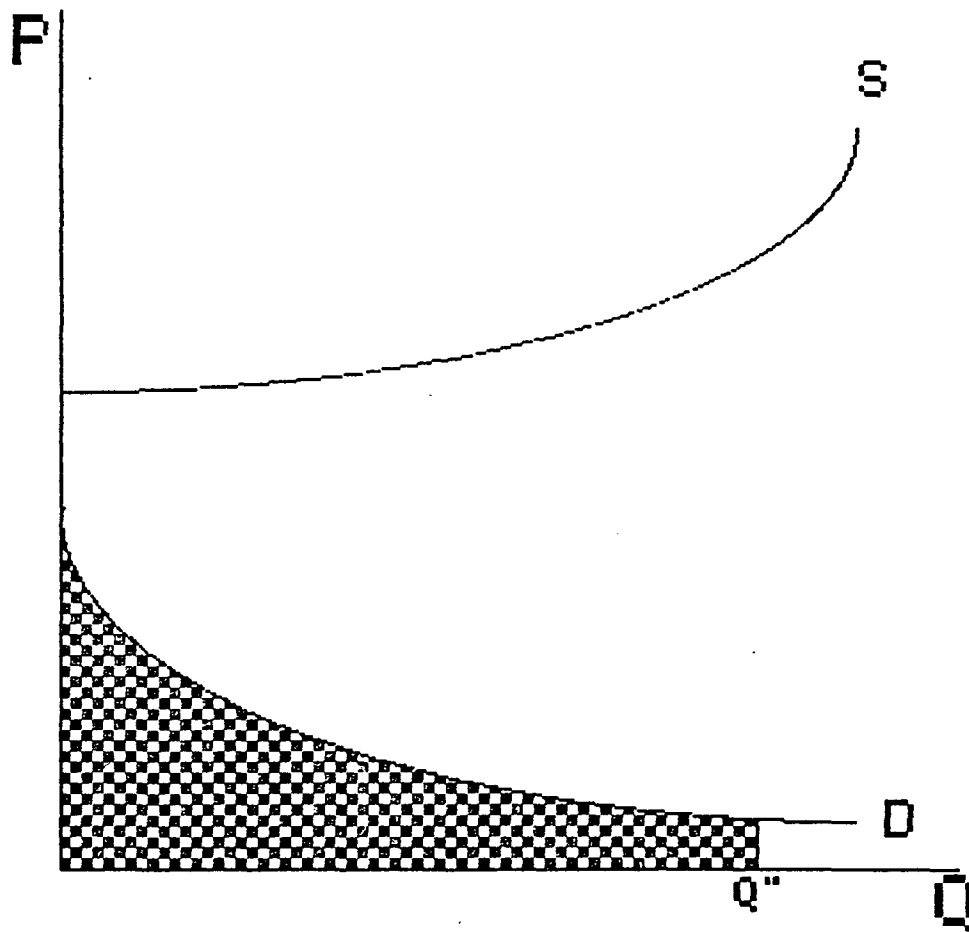## WASTE ASSIMILATION BENEFIT PROVIDED BY THE ECOSYSTEM

FIGURE 3

WASTE ASSIMILATION
BENEFIT PROVIDED BY
THE ECOSYSTEM

This is essentially the difficulty with the pioneering and influential study of the value of estuarine wetlands by Gosselink, Odum, and Pope (GOP, 1974). They claim that an acre of estuarine wetland provides benefits which would cost $2,500 per year if produced by man-made treatment plants. Shabman and Batie (1977) are justifiably critical of this figure:

> ". . . the use of alternative estimates should be governed by three considerations: (1) the alternative considered should provide the same services; (2) the alternative selected for the cost comparison should be the least-cost alternative; and (3) there should be substantial evidence that the service would be demanded by society if it were provided by the least-cost alternative. GOP failed to subject their estimate to any of these important tests."

Park and Batie (1979) contend that GOP not only failed to test whether the least-cost alternative would be demanded, but that their identification of waste treatment plants as the least expensive alternative may be incorrect. They argue that recent evidence suggests that adjustments in agricultural practices (e.g., restriction on the application of fertilizers which "run off" into estuarine waters) may be a less costly alternative to the construction of treatment plants. The criticism of the work of GOP is not to suggest that waste assimilation is not an important service provided by wetlands; however, care must be taken when determining just how society values that service.

Problems have also plagued efforts to value benefits which might be provided by aquatic ecosystems sometime in the future but which are not currently provided. Instead of valuing the option to use a resource as an input to production in the future in the way suggested in Chapter 5, some studies have calculated benefits as if the resource were already being used. What is missing here is an estimation of the likelihood that the resource will ever be

used and the timing of its use. Gupta and Foster (GF, 1976) attempt to value wetlands as a potential source of water supply for the state of Massachusetts and find that the state's wetlands could provide an annual benefit of $2,800 per acre. Unfortunately, GF's estimated benefit of wetlands' preservation in this regard is calculated as though the cost savings of using wetlands instead of current sources were already realized. Their finding, that wetlands would provide a cheaper supply of water for Massachusetts, can be questioned in two respects. First, if wetlands are a cheaper alternative to current sources, why are they not used? Second, if it is the existence of institutional barriers which block their use, why won't those barriers continue to preclude the tapping of wetlands as a supply of water in the future? Although it is certainly true that the preservation of wetlands may be valuable because the option to use them as a water source would be retained, this is not the benefit GF estimate. As a final point, their estimate of the total value of undeveloped wetlands may be plagued by double counting problems. If water were taken from Massachusetts' wetlands, would the same wetlands continue to generate the recreational and amenity benefits they add to the water supply benefits?

We now turn to the commercial harvest example. A substantial amount of previous empirical work has sought to value the environment as input for this purpose in ways not fully consistent with the deceptively simple approach discussed thus far and summarized in Figure 1. The estimated benefits variously fail to analyze changes in the relevant cost structure, ignore price effects of a change in production, and rely on ad hoc measures like total or net revenue. As a measure of change in social welfare, revenue figures exhibit at

least two problems. First, they do not reflect the opportunity cost of producing goods and services. Second, demand for many fish and shellfish species is relatively price inelastic (Bell, 1970), so an increase in production due to an environmental improvement results in a decrease in total revenue, incorrectly implying that the improvement does not lead to a welfare gain. About the best that can be said for the revenue calculations (with or without price effects) is that they are not relevant to the determination of a change in combined consumer and producer surplus--our preferred welfare measure.

A Council on Environmental Quality (CEQ, 1970) study illustrates the same difficulties in a somewhat different way. The study reports that, due to the practice of ocean dumping, one-fifth of the nation's shellfish beds are contaminated and closed. Assuming the closed shellfish beds would be as productive as their open counterparts, the study concludes that an improvement in water quality would result in a 25 percent increase in quantity produced and a subsequent 25 percent increase in total revenues. The increase in total revenues are claimed as the gain to society of cleaning up the shellfish beds. However, as long as demand is not perfectly elastic, an additional 25 percent in the amount of shellfish supplied to the market could only be sold if the price of shellfish fell. The estimate of CEQ of an additional $63 million in shellfish revenues (the additional 25 percent) is clearly an overstatement. But in any case the revenue figures do not reflect costs or willingness to pay for nonmarginal units and, hence, are not adequate measures of welfare.

An important question to address, in valuing commercial fishing benefits, is this: What is the contribution of the ecosystem to the production process? It is a question some studies have failed to address. Thus, GOP (1974),

in assessing the value of wetlands as a fish nursery, divide annual dockside values of fish products landed by the total number of wetland acres to arrive at a value per acre in production of fish. Imputing all of the revenue from commercial fishing to wetland acreage, however, ignores the contribution of other fishing inputs like labor and capital.

The more recent study by Lynne, Conroy, and Prochaska (LCP, 1981) suggests that it may be possible to isolate the contribution of environmental inputs to production. They develop a bioeconomic model in which human effort and marshland are distinct inputs in the production of blue crab off Florida's Gulf Coast. The population of blue crabs is assumed to be a function of the quantity of local marshland acres. Since the successful harvesting of the crabs is modeled to be dependent on their population level, marshlands, which act to define the carrying capacity for blue crabs, appear as an input in the production function. The reduced form production function is estimated according to the ordinary least-squares criterion; and, using the appropriate estimated coefficients, a marginal product for an acre of wetlands is calculated. Finally, the value of the marginal product for an acre is computed using current dockside prices. The study is laudable for valuing both marshland acreage and human input in the production of blue crabs. However, the authors' contention that the value of the marginal product is the relevant measure of benefits provided by wetlands is incorrect. Let us take up the analysis at this point and develop an example in which notions of consumer and producer surplus are correctly employed, as in Figure 1, to evaluate the commercial fishing benefits produced by the marshland.

In keeping with the spirit of LCP, consider the optimization problem faced by a price-taking firm or industry where price is P and the unit cost of the human effort input is W:

(1) $\max_{X_1}$ P F$(X_1, \bar{X}_2)$ - W X$_1$.

The production process is posited to be a function, F$(\cdot)$, of two inputs: one (X_1) which captures the efforts of man to harvest shellfish and another (X_2) which represents the contribution of an ecosystem variable like marsh-land acreage. The bar over X$_2$ indicates that, for the time being, the acreage is fixed. Although we, like LCP, model human effort as a single input, the number of traps set, one many prefer to explicitly model the use of several inputs so that substitution among them can be studied.

We assume that the production of blue crabs can be represented as a Cobb-Douglas process. Although the Cobb-Douglas form is no doubt a simplification of the true production process (and is probably a poor approximation to reality for extreme values of either input), we use it here because our main purpose is to demonstrate the procedure for calculating changes in combined consumer and producer surplus. Therefore, substituting for the production function in equation (1) the Cobb-Douglas form and noting that cost minimization is the dual problem to profit maximization, the optimization problem can be rewritten as

(2) $\min_{X_1, \lambda} \mathcal{L}$ = W X$_1$ + $\lambda(Q - A X_1^a \bar{X}_2^b)$

where λ is the Lagrange multiplier; Q is output; and A, a, and b are parameters. Differentiating the Lagrangian with respect to the effort variable and the Lagrange multiplier yields

$$(3) \qquad \frac{\partial \mathcal{L}}{\partial X_1} = W - \lambda A \bar{X}_2^b a X_1^{a-1} = 0$$

$$(4) \qquad \frac{\partial \mathcal{L}}{\partial \lambda} = Q - A X_1^a \bar{X}_2^b = 0.$$

Since the production function is characterized by only one decision variable, X_1, equation (4) is the only one needed to solve for the cost function, $C(\cdot)$.

$$(5) \qquad X_1 = \left[\frac{Q}{A \bar{X}_2^b} \right]^{1/a}$$

$$(6) \qquad C(W, Q, \bar{X}_2) = W A^{-1/a} \bar{X}_2^{-b/a} Q^{1/a}.$$

Differentiating the cost function with respect to output generates the marginal cost expression

$$(7) \qquad MC = \frac{\partial C}{\partial Q} = \frac{W}{a} A^{-1/a} \bar{X}_2^{-b/a} Q^{(1-a)/a}.$$

The blue crab industry also presumably faces a demand curve for its product. A simple constant elasticity demand function is given in (8), and the corresponding inverse demand function in (9):

(8) $Q = KP^{-m}$

(9) $P = K^{1/m} Q^{-1/m}$

where K is a parameter and m is the (constant) elasticity. The profit-maximizing firms will equate price and marginal cost so that the equilibrium level of blue crabs sold is given by

(10) $Q = \left[\frac{a}{w} K^{1/m} A^{1/a} \bar{X}_2^{b/a} \right]^{ma/[m+(1-m)a]}$.

The result in (10) holds for all relevant values of marsh acreage, $\bar{X}_2$, available for the biological promotion of the blue crab population. Therefore, we first calculate the equilibrium output associated with various levels of wetland acreage, then we compute the equilibrium price corresponding to the output by use of equation (9).

We proceed to calibrate the parameters of the model in order to construct an example which is reasonably compatible with the price, input, and output data used by LCP. We also incorporate their econometric finding that the marginal product of an acre of marsh is roughly 2-1/2 pounds of blue crab (annually). Although the demand for shellfish has been found to be relatively price-inelastic, as we noted earlier, we assume in this case a high elasticity since the Gulf Coast fishery is presumably not the sole source of blue crab in the market. Welfare gains associated with an increase in marshland habitat (remember we are considering only gains in the blue crab industry for purposes of this example) are calculated as the change in consumer and producer surplus. These measures are presented in Table 1. For example, for a demand elasticity of -2.05, the net gain associated with an increase from 25,000

TABLE 1

Welfare Gain Associated with an Increase in Wetland Acreage
(From an Initial Base of 25,000 Acres)

Elasticity (m)	Wetland acreage ($\overline{X}_2$)	Number of traps (X_1)	Change in combined surplus
2.05	100.000	33.610	191,389
2.05	200,000	33,332	294,290
2.05	300,000	33,170	356,843
2.05	400,000	33,056	402,316
2.05	500,000	33,000	435,829

acres to 100,000 is $191,389. Successive increments in acreage add less to estimated benefits due to diminishing returns to the marshland input.

The results of a sensitivity analysis, in which different price elasticities of demand [ranging from (-.25) to (-3.45)] are used to calibrate the model, indicate that, in this particular model, the estimates of welfare gain are reasonably robust to the choice of an assumed price elasticity.

The purpose of this exercise has been to demonstrate that a theoretically correct measure of welfare can be constructed and calculated on the basis of empirical information about the impact on product supply (given demand) of a change in ecosystem characteristics (here the number of wetland acres) which, in turn, might be related to pollution control.

Of course, this has been a hypothetical exercise; and, in an actual case study, one would econometrically estimate the demand and production functions necessary to conduct the welfare analysis. Moreover, if the estimated demand function includes an income variable, simple Marshallian consumer surplus is no longer the appropriate welfare measure. Fortunately, for a variety of functional forms for the demand function, exact surplus measures are known and available.

A still more recent study, by Kahn and Kemp (KK, 1985), appears to follow the procedure we have outlined, though they use it to calculate a welfare loss. Specifically, they are concerned with the effect the decline in submerged aquatic vegetation (SAV) is having on the various fisheries supported by Chesapeake Bay. SAV serves as an important link in the estuarine food chain, and KK attempt to quantify the welfare loss primarily to the striped bass commercial fishery and, also, to other commercial and sport fisheries stemming from the reduction in SAV caused by agricultural runoff, discharges

from sewage treatment plants and soil erosion, and the consequent reduction in the carrying capacity of the Bay. Unlike LCP, KK are fortunate to have population data on the striped bass. With this, they can estimate a supply function which includes a population variable for the fish and an equation which relates SAV to fish. After estimating a demand function for striped bass, KK calculate the losses in consumer and producer surplus following incremental reductions in SAV. One criticism that can be made of their procedure is that, since demand is estimated as a function of per capita income, a more exact welfare measure than Marshallian consumer surplus could have been calculated. Just for purposes of comparison with the welfare gains that we calculated for the Florida Gulf Coast blue crab fishery, we observe that a 50 percent reduction in SAV is associated with an annual loss of approximately $4 million. This is substantially larger than the numbers in our example. It is important to note that KK are casting a wider net, so to speak: both commercial and sport fishing, for several species, are considered.

The studies just described are limited by their static nature. Both examine the contribution of an environmental input to production assuming the fishery is in bioeconomic equilibrium (i.e., the harvest rate of the marketed species equals its growth rate). To the extent that their data are comprised of observations for years in which the fisheries were not in a steady state, the regression coefficients they obtain will be biased as parameters of steady-state models. In addition, static approaches to fisheries economics fail to evaluate the stream of benefits generated by fisheries as they move from one equilibrium to the next. As demonstrated in Chapter 3, the higher

trophic levels of damaged ecosystems may respond slowly to pollution control measures, and attempts to value control need to take this into account.

The need for dynamic analysis arises from the recognition that fishery resources constitute capital assets which yield a stream of benefits over time, and it is in this framework that we can view proposed environmental cleanup policies as potential investments. Although much of the literature now recognizes the dynamic nature of fishery resources, with a few articles even explicitly recognizing the dynamic links between predator and prey species (see Clark, 1976, and Ragozin and Brown, 1985), the literature has not considered the management of fisheries' environmental problems in a dynamic context.

A framework for finding an optimal management strategy when a fishery is confronted with pollution and open-access problems might look something like the following. The management problem is one of simultaneously determining harvesting and pollution control policies to maximize the present discounted value of net benefits generated by the fishery. In the most general notation, i.e., making no assumptions about the forms of economic or biologic functions, the management problem is

$$(11) \quad \max_{E(t),Z(t)} \sum_{t=0}^{\infty} (1 + r)^{-t} NB[E(t), Z(t), X(t)]$$

subject to

$$(12) \quad X(t+1) - X(t) = f\{E(t), Q[Z(t)], X(t)\}$$

and

$$(13) \quad X(0) = X_0$$

-16-

where r is a discount rate, NB(·) is a net benefit function (e.g., combined consumer and producer surplus), E is fishing effort, Z is pollution control, X is the stock of the harvested species, and Q is the level of environmental quality. Further realism may be given to the model by including additional equations of motion [like equation (12)] which represent the growth rates of other species in the ecosystem and establish links between distinct levels of the food chain. Modeling species interaction may be of particular importance if pollution directly affects growth rates at the lower trophic levels, as demonstrated in Chapter 3. However, the introduction of biological inter-action among species also poses the problem of selecting an appropriate model from the available alternatives (see May, 1973, for a description of the various ways in which species interaction may be modeled). Interactions can be complex and models like the Lotka-Volterra used in Chapters 3 and 4 and also in the studies reviewed in this section which imply simple feeding hierarchies rather than complex food webs may be misleading (see Harte, 1985).

A key feature of the solution of the optimization problem stated in equations (11) through (13) may be the interdependence of the two control variables, allowable fishing effort, and pollution control. For example, if the level of the fish stock is below the optimum, the derived solution to the management problem may include the enactment of stringent pollution controls to enable the fish population to recover. The solution may also include concurrent restrictions on fishing effort (possibly even prohibition) so that the eventual benefits of costly pollution control may be realized.

The fisheries management problem is further complicated by the fact that decisions must be made in the face of uncertainty. As discussed in Chapter 4,

uncertainty pervades the modeling of species interaction; and this is compounded by uncertainty about ecosystem responsiveness to pollution control. When uncertainty about the values of economic variables is introduced, the optimization problem becomes a very difficult stochastic control problem indeed. If it is the case that uncertainty about the parameters of the model can be reduced by research or the acquisition of information through experience, management strategies should ideally be evaluated with the aid of closed-loop models in which policy decisions are subject to revison as new information becomes available, as discussed in Chapter 5 (see also Rausser, 1978).

II.   Aquatic Ecosystems as a Final Good

When an aquatic ecosystem is conceived of as a final good the benefits of enhancing the ecosystem typically take the form of improved opportunities for water-related recreation. These benefits can be estimated using the methodologies discussed in Chapter 6—either contingent valuation/behavior experiments or the revealed preference approach based on fitting demand functions for visiting alternative recreation sites (also called the "travel-cost" approach). Some of the methodological issues involved in contingent valuation experiments are discussed in Cummings, Brookshire and Schulze (1986), Hanemann (1985), and Carson and Mitchell (forthcoming). Issues involved in the travel-cost approach are discussed in Bockstael, Hanemann and Strand (1984) and Smith and Desvousges (1986).

The main challenge confronting practitioners of travel-cost studies is the need to handle the allocation of water-based recreation activities among multiple sites differing in their environmental quality attributes in a manner consistent with the utility maximization hypothesis. Two particular aspects

stand out--the selection of appropriate functional forms for the ordinary demand functions, and the need to deal with corner solutions. Taking the question of functional forms first, the problem is to select a set of functions for the ordinary demands, $x_i = h^i (p,q,y)$, $i = 1,..,N$, defined at the beginning of Chapter 6. In this context x_i is the number of visits to recreation site i by a household over some period of time (e.g., the fishing season), $p = (p_1,...,p_N)$ where p_i is some measure of the cost of visiting the ith site, $q = (q_1,..,q_N)$ where q is some vector of attributes of the ith site (including water quality, etc.) and y is either the household's total income or its total expenditure on recreation activities. The problem is that, if these demand functions are to be consistent with some utility maximization hypothesis, they must satisfy certain economic integrability conditions, including (i) the adding up condition and (ii) the symmetry and (iii) negative semidefiniteness of the matrix of Slutsky terms, $S = \{ s_{ij} \}$, where

$$ s_{ij} = \frac{\partial h^i (p,q,y)}{\partial p_j} + h^j (p,q,y) \frac{\partial h^i (p,q,y)}{\partial y} . \tag{14}$$

These requirements are by no means trivial and impose significant restrictions on the eligible functional forms. For example a demand system of the form

$$ \ln x_i = \alpha_i - \beta_i p_i + \gamma_i y \qquad i = 1,..,N \tag{15a}$$

where
$$ \alpha_i = \alpha_0 + \sum_k \delta_{1k} q_{ik} \tag{15b}$$

$$ \beta_i = \beta_0 + \sum \delta_{2k} q_{ik} \tag{15c}$$

$$ \gamma_i = \gamma_0 + \sum \delta_{3k} q_{ik} \tag{15d}$$

which is employed in Smith and Desvousges (1986), would appear to violate the symmetry of the s_{ij} terms. Other generalizations of the semi-log form to systems of multiple demand equations are examined by Hanemann and Lafrance (1983), where it is shown that the symmetry conditions place very stringent

(and empirically implausible) restrictions on the underlying direct utility
function. This does not mean that there are no suitable functional forms:
systems such as the Linear Expenditure System--Binkley and Hanemann (1975)--
and other members of the Generalized Gorman Polar Form family of indirect
utility functions

$$v(p, q, y) = F\left(\frac{y}{b(p, q)}\right) + a(p, q)$$ (16)

can certainly be employed.

The second issue--the phenomenon of <u>corner</u> solutions--is more trouble-
some. This refers to a situation where some of the x_i's are zero--a household
visits some of the available sites, but <u>not</u> all of them. The conventional theory
of consumer behavior is developed under the assumption of an <u>interior</u> solution
to the utility maximization problem (1) in Chapter 6--i.e., a solution where
all the x_i's are positive. Modifying this theory to deal with non-consumption
of certain goods (non-visitation of certain sites)--a phenomenon that is
overwhelmingly apparent in micro-data sets--is a rather complex task. The
problems involved, and some possible solutions, are examined in Chapters
8-10 of Bockstael, Hanemann and Strand.

A common approach to modelling corner solutions is to decompose consumer
choices into two elements: the selection of a <u>total</u> level of recreation activity,
$\bar{x} = \sum x_i$, and then the allocation of this total among the alternative possible
sites based on some type of shares model

$$x_i = \Pi_i (p, q, y) \, \bar{x} \qquad i = 1, \ldots, N$$ (17)

where Π_i, the share of total visits assigned to the ith site, satisfies

$$\Pi_i \geq 0, \qquad \sum \Pi_i = 1.$$ (18)

Statistical models such as logit and probit can be used to estimate the share
equations, and these models can be related to a utility maximization hypothesis.
But, at the present time, it is often difficult to obtain a utility-theoretic

justification for the "macro visitation equation" determining $\bar{x}$, and to integrate it with the share equations in a theoretically consistent manner. That is to say, one would like the determination of $\bar{x}$ and $\bar{\pi}_1, \ldots, \bar{\pi}_N$ to originate in a single, simultaneous utility maximization procedure. Some models which permit this have recently been developed, but they are relatively difficult to estimate. The resolution of these issues represents one of the frontiers of research for the travel cost approach.

BIBLIOGRAPHY

Bell, F. W. The Future of the World's Fishery Resources: Forecasts of De-
    mand, Supply and Prices to the Year 2000 with a Discussion of Implications
    for Public Policy, U. S. National Marine Fisheries Service, No. 65-1, 1970.

Clark, C. W. Mathematical Bioeconomics. New York: John Wiley and Sons, Inc.,
    1976.

Council on Environmental Quality. Ocean Dumping: A National Policy. Washing-
    ton, D. C.: Government Printing Office, 1970.

Gosselink, J. G., Odum, E. P., and Pope, R. M. The Value of the Tidal Marsh,
    Center for Wetland Resources, Louisiana State University, No. LSU-SG-74-03,
    1974.

Gupta, T. R., and Foster, J. H. "Economics of Freshwater Wetland Preservation
    in Massachusetts," in J. S. Larson, ed., Models for Assessment of Fresh-
    water Wetlands, Water Resources Center, University of Massachusetts,
    No. 32, 1976.

Harte, J. Consider a Spherical Cow. Los Altos: William Kaufmann, Inc., 1985.

Kahn, J. R., and Kemp, W. M. "Economic Losses Associated with the Degradation
    of an Ecosystem: The Case of Submerged Aquatic Vegetation in Chesapeake
    Bay," Journal of Environmental Economics and Management, Vol. 12 (1985),
    pp. 246-63.

Lynne, G. D., Conroy, P., and Prochaska, F. J. "Economic Valuation of Marsh
    Areas for Marine Production Processes," Journal of Environmental Economics
    and Management, Vol. 8 (1981), pp. 175-86.

May, R. M. Stability and Complexity in Model Ecosystems. Princeton, New
    Jersey: Princeton University Press, 1973.

Park, W. M., and Batie, S. S. "Methodological Issues Associated with Estimation of Economic Value of Coastal Wetlands in Improving Water Quality," Virginia Polytechnic Institute and State University, No. VPI-SG-79-09, 1979.

Ragozin, D. L., and Brown, G., Jr. "Harvest Policies and Nonmarket Valuation in a Predator-Prey System," Journal of Environmental Economics and Management, Vol. 12 (1985), pp. 155-68.

Rausser, G. C. "Active Learning, Control Theory, and Agricultural Policy," American Journal of Agricultural Economics, Vol. 60 (1978), pp. 476-90.

Shabman, L. A., and Batie, S. S. "Estimating the Economic Value of Natural Coastal Wetlands: A Cautionary Note," Coastal Zone Management, Vol. 4 (1977), pp. 231-47.

Binkley, Clark S. and Hanemann, W. Michael, (1975) The Recreation Benefits of Water Quality Improvement. Cambridge, Mass.:Urban Systems Research & Engineering, Inc.

Bockstael, Nancy E., Hanemann, W. Michael and Strand, Jr., Ivar E., (1984) Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models. University of Maryland, Department of Agricultural and Resource Economics, College Park, MD.

Carson, Richard T. and Mitchell, Robert C., Using Surveys To Value Public Goods: The Contingent Valuation Method, Washington, D.C.: Resources for the Future, Inc. (forthcoming).

Cummings, R.G., Brookshire, D.S. and Schulze, W.D., (1986) Valuing Environmental Goods: An Assessment of the Contingent Valuation Method Tolowa, NJ: Rowman and Allanheld.

Hanemann, W. Michael (1985), "Some Issues in Discrete - and Continuous - Response Contingent Valuation Studies," Northeastern Journal of Agricultural Economics , April 1985.

_____ and Lafrance, Jeffery T., (1983) "On the Integration of Some Common Demand Systems," Staff Paper No. 83-10, Dept. of Agricultural Economics and Economics, Montana State University, Bozeman.

Smith, V. Kerry and Desvousges, William L. (1986), <u>Measuring Water Quality Benefits</u>, Boston: Kluwer-Nijhoff Publishing Co.)

Chapter 8.

Further Work

Our present intention is to proceed in two areas: (1) comparative analysis of models for policy evaluation; and (2) development of a case study.

The first task, the comparative analysis, is intended to further integrate the ecologic and economic models developed in earlier chapters, and to compare the results obtained with those of variant versions of the models. Both aspects of this task are important. The first involves a tighter linking (than any in the present report) of a model of ecosystem recovery with a model of dynamic optimization under uncertainty. The idea is to develop the capability to evaluate control policies leading to ecosystem recovery, taking account of the (probabilistic) state of the system over time and at any point in time.

The second aspect of this task, comparative analysis of different models, is dictated by our lack of knowledge about population dynamics in a recovering aquatic ecosystem. In chapters 3 and 4 these dynamics were described by perhaps the simplest model for the purpose, the Lotka-Volterra. This was sufficient to obtain interesting results about qualitative features of recovery dynamics and the propagation of uncertainty. But as we move toward application (as in the case study described below) it becomes important to determine whether the results are robust, i.e., whether they continue to hold for equally plausible, though more complex, specifications of ecosystem population dynamics. Further, we need to explore the notion of robustness itself. Two models may yield seemingly quite different predictions about the nature and timing of recovery, yet imply the same

ranking of _policy_ alternatives. For example, one model may predict recovery of a fish population to 50% of its pre-pollution level (ignoring uncertainty) within five years of the imposition of some control measure, whereas another may predict recovery to just 10%. But the net present value of control may be positive in both cases. In any event, considerable further work is needed, in our judgment, on model development, integration, and comparative analysis, before we are ready to tackle a case study.

Turning now to the case study, we wish to pose a basic question: What do we want to get out of a case study? Two things, it seems to us. First, of course, we want quantitative results. What are the benefits of a particular control option? Second, however, we want to know what the results depend on. Partly, this is traditional sensitivity analysis. How are results affected by changes in assumptions about the discount rate, about a parameter describing interaction between the first and second trophic levels, and so on. But more importantly, we want to try to establish links between results and the types of _models_ used to generate them. This task clearly links back to our proposed work in the first area, comparative analysis of models for policy evaluation. The difference is that now we are proposing to go through the exercise in a real case, with real numbers.

With these objectives in mind, we wish to propose a "double-barreled" study. First, we would look at a relatively simple lake ecosystem, and one for which there also exists fairly good data on pollution control and subsequent recovery. A leading candidate here is Lake Washington, in the state of Washington. The idea would be to "field-test" our modeling approach in a relatively favorable setting.

Second, we would like to tackle San Francisco Bay. The Bay is of

course a much larger and more complex aquatic ecosystem, a marine estuary with substantial wetlands. Further, existing data are less reliable than for Lake Washington. Yet even with these difficulties, we feel the Bay is an appropriate subject for study by this project, for several reasons. First, it is economically important, a major influence on the natural resource base (including climate) of a metropolitan area of more than five million people. Second, the Bay is the subject of considerable current research and policy interest, at both the state and national levels. Third, a related point, the Bay ecosystem includes the major remaining wetlands in Northern California, and wetlands are themselves the subject of much current interest. Fourth, a study of San Francisco Bay would nicely complement existing work on the major east coast marine estuarine system, the Chesapeake Bay. Fifth, clearly travel costs would be minimized by choice of the Bay. Sixth, and finally, despite, or perhaps because of, the difficulties, we regard the proposed study as an exciting challenge.

We should note that, again because of the magnitude of the task and the potential difficulties, we do not propose to complete a study of the Bay within 12 to 18 months following submission of the final report on the current study. But we certainly would anticipate completion of parts of the task, which might stand on their own as interesting and useful research results.