"CONCEPTUAL AND METHODOLOGICAL CHALLENGES IN ENVIRONMENTAL AND NATURAL RESOURCES RESEARCH"

Gilberto C. Gallopin
9 June 1995

Invited Conference at the International Symposium on Statistics in Agriculture and Environmental Research -Satellite Conference"; CIAT/COLCIENCIAS; CIAT, Palmira, Colombia, June 7-9, 1995

INTRODUCTION

Rather than trying to cover this subject systematically, which would be obviously impossible in the time available, I will discuss some of the general properties of environmental and natural resources system that are reelevant to the discussions of this Symposium, and then I will illustrate a few of the challenges I believe more significant. I hope in this way to share with you a general , but by no means complete, panorama of the subject, and some flashes suggesting the very wide range of issues involved. I will be talking from the viewpoint of the user of biometrical and statistical techniques, not as a biometrist.

Perhaps it will be useful to start by defining what we mean by environment and natural resources. Environment, in particular, is a concept that has been plagued by confusion and ambiguity, and even its definition has been subject to international political and ideological debates (since the 1972 Stockholm Conference on the Human Environment).

In its basic (and Western dualistic) sense, the concept on environment is relative to a system and to a level of aggregation.

THE ENVIRONMENT AND THE NATURAL RESOURCES

The **environment of a system** includes everything that is not contained in the system and that affects the functioning of the system or is affected by it. This is a functional definition. As it is easily seen, it is to some degree inherently arbitrary, depending upon what I chose to include within the system and what I leave out (Figure 1).

There is also the issue of defining the level of resolution or the level of aggregation of the system. If we want to discuss the human environment (i.e. the environment of a human system), when we define the system of interest as the whole of mankind, then its environment is composed of the nonliving elements and forces plus all the different species of non-human organisms, interacting in various ways.

However, if we are interested in the environment of an individual person, or the environment of a human group (e.g. a community, or the poor population), immediately the other people, or groups of people, with which it interacts becomes a part of the environment of that system, and social, economic and political factors are legitimately included in the environment, in addition to the physical and biological ones.

Having said that (and thus having protected myself from accusations or biological reductionism) I will concentrate in the rest of my presentation upon the biophysical components of the environment.

There are many possible ways to classify environmental factors according to different criteria (I have collected about a dozen meaningful alternative definitions). I will not unload them upon you; if anybody is interested I could provide a copy of a paper of mine.

Here I want to highlight only **two aspects of the concept of environment**. The first is its relational nature, and the second is that some parts of the environment have direct economic connotations.

• **Relational nature**. A very significant (and often forgotten in practice) fact is that the environmental components interact, not only with the system of interest, but also between themselves; they also have a systemic nature; they are part of ecological systems at various levels. The important point here is that attempts to deal only with individual components of ecological systems as if they were isolated, may often backfire because of multiple reverberations of changes through causal chains within the system.

The second important point is that both human and biological systems(and many non-living components of the environment) are open systems, depending on the active exchanges of matter, energy and information with their environment in order to subsist, at the same time using and dissipating external energy. The Earth itself is such an open system, depending of the solar energy flux for the maintenance and evolution of life, and for the atmospheric and oceanic dynamics. An individual organism is, likewise, an open system.

It is clear that the survival of any living system (including humans) depends on the maintenance of its exchanges with its environment.

• **Economic relevance.**

Some components of the human environment have direct economic significance, in the sense hat they serve as material or non-material inputs to some economic or productive process. Those are then called **natural resources** (Figure 2).

Some are **non-renewable** (they are not replenished, or they are replenished at a negligible rate). Fossil fuels are a clear example. No matter the rate of use by humans, they will be exhausted eventually.

Other natural resources replenish or regenerate themselves. Of those, some are **non-exhaustible**, such as solar and wind energy; no matter haw much is consumed, the source is not affected.

Exhaustible, renewable natural resources are mostly based or depending upon living systems: a fishery, a forest, agricultural land, are common examples. Those natural resources can be used indefinitely provided the rate of use does not exceed their rate of renewal, and their utilization does not destroy the basic renewal mechanisms. Otherwise, those natural resources can be exhausted and destroyed, often irreversibly. This is a real problem in Latin America, the so called "mining" of the renewable natural resources, leading to desertification, soil degradation, etc.

Natural resources are usually valued economically. It should not be forgotten that ecological systems also provide important **services** including essential conditions for life (including human life) that are usually not recognized nor valued. Those include usually functions rather than material components, such as hydrological regulation, waste dilution capacity, recycling of essential nutrients, soil formation, maintenance of atmospheric chemical composition, maintenance of a vast genetic library, pollination of crops, maintenance of aesthetic qualities of the environment, regulation of climate.

One significant distinction between natural resources and ecological services is that resources are usually managed (even if sometimes management actions are limited to extraction) while ecological services are usually not managed (but they still are often impacted by human activities).

The range of sampling and analysis requirements in environmental and natural resources research is very wide, as exemplified in the following four environmental studies each with a different objective[1]:

1. To obtain warning of abiotic environmental deterioration at the site of an effluent by monitoring to detect changes in species composition.

2. To determine the impact effects if any, of existing point-source pollution by

---

[1] Green, R.H. 1979. "Sampling Design and Statistical Methods for Environmental Biologists" John Wiley & Sons, New York.

assessing the spatial pattern of species composition in the adjacent area.

3. To classify a series of habitats on the basis of the abiotic environments and biota (living organisms), for assignement to different categories of multiple use.

4. To determine whether a biological community (a relatively stable assemblage of species in a given area) has changed over time, given a long series of annual species lists.

The four objectives are fairly representative of typical environmental and natural resource research. They share some similarities, but they represent different sampling and analysis problems.

For instance, the spatial pattern of biota in two cases contains part of all the information of interest, while in other case is a source of noise that obscures the required information, and in another is irrelevant (see Table 1).

Therefore, the appropriate distribution of sampling in time and space differs among the four objectives.

Even if the same type of environment were involved in all four cases, the nature of each problem suggests a different approach in sampling design, statistical analysis, sets of hypothesis, and presentation of results.

THREE BASIC APPROACHES

I believe it is useful for the present purposes, to identify three basic approaches to environmental and natural resource research: **descriptive, explanatory/predictive,** and **normative.** This classification is somewhat arbitrary, but is may help to think about the different kinds of challenges and of possible responses. The three approaches are often stages in a sequence in a given research.

Descriptive studies focus on questions such as: which are the constitutive elements of the environmental system? What is its basic structure of relations? How abundant is the resource? How is the system changing?

Statistical approaches and techniques that are usually employed in the description and characterization of environmental and natural resource system include:

- Sampling theory

- A number of multivariable methods for categorization, such as numerical taxonomy, cluster analysis, (using both probabilistic or information theory measures), factor analysis, discriminant analysis, principal components analysis.

- Time-series analysis, spectral analysis.

- Single and multiple linear regression analysis (often used to estimate variables difficult to measure directly)

- Various non-parametric method

- Data transformations of various types,

Those kinds of studies may include:

- The spatial characterization of ecosystem types, land use types, degraded areas, distribution of the abundance of timber, spatial distribution of pollutants, etc.

- Baseline studies, that is, a characterization of the condition of a given ecosystem, landscape or region previous to a human intervention (development project, a dam, a nuclear reactor)

- Assessment of the abundance and temporal variability of a renewable natural resource, such as a fishery stock.

- Assessment of the degree and reality of changes in the value of some variable (e.g. climatic change, biodiversity)

- The determination of the boundaries between different ecosystems.

- The determination of similarities and differences between different ecosystems.

- Estimation of the concentration and spatial distribution of a pollutant

- Characterization of a plant community in terms of presence/absence of species, or in terms of their relative abundance.

Some commons problems with environmental data are:

- The data rarely satisfy the assumptions of the statistical methods commonly thought to non-statistitians. Assumptions of independent and normal error distributions, homogeneity of variance, and additivity of effects are often violated by environmental data (for example, samples of abundances of organisms tend to produce stewed distributions, multiplicative rather than additive effects, and heterogeneity of variances).

- Missing observations for some variables at certain times and places are common.

- The data sets to be used for hypothesis testing may be a mixture of different kinds of variables: nominal, ordinal, cardinal.

- Strongly non-linear relations between different variables often exist.

- In a number of cases, replicates are not available (some ecosystems or situations may be unique, without a comparable situation allowing estimates of within-groups variations).

- Problems with defining the boundaries of the environmental or ecological system (e.g. oceanic ecosystems, definition of the individual in some plant ecology studies).

In a number of cases appropriate solutions can be found, but this often requires a level of statistical expertise not commonly available to the researchers in ecology, the environment, or natural resources.

Appropriate transformations, careful exploratory data analysis, non-parametric methods, simultaneous use of several analysis methods based on different assumptions ("the truth is the intersection of independent lies"), Monte Carlo techniques (e.g.. simulating a number of data set satisfying the null hypothesis but having the same properties that violate the assumptions and using them to test the alternative hypothesis), and residual analysis, are of much use here.

Those who use statistical methods in environmental studies (and certainly not everybody does) tend to fall into one of two extremes regarding the assumptions underlying the methods they use. Either they ignore the fact that there are assumptions at all and apply the methods mechanically or they are paranoid about them and rely only on non-parametric methods, without assessing the likelihood and consequences of violating the assumptions.

Having referred to some of the problems, I want to mention just a couple of interesting challenges in the description of environmental systems.

- A common and important problem arising in ecology, land use, and environmental assessment, is the problem of **inferring temporal sequences from synchronic information** (e.g. current sampling of a set of farm plots abandoned at different times in the past).

- Another challenge is associated with the fast growing field of **GIS**, and the treatment of unordered spatial variability, accuracy and confidence limits of cartography, etc. Important statistical advances have been performed and more can be expected.

- The question of the **continuous or discrete nature of ecological communities and processes**. This is a long controversy in ecology. Holling recently presented analysis indicating that many ecological components and processes are "lumpy"; clustered into non-overlapping groups.

Explanatory/predictive studies concentrate upon understanding the functioning of the environmental system, its causal structure, and the determinants and consequences of change.

Statistical approaches and techniques most commonly used here include:

- Time-series analysis.
- Single and multiple linear regression analysis.
- Non linear regression and curve fitting.
- Parameter estimation of complex models (including simulation models).
- Monte Carlo simulations.
- Correlation analysis.
- Experimental design
- Various non-parametric methods (Correlation and Regression Trees, etc.).
- Path correlation.
- ANOVA, ANCOVA.

Those <u>kinds of studies</u> may include:

- Assessing the potential environmental impact of a new development or human action.
- Predicting the productivity of a tree plantation, or a fishery, under different treatment or environmental conditions.

- Explaining historical changes in a plant community.

- Exploring the behavior of the resource system in conditions never encountered in the past.

- Implementing deterministic of stochastic simulation models of the environmental or resource system for understanding, forecasting or decision-making.

- Performing "what if" simulated experiments that cannot carried out in reality (because of their magnitude, inherent risk, or cost).

Some common problems with those studies (in addition to the ones already mentioned in relation to the descriptive studies) are:

- Controlled experiments are often impossible in environmental systems, particularly at the ecosystem level. This is because the high complexity of those systems, the fact that many variables cannot be controlled, and the fact that variability and heterogeneity are often essential characteristics of ecological systems.

Those factors explain why experimental design and analysis of variance are less relevant to environmental and natural resources research that to agricultural research. Only in a few cases, (particularly in systems with a few populations), in dosage studies, and in laboratory studies, are experiments central.

- Situations or systems states that contains information are often non-repeatable.

- Ecological systems may evolve in time and exhibit sudden changes in behavior. Not only variables and parameter values may change; the system's structure may also change.

- Even the simplest causal relations between variables are strongly non-linear (e.g. the basic ecological equation describing the growth of a single population- the logistic model- is non-linearizable).

- Confounding of variables is often characteristic of spatial as well as temporal relationships. For instance, in streams environmental variables such as current velocity, depth, sediments, particle size, and organic content are strongly intercorrelated. Therefore, descriptive field sampling alone cannot show whether a statistical relationship between a species distribution and, say, sediment type has direct causal basis. Transects across environmental gradients (where many variables change simultaneously) are particularly vulnerable to the confounded variable problem.

In general the above kinds of problems are more difficult to solve than those associated with descriptive studies. For instance, finding the parameter values that minimize a residual sum of squares in the case of strongly non-linear models must often be done iterativelly and furthermore there is no guarantee that the calculated values represent a global minimum; many tests of hypothesis become invalid or approximate; confidence limits are asymmetric and often impossible to calculate; *ad hoc* methods may be necessary.

Nonetheless, some solutions to some of those problems exist, but are often unfamiliar even to many statisticians. Monte Carlo simulation, nonlinear curve fitting and parameter estimation methods, Bayesian inference, and robust estimation. Some of the problems such as the non-repeatability lack any satisfactory statistical solution. The problem of confounding of variables cam be solved through experimentation, but often this is not possible.

Perhaps the most important limitation, from the viewpoint of the researcher in environmental and natural resources problems, is that there is no standard set of methods and procedures for dealing with those kinds of problems, and often a highly sophisticated statistical knowledge is required.

As examples of some interesting challenges derived from this area, I will mention:

- The problem of complex model validation. Simulation models are often used in environmental and resource management research. They may contain many hypothesis in the form of functional relationships between variables, as many as hundreds of them; in some cases even thousands. Up to date, no satisfactory ways of validating those models exist. One usual way of "validating" simulation models is to show that the behavior of the model reproduces historical data. Obviously, a believable model should predict or mimic qualitative properties of the temporal and spatial pattern characteristic of the historical behavior of the system. But practically any complex model can be "tuned" to fit practically any given pattern of historical data. There are some ways to mitigate this problem (such as calibrating the model with a subset of the historical data, and using the model to predict the other subset) but they are not very satisfactory. Another problem is that, even if each individual relationship was rigorously tested for significance at, say, $\alpha=0.05$, the likelihood that one or more of the hypothesis is false increases with the number of hypothesis (equations) in the model. Thus any complex simulation model is bound to contain incorrect hypothesis.

A third problem is that in environmental and natural resources research, models are usually built in order to anticipate the behavior of the system under new conditions often never encountered in the past (this is typically the case in environmental impact assessment) and therefore historical data may not contain the relevant information. Another problem is error transmission in complex simulation models. Finally, and possibly the most difficult challenge, is to define methods that allow the assessment of the goodness of the model's structure of variables and relations. A promising approach is Klir's General Systems Problem Solver[2], but so far little statistical analysis has been associated with it. Regarding structure, it is often assumed that the more detailed the model is, the better is its descriptive and predictive power. However, this is usually not the case. Constanza and Sklar[3] reviewed 87 mathematical models for freshwater wetlands and shallow water bodies. For about half they could measure descriptive accuracy in terms of the correlations between model predictions and historical data. They measured model complexity in terms of an "articulation index" based on the number of model components, temporal resolution of predictions and spatial resolution. The overall model effectiveness was defined as the product of accuracy and articulation index. They found that effectiveness thus defined was a dome-shaped function of model complexity: it is possible to predict a few things very well, or a lot of things very poorly; between those extremes there is a balance where reasonable accuracy is maintained without too much loss of biological "realism" (Figure 3).

- A second conceptual and methodological challenge arising from environmental research gives rise to a non-traditional way of interpreting experiments. It has been said before that too often experiments cannot be performed in ecosystems; in some other cases there is a scale problem: one cannot make inferences about the environmental impact of a large dam by using small experimental reservoirs; they are simply different types of ecosystems, with different organization and behavior. However, sometimes useful natural trials exist wherever there are examples of the case being modelled but that exhibit different behavior or has been exposed to similar kinds of disturbance. One important challenge is to derive methods to use those "natural experiments" in a rigorous way.

Another associated challenge is how to use the information being generated by the human intervention or development project that is being assessed. Often in environmental assessment, the development project itself (e.g. building a dam) is the only possible experiment. But the information value of the project as shedding further light on the understanding of the environmental system and in designing adaptive management activities (perhaps using information generated by the early stages of the project to modify the later stages) is usually not considered. This is very valuable information, and very expensive if the cost of the project is taken into account.

This looks to me as an interesting area for statistical inference research and a different way of viewing experimental design.

---

[2] Klir, G.J. 1985. "Architecture of Systems Problem Solving". Plenum Press, New York.

[3] Constanza, R. and Sjlar, F.H. 1983. "Mathematical Models of Freshwater Wetland and Shallow Water Ecosystems: An Articulated Review". Paper prepared for SCOPE Int. Conf. on Freshwater Wetlands and Shallow Water Bodies, Talinn, USSR, August 1983.

Normative Research in environmental studies concentrates on value-ladden issues such as which is the optimal rate of harvest or utilization of the resource, what is its sustainable use, what is the proper way to value the resources and environmental functions, the assessement of environmental quality, and so on.

A fundamental distinction must be made between assessing the **state or condition** of an environmental variable or resource, which can be done rather objectively, and assessing its **quality or value.** The latter always embody a value judgement, either explicit or implicit. This is clearly illustrated by two alternative ways that have been used to assess the impact of atmospheric pollution on humans: one is by estimating the losses in terms of decreasing life expectancy, and the other by estimating the losses to the economy of the reduced working hours due to the sickness leave of workers.

Statistical approaches are less used here in general. In the case of highly engineered natural resource systems, such as annual crops, agroforestry, tree plantations, experimental design to determine the sustainable yields and the maximum sustainable yields are often employed. Sampling techniques for renewable resources such as fisheries or forests are highly developed and sometimes very sophisticated.

Time-series analysis are also employed, as well as deterministic or stochastic simulation models.

However, this is more the province of statistical-decision theorists, and econometrics concepts such as "utility functions", "risk-aversion", economic efficiency, inter-temporal valuation (discounting in time), option value, are relevant here.

As this is not within the central focus of this Symposium, I would not develop it further.

SUSTAINABILITY: A CENTRAL CHALLENGE

I would like to end my presentation by emphasizing that all approaches converge together in the issue of **sustainability** (even when limited only to the biophysical environmental sustainable of human activities).

Descriptive, explanatory/predictive, and normative research are needed to assess sustainability and to understand the roots of sustainability. While there is a long tradition in ecology of estimating sustainable yields and maximum sustainable yields of single populations, or systems composed by few populations, long term sustainability of complex environmental and natural resource systems (and of socio-ecological systems) requires a level of understanding that has not yet been reached.

Criteria to define when and whether a human activity is environmentally sustainable need much more development. Currently, it is easier to demonstrate that some activity is non-sustainable, then the opposite.

An important point is the distinction between sustainability in time of an **output of a system** (eg. agricultural yield, fish catch) and sustainability of the **system itself** (eg. farming system, productive ecosystem).

Another important point often neglected, is that in the current changing world sustainability does not imply necessarily the capacity to produce cucumbers, or rice, forever without diminishing returns. It must imply the capacity to change, and to cope with change, in an adaptive and proactive way.

This means that research should concentrate on identifying and nurturing the basic **sources of renewal and regeneration** in the environmental or natural resource system, rather than simply focus on the constancy of the output.

Sustainability analysis and prediction is probably one of the most significant current challenges in environmental and natural resources research, and also for statistical research in those areas.

Sustainability is certainly also one of the most significant social and cultural challenges faced by mankind today.

Table 1. Sampling and analysis properties for four environmental studies objectives

Objective	Study 1: Warning of environmental deterioration	Study 2: Pattern and point-source pollution	Study 3: Habitat classification	Study 4: Community change from species list
Spatial biological pattern	Noise	Information	Information	Not relevant
Abiotic environment spatial pattern	Noise	Noise	Information	Not relevant
Pollutant	To be detected	Exists	May exist along with other environmental factors	Perhaps relevant to subsequent evaluation
Distribution of samples in space	Experimental and control station, each with replication	Many stations on a grid, each with replication	In different habitats, preferably with replication	Not relevant
Temporal biological change	Information	Noise	Noise	Information
Abiotic environmenta l temporal change	Noise	Noise	Noise	Information
Distribution of samples over time	Equal intervals, to continue indefinitely	At least two different times	Preferably two different times	Long series, preferably at equal intervals

Source: modified from Green, R.H. 1979. "Sampling Design and Statistical Methods for Environmental Biologists" John Wiley & Sons, New York (page 3).

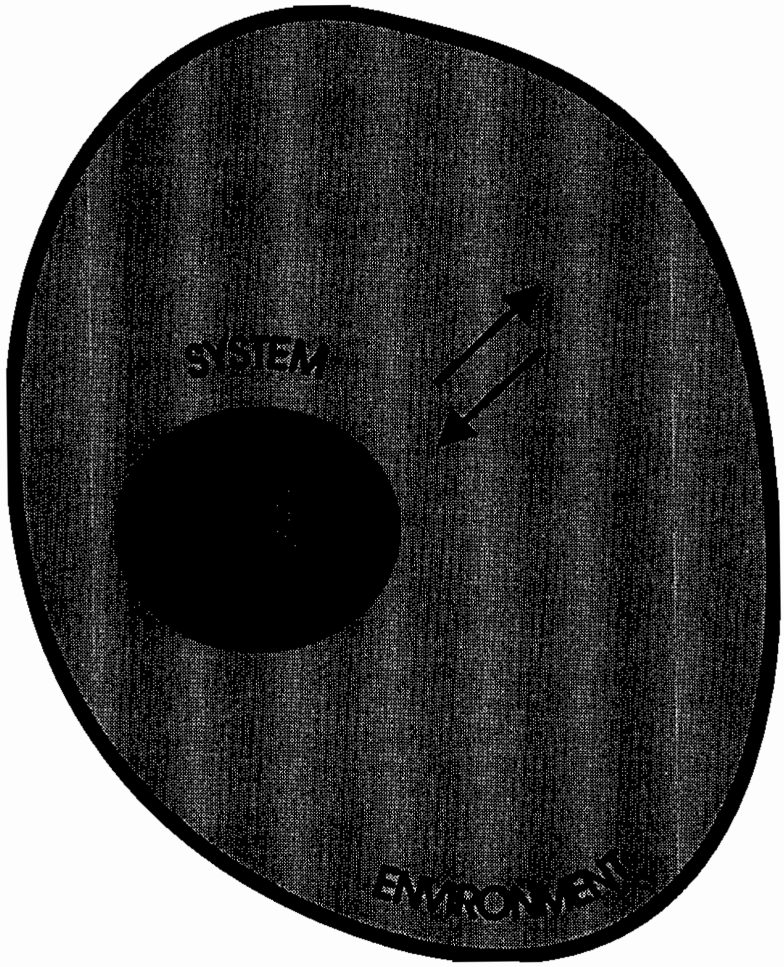

Figure 1. The functional representation of the environment of a system.

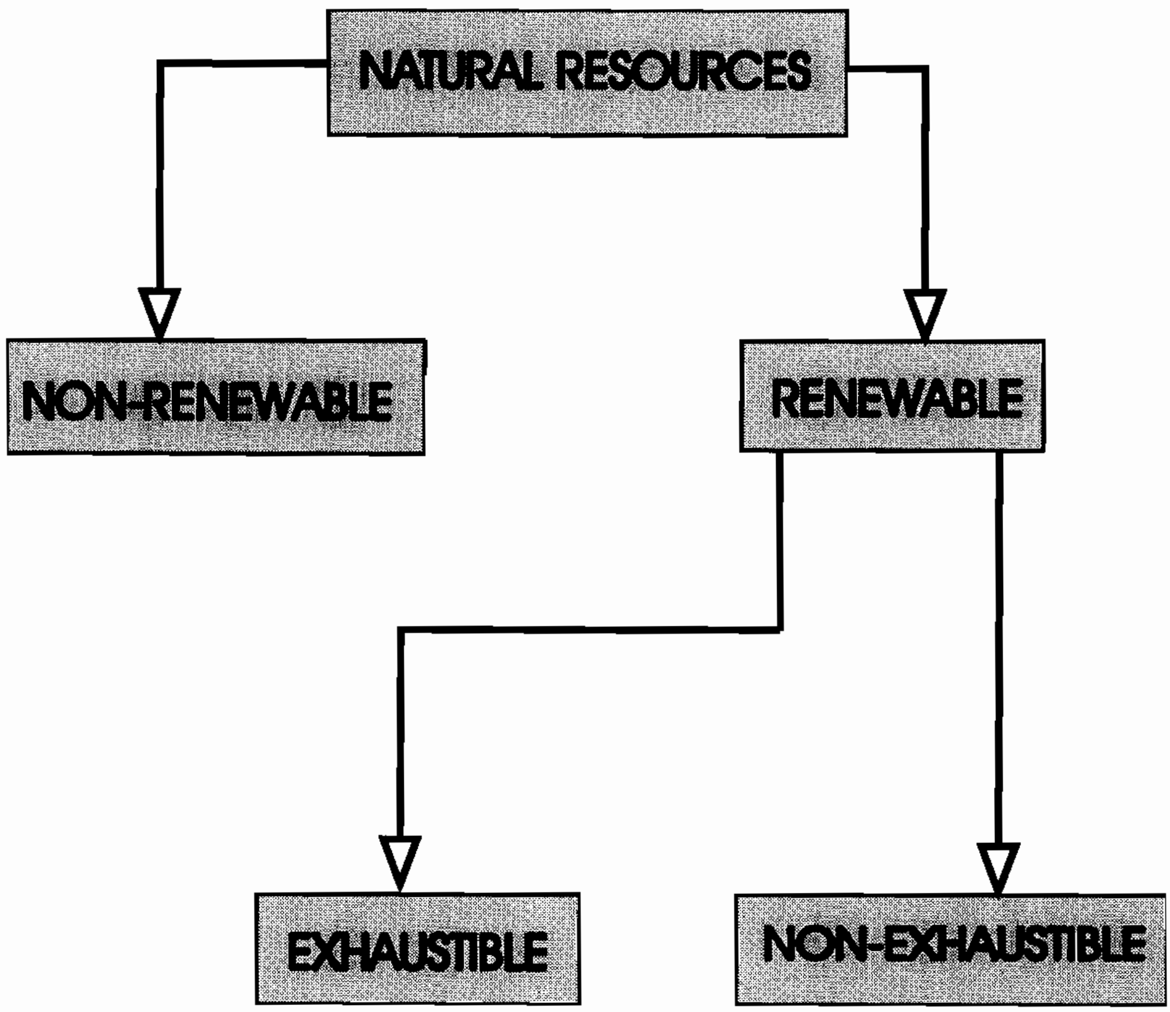

Figure 2. A classification of natural resources.

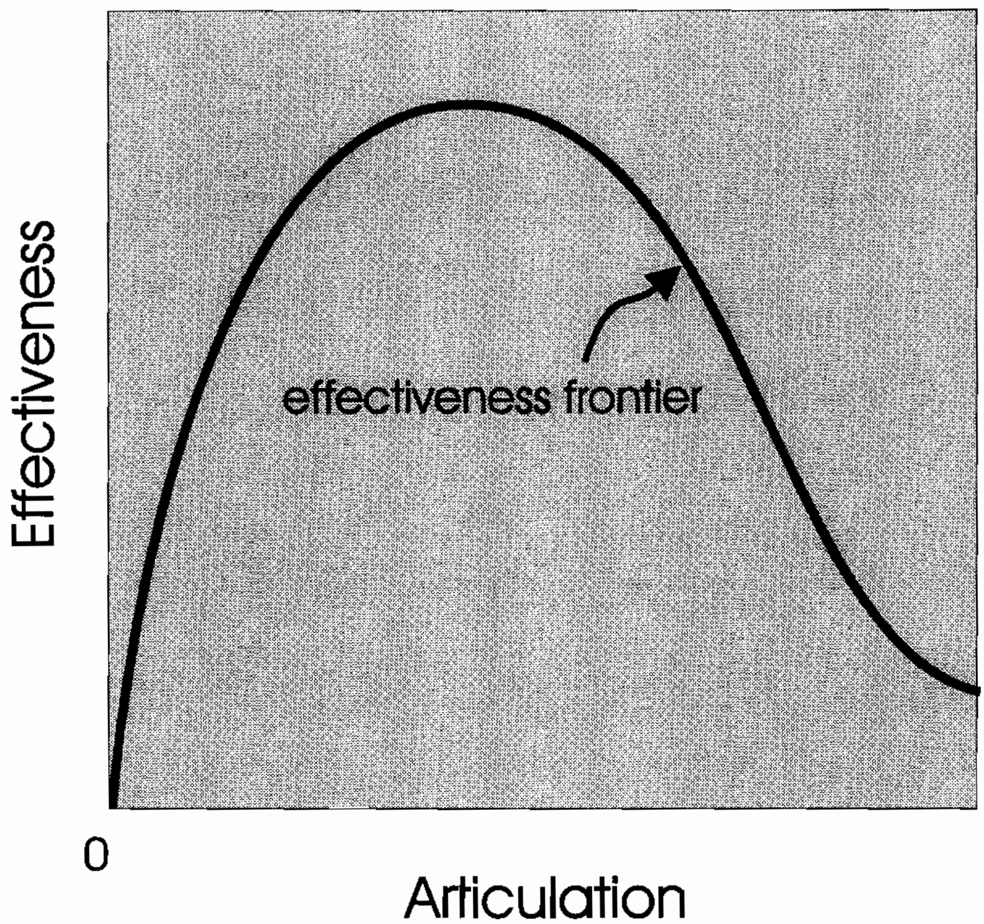

Figure 3. Predictive capacity of models as a function of model complexity (articulation). Source: Constanza, R. and Sjlar, F.H. 1983. "Mathematical Models of Freshwater Wetland and Shallow Water Ecosystems: An Articulated Review". Paper prepared for SCOPE Int. Conf. on Freshwater Wetlands and Shallow Water Bodies, Talinn, USSR, August 1983.