5.10   Selecting indicators

Contents

Introduction

While technical quality is a vital component of an overall environmental sampling program the core issue should always be quality of decisions. Assessment methods need to be considered in the context of the specific program.

This document is based on the following assumptions:

- *A priori* there are no core indicators

- Indicators must be chosen on the basis of the hypotheses of the program

- The "best" method is not necessarily the appropriate method (ie. The "gold standard" is a point of comparison not a given)

- Field and analytical methods must be comparable in terms of precision

- Analysis cannot compensate for poor experimental design

- Design, sampling, analysis, and reporting need to be integrated

- Quality control is a necessary component of any sampling program.

In this document we do not include discussion on the use and management of recreational waters as this issue is undergoing renewed discussion and is covered comprehensively elsewhere. In general the design and implementation of such programs are more focused and prescribed. Useful sources include:

- ANZECC Water Quality Guidelines Chapter 5
- The "Annapolis Protocol" which discusses alternative strategies for monitoring and reporting recreation quality
- The US EPA's manual for management of RECREATIONAL WATERS

Guidelines for selecting and using indices.

1. Determine the goals of the experiment.

2. Define, as clearly as possible, the abstract concept that is to be assessed (using conceptual and other models).

3. Keeping in mind the goals of the experiment, decide what aspect(s) of the underlying concept are important in terms of the management or scientific objectives.

4. Identify indices that have been developed to quantify this aspect(s).

5. Assess the sensitivity of the indices.

6. Examine the range and scale of the indices. Is it easy to tell whether an observed value represents a desirable level, or whether an observed change in the index represents an important change in the achievement of management objectives (the "effect size" of interest)?

7. Ensure that the bias, precision and standard error of all selected indices are well understood and predictable in relation to the effect size of interest. Each index should not be overly sensitive to small measurement errors. Large, unquantifiable biases are particularly troublesome and can easily arise if the index depends on complex mathematical functions of the measured values, but can also arise in simple indices, such as the species richness.

Selecting Environmental indicators.

The U.S. Environmental Protection Agency (USEPA) and the U.S. Geological Survey (USGS developed the following definition of "environmental indicator" as a " ... measurable feature or features that provide managerially and scientifically useful evidence of environmental and ecosystem quality or reliable evidence of trends in quality." Thus, environmental indicators must be:

- measurable with available technology,
- scientifically valid for assessing or documenting ecosystem quality, and

- useful for providing information for management decision making.

Environmental indicators encompass a broad suite of measures that include tools for assessment of chemical, physical, and biological conditions and processes at several levels. These characteristics of environmental indicators have helped define the scope of the group activities.

Early detection of acute and chronic changes (ANZECC 8.1.1.2)

"An early warning indicator can be described as a measurable biological, physical or chemical response in relation to a particular stress, prior to significant adverse affects occurring on the system of interest. The underlying concept of early warning indicators is that effects can be detected, which are in effect, precursors to, or indicate the onset of, actual environmental impacts. Such 'early warning' then provides an opportunity to implement management decisions before serious environmental harm occurs.

"Ideal attributes of early warning indicators have been extensively discussed elsewhere (Cairns & van der Schalie 1980, Cairns et al 1993, McCormick & Cairns 1994), and have been summarised in a modified form by van Dam et al (in press), as presented below. To have potential as an early warning indicator, a particular response should be:

1. *anticipatory*: should occur at levels of organisation, either biological or physical, that provide an indication of degradation, or some form of adverse effect, before significant 'serious' sounds a bit emotive environmental harm has occurred;

2. *sensitive*: in detecting potential significant impacts prior to them occurring, an early warning indicator should be sensitive to low levels, or early stages of the stressor;

3. *diagnostic*: should be sufficiently specific to a stressor, or group of stressors, to increase confidence in identifying the cause of an effect;

4. *broadly applicable*: alternatively, an early warning indicator should predict potential impacts from a broad range of stressors;

5. *correlated to actual environmental effects*: knowledge that continued exposure to the stressor, and hence continued manifestation of the response, would eventually lead to significant environmental effects is important;

6. *timely and cost-effective*: should provide information quickly enough to initiate effective management action prior to significant environmental impacts occurring, and be inexpensive to measure while providing the maximum amount of information per unit effort;

7. *regionally and socially relevant*: should be relevant to the ecosystem being assessed, and of obvious value to, and observable by stakeholders, or predictive of a measure that is;

8. *easy to measure*: should be able to be measured using a standard procedure with known reliability and low measurement error;

9. *constant in space and time*: should be capable of detecting small changes, and clearly distinguishing that a response is caused by some anthropogenic source, not by natural factors as part of the natural background (ie high signal : noise ratio);

10. *nondestructive*: measurement of the indicator should be nondestructive to the ecosystem being assessed.

"The importance of the above attributes cannot be over-emphasised, as any assessment of actual or potential environmental degradation will only be as effective as the indicators chosen to assess it (Cairns et al 1993). However, it should be stressed that it is impossible for an early warning indicator to possess all the above attributes. In many cases some of them will conflict, or not be achievable. For example, a biochemical biomarker might provide an excellent indication of exposure to a particular pollutant, but might not be correlated to effects at higher levels of biological organisation (e.g. ecosystems). Moreover, the biomarker may not be applicable to other pollutants. Similarly, a long term monitoring program might provide excellent baseline data from which small perturbations will be obvious, but may be neither time- nor cost-efficient. Subsequently, not all the attributes will be achievable for each indicator. Therefore, decisions are required as to which attributes are more appropriate and achievable for a particular purpose, and indicators chosen based on those attributes. Particular attributes can be prioritised, and this is further discussed below".

Standard Selection Criteria.

Environmental indicators should be able to satisfy predetermined selection criteria to ensure their viability. These criteria provide a series of guidelines that shape the decision making process, which results in an indicator that meets the needs of the program. It is important to put the selection criteria into a standardized format that can be useful for nationwide programs. Standardization of the selection criteria streamlines the indicator selection process, reduces costs, prevents duplication of effort, and provides a consistency, thereby increasing the potential for cross-program comparisons.

Criteria Categories

Scientific validity is the foundation for determining whether data can be compared with reference conditions or other sites. Data collected from a sampling site become irrelevant if they cannot be easily compared with conditions found at a site determined to be minimally impaired. Factors must be balanced when considering the scientific validity of an indicator and its application in real-world situations. An indicator must not only be scientifically valid, but its application must be practical (that is, not too costly or too technically complex) when placed within the constraints of a monitoring program. Of primary importance is that the indicator must be able to address the questions that the program seeks to answer.

For discussion purposes, these criteria have been divided into three categories--scientific validity (technical considerations) practical considerations, and programmatic considerations. Although discussed separately, these categories are not entirely separate entities, but rather portions of characteristics that provide some guidance in the indicator-selection process.

Table 2. Relations among measurable water-quality characteristics and selected water-quality issues
[●, primary characteristics for each issue; ○, secondary characteristics useful for evaluating each issue]

Water-quality characteristics	Pathogens	Nutrient enrichment	Trace elements	Pesticides	Industrial organics	Suspended sediment	Salinity	Temperature	Acidity	Dissolved oxygen	Ecological condition
Field:											
Streamflow	○	○	○	○	○	○	○	○	○	○	○
Dissolved Oxygen		○	○	○	○			○		●	○
pH and alkalinity		○	○	○	○		○		●	○	○
Specific conductance		○	○	○	○		●				○
Temperature	○	○	○	○	○			●		○	○
Habitat characteristics		○	○	○	○	○	○	○	○	○	●
Biological communities		○	○	○	○	○	○	○	○	○	●
Laboratory:											
Suspended Sediment	○	○	○	○	○	●					○
Major constituents											
Dissolved solids		○	○			●			○		○
Major ions and metals		○	○				○		○		○
Nutrients		●					○		○	○	○
Organic carbon		○	○	○	○				○	○	○
Trace elements			●								
Organic contaminants											
Dissolved pesticides				●			○				○
Hydrophobic pesticides				●		○					○
Polychlorinated biphenyls					●	○					○
Other semivolatile organic contaminants					●						○
Volatile organic contaminants				●	●						○
Biological											
Bacteria	●										○
Chlorophyll		○							○	○	
Species identification											●

Scientific Validity

As with any monitoring or bioassessment program, the data collected must be scientifically valid for it to be useful.

Measurements of environmental indicators should produce data that are valid and quantitative or qualitative and allow for comparisons on temporal and spatial levels. This is particularly important for comparisons with the reference condition. Interpretation of measurements must accurately discern between natural variability and the effects induced by anthropogenic stressors. This requires a level of sensitivity and resolution sufficient to detect ecological perturbations and to indicate not only the presence of a problem, but to provide early warning signs of an impending impact. The methodology should be reproducible and provide the same level of sensitivity regardless of geographic location. It also should have a wide geographic range of application and a set of reference-condition data that can be used for comparisons.

Practical Considerations

The success of a sampling program is dependent on the ability to collect consistent data over the long term; consistency is directly related to the practical application of the prescribed methodologies. The practical considerations include monitoring costs, availability of experienced

personnel, the practical application of the technology, and the environmental impacts caused as a result of monitoring.

A cost-effective procedure should supply a large amount of information in comparison to cost and effort. Of significant importance is the acknowledgment that not every quantitative characteristic needs to be measured unless it is required to answer the specific questions. It may be more important to have a range of qualitative and quantitative data from a large number of sites than it is to have a small number of quantitative parameter measurements from a small number of sites. Cost effectiveness may be dependent on the availability of experienced personnel and the ability to find or detect the indicating parameters at all locations. State-of-the-art technology is useless in a biomonitoring program if experienced personnel are in short supply or the data cannot be collected at all the stations. Equally important is the ability to collect the data with limited impact to the environment. Some collection procedures (for example, using rotenone to collect fish) are very effective, but minor miscalculations can cause significant environmental damage. These methodologies should be replaced with less destructive procedures.

Programmatic Considerations

Stated objectives of a program are an important factor in selecting indicators. Sampling and analysis programs should be structured around questions to be addressed. The term "programmatic considerations" simply means that the program should be evaluated to confirm that the original objectives will be met once the data have come together. If the design and the data being produced by a program do not meet the original objective(s) within the context of scientific validity and resource availability, then the selected indicators and uncertainty specifications should be re-evaluated.

Another important consideration is the ease with which the information obtained can be communicated to the public. Although it is essential to present information for decision makers, scientists, or other specialized audiences, information for the general public needs to be responsive to public interests and summarized for clarity.

Potential errors.

Guidelines for developing a measurement protocol

Following are recommended guidelines for improving the quality and value of measurements. As with any general guidelines, these will need to be adapted to specific applications, and are intended more as an initial checklist of important items to consider than as an inflexible set of rules.

1. Determine what parameter needs to be measured. In so doing, pay attention both to the relevance of the quantity being measured and the practical difficulties in obtaining reliable measurements or estimates. Consider both the concept and its measurement.

2. Determine the demands that need to be placed on the quality of the resulting measurements or estimates.

3. Devise a measurement system that will meet these demands. If this task is impossible, reassess the management objectives and strategies and revisit guidelines 1 and 2.

4. Assess the accuracy of the proposed measurement system by taking repeated measurements of known quantities under a variety of conditions.

5. Establish an unambiguous protocol for taking measurements, and ensure its proper implementation.

6. Implement a system of periodic checks on the continuing performance of the measurement system in light of internal changes and external demands. Watch for subtle increases in the demands placed on the system.

7. Look for ways to develop incremental improvements while maintaining the integrity of any long-range data series. When implementing changes, phase them in, running new and old methods in parallel during a transition period.

The sources of error generation are discussed above.

Reducing Measurement Errors

Although some measurement errors are inevitable, they can often be reduced substantially. The resulting benefits to the experiment can be considerable. Following are some guidelines for achieving this goal.

Guidelines for reducing measurement errors

1.   Counting

•   Ensure that experienced workers train new personnel.

•   Where feasible, mark or discard items previously counted to reduce double counting.

•   Anticipate undercounting. Try to assess its extent by taking counts of populations of known size.

•   Try to reduce errors by taking counts only in favourable conditions and by implementing a rigorous protocol.

2.   Physical measurements

•   Instruments should be calibrated before first use, and periodically thereafter.

•   Personnel should be trained in the use of all measuring devices.

•   Experienced personnel, as part of an overall quality control program, should spot-check measurements, particularly those taken by new personnel.

- Incorporate new equipment where appropriate (e.g., lasers and ultrasound, for distance measurements).

3. Remeasurement

- Watch for the transfer of errors from previous measurements (e.g., a mistaken birth from an item erroneously marked as dead).

- Reduce errors in relocating the site of previous measurements through more careful marking, use of modern electronic GPS technology, etc.

- Ensure that bias is not propagated through the use of previous measurements as guides to subsequent ones. (This issue is particularly troublesome in subjective estimates.)

4. Visual estimates

- Ensure that all visual estimates are conducted according to rigorous protocols by well-trained observers.

- Pay particular attention to observer bias. When bringing a new observer into the program, ensure that an experienced observer(s) backs up his/her results.

- If sites or times are to be selected as part of the collection of visual estimates, eliminate selection bias by providing a protocol for site- or time-selection. Do not, for example, let vegetation samplers pick modal sites.

5. Data handling

- Record data directly into electronic form where possible.

- Back up all data frequently.

- Use electronic data screening programs to search for aberrant measurements that might be due to a data handling error.

- Design any manual data-recording forms and electronic data-entry interfaces to minimize data-entry errors. In the forms, include a field for comments, encourage its use, and ensure that the comments are not lost or ignored.

Summary

"A century ago, the World's renewable resources seemed so limitless that we asked very little of our measurements of the resources and their support systems. With new requirements imposed by legislation, increased community expectations and with increased harvesting capacity, we are increasing our demands on measurement systems. [The recent controversy over the licensing of saltwater recreational fishermen may be less about a breakdown in the quality of the measurement procedures, as about the fact that our management expectations have increased beyond the capacity of the measurement system.] Assess continually the adequacy of a

measurement system to improve it where possible and to point out when its limitations may be exceeded. The assessment should include:

- the choice of quantities to be measured;

- the procedures and equipment for taking the measurements;

- any associated sampling;

- the processing, storage, and analysis of the data; and

- the demands and expectations of the resource managers or others who use the results.

"Good quality control procedures are an essential component of any measurement process.

This applies even if the measurements are to be used solely for indicating trends. A well-designed protocol must be followed, and the reliability of the data must be commensurate with management needs. Just as expectations can rise without notice, so can a gradual deterioration in quality go undetected.

"Consider, for example, the monitoring of spawning habitat in a remote lake. Workers faced with a rising wind on the long stretch of water back to the landing would be tempted to cut corners if they believed that no one valued their work or would ever check on their accuracy. For field measurements, often taken in isolated conditions, quality checks will inevitably be infrequent but should not be ignored.

"The quality of the data will depend critically on the reliability and commitment of the field staff; sound, quality management practices will foster the required spirit.

The myths of sampling should now have been dispelled:

> *It is a waste of time to worry about measurement errors. I have enough practice in my field to have reduced measurement errors to a negligible size.*

> Measurement errors in many field studies are large, and inadequate attention to them has led to major management disasters.

> *If I know that my measurements are not perfect, then I should take several, and average them, maybe throwing out the odd one that is far from the others.*

> Taking repeated measurements allows the researcher to assess the average size of the chance errors. Averaging these measurements will usually reduce the impact of the chance errors. However, aberrant measurements should be singled out for special attention, not casually or routinely discarded. They could provide valuable insight, and are an important part of the information collected. In addition, averaging will not reduce any systematic bias.

> *I have the resources only to make a subjective guess at the abundance of some minor species. Surely this will be adequate. After all, I am only looking for trends. If the measurement errors are large, and are consistently present, can't we ignore them when we are looking for trends?*

> We need to know enough about the errors to be able to distinguish between a trend in the quantity being measured and in the measurement errors. Furthermore, a false estimate of the historical state of the forests could lead to inappropriate management actions. Trend indicators are often set up when it seems too difficult or costly to implement the rigorous procedures required to produce unbiased abundance estimates. Trend indicators demand

almost as much rigour, and measurement procedures must be rigorous enough to rule out any cause for a trend other than a change in abundance.

*I don't have to worry about measurement errors. I always take repeated observations and use standard statistical techniques to deal with them. If my measurements do contain large chance errors, then can't I just take repeated measurements, do a routine statistical analysis, and quote a p-value to silence the pesky biometrician*s?

The standard statistical analysis procedures require specific assumptions about the measurement errors. Violated assumptions lead to questionable analyses and management decisions. In addition you need to be concerned about randomization, pseudoreplication and the power of the tests so a "p-value" is less than adequate to control for measurement errors.

I have an important job to do. I don't have the time or luxury of worrying about statistical niceties like academics and scientists. I need to get on with managing.

Thorough attention to measurement errors and other "statistical niceties" will help, not hinder, the ongoing development of improved management strategies. An understanding of the principles of statistics and the relevant analyses will also prevent wasted effort and invalid conclusions being drawn from the data.