Too Little Information: The Fate of Biological Data (Newsletter)*

April 17, 2011

It is widely accepted that raw data need to be converted to information (commonly by statistical analyses) and the results interpreted to form knowledge before informed decisions can be made. With biological data this process is not followed as frequently as it should. Modern spatial analyses and statistical models can provide valuable and useful information that is otherwise lost.

Biological data are counts (or presence/absence), proportions, and frequencies. They are not continuous variables with a true zero so the familiar parametric statistics cannot be used. Sometimes the variable of interest is an index such as species richness. Plant and animal data are dependent upon location and time, including the time of day.

Three of the questions that collection and analysis of biologic data can answer are: 1) Why is this plant or animal species (or community of species) found at this location and time? 2) Where else is this species (or community) located? 3) Why is this species (or community) not in this location that seems similar to where we find it?

Answering these questions requires collection of known or likely explanatory variables when the biotic data are gathered. Topographic, geologic, chemical, hydrologic, basin and stream network topology are categories of known or likely variables that influence the biota observed at any location and time. In addition to modeling the interactions of these explanatory and response variables in a graphical spatial analysis system, statistical models have been developed to perform linear and non-linear regressions on categorical and nominal data mixed with continuous data. Bayesian inference techniques allow robust population estimates from mark and recapture studies of animals.

These analytical techniques provide knowledge and insights not otherwise available to operators, regulators, and other decision-makers. They can be applied to evaluate grazing, logging, mining, energy production (hydroelectric, geothermal, wind) and distribution (electrical transmission lines, oil and gas pipelines), agricultural practices, and reclamation/restoration programs.

Three examples briefly described illustrate the range of possibilities for policy and operational decisions to be better based on science.

Wolves have been re-introduced into the Pacific northwest. They are found around grazing cattle, particularly calves, and around deer and elk herds. But, are they evenly distributed or clumped? If the latter, what environmental conditions can explain their distribution?

Salmonid recovery efforts in the Columbia and Salmon Rivers assume that if dams are removed and other blocked habitats opened populations of the five will become established. Will conditions in these areas be sufficiently similar to river reaches with established populations to anticipate re-population? Similarly, if a population of resident fish is present in a stream with known water chemistry, why are they not present in a nearby drainage with similar water chemistry?

Cheatgrass has become well established across western rangelands, from valley floors to higher elevations. Yet the coverage is not uniform and may change over time. What explanatory variables are associated with areas of high (or low) cheatgrass coverage? If areal coverage declines (rather than expanding as it has in the past), what conditions might explain this distributional change?

The above do not preclude applying these analytical and statistical models to baseline data for environmental impact assessments or demonstrating compliance with environmental permit conditions. Instead of relying on "best professional judgment", diversity indices, and similar approaches to explaining biotic data use of robust spatial analyses and statistical models result in interpretations that are technically sound and legally defensible.